Biological age reveals personal skin aging variations
- An innovative skin aging index based on non-invasive internal skin RNA expression

Shodai Tanaka[1], Naoki Oya[1], Hiroki Yanagisawa[2], Yasuko Yoshida-Amano[2], Yoshito Takahashi[3], Takayoshi Inoue[1*]

[1] Biological Science Research, Kao Corporation, 2-1-3, Bunka, Sumida, Tokyo, 131-8501, Japan
[2] Skincare Products Research, Kao Corporation, 2-1-3, Bunka, Sumida, Tokyo, 131-8501, Japan
[3] Biological Science Research, Kao Corporation, 5-3-28, Kotobuki-cho, Odawara, Kanagawa, 250-0002, Japan

* Takayoshi Inoue, +8170-3297-0982, inoue.takayoshi@kao.com

Abstract

Progress of aging varies among individuals. Although chronological age is one of the most general indexes of aging, it doesn't always reflect a degree of aging accurately. Recently, the concept of "biological age", estimated by combining comprehensive biological information associated with aging and machine learning, has been attracting global attention, because of its usefulness in aging research. Therefore, we considered that biological age in the skin would precisely reflect the degree of skin aging, which cannot be determined by chronological age alone, in addition to providing insights into skin aging.

Skin Surface Lipid-RNA (SSL-RNA) monitoring, our original technology, enables the non-invasive acquisition of comprehensive gene expression profile in the skin. To examine whether SSL-RNA profiles associate with skin aging, we extracted genes significantly correlated with chronological age and gene ontology analysis was performed. As a result, these genes showed consistent changes with previously reported aging-related biological functions. Next, we predicted chronological age by machine learning and defined it as biological age based on SSL-RNA profile. Then, we examined the association of biological age with skin aging. As a result, significant correlation was obtained between skin aging conditions and biological age, but not with chronological age. These findings suggest that biological age reflects the degree of skin aging more accurately than chronological age.

We consider that biological age is an innovative index useful for examination of skin aging that provides more functional understanding of that factor in individuals in a timely manner as well as a personalized cosmetic future plan.

Keywords: Skin Surface Lipid-RNA, Noninvasive, Aging, Biological age, Transcriptome, Machine learning

1. Introduction.

Aging is an inevitable process, and its degree of progression and phenotype vary among individuals, owing to genetic predispositions and environmental factors. Skin aging is no exception, and the effects of ultraviolet radiation can further accelerate and complicate this process, leading to wrinkles, hyperpigmentations, and other appearance-impairing phenotypes. Chronological age (the number of years since birth) is one of the most common indexes of aging, but elapsed time is not always accurate indicator of the degree or phenotype, so there can be variations in aging even if among the same chronological age individuals. In recent years, the concept of "biological age", an estimate of the degree of aging progression based on the amount of decline in physical functions, has been attracting global attention [1]. Biological age has been estimated by combining comprehensive biological information, such as DNA methylation and RNA expression, associated with aging, and machine learning which is capable of handling "big data" [1,2]. It was hypothesized that by estimating biological age in the skin, it would be possible to obtain an index that would precisely reflect the degree of skin aging, which cannot be determined by chronological age alone, in addition to providing insights into skin aging.

Our previous studies have found that human mRNA is present in sebum and have termed it "skin surface lipid-RNA" (SSL-RNA). In addition, transcriptome analysis technology led us to obtain comprehensive gene expression profile in the skin from SSL-RNA, acquired by non-invasive method that comprises simply wiping sebum using an oil-blotting film [3]. Furthermore, differentially expressed genes involved in pathology between atopic dermatitis patients and healthy subjects could be detected in SSL-RNA [3], and gene expression changes were detected to be consistent with physical changes associated with circadian rhythms and menstrual cycles, suggesting that SSL-RNA reflects various skin and body conditions.

In the present study, whether SSL-RNA reflects age-related changes was first examined. In addition, to understand the personal skin aging, chronological age was predicted by combination SSL-RNA profiles and machine learning, and whether the predicted age was a reasonable value related to skin aging, as a biological age based on SSL-RNA, was examined. Furthermore, whether biological age based on SSL-RNA can identify individual differences in skin aging that cannot be detected by chronological age alone was determined.

2. Materials and Methods.

2-1. Subjects

A clinical study was performed with healthy Japanese women aged 20 to 59 years. The numbers of subjects in each age group are shown in Table 1. The study was reviewed and approved by the Clinical Research Ethics Committee of Kao Corporation, and subjects were fully informed about the study and provided written, informed consent. Sebum samples were obtained from all subjects' whole faces with a single sheet of oil-blotting film, and sebum samples were stored at −80°C until RNA extraction. Age-related skin parameters were measured by the following methods: (i) using a Cutometer® to determine skin visco-elasticity (R2: Ua/Uf), net elasticity (R5: Ur/Ue), portion of visco-elasticity (R6: Uv/Ue), and portion of elasticity (R7: Ur/Uf); (ii) three-dimensional analysis of eye corner replicas to determine center line average roughness (Ra), 10-point average roughness (Rz), and the highest peak (Rmax); (iii) stratum corneum tape stripping for the stratum corneum cell area and relative CML (Nε-(Carboxymethyl)lysine) in the stratum corneum; (iv) using an AGE (Advanced Glycation End Products) Reader® to assess the skin autofluorescence that indicates the skin glycation level; (v) visual evaluation for the hyperpigmentation score.

Table 1. The number of subjects

chronological age	sample size
20s	37
30s	40
40s	39
50s	38

2-2. Acquisition of SSL-RNA profiles

SSL-RNA was extracted from sebum samples using QIAzol reagent (QIAGEN) and chloroform, and then purified with the RNeasy Mini kit (QIAGEN). Extracted RNA was subjected to an Ion AmpliSeq Transcriptome Human Gene Expression Kit (Thermo Fisher Scientific), with a slightly modified protocol. Briefly, SSL-RNA was reverse-transcribed and then multiplex amplification was performed. The amplicons were purified with AMPure XP (Beckman Coulter) and then confirmed by Agilent 4200 TapeStation (Agilent). Subsequently, the final library was constructed after digesting the excess primer and ligating adaptors. Libraries were quantified using an Ion Library TaqMan Quantitation Kit (Thermo Fisher Scientific) and diluted libraries were sequenced with Ion GeneStudio S5 Prime (Thermo

Fisher Scientific). Sequence data were subjected to the primary analysis using the AmpliSeq RNA plugin of the Ion Torrent Suite Software Plugins (Thermo Fisher Scientific).

2-3. Transcriptome analysis

For data quality control (QC), samples with percentage of genes detected (Targets Detected) values greater than 20% (calculated from the AmpliSeq RNA plugin) were selected, and genes with non-zero read counts in more than 90% of the samples were selected. Normalized counts were obtained from read counts data by normalization with DESeq2 R package. Normalized counts were subsequently converted to $\log_2$(normalized counts + 1) for approximating to normal distribution. Spearman's rank correlation analysis between chronological age and each gene expression was performed, then extracted significantly correlated genes (p-value < 0.05).

2-4. Gene ontology (GO) analysis

For genes significantly correlated with chronological age as described above, gene ontology enrichment analysis was performed using PANTHER (http://pantherdb.org/) and a statistical analysis to calculate false discovery rate (FDR) was with Fisher's exact test. GO terms significantly enriched in the gene sets were extracted (FDR < 0.05).

2-5. Machine learning

Supervised machine learning was conducted with the classification and regression training (caret) R package. For data that passed QC in the Transcriptome analysis section, read counts data were normalized by reads per million (RPM) method and converted to $\log_2$(RPM + 1) for approximate the normal distribution. Chronological age regression models were constructed based on gene expression that were significantly correlated with chronological age as features. Specifically, the samples (RPM data) were divided into 50% training data and 50% test data through stratified sampling. As data pre-treatment, missing value imputation by the k-nearest neighbor algorithm (k = 5) and standardization were performed for the training data features. Regression model were constructed by 10-fold cross-validation of the training data, and the optimal parameters of the indicated 10 algorithms were selected by a grid search: linear multiple regression, penalized linear regression (Lasso, ridge

regression, and elastic net), partial least squares (PLS) regression, decision tree, random forest, linear kernel support vector machine (SVM), polynomial kernel SVM, and Gaussian kernel SVM. The root mean square error (RMSE) of the measured and predicted values were used as precision evaluation indices, and the algorithm and parameter combination with the lowest RMSE among the constructed models was adopted as the best model. Finally, the generalization performance of the constructed models was verified using test data after the same pre-treatment as that for the training data.

2-6. Examination of the relationships between chronological/biological age and skin parameters

Pearson's correlation analysis between chronological age and age-related skin parameters in all subjects was performed. For the hyperpigmentation score only, Spearman's rank correlation analysis was performed. In addition, skin parameters for each age group were statistically compared using the Tukey-Kramer test. For each of the chronological age groups (20s, 30s, 40s, and 50s), skin parameters for the oldest 25% and youngest 25% of the population at biological age, were statistically compared using the Welch's t-test, and the Mann-Whitney U-test was only used for the hyperpigmentation score. In addition, For the test data described in Machine learning section, Spearman's rank correlation analysis was performed between biological age and skin parameters and between chronological age and skin parameters, at chronological age in the 20s, 30s, 40s, and 50s.

3. Results.

3-1. Subjects' skin aging characteristics

For the investigation with the aim of understanding skin aging, it was to be confirmed that the subjects constituted a population with general skin aging characteristics. Correlation analysis between chronological age and age-related skin parameters was performed for all subjects aged 20-59 years. The results indicated that all skin parameters, that is skin elasticity (R2, R5, R6, and R7) measured with a Cutometer®, skin surface roughness (Ra, Rz, and Rmax) obtained via three-dimensional analysis of skin replica, the stratum corneum cell area and stratum corneum CML relative value measured on stratum corneum tape stripping, the

skin glycation level measured with an AGE Reader, and the hyperpigmentation score judged visually by an expert evaluator, showed significant correlations with chronological age ($p < 0.05$; Fig.1A-K). Furthermore, comparisons of the skin parameters for each age group tended to show the more difference between subjects in their 30s and 40s, and the distribution tended to be greater for subjects in their 40s than for those in their 30s (Fig.1L-N). It is known that these skin parameters, represented by surface roughness (wrinkles) and hyperpigmentation, change with aging [4], and this tendency was found in our results. These results show that the subject population in this study had the general characteristics of skin aging, with changes being more marked in subjects in their 40s.

3-2. Age-related changes in SSL-RNA

As it was to be confirmed that the subjects have characteristics of skin aging, it was deemed possible to investigate skin aging in this population. In previous studies about biological age, the degree of aging was estimated based on biomolecular information that changes with aging. In the present study, whether SSL-RNAs expression show age-related changes was examined, for which the expression levels of 2,323 genes in 113 samples that had passed QC were normalized and approximated to normal distributions by logarithmic transformation. Spearman's rank correlation coefficient (*rho*) and significance probability value (*p*-value) between normalized expression levels for each gene and chronological age were calculated to extract genes that change with aging. As a result, 189 genes showing significant positive correlations with chronological age ($rho > 0$ and $p < 0.05$) and 179 genes showing significant negative correlations with chronological age ($rho < 0$ and $p < 0.05$) were obtained.

Gene ontology (GO) analysis was performed for each of the 189 positively and 179 negatively correlated genes. The results indicated that 386 GO terms, such as "response to organic substance (GO:0010033)" and "immune system process (GO:0002376)", were significantly enriched in the positively correlated genes (FDR < 0.05). On the other hands, 153 GO terms, such as "macromolecular catabolic process (GO:0009057)" and "oxidative phosphorylation (GO:0006119)", were significantly enriched in genes showing a negative correlation (FDR < 0.05). In a previous transcriptome analysis with skin biopsies of healthy, Caucasian females, it was found that the expression of genes involved in the immune response and cellular senescence increases with age, whereas the expression of genes

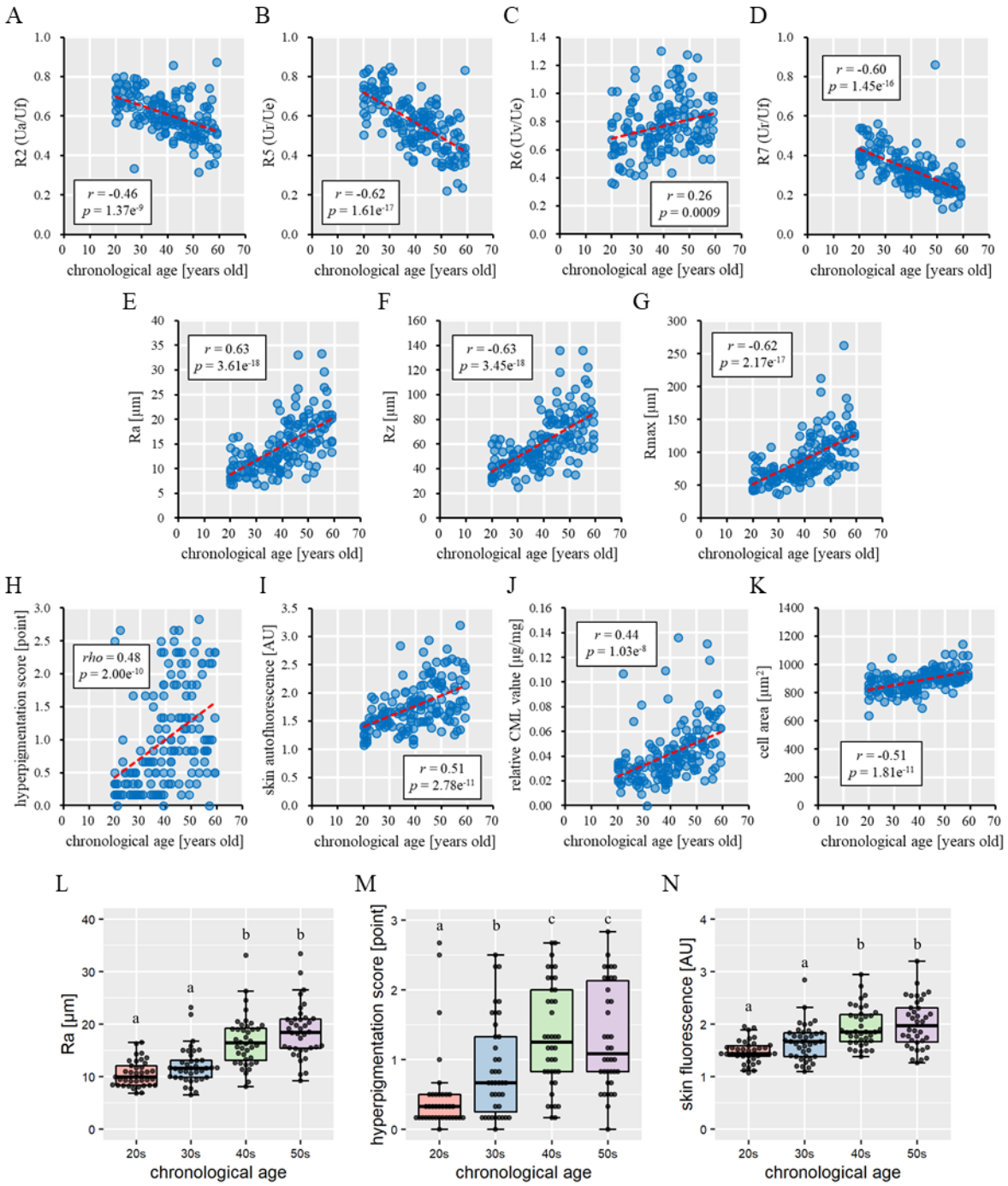

Figure 1. Subjects' skin aging characteristics

(A-K) Scatter plots between chronological age and each skin parameters in all subjects (20-59 years old). The dashed red lines indicate linear trendlines. Pearson's correlation coefficient (r) and Spearman's rank correlation coefficient (rho), and probability value (p) were calculated and shown. (L-K) Boxplots of skin parameters for each chronological age group (20s, 30s, 40s, and 50s). Skin parameters were statistically compared using Tukey-Kramer test, and different alphabets indicate that there is a statistically significant difference ($p < 0.05$) between the two groups.

associated with metabolism, including mitochondrial function, decreases [5]. In the present study, it was found that immune response functions were enriched in genes showing positive correlations. Moreover, essential transcription factors regulating the immune response, such as NFKB1/2 (nuclear factor kappa B subunit 1/2) and RELA (v-rel avian reticuloendotheliosis viral oncogene homolog A), typical cytokines and matrix metalloproteinases, such as TNF (tumor necrosis factor) and MMP9 (matrix metalloproteinase 9), and the cellular senescence factor CDKN1A (cyclin dependent kinase inhibitor 1A) were changed with aging. In addition, mitochondrial functions, such as oxidative phosphorylation and ATP metabolism, were enriched in negatively correlated genes, clearly showing that differential gene expression in SSL-RNA was consistent with previously reported age-related changes. In summary, it was suggested that SSL-RNA provides information about age-related changes.

3-3. Construction of chronological age regression model by machine learning

Up to this point, genes with differential expression with aging (i.e., with a significant correlation with chronological age) were identified in SSL-RNA, and it was suggested that these gene sets provide information about age-related information. In this study, to investigate the possibility of obtaining an age-related index based on SSL-RNA, we attempted to construct a chronological age, a general index of aging, regression model based on all 368 genes identified in the previous section as features. The 113 samples that passed QC and for which the expression levels were normalized and transformed to $\log_2(\mathrm{RPM} + 1)$ were split into 50% training data (n = 58) and 50% test data (n = 55), such that the chronological age distribution was consistent. In addition, each feature of the training data was pre-treated via missing value imputation and standardization. With these training data, model construction and optimal parameter selection were performed by 10-fold cross-validation for the 10 algorithms described in the Machine learning section. As a result, the model using partial least squares (PLS) regression had the lowest resampled RMSE, and thus, the algorithm and selected parameters combination were considered to present the best model. Test data, which were not involved in the model construction, were then inputted to verify the generalization performance of the constructed model. The Pearson's correlation coefficient (r) between the predicted values and measured values (chronological age) was

calculated, and a strong correlation was found, r = 0.77, indicating that the constructed model had high generalization performance. These findings show that machine learning can be used for regression analysis of chronological age based on SSL-RNA.

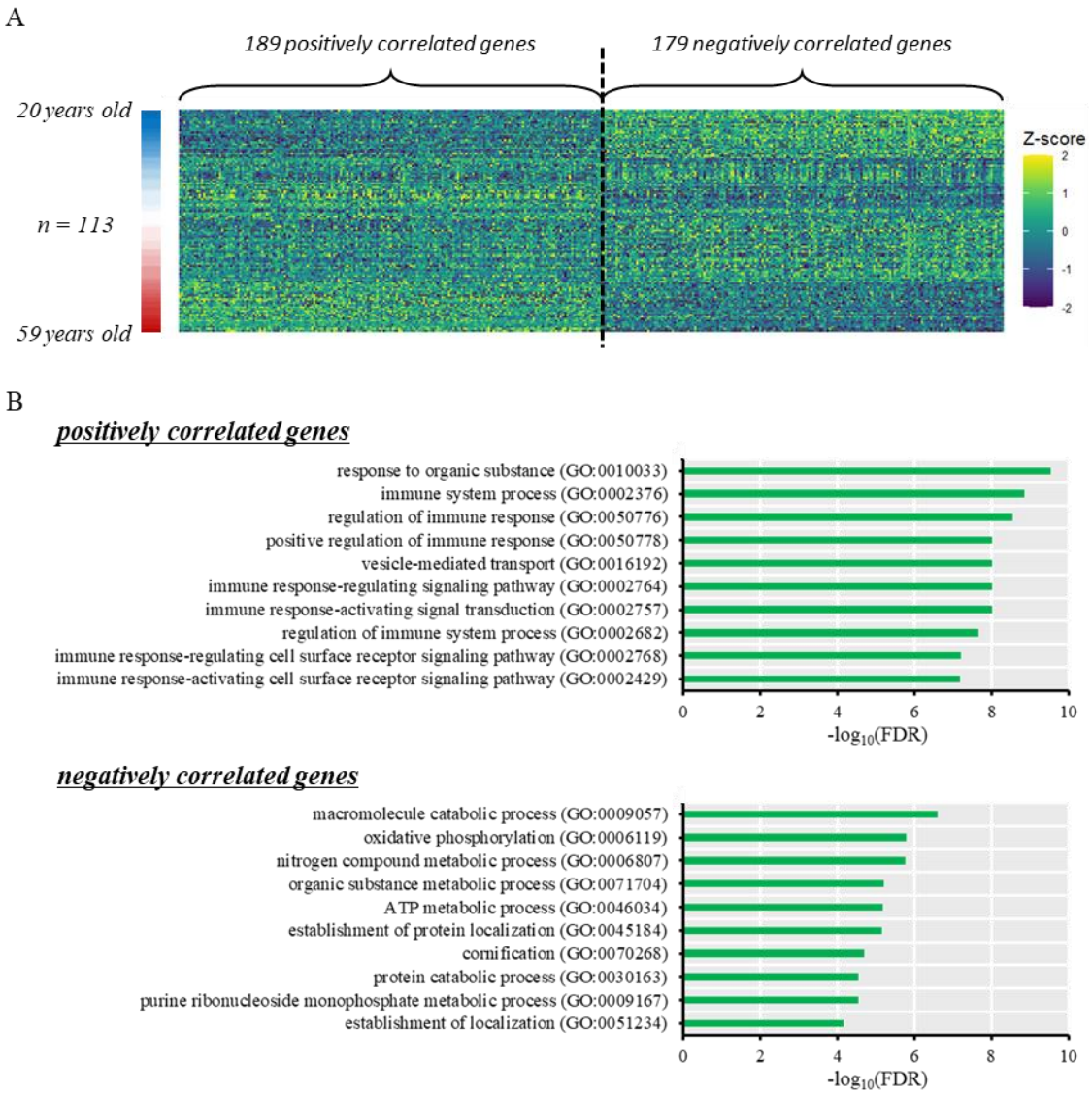

Figure 2. Age-related changes in gene expression and biological processes in SSL-RNA

(A) Heatmap of gene expression which significantly correlated with chronological age ranging from 20 to 59 years old. Values are shown as standardized z-score of gene expression ($\log_2$(normalized counts + 1)). The yellow and blue colors show the high and low expression, respectively. (B) Bar charts showing top 10 of unique significant biological process (BP). The upper panel shows top 10 of BP enriched in positively correlated genes and the lower panel shows top 10 of BP enriched in negatively correlated genes. Values are shown as common logarithmic transformed the reciprocal of FDR.

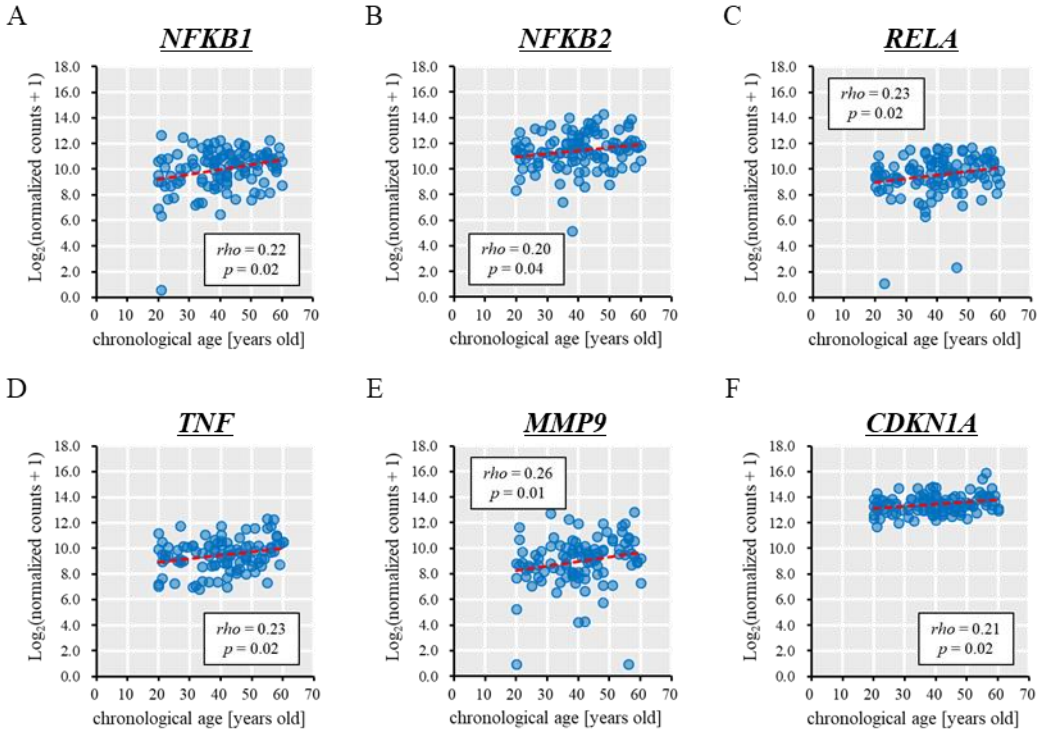

Figure 3. Correlation between gene expression and chronological age
(A-F) Scatter plots between chronological age and each gene expression ($\log_2$(normalized counts + 1)). The dashed red line indicates a linear trendline. Spearman's rank correlation coefficient (*rho*), and probability value (*p*) were calculated and shown.

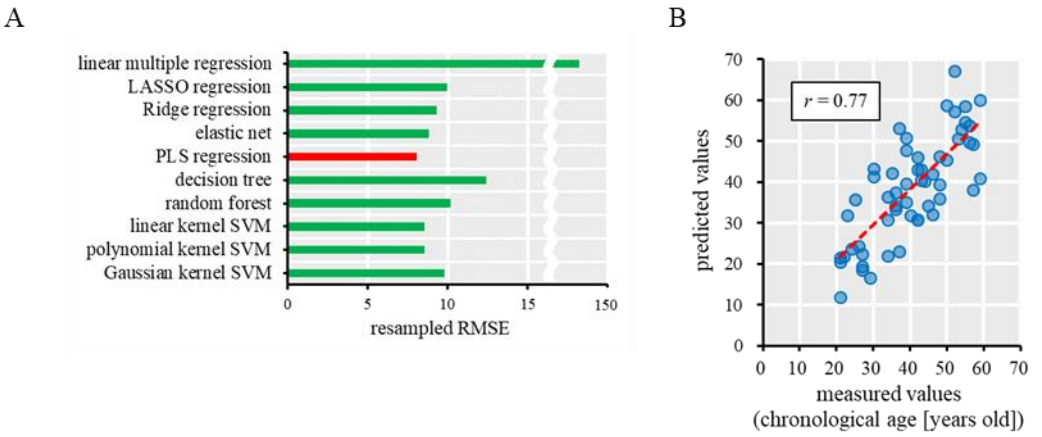

Figure 4. Construction of chronological age regression model based on SSL-RNA
(A) Bar chart showing resampled RMSEs calculated by 10-fold cross validation for the 10 algorithms in the training data. Red bar shows an algorithm with the lowest resampled RMSE. (B) Scatter plot between measured values (chronological age) and predicted values. The dashed red line indicates a linear trendline. Pearson's correlation coefficient was calculated and shown.

3-4. The definition of biological age and examination of the relationship with skin aging

Although it was shown that machine learning can be used for regression analysis of chronological age based on SSL-RNA, there were some subjects with high predicted values and others with low values despite being in the same chronological age group. As the constructed regression model based on gene expression that changed with aging as the feature, the differences in predicted values show the degree of deviation from the average gene expression in that age group. Therefore, it was considered that differences in predicted values might reflect differences in the degree of aging in individuals. In this context, whether this index is related to skin aging was examined when the chronological age predicted by machine learning was taken to be the biological age based on SSL-RNA. For the test data in the previous section, correlation analysis between the predicted age and skin parameters revealed significant ($p < 0.05$) correlations for all items except the stratum corneum cell area and R6. These findings suggest that the chronological age predicted by machine learning is associated with skin aging. Therefore, this value was defined as the biological age based on SSL-RNA.

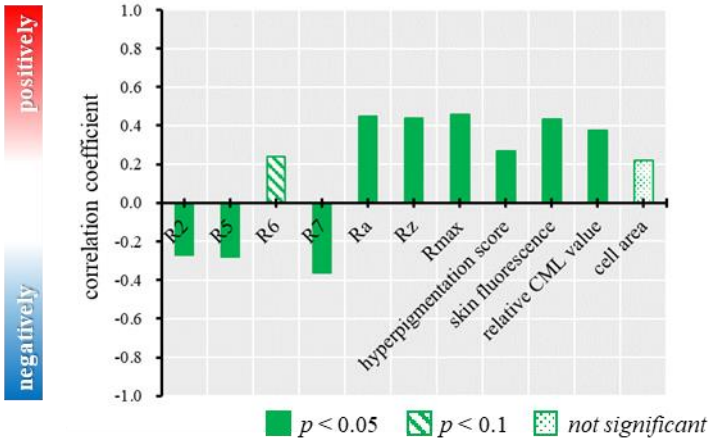

Figure 5. Relationships between predicted chronological age and age-related skin parameters

(A) Bar chart showing correlation coefficient between predicted chronological age and skin parameters. Fill patterns show probability values in statistical tests.

3-5. Verification for superiority of biological age based on SSL-RNA over chronological age

To verify that biological age based on SSL-RNA reflects the degree of skin aging more accurately than chronological age, which is simply elapsed time, whether differences in

biological age alone could detect differences in the degree of skin aging between two groups with no difference in chronological age was examined. Therefore, based on a group with chronological age in their 40s, which is considered that skin aging starts to be apparent, the degree of skin aging was compared between groups with different biological ages within the 40s. Specifically, with test data, age-related skin parameters were compared between the four subjects in the top 25% of biological age (*older*) and the four subjects in the bottom 25% of biological age (*younger*), in their 40s (n = 14). As a result, in two groups with no significant differences in chronological age but with significant differences in biological age, *older* (i.e. group has high biological age) showed slightly lower R5, and significantly or slightly higher Ra, Rmax, and AGE Reader measurements than *younger*. It was therefore confirmed that even in the same chronological age group, subjects with a higher biological age tend to have a lower skin visco-elasticity, a rougher skin surface, and a more advanced skin glycation, suggesting their tendency to have more advanced skin aging. These results suggest that biological age can be used to detect differences in the degree of skin aging between two groups with no difference in chronological age.

In addition, a study was performed with all subjects in their 40s. As detailed in Section 3-1, skin parameters related to skin aging show significant correlations with chronological age. However, with the test data, a restriction to subjects in their 40s results in loss of significant correlation between the chronological age and skin parameters, except skin surface roughness (Ra and Rz). Nevertheless, with respect to biological age, when correlation analysis with skin parameters was performed solely for subjects with chronological ages in their 40s, significant correlations were detected in skin visco-elasticity (R2, R5, and R7), skin surface roughness (Ra, Rz, and Rmax), and skin glycation (AGE Reader measurements). These findings suggest that biological age reflects the degree of skin aging more accurately than chronological age.

4. Discussion.

The most generally recognized index of aging is the number of years since birth, that is, chronological age. However, considering that some people require nursing care while others enjoy sports activities at the same chronological age, this value, which is simply elapsed time, is not a perfect aging index. In this context, the concept of "biological age" has been proposed to more accurately estimate the decline in physical functions due to aging. In recent years,

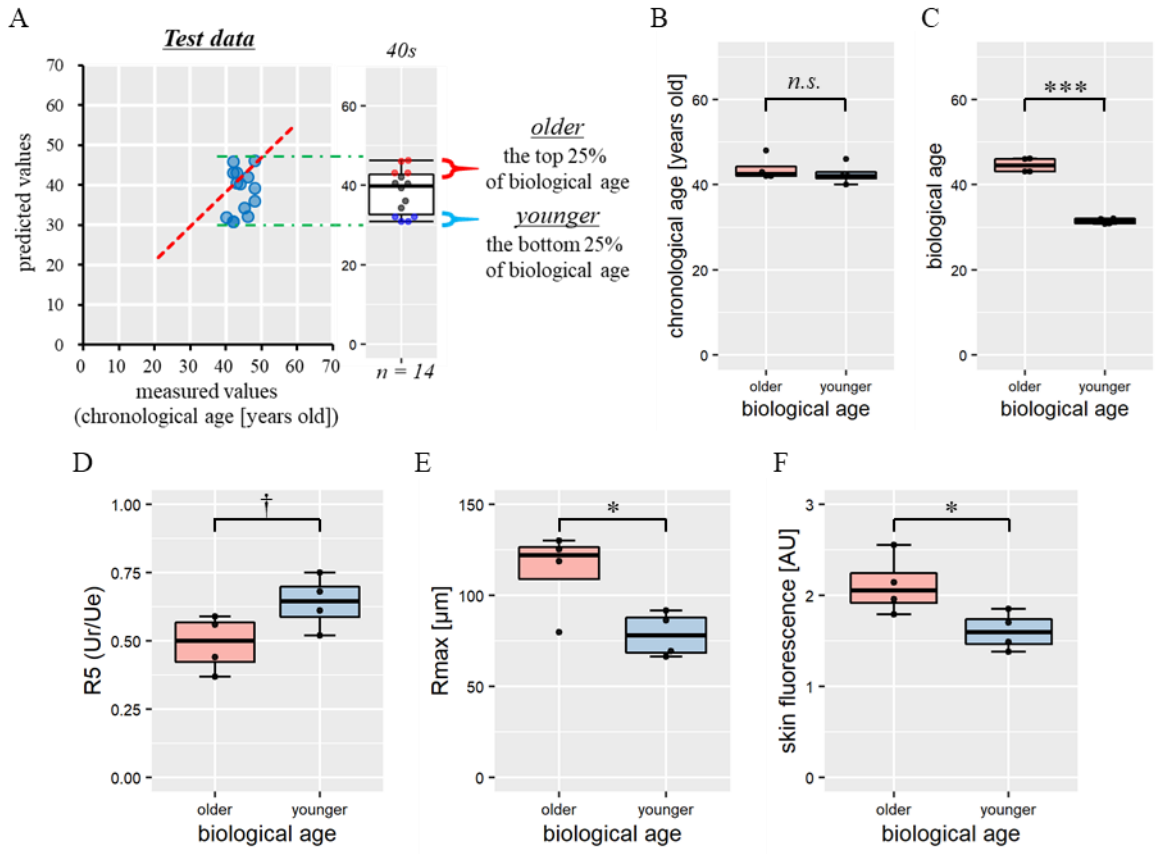

Figure 6. Comparison of skin parameters in subjects with high or low biological age

(A) Definition of populations with relatively high and low biological age in a group with chronological age in their 40s (n = 14). (B-F) Boxplots showing (B) chronological age, (C) biological age, and (D-F) skin parameters in two biological age groups (*older* and *younger*). Welch's *t*-test was perfomed for two groups : ***; $p < 0.001$, *; $p < 0.05$, †; $p < 0.1$, *n.s.*; not significant.

rapid advances in analytical technologies and the development of computational methods, such as machine learning, have enabled us to acquire a vast amount of biological information and the extraction of useful information. In connection with this, biological age has been estimated based on biological "big data", such as transcriptomic, proteomic, and metabolomics [2]. Above all, the epigenetic clock has attracted global attention as the most promising estimation method, having been proposed by Horvath *et al*. in 2013. In the epigenetic clock, the degree of DNA methylation is used as a feature to construct a chronological age regression model, and the predicted value is defined as the biological age based on the degree of DNA methylation [1]. The same estimation method was considered

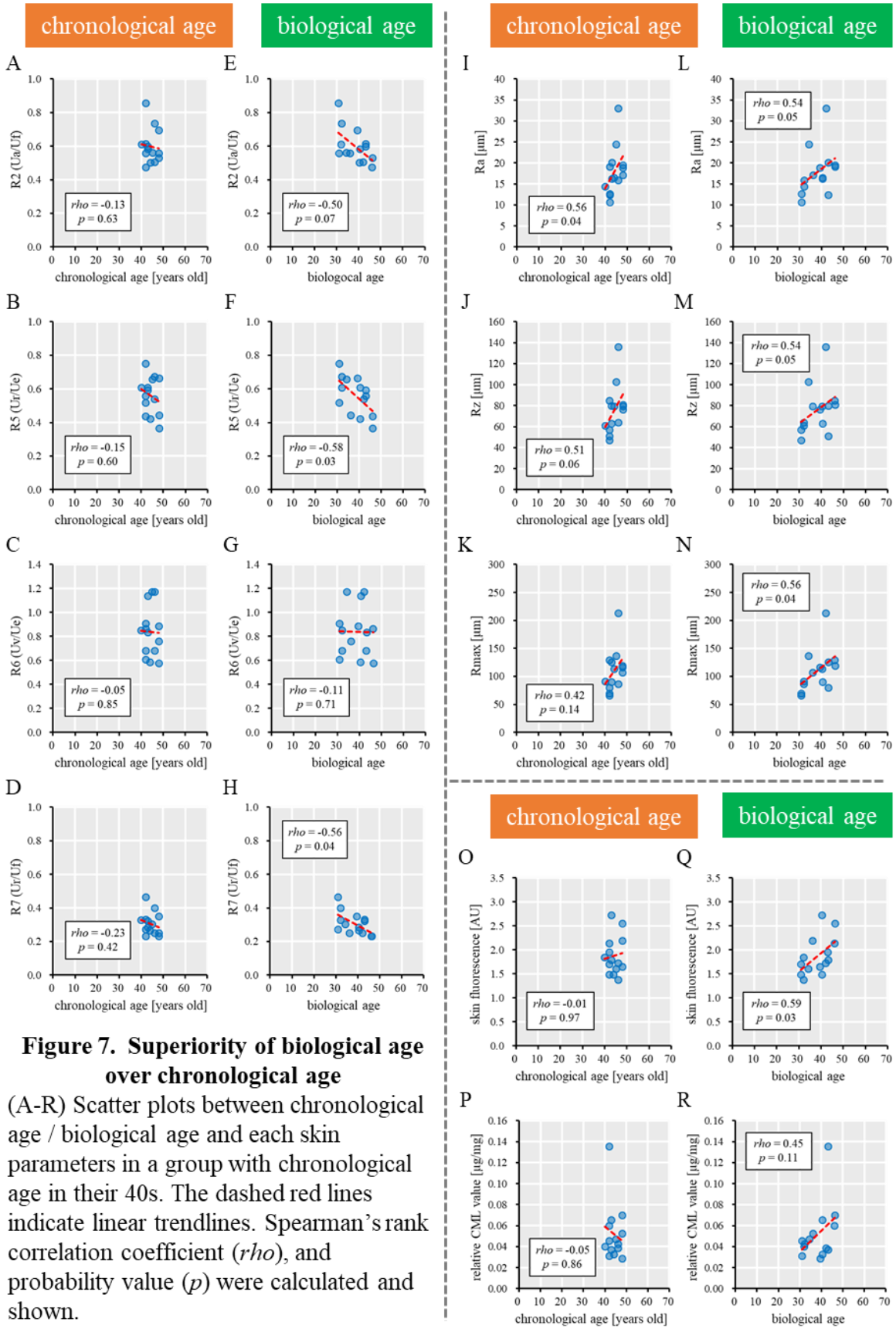

Figure 7. Superiority of biological age over chronological age

(A-R) Scatter plots between chronological age / biological age and each skin parameters in a group with chronological age in their 40s. The dashed red lines indicate linear trendlines. Spearman's rank correlation coefficient (*rho*), and probability value (*p*) were calculated and shown.

appropriate for the present study, and chronological age regression model based on the SSL-RNA profiles was constructed. As a result, high generalization performance of the constructed model was shown by the test data, which were not involved in model construction. Therefore, when that predicted chronological age is taken to be the biological age based on SSL-RNA, whether the biological age is associated with age-related skin parameters was examined. As a result, significant correlations with numerous skin parameters were detected, and taking this predicted value as the biological age based on SSL-RNA was considered reasonable. Furthermore, the verification of whether the estimated biological age reflected skin aging more accurately than chronological age showed that between two groups with no difference in chronological age but differences in biological age could detect differences in skin aging. In other words, the findings suggest that the genes used for estimation in this study, as features, are closely linked to skin aging, and it would be meaningful to assess the detected genes in more detail as target molecules for anti-skin aging technology.

The phenotype of skin aging, represented by wrinkles and hyperpigmentation, is difficult to improve once it becomes apparent, and preventive measures such as ultraviolet protection and moisturizing are important. However, in practice, it is difficult to recognize skin problems before the phenotype is apparent, and people therefore cannot readily realize the importance and effectiveness of preventive measures. Biological age based on SSL-RNA is estimated from gene expression information in the skin and can be used as an index of skin aging before the skin aging phenotype becomes apparent. It offers an advantage over methods for calculating the apparent age of the skin from facial photographs. In addition, in contrast with methods involving biopsy, sebum can be collected repeatedly from the same site, making it possible to track gene expression information in the skin, including biological age. Therefore, by repeatedly collecting sebum, the biological age and its progression at the time of collection can be estimated, and this method therefore offers a new index that can objectively be used to assess efforts made to prevent skin aging when the effects are not readily realized. Furthermore, research on tracking skin aging prevention and alleviation measures from the perspective of biological age should further elucidate the significance of this index. In other words, whereas chronological age is irreversible, it might be possible to improve biological age by reassessing care habits and environmental factors.

5. Conclusion.

This study showed that SSL-RNA, which constitutes non-invasively collected comprehensive gene expression internal skin, provides information on age-related changes in the skin. Furthermore, by combining this information on age-related changes with machine learning, it was determined to be possible to estimate individual differences in the degree of skin aging, which cannot be determined by chronological age alone, as the biological age. We consider that biological age is an innovative index useful for examination of skin aging that provides more functional understanding of that factor in individuals in a timely manner as well as a personalized cosmetic future plan.

6. Conflict of Interest Statement.

NONE.

7. References.

1. Horvath, S. (2013). "DNA methylation age of human tissues and cell types." Genome Biol **14**(10): 3156.
2. Jylhävä, J., et al. (2017). "Biological Age Predictors." EBioMedicine **21**: 29-36.
3. Inoue, T., et al. (2022). "Non-invasive human skin transcriptome analysis using mRNA in skin surface lipids." Commun Biol **5**(1): 215.
4. Farage, M. A., et al. (2008). "Intrinsic and extrinsic factors in skin ageing: a review." Int J Cosmet Sci **30**(2): 87-95.
5. Kimball, A. B., et al. (2018). "Age-induced and photoinduced changes in gene expression profiles in facial skin of Caucasian females across 6 decades of age." Journal of the American Academy of Dermatology **78**(1): 29-39.e27.