ARTICLES
https://doi.org/10.1038/s42255-020-00270-x

nature metabolism

Check for updates

Three-dimensional facial-image analysis to predict heterogeneity of the human ageing rate and the impact of lifestyle

Xian Xia[1,2,3,16], Xingwei Chen[1,2,3,16], Gang Wu[1], Fang Li[1], Yiyang Wang[1,2,3], Yang Chen[1,2,3], Mingxu Chen[1,2,3], Xinyu Wang[1,2,4], Weiyang Chen[1], Bo Xian[1], Weizhong Chen[1], Yaqiang Cao[1], Chi Xu[1], Wenxuan Gong[1,2,3], Guoyu Chen[1,2,3], Donghong Cai[1,2,3], Wenxin Wei[5], Yizhen Yan[1,2,3], Kangping Liu[2], Nan Qiao[6], Xiaohui Zhao[6], Jin Jia[6], Wei Wang[7], Brian K. Kennedy[8,9,10,11], Kang Zhang[12], Carlo V. Cannistraci[13,14], Yong Zhou[15] ✉ and Jing-Dong J. Han[1,2] ✉

Not all individuals age at the same rate. Methods such as the 'methylation clock' are invasive, rely on expensive assays of tissue samples and infer the ageing rate by training on chronological age, which is used as a reference for prediction errors. Here, we develop models based on convoluted neural networks through training on non-invasive three-dimensional (3D) facial images of approximately 5,000 Han Chinese individuals that achieve an average difference between chronological or perceived age and predicted age of ±2.8 and 2.9 yr, respectively. We further profile blood transcriptomes from 280 individuals and infer the molecular regulators mediating the impact of lifestyle on the facial-ageing rate through a causal-inference model. These relationships have been deposited and visualized in the Human Blood Gene Expression—3D Facial Image (HuB-Fi) database. Overall, we find that humans age at different rates both in the blood and in the face, but do so coherently and with heterogeneity peaking at middle age. Our study provides an example of how artificial intelligence can be leveraged to determine the perceived age of humans as a marker of biological age, while no longer relying on prediction errors of chronological age, and to estimate the heterogeneity of ageing rates within a population.

Ageing is a major risk factor for many complex human diseases. Rates of biological ageing are highly variable among individuals of the same chronological age. To quantify ageing, ideally, biological age rather than chronological age should be assessed; however, because of the lack of a gold standard for biological age, the population average of chronological age is often assumed to be the standard of biological age. The deviators or outliers of the average or standard curve are defined and often substantiated as fast or slow agers by other physiological or molecular parameters. On the basis of this assumption, many methods have been developed to quantify ageing. For example, predicting age using the transcriptome in human peripheral blood has a mean absolute difference (MAD) between chronological age and predicted biological age of 7.8 yr (ref. [1]), using the proteome has a Pearson correlation coefficient (PCC) of 0.93–0.97 (ref. [2]) and using the DNA methylome has a MAD of 4.9 yr in whole blood from 656 human cohorts[3] and 3.6 yr in heterogeneous tissues[4]. However, because the transcriptome, DNA methylome and proteome have to be measured in blood cells or other tissues, invasiveness and high costs preclude their application to large-scale screens and routine physical examinations.

Facial images have been used in traditional Chinese medicine as a major diagnosis tool for evaluating health and disease status. This practice originated more than 2,000 years ago[5], and is increasingly being used by Western-medicine clinicians to diagnose developmental syndromes[6]. Our previous work has established 3D human facial image as an ageing marker, generated a partial least squares regression (PLSR) model and used the difference between chronological and predicted age (AgeDiff) to identify outliers in the ageing rate[7].

Many biological pathways are known to be associated with ageing or age[8], but no biological process has been directly linked to differential ageing rates (or AgeDiff) across individuals. Indeed, the ageing rate might be associated with a subset of ageing-related

[1]CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences Center for Excellence in Molecular Cell Science, Collaborative Innovation Center for Genetics and Developmental Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China. [2]Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Center for Quantitative Biology (CQB), Peking University, Beijing, China. [3]University of Chinese Academy of Sciences, Beijing, China. [4]School of Life Science and Technology, ShanghaiTech University, Shanghai, China. [5]Department of Hepatic Surgery, Eastern Hepatobiliary Surgery Hospital, Second Military Medical University, Shanghai, China. [6]Accenture China Artificial Intelligence Lab, Shenzhen, China. [7]School of Medical and Health Sciences, Edith Cowan University, Perth, Western Australia, Australia. [8]Departments of Biochemistry and Physiology, National University of Singapore, Singapore, Singapore. [9]Centre for Healthy Ageing, National University Health System, Singapore, Singapore. [10]Singapore Institute for Clinical Sciences, A*STAR, Singapore, Singapore. [11]Buck Institute for Research on Aging, Novato, CA, USA. [12]Faculty of Medicine, Macau University of Science and Technology, Macau, China. [13]Biomedical Cybernetics Group, Biotechnology Center (BIOTEC), Center for Molecular and Cellular Bioengineering (CMCB), Center for Systems Biology Dresden (CSBD), Cluster of Excellence Physics of Life (PoL), Department of Physics, Technische Universität Dresden, Dresden, Germany. [14]Center for Complex Network Intelligence (CCNI) at the Tsinghua Laboratory of Brain and Intelligence (THBI) and Department of Bioengineering, Tsinghua University, Beijing, China. [15]Clinical Research Institute, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. [16]These authors contributed equally: Xian Xia, Xingwei Chen. ✉e-mail: yongzhou78214@163.com; jackie.han@pku.edu.cn

changes that are more modifiable by environment, diet, lifestyle or variations in genetic background across different individuals than the genetically hardcoded programme that is fatefully played out for a species. Thus, an understanding of the molecular events contributing to AgeDiff is actionable and pertinent to understanding healthy ageing.

Recently, deep CNNs have been successfully used in the classification of Alzheimer's disease[9] and skin cancer[10]. Benefiting from these highly efficient CNN architectures, we trained three classic CNNs (Inception, Visual Geometry Group (VGG) and ResNet) using the large datasets in ImageNet. We then applied them to 3D facial images from a large cohort of 4,719 individuals so they could learn age by transfer learning. In combining the results from the 3 CNNs, the accuracy of age and perceived-age prediction reached a MAD of 2.79 and 2.90 yr, respectively. This allowed us to examine ageing-rate heterogeneity across individuals during ageing, which was otherwise missed by our previous linear PLSR model. Our CNN for perceived-age prediction is designed to learn a representation of biological age instead of chronological age, which is supported by its detection of much-stronger associations of lifestyle and health parameters with ageing than those found by a chronological-age predictor. We used a statistical test[11], coupled with peripheral blood mononuclear cell (PBMC) transcriptome analysis, and built a causal-inference network that reveals circulation factors mediating lifestyle impacts on facial morphologies from blood transcriptomes. We deposited and visualized the associations and inferred causal relationships in the HuB-Fi (http://www.picb.ac.cn/hanlab/hub-fi/) database for researchers and the public to explore.

Results

Deep CNN predicts accurate age and perceived age. To construct a highly accurate age-estimating model from 3D facial images, we collected 4,719 facial images from Han Chinese individuals in North China (Jidong) with the 3dMDface System, together with corresponding age, sex, physical-examination and questionnaire information (Supplementary Table 1 in the Supplementary Information). We developed a fully automated pipeline for analysis of 3D facial images, including pose correction, landmark identification and projection of depth and red, green and blue (RGB) colour of a 3D image onto two-dimensional (2D) images with four channels, for face registration and deep learning. We then used these 2D images to train the CNN model to predict chronological age (FaceCnnAge) (Fig. 1a, Extended Data Fig. 1a and Methods). The predictor achieved a MAD of 2.79 yr by 10-fold cross-validation (Methods and Fig. 1b, left), compared with 4.47 yr using a PLS model (FacePlsAge) (Fig. 1b and Extended Data Fig. 1b). Because a CNN trained on chronological age by optimizing prediction accuracy for chronological age might not be optimized to detect biological age, we trained another CNN model to predict perceived age, which might better reflect the health state[12], termed FaceCnnPerceivedAge. This model also achieved a high accuracy, with a MAD of 2.90 yr for perceived age and 4.10 yr for chronological age (Fig. 1b). We then examined the consistency of AgeDiffs or outliers (|AgeDiff| > MAD) given by different models and found that they are highly correlated (Supplementary Table 2) or significantly overlapped ($P < 0.05$), especially between the two CNN models (Extended Data Fig. 1b).

To validate the CNN models in independent cohorts, we predicted the ages of people in 332 images from the 2012 Beijing cohort (Supplementary Table 3; these images have only half the resolution of those in the other cohorts) and in 358 images from the 2015 Beijing cohort (Supplementary Table 4) with the CNN models that learnt from the large Jidong cohort. We achieved a MAD of 3.85 yr for data collected in 2012 and of 3.92 yr for data collected in 2015 for FaceCnnAge, and a MAD of 6.66 yr for 2012 and of 4.12 yr for 2015 for FaceCnnPerceivedAge (Extended Data Fig. 1c). As a comparison, the PLS model achieved MAD = 6.15 and 4.81 yr for the 2012 and 2015 Beijing cohorts, respectively. These results suggest that the CNN models achieve superior accuracy not only over all other linear models by cross-validation in the same cohort but also in at least two other independent cohorts. It should be noted that, due to the small sample sizes of the Beijing cohorts, we could not do the opposite, that is train the CNN models on the Beijing cohorts and validate them on the Jidong cohort.

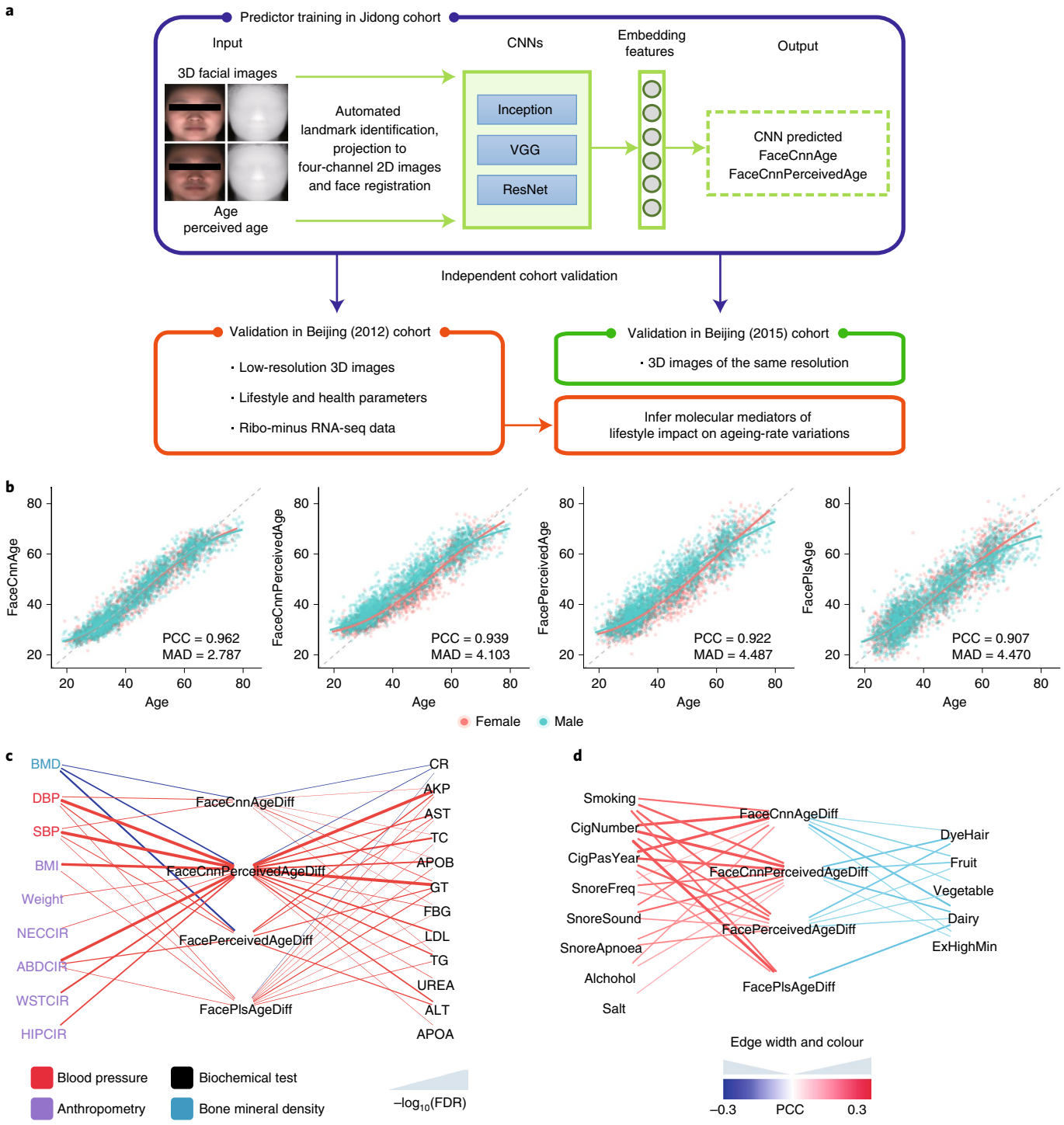
Health and lifestyle parameters associated with AgeDiffs. To examine the biological relevance of AgeDiffs, we next examined the association of AgeDiffs with health parameters corrected for age (false-discovery rate (FDR) < 0.1) in the Jidong cohort. Indices related to obesity, blood pressure, transglutaminase, alkaline phosphatase and cholesterol were among the factors that were most strongly associated with all AgeDiffs, whereas bone mineral density and creatinine were significantly negatively associated with three out of four AgeDiffs. Interestingly, among all four AgeDiffs, FaceCnnPerceivedAgeDiff associated with remarkably more health parameters than other AgeDiffs did (Fig. 1c), indicating that it could be a superior predictor of health, even when compared with the model's training data—human-perceived age.

To uncover the associations of lifestyles factors with differential ageing rates, we determined the association between quantitative lifestyle parameters with AgeDiffs in the Jidong cohort. Smoking, number of cigarettes and years of passive smoking were linked to all four AgeDiffs as the factors that most strongly increased the ageing rate. Frequency of snoring, sound of snoring

Fig. 1 | Accuracy of age predictors, and the health parameters and lifestyles associated with AgeDiffs. a, The large amount of imaging data we collected and the fully automated pipeline for analysis of 3D facial images (landmark identification and face registration) we developed here enabled us to train CNNs to predict chronological and perceived age in a cohort of 4,719 individuals (Jidong) by cross-validation and by validation in two independent cohorts of ~300 individuals each (Beijing 2012 and Beijing 2015). The ImageNet pretrained weights of Inception (Inception_v1), VGG (VGG_16) and ResNet (ResNet_50) were transferred to learn age or perceived age. Average faces of the cohort are used for illustration. Ageing rate, as defined by predicted versus chronological age (AgeDiff), was examined for associations with health and lifestyle in the Jidong cohort. Ribo-minus RNA-seq of blood PBMCs was generated for the Beijing 2012 cohort, and was used to infer the molecular mediators of lifestyle impact on ageing-rate variations. **b**, Correlation of predicted age from four approaches with chronological age. **c**, Association-strength network of 4 AgeDiffs and health parameters, corrected for chronological age, filtered by FDR < 0.1. BMD, bone mineral density; DBP, diastolic blood pressure; SBP, systolic blood pressure; BMI, body mass index; NECCIR, neck circumference; ABDCIR, abdominal circumference; WSTCIR, waist circumference; HIPCIR, hip circumference; CR, creatinine; AKP, alkaline phosphatase; ALT, alanine aminotransfersase; AST, aspartate transaminase; TC, total cholesterol; APOA, apolipoprotein A; APOB, apolipoprotein B; GT, transglutaminase; FBG, fasting blood glucose; LDL, low-density lipoprotein; TG, triglycerides. **d**, Association network between lifestyle and AgeDiffs in the Jidong cohort. Edges are filtered by FDR < 0.1. Smoking, smoker or non-smoker; CigNumber, average number of cigarettes per day; CigPasYear, number of years of passive smoking; SnoreFreq, frequency of snoring per week; SnoreSound, whether the sound of a snore is louder than speaking voice; SnoreApnoea: whether snoring is coupled with apnoea; Alchohol, average intake of alcohol per day; Salt, average intake of salt per day; DyeHair, whether have hair dyed in the last three months; Fruit, frequency of consuming fruits; Vegetable, frequency of consuming vegetables; Dair, frequency of consuming dairy products; ExHighMin, average length of vigorous exercise per day.

and apnoea during snoring were linked to all but FacePlsAgeDiff. Among factors that decreased the ageing rate, dairy intake was linked to all four AgeDiffs. The associations identified by FaceCnnPerceivedAgeDiff essentially captured all those found by FaceCnnAgeDiff or by FacePerceivedAgeDiff, but captured a greater number of associations (Fig. 1d). This suggests that both are optimized to accurately detect the ageing rate, with FaceCnnPerceivedAgeDiff being a better estimate of the biological ageing rate than all other age predictors are, including perceived age itself. Similar associations can be observed in the smaller Beijing 2012 cohort (Extended Data Fig. 2).

AgeDiff heterogeneity peaks at middle age. In contrast to the s.d. expected from randomly guessing a number between 20–85, which is highest at young and old age (Extended Data Fig. 3a), the heterogeneity of the ageing rate precisely captured by FaceCnnAgeDiff and FaceCnnPerceivedAgeDiff peaked at middle age (Fig. 2a). FacePerceivedAgeDiff also peaked at middle age, although it plateaued afterwards in males. FacePlsAgeDiff, however, is very different in that it shows a monotonic increase in s.d. with increasing chronological age and an s.d. that appears random at old age. These patterns are insensitive to the bin size (moving window) (Methods and Extended Data Fig. 3b). Similar patterns can be observed in the



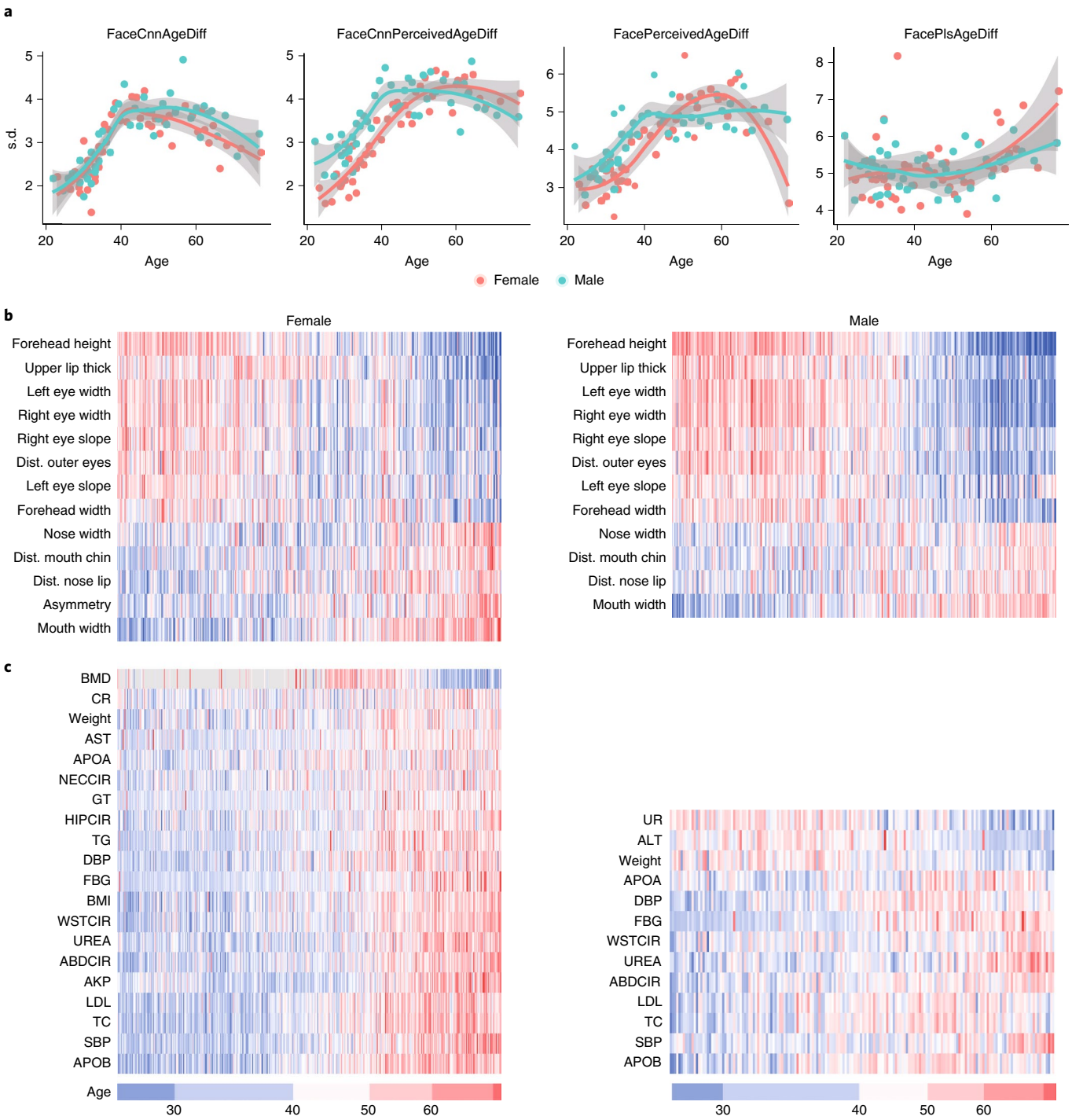


Fig. 2 | AgeDiff heterogeneity peaks at middle age. a, Relationship between age and the s.d. of 4 AgeDiffs with a bin size of 100. Data are presented as mean ± s.d. **b**, Heatmap of facial features in the Jidong cohort sorted by increasing chronological age (PCC with age, FDR < 0.1 and |PCC| > 0.1 in females and in males). Features are ranked by PCC from lowest to highest in females and in males. **c**, A heatmap of health parameters in the Jidong cohort sorted by increasing chronological age (PCC with age, FDR < 0.1 and |PCC| > 0.1). Features are ordered as in **b**. UR, uric acid; see Fig. 1c for all other abbreviations.

smaller Beijing cohort. Using broken-stick regression to pinpoint the breakpoint in the regression line, we found that all four AgeDiffs predominately show a breakpoint peaking at middle age (Extended Data Fig. 4a,b).

Previous findings show that for human brain ageing transcriptomes, unlike those of very young or very old individuals, middle-aged individuals display high heterogeneity[13]. We observed a similar pattern when all samples were visualized by a heatmap of age-related quantitative facial features (Fig. 2b) and age-related health parameters (Fig. 2c) in the Jidong cohort and also in the Beijing 2012 cohort (Extended Data Fig. 3c).

Transcriptomic age is consistent with facial age. Using the cohort that we collected from the Beijing area (Beijing 2012), we extracted and sequenced the ribo-minus RNA from PBMCs of 280 individuals with matching 3D facial images (Supplementary Table 5). We

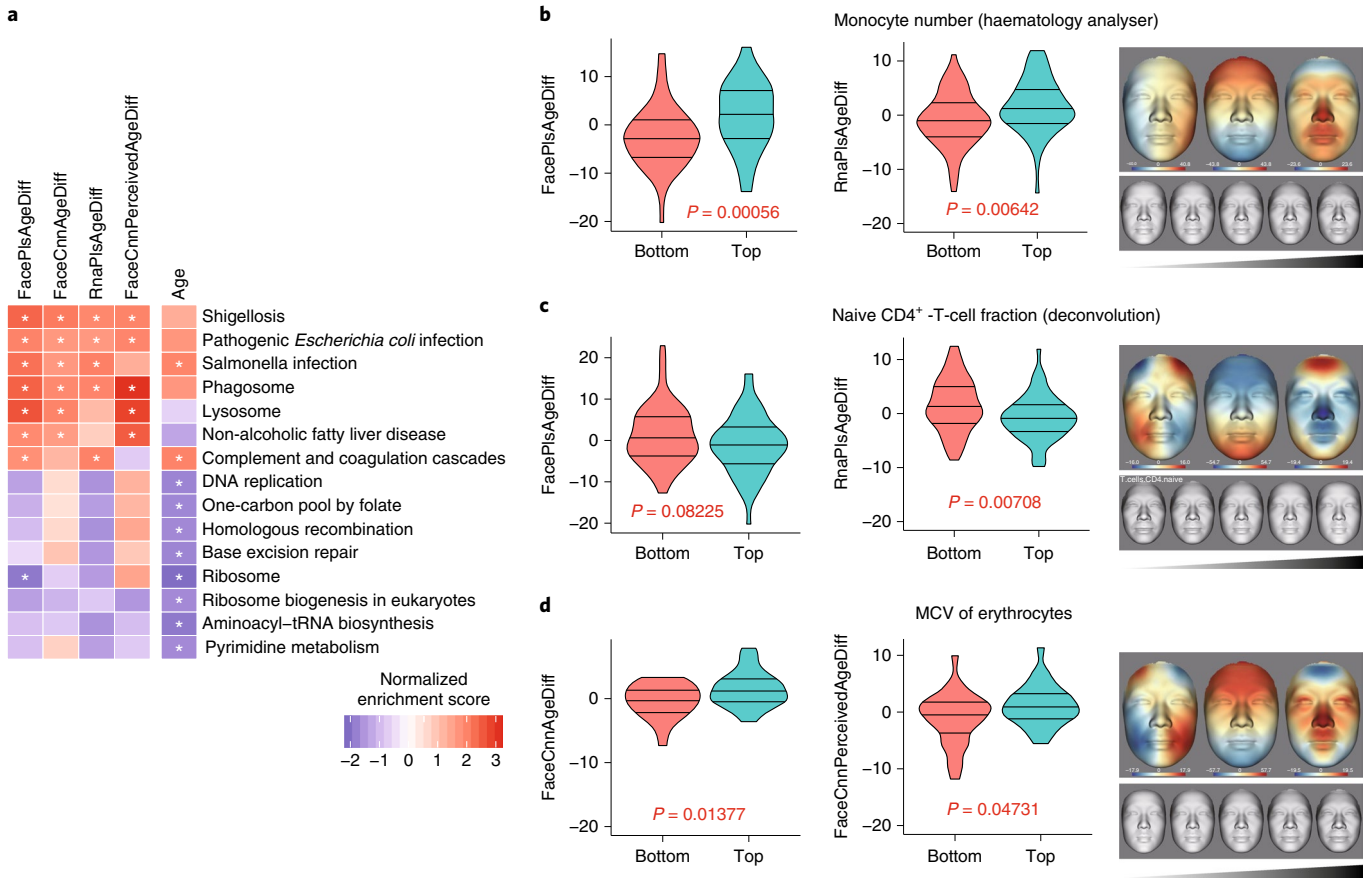


Fig. 3 | Inflammation is related to AgeDiff at the molecular and cellular level. a, KEGG pathway enrichment scores for chronological age and AgeDiffs. Asterisks denote FDR < 0.1. **b**, Distribution of FacePlsAgeDiff (left) or RnaPlsAgeDiff (middle) between the top and bottom 20% of samples on monocyte number derived from haematology analyser (two-sided *t*-test), and the projections on a 3D facial image (right; color scale in arbitrary units). **c**, Distribution of FacePlsAgeDiff (left) or RnaPlsAgeDiff (middle) between the top and bottom 20% of samples on deconvoluted naive CD4+ T cells (two-sided *t*-test), and the projection on 3D facial image (right). **d**, Distribution of FaceCnnAgeDiff (left) or FaceCnnPerceivedAgeDiff (middle) between the top and bottom 10% of samples on MCV (two-sided *t*-test), and the projection on a 3D facial image (right).

performed PLSR analysis on transcriptomes to predict chronological age in the same way that we did for 3D facial images[7]. The MAD of the predicted age (RnaPlsAge) by all transcribed genes and the chronological age was 5.68 yr (Extended Data Fig. 5a–d). Like the s.d. of AgeDiff detected by CNN in both the large and the small cohort, RnaPlsAgeDiff in this small cohort peaked at middle age (around 40–50 yr old) (Extended Data Fig. 3d,e). Similarly, with broken-stick regression, we found that all four AgeDiffs predominately showed a breakpoint peaking at middle age in the Beijing 2012 cohort (Extended Data Fig. 4c,d).

We also applied the FaceCnnAge and FaceCnnPerceivedAge CNN models that learnt from the large Jidong cohort to the 280 samples and compared them with PLSR-derived models FacePlsAge (trained on the 280 3D facial images) and RnaPlsAge. Despite being derived from different machine-learning algorithms, different training sets and different data types, the AgeDiffs predicted by various models significantly correlated with each other, with higher correlation of RNA with CNN-based models than with the facial PLS model (Supplementary Table 6), especially among outliers (Supplementary Table 7). Moreover, the outliers identified by RnaPlsAge significantly overlapped with the two CNN-based models in both fast and slow agers ($P < 0.05$, Extended Data Fig. 5e). This confirms that the outliers detected by analysis of the transcriptomes of PBMCs are reflected in 3D facial features.

Inflammation is related to AgeDiff at the transcriptome level. To identify the blood transcriptome differences associated with differential facial-ageing rates, we compared the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways with gene expression that was correlated with at least two of the four estimated AgeDiffs or with age (Fig. 3a). Among pathways positively correlated with AgeDiffs (FDR < 0.1), most were also positively correlated with age and inflammation, such as those linked to *Escherichia coli*, salmonella infection and shigellosis. And, similarly to age, AgeDiffs tended to be consistently negatively correlated with biogenesis of ribosome and transfer RNA. Although all four AgeDiffs were uncorrelated with chronological age after age correction, the molecular pathways that correlated with age and the four AgeDiffs were largely consistent, indicating that the acceleration or deceleration of ageing detected by AgeDiffs measurements is largely consistent with the ageing process at the pathway level.

We then sought to further verify the AgeDiff–inflammation association by examining whether inflammatory gene sets are positively related to AgeDiffs. Cytokine expression was positively related to FacePlsAgeDiff, RnaPlsAgeDiff and FaceCnnAgeDiff (Kolmogorov–Smirnov (KS) test, nominal $P < 0.05$), but not FaceCnnPerceivedAgeDiff. Innate-immunity-related genes, such as those linked to antigen processing and presentation, were positively related to all four AgeDiffs (KS test, Extended Data Fig. 6a). In summary, all four AgeDiffs were significantly positively related

to inflammation and innate-immunity processes, which are also important features of ageing per se.

Inflammation is related to AgeDiff at the cell-type level. We also considered whether cell fractions are differentially associated with AgeDiffs. Using CIBERSORT[14], we found that the monocyte fraction was significantly positively associated with FacePlsAgeDiff, and the direction was consistent, although not significant, with its association with the other three AgeDiffs across all individuals (Extended Data Fig. 6b). Consistently, monocyte count was even more significantly associated with FacePlsAgeDiff (*t*-test, $P = 0.00056$ between the top and bottom 20% of samples, Fig. 3b). The PLSR components 1 and 2 of 3D facial images regressed to monocyte fractions most significantly mapped to shrinkage of the forehead (Fig. 3b), which is in agreement with the negative association between inflammatory biomarkers, including interleukin-6, osteoprotegerin and tumour necrosis factor-α, and brain volume as measured by magnetic resonance imaging[15].

In contrast, naive $CD4^+$ T cells were consistently negatively associated with all four AgeDiffs, although the association was significant for only FacePlsAgeDiff and for RnaPlsAgeDiff (Fig. 3c and Extended Data Fig. 6b). The PLSR components 1 and 2 of 3D facial changes to this cell fraction showed mapping to a reduction of under-eye puffiness (Fig. 3c, *z* axis).

In addition, both CNN-derived AgeDiffs were significantly associated with the mean corpuscular volume of erythrocytes (MCV), a known chronic-illness indicator that is positively associated with ageing[16], nutrition[17] and alcohol abuse[18], between more-extreme samples (compared between the top and bottom 10% of samples; the association was marginal if the comparison was between the top and bottom 20% of samples). The facial pattern was similar to that associated with the monocyte fraction (Fig. 3d).

The impacts of lifestyle on AgeDiff and the mediating transcriptome regulators. Our parallel measurement of lifestyles, blood cell transcriptome and AgeDiffs for the same individuals in the Beijing 2012 cohort offered an opportunity to examine the molecular mediators of the impact of lifestyle on facial AgeDiff. We thus generated a tripartite network of lifestyle–transcriptome–AgeDiff using a causal-inference framework[11,19] querying all transcriptome clusters, ENCODE transcription factors, signalling and epigenetic factors and cytokines as potential mediators ($P < 0.05$ and FDR < 0.1, Methods). Cytokines semaphorin 6B (encoded by *SEMA6B*) and granulin (encoded by *GRN*) were inferred to be the mediators by which smoking increases FaceCnnAgeDiff and FacePlsAgeDiff, respectively; *SEMA6B* was also inferred to be a mediator by which alcohol drinking (measured in 'Drunkenness days per week') increased FaceCnnAgeDiff. ZZ-type zinc-finger-containing protein 3 (encoded by *ZZZ3*), involved in chromatin organization, was inferred to be the mediator of yoghurt's negative effect on FaceCnnAgeDiff, and to lower *SMAD1* to negatively affect FaceCnnPerceivedAgeDiff. RnaPlsAgeDiff had its own separate subnetwork, in which consumption of stem and root crops (for example, potato) were positively related to RnaPlsAgeDiff, mediated by tumour protein p53 (encoded by *TP53*) (Fig. 4a, $P < 0.05$ and FDR < 0.1). Interestingly, semaphorin is a chemokine, while granulin (cleaved from GRN, the granulin precursor) is also a secreted factor playing important roles in the development of the central nervous system[20], wound healing[21], tumorigenesis[22] and neurodegenerative disease[23]. Overall, the network can be partitioned into four modules on the basis of connectivity density: the first, mostly affecting FacePlsAgeDiff and partially affecting FaceCnnAgeDiff, is enriched for proteolysis; the second, mostly affecting FaceCnnAgeDiff, is enriched for glycoprotein, protease and lysosome; the third, mostly affecting FaceCnnAgeDiff, is enriched for transmembrane proteins and zinc-finger transcription factors; and the fourth, exclusively affecting RnaPlsAgeDiff, is enriched for response to antibiotics and ultraviolet radiation (Fig. 4a).

Ribo-minus RNA sequencing (RNA-seq) allowed us to simultaneously examine coding and non-coding RNA changes during human ageing. We found 935 long non-coding RNAs (lncRNAs) that were commonly expressed (fragments per kilobase of transcript per million reads mapped (FPKM) > 2) in at least a quarter of samples. Among them, 62 and 210 were up- and downregulated by age, respectively (FDR < 0.1) (Extended Data Fig. 7a). Twenty-three lncRNAs appeared in the causal-inference network (Fig. 4a). We found 5,002 circular RNAs (circRNAs) commonly expressed (transcripts per million (TPM) > 2) in at least a quarter of samples. Among them, 41 and 8 were up- and downregulated with age, respectively (FDR < 0.1) (Extended Data Fig. 7b,c), and 45 of them appeared in the causal-inference network (Fig. 4a). Despite a trend of a higher fraction of total circRNAs during ageing (Extended Data Fig. 7d, PCC = 0.26), consistent with global circRNA accumulation with age[24], total circRNA level shows no significant association with AgeDiffs, implicating it to more likely be a consequence of ageing.

We applied the same causal-inference approach to only the outliers defined by all four AgeDiffs because we assumed that some associations might be significant only in more-extreme samples. Indeed, we found a very different set of associations among the outliers as compared with those of the whole cohort. Ice-cream intake was inferred to be positively related to FacePlsAgeDiff through two circRNAs and to FaceCnnAgeDiff through AE binding protein 1 (encoded by *AEBP1*), chromobox 5 (encoded by *CBX5*) and sterol regulatory element-binding transcription factor 1 (encoded by *SREBF1*). Yoghurt was negatively related to FaceCnnAgeDiff and FaceCnnPerceivedAgeDiff through a few lncRNAs and circRNAs among the outliers ($P < 0.05$ and FDR < 0.1, Fig. 4b). SREBF1 modulates cellular cholesterol metabolism, regulates lipogenesis and glycolysis[25] and has an important role in the immune system[26,27].

To facilitate the full utilization of this dataset, we developed the HuB-Fi database for querying and visualizing health status and transcript changes associated with facial-ageing features, and the impacts of different lifestyles.

Discussion

In this study, we developed deep-learning CNN models that made age predictions based on high-resolution 3D facial images with high accuracy in a large cohort of ~5,000 people. CNN models identified a larger number of associations, that were stronger in significance, of ageing with health parameters and lifestyles, demonstrating that these models are a high-accuracy method for estimation of the ageing rate that can be complementary to other methods. Despite there being relatively fewer samples of very old ages (>70 yr) included for age prediction in our work, we consistently found many lifestyle and health parameters associated with various AgeDiffs, in particular the AgeDiff given by the artificial intelligence (AI)-based perceived-age predictor (FaceCnnPerceivedAgeDiff). The observation that the AI model often outperforms the PLS model is probably because the AI predictor is less sensitive to imbalanced data structures. In fact, FaceCnnPerceivedAgeDiff not only is associated with many more health parameters but also is more highly significantly associated with blood pressure than are the other three AgeDiffs (Supplementary Table 8). This suggests that our age predictors, in particular the AI-based predictor of perceived age, are superior health estimators. Unlike using chronological age in machine learning, the unique advantage of using perceived age as a health biomarker has been illustrated before by the Christensen group in Danish twins[12]. Here, we trained an AI predictor to learn how humans perceive other people's ages, which reliably captured human perception and even corrected some human errors through regularization, hence generating a health predictor that is even more accurate than perceived age (the training standards). Although CNNs

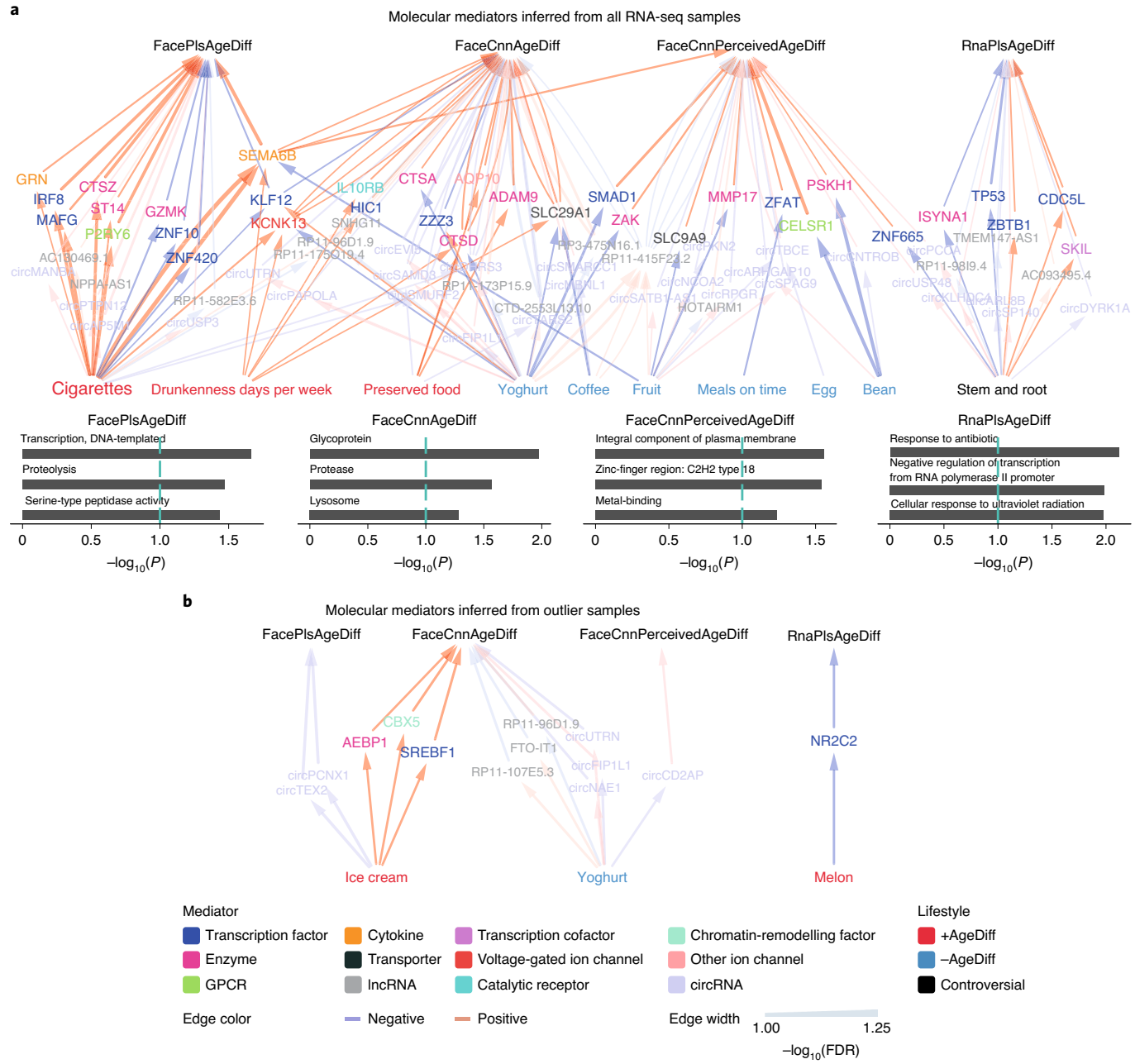


Fig. 4 | Inferred molecular mediators of AgeDiffs. a,**b**, Network of inferred causal relationships (causal-inference test with FDR correction, Methods) from lifestyle factors to AgeDiffs via molecular mediators (transcription factors, cytokines, regulatory genes and commonly expressed non-coding RNAs) among all samples (**a**) or outlier samples (**b**) in the Beijing 2012 cohort. Opaque nodes and edges indicate non-coding RNA mediators.

could give a much more accurate prediction of chronological age by optimizing chronological-age prediction, such prediction might ignore the deviations of biological age from chronological age. This is partially circumvented by training the CNN to learn perceived age, even though perceived age may not be a true reflection of biological age. The PerceivedAge s.d. generally increased with chronological age (PCC = 0.248, $P < 2.2 \times 10^{-16}$). This is perhaps because the judges of age were all young people who are not accustomed to judging older people's ages from experience, and even by random guessing, a larger number (representing age in this case) will have a larger range of variance than a smaller number will. However, although the PLS model does not predict chronological age as accurately as the CNN does, its simplicity may capture certain aspects of biological age that are ignored by the CNN. Future studies of a panel of gold standards of biological age will allow CNNs to be trained to specifically estimate biological age, and a better judgement to be made of the utility of different models for predicting biological age.

As for the specific pipeline and parameters we chose in this study, we first tested the most widely used CNN architectures, GoogleNet, VGG with 16 layers and ResNet with 50 layers. We found very similar performance between ResNet with 101, 152 and 200 layers and ResNet with 50 layers. Therefore, in the final model, we included in the ensemble the results of only ResNet with 50 layers. Using just RGB images, the deep CNN models could predict age with high accuracy, with a MAD of 2.90 yr when using GoogleNet, a MAD of 3.27 yr when using VGG and a MAD of 3.13 yr when using ResNet.

Similarly, we also predicted age with only depth information, which produced a MAD of 3.50 yr when using GoogleNet, a MAD of 3.75 yr when using VGG and a MAD of 3.70 yr when using ResNet. An ensemble of the three CNNs reaches a MAD of 2.82 yr for RGB images and a MAD of 3.35 yr for depth images. The final ensemble of the RGB and depth images yielded our reported MAD of 2.79 yr.

We further collected parallel data from 280 Han Chinese individuals in Beijing. We applied the CNN models trained on the large cohort to this smaller cohort and compared them with linear facial-image and transcriptome-age prediction models derived from the small cohort. Outliers identified by all four AgeDiffs significantly overlapped ($P < 0.05$).

At the functional level, high AgeDiffs all converged on the upregulation of inflammation and innate-immunity function, which is also supported by the upregulation of monocyte counts, a major player in innate immunity, and MCV, an indicator of chronic illness. Our analysis highlights the functional similarity between RNA-derived and 3D-facial-image-derived AgeDiffs not only at the molecular and functional level, but also at the cellular level.

The age-related facial features, health parameters and gene expression, as well as various AgeDiffs (except FacePlsAgeDiff) all revealed a large heterogeneity of the ageing rate at middle age, with a bimodal distribution of the old and young patterns. This suggests that young and old are the alternative steady states, and middle age is a transition stage (Fig. 2 and Extended Data Fig. 3), implying that middle age is the ideal stage for ageing interventions. Incidentally, it also avoids the complications of growth/maturation in the young group and the impact of disease status in the older group, and supports ageing-related risk screening for middle-aged people, a target group in which the benefits of implementation of therapies are likely to outweigh the potential harms of screening[28].

Finally, by scrutinizing the lifestyle choices responsible for acceleration or deceleration of AgeDiff and their potential molecular mediators in the blood, we found many potential causal associations between lifestyle, transcript and AgeDiff.

Overall, our analyses based on various age predictors show that humans age at different rates both in the blood and on the face, but coherently and with heterogeneity peaking at middle age. The differences can be attributed to different lifestyles, which through impacting circulating factors may affect facial morphology. More-accurate CNN-based predictors enabled more-sensitive and statistically robust discoveries of associations between facial features, health and lifestyle.

Despite some obvious limitations in our current study that need to be addressed in future studies, such as that individuals of advanced age were not included, that only Han Chinese people were included and that the current model is based only on a high-resolution 3D imaging system, our study represents a step closer to the ultimate goal of identifying actionable lifestyle choices and their molecular mediators to target ageing. The AI technology we developed here aims to easily and widely monitor ageing rates, which will greatly benefit the ever-growing ageing populations and development of effective intervention strategies. However, as this system is based on facial images, the technology is not only sensitive in regard to personal identification but potentially can be abused for unintended purposes; hence, it should be carefully guarded against any unethical use. With an increased sample size, we expect that other lifestyle parameters, their molecular mediators and facial signatures will be discovered by the same type of study in the near future.

Methods

Ethics statement. The Jidong study was carried out according to the guidelines of the Declaration of Helsinki[29]. Approval was obtained from Ethical Committees of the Staff Hospital of Jidong oil-field of China National Petroleum Corporation. The approval will be renewed every 5 years. Written informed consent was obtained from each of the participants. The Beijing study was approved by the ethics committee of the Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences. For the detailed recruitment procedure, please refer to the Reporting Summary.

Data acquisition. We collected 3D facial images of the Han Chinese cohort at Tangshan in Heibei province using the 3dMDface System (www.3dmd.com). As a requirement, volunteers sat straight and looked forward to the 3dMDface device for 1 min, without any facial expression, for the 3D facial images to be taken. The 3dMDface system returned OBJ-formatted 3D facial surfaces with point clouds and corresponding texture 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) were collected from questionnaires, weight, height and blood pressure were measured and routine blood tests were done.

Three-dimensional facial images, blood and routine physical indicators (full blood count, blood basic metabolic panel and anthropometry) were collected at Beijing Centers for Diseases Control and Prevention (CDC) and Centers for Preventive Medical Research. All participants provided written informed consent prior to this study. All participants (Han Chinese people, 169 females and 163 males) in this study had no history of facial surgery or facial abnormalities. We used the 3dMDface system to capture 3D facial images of about 40,000 vertices per face. The detailed procedure of photography capture has been described in previous work[7].

Human-perceived age. To compare human-perceived age and CNN-predicted age, we gathered 50 volunteers to evaluate the ages of all samples. We randomly shuffled the images and asked the volunteers to guess the ages of participants by looking at their facial images. Each image was evaluated by 5.33 volunteers on average (at least 3 and no more than 6), and the average guess of the volunteers was used as the human-perceived age of each face.

Age prediction using 3D facial images. *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.

Data structure of 3D facial images. The 3D facial images were stored in the OBJ file format and with corresponding texture images. The OBJ-format file contains four types of information for a 3D face surface with N points: (1) N lines of 'v' plus XXX (where X is a floating-count number), representing coordinates of the N points; (2) N lines of 'vn' plus XXX, representing the normal vector of each of the N points; (3) N lines of 'vt' plus XX, representing coordinates of each point in the corresponding texture image; and (4) multiple lines starting with 'f' and three groups of integers, representing that these three indexes form a triangle, which further constitutes multiple triangle meshes in the 3D facial images.

Nose-tip annotation. The nose tip is the most salient and stable feature point on 3D facial images. We first detected the nose-tip point on a face using sphere fitting[30]. We used a sphere with a radius equal to r to fit regions over the 3D face and took the point with the smallest least square difference as the nose-tip point. To calculate the least square of every point p with coordinate (x_p,y_p,z_p), we first summarized all points on the 3D face with a distance of <1.5 cm from point p, and subjected them to the following formula:

$$\sqrt{(x_p - x_i)^2+(y_p - y_i)^2+(z_p - z_i)^2} < 1.5$$

where i represents the ith vertex on the 3D face. Then, we assumed that there were, in total, N points with a distance of <1.5 cm to point p. These points were annotated as $p_i = (x_i, y_i, z_i)$, $1 \le i \le N$. At the same time, we assumed that the centre of the sphere had coordinate (o_x,o_y,o_z); then, the sphere could be represented using the following formula:

$$(x - o_x)^2+(y - o_y)^2+(z - o_z)^2 = r^2$$

where r represents the radius of the sphere. For points $p_i = (x_i,y_i,z_i)$, we calculated the distance between the point and the sphere, which can be represented as the error:

$$\varepsilon_i = \left|(x_i - o_x)^2+(y_i - o_y)^2+(z_i - o_z)^2 - r^2\right|$$

$$\varepsilon_i = \left|\left(x_i^2 + y_i^2 + z_i^2\right) - \left(2o_x x_i + 2o_y y_i + 2o_z z_i\right) + \left(o_x^2 + o_y^2 + o_z^2 - r^2\right)\right|$$

Then to clearly represent the errors, we reformulated the error as:

$$\varepsilon_i = |Y - (w_1 X_1 + w_2 X_2 + w_3 X_3 + w_0)|$$

where $w_1 = 2o_x$, $w_2 = 2o_y$, $w_3 = 2o_z$, $w_0 = r^2 - o_x^2 - o_y^2 - o_z^2$, $Y = x_i^2 + y_i^2 + y_i^2$, $X_1 = x_i$, $X_2 = y_i$ and $X_3 = z_i$; then, ε_i can be regarded as the error of the linear regression, with X_1, X_2 and X_3 as the independent variable and Y as the dependent variable. Taking the derivatives of the least square loss function, we generated the

solution for the regression. First, we transformed the regression problem into the matrix formula:

$$\begin{bmatrix} x_1 & y_1 & z_1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_N & y_N & z_N & 1 \end{bmatrix} \begin{bmatrix} 2o_x \\ 2o_y \\ 2o_z \\ r^2 - o_x^2 - o_y^2 - o_z^2 \end{bmatrix} = \begin{bmatrix} x_1^2 + y_1^2 + y_1^2 \\ \vdots \\ x_N^2 + y_N^2 + y_N^2 \end{bmatrix}$$

To simplify the equation, we assumed:

$$X = \begin{bmatrix} x_1 & y_1 & z_1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_N & y_N & z_N & 1 \end{bmatrix}$$

$$W = \begin{bmatrix} 2o_x \\ 2o_y \\ 2o_z \\ r^2 - o_x^2 - o_y^2 - o_z^2 \end{bmatrix}$$

$$y = \begin{bmatrix} x_1^2 + y_1^2 + y_1^2 \\ \vdots \\ x_N^2 + y_N^2 + y_N^2 \end{bmatrix}$$

Our regression can be written as:

$$XW = y$$

Then we generated the solution for W:

$$W = \begin{bmatrix} 2o_x \\ 2o_y \\ 2o_z \\ r^2 - o_x^2 - o_y^2 - o_z^2 \end{bmatrix} = \left(X^T X\right)^{-1} X^T y$$

After getting the values for W, we calculated the centre and the radius of the sphere, with which we calculated the least square loss of each point on the 3D face:

$$\left(o_x, o_y, o_z\right) = \left(\frac{W[1]}{2}, \frac{W[2]}{2}, \frac{W[3]}{2}\right)$$

$$r = \sqrt{W[4] + \frac{W[1]^2 + W[2]^2 + W[3]^2}{4}}$$

Pose correction of the 3D face. Although the volunteers were asked to sit straight and look forward, there were still differences in poses among 3D faces, including looking slightly left or right looking and looking up or down. As we had already identified the location of the nose tip, we corrected the face pose using points around the nose tip. Assuming that the coordinate of the nose tip was (x_{nose},y_{nose}, z_{nose}), points on the 3D face with a distance to the nose tip of <5 cm were used to correct the pose; that is, points were subjected to:

$$\sqrt{(x_{nose} - x_i)^2 + (y_{nose} - y_i)^2 + (z_{nose} - z_i)^2} < 5$$

We assumed there were in total M points, $p_i = (x_i, y_i, z_i), 1 \le i \le M$, satisfying the requirements.

We then wrote these M points in matrix format as:

$$X = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_M \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_M & y_M & z_M \end{bmatrix}$$

We applied principal component analysis (PCA) to these data, and average coordinates of each axis were calculated:

$$\bar{x} = \frac{1}{M}\sum_{i=1}^{M} x_i,\ \bar{y} = \frac{1}{M}\sum_{i=1}^{M} y_i,\ \bar{z} = \frac{1}{M}\sum_{i=1}^{M} z_i$$

Each sample's coordinates minus the average coordinates of each axis were calculated:

$$X_1 = \begin{bmatrix} x_1 - \bar{x} & y_1 - \bar{y} & z_1 - \bar{z} \\ x_2 - \bar{x} & y_2 - \bar{y} & z_2 - \bar{z} \\ \vdots & \vdots & \vdots \\ x_M - \bar{x} & y_M - \bar{y} & z_M - \bar{z} \end{bmatrix}$$

Then the covariance matrix of X_1 was calculated:

$$C = \begin{bmatrix} \mathrm{cov(x,x)} & \mathrm{cov(x,y)} & \mathrm{cov(x,z)} \\ \mathrm{cov(x,y)} & \mathrm{cov(y,y)} & \mathrm{cov(y,z)} \\ \mathrm{cov(x,z)} & \mathrm{cov(y,z)} & \mathrm{cov(z,z)} \end{bmatrix}$$

where

$$cov(x,y) = \frac{\sum_{i=1}^{M}(x_i - \bar{x})(y_i - \bar{y})}{M - 1}$$

Finally, the eigenvalue and corresponding eigenvector were calculated:

$$C\gamma = \lambda\gamma$$

where λ is the eigenvalue and γ is corresponding eigenvector. The maximum eigenvalue λ_1 and corresponding eigenvector γ_1 represent the direction of maximum variance of the data, which is the direction along the nose bridge. The second maximum eigenvalue λ_2 represents the direction of second maximum variance of the data, which is the direction of the nose from left to right. The minimum eigenvalue λ_3 and corresponding eigenvector γ_3 represent the direction of minimum variance of the data, which is the direction from the inner part of the head to the outer part of the head along the eye sight direction. The direction of γ_1 was considered to be the y axis, the direction of γ_2 was the x axis and the direction of γ_3 was the z axis. We rotated the coordinates of the 3D faces into the coordinate system with the three eigenvectors γ_1, γ_2 and γ_3 as the x, y and z axis, respectively, which were used to correct the pose using the following formula:

$$X_{new} = UX_1 = U(X - [\bar{x} \quad \bar{y} \quad \bar{z}])$$

where

$$U = [\gamma_2 \quad \gamma_1 \quad \gamma_3] = \begin{bmatrix} \gamma_{21} & \gamma_{11} & \gamma_{31} \\ \gamma_{22} & \gamma_{12} & \gamma_{32} \\ \gamma_{21} & \gamma_{13} & \gamma_{33} \end{bmatrix}$$

Projection to the X–Y plane. Our 3D facial images were projected onto the X–Y 2D plane with the z buffer and scan line algorithms, with sampling resolution at 0.1 cm. All RGB pixel information was kept in the projected 2D facial image.

Eye-corner and mouth-corner detection. Besides detecting the nose tip, we detected another six landmarks (outer corner of left eye, inner corner of left eye, outer corner of right eye, inner corner of right eye, left mouth corner and right mouth corner) on each face. We first annotated these 6 landmarks manually for 50 females and 50 males, and trained the machine-learning algorithm (PCA) to recognize these landmarks. Then, all 100 samples were projected onto the X–Y plane using the corresponding six landmarks. For each of the 6 landmarks, the outer corner of the left eye for example, blocks centred at the landmark of size 21 × 21 pixels were used to extract the RGB texture and depth information. Then, for each landmark, feature vectors with 1,764 values were extracted to represent the landmark. The outer corners of the left eyes for all 100 samples were represented by feature vectors:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

where $m = 1{,}764$ and $n = 100$. Similarly, we first subtracted the matrix by the average of each feature:

$$X_1 = \begin{bmatrix} x_{11} - \overline{x_{*1}} & x_{12} - \overline{x_{*2}} & \cdots & x_{1m} - \overline{x_{*m}} \\ x_{21} - \overline{x_{*1}} & x_{22} - \overline{x_{*2}} & \cdots & x_{2m} - \overline{x_{*m}} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} - \overline{x_{*1}} & x_{n2} - \overline{x_{*2}} & \cdots & x_{nm} - \overline{x_{*m}} \end{bmatrix}$$

where

$$\overline{x_{*i}} = \frac{1}{n}\sum_{j=1}^{n} x_{ji}$$

representing the average of the ith feature over all samples, where j is the index used to iterate through all the samples. Similarly to the pose correction, we then calculated the covariance matrix of X_1 and selected the maximum 16 eigenvalues (represented by v) and used their corresponding eigenvectors to construct the feature subspace:

$$U = [v_1 \quad v_2 \quad \cdots \quad v_{16}]$$

We could then easily reflect the 1,764 features of the outer corner of the left eye for 100 samples into the subspace. For other faces, to annotate the outer corner

the of left eye, we reflected the features of all candidate points into the established subspace. Then, we represented the coordinates of the candidate points in the subspace:

$$x_{\text{test\_new}} = U(x_{\text{test}} - [\bar{x}_{*1} \quad \bar{x}_{*2} \quad \cdots \quad \bar{x}_{*m}])$$

$$x_{\text{test\_new}} = [w_1 \quad w_2 \quad w_3 \quad \cdots \quad w_{16}]$$

where $[\bar{x}_{*1} \quad \bar{x}_{*2} \quad \cdots \quad \bar{x}_{*m}]$ is the average vector in the training set. We next defined the distance between candidate points and the outer corners of the left eye in the training set with the following formula:

$$d = \sum_{i=1}^{16} \frac{w_i^2}{var(w_i^{\text{train}})}$$

where $\text{var}(w_i^{\text{train}})$ is the variance of w_i in the training set. The point that reached the smallest distance from the training set was regarded as the outer corner of the left eye in the test image. Similar processes were applied to the other five landmarks.

Additional landmark detection. In addition to the seven landmarks detected above (nose tip, outer corner of left eye, inner corner of left eye, outer corner of right eye, inner corner of right eye, left mouth corner and right mouth corner), we detected landmarks such as the chin point, upper lip point and lower lip point. We localized points of these landmarks on the 3D face on the basis of prior knowledge. For example, the upper lip point is below the nose tip, with z-axis local maximum, and with the high red colour intensity under it because the lips have higher red colour intensity than the face skin; similarly, the lower lip point is below the upper lip point, with z-axis local maximum, and with high red colour intensity above it; and the chin point is below the lower lip point, showing a small angle from the point to the points above and below.

Preparation of 3D faces for deep learning. All 3D facial images were first preprocessed for landmark detection, posture correction and registration. Different 3D faces have different numbers of points and different sizes. On the 2D face, we used dlib (version 19.8.0)[31] to detect the left and right profiles of the face, and then cropped the faces according the profiles and landmarks, using the left- and right-most edges as left and right boundaries, with the bottom boundary set at the chin point and the upper boundary by the coordinates of the chin point and nose-tip point (assuming the y axis of the nose tip is y_1, and the y axis of the chin point is y_2, where $y_2 > y_1$, then the y axis of the upper boundary is $y_1 - \sqrt{2}(y_2 - y_1)$). During the final cropping of the image, we extended each boundary by 2 pixels, and the depth image was cropped at the same place. All RGB images and deep images were resized to 224 × 224 pixels before using deep learning.

Deep CNNs. We combined 3 popular CNN architectures, GoogleNet[32], VGG[33] and ResNet[34], in Python to predict age. For all three CNNs, we used one node as the last layer.

Ensemble of deep CNNs. To further improve our accuracy, 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 $y = (GoogleNet + VGG + ResNet)/3$, where *GoogleNet*, *VGG* and *ResNet* represent the ages predicted by these three CNNs, respectively, and y represents the emsemble predicted age.

Loss functions and CNN training. For age prediction, we replaced the last layer of the three CNNs with one node without any activation functions, and we used the mean absolute difference between the predicted age and actual age as the loss function:

$$L = \frac{1}{batch\_size} \sum_{i=1}^{batch\_size} |y_i - \widehat{y_i}|$$

where *batch_size* is one of the parameters when training the CNN, indicating the number of samples in each weight-update process. Here, we set *batch_size* equal to 32. y_i represents the actual age of the ith sample and $\widehat{y_i}$ represents the predicted age of the ith sample.

We then used the Adam (as implemented in TensorFlow version 1.8.0)[35] algorithm to update the weights, with initial learning rate set at 0.0001 and 80% probability to keep nodes in dropout layers.

Data augmentation. During training of deep CNN models, we reflected all samples from left to right, which can increase the amount of training data.

Tenfold cross-validation. To validate the prediction ability for age using 3D facial images, we divided all our 4,719 samples into 10 groups randomly for 10-fold cross-validation. 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. When evaluating the models with independent datasets, we used all ten models generated from the ten cross-validation to test the datasets independently, and used the average performance level as the final prediction capability.

Age prediction using 3D facial images with PLS model. The prediction of age with PLS has been described in detail in previous work[7].

AgeDiffs association with health parameters. Age-corrected AgeDiffs were used to calculate the Pearson's correlation coefficient and corresponding P values. Benjamini–Hochberg correction was used for FDR calculation.

RNA isolation, sequencing and normalization. We drew 10 ml of total fasting blood from each participant in the early morning. Plasma and PBMCs were separated using a Sigma-Aldrich HISTOPAQUE-1077 kit. RNA was extracted with TRIzol (Invitrogen) using the protocol provided by the manufacturer. Samples with RNA integrity number (RIN) > 8 were processed with the Illumina TruSeq stranded total RNA with Ribo-Zero sample preparation kit and then sequenced on an Illumina HiSeq 2000 machine according to the manufacturer's instructions. Paired-end reads (126 bp) were mapped to the GRCh38 reference human genome using STAR9 version 2.4.0d.

RNA normalization. Only reads with a unique match to the reference genome were kept. Protein-coding gene quantification was performed using GENCODE version 24 annotation[36]. The normalization and quantification were conducted by Cufflinks (version 2.2.1)[37].

De novo identification of exonic circRNA. A computational pipeline (CIRCexplorer2 (ref. [38]), version 2.3.5) was used to predict exonic circRNA. Human PBMC ribo-minus paired-end RNA-seq raw data were mapped to the hg38 human reference genome using TopHat-Fusion (implemented in TopHat version 2.1.1), and then were parsed by CIRCexplorer2 to obtain circRNA back spliced junction reads. CircRNA read counts from all samples were merged together and scaled to TPM.

Age prediction using transcriptome. The 'pls' package[39] in R was used for PLS regression. Both chronological age and gene-expression levels ($\log_2$ scaled) were combined into one matrix as the input. The leave-one-out (LOO) method (predictors were trained using all but one sample and then used to predict the age of the left-out sample) 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 and training the PLS model on such samples, and then was evaluated using both MAD and PCC.

AgeDiff correction. To study age-independent associations of AgeDiff, we corrected AgeDiff by fitting to a polynomial model to age as follows, with span 0.75 and degree 2:

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

Unlike other AgeDiffs, there are significant differences between sexes in FacePerceivedAgeDiff and FaceCnnPerceivedAgeDiff ($P < 0.05$), and we corrected these two AgeDiffs in males and females separately.

Correlation between AgeDiff SD and chronological age. A non-overlapping sliding-window approach was used to study the relationship between AgeDiff SD and age. Samples were sorted by age, and then a non-overlapping sliding window with different bin size (100 or 20 in Jidong cohort, 20 or 10 in Beijing cohort) was used to group samples and calculate mean of age and s.d. of AgeDiffs in each sliding window.

Term-enrichment analysis. Gene set enrichment analysis (GSEA)[40] and Fisher's exact test in DAVID[41] were utilized for the term-enrichment test. KEGG pathway genes were downloaded from the KEGG database using the REST API. The cytokine genes were collected from the ImmPort database[42] and senescence-associated secretory phenotype (SASP) genes were collected from a review of SASP[43].

Projection of cell fraction and gene-expression features on the average 3D facial image. With the 'pls' package in R, the cell fraction or another feature was treated as the dependent variable, while the 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 average 3D facial image synthesized from the whole cohort.

Lifestyle–regulator–AgeDiff causal-relationship inference. The causal-inference test of Millstein et al.[11] was adopted with the R package 'cit'[44]. Briefly, each of the lifestyle–gene expression–AgeDiff relationships was tested separately to classify them into mediated by, consequential to or independent of gene expression.

A causal inference met the following four criteria: (1) lifestyle and AgeDiff were correlated; (2) lifestyle was associated with gene expression after adjusting for AgeDiff; (3) gene expression was associated with AgeDiff after adjusting for lifestyle; and (4) lifestyle was independent of AgeDiff after adjusting for gene expression. 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 $P < 0.05$ and FDR < 0.1 (calculated by the 'cit' package[44]) were retained. The network was visualized by Cytoscape (version 3.5.0).

The input regulatory genes included: (1) human signal-transduction-pathway gene sets downloaded from KEGG pathway (https://www.genome.jp/kegg/pathway.html) and Reactome (https://www.reactome.org/) databases; (2) human transcription factors, enzymes, transporters, receptors and ion channels downloaded from the Animal Transcription Factor Database (http://www.bioguo.org/AnimalTFDB/), the human DEPhOsphorylation Database (http://www.depod.bioss.uni-freiburg.de/) and IUPHAR/BPS database (http://www.guidetopharmacology.org/); and (3) cytokines and SASPs from ImmPort database[42] (https://www.immport.org/shared/genelists) and a review on SASPs[43].

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The regulatory genes used in building the causal-inference network are from the KEGG pathway (https://www.genome.jp/kegg/pathway.html), Reactome (https://www.reactome.org/), Animal Transcription Factor Database (http://www.bioguo.org/AnimalTFDB/), human DEPhOsphorylation Database (http://www.depod.bioss.uni-freiburg.de/), IUPHAR/BPS database (http://www.guidetopharmacology.org/) and ImmPort[42] (https://www.immport.org/shared/genelists). ImageNet (http://image-net.org/) was used for pretraining.
Results of facial-image, transcriptome and lifestyle associations are searchable at http://www.picb.ac.cn/hanlab/hub-fi/. Three-dimensional images and other metadata sensitive to personal identification cannot be publicized or shared according to our participant consent agreement. Individual sequencing raw data, as they contain genetic information, will be available on request under the condition of approval of the ethics committee of Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences abiding China Human Genetic Resource law. Mapped read counts and FPKM expression values of coding genes from the RNA-seq are deposited to the public repository NODE at https://www.biosino.org/node/project/detail/OEP001041.

Received: 25 January 2020; Accepted: 24 July 2020;
Published online: 7 September 2020

References

1. Peters, M. J. et al. The transcriptional landscape of age in human peripheral blood. *Nat. Commun.* **6**, 8570 (2015).
2. Lehallier, B. et al. Undulating changes in human plasma proteome profiles across the lifespan. *Nat. Med.* **25**, 1843–1850 (2019).
3. Hannum, G. et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* **49**, 359–367 (2013).
4. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, 3156 (2013).
5. Unschuld, P. U. & Tessenow, H. *Huang Di Nei Jing Su Wen* (University of California Press, 2011).
6. Chen, W., Xia, X., Huang, Y., Chen, X. & Han, J.-D. J. Bioimaging for quantitative phenotype analysis. *Methods* **102**, 20–25 (2016).
7. Chen, W. et al. Three-dimensional human facial morphologies as robust aging markers. *Cell Res.* **25**, 574–587 (2015).
8. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The Hallmarks of Aging. *Cell* **153**, 1194–1217 (2013).
9. Gao, X. W., Hui, R. & Tian, Z. Classification of CT brain images based on deep learning networks. *Computer Methods Programs Biomed.* **138**, 49–56 (2017).
10. Esteva, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115–118 (2017).
11. Millstein, J., Zhang, B., Zhu, J. & Schadt, E. E. Disentangling molecular relationships with a causal inference test. *BMC Genet.* **10**, 23 (2009).
12. Debrabant, B. et al. DNA methylation age and perceived age in elderly Danish twins. *Mechanisms Ageing Dev.* **169**, 40–44 (2018).
13. Lu, T. et al. Gene regulation and DNA damage in the ageing human brain. *Nature* **429**, 883–891 (2004).
14. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
15. Jefferson, A. L. et al. Inflammatory biomarkers are associated with total brain volume. *Neurology* **68**, 1032–1038 (2007).
16. Frąckiewicz, J. et al. Hematological parameters and all-cause mortality: a prospective study of older people. *Aging Clin. Exp. Res.* **30**, 517–526 (2018).
17. Chatthanawaree, W. Biomarkers of cobalamin (vitamin B12) deficiency and its application. *J. Nutr. Health Aging* **15**, 227–231 (2011).
18. Conigrave, K. M., Davies, P., Haber, P. & Whitfield, J. B. Traditional markers of excessive alcohol use. *Addiction* **98**, 31–43 (2003).
19. Liu, Y. et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat. Biotechnol.* **31**, 142–147 (2013).
20. Ahmed, Z. et al. Accelerated lipofuscinosis and ubiquitination in granulin knockout mice suggest a role for progranulin in successful aging. *Am. J. Pathol.* **177**, 311–324 (2010).
21. He, Z., Ong, C. H. P., Halper, J. & Bateman, A. Progranulin is a mediator of the wound response. *Nat. Med.* **9**, 225–229 (2003).
22. Elkabets, M. et al. Human tumors instigate granulin-expressing hematopoietic cells that promote malignancy by activating stromal fibroblasts in mice. *J. Clin. Invest* **121**, 784–799 (2011).
23. Chitramuthu, B. P., Bennett, H. P. J. & Bateman, A. Progranulin: a new avenue towards the understanding and treatment of neurodegenerative disease. *Brain* **140**, 3081–3104 (2017).
24. Knupp, D. & Miura, P. CircRNA accumulation: a new hallmark of aging? *Mechanisms Ageing Dev.* **173**, 71–79 (2018).
25. Ruiz, R. et al. Sterol regulatory element-binding protein-1 (SREBP-1) is required to regulate glycogen synthesis and gluconeogenic gene expression in mouse liver. *J. Biol. Chem.* **289**, 5510–5517 (2014).
26. Oishi, Y. et al. SREBP1 contributes to resolution of pro-inflammatory TLR4 signaling by reprogramming fatty acid metabolism. *Cell Metab.* **25**, 412–427 (2017).
27. Li, S. et al. Metabolic phenotypes of response to vaccination in humans. *Cell* **169**, 862–877.e817 (2017).
28. Schoenborn, N. L. et al. Preferred clinician communication about stopping cancer screening among older US adults: results from a national survey. *JAMA Oncol.* **4**, 1126–1128 (2018).
29. World Medical Association Inc Declaration of Helsinki. Ethical principles for medical research involving human subjects. *J. Indian Med. Assoc.* **107**, 403–405 (2009).
30. Guo, J., Mei, X. & Tang, K. Automatic landmark annotation and dense correspondence registration for 3D human facial images. *BMC Bioinformatics* **14**, 232 (2013).
31. King, D. E. Dlib-ml: a Machine Learning Toolkit *J. Mach. Learn. Res.* **10**, 1755–1758 (2009).
32. Szegedy, C. et al. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 1–9 (2015).
33. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. Preprint at https://https://arxiv.org/abs/1409.1556 (2014).
34. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. in *IEEE Conference on Computer Vision and Pattern Recognition* 770–778 (2016).
35. Kingma, D. & Ba, J. Adam: A method for stochastic optimization. Preprint at https://arxiv.org/abs/1412.6980 (2014).
36. Harrow, J. et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
37. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
38. Zhang, X. O. et al. Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res.* **26**, 1277–1287 (2016).
39. Wold, S., Sjöström, M. & Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometrics Intell. Lab. Syst.* **58**, 109–130 (2001).
40. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
41. Dennis, G. et al. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* **4**, P3 (2003).
42. Bhattacharya, S. et al. ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. *Sci. Data* **5**, 180015 (2018).
43. Coppé, J.-P., Desprez, P.-Y., Krtolica, A. & Campisi, J. The senescence-associated secretory phenotype: the dark side of tumor suppression. *Annu Rev. Pathol.* **5**, 99–118 (2010).
44. Millstein, J., Chen, G. K. & Breton, C. V. cit: hypothesis testing software for mediation analysis in genomic applications. *Bioinformatics* **32**, 2364–2365 (2016).

Acknowledgements

This work was supported by grants from the National Natural Science Foundation of China (91749205), China Ministry of Science and Technology (2016YFE0108700) and Shanghai Municipal Science and Technology Major Project (2017SHZDZX01) to J.-D.J.H.

Author contributions

J.-D.J.H. conceived and designed the study and analyses. Y.Z. set up the cohort and designed and collected baseline data. C.V.C., together with J.D.J.H., conceived the

paradigm-shift idea to train the CNN on perceived age instead of chronological age for an AI-based perceived-age predictor. X.X. and X.C. analysed the data, with help from Y.W. and C.X. Y. Cao, W. Wei, G.C. Y.Y., X.X., K.L. and D.C. helped in collecting and preprocessing data. N.Q., X.Z. and J.J. helped to set up GPU, AI systems and training. G.W. and F.L. performed the PBMC-isolation and RNA-extraction experiments. Weiyang Chen provided FacePlsAge. B.X., Weizhong Chen, Y. Cao, C.X. and W.G. helped to recruit the Beijing cohorts. G.C. quantified the wrinkle and symmetry on faces. Y. Chen analysed circRNA. X.W., M.C. and D.C. helped with data analysis. X.X., X.C., J.D.J.H, C.V.C., K.Z., B.K.K. and W. Wang wrote the manuscript. All authors contributed to preparation of the manuscript. G.W., F.L. and Y.W. contributed equally.

Competing interests

Authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s42255-020-00270-x.

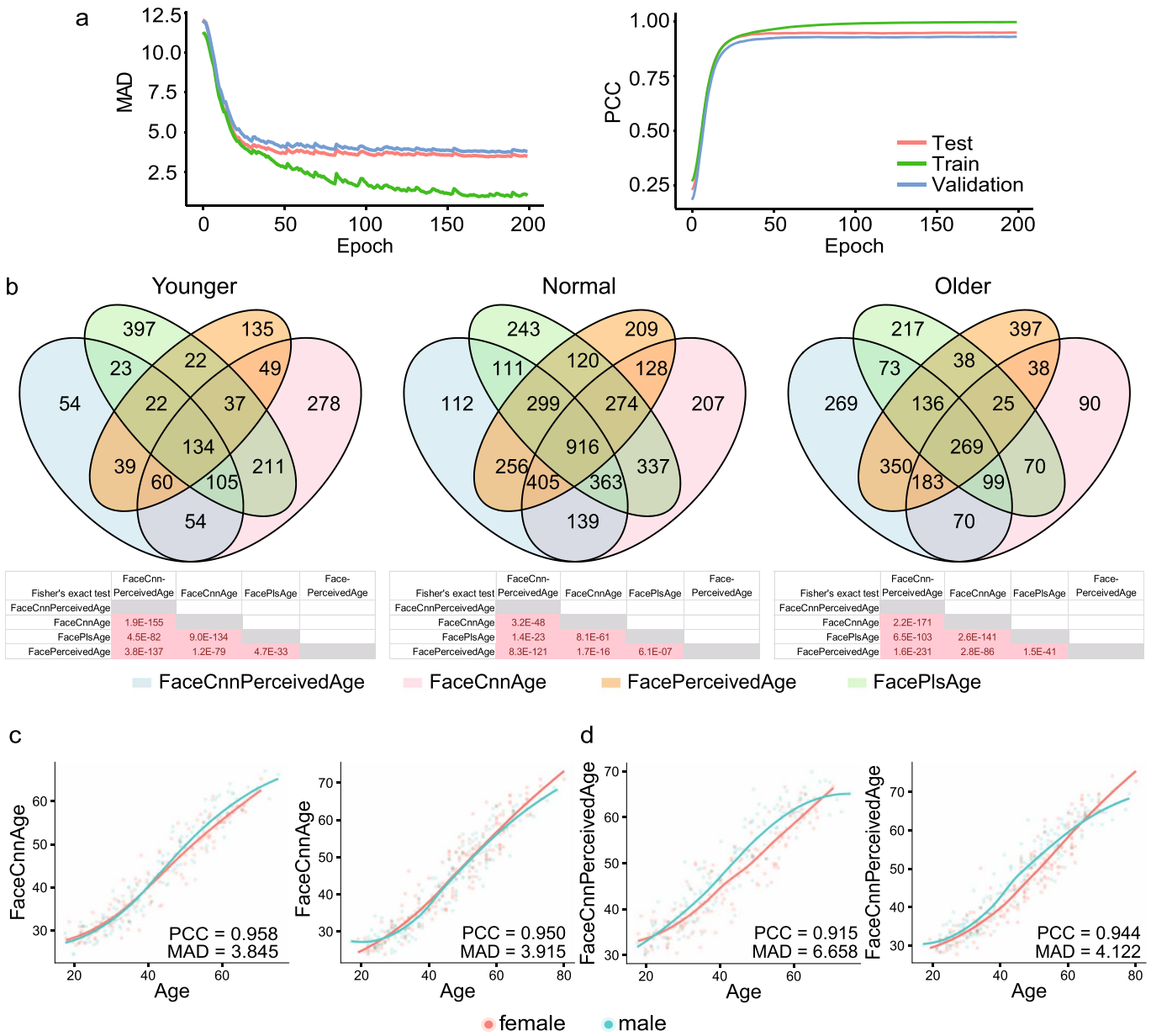
Supplementary information is available for this paper at https://doi.org/10.1038/s42255-020-00270-x.

Correspondence and requests for materials should be addressed to Y.Z. or J.-D.J.H.

Peer review information Primary Handling Editor: Pooja Jha.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

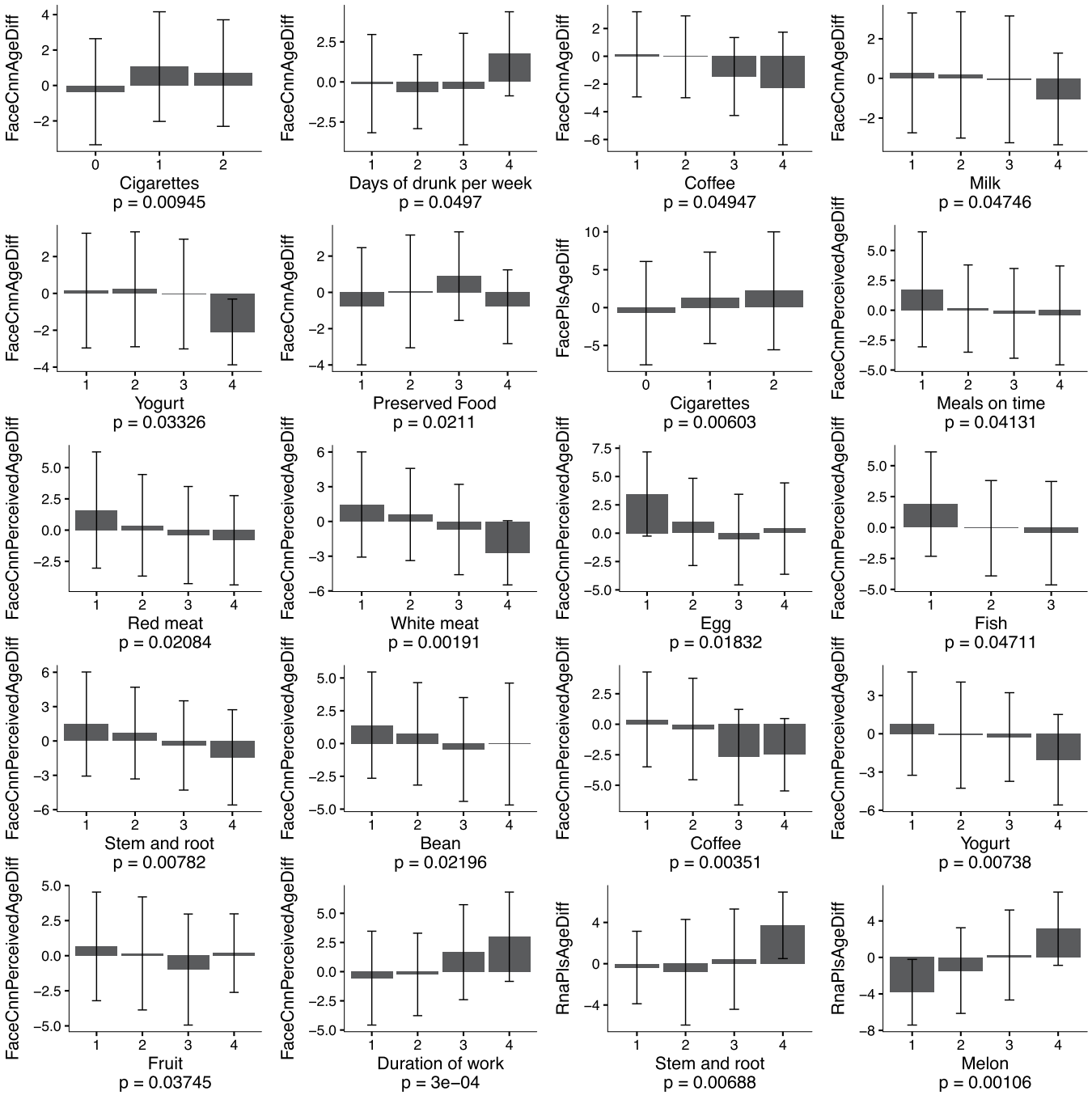


Fisher's exact test	FaceCnn-PerceivedAge	FaceCnnAge	FacePlsAge	Face-PerceivedAge
FaceCnnPerceivedAge				
FaceCnnAge	1.9E-155			
FacePlsAge	4.5E-82	9.0E-134		
FacePerceivedAge	3.8E-137	1.2E-79	4.7E-33	

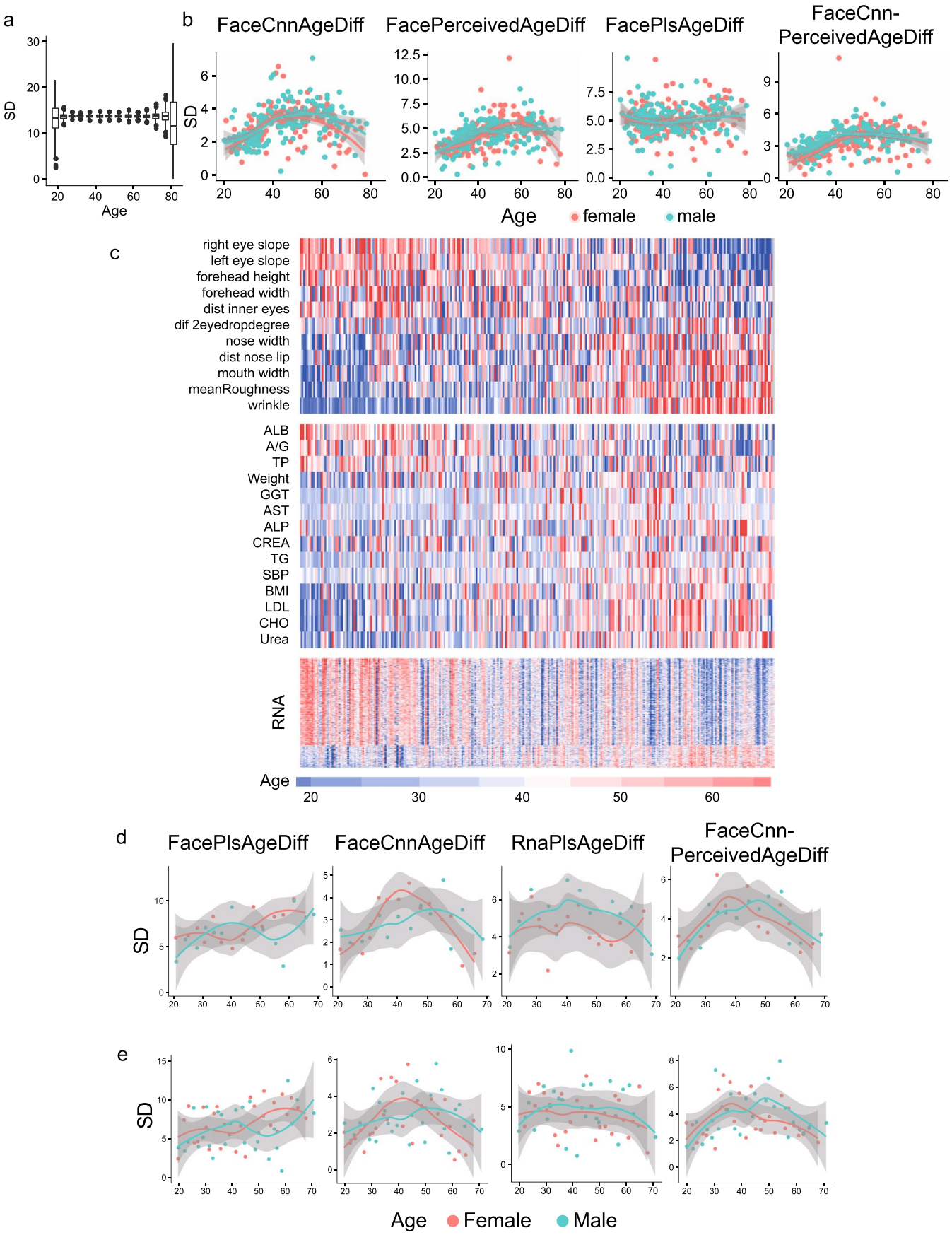
Fisher's exact test	FaceCnn-PerceivedAge	FaceCnnAge	FacePlsAge	Face-PerceivedAge
FaceCnnPerceivedAge				
FaceCnnAge	3.2E-48			
FacePlsAge	1.4E-23	8.1E-61		
FacePerceivedAge	8.3E-121	1.7E-16	6.1E-07	

Fisher's exact test	FaceCnn-PerceivedAge	FaceCnnAge	FacePlsAge	Face-PerceivedAge
FaceCnnPerceivedAge				
FaceCnnAge	2.2E-171			
FacePlsAge	6.5E-103	2.6E-141		
FacePerceivedAge	1.6E-231	2.8E-86	1.5E-41	

Extended Data Fig. 1 | Age and perceived age prediction using CNN. a, Deep learning performance in training, validation and testing datasets for age prediction. Average loss (mean average difference (MAD), upper panel) and accuracy (Pearson Correlation Coefficient (PCC), lower panel) were plotted over training epochs. One epoch indicates the network weights updates over the whole training dataset for one time. **b**, Overlap between FaceCnnAge, FaceCnnPerceivedAge, FacePerceivedAge, and FacePlsAge in younger (left), normal (middle), and older (right) samples. The lower table shows the one-tailed Fisher's exact test p-values with light red indicate $p < 0.05$. **c**, **d**, Independent validation of FaceCnnAge (c) and FaceCnnPerceivedAge (d) on 332 facial images collected in Beijing, 2012 (left) and 358 facial images collected in Beijing, 2015 (right).

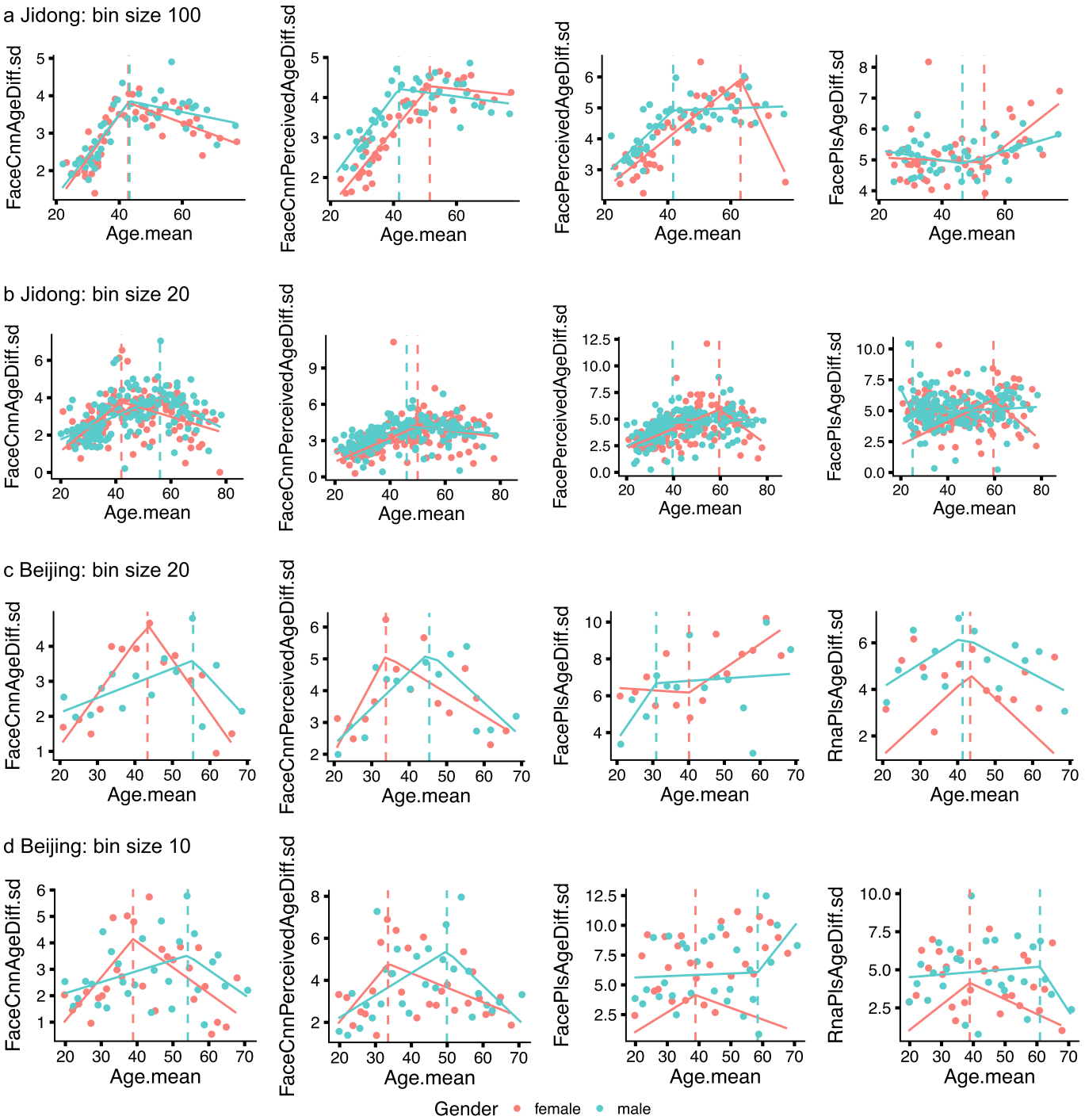


Extended Data Fig. 2 | AgeDiff-associated lifestyles in Beijing (2012) cohort by ANOVA test. The p value is derived from ANOVA (n=341). All results with $p < 0.05$ are shown here. Data are presented as mean +/- SD.

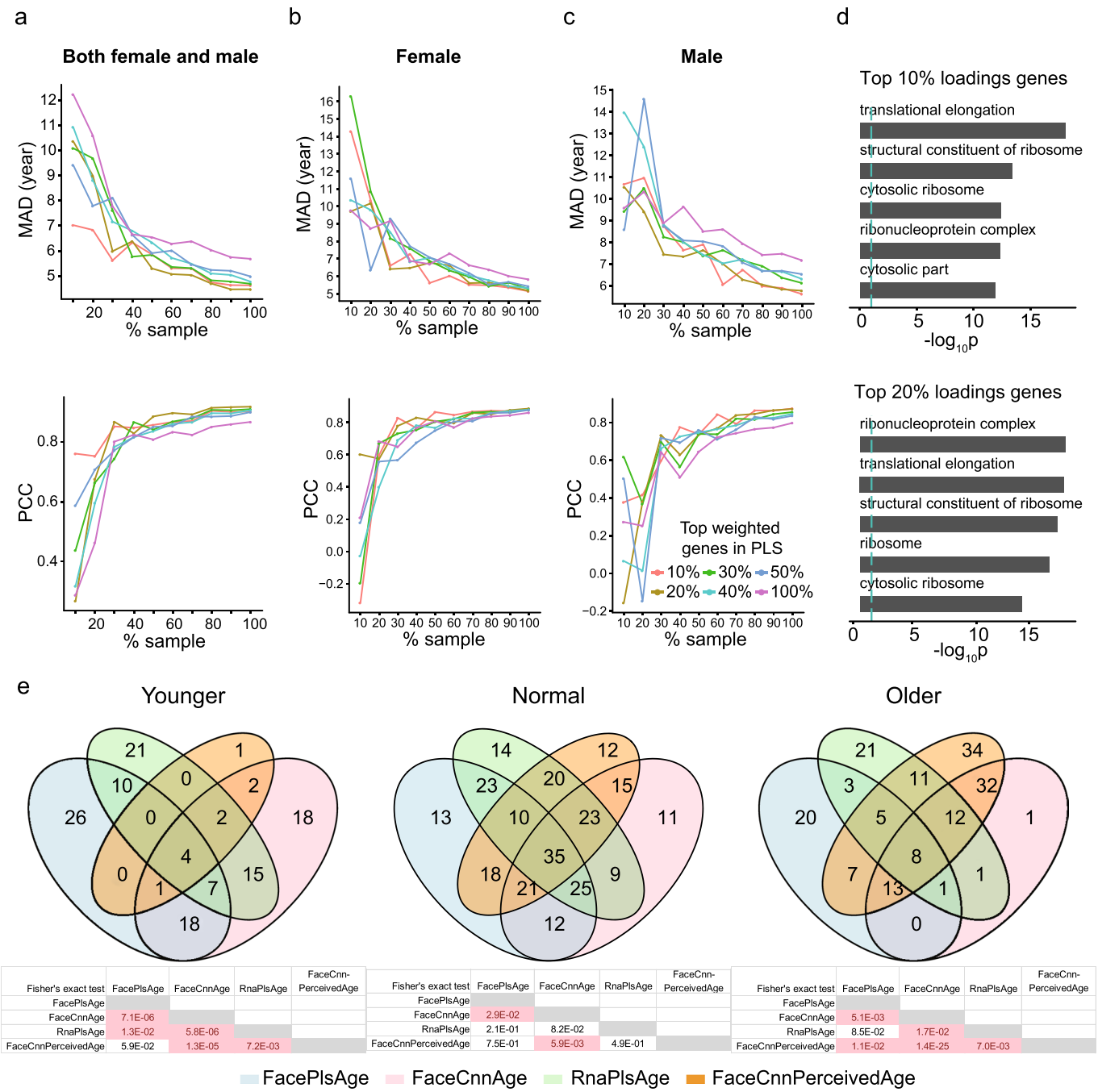


Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Heterogeneity of aging rate at different ages in Jidong and Beijing cohort. a, The standard deviation of randomly guess a number between 20–85 (n=4719, the boxes show 25%, 50% and 75% quantile and whiskers show maximum and minimum value). **b**, Relationship between age and the standard deviation of four AgeDiffs with bin size as 20 in Jidong cohort. Data are presented as mean +/- SD. **c**, Heatmap of aging-related facial features, health parameters and RNAs in Beijing (2012) cohort sorted by increasing chronological age (PCC with age, FDR < 0.05). Features were ranked by PCC from low to high (ALB: albumin, A/G: albumin/ globulin, TP: total protein, GGT: glutamyl transpeptidase, ALP: alkline phosphatase, CREA: creatinine, CHO: total cholesterol). **d**, **e**, Relationship between age and the standard deviation of four AgeDiffs with bin size as 20 (**d**) and 10 (**e**) in Beijing (2012) cohort. Data are presented as mean +/- SD.



Extended Data Fig. 4 | Broken stick regression of SD of AgeDiffs against age. a, **b**, Broken stick regression in Jidong cohort with bin size 100 (**a**) and 20 (**b**). **c**, **d**, Broken stick regression in Beijing (2012) cohort with bin size 20 (**c**) and 10 (**d**).

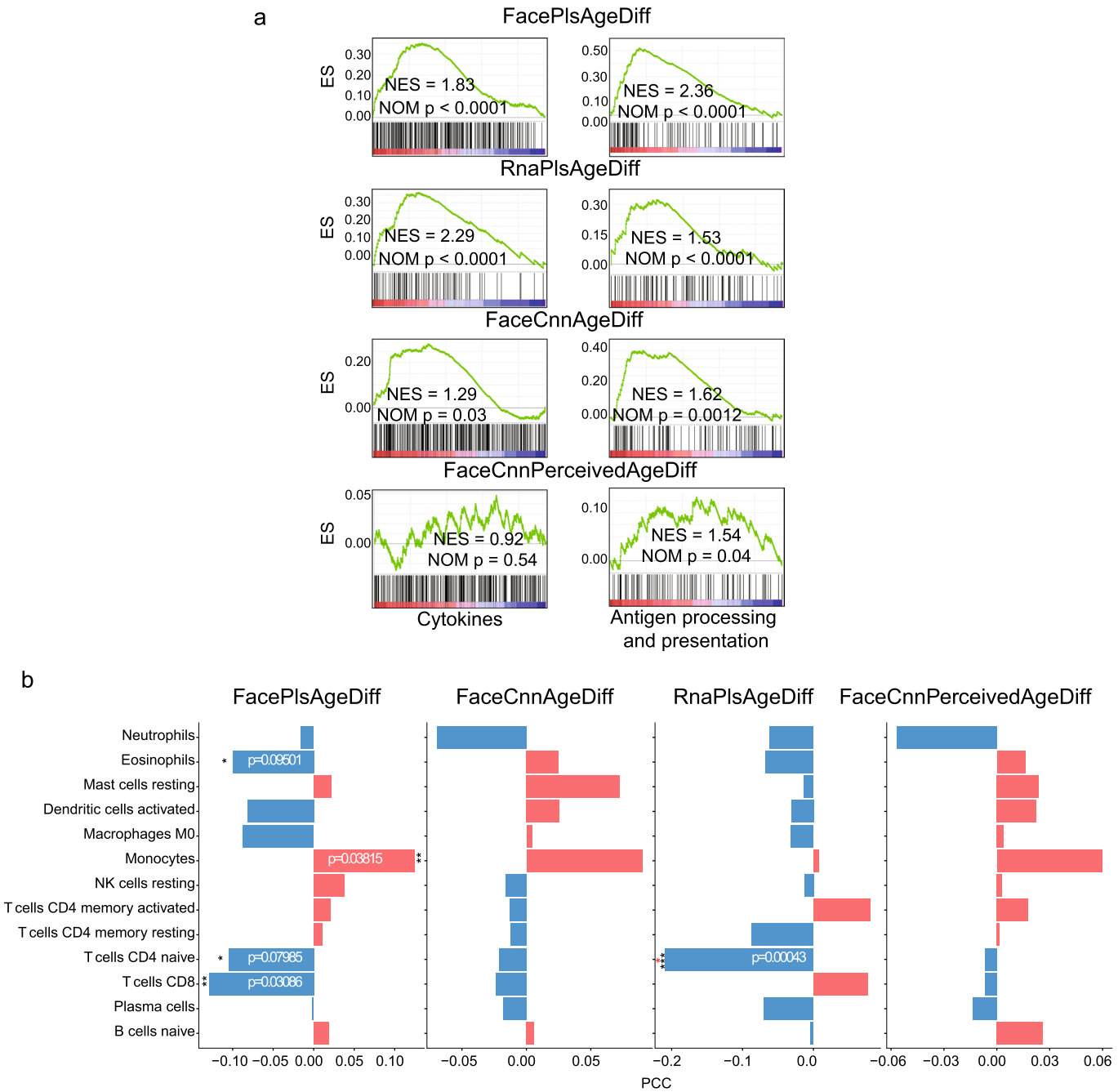


Fisher's exact test	FacePlsAge	FaceCnnAge	RnaPlsAge	FaceCnn-PerceivedAge
FacePlsAge				
FaceCnnAge	7.1E-06			
RnaPlsAge	1.3E-02	5.8E-06		
FaceCnnPerceivedAge	5.9E-02	1.3E-05	7.2E-03	

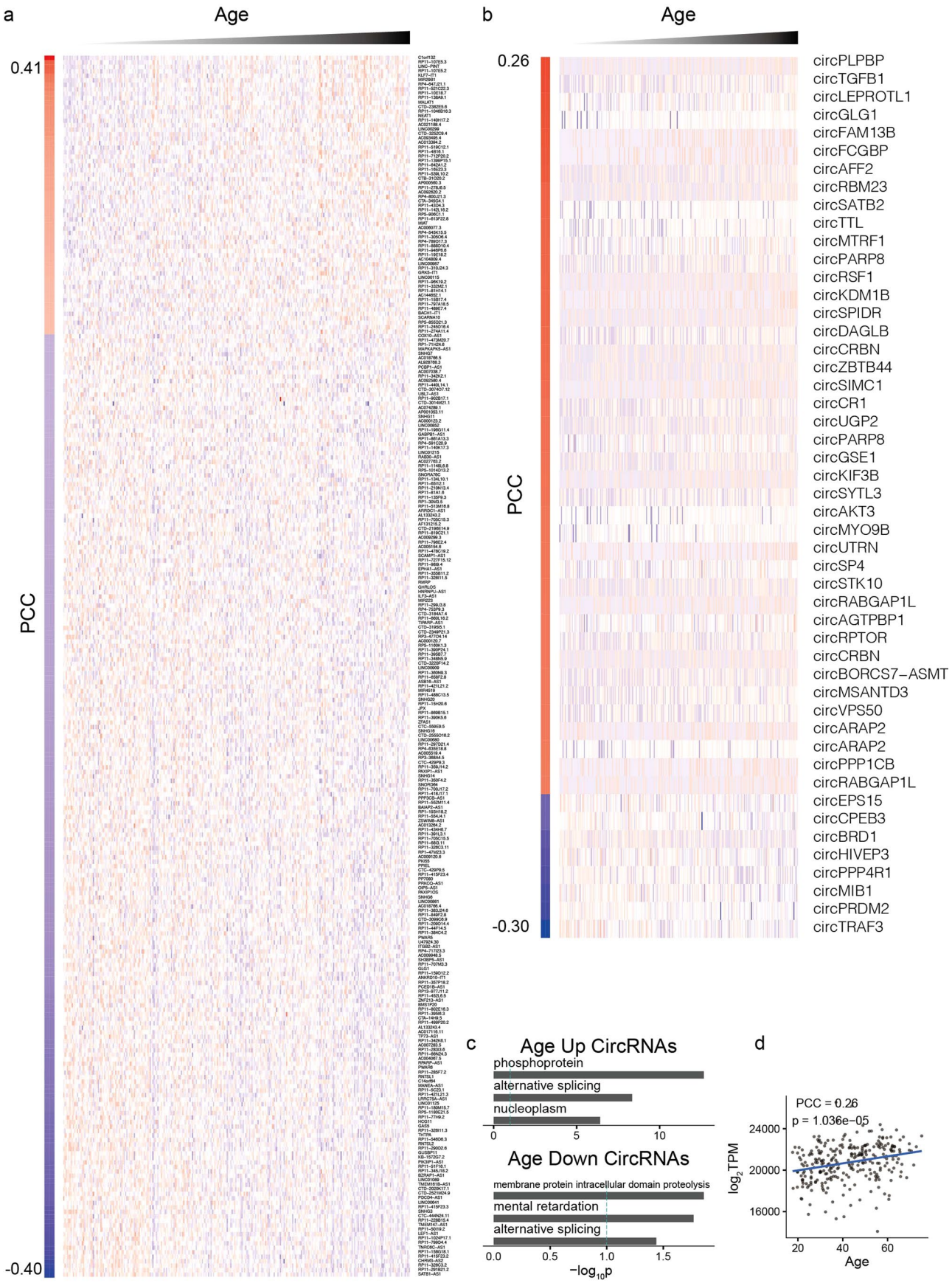
Fisher's exact test	FacePlsAge	FaceCnnAge	RnaPlsAge	FaceCnn-PerceivedAge
FacePlsAge				
FaceCnnAge	2.9E-02			
RnaPlsAge	2.1E-01	8.2E-02		
FaceCnnPerceivedAge	7.5E-01	5.9E-03	4.9E-01	

Fisher's exact test	FacePlsAge	FaceCnnAge	RnaPlsAge	FaceCnn-PerceivedAge
FacePlsAge				
FaceCnnAge	5.1E-03			
RnaPlsAge	8.5E-02	1.7E-02		
FaceCnnPerceivedAge	1.1E-02	1.4E-25	7.0E-03	

Extended Data Fig. 5 | Age prediction using transcriptome. a–c, Mean absolute difference (MAD) (top panel), Pearson correlation coefficient (PCC) (bottom panel) saturation analysis and correlation against chronological age of transcriptomes PLS age prediction for all the samples (**a**), female (**b**) and male (**c**), respectively. **d**, Enriched GO biological processes terms of PLS top 10% (upper) or 20% (lower) loading genes. P values are derived from hypergeometric test (Methods). **e**, Overlap between FacePlsAge, FaceCnnAge, RnaPlsAge, and FaceCnnPerceivedAge in younger (left), normal (middle), and older (right) samples. The lower table shows the one-tailed Fisher's exact test p-values with light red highlight indicate $p < 0.05$.



Extended Data Fig. 6 | Associations between cell types and AgeDiffs. a, Cytokines (left panels) and antigen processing and presentation (right panels) enrichment scores as a function of four AgeDiffs. P values are derived from permutation test (Methods). **b**, Association of RNA-seq deconvoluted cell type fractions and AgeDiffs (* p<0.1, ** p<0.05, *** p<0.01 derived from two-sided t test, and * Benjamini-Hochberg correction derived FDR < 0.1).



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Expression profile of expressed non-coding RNAs against chronological age. a, The heatmap of expressed lncRNAs (FPKM > 2) significantly related to chronological age (FDR < 0.1). The samples (columns) were sorted by age and lncRNAs were sorted by PCC of expression to age from high to low. **b**, The heatmap of expressed circRNAs (TPM > 2) significantly related to chronological age (FDR < 0.1). The samples (columns) were sorted by age and circRNAs were sorted by PCC of expression to age from high to low. **c**, Top three enriched terms for parent genes of age-up (top) and age-down circRNAs. P values are derived from hypergeometric test (Methods). **d**, The correlation between age and total expression level of circRNAs. P value is derived from two-sided t test.

nature research

Corresponding author(s): Jing-Dong Jackie Han

Last updated by author(s): Jul 1, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	3dMDface system (www.3dmd.com) was used for 3D facial images collection.
Data analysis	python (version 2.7.13) and R (version 3.5.0) were used for data analysis. Cytoscape (version 3.5.0) was used for network visualization. RNA-seq analysis: Mapping was done by STAR (version 2.4.0d). Quantification was done by Cufflinks (version 2.2.1). Circular RNAs were quantified by CIRCexplorer2 (version 2.3.5) and Tophat-Fusion (implemented in TopHat version 2.1.1). dlib (version 19.8.0) and TensorFlow (version 1.8.0) was used for facial image processing.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The regulatory genes in causal inference network building are from KEGG pathway (https://www.genome.jp/kegg/pathway.html), Reactome (https://www.reactome.org/), Animal Transcription Factor Database (http://www.bioguo.org/AnimalTFDB/), the human DEPhOsphorylation Database (http://www.depod.bioss.uni-freiburg.de/), IUPHAR/BPS database (http://www.guidetopharmacology.org/) and ImmPort35 (https://www.immport.org/shared/genelists). ImageNet (http://image-net.org/) were used for pre-train.
Facial image, transcriptome and lifestyles associations results are searchable at http://www.picb.ac.cn/hanlab/hub-fi/. 3D images and other personal identification sensitive meta data cannot be publicized or shared according to our participant consent agreement. Individual sequencing raw data, as they contain genetic information, will be available on request under the condition of approval of the ethics committee of Shanghai Institute of Nutrition and Health, Chinese Academy of

Sciences abiding China Human Genetic Resource law. Mapped read counts and FPKM expression values of coding genes from the RNA-seq are deposited to the public repository NODE at https://www.biosino.org/node/project/detail/OEP001041.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	In CNN model training for 3D facial images, we use epoch VS MAD/PCC plot to show the sample size is sufficient for age prediction in current setup (Extened Data Fig. 1a). In PLS model training for RNA-seq, we use saturation analysis (percentage of sample size VS MAD/PCC) to show the sample size is sufficient for age prediction among all samples Extened Data Fig. 5a).
Data exclusions	In Jidong cohort study, we originally had more than 5000 samples and we excluded the samples with no age information or low quality 3D facial images, which ended up with 4719 samples. In Beijing (2012) cohort study with RNA-seq, 286 samples were sent for RNA sequencing but 6 of them did not pass the quality control therefore excluded for all the downstream analysis.
Replication	We use Beijing (2012) and Beijing (2015) as independent test datasets for the two CNN based models. And the result is consistent in these two datasets as CNN based models are more accurate than PLS based model in age prediction.
Randomization	This study does not have experimental groups. All the covariates can be controlled by linear model.
Blinding	Blinding was not relevant in this study because there was no treatment to the participants.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	The age and gender distribution of Jidong, Beijing (2012) and Beijing (2015) cohorts for facial image model training and tests are shown in Table S1, S3, S4, respectively. The age and gender distribution for Beijing (2012) cohort with corresponding RNA-seq data is listed in Table S5. All participants in this study were Han Chinese. No treatment was adopted in this study.
Recruitment	Jidong cohort were recruited in Staff Hospital of Jidong oil-field when they took annual physical examination. The population structure in this area is relatively young, which may lead to under representation of aging rate associations in the old ages. Beijing (2012) and Beijing (2015) cohorts were both recruited by advertisement voluntarily with requirement of being healthy. No clear impact of this recruitment strategy was noted.
Ethics oversight	Jidong study was approved by Ethical Committees of the Staff Hospital of Jidong oil-field of China National Petroleum Corporation. Beijing study was approved by the ethics committee of the Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences.

Note that full information on the approval of the study protocol must also be provided in the manuscript.