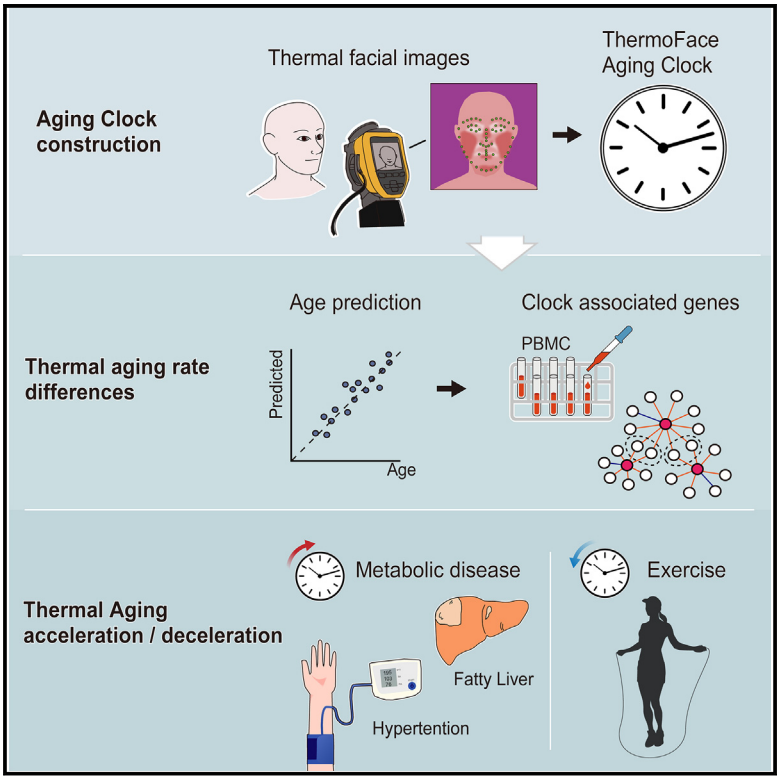
Cell Metabolism
Clinical and Translational Report

Thermal facial image analyses reveal quantitative hallmarks of aging and metabolic diseases

Graphical abstract



Highlights

- Collected >2,800 thermal facial images and developed ThermoFace aging clock models
- ThermoFace is related to metabolic indicators, DNA repair, ATPase, and lipolysis
- ThermoFace aging rate increases with metabolic diseases and decreases with adequate sleep
- ThermoFace can rapidly assess aging and metabolic diseases and intervention

Authors

Zhengqing Yu, Yong Zhou, Kehang Mao, ..., Hongxiao Liu, Yi Wang, Jing-Dong J. Han

Correspondence

jackie.han@pku.edu.cn

In brief

By collecting and analyzing thermal facial images, Yu et al. find that human facial temperature distribution contains information on aging and metabolic states and has potential for rapid screening of aging and metabolic disorders and for revealing intervention strategies.

Yu et al., 2024, Cell Metabolism *36*, 1482–1493
July 2, 2024
https://doi.org/10.1016/j.cmet.2024.05.012

CellPress

CellPress

Cell Metabolism

Clinical and Translational Report

Thermal facial image analyses reveal quantitative hallmarks of aging and metabolic diseases

Zhengqing Yu,[1,8] Yong Zhou,[2,8] Kehang Mao,[1,8] Bo Pang,[3,8] Kai Wang,[5] Tang Jin,[5] Haonan Zheng,[1] Haotian Zhai,[1] Yiyang Wang,[1] Xiaohan Xu,[4] Hongxiao Liu,[4] Yi Wang,[6] and Jing-Dong J. Han[1,5,7,9,*]

[1]Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Center for Quantitative Biology (CQB), Peking University, Beijing, China

[2]Clinical Research Institute, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[3]Clinical Laboratory, Guang'anmen Hospital, China Academy of Chinese Medical Sciences, Beijing, China

[4]Department of Rheumatology, Guang'anmen Hospital, China Academy of Chinese Medical Sciences, Beijing, China

[5]International Center for Aging and Cancer (ICAC), Hainan Medical University, Haikou, China

[6]Kailuan Majiagou Hospital, Tangshan, Hebei Province, China

[7]Peking University Chengdu Academy for Advanced Interdisciplinary Biotechnologies, Chengdu, China

[8]These authors contributed equally

[9]Lead contact

*Correspondence: jackie.han@pku.edu.cn

https://doi.org/10.1016/j.cmet.2024.05.012

SUMMARY

Although human core body temperature is known to decrease with age, the age dependency of facial temperature and its potential to indicate aging rate or aging-related diseases remains uncertain. Here, we collected thermal facial images of 2,811 Han Chinese individuals 20–90 years old, developed the ThermoFace method to automatically process and analyze images, and then generated thermal age and disease prediction models. The ThermoFace deep learning model for thermal facial age has a mean absolute deviation of about 5 years in cross-validation and 5.18 years in an independent cohort. The difference between predicted and chronological age is highly associated with metabolic parameters, sleep time, and gene expression pathways like DNA repair, lipolysis, and ATPase in the blood transcriptome, and it is modifiable by exercise. Consistently, ThermoFace disease predictors forecast metabolic diseases like fatty liver with high accuracy (AUC > 0.80), with predicted disease probability correlated with metabolic parameters.

INTRODUCTION

Aging is a major risk factor for many complex human diseases.[1,2] The rate of biological aging is heterogeneous across different individuals and, thus, many aging clocks have been developed to monitor the aging rate at various levels, from molecular to phenotypical.[3–7] Due to the complicated sample preparation and data acquisition steps needed for most molecular clocks, it is hard to provide real-time or on-the-fly aging rate assessment for an individual. How to obtain data rapidly and inexpensively, derive a clock accurately based on large cohorts, and make the clocks applicable to general populations is still a challenge.

Body temperature, an often overlooked but important factor in aging and longevity, is a key homeostatic parameter that affects cell function and organism survival and is closely related to metabolism, protein denaturation, membrane fluidity, ion fluxes, and enzyme properties.[8] In many endothermic species, individuals with lower body temperatures live longer and age more slowly than those with higher body temperatures.[9] Cold temperature increases lifespan and delays aging in invertebrate organisms.[10–12] Transgenic mice with reduced core body temperature have extended lifespans relative to normal body temperature mice.[13] The decrease in body temperature under lifespan-extending calorie restriction (CR) is associated with metabolic rate.[14] Metabolic rate is regulated at the organ (e.g., cardiovascular regulation rate) and cellular (e.g., mitochondrial capacity) level and is always influenced by ambient temperature.[15] Interestingly, human core body temperature tends to decrease with age.[16] High body temperature often implies high metabolic rates[17,18] and can be induced by psychological and metabolic stress.[19] For example, individuals with type 1 diabetes have been shown to have an impaired ability to dissipate heat, with impaired thermoregulation,[20] and in acute human stroke there is a positive association between body temperature and initial stroke severity, infarct size, mortality, and outcome.[21] Infrared temperature images of the human body have been tested for screening cardiovascular diseases (CVDs), with the best goodness of fit (adjusted r^2) reaching 0.563 for abnormal total cholesterol and 0.568 for high-density lipoprotein (HDL) based on the average surface temperature from head to toe.[22]

The human face contains a wealth of information. We have demonstrated previously that three-dimensional (3D) facial images can be used to reliably assess the aging rate of an individual, especially when coupled with deep learning.[23,24] The facial

vascular and fat distribution is enriched and closely correlated with body surface temperature.[25,26] Yet the current analyses of facial temperature are limited to a rough association between emotion and the total (or average) facial area temperature.[27,28] Research on facial recognition by thermal imaging has gradually come into view, and some studies can roughly confirm the identity of individuals through facial recognition.[29,30] Precise quantification of temperature distribution features of the human face with aging and various diseases has not been explored, and it is still a mystery whether facial thermographic sub-regional features can be used as a marker of aging or diseases, as advocated by traditional Chinese medicine.[31] Here, we collected thermal facial images of 2,811 individuals and developed the "ThermoFace" (TF) method that integrates a facial recognition and temperature extraction system by matching infrared and visible facial images and generates thermal-image-based aging clocks. We further identified lifestyle, health, and molecular features associated with accelerated and decelerated TF aging. The fast and convenient thermal infrared data acquisition and the accuracy of our models make TF an ideal monitoring and screening tool for healthy aging.

RESULTS

Overview of thermal facial image data and the TF algorithm

We collected 2,811 thermal facial images and parallel two-dimensional (2D) images, using the Fluke Ti401pro thermal image camera (details in STAR Methods), from Han Chinese in Jidong, China, in 3 batches from 2020 to 2022, with ages varying from 21 to 88 years in both sexes (female n = 1,339, male n = 1,472) (Figures 1A, S1A, and S1B; Table S1). Based on these data, we developed TF, a thermal infrared face mesh recognition and area division algorithm to align faces across infrared facial images (Figure 1A; STAR Methods). This is achieved by first identifying 54 feature landmarks on the face from the thermal facial images and 486 feature landmarks on corresponding 2D visible images. 2D visible image landmarks formed the regional division template, FaceMesh (STAR Methods); then, using shared feature landmarks in both thermal and visible images, we fit FaceMesh to each thermal facial image. To enable comparison of different thermal facial images, the FaceMesh is standardized to a grid of 897 triangles by Delaunay triangulation[32] constructed from landmarks, which we termed a TF mesh (Figure 1A; STAR Methods).

To maximize the information retained in each TF mesh grid, we kept the max value (Max), min value (Min), and temperature standard deviation (SD) of each triangle (Figure 1A). Also, to avoid the noise caused by the environment temperature, we implemented a strict set of criteria (STAR Methods) in the sample collection process and, most importantly, applied Z score normalization to TF samples to avoid the overall temperature difference and enable cross-sample comparisons; we term these normalized images TF, based on which all the following analyses are performed. The TF samples of batch2020 and batch2021 have a similar distribution on the principal component analysis (PCA) plot (Figure S1C). For a subset of 102 and 82 samples in females and males, respectively, who have thermal images in both years, the TF patterns are highly correlated between the 2 years (Spearman correlation [RCC] = 0.92; Figure S1D). For a subset of 41 subjects, 2–3 TF images from the same sample collected over 30 min to 2 h are clustered by PCA, with within-individual distances significantly smaller than between-individual distances (Figure S1E; $p < 2.2e{-}16$). These indicate that the TF patterns are relatively stable individual-wise and might be regarded as individual signatures.

Aging-associated thermal facial pattern changes

We first examined the average TF-normalized pattern in each decade. This shows an obvious decrease in the nose, cheeks, and eyebrows with age in both females and males. However, the decrease starts at around 50 years in females but around 60 years in males (Figure S1F). We then clustered all the triangles in the TF Max using a soft clustering algorithm[33] (Figure 1B). Four clusters can be clearly observed: cluster 1, highlighting the lower nasolabial folds and lower cheeks, shows a near-linearly age-dependent increase in temperature; cluster 2, highlighting the nose and nasal-mouth region, shows an abrupt, age-dependent temperature drop starting from 50 years old; cluster 3, highlighting the forehead, eyes, and upper nasal labial folds, shows an opposite pattern to cluster 2; and cluster 4, highlighting the eye corners, shows the lowest temperature in middle age (Figure 1B). The TF Min patterns are similar to Max, but SD patterns are different (Figures S2A, S2B, S2D, and S2E). To verify the correlation between the average normalized temperature and chronological age, we plotted the average normalized temperature for the nose area (cluster 2) against age and observed that it indeed decreased with increasing age, particularly after 50 years old (Figure S2C). Average temperature of other clusters, even the age-increased cluster 1, also slightly drops with age, suggesting the age-dependent increases in the normalized TF are due to the relatively larger temperature drops in other facial areas (Figure S2C). Overall, our reconstructed average face profiles reveal previously unrecognized patterns of facial temperature change during aging.

TF aging clocks

To quantitatively detect aging-induced TF patterns, we performed partial least-squares regression (PLSR) (STAR Methods), which maximizes the variance of age-dependent changes when regressing based on age.[34] Because of differences in TF aging between men and women (Figures S1F and S3A), we analyzed them separately. The cumulative variances in the TF data and age explained by the first three components are 52.23% and 62.32% in females and males, respectively. The first three PLS components' prediction is highly correlated with age (Pearson correlation coefficient [PCC] = 0.79 and 0.72 for males and females, respectively) (Figure 2A).

With the increase in PLS component 1, which is highly correlated with age (PCC = 0.66 and 0.53 for males and females, respectively), there is the large decrease in temperature of the nose and a small decrease in the cheeks (Figure 2A), as shown by cluster 2 in the TF clustering (Figure 1B). However, this pattern disappears when the top 3 PLS components are analyzed together due to the orthogonal changes of PLS2 and 3 to PLS1. Instead, the temperature increase and high variability around the eyes and mouth (nasolabial folds) with age remain the most common salient features revealed by PLS1 and the

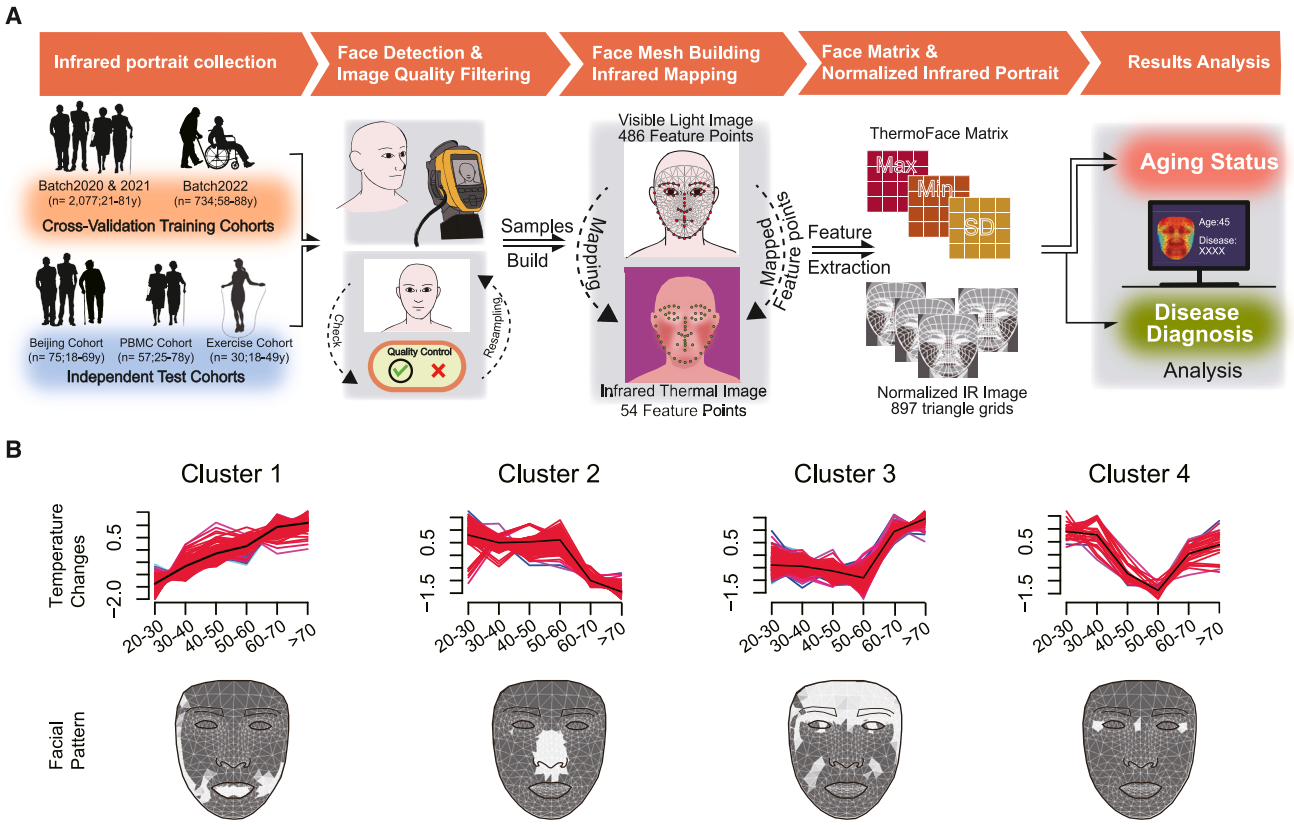


Figure 1. ThermoFace and thermal facial aging patterns

(A) Schematic diagram of thermal facial image collection and the ThermoFace pipeline for data processing and analysis.

(B) Fuzzy c-means clustering identified 4 distinct temporal patterns of temperature changes. The x axis represents six age stages, while the y axis represents $\log_2$-transformed, normalized intensity ratios in each stage. The facial areas of each cluster are displayed as white on gray background.

See also Figures S1 and S2.

top 3 PLS components (Figure 2A), which is often associated with various diseases (see below).

Using the first 13 components, which explained 82.42% and 85.49% of the total variance in the female and male TF data, respectively, we generated the TF PLSR clock. We also trained another linear model using elastic net (EN). Based on 10-fold cross-validation, PLSR has a slightly better performance in predicting chronological age for both females and males with a mean absolute deviation (MAD) of 6.58 (PCC = 0.80, $p < 2.2e{-}16$) and 6.45 years (PCC = 0.84, $p < 2.2e{-}16$), respectively, while EN possesses an MAD of 6.67 (PCC = 0.80, $p < 2.2e{-}16$) and 6.58 years (PCC = 0.80, $p < 2.2e{-}16$).

We also constructed a TF-based convolutional neural network (CNN) model, using the same method we previously used to build a 3D facial image aging model.[23] Based on 10-fold cross-validation, the CNN model is better than linear models, with MAD decreasing to 5.17 years in females and 5.08 years in males (PCC_{female} = 0.90, $p < 2.2e{-}16$; PCC_{male} = 0.90, $p < 2.2e{-}16$) (Figures 2B and S5A). All three TF-based clocks show that the high correlation between the TF predictions for the same subjects over 2 years and the high within-individual correlation confirm that TF can be utilized to estimate the aging status of individuals (Table S2). Saturation analysis shows that MAD and PCC both plateau at approximately 60% of the current sample size for TF PLSR and TF EN models (Figure S4A), but for the ThermoFace CNN (TF.CNN) model, only PCC plateaus; the MAD still shows a slight decrease when approaching 100% of sample size, indicating that more samples may still increase its accuracy, but not dramatically. Furthermore, an independent cohort of 75 healthy individuals (without major chronic diseases) was recruited at Beijing's Guang'anmen Hospital (Beijing cohort, $n = 75$, age range [18,69]); adjusting for the differential age distribution (Figure S4B), the cAgeDiff from 3 TF-based clocks trained on the Jidong cohort exhibited similar performance on this independent cohort (TF PLSR model, MAD of 4.40, PCC = 0.91, $p < 2.2e{-}16$; TF EN model, MAD of 4.47, PCC = 0.92, $p < 2.2e{-}16$; TF.CNN model, MAD of 5.18, PCC = 0.86, $p < 2.2e{-}16$) (Figure S4C). Both analyses above suggest the saturation and stability of the models. However, none of the models are as accurate as the CNN models trained on the relatively smaller sample size of 3D facial images (n = 2,381) collected in the same cohort, which have MADs of 3.09 and 3.15 years for females and males, respectively (Figures S5B and S5C). For convenient comparison with the 3D facial clock in the subsequent analysis, we collectively refer to the three clocks derived from different methods as the "thermal facial

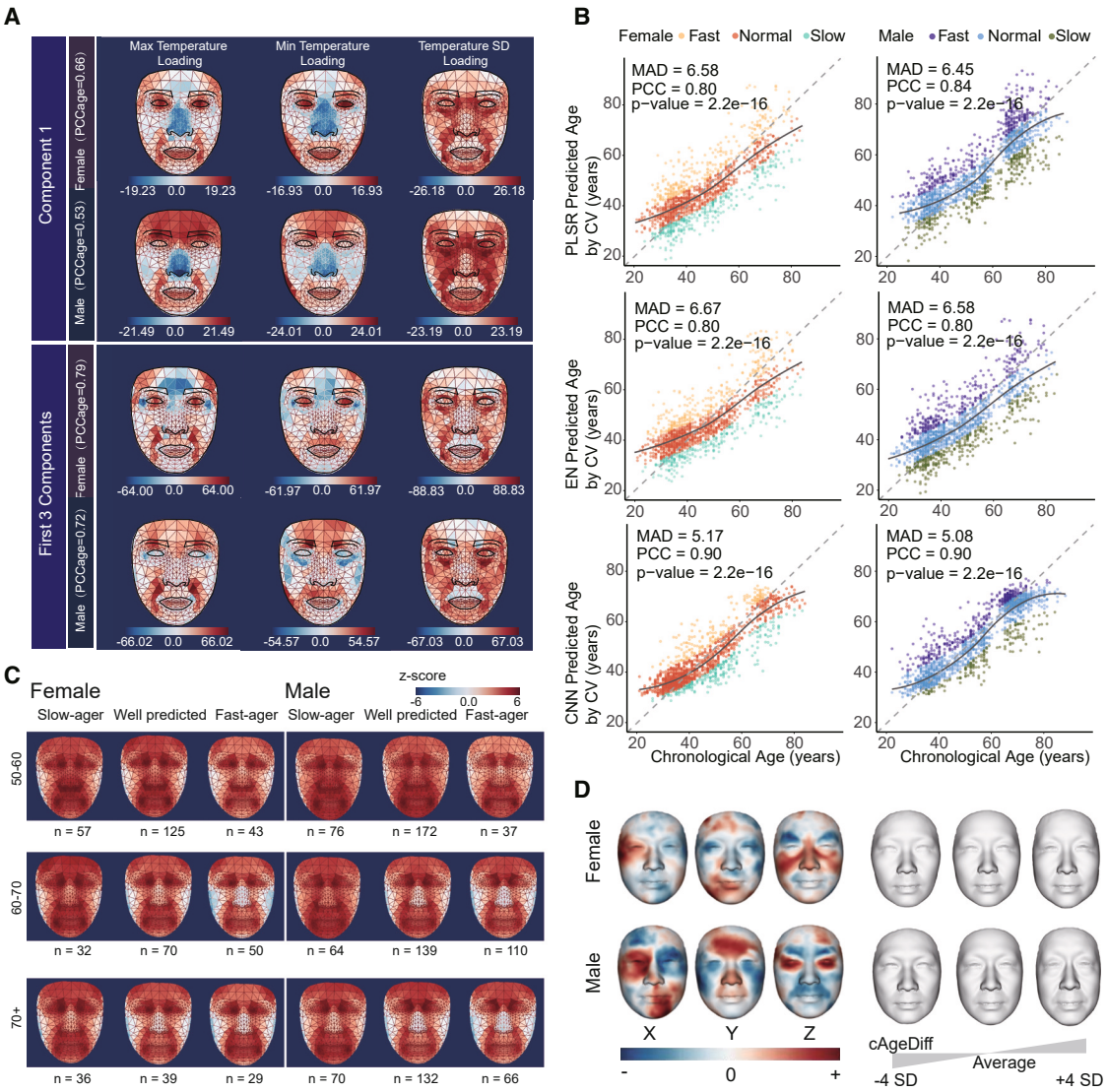


Figure 2. ThermoFace aging clocks

(A) Heatmap of ThermoFace effects displaying efficient values of age-correlated PLS component 1 and the combination of the first 3 components for females and males. The effect coefficients were multiplied by 1,000 and are derived from the PLSR model using all samples used for training. Red and blue denote, respectively, higher and lower values for Max, Min, and SD of temperature.

(B) Correlation of ThermoFace age predicted by three different models with chronological age. Predictors are trained separately in females and males, using 10-fold cross-validation (CV) method (STAR Methods) to predict the age of the held-out samples. Trend line is loess fitted. Fast, normal, and slow aging groups are labeled with different colors, categorized by |cAgeDiffs| > 5 years.

(C) The normalized Max temperature average profiles of the predicted fast agers, slow agers, and well-predicted female and male subjects in age groups older than 50 years. *n* indicates the number of subjects in each class of each age group.

(D) Projections of ThermoFace PLSR cAgeDiff on a 3D facial image (left: color scale in arbitrary units). Red and blue denote, respectively, higher and lower values along x, y, and z axes (left). Synthesized female and male average profiles between −4 and +4 SD of loading values of PLS component 1 and 2 (right).

See also Figures S3–S6.

clocks.'' We also tested using different TF pattern clusters to generate cluster-specific models and found that, individually, they are all inferior to the whole face-based clocks (Table S3), implying that the information contained in the whole face is richer and more comprehensive for aging assessment.

To explore the biological relevance of the thermal facial aging rate versus the 3D facial aging rate, we first utilize the difference between chronological age and predicted age from cross-validation models (AgeDiff) to identify outliers in the aging rate (STAR Methods). We find that, as in many other clock models, AgeDiff slightly overpredicts in young age and underpredicts in old age; to make all four AgeDiffs independent of chronological age, we corrected age effect on AgeDiff by loess to obtain cAgeDiff (Figure S5D; STAR Methods). The lack of correlation between the corrected cAgeDiff and chronological age makes the cAgeDiff comparable across different chronological ages

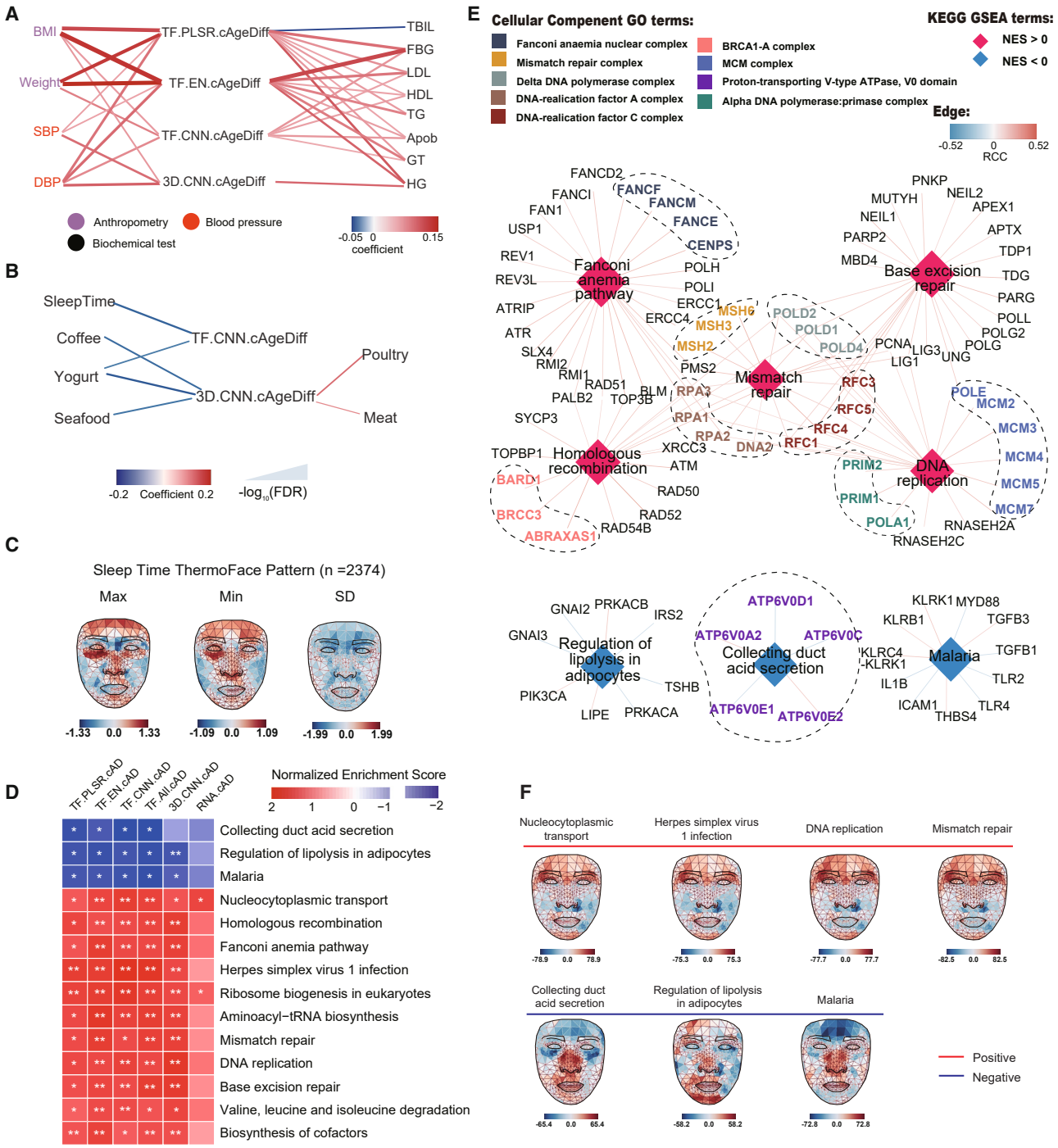


Figure 3. cAgeDiff heterogeneity difference and molecular pathways associated with cAgeDiff

(A and B) Association of cAgeDiff with anthropometric, blood pressure, and biochemical test (A) and lifestyles (B), determined by multi-regression coefficient with Benjamini-Hochberg corrected *p* value. The edge width and color intensity represent the strength of association.

(C) Patterns corresponding to the first two principal components on the ThermoFace regressing to sleep time with PLSR. Sleep time is divided into 5 grades according to the length of time (STAR Methods). The panel shows the Max temperature value PLSR loading multiplied by 1,000. *n* refers to the number of samples used to generate the ThermoFace feature.

(D) KEGG pathway enrichment scores (NES) for thermal cAgeDiff-associated pathways. $^{*}p < 0.05$, $^{**}p < 0.01$. GSEA enrichment of KEGG pathway over genes ranked by their expression RCC to each cAgeDiff across 57 individuals. All four thermal cAgeDiffs must have significant associations to be included, enrichment scores for RNA-seq and 3D cAgeDiffs are shown for comparison.

(legend continued on next page)

(Figure S5D). Based on cAgeDiff, our TF PLSR clock detected faster and slower agers (|cAgeDiff| > 5 years) and, indeed, shows an aging and opposite-to-aging pattern, respectively (Figure 2C). The consistency of outliers given by the different models significantly overlapped ($p < 0.05$), especially between the three TF clock models, while the overlap of the TF.CNN model to the 3DFace CNN model (3D.CNN) model is larger than the two linear TF clocks (TF.PLSR and TF.EN) (Figure S4D; Table S4). Furthermore, the projection of TF cAgeDiff on the 3D face in the x axis direction shows an aging pattern of contraction of the upper half of the face and expansion of the lower half of the face, which is consistent with the 3D facial aging pattern (Figures 2D, S6A, and S6B). A clear difference between men and women appears in the z axis, where women display a prominence in the cheek and nose regions and depression above the brow bone, while men show a depression in the overall face, except for the eye and periocular regions, which become more prominent during aging (Figure 2D). This is perhaps related to the different patterns of facial fatty tissue accumulation during the aging process in males and females. These correlations with 3D aging pattern further confirm that the TF age predictor recapitulates at least part of the 3D facial aging patterns.

Health and lifestyle parameters associate with cAgeDiffs

To compare aging differences, we compiled parameters—including routine anthropometry, blood pressure, biochemical test, and lifestyles—that were collected for over 50% of the participants. The parameters can be clustered into 3 groups (Figures S7B, S7C, and S8). One of these groups (cluster 1) is highly positively correlated with cAgeDiffs and includes blood platelets, white blood cells, and lymphocyte count in both sexes, and some other parameters are sex-specific, grouped with age and cAgeDiff (Figures S7B and S7C). Analyzing the relationship between each parameter and cAgeDiff in greater depth by multiple regression analysis (STAR Methods), we find that the 3 TF cAgeDiffs are all significantly positively correlated with metabolic disease indicators more than the 3D facial clock cAgeDiff (Figure 3A), such as BMI, fasting blood glucose, and apolipoprotein B ($p < 0.05$ and false discovery rate [FDR] < 0.1; Table S6), suggesting that metabolic states are better reflected on thermal facial patterns than in 3D structure.

On the other hand, the 3D facial clock is more associated with lifestyles than thermal facial clocks, and they both have yogurt intake as a counteracting factor. However, thermal clock cAgeDiff is significantly negatively correlated with sleep time, but not 3D clock (Figure 3B). Sleep time shown by the first two components of PLSR, based on a combination of temperature Max, Min, and SD patterns (with predicted to actual sleep time PCC of 0.415, $p < 2.2e{-}16$), manifests as an increase of both Max and Min temperature in the forehead and under the eye and nose, a decrease in the cheeks, and a dramatic decrease of temperature SD in nearly all facial areas (Figure 3C). Compared with sleep time, yogurt intake—shown by the first two components of PLSR (with predicted to yogurt intake level PCC of 0.465, $p < 2.2e{-}16$)—shared the same increase of temperature in the undereye area and nose and a dramatic decrease of temperature SD in all facial areas, but its decrease in the cheeks, especially in the nasolabial folds, was more dramatic, and the increase in nose temperature is concentrated on the tip and wing of the nose rather than the bridge (Figure S20). These suggest a strong impact of sleep and yogurt intake on the thermal facial pattern and steady-state regulation.

When breaking down the associations into different age groups, we find the thermal facial cAgeDiff association ($p < 0.05$ and FDR < 0.1) with metabolic indicators exists similarly in all age groups; in females, the cAgeDiff associations for all three thermal facial clocks are stronger than those for the 3D facial clock, but the cAgeDiff association strengths for the thermal clocks and the 3D clock are similar in males (Figure S9A). In contrast, the cAgeDiff associations with lifestyles are very different in different age groups—and sometimes opposite (Figure S9B). In particular, meals-on-time is negatively associated with thermal clock cAgeDiffs in young people (20–30 years) but strongly positively associated in old people (over 60 years) (Figure S9B), suggesting that the same food intake behavior may result in a different metabolic state in young and old people.

Blood transcriptome changes associated with TF cAgeDiffs

We further explored the molecular underpinnings associated with TF aging by collecting peripheral blood mononuclear cell (PBMC) transcriptomic data of 57 healthy individuals from Guang'anmen Hospital and built the PBMC cohort (mean age 49.63 years, age interval [23,78]; STAR Methods). Thermal facial images and 3D facial images were also taken at the same time, with the volunteers in a calm state. For comparison, we also built 3D facial clocks and an RNA sequencing (RNA-seq) clock for the same cohort as described previously.[23] The blood samples correspond to TF average cAgeDiff intervals in [−10.1, 11.7] and an average thermal facial age MAD of 4.77 years. Pathway association by gene set enrichment analysis (GSEA) and detailed molecular network analyses shows that the great majority of the TF cAgeDiff positively associated pathways are linked into 5 DNA repair pathways, 3 RNA processing and transport pathways, and the Herpes virus infection response pathway, which consists of 57 virus-defense-related zinc-finger transcription factors ($p < 0.05$; Figures 3D, 3E, and S11–S13). All of these pathways are highly energy consuming.[35–38] On the other hand, only 3 pathways are found to be negatively associated with TF cAgeDiff: regulation of lipolysis in adipocyte; collecting duct acid secretion, which consists of 5 ATPases; and the malaria pathway, which consists

(E) Network among genes within the KEGG pathways correlated with thermal AgeDiffs determined by GSEA. The color of each edge represents the RCC value of the gene, with the average thermal cAgeDiff. Edges indicate the membership of a gene to a pathway. Each circle marks a gene complex (comprising at least 3 genes) identified through cellular component GO terms.

(F) ThermoFace patterns of thermal cAgeDiff positively and negatively associated pathways. The panel shows the Max temperature PLSR first 3 components' total loading values to each pathway's *Z* score-normalized median gene expression values multiplied by 1,000 among 57 individuals (PBMC cohort). The red and blue bars above the images denote the positive and negative GSEA NES (>1.5 and <−1.5) of the pathway to the average thermal cAgeDiff, respectively.
See also Figures S7–S13.

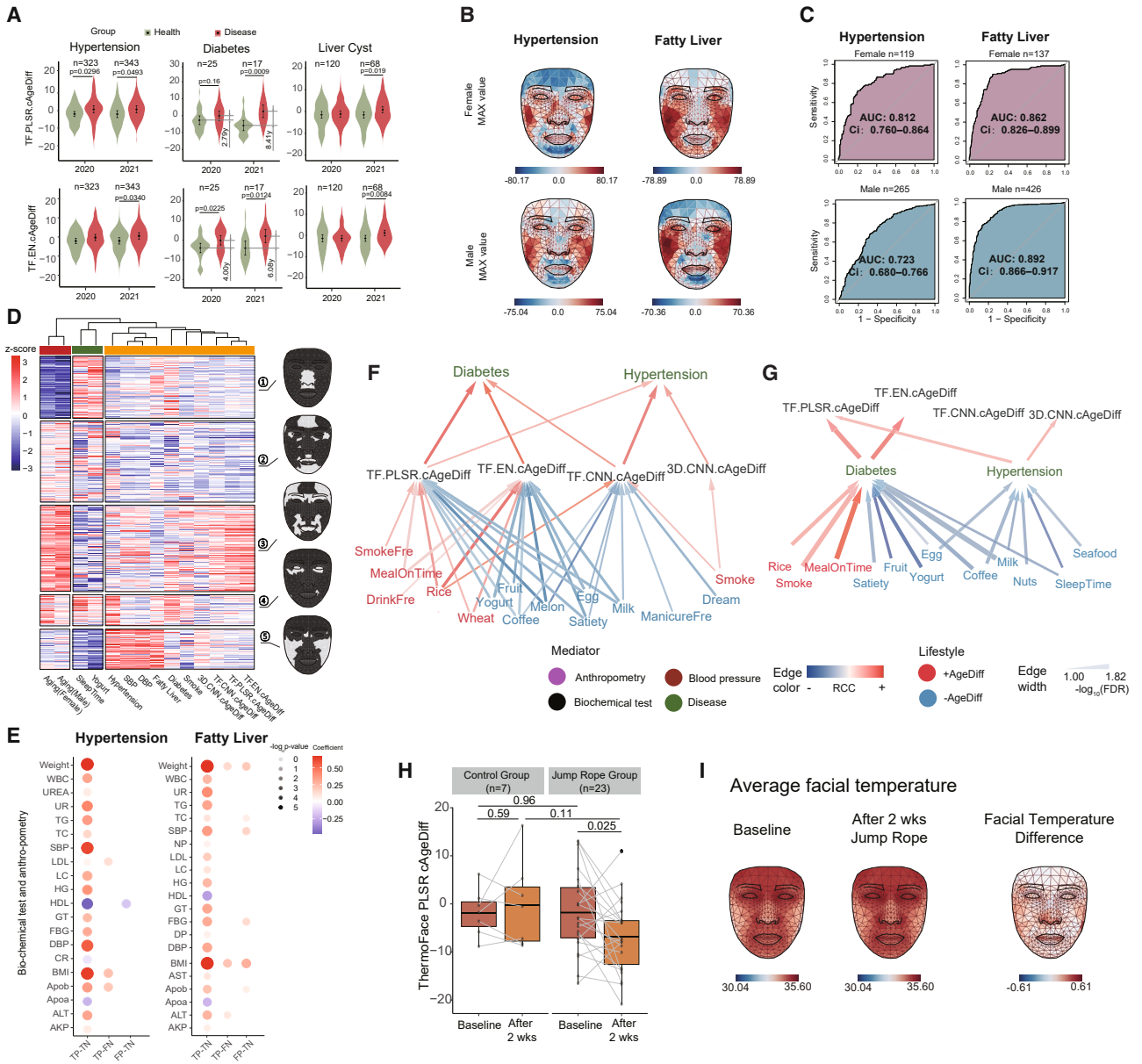


Figure 4. ThermoFace association with metabolic diseases and mapping to thermal facial patterns

(A) ThermoFacePLSRcAgeDiff and ThermoFaceENcAgeDiff of disease and age-matched control individuals in two batches; green and red indicates healthy and disease sample results, respectively. *n* is the number of disease samples. Gray lines mark the mean difference in years. Significant *p* values are determined by unpaired two-sided Student's t test.

(B) ThermoFace patterns of first two principal components of a disease. The panel shows the Max temperature value loading multiplied by 1,000.

(C) ThermoFace PLS-DA disease bi-classification ROC curve for the best prediction of each disease among cross-validation models (Figure S19A). CI with 95% confidence interval. *n* is the number of disease samples.

(D) Euclidean distance-based clustering of aging or disease ThermoFace patterns. The facial area of the corresponding cluster is displayed on the right as white on black background; see Table S5 for abbreviations correspondence.

(E) Bubble plot of logistic regression model correlation coefficient of factors influencing disease prediction for TP (1) versus TN (0), TP (1) versus FN (0), and FP (1) versus TN (0). Color shows slop coefficient; dot intensity indicates Benjamini-Hochberg adjusted two-sided Student's t test *p* value. See Table S5 for abbreviation descriptions.

(F and G) Network of inferred causal relationships (causal-inference test with FDR < 0.1; STAR Methods) from lifestyle factors to disease via cAgeDiffs mediators (F) and relationships from lifestyle factors to cAgeDiffs via disease mediators (G) among all samples.

(legend continued on next page)

of proteins involved in detecting, responding to, and killing malaria-infected cells ($p < 0.05$, Figure 3D, 3E, and S11–S13). These pathways, in general, release heat.[39–42]

The TF projections of the cAgeDiff-associated pathways show that the pathways with a positive association reflected an increase in temperature in the mid-face, eye, and forehead regions, reminiscent of the cluster 3 pattern in Figure 1B, while pathways with a negative association showed a decrease in temperature on forehead and around the eyes and an increase in the middle of the face, in particular the nose, reminiscent of the cluster 2 pattern in Figure 1B (Figure 3F; top 20 genes in each pathway are shown in Table S9).

Metabolic diseases accelerate TF aging clocks

Aging is accompanied by the onset of aging-related diseases, with a gradual increase in the incidence of various diseases in middle age and beyond. In our cohort, fatty liver occurs earlier in men than women, and women are more likely to develop intrahepatic calcification (Table S7).

Hypertension, diabetes mellitus, and liver cysts show significantly accelerated aging compared with the healthy group of the same age distribution (Benjamini-Hochberg corrected t test, $p < 0.05$), and similar results are observed when the data are split into two non-overlapping batches (Figures 4A and S15; Table S8). Diabetes causes the most pronounced increase in cAgeDiffs, with the diabetic population aging an average of 6.28 years more relative to the healthy population in the TF prediction and 2.86 years in the 3D face prediction (Figure 4A; Table S8). This implies TF is more associated with health and disease than 3D face.

We then directly trained classifiers to distinguish various diseases from healthy age- and sex-matched controls. We find diseases predictable by thermal facial images (AUC > 0.80) are hypertension in females (AUC = 0.817, accuracy = 0.782, specificity = 0.800, sensitivity = 0.723) and fatty liver in both males (AUC = 0.892, accuracy = 0.800, specificity = 0.765, sensitivity = 0.863) and females (AUC = 0.812, accuracy = 0.797, specificity = 0.898, sensitivity = 0.756) (Figures 4C, S16A, S17A, and S17C; Table S10). Saturation analysis conducted by randomly selecting data shows that AUC and predictor accuracy plateau after 60% or 80% of the sample size for hypertension, but not completely saturated for fatty liver, thus there is still room for improvement (Figure S16B), indicating that there is a great potential for using TF to detect these diseases. Reserving part of the data as an independent validation dataset, the models gave a similar predictive performance on this independent dataset as cross-validations (Figure S17B), demonstrating the stability and statistical power of the models. TF disease predictors also showed some ability to differentiate between diseases (Table S11). TF disease predictors reached better performance when separately predicted for each sex (Figures 4C and S17A), and facial patterns are clearer than predictors trained from both sexes (Figures 4B, S18A, and S18B). In addition, the effect of taking medication or not on disease predictors was negligible (Figure S18C). When the same method was used to examine the effectiveness of the 3D face in predicting diseases, it was found the 3D face predictor was inferior to the TF predictor, especially for the determination of hypertension (Figure S19). Although displaying a differential increase in temperature on cheeks, forehead, and peri-mouth regions, hypertension shows increased temperature in the eye area in both sexes and a relatively colder nose in males (Figure 4B), which are similar to the pattern of cAgeDiffs and aging, and also increased temperature on the cheeks, which is opposite to the pattern of cAgeDiffs and aging (compare to Figures 2A and 2C). Interestingly, these disease TF patterns are more similar to the inflammation-related molecular pathway expression patterns than cAgeDiff and aging (compare to Figure 3F), suggestive of chronic inflammation as a major contributor to the disease TF patterns. In TF regression prediction for systolic blood pressure (SBP) and diastolic blood pressure (DBP) (with a predicted to SBP PCC of 0.620, $p < 2.2e{-}16$ and predicted to DBP PCC of 0.573, $p < 2.2e{-}16$), the TF features of SBP and DBP have similar patterns of increased temperature on the cheeks and decreased temperature on the forehead area, and both are similar to hypertension TF patterns (Figure S18A). However, unlike in hypertension, which has a temperature rise in the periocular region, the TF pattern of SBP and DBP exhibited lower temperatures in the eyes and periorbital regions, which might suggest the involvement of additional factors contributing the hypertension disease other than blood pressure itself (Figure S20). BIC-SKMeans[43] clustering (Figure 4D) of the impact on TF of aging and cAgeDiff versus diseases and cAgeDiff-associated lifestyles (Figure S20) reveals that most diseases and smoking cluster with various cAgeDiffs in the orange cluster, which are similar to aging (red cluster) but less pronounced and opposite to the sleep time and yogurt intake cluster (green cluster). Specifically, region 1 is concentrated in the nose region, which shows a strong decrease in temperature with aging and weaker decrease with cAgeDiffs, smoking, and hypertension; region 2 contains the forehead center, outer eye corner regions, and chin, and an increase in temperature with aging but a decrease with cAgeDiffs, smoking, and diseases; region 3 contains side of forehead, temples, upper eye lids, nasolabial folds, and mouth, where both aging and cAgeDiffs strongly enhance the temperature, similar to hypertension, diabetes, and smoking; region 4 mainly includes eyeballs and under eye, where hypertension, fatty liver, diabetes, and smoking all strongly enhance the temperature, similar to aging and thermal cAgeDiff patterns, thus making it the most consistent aging and aging-acceleration pattern of all conditions examined (followed by the opposite in region 1); and region 5 includes the cheek area, where temperature weakly decreases with aging and cAgeDiffs but is strongly increased by hypertension, fatty liver, and diabetes, similar to 3D facial cAgeDiff (Figure 4D). As expected, sleep time and yogurt intake have the opposite patterns in region 1 and region 3 compared with age and cAgeDiff patterns. Additionally, in region 5, sleep time and

(H) ThermoFace PLSR cAgeDiffs before and after jump rope training compared with control group. Paired t test p values are based on baseline (initial assessment) and 2 weeks later in the control and jump rope groups, separately, and unpaired t test p values between control groups and jump rope group. n indicates the number of subjects.

(I) Visualization of average ThermoFace temperature and difference between two average temperature plots. Color bar shows temperature (°C).

See also Figures S14–S22.

yogurt intake both have the strongest temperature decrease on the cheeks, opposite to the strong temperature increase of hypertension and fatty liver (Figure 4D). These suggest a partial or disconcerted aging acceleration induced by these aging-related metabolic syndromes and smoking, which is more similar to aging acceleration than aging itself.

Multi-regression models corrected by age and sex reveal that higher body weight, blood pressure (SBP and DBP), and BMI and lower HDL are the most influential physiological factors for disease prediction in patients with fatty liver and hypertension (true positives [TPs]) compared with age-matched healthy controls (true negatives [TNs]) (Figure 4E). Interestingly, healthy individuals predicted to have a disease (false positives [FPs]), or individuals with a disease but predicted to be healthy (false negatives [FNs]), often have the expected disease or health patterns of these physiological parameters (Figure 4E). Thus, the FPs and FNs of models can be partially explained by metabolic status of the subjects and, in turn, may guide disease prevention. Similar associations can be seen by the significant differential levels of these parameters when pairwise-comparing FP versus TN or TP versus FN (Figure S21).

We then asked whether the metabolic disease can mediate the lifestyle impact on TF aging rates (cAgeDiffs) or the other way around by causal inference test (CIT) (STAR Methods). As inferred by CIT analysis, diabetes and hypertension can both mediate the impact of various lifestyles on TF aging acceleration (Figure 4G) and be promoted by TF aging acceleration (Figure 4F), with most causal lifestyle factors shared between the two models, except adequate sleep, which seems to only counteract the diseases' impact on TF aging acceleration (Figure 4G), but not the other way around (Figure 4F), which implies that adequate sleep can alleviate the effects on aging caused by these diseases, but not vice versa. Similar patterns can be found when separating individuals by sex and by age, specifically middle-aged and older individuals (Figure S22), suggesting the robustness of these associations.

Exercise reduces thermal facial aging

Jump training enhances cardiovascular health, effectively burns calories, provides full-body exercise, and improves coordination and agility.[44,45] Previous research has demonstrated the benefits of jump rope training on improving overall health, including heart, bone, respiration, and metabolic health (https://runrepeat.com/benefits-of-jumping-rope).[46] In order to investigate whether TF responds to anti-aging interventions, we established an exercise cohort at Hainan Medical University based on jump rope training, with >10 min and >800 jump times between 5:30 p.m. and 6:30 p.m. for 2 weeks (STAR Methods) (n = 30, age range [18,49] years, 7 in control, 23 in jump rope group; t test shows non-significant age distribution [$p > 0.05$]). All thermal images were taken at a calm state without any acute effect of physical activity. After 2 weeks of training, there was a significant decrease in TF PLSR cAgeDiff, with an average decrease of 5 years (Figure 4H; t test, $p = 0.025$) in the training group, or 5.5 years when compared with the controls after 2 weeks ($p = 0.11$), while a slight but insignificant increase of 0.5 years was observed in the control group. cAgeDiff of baseline between the control group and the jump rope group showed no significant difference (t test, $p > 0.05$). For EN and CNN clocks, comparing the before and after jump rope training in the training group, there is a decrease in the average cAgeDiff, but the decrease is statistically insignificant. However, we noticed that the control group had an opposite but insignificant change in cAgeDiff after 2 weeks, indicating a slight upward shift in baseline measured by these two clocks. Considering this, comparing the cAgeDiff between the jump rope training group and the control group after 2 weeks (column 4 versus 2), the EN cAgeDiff shows a significant decrease ($p = 0.027$) and the CNN cAgeDiff shows a marginally significant decrease ($p = 0.071$) (Figure S14). Visualization of baseline and the jump rope group's mean facial temperatures after 2 weeks shows that the facial temperatures of the jump rope group (at a calm state) were overall higher than those of the non-exercise group (Figure 4I); the differential thermal pattern ($p < 0.05$) indicates that temperatures in the nose region of the jump rope exercise group were significantly elevated in contrast to the age-dependent decrease in this region (cluster 2 in Figure 1B). This indicates the possibility of mediating thermal facial aging with something as simple as 2 weeks of daily jump rope exercise.

DISCUSSION

In this study, we established that thermographic facial images can be used to estimate a thermal biological age and the extent or rate of aging, and they can promisingly predict several metabolic diseases, even in independent cohorts. In recent studies,[47,48] distinctions between aging markers and aging rate markers have been proposed. Our data demonstrated that the differential TF AgeDiffs (aging rate) of people with the same chronological age are associated with their health and disease states and that a specific health intervention, such as jump rope, could decrease the TF AgeDiffs, indicating that TF ages can be used to assess biological age, placing them within the category of biomarkers of biological age. We developed a complete set of standardized extraction methods for infrared facial information based on infrared and matching 2D visible images of faces acquired by infrared thermography—TF. We quantified the changes of facial thermographic features and patterns with aging. Our results indicate that the 3D facial aging clock predicts the chronological age better than the TF aging clock (with a lower MAD), which may be due to the larger amount of information it contains. Each 3D facial image contains 7,200 data points, while TF currently contains 897 data points. Nonetheless, our further findings show that TF images have a wider association with physiological parameters than do 3D facial images. Our TF cAgeDiffs reveal similar molecular associations with inflammation/infection signatures and ribosome levels in the blood cells as other published aging clocks, but show stronger associations with metabolic health and diseases, consistent with the notion that body temperature distribution is a direct readout of cellular metabolism. Thus, it is not surprising that TF, but not 3D facial images, can well predict several metabolic diseases. Overall, despite lacking the same level of precision in estimating the chronological age as 3D facial images, TF is more related to an individual's health, hence biological age.

The association of periocular temperature increases with various metabolic diseases points to a common feature of metabolic disorders. At least in the case of diabetes, ocular

phenotypes are not only a manifestation of the disease but are also now fully established as an early diagnosis marker.[49] It would be interesting to see whether the eye area—due to its intense vascularization and metabolic activities—is the most sensitive to fluctuations in metabolic status. Other than this common feature, different diseases display differential temperature changes on the cheeks, forehead, and peri-mouth regions, suggesting the possibility of differentiating these diseases based on thermal facial patterns.

Limitations of the study

Facial thermography has been shown to have a correlation with participant emotion,[50] outdoor temperatures, and seasonal influences throughout the year, all of which might influence the result; as such, image collections were carried out in the morning in air-conditioned rooms with room temperature between 20°C and 25°C after participants enter a calm state of mind and body. Thus, in this study, we tried our best to avoid the influence of emotion and environment temperature changes, which by themselves are interesting topics to explore in the future. In this study, the thermal images of all but the Hainan jump rope cohort were taken in the early morning before breakfast, while those from the Hainan jump rope cohort were taken at random times between 9 a.m. and 12 p.m. or 3 and 6 p.m. Because circadian fluctuations were largely excluded when the clocks were built, it is unclear whether images taken in the afternoon or after meals may introduce confounding factors, and this awaits further investigation.

Our current results indicate that thermal facial aging can be altered by a 2-week daily jump rope exercise; limitations in sample size and tracking duration necessitate further in-depth investigation into the long-term metabolic effects of sustained physical activity.

Additionally, other than the jump rope cohort, which is from a subtropical region, this study was mainly based on the thermal facial imaging samples in the regions of Hebei and Beijing, which share similarities in terms of climate. Our current study only involves Han Chinese. To extend it to other ethnic groups, populations from different ethnic backgrounds and climates to need to be tested for the effectiveness of TF as an aging and health assessment across diverse human populations.

STAR★METHODS

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

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cmet.2024.05.012.

ACKNOWLEDGMENTS

This work was supported by grants from the National Natural Science Foundation of China (92374207, 92049302, 32088101, and 32330017) and Ministry of Science and Technology of the People's Republic of China (2020YFA0804000) to J.-D.J.H.

AUTHOR CONTRIBUTIONS

J.-D.J.H. conceived and designed the study. Y.Z. set up the cohort and collected data from Jidong. B.P. and H.L. collected data from Beijing. Yi Wang provided Majiagou data. K.W. and T.J. collected data from Hainan. Z.Y. analyzed the ThermoFace data, guided by J.-D.J.H. Z.Y. implemented the ThermoFace algorithm. K.M. analyzed the thermal age association in RNA-seq data. H. Zheng and H. Zhai processed RNA-seq data. Yiyang Wang and K.M. analyzed the 3D facial data. Z.Y. and J.-D.J.H. wrote the manuscript with help from K.M.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 26, 2023
Revised: March 19, 2024
Accepted: May 22, 2024
Published: July 2, 2024

REFERENCES

1. López-Otín, C., and Kroemer, G. (2021). Hallmarks of health. Cell *184*, 33–63. https://doi.org/10.1016/j.cell.2020.11.034.

2. López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The hallmarks of aging. Cell *153*, 1194–1217. https://doi.org/10.1016/j.cell.2013.05.039.

3. Krištić, J., Vučković, F., Menni, C., Klarić, L., Keser, T., Beceheli, I., Pučić-Baković, M., Novokmet, M., Mangino, M., Thaqi, K., et al. (2014). Glycans are a novel biomarker of chronological and biological ages. J. Gerontol. A Biol. Sci. Med. Sci. *69*, 779–789. https://doi.org/10.1093/gerona/glt190.

4. Lehallier, B., Shokhirev, M.N., Wyss-Coray, T., and Johnson, A.A. (2020). Data mining of human plasma proteins generates a multitude of highly predictive aging clocks that reflect different aspects of aging. Aging Cell *19*, e13256. https://doi.org/10.1111/acel.13256.

5. Peters, M.J., Joehanes, R., Pilling, L.C., Schurmann, C., Conneely, K.N., Powell, J., Reinmaa, E., Sutphin, G.L., Zhernakova, A., Schramm, K., et al. (2015). The transcriptional landscape of age in human peripheral blood. Nat. Commun. *6*, 8570. https://doi.org/10.1038/ncomms9570.

6. Robinson, O., Chadeau Hyam, M., Karaman, I., Climaco Pinto, R., Ala-Korpela, M., Handakas, E., Fiorito, G., Gao, H., Heard, A., Jarvelin, M.R., et al. (2020). Determinants of accelerated metabolomic and epigenetic aging in a UK cohort. Aging Cell *19*, e13149. https://doi.org/10.1111/acel.13149.

7. Xia, X., Wang, Y., Yu, Z., Chen, J., and Han, J.J. (2021). Assessing the rate of aging to monitor aging itself. Ageing Res. Rev. *69*, 101350. https://doi.org/10.1016/j.arr.2021.101350.

8. Morrison, S.F. (2016). Central control of body temperature. F1000Res. *5*. F1000 Faculty Rev-880. https://doi.org/10.12688/f1000research.7958.1.

9. Flouris, A.D., and Piantoni, C. (2015). Links between thermoregulation and aging in endotherms and ectotherms. Temperature (Austin) *2*, 73–85. https://doi.org/10.4161/23328940.2014.989793.

10. Rachakatla, A., and Kalashikam, R.R. (2022). Calorie restriction-regulated molecular pathways and its impact on various age groups: an overview. DNA Cell Biol. *41*, 459–468. https://doi.org/10.1089/dna.2021.0922.

11. Colman, R.J., Anderson, R.M., Johnson, S.C., Kastman, E.K., Kosmatka, K.J., Beasley, T.M., Allison, D.B., Cruzen, C., Simmons, H.A., Kemnitz, J.W., and Weindruch, R. (2009). Caloric restriction delays disease onset and mortality in rhesus monkeys. Science *325*, 201–204. https://doi.org/10.1126/science.1173635.

12. Xiao, R., Zhang, B., Dong, Y., Gong, J., Xu, T., Liu, J., and Xu, X.Z.S. (2013). A genetic program promotes C. elegans longevity at cold temperatures via a thermosensitive TRP channel. Cell *152*, 806–817. https://doi.org/10.1016/j.cell.2013.01.020.

13. Conti, B., Sanchez-Alavez, M., Winsky-Sommerer, R., Morale, M.C., Lucero, J., Brownell, S., Fabre, V., Huitron-Resendiz, S., Henriksen, S., Zorrilla, E.P., et al. (2006). Transgenic mice with a reduced core body temperature have an increased life span. Science *314*, 825–828. https://doi.org/10.1126/science.1132191.

14. Ravussin, E., Redman, L.M., Rochon, J., Das, S.K., Fontana, L., Kraus, W.E., Romashkan, S., Williamson, D.A., Meydani, S.N., Villareal, D.T., et al. (2015). A 2-year randomized controlled trial of human caloric restriction: feasibility and effects on predictors of health span and longevity. J. Gerontol. A Biol. Sci. Med. Sci. *70*, 1097–1104. https://doi.org/10.1093/gerona/glv057.

15. Landsberg, L. (2012). Core temperature: a forgotten variable in energy expenditure and obesity? Obes. Rev. *13*, 97–104. https://doi.org/10.1111/j.1467-789X.2012.01040.x.

16. Waalen, J., and Buxbaum, J.N. (2011). Is older colder or colder older? The association of age with body temperature in 18,630 individuals. J. Gerontol. A Biol. Sci. Med. Sci. *66*, 487–492. https://doi.org/10.1093/gerona/glr001.

17. Heldmaier, G., and Ruf, T. (1992). Body temperature and metabolic rate during natural hypothermia in endotherms. J. Comp. Physiol. B *162*, 696–706. https://doi.org/10.1007/BF00301619.

18. Geiser, F. (2004). Metabolic rate and body temperature reduction during hibernation and daily torpor. Annu. Rev. Physiol. *66*, 239–274. https://doi.org/10.1146/annurev.physiol.66.032102.115105.

19. Kataoka, N., Hioki, H., Kaneko, T., and Nakamura, K. (2014). Psychological stress activates a dorsomedial hypothalamus-medullary raphe circuit driving brown adipose tissue thermogenesis and hyperthermia. Cell Metab. *20*, 346–358. https://doi.org/10.1016/j.cmet.2014.05.018.

20. Kenny, G.P., Sigal, R.J., and McGinn, R. (2016). Body temperature regulation in diabetes. Temperature (Austin) *3*, 119–145. https://doi.org/10.1080/23328940.2015.1131506.

21. Reith, J., Jørgensen, H.S., Pedersen, P.M., Nakayama, H., Raaschou, H.O., Jeppesen, L.L., and Olsen, T.S. (1996). Body temperature in acute stroke: relation to stroke severity, infarct size, mortality, and outcome. Lancet *347*, 422–425. https://doi.org/10.1016/s0140-6736(96)90008-2.

22. Thiruvengadam, J., Anburajan, M., Menaka, M., and Venkatraman, B. (2014). Potential of thermal imaging as a tool for prediction of cardiovascular disease. J. Med. Phys. *39*, 98–105. https://doi.org/10.4103/0971-6203.131283.

23. Xia, X., Chen, X., Wu, G., Li, F., Wang, Y., Chen, Y., Chen, M., Wang, X., Chen, W., Xian, B., et al. (2020). Three-dimensional facial-image analysis to predict heterogeneity of the human ageing rate and the impact of lifestyle. Nat. Metab. *2*, 946–957. https://doi.org/10.1038/s42255-020-00270-x.

24. Chen, W., Qian, W., Wu, G., Chen, W., Xian, B., Chen, X., Cao, Y., Green, C.D., Zhao, F., Tang, K., and Han, J.D. (2015). Three-dimensional human facial morphologies as robust aging markers. Cell Res. *25*, 574–587. https://doi.org/10.1038/cr.2015.36.

25. Salamunes, A.C.C., Stadnik, A.M.W., and Neves, E.B. (2017). The effect of body fat percentage and body fat distribution on skin surface temperature with infrared thermography. J. Therm. Biol. *66*, 1–9. https://doi.org/10.1016/j.jtherbio.2017.03.006.

26. Weinstein, S.A., Weinstein, G., Weinstein, E.L., and Gelb, M. (1991). Facial thermography, basis, protocol, and clinical value. Cranio *9*, 201–211. https://doi.org/10.1080/08869634.1991.11678368.

27. Wang, S., Liu, Z., Lv, S., Lv, Y., Wu, G., Peng, P., Chen, F., and Wang, X. (2010). A natural visible and infrared facial expression database for expression recognition and emotion inference. IEEE Trans. Multimedia *12*, 682–691. https://doi.org/10.1109/TMM.2010.2060716.

28. Nakanishi, R., and Imai-Matsumura, K. (2008). Facial skin temperature decreases in infants with joyful expression. Infant Behav. Dev. *31*, 137–144. https://doi.org/10.1016/j.infbeh.2007.09.001.

29. Lin, S.D., Chen, L., and Chen, W. (2021). Thermal face recognition under different conditions. BMC Bioinformatics *22*, 313. https://doi.org/10.1186/s12859-021-04228-y.

30. Głowacka, N., and Rumiński, J. (2021). Face with mask detection in thermal images using deep neural networks. Sensors (Basel) *21*, 6387. https://doi.org/10.3390/s21196387.

31. Li, N., Yu, J., Mao, X., Zhao, Y., and Huang, L. (2022). The research and development thinking on the status of artificial intelligence in traditional Chinese medicine. Evid. Based Complement. Alternat. Med. *2022*, 7644524. https://doi.org/10.1155/2022/7644524.

32. Shewchuk, J.R. (1996). Workshop on applied computational geometry (Springer), pp. 203–222.

33. Kumar, L., and E Futschik, M. (2007). Mfuzz: a software package for soft clustering of microarray data. Bioinformation *2*, 5–7. https://doi.org/10.6026/97320630002005.

34. Wold, S., Sjöström, M., and Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst. *58*, 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1.

35. Qin, L., Fan, M., Candas, D., Jiang, G., Papadopoulos, S., Tian, L., Woloschak, G., Grdina, D.J., and Li, J.J. (2015). CDK1 enhances mitochondrial bioenergetics for radiation-induced DNA repair. Cell Rep. *13*, 2056–2063. https://doi.org/10.1016/j.celrep.2015.11.015.

36. Palmer, C.S. (2022). Innate metabolic responses against viral infections. Nat. Metab. *4*, 1245–1259. https://doi.org/10.1038/s42255-022-00652-3.

37. Licatalosi, D.D., and Darnell, R.B. (2010). RNA processing and its regulation: global insights into biological networks. Nat. Rev. Genet. *11*, 75–87. https://doi.org/10.1038/nrg2673.

38. Lee, C.-D., and Tu, B.P. (2017). Metabolic influences on RNA biology and translation. Crit. Rev. Biochem. Mol. Biol. *52*, 176–184. https://doi.org/10.1080/10409238.2017.1283294.

39. Puchades, C., Sandate, C.R., and Lander, G.C. (2020). The molecular principles governing the activity and functional diversity of AAA+ proteins. Nat. Rev. Mol. Cell Biol. *21*, 43–58. https://doi.org/10.1038/s41580-019-0183-6.

40. Yang, A., and Mottillo, E.P. (2020). Adipocyte lipolysis: from molecular mechanisms of regulation to disease and therapeutics. Biochem. J. *477*, 985–1008. https://doi.org/10.1042/BCJ20190468.

41. Arruda, A.P., Ketzer, L.A., Nigro, M., Galina, A., Carvalho, D.P., and de Meis, L. (2008). Cold tolerance in hypothyroid rabbits: role of skeletal muscle mitochondria and sarcoplasmic reticulum Ca2+ ATPase isoform 1 heat production. Endocrinology *149*, 6262–6271. https://doi.org/10.1210/en.2008-0564.

42. Schumann, R.R. (2007). Malarial fever: Hemozoin is involved but Toll-free. Proc. Natl. Acad. Sci. USA *104*, 1743–1744. https://doi.org/10.1073/pnas.0610874104.

43. Sun, N., Yu, X., Li, F., Liu, D., Suo, S., Chen, W., Chen, S., Song, L., Green, C.D., McDermott, J., et al. (2017). Inference of differentiation time for single cell transcriptomes using cell population reference data. Nat. Commun. *8*, 1856. https://doi.org/10.1038/s41467-017-01860-2.

44. Pouyafar, M., Askari, R., Hoseini Kakhk, S.A., Damavandi, M., and Maleki, A. (2021). Comparing the effects of eight weeks of whole body vibration exercise combined with rope skipping at two different intensities on physical performance of older men: a randomized single-blind clinical trial. Salmand *16*, 376–395. https://doi.org/10.32598/sija.2021.16.3.2885.2.

45. Tang, Z., Ming, Y., Wu, M., Jing, J., Xu, S., Li, H., and Zhu, Y. (2021). Effects of caloric restriction and rope-skipping exercise on cardiometabolic health: a pilot randomized controlled trial in young adults. Nutrients *13*, 3222. https://doi.org/10.3390/nu13093222.

46. Sung, K.-D., Pekas, E.J., Scott, S.D., Son, W.-M., and Park, S.-Y. (2019). The effects of a 12-week jump rope exercise program on abdominal adiposity, vasoactive substances, inflammation, and vascular function in adolescent girls with prehypertension. Eur. J. Appl. Physiol. *119*, 577–585. https://doi.org/10.1007/s00421-018-4051-4.

47. Miller, R.A., Li, X., and Garcia, G. (2023). Aging rate indicators: speedometers for aging research in mice. Aging Biol. *1*. https://doi.org/10.59368/agingbio.20230003.

48. Moqri, M., Herzog, C., Poganik, J.R., Biomarkers of Aging Consortium, Justice, J., Belsky, D.W., Higgins-Chen, A., Moskalev, A., Fuellen, G., Cohen, A.A., et al. (2023). Biomarkers of aging for the identification and evaluation of longevity interventions. Cell *186*, 3758–3775. https://doi.org/10.1016/j.cell.2023.08.003.

49. Vujosevic, S., Aldington, S.J., Silva, P., Hernández, C., Scanlon, P., Peto, T., and Simó, R. (2020). Screening for diabetic retinopathy: new perspectives and challenges. Lancet Diabetes Endocrinol. *8*, 337–347. https://doi.org/10.1016/S2213-8587(19)30411-5.

50. Zajonc, R.B., Murphy, S.T., and Inglehart, M. (1989). Feeling and facial efference: implications of the vascular theory of emotion. Psychol. Rev. *96*, 395–416. https://doi.org/10.1037/0033-295x.96.3.395.

51. Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M.G., et al. (2019). Mediapipe: A framework for building perception pipelines. Preprint at arXiv. https://doi.org/10.48550/arXiv.1906.08172.

52. Abdrakhmanova, M., Kuzdeuov, A., Jarju, S., Khassanov, Y., Lewis, M., and Varol, H.A. (2021). Speakingfaces: A large-scale multimodal dataset of voice commands with visual and thermal video streams. Sensors (Basel) *21*, 3465. https://doi.org/10.3390/s21103465.

53. Deng, J., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In IEEE conference on computer vision and pattern recognition, pp. 248–255.

54. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

55. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. (Camb) *2*, 100141. https://doi.org/10.1016/j.xinn.2021.100141.

56. Leek, J.T., ohnson, W.E., Parker, H.S., Fertig, E.J., Jaffe, A.E., Storey, J.D., Zhang, Y., and Torres, L.C. (2019). sva: Surrogate variable analysis. R package, 882–883. version 3.

57. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504. https://doi.org/10.1101/gr.1239303.

58. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

59. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. *37*, 907–915. https://doi.org/10.1038/s41587-019-0201-4.

60. Robin, X., Turck, N., and Hainard, A. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics *12*, 1–8.

61. Hastie, T., and Qian, J. (2014). Glmnet vignette (GitHub), pp. 1–30.

62. Millstein, J., Chen, G.K., and Breton, C.V. (2016). Cit: hypothesis testing software for mediation analysis in genomic applications. Bioinformatics *32*, 2364–2365. https://doi.org/10.1093/bioinformatics/btw135.

63. Williams, J.R. (2008). The Declaration of Helsinki and public health. Bull. World Health Organ. *86*, 650–652. https://doi.org/10.2471/blt.08.050955.

64. He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.

65. Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. Preprint at arXiv. https://doi.org/10.48550/arXiv.1409.1556.

66. Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015. p. 1–9.

67. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. *7*, 562–578. https://doi.org/10.1038/nprot.2012.016.

68. Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118–127. https://doi.org/10.1093/biostatistics/kxj037.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Ficoll-Paque PLUS	Cytiva	17144002
TRIzol Reagent	Invitrogen	15596018CN
Critical commercial assays		
DNF-471 STANDARD SENSITIVITY RNA ANALYSIS KIT	AATI	DNF-471-0500
Optimal Dual-mode mRNA Library Prep Kit	BGI	LR00R96
Qubit ssDNA Assay Kit	Invitrogen	Q10212
DNBSEQ-T7RS High-throughput Sequencing Set (FCL PE150)	MGI	1000028454; RRID:SCR_024847
Deposited data		
PBMC RNA-seq data	This paper	Project ID: PRJCA025703; OMIX ID: OMIX006341
Software and algorithms		
Python (version 3.6)	Python Software Foundation	https://www.python.org/; RRID:SCR_008394
ThermoFace	This paper	https://github.com/JackieHanLab/ThermoFace;
Mediapipe	Lugaresi et al.[51]	https://github.com/google/mediapipe
MATLAB (version R2010b)	Matlab Software Foundation	http://www.mathworks.com/products/matlab/; RRID:SCR_001622
SmartView (Version 4.3)	Fluke Corporation	https://www.fluke.com/en-us/support/software-downloads/software-for-fluke-infrared-cameras
SpeakingFaces	Abdrakhmanova, et al.[52]	https://github.com/IS2AI/SpeakingFaces
ImageNet	Deng et al.[53]	http://image-net.org/
Mfuzz	Kumar and E Futschik[33]	http://mfuzz.sysbiolab.eu/; RRID:SCR_000523
scikit-learn (version 1.1.1)	N/A	https://scikit-learn.org/; RRID:SCR_002577
STAR (version 2.4.0)	Dobin et al.[54]	https://code.google.com/archive/p/rna-star/; RRID:SCR_004463
R (version 4.1.1)	R Core Team	https://www.r-project.org/; RRID:SCR_001905
pheatmap (version 1.0.12)	N/A	https://github.com/raivokolde/pheatmap; RRID:SCR_016418
clusterProfiler(version 4.2.2)	Wu et al.[55]	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html; RRID:SCR_016884
sva (version 3.42.0)	Leek et al.[56]	https://bioconductor.org/packages/release/bioc/html/sva.html; RRID:SCR_012836
Cytoscape (version 3.8.2)	Shannon et al.[57]	https://cytoscape.org/; RRID:SCR_015784
FeatureCounts(version 2.0.3)	Liao et al.[58]	https://subread.sourceforge.net/featureCounts.html;RRID:SCR_012919
HISAT2 (version 2.1.0)	Kim et al.[59]	https://daehwankimlab.github.io/hisat2/manual/; RRID:SCR_015530
Samtools (version 1.14)	N/A	http://samtools.sourceforge.net/; RRID:SCR_002105
mixOmics (version 6.18.1)	N/A	https://www.bioconductor.org/packages/release/bioc/html/mixOmics.html; RRID:SCR_016889
pROC (version 1.18.5)	Robin et al.[60]	https://cran.r-project.org/web/packages/pROC/index.html; RRID:SCR_024286
pls (version 2.8-2)	Wold et al.[34]	https://github.com/khliland/pls
glmnet (version 4.1-7)	Hastie et al.[61]	https://cran.r-project.org/web/packages/glmnet/index.html; RRID:SCR_015505

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
cit	Millstein et al.[62]	https://cran.r-project.org/web/packages/cit/index.html
Other		
Fluke Ti401pro	Fluke Corporation	https://www.fluke-direct.com/
3dMDface System	3dMD	https://www.3dmd.com/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for reagents should be directed and will be fulfilled by the lead contact (Jing-Dong J. Han, jackie.han@pku.edu.cn).

Materials availability

The study did not generate new unique reagents.

Data and code availability

- All mapped and batch corrected RNA-seq reads in this study are uploaded in CNCB (https://www.cncb.ac.cn/ or https://ngdc.cncb.ac.cn/bioproject/). Project: PRJCA025703, OMIX ID: OMIX006341.
- The package is available at https://github.com/JackieHanLab/ThermoFace.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Participants for this study were recruited from multiple cohorts and were carried out according to the guidelines of the Declaration of Helsinki[63]:

1. Jidong Cohort: Three batches were recruited from 2020 to 2022 during annual physical examinations at the Staff Hospital of Jidong Oil-field and Kailuan Majiagou Hospital. The cohort comprises individuals of both sexes, ranging in age from 21 to 88 years. The cohort consists of 1,339 females and 1,472 males (refer to Figure 1A, S1A, and S1B; Table S1).
2. Beijing Cohort: 75 healthy individuals were recruited from Guang'anmen Hospital, with ages ranging from 18 to 69 years old (refer to Figure 1A).
3. PBMC Cohort: 57 healthy individuals were recruited from Guang'anmen Hospital, with ages ranging from 25 to 78 years old (refer to Figure 1A).
4. Exercise Cohort: 30 healthy individuals were recruited from Hainan Medical University, with ages ranging from 18 to 49 years old, 7 in control group, 23 in jump rope group (refer to Figure 1A).

All participants included in this study are of Han Chinese ethnicity.

Recruitment for all cohorts was conducted through voluntary advertisement, with participants required to be in a non-emergency state, self-reporting as healthy, and without major chronic diseases.

Ethical approval for each cohort was obtained from respective hospitals or university.

METHOD DETAILS

Data acquisition

We collected Thermal Infrared facial images of the Han Chinese cohort at Tangshan in Heibei province (Jidong cohort) and Beijing (Beijing cohort and PBMC cohort) using the Fluke Ti401pro (640 x 480 resolution, Thermal sensitivity $\leq$0.075 °C at 30 °C target temp) (https://www.fluke-direct.com/). The Jidong cohort was recruited in Staff Hospital of Jidong oil-field and Kailuan Majiagou Hospital during their annual physical examination. The Beijing cohort and PBMC cohort were recruited in Guang'anmen Hospital. Ambient temperature (Te) at the time of image acquisition was recorded by the camera. As a requirement, volunteers sat straight and looked forward to the Thermal image device for 1 minute, without any facial expression, for the Thermal Infrared facial images to be taken. The Fluke system returned IS2-formatted images. Also, for each participant, baseline information including age, sex, education level, anthropometry, and lifestyle factors (for example, frequency of smoking, alcohol drinking, diet, and exercise) was collected from questionnaires, weight, height, and blood pressure was measured and routine blood tests were done.

At the same time, we collected 3D facial images of the same volunteers using the 3dMDface System (https://3dmd.com/). Acquisition requirements are the same as for IR. The 3dMDface system returned OBJ-formatted 3D facial surfaces with point clouds and corresponding texture images. The detailed procedure of photography capture has been described in our previous work.[23]

The standard for collecting and inspecting results of infrared facial thermal imaging data included: 1. Once the infrared camera was set up and the imaging system was available, the camera was not moved again; 2. The imaging location was as spacious as possible and the test volunteer should not lean against the wall; 3. During the image acquisition, only the test volunteer was present; if there were multiple people, the data was invalid and the facial data of the test volunteer in the acquired image was more than 35% of the frame; 4. Maintained adequate indoor lighting and avoided taking pictures with backlight. If the external sunlight was bright, curtains were pulled. Visible light data without clear facial features of the test subject were considered invalid. 5. Ensured that the time fluctuation of each image acquisition was minimized. Recorded different collection time periods in the table using the 24-hour format, such as 2021.04.03.09:34.06. Maintained a constant temperature environment (between 20-25 °C with an air conditioner) and ensured that the test subject was not directly facing the air conditioner. Recorded the environmental temperature before each shoot and adjusted the imager's background temperature accordingly. 7. Prior to image collection, ensured that test volunteers refrained from intense activities or consuming beverages. Documented volunteers with fever or makeup. Allowed at least 15 minutes of buffer time for the test subject in cases of significant indoor-outdoor temperature differences. 8. Directed the test subject to maintain a neutral facial expression, looking straight into the lens without smiling, tilting, or opening their mouth. Ensured consistent distance between each test subject and the lens. 9. Conducted laser calibration before each sampling batch, proceeding with image capture only after aligning the visible light and infrared contours of the object. 10. Activated the fill light option during data collection to prevent facial backlighting. 11. Immediately reviewed collected visible light images to ensure clear facial recognition. Retook images lacking clear facial features. 12. Ensured the test volunteer's forehead was fully exposed, providing a hair band for those with longer bangs to lift their hair. No glasses were allowed.

When analyzing the data, we excluded onsite drinking, smoking or makeup to ensure that our result is not influenced by transient states. Also, facial images with obvious facial hair were excluded from further analyses.

Exercise cohort data acquisition

Volunteer of exercise cohort was recruited under the guidance of the standard for collecting and inspecting results of infrared facial thermal imaging data. When collecting exercise cohort data, the volunteers were randomly assigned to "Control group" and "Jump Rope Group". While the control group was not required to do anything to alter their lifestyle, the jump rope group's volunteers were required to adhere to the daily jump rope training of more than 10 minutes and no less than 800 jump times during the time between 5:30 pm and 6:30 pm for two weeks, and volunteers were encouraged to take the initiative to increase the training minutes as well as the capacity after adaptation. After two weeks, both the control and the jump rope group were recalled for thermal image taking.

Data structure of IS2

The thermal infrared images were captured and stored in IS2 format. It could be opened by SmartView 4.3 software to view the image quality directly. The uncompressed folder contained IR.data and visible images, one of which was in high resolution. IR.data contained the raw data collected by the thermal infrared CCD, which needed to be corrected for Tc (Original temperature of CCD signal) because the transmittance and background temperature had an important effect on the accuracy of the thermal imaging camera.

The corrected temperature Tb was

$$Tb = e * t * Tc + (2 - e - t) * Te$$

where Tc was original temperature of CCD signal, *e* was the Efficiency, t was the Transmission, and Te was the ambient temperature at the time of image shooting (°C), calculated in batch using MATLAB R2010b, and the temperature matrix was output in the range of 27-38 degrees Celsius and saved in PNG format.

Thermal imaging noise removal

Photonic noise exists in the photosensitive material of the IR detector and needed to be removed. Fluke Ti401pro captured images exist pretzel noise, which means the noise was small and had a clear boundary with the surrounding points. The image was filtered by the median filtering method, using cv2.medianBlur from the opencv2 (Python) package.

Landmark recognition

Detection of facial feature points (landmarks) and accurate standardization of facial regions provided the framework for the construction of facial region delineation for the face network. We used the manually labeled feature points dataset provided by the GitHub project Speaking Faces (based on python 3.6)[52] and based on thermal face labeled feature point the functions in dlib package (based on python 3.6) dlib.simple_object_detector() and dlib.shape_predictor() was trained firstly to generate an SVM thermal face 54 feature points detection model that is used to detect 54 feature landmarks on our thermal facial images as shown in Figure 1A.

Visible image facial raster network construction

To limit information loss and retain most of the spatial structure, the region segmentation of the face was performed using the face mesh building part of Google Mediapipe.[51] MediaPipe Face Mesh provided a face geometry segmentation method that can estimate

468 3D face feature locations in real time from 2D visible photos. Machine learning (ML) was used to infer the 3D surface geometry and only a 2D camera input is required. From MediaPipe Face Mesh we obtain a 468-point FaceMesh for each visible image. Mediapipe used mean absolute distance between the ground truth vertex locations and the predictions, normalized by interocular distance (IOD), defined as the distance between the eye centers (estimated as midpoints of eye corner connecting segments to avoid gaze direction dependence). The estimated IOD mean absolute distance error of FaceMesh was 2.56%.

Infrared and visible feature points registration

The construction of Face Mesh could only be based on the visible face images, the collection of facial infrared images and facial visible images Spatio-temporal overlap, in order to achieve infrared facial region Face Mesh segmentation, we developed a method to match visible feature points and thermal infrared feature points as the following. First, the corresponding feature points were extracted in FaceMesh for thermal infrared detection, and 10 points in the thermal image were extracted (points 50, 49, 51, 52, 37, 40, 43, 46, 1, 17, 31, 34, 9, 13, 5) po={(x_{1o},y_{1o}),(x_{2o},y_{2o}),…,(x_{10o},y_{10o})}, corresponding to the points p_c={(x_{1c},y_{1c}),(x_{2c},y_{2c}),…,(x_{10c},y_{10c})} in FaceMesh. Then the *Procrustes* transformation was applied to obtain the corresponding face mesh of the thermal face. *Procrustes* transformed by iteration such that:

$$\underset{s,\theta,t}{\operatorname{argmin}} \sum_{i=1}^{n} \left\| s \cdot \boldsymbol{R} \cdot {p_{io}}^{T} + \boldsymbol{t} - p_{ic} \right\|^2$$

where, $\mathbf{R} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$. The other points detected by FaceMesh are transformed correspondingly according to the obtained Procrustes transformation matrix to obtain the infrared light region feature points of ThermoFace.

Delaunay mesh construction

To cut the thermal imaging facial area, the Delaunay triangle network was constructed based on 468 feature points:

a) Computed the enclosing box of the region and generated two hyper-triangles as the initial triangular mesh. Generated boundary points based on the boundary curve.
b) Inserted the boundary points into the triangular mesh one by one. After each insertion step, a new Delaunay triangular mesh was obtained using the Bowyer-Watson algorithm.
c) Removed the auxiliary vertices inserted in the first step (and also removed the triangles connected to them) to obtain a Delaunay triangulated mesh with respect to all boundary points (i.e., Delaunay triangulation of the boundary points).
d) For the triangles inside the region, the center of gravity (centroid) was inserted into the triangle mesh according to certain conditions and adjusted according to the Bowyer-Watson algorithm to obtain a new Delaunay triangle profile. Until there were no more center of gravity points to be inserted, at which point the complete Delaunay triangulated mesh of the region was obtained.

By the above procedure, 897 triangular regions of the thermal image were obtained, and the maximum value ($TMax_1$,$TMax_2$,…, $TMax_n$), minimum value($TMin_1$,$TMin_2$,…,$TMin_n$)and standard deviation (TSD_1,TSD_2,…,TSD_n)of the temperature were extracted by cutting the facial regions one by one using opencv2.

Fuzzy C-means clustering for ThermoFace

ThermoFace regions of samples separated into 6 age groups were grouped into different clusters using Mfuzz package in R with fuzzy c-means algorithm.[33] Age groups were used as time stages.

Visualization of ThermoFace projection

The value of each area (e.g., maximum or minimum temperature, coefficients intensity) was linearly scaled to 10-255 by

$$V_p = y(V_T) = \alpha V_T + b$$

where V_T was temperature value, V_p was Image corresponding color value, α was the scaling factor and b was the scaling intercept, and setting the original data 0 value was the midpoint. Then picking the color corresponding to V_p from the reference file ('jet', 'Rdbu', or 'seismic', from Matplotlib, python) to fill the standard ThermoFace surface mesh.

MAD Definition

In this study, the mean absolute deviation (MAD) was employed as the evaluation metric for the age model. MAD represented the average absolute difference between the predicted age and the actual age for all samples within a model,

$$MAD = \frac{1}{n} \sum_{i=1}^{n} |x_i - f(x_i)|$$

Where n was the number of samples, x_i was the actual age value, and $f(x_i)$ was the predicted age.

Cross Validation

To validate the prediction ability for age using ThermoFace matrix, we divided all our 2,811 samples into 10 groups randomly for 10-fold cross-validation. The model was trained with nine of the ten groups as a training set, a new linear mode was built based on training data and tested in the one not used for training data. MAD was the average MAD of each model for 10 cross-validation.

CNN model

During each evaluation process, the model was trained with eight of the ten groups as a training set, hyperparameters were determined with one of the remaining groups as the validation set and the predictive performance was tested on the last group as the test set. Finally, we used all ten models generated from the 10 cross-validation to test the datasets independently, and used the average performance level as the final prediction capability.

Chronological age

Birth dates were obtained by the Chinese-government-issued official Resident Identity Card. Age was rounded to the nearest year for model training.

PLSR based age predictor

The 'pls' package[34] in R was used for PLS regression. Both chronological age and ThermoFace matrix (z-score normalized within samples) were combined into one matrix as the input. The 10-fold cross-validation (CV) method was applied to obtain the age prediction of each sample. MAD was calculated by the difference between predicted age and chronological age. Saturation analysis was conducted by randomly selecting a designated percentage of all the samples with the same age distribution of all samples and training the PLS model on such samples, and then was evaluated using both MAD and PCC. Components number used was determined by examining RMSE calculated in each PLSR model which was also verified by 10-fold CV.

Elastic Net based age predictor

The 'glmnet' package[61] in R was used for Elastic Net regression. Input to train was as same as PLSR based age predictor. Saturation analysis was conducted as well. Best alpha and lambda parameters were determined by EN model using all the samples as training data and applied to CV models.

Preparation of ThermoFace faces for deep learning

The affine transformation (Affine) was performed first. The equation was shown below, where the matrix composed of a,b,c,d was the rotation matrix, t_x,t_y was the translation transformation, x,y were the matrix elements before the transformation, and x_t,y_t was the matrix elements after the transformation.

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$

The affine transformation was applied to the image using the OpenCV.warpAffine function. Since warpAffine could only perform rectangular transformations, it found the minimum outer rectangular border in each face mesh triangle, cropped the rectangular area from the source image, and applied the affine transform to the cropped image to obtain the output image.

To create the output triangle mask, which preserved the color of all pixels inside the triangle, face mesh standard surface mesh map as input to 'affine' for it to transform all the samples corresponding to the triangles and the temperature in their regions, match the corresponding regions of the face mesh template. Repeated the same number of times as the number of triangles to obtain the normalized facial temperature distribution results independent of the facial structure morphology.

The resulting images were cut according to the top and bottom boundary points and subsequently resized to 224 × 224 pixels before using deep learning.

Data augmentation

During training of deep CNN models, we flipped all samples from left to right, which could increase the amount of training data.

Deep CNNs

We combined 3 CNN architectures, GoogleNet, VGG and ResNet,[64–66] in Python to predict age. For all three CNNs, we used one node as the last layer for regression.

Ensemble of deep CNNs

We used an ensemble of three CNNs, GoogleNet, VGG with 16 layers and ResNet with 50 layers to predict age. We trained and generated the results of the three networks independently, and generated an ensemble by taking their average, with the formula

$$y = (\text{GoogleNet} + \text{VGG} + \text{ResNet})/3$$

where *GoogleNet*, *VGG* and *ResNet* represented the ages predicted by these three CNNs, respectively, and *y* represented the ensemble predicted age.

Age prediction using 3D facial images with CNN model

The prediction of age with CNN had been described in detail in previous work.[23]

AgeDiff correction

To study age-independent associations of AgeDiff, we corrected AgeDiff by fitting to a polynomial model to age as follows, with span 0.8 and degree 2:

$$\mathrm{cAgeDiff} = \mathrm{AgeDiff} - \mathrm{loess}(\mathrm{AgeDiff} \sim \mathrm{Age})$$

cAgeDiffs association with health parameters

cAgeDiffs Pearson's correlation coefficient and corresponding p-values to each health parameter were calculated. Benjamini–Hochberg correction was used for FDR calculation. R package 'pheatmap' was used for hierarchical clustering of correlations.

Correlation between cAgeDiff SD and chronological age

A non-overlapping sliding-window approach was used to study the relationship between cAgeDiff SD and age. Samples were sorted by age, and then a non-overlapping sliding window with bin size 25 was used to group samples and calculate mean of age and SD of cAgeDiffs in each sliding window.

Amplification, normalization, and sequencing of RNA

Each participant had 10 ml of total fasting blood drawn in the early morning. Venous blood samples in EDTA-K3 anticoagulated tubes were obtained from the subjects. Peripheral blood mononuclear cells (PBMCs) were isolated by Ficoll density gradient centrifugation (Ficoll-Paque PLUS, Cytiva). RNA was extracted using TRIzol (Invitrogen), following the manufacturer's instructions. The DNF-471 STANDARD SENSITIVITY RNA ANALYSIS KIT was used to prepare samples with RNA integrity numbers (RIN) greater than 8 before they were sequenced on a MGIDNBSEQ-T7 device in accordance with the manufacturer's recommendations. Using STAR(version 2.4.0), paired-end reads (126 bp) were mapped to the GRCh38 reference human genome.[54]

RNA normalization

Only reads with a unique match to the reference genome were kept. GENCODE v43 annotation was used to quantify the protein-coding genes. Cufflinks carried out the quantification and normalization.[67]

Ribominus RNA-seq of the PBMC samples

The raw clean data was first aligned to GRCh38 genome using Hisat2 (version 2.2.1)[59] with the exclusion of reads that are flagged as poor quality based on the QSEQ filter. Fragment counting and gene annotation were done by using featureCounts (version 2.0.3)[58] against GENCODE v43. Read counts were then transformed into transcripts per million (TPM). Then ComBat function from the 'SVA' R package[68] was used to mitigate any batch-related discrepancies between two batches.

KEGG pathway association enrichment analysis

We first filtered the transcript expression matrix in TPM (transcripts per million) to retain only transcripts with log(TPM+1)>1 in at least 3 samples and then ranked the transcripts based on the Spearman correlation coefficient (RCC) between cAgediff and the log(TPM+1) values of transcripts. ResKEGG function from the clusterprofiler package in R[55] was used to perform KEGG pathway enrichment analysis on the filtered transcripts. Only pathways significantly associated with 3 thermal cAgediffs and the average of them (p-value < 0.05) were retained as significantly associated with thermal cAgeDiffs, which were further corrected for age within the PBMC cohort to ensure their independence to age.

SleepTime grading

Sleep time was divided into 5 grades according to the length of time. Grade 1-5 represent for sleep time less than 4 hours, 4∼5.9 hours, 6∼6.9 hours, 7∼7.9 hours and 8∼9.9 hours, respectively.

Yogurt intake grading

Yogurt intake habit was divided into 4 grades according to frequency. Grade 1-4 represented for hardly, occasionally, often and almost daily, respectively.

PLS-DA disease prediction

ThermoFace Matrix data and PLS-DA in 'mixOmics' R package were used for constructing disease predictor. The classification label of the disease was based on the clinically diagnosed disease from the questionnaire, with the classification of mild and severe fatty liver based on the ultrasound findings of "expected mild fatty liver" and "expected severe fatty liver", respectively. The corresponding data for the healthy population were those who did not have hypertension, diabetes, liver cyst, fatty liver, cerebrovascular disease and cardiovascular disease, and their age distribution is the same with the disease population. Similarly, we split the data 10 folds, cross-validate, and took the data as a test set of results to represent the results given by the disease predictor to draw ROC curves

and confidence interval tests. The classification performance of the PLS-DA model was evaluated by the perf function using 5-fold cross-validation with 10 repetitions. The number of repetitions was necessary to ensure a good estimate of the classification error rate. Based on the max.dist performance results, we could decide the ncomp(component number) value of the final PLS model.

Testing the effect of medication taking on predictors

To test if medicine intake would influence the accuracy of ThermoFace disease detector, we applied roc.test in R package 'pROC'. Patients with and without medicine were separated into two groups, with the same age distribution of healthy control to redraw ROC curve. The curves of the same disease with or without medicine taking were combined to test for the significance of difference between disease and control. To calculate corresponding *P value*s, 'bootstrap' method was used with boot.n=200.

Projection of features on the ThermoFace mesh and average 3D facial image

With the 'pls' package in R, the disease or the ThermoFace PLSR cAgeDiff was treated as the dependent variable, while the ThermoFace matrices or 3D facial images were treated as independent variables for PLS regression. The loadings of the first two components weighed by the score of each component were combined and projected to the standard surface mesh feature map or average 3D facial image synthesized from the whole cohort.

Questionnaire parameters filter

We collected medical examination data and questionnaires of lifestyles. In Figures S9 and S10, we first filtered all parameters for all samples, filtering out parameters with null values in 30% of the samples, followed by filtering the samples with null values in 20% of the parameters, and finally comparing the parameter differences for each row of the remaining samples, the parameters with samples with a difference in parameters fewer than 10% of total samples were filtered out. 432 male and 546 female samples remained. Because some questionnaire information is missing for Batch2022, we had to split the data into two, in Figures S7B and S7C, the data from Batch2020 and Batch2021 were used for filtering, but in Figure S8, only the data from Batch2022 were used, so differences exist in Figures S7B, S7C, and S8 parameters. 560 males and 173 females from Batch2022 remained in Batch2022 in Figure S8.

Lifestyle bias comparison of predicted patient groups

Wilcox.test was used for pairwise comparison of the significance of the differences in lifestyle habits for the four predicted outcomes (DD, DH, HD, HH). Benjamini-Hochberg correction was used for FDR calculation. $\log_2$FC=$\log_2$(treatLifestyleFre)-$\log_2$(conLifestyleFre), treatLifestyleFre and conLifestyleFre were the mean of the treated and control, where treatLifestyleFre was the treated lifestyle frequencies, conLifestyleFre was the control lifestyle frequencies.

Lifestyle–regulator–cAgeDiff causal-relationship inference

The causal-inference test was adopted with the R package 'cit'[62]. Details in our previous paper. A causal inference met the following four criteria: (1) lifestyle and cAgeDiff were correlated; (2) lifestyle was associated with regulator after adjusting for cAgeDiff; (3) regulator was associated with cAgeDiff after adjusting for lifestyle; and (4) lifestyle was independent of cAgeDiff after adjusting for regulator. To summarize the p-values of the four tests, the intersection-union test was used as the p-value for the whole causal-inference test. Significant results with p-value< 0.05 and FDR < 0.1 (calculated by the 'cit' package[62]) were retained. The network was visualized by Cytoscape (version 3.5.0).

Datasource for Thermoface details

ThermoFace was used for thermal facial image extraction and ThermoFace Matrix construction. SpeakingFaces (https://github.com/IS2AI/SpeakingFaces) was used for thermal face 56 feature points SVM predictor building. ImageNet (http://image-net.org/) was used for CNN pre-train. Thermal images, 3D images, and other personal identification sensitive meta data could not be publicized or shared according to our participant consent agreement.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were performed with R version 4.1.1, using two-tailed Student's t-tests or Wilcoxon tests as appropriate. Pearson correlation and Spearman rank-correlation coefficient were used to measure the correlations. All of the statistical tests performed were indicated in the figure legends. The data were presented as mean or individual points with box plots or violin plots that display the median and quartiles. Error bars represented standard error intervals around the mean, as specified in the figure legends.