Am J Transl Res 2026;18(1):814-825
www.ajtr.org /ISSN:1943-8141/AJTR0170692

Original Article

Transcriptome-wide analysis of genes associated with collagen III in human skin

Mengyun Chen[1,2,3,4*], Gang Chen[1,2,3,4*], Huayang Tang[1,2,3,4], Jindian Zha[5], Liming Hou[6], Jun Wang[6], Jun Zhao[7,8,9], Xiaodong Zheng[1,2,3,4,9]

[1]Department of Dermatology, The First Affiliated Hospital, Anhui Medical University, Hefei, Anhui, P. R. China; [2]Institute of Dermatology, Anhui Medical University, Hefei, Anhui, P. R. China; [3]Key Laboratory of Dermatology (Anhui Medical University), Ministry of Education, Hefei, Anhui, P. R. China; [4]Collaborative Innovation Center of Complex and Severe Skin Disease, Anhui Medical University, Hefei, Anhui, P. R. China; [5]School of Health Management, Anhui Medical University, Hefei, Anhui, P. R. China; [6]Xiamen 943 Cosmetics Co., LTD., Xiamen, Fujian, P. R. China; [7]Department of Microbiology, Anhui Medical University, Hefei, Anhui, P. R. China; [8]Wuhu Interferon Bio-products Industry Research Institute Co., Ltd. Wuhu, Anhui, P. R. China; [9]Anhui Province Key Laboratory of Synthetic Biological Protein, Hefei, Anhui, P. R. China. [*]Equal contributors.

Received December 11, 2025; Accepted January 19, 2026; Epub January 25, 2026; Published January 30, 2026

Abstract: Objectives: To gain a deeper understanding of the association between several widely used skin aging-related proteins and all other genes within the genome. Methods: Skin transcriptome sequencing data of 142 healthy controls across three different populations were mapped to the NCBI build 37 reference genome. Quantified gene abundances were obtained, and missing data were imputed. Principal component analysis was performed to identify sample stratification. Finally, linear regression analyses were conducted to investigate the relationships between the expression levels of these widely used proteins and all other genes. Results: Following realignment and abundance calculation, a total of 63,677 genes were obtained, and 20,476 genes remained after the quality control. Principal component analysis (PCA) results indicated sample stratification among different populations. Upon performing linear regression analyses across the separate datasets, consistent results were observed. The analysis revealed five protein-coding genes significantly associated with collagen III (P value < 2.44E-6, $R^2 > 0.5$). Two protein-coding genes, including collagen III, were positively correlated with fibronectin, and five protein-coding genes were positively correlated with elastin. Enrichment analysis demonstrated that extracellular matrix (ECM)-receptor interaction was the most significant pathway associated with genes linked to collagen III and fibronectin ($P_{\_adjust}$ = 1.29E-10). Conclusions: These findings highlight the crucial role of ECM-receptor interaction in the context of collagen III- and fibronectin-related genes, providing valuable insights into the underlying biological mechanisms.

Keywords: Aging, collagen III, fibronectin, elastin, extracellular matrix proteins

Introduction

Aging, an enduring and universal aspect of human existence, has long captivated scientific inquiry due to its profound impact on physiological function and overall health. As a complex and multifaceted process, aging involves intricate molecular and cellular mechanisms that contribute to progressive declines in tissue and organ function. Recently, comprehensive reviews have illuminated key pathways underlying this phenomenon, notably the interplay between mitochondrial dysfunction, inflammaging, and sarcopenia [1]. Similarly, integrative analyses of transcriptomic and metabolomic data have revealed the critical role of the hepatokine FGF21 in liver aging, offering novel insights into systemic metabolic changes during the aging process [2]. These studies collectively emphasize the intricate biological networks that drive aging, setting the stage for a deeper exploration of targeted interventions to mitigate its effects.

The skin is a crucial organ that covers the surface of the human body and is directly exposed to the external environment. As we age, the skin undergoes significant structural and functional

https://doi.org/10.62347/ERKN7344

changes. Many factors contribute to skin aging, both intrinsic and extrinsic, with several remaining unknown. One such extrinsic factor is ultraviolet radiation [3], which can cause premature skin aging. Pollution is another extrinsic factor that can accelerate the aging process. Intrinsic factors include reactive oxygen species [4], clonal hematopoiesis in the elderly [5], loss of epigenetic information [6], and alterations in gene expression [7]. Together, these factors contribute to skin aging.

Skin cells exhibit complex alterations during aging and are frequently accompanied by changes in gene expression related to extracellular matrix (ECM) proteins [8-10]. Collagen, one of the most abundant extracellular matrix proteins, has been shown in many studies to exert significant anti-aging effects when supplemented [11-13] and is widely used in products from many well-known cosmetic brands, most commonly in the form of type I and III collagen. Collagen bundles are resistant to pressure deformation due to their rotating structures and play a critical role in maintaining skin elasticity [14]. As a result, collagen is considered an ideal biomedical and cosmetic skin care material [15]. As an essential component of the skin, collagen plays an important role in injection fillers, wound dressings, functional skin care products, and general cosmetic applications because of its good supporting, repairing, moisturizing, and whitening properties.

In addition to collagen, elastin (ELN) and fibronectin (FN1) also play important roles in the anti-aging process. One study reported that cardiac-specific overexpression of fibronectin type III domain-containing 5 (FNDC5) attenuated aging-related cardiac remodeling and dysfunction in mice [16].

Fibronectin is a high-molecular-weight glycoprotein that typically exists as a dimer of approximately 500 kDa. It is characterized by variable molecular conformations and multiple isoforms generated through alternative splicing. Fibronectin is a common ECM protein involved in cell adhesion, diffusion, migration, proliferation, and apoptosis, and preferentially binds cells through interactions with integrins, other fibronectin subunits, collagen, heparin, fibrin, matrix metalloproteinases, and growth factors [17]. Fibronectin fibrils possess unique mechanical properties that enable them to modulate the mechanotransduction signals perceived and transmitted by cells. By binding to other ECM proteins, including collagen, elastin, and proteoglycans, fibronectin promotes ECM maturation and tissue specificity [18].

Elastin (ELN) is a protein essential for the elasticity of the native ECM, accounting for approximately 2-4% of the dry weight of the skin and providing elasticity to various organs [19]. Elastin fibers confer elasticity and resilience to the skin, enabling it to adapt to internal physiological and external environmental pressures. Elastin is also biocompatible and non-immunogenic, making it an ideal material for wound dressings [20]. In our previous study, we designed an elastin-like recombinant polypeptide (ELR) capable of being absorbed through the skin based on the properties of the hexapeptide VGVAPG. Continuous use of this ELR could significantly improve skin elasticity and reduce wrinkles [21].

In this study, we analyzed transcriptome sequencing data from human skin samples from the Chinese Han population [22], as well as data from two publicly available skin transcriptome sequencing studies [23-25]. Our goal was to examine the relationships between several widely used skin aging-related proteins-particularly collagen III, fibronectin, and elastin-and all the other genes present in the genome.

Materials and methods

Data sources

Transcriptome sequencing data from 20 healthy Chinese Han individuals were extracted from our previous study (SRP065758) [22]. Two additional transcriptome sequencing datasets, comprising 84 healthy Detroit Caucasian samples and 38 healthy German Caucasian samples, were obtained from GEO datasets SRP035988 [23] and GSE121212 [24, 25], respectively. The present study was conducted in accordance with the principles of the Declaration of Helsinki.

Realignment and estimation of gene expression levels

All data processing and analyses were conducted in a Linux environment. First, all FASTQ files were realigned to the NCBI build 37 reference genome using STAR [26]. RNA-seq FASTQ files from the Detroit samples were generated using Illumina GAII platform and consisted of single-end reads, whereas files from the Chinese and German samples were generated using the Illumina HiSeq 2500 platform and consisted of paired-end reads. Gene and isoform abundances were quantified from both single-end and paired-end RNA-seq data using the software package RSEM [27]. For paired-end data from the Chinese and German samples, the additional parameter "paired-end" was specified. Gene expression levels were represented as transcripts per million (TPM) [28], and values were obtained for 63,677 genes annotated in the Ensembl database (Homo_ sapiens.GRCh37.75.gtf).

Missing data imputation and principal component analysis

A local Python script was used to remove genes with more than 20% missing data. Subsequently, a k-nearest neighbors (KNN) approach (K = 10, approximately equal to the square root of the total sample number) was used to impute the remaining missing values.

Batch effect correction

To mitigate potential batch effects arising from differences in geographic origin (America, Germany, and China), batch correction was applied to the TPM expression matrix using pyComBat, a Python implementation of the ComBat empirical Bayes method [29]. The raw TPM values were first log2-transformed by adding a pseudocount of 1 (log2(TPM + 1)) to stabilize variance and approximate normality. Samples were assigned to three batches based on their geographic origin: batch 1 (America, n = 84), batch 2 (Germany, n = 38), and batch 3 (China, n = 20). The ComBat algorithm was performed directly on the log-transformed matrix (genes as rows, samples as columns) using default parameters (parametric adjustment enabled, no additional covariates). The resulting batch-corrected matrix was z-score standardized (mean = 0, standard deviation = 1 per gene) prior to downstream analyses. The log2-transformed matrix before correction and the batch-corrected matrix after correction were both saved for reproducibility (Supplementary Files 1 and 2).

Principal component analysis

Principal component analysis (PCA) was performed on the standardized log2-transformed data before and after batch correction to evaluate the impact of batch effects and the effectiveness of the correction. PCA was implemented using scikit-learn, with the first two principal components retained. Samples were colored by geographic origin to visualize clustering patterns.

Correlation and bioinformatics analysis

To identify genes co-expressed with COL3A1 and other aging-related extracellular matrix proteins, and to elucidate the associated biological processes, we implemented the following analytical workflow.

First, pairwise correlations were calculated between the expression of COL3A1 and all other genes (n = 20,475, excluding COL3A1 itself) using the batch-corrected log2-transformed TPM matrix. Both the Pearson correlation coefficient (assessing linear relationships) and Spearman rank correlation coefficient (assessing monotonic relationships) were computed using the pearsonr and spearmanr functions from SciPy.

Linear regression and bioinformatics analysis of aging-related extracellular matrix proteins

Linear regression analysis was performed across all three datasets using the expression levels of multiple aging-associated extracellular matrix proteins (including COL3A1 and others) as dependent variables. Linear regression analysis was performed using the Python function linregress to evaluate associations between each selected protein and all other genes. Manhattan plots and quantile-quantile (Q-Q) plots were generated for each regression analysis using the R package qqman, and genomic inflation factors (λ) were calculated to assess potential inflation of test statistics.

Candidate genes significantly associated with the aging-related proteins were then subjected to Gene Ontology (GO) and pathway enrichment analysis using the R package clusterProfiler. Enrichment analyses were performed with default parameters, and results were adjusted for multiple testing using the Benjamini-

Hochberg false discovery rate (FDR) method (FDR < 0.05).

Result

Quality control

A total of 215,170 isoforms and 63,677 genes (gene IDs) were obtained after realignment and abundance calculation. A TPM matrix containing 56,635 genes (gene names) was then generated using a local Perl script. Among these, 7,042 duplicated genes were identified; for each duplicated entry, the former record was removed and the latter retained. After removing genes with more than 20% missing data across all samples, the remaining missing values were imputed using the KNN method. Ultimately, 20,476 genes remained for subsequent analysis (Supplementary Table 1). Principal component analysis (PCA) of the log2-transformed data (log2(TPM + 1)) before batch correction revealed distinct clustering of samples by geographic origin (America, Germany, and China), with clear separation along the first two principal components (**Figure 1A**). This pattern indicated substantial batch effects attributable to sample collection site or technical differences. Following pyComBat batch correction, the geographic clustering was substantially reduced, and samples from different origins showed improved overlap and more homogeneous distribution in the PCA space (**Figure 1B**).

Correlation analysis with COL3A1

Collagen is a major component of the extracellular matrix, and collagen III is one of the collagen proteins present in the skin, with expression levels known to decrease with age. Pairwise correlations between COL3A1 expression and all other genes (n = 20,475) were computed using the batch-corrected log2-transformed TPM matrix. Strong positive correlations were observed with several extracellular matrix-related genes. The top 11 genes ranked by Pearson correlation coefficient are summarized in **Table 1**. All associations were highly significant ($P < 1 \times 10^{-34}$), with Pearson r values ranging from 0.815 to 0.951. Notably, collagen family members such as COL1A2 (r = 0.951451, $P = 1.61 \times 10^{-73}$), COL5A2 (r = 0.941399, $P = 5.96 \times 10^{-68}$), COL1A1 (r = 0.923049, $P = 5.98 \times 10^{-60}$), and COL5A1 (r = 0.876161, $P = 3.26 \times 10^{-46}$) exhibited the strongest associations. Spearman rank correlations showed consistent trends, though slightly lower in magnitude for some genes (**Table 1**). The complete correlation results are provided in Supplementary Table 2.

Linear regression and bioinformatics analysis of aging-related extracellular matrix proteins

Linear regression analyses were also performed between the expression level of *COL3A1* and those of 20,765 other genes across the three populations (Supplementary Table 3). The expression levels of 12 protein-coding genes and one long noncoding RNA (lincRNA; gene name: *RP11-572C15.5*, gene ID: ENST00000610103) were found to be significantly associated with *COL3A1* (**Table 2**). All of these associations were positive (*P* value < 1E-4, slope > 0, $R^2 > 0.5$). After applying Bonferroni's correction (*P* value < 2.44E-6), five protein-coding genes, including *FN1*, remained significantly associated with *COL3A1*. Manhattan plots were generated for each population (Supplementary Figure 1), and moderate inflation was observed in the Q-Q plots (Supplementary Figure 2) with genomic inflation factors of λ_China = 1.8485, λ_America = 3.764, and λ_Germany = 4.5721 for the Chinese, American, and German populations, respectively. Notably, collagen I and collagen V genes showed extremely significant positive correlation with *COL3A1* across all three databases ($R^2 > 0.6$). Among these genes, the expression levels (TPM) of collagen I and *SPARC* were very similar to those of *COL3A1* (**Figure 2**; **Table 3**).

Linear regression analyses were also performed to assess associations between fibronectin and elastin expression levels and those of other genes (Supplementary Tables 4 and 5). The results showed that six protein-coding genes, excluding *COL3A1*, were positively correlated with fibronectin, and notably, five of these six genes were collagen genes (**Table 4**). After Bonferroni's correction, only *COL3A1* and *COL5A2* remained significantly associated with fibronectin across all three populations. In addition, 15 protein-coding genes were identified as being positively correlated with elastin. After applying Bonferroni's correction, five protein-coding genes remained significantly associated with elastin. Among these, one gene, FBN1 (fibrillin 1), showed strong and consistent associations across all populations ($P_{\_China}$ =

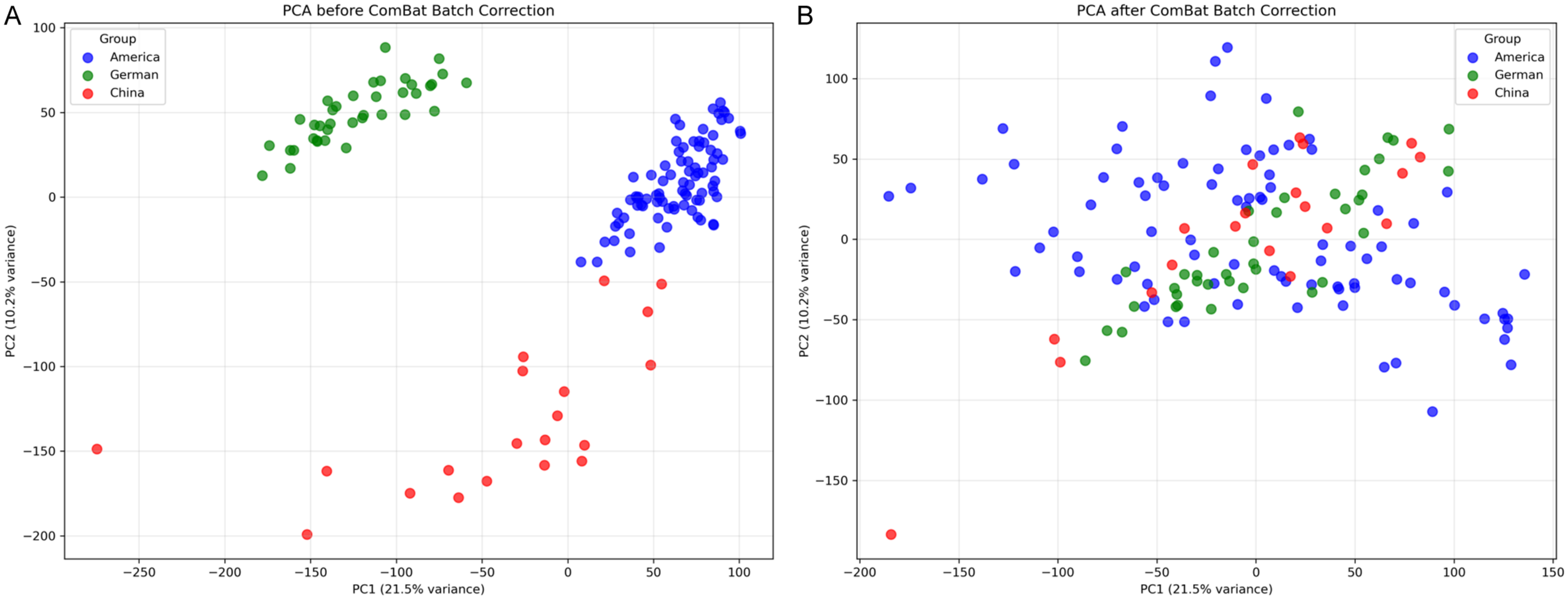


Figure 1. Principal component analysis (PCA) of the log2-transformed data before batch correction (A) and principal component analysis (PCA) after pyComBat batch correction (B).

Table 1. Top 11 genes most positively correlated with COL3A1 in the batch-corrected TPM matrix

Gene	Pearson_r	Pearson_p	Pearson_se	Spearman_r	Spearman_p
COL1A2	0.951451	1.61E-73	0.026014	0.948888	5.40E-72
COL5A2	0.941399	5.96E-68	0.028507	0.94542	4.74E-70
COL1A1	0.923049	5.98E-60	0.032512	0.915786	2.55E-57
SPARC	0.919273	1.50E-58	0.033267	0.915669	2.80E-57
COL5A1	0.876161	3.26E-46	0.040738	0.861143	5.70E-43
RP11-572C15.6	0.856194	5.53E-42	0.043665	0.853435	1.89E-41
CCDC80	0.855301	8.25E-42	0.04379	0.848199	1.82E-40
PPIC	0.824793	1.75E-36	0.047788	0.812495	1.28E-34
ITIH5	0.819891	1.01E-35	0.048387	0.786927	3.76E-31
KDELR3	0.816389	3.41E-35	0.048808	0.815192	5.13E-35
C1QTNF3	0.814777	5.91E-35	0.049	0.801074	5.23E-33

Table 2. Linear regression tests for genes related to collagen III in three different populations

Gene	Chinese			America			Germany		
	slope	r^2	*p*_value	slope	r^2	*p*_value	slope	r^2	*p*_value
ADAMTS2	68.3375	0.5948	6.86E-05	110.4874	0.7676	1.04E-27	88.7610	0.8518	1.69E-16
C1QTNF3	47.1087	0.5815	9.27E-05	84.5749	0.6974	5.44E-23	33.1258	0.6671	4.02E-10
COL1A1	0.8288	0.6029	5.69E-05	0.5389	0.8162	6.67E-32	1.0625	0.9327	1.08E-22
COL1A2	0.8415	0.7632	4.89E-07	0.7426	0.9280	1.31E-48	1.1084	0.9476	1.22E-24
COL5A1	21.3674	0.6366	2.50E-05	19.5673	0.7545	9.87E-27	22.6287	0.9097	2.20E-20
COL5A2	35.5948	0.7754	3.03E-07	65.9976	0.9341	3.37E-50	31.2108	0.9126	1.23E-20
CTHRC1	10.3234	0.7046	3.71E-06	27.4461	0.5568	3.79E-16	17.7882	0.6467	1.19E-09
FN1	4.6950	0.7303	1.61E-06	8.8374	0.5685	1.25E-16	5.9737	0.6995	6.24E-11
KDELR3	73.2338	0.6736	9.27E-06	119.0933	0.6930	9.91E-23	60.3685	0.6631	4.99E-10
RP11-572C15.6	22.0445	0.5804	9.49E-05	29.4488	0.6925	1.06E-22	13.5544	0.8330	1.48E-15
SERPINH1	8.5457	0.6426	2.15E-05	21.0313	0.5165	1.39E-14	18.1860	0.6248	3.58E-09
SFRP2	7.2209	0.7697	3.80E-07	8.2718	0.6411	6.18E-20	5.2468	0.5957	1.40E-08
SPARC	0.8000	0.7528	7.25E-07	0.6535	0.8297	2.93E-33	1.1511	0.8970	2.37E-19

1.2E-6, $R2_{\_China}$ = 0.74; $P_{\_America}$ = 1.52E-14, $R2_{\_America}$ = 0.68; $P_{\_Germany}$ = 4.87E-14, $R2_{\_Germany}$ = 0.08) (**Table 5**).

ECM-receptor interaction is the major pathway associated with COL3A1 and FN1

Because the expression levels of *COL3A1* and *FN1* were significantly associated (**Tables 2** and **3**), Gene Ontology (GO) enrichment and pathway analyses were performed using genes associated with these two genes (*P* value < 1E-4), after removal of the lincRNA gene. As expected, ECM-receptor interaction was identified as the most significant pathway associated with these genes ($p_{\_adjust}$ = 1.29E-10). Three additional pathways reached genome-wide significance: protein digestion and absorption ($p_{\_adjust}$ = 6.18E-9), focal adhesion ($p_{\_adjust}$ = 1.6E-8), and amoebiasis ($p_{\_adjust}$ = 1.6E-8) (**Figure 3A**; Supplementary Table 6). Thirteen Gene Ontology terms met the significance criteria ($p_{\_adjust}$ < 5E-8, gene count > 3), of which five were directly related to the extracellular matrix (**Figure 3B**; Supplementary Table 7).

KEGG pathway and GO enrichment analyses were also conducted for elastin-related genes. No pathways reached genome-wide significance; in association with these genes, only moderately significant pathways ($p_{\_adjust}$ < 0.05, gene count > 2) were identified, including protein digestion and absorption (p_adjust = 0.0015), ECM-receptor interaction ($p_{\_adjust}$ = 0.0015), and focal adhesion ($p_{\_adjust}$ = 0.0127) (Supplementary Table 8). Similar results were obtained for the GO enrichment analysis, 21 moderately significant GO terms with gene

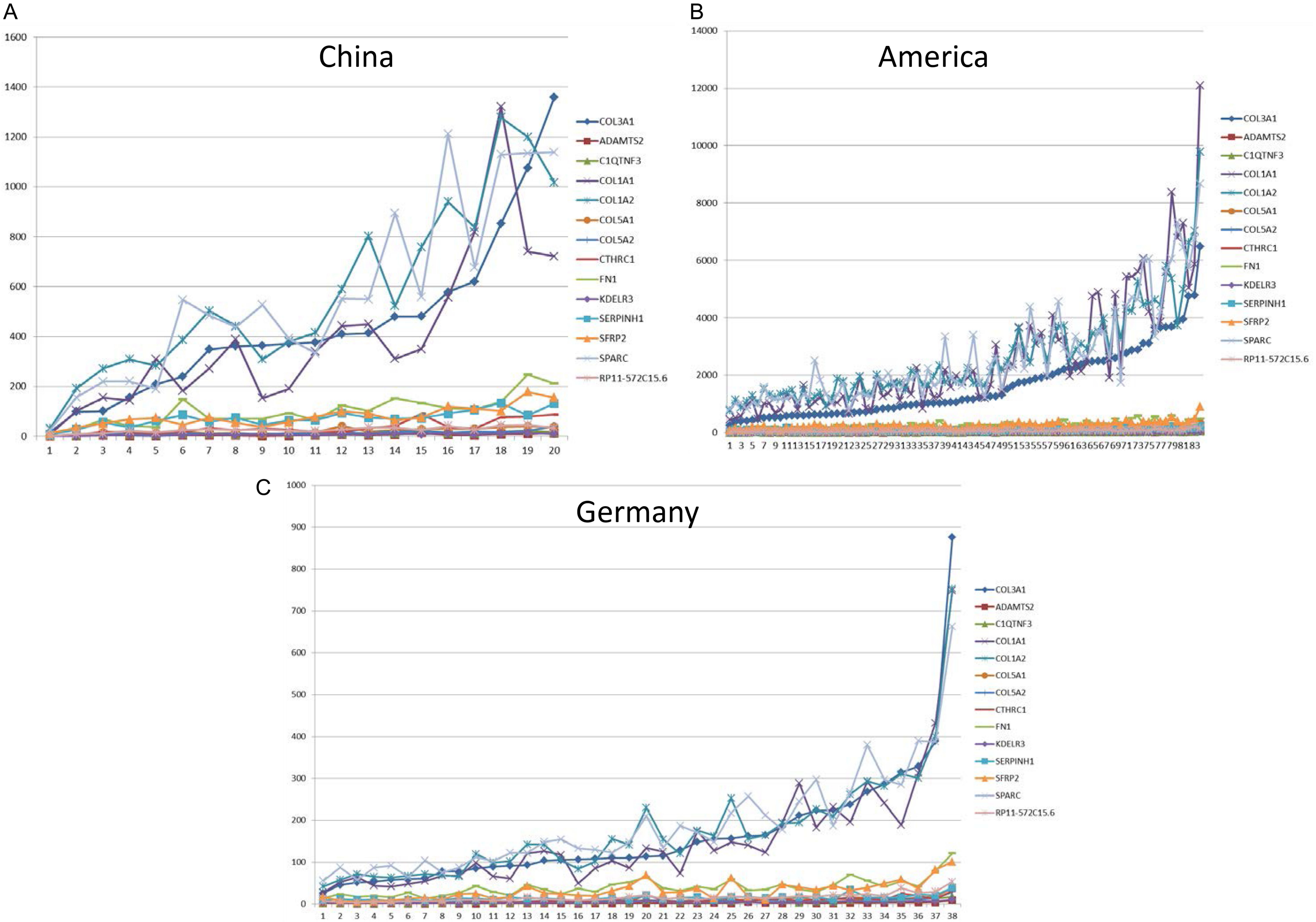


Figure 2. Expression levels of 12 protein-coding genes and one long noncoding RNA that are strongly associated with collagen III across three different populations.

Table 3. Expression levels (represented as TPM) of 14 significantly associated genes in three different populations

Gene	Chinese		America		Germany	
	Mean	SD	Mean	SD	Mean	SD
COL3A1	445.5505	331.5389	1640.805	1228.273	163.1695	147.2035
ADAMTS2	5.9005	3.741607	17.84321	9.739771	2.158684	1.530641
C1QTNF3	11.508	5.366633	19.36571	12.1281	6.527895	3.629546
COL1A1	397.8595	310.6134	2547.631	2058.96	150.9292	133.7992
COL1A2	573.5195	344.207	2616.807	1593.246	173.7679	129.2751
COL5A1	20.4485	12.37967	96.19762	54.52518	9.468684	6.204475
COL5A2	13.618	8.201584	32.40667	17.98736	7.981053	4.505575
CTHRC1	34.823	26.95808	63.31643	33.39297	8.948684	6.654646
FN1	101.317	60.34712	283.069	104.796	39.49974	20.60953
KDELR3	6.822	3.715675	16.68083	8.5854	4.004737	1.9857
RP11-572C15.6	26.095	11.45757	65.85583	34.70752	15.16289	9.911675
SERPINH1	71.7625	31.0992	110.253	41.97254	14.83026	6.397889
SFRP2	79.2545	40.28158	261.1256	118.8946	31.58289	21.65358
SPARC	568.7905	359.56	2647.028	1712.028	187.2745	121.1159

Table 4. Linear regression tests for genes related to fibronectin in three different populations

Gene	Chinese			America			Germany		
	slope	r^2	*p*_value	slope	r^2	*p*_value	slope	r^2	*p*_value
COL1A2	0.139226	0.630615	2.91E-05	0.051785	0.619855	6.64E-19	0.144798	0.824936	3.45E-15
COL3A1	0.155554	0.730327	1.61E-06	0.064332	0.568522	1.25E-16	0.117096	0.699492	6.24E-11
COL5A1	4.109978	0.71086	3.05E-06	1.368256	0.506803	3.16E-14	2.875141	0.74919	2.33E-12
COL5A2	6.378078	0.751384	7.64E-07	4.607321	0.625377	3.63E-19	4.048179	0.78322	1.66E-13
COL6A3	1.856516	0.595645	6.73E-05	1.765701	0.637982	8.84E-20	1.375148	0.837583	8.88E-16
NID1	13.9956	0.682342	7.23E-06	12.68837	0.504161	3.95E-14	17.13189	0.810853	1.4E-14
SH2B3	24.70566	0.668036	1.09E-05	48.63185	0.509385	2.55E-14	28.63564	0.581275	2.66E-08

Table 5. Linear regression tests for genes related to elastin in three different populations

Gene	Chinese			America			Germany		
	slope	r^2	*p*_value	slope	r^2	*p*_value	slope	r^2	*p*_value
ADAMTS2	3.875105	0.715865	2.60E-06	3.710345	0.669513	2.06E-21	3.048917	0.808356	1.78E-14
C1QTNF3	2.741639	0.737178	1.27E-06	2.578451	0.501344	4.99E-14	1.103955	0.595899	1.39E-08
CD248	0.496073	0.651028	1.72E-05	0.609921	0.732286	3.5E-25	1.197961	0.550888	9.63E-08
COL5A1	1.093839	0.624414	3.39E-05	0.724803	0.800695	1.87E-30	0.786154	0.88306	2.34E-18
COL6A1	0.125534	0.606401	5.24E-05	0.206539	0.599153	5.94E-18	0.26113	0.77815	2.52E-13
COL6A2	0.10557	0.795619	1.28E-07	0.143913	0.596135	8.1E-18	0.103899	0.768565	5.42E-13
COL8A2	6.030715	0.581658	9.23E-05	13.02543	0.590645	1.41E-17	4.361812	0.75562	1.46E-12
FBN1	0.949153	0.738738	1.20E-06	0.682836	0.515417	1.52E-14	0.233343	0.797376	4.87E-14
FSTL1	0.51416	0.734822	1.38E-06	0.283146	0.59015	1.49E-17	0.130718	0.812881	1.15E-14
GALNT16	3.908387	0.701584	4.07E-06	3.084949	0.539936	1.77E-15	1.782757	0.61851	4.84E-09
HSD3B7	3.206404	0.839646	1.41E-08	4.676286	0.536396	2.44E-15	2.088651	0.519531	3.33E-07
ITIH5	1.98818	0.691006	5.61E-06	1.900723	0.569429	1.14E-16	1.277689	0.750796	2.08E-12
LRP1	0.691353	0.715069	2.66E-06	0.77777	0.525812	6.2E-15	0.445566	0.601433	1.08E-08
TCF7L1	2.804929	0.586954	8.20E-05	6.723257	0.550184	6.98E-16	12.30984	0.668577	3.71E-10
THY1	0.70467	0.682596	7.18E-06	0.780422	0.618496	7.7E-19	1.970119	0.661703	5.4E-10

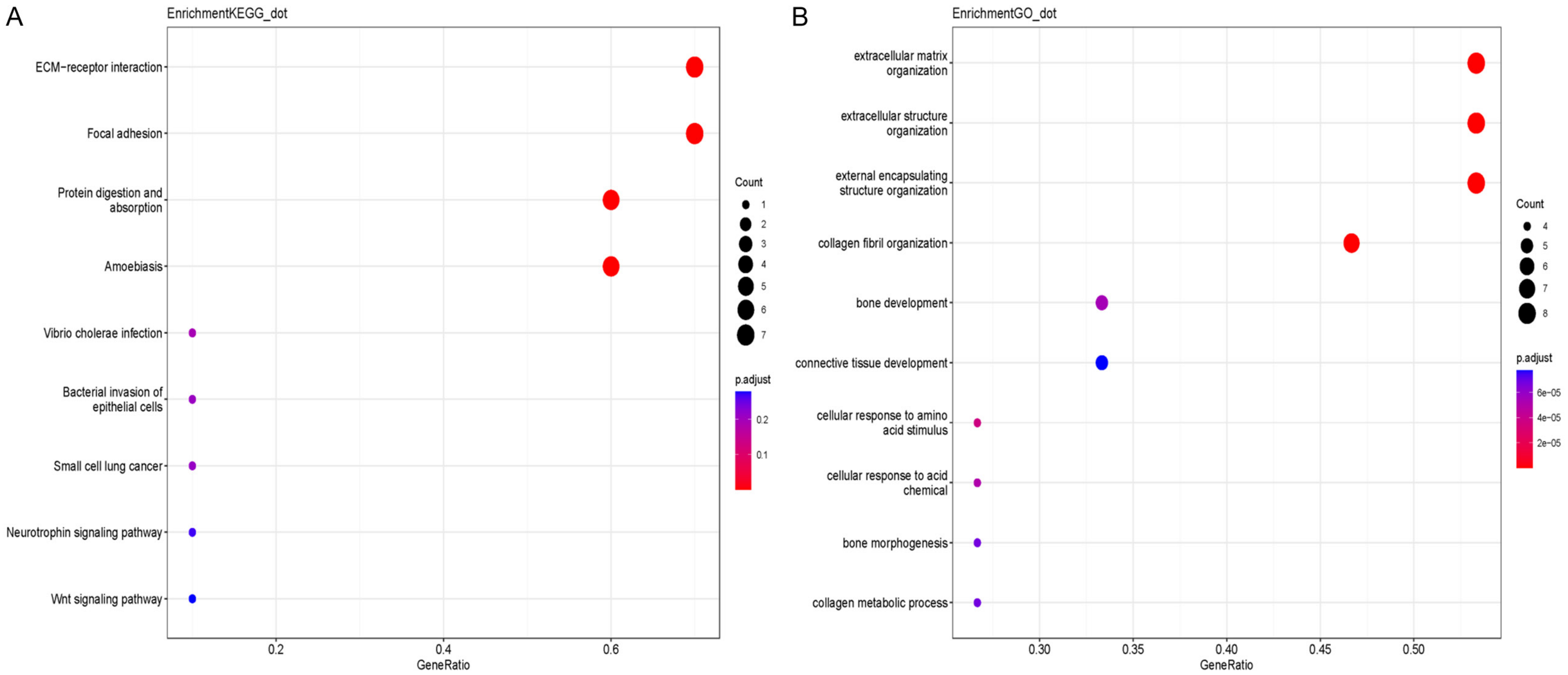


Figure 3. KEGG pathway enrichment and Gene Ontology (GO) enrichment analysis for genes associated with collagen III and fibronectin.

counts greater than two were observed, including extracellular matrix structural constituent conferring tensile strength ($p_{\_adjust}$ = 1.06E-6), extracellular matrix structural constituent ($p_{\_adjust}$ = 3.2E-6), collagen-containing extracellular matrix ($p_{\_adjust}$ = 4.31E-6), et al. (Supplementary Table 9).

Discussion

Protein-protein interactions can be detected using various experimental methods, such as GST pull-down assays, co-immunoprecipitation, and yeast two-hybrid systems [30]. However, these approaches may fail to detect transient protein-protein interactions with low affinity, indirect interactions that require intermediary proteins, or interactions affected by experimental constraints. In the present study, we directly analyzed gene expression levels in human skin tissues and assessed correlations between proteins using linear regression analysis. This strategy enables the identification of potential protein-protein interactions, including indirect associations, based on shared expression patterns of the corresponding genes.

Our results indicate that five proteins are significantly associated with type III collagen and two proteins are significantly associated with fibronectin, including collagen I, collagen III, collagen V, fibronectin, *SFRP2*, and *SPARC*. The first four proteins are all extracellular matrix proteins, while *SFRP2* is involved in ECM organization (**Figure 3B**). The *SPARC* gene encodes a cysteine-rich acidic matrix-associated protein that has been implicated in numerous cellular processes, including cell-ECM interactions and ECM assembly [31]. The extracellular matrix is a dynamic, three-dimensional network of macromolecules that provides structural support to cells and tissues and is composed primarily of collagen, proteoglycans, elastin, fibronectin, laminin, and other glycoproteins [32, 33]. ECM proteins typically contain multiple independently folded domains with highly conserved sequences and arrangements. Some of these domains bind to adhesion receptors, such as integrins, which mediate cell-matrix adhesion and signal transduction. In addition, ECM proteins bind soluble growth factors and regulate their distribution, activation, and presentation to cells, thereby integrating complex multivalent signals in a spatially organized and regulated manner [34]. The extracellular matrix, which accounts for more than 70% of the skin, serves as the central hub for skin repair and regeneration, and its synthesis and function are crucial for wound healing and dermal regeneration [35].

Previous studies have shown that reducing matrix metalloproteinase-2 (*MMP-2*) levels in older individuals can soften carotid arteries by slowing elastin degradation and increasing the availability of molecules that promote vascular relaxation, ultimately improving cardiovascular health [36]. In the present study, we identified five proteins that were significantly and positively correlated with elastin, including *C1QTNF3*, *COL6A2*, *FBN1*, *FSTL1*, and *HSD3B7*. Among these genes, *FBN1* encodes a preproprotein that is proteolytically processed to generate two proteins: the extracellular matrix component fibrillin-1 and the hormone asprosin. Fibrillin microfibrils provide structural support to connective tissues throughout the body. In tissues such as the lungs, blood vessels, and skin, these microfibrils form a scaffold that anchors elastin fibers, thereby conferring strength and elasticity [37].

In summary, we conducted linear regression analyses of skin transcriptome data and identified genes whose expression levels are strongly correlated with widely used aging-related proteins. It is important to note, however, that this analytical approach does not provide direct evidence of physical protein-protein interactions. Therefore, further experimental validation will be required to confirm these associations and elucidate their functional significance.

Acknowledgements

We thank all participants from three different countries. We also thank the previous researchers for free access to these valuable genetic data sets. This work was supported by scientific research project of colleges and universities in Anhui Province (2022AH051155, 2022AH050761), Promotion Program of Basic and Clinical Cooperative Research of Anhui Medical University (2023xkjT047, 2023xkjT-035), Major Scientific Research Projects of Universities in Anhui Province (No. 2024AH-040109), Doctoral Research Fund of Anhui Medical University (XJ202012). Anhui Province Key Research and Development Program (2023s07020021), Wuhu City Industrialization

Project for Patented Inventions and Technological Achievements (202419).

Disclosure of conflict of interest

None.

Abbreviations

ECM, extracellular matrix; FNDC5, fibronectin type III domain-containing 5; ELR, elastin-like recombinant polypeptide; TPM, transcripts per million; KNN, k-nearest neighbors; PCA, principal component analysis; lncRNA, long noncoding RNA.

Address correspondence to: Dr. Xiaodong Zheng, Department of Dermatology, The First Affiliated Hospital of Anhui Medical University, Hefei 230032, Anhui, P. R. China. Tel: +86-551-65138576; E-mail: zhengxiaodong@ahmu.edu.cn; Dr. Jun Zhao, Department of Microbiology, Anhui Medical University, Hefei 230032, Anhui, P. R. China. Tel: +86-551-65119667; ORICD: 0000-0003-3012-4998; Fax: +86-551-65119667; E-mail: junzhaomedical@163.com

References

[1] Xu X and Wen Z. The mediating role of inflammaging between mitochondrial dysfunction and sarcopenia in aging: a review. Am J Clin Exp Immunol 2023; 12: 109-126.

[2] Wang W, Qian J, Shang M, Qiao Y, Huang J, Gao X, Ye Z, Tong X, Xu K, Li X, Liu Z, Zhou L and Zheng S. Integrative analysis of the transcriptome and metabolome reveals the importance of hepatokine FGF21 in liver aging. Genes Dis 2023; 11: 101161.

[3] Salminen A, Kaarniranta K and Kauppinen A. Photoaging: UV radiation-induced inflammation and immunosuppression accelerate the aging process in the skin. Inflamm Res 2022; 71: 817-831.

[4] Davalli P, Mitic T, Caporali A, Lauriola A and D'Arca D. ROS, cell senescence, and novel molecular mechanisms in aging and age-related diseases. Oxid Med Cell Longev 2016; 2016: 3565127.

[5] Mitchell E, Spencer Chapman M, Williams N, Dawson KJ, Mende N, Calderbank EF, Jung H, Mitchell T, Coorens THH, Spencer DH, Machado H, Lee-Six H, Davies M, Hayler D, Fabre MA, Mahbubani K, Abascal F, Cagan A, Vassiliou GS, Baxter J, Martincorena I, Stratton MR, Kent DG, Chatterjee K, Parsy KS, Green AR, Nangalia J, Laurenti E and Campbell PJ. Clonal dynamics of haematopoiesis across the human lifespan. Nature 2022; 606: 343-350.

[6] Yang JH, Hayano M, Griffin PT, Amorim JA, Bonkowski MS, Apostolides JK, Salfati EL, Blanchette M, Munding EM, Bhakta M, Chew YC, Guo W, Yang X, Maybury-Lewis S, Tian X, Ross JM, Coppotelli G, Meer MV, Rogers-Hammond R, Vera DL, Lu YR, Pippin JW, Creswell ML, Dou Z, Xu C, Mitchell SJ, Das A, O'Connell BL, Thakur S, Kane AE, Su Q, Mohri Y, Nishimura EK, Schaevitz L, Garg N, Balta AM, Rego MA, Gregory-Ksander M, Jakobs TC, Zhong L, Wakimoto H, El Andari J, Grimm D, Mostoslavsky R, Wagers AJ, Tsubota K, Bonasera SJ, Palmeira CM, Seidman JG, Seidman CE, Wolf NS, Kreiling JA, Sedivy JM, Murphy GF, Green RE, Garcia BA, Berger SL, Oberdoerffer P, Shankland SJ, Gladyshev VN, Ksander BR, Pfenning AR, Rajman LA and Sinclair DA. Loss of epigenetic information as a cause of mammalian aging. Cell 2023; 186: 305-326, e27.

[7] Lago JC and Puzzi MB. The effect of aging in primary human dermal fibroblasts. PLoS One 2019; 14: e0219165.

[8] Shoko T, Maharaj VJ, Naidoo D, Tselanyane M, Nthambeleni R, Khorombi E and Apostolides Z. Anti-aging potential of extracts from sclerocarya birrea (A. Rich.) Hochst and its chemical profiling by UPLC-Q-TOF-MS. BMC Complement Altern Med 2018; 18: 54.

[9] Na J, Bak DH, Im SI, Choi H, Hwang JH, Kong SY, No YA, Lee Y and Kim BJ. Anti-apoptotic effects of glycosaminoglycans via inhibition of ERK/AP-1 signaling in TNF-alpha-stimulated human dermal fibroblasts. Int J Mol Med 2018; 41: 3090-3098.

[10] Li L, Hwang E, Ngo HTT, Lin P, Gao W, Liu Y and Yi TH. Antiphotoaging effect of prunus yeonesis blossom extract via inhibition of MAPK/AP-1 and regulation of the TGF-betaI/Smad and Nrf2/ARE signaling pathways. Photochem Photobiol 2018; 94: 725-732.

[11] Pu SY, Huang YL, Pu CM, Kang YN, Hoang KD, Chen KH and Chen C. Effects of oral collagen for skin anti-aging: a systematic review and meta-analysis. Nutrients 2023; 15: 2080.

[12] Figueres Juher T and Bases Perez E. An overview of the beneficial effects of hydrolysed collagen intake on joint and bone health and on skin ageing. Nutr Hosp 2015; 32 Suppl 1: 62-66.

[13] Kim H, Choi N, Kim DY, Kim SY, Song SY and Sung JH. TGF-beta2 and collagen play pivotal roles in the spheroid formation and anti-aging of human dermal papilla cells. Aging (Albany NY) 2021; 13: 19978-19995.

[14] Peng W, Li D, Dai K, Wang Y, Song P, Li H, Tang P, Zhang Z, Li Z, Zhou Y and Zhou C. Recent progress of collagen, chitosan, alginate and other hydrogels in skin repair and wound dressing applications. Int J Biol Macromol 2022; 208: 400-408.

[15] Ye T, Xiang Q, Yang Y and Huang Y. Research, development and application of collagen: a review. Sheng Wu Gong Cheng Xue Bao 2023; 39: 942-960.

[16] Hu C, Zhang X, Hu M, Teng T, Yuan YP, Song P, Kong CY, Xu SC, Ma ZG and Tang QZ. Fibronectin type III domain-containing 5 improves aging-related cardiac dysfunction in mice. Aging Cell 2022; 21: e13556.

[17] Patten J and Wang K. Fibronectin in development and wound healing. Adv Drug Deliv Rev 2021; 170: 353-368.

[18] Dalton CJ and Lemmon CA. Fibronectin: molecular structure, fibrillar structure and mechanochemical signaling. Cells 2021; 10: 2443.

[19] Aamodt JM and Grainger DW. Extracellular matrix-based biomaterial scaffolds and the host response. Biomaterials 2016; 86: 68-82.

[20] Wen Q, Mithieux SM and Weiss AS. Elastin biomaterials in dermal repair. Trends Biotechnol 2020; 38: 280-291.

[21] Wu K, Liu Z, Wang W, Zhou F, Cheng Q, Bian Y, Su W, Liu B, Zha J, Zhao J and Zheng X. An artificially designed elastin-like recombinant polypeptide improves aging skin. Am J Transl Res 2022; 14: 8562-8571.

[22] Dou J, Zhang L, Xie X, Ye L, Yang C, Wen L, Shen C, Zhu C, Zhao S, Zhu Z, Liang B, Wang Z, Li H, Fan X, Liu S, Yin X, Zheng X, Sun L, Yang S, Cui Y, Zhou F and Zhang X. Integrative analyses reveal biological pathways and key genes in psoriasis. Br J Dermatol 2017; 177: 1349-1357.

[23] Li B, Tsoi LC, Swindell WR, Gudjonsson JE, Tejasvi T, Johnston A, Ding J, Stuart PE, Xing X, Kochkodan JJ, Voorhees JJ, Kang HM, Nair RP, Abecasis GR and Elder JT. Transcriptome analysis of psoriasis in a large case-control sample: RNA-seq provides insights into disease mechanisms. J Invest Dermatol 2014; 134: 1828-1838.

[24] Tsoi LC, Rodriguez E, Degenhardt F, Baurecht H, Wehkamp U, Volks N, Szymczak S, Swindell WR, Sarkar MK, Raja K, Shao S, Patrick M, Gao Y, Uppala R, Perez White BE, Getsios S, Harms PW, Maverakis E, Elder JT, Franke A, Gudjonsson JE and Weidinger S. Atopic dermatitis is an IL-13-dominant disease with greater molecular heterogeneity compared to psoriasis. J Invest Dermatol 2019; 139: 1480-1489.

[25] Tsoi LC, Rodriguez E, Stolzl D, Wehkamp U, Sun J, Gerdes S, Sarkar MK, Hubenthal M, Zeng C, Uppala R, Xing X, Thielking F, Billi AC, Swindell WR, Shefler A, Chen J, Patrick MT, Harms PW, Kahlenberg JM, Perez White BE, Maverakis E, Gudjonsson JE and Weidinger S. Progression of acute-to-chronic atopic dermatitis is associated with quantitative rather than qualitative changes in cytokine responses. J Allergy Clin Immunol 2020; 145: 1406-1415.

[26] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M and Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013; 29: 15-21.

[27] Li B and Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 2011; 12: 323.

[28] Wagner GP, Kin K and Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci 2012; 131: 281-285.

[29] Johnson WE, Li C and Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 2007; 8: 118-127.

[30] Fields S and Song O. A novel genetic system to detect protein-protein interactions. Nature 1989; 340: 245-246.

[31] Arnold SA and Brekken RA. SPARC: a matricellular regulator of tumorigenesis. J Cell Commun Signal 2009; 3: 255-273.

[32] Theocharis AD, Skandalis SS, Gialeli C and Karamanos NK. Extracellular matrix structure. Adv Drug Deliv Rev 2016; 97: 4-27.

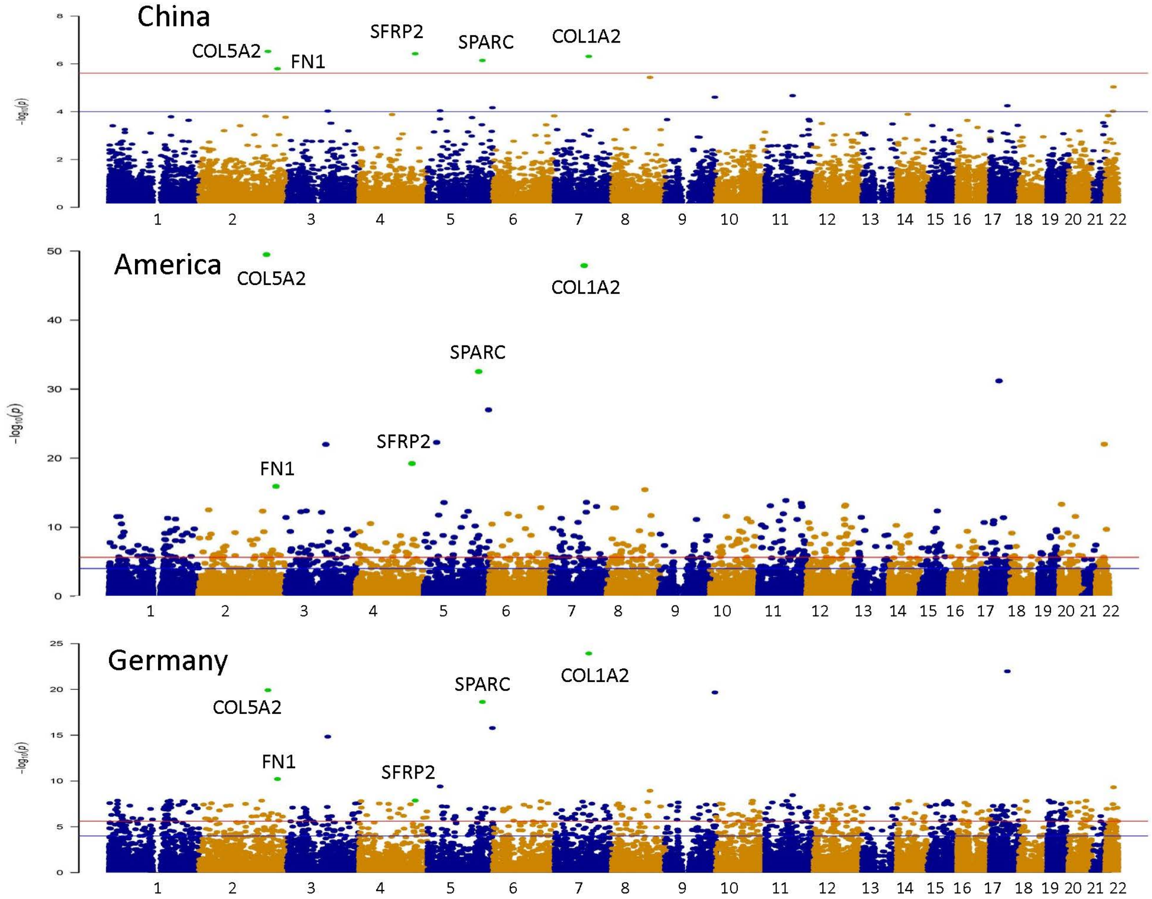
[33] Karamanos NK, Theocharis AD, Piperigkou Z, Manou D, Passi A, Skandalis SS, Vynios DH, Orian-Rousseau V, Ricard-Blum S, Schmelzer CEH, Duca L, Durbeej M, Afratis NA, Troeberg L, Franchi M, Masola V and Onisto M. A guide to the composition and functions of the extracellular matrix. FEBS J 2021; 288: 6850-6912.

[34] Hynes RO. The extracellular matrix: not just pretty fibrils. Science 2009; 326: 1216-1219.

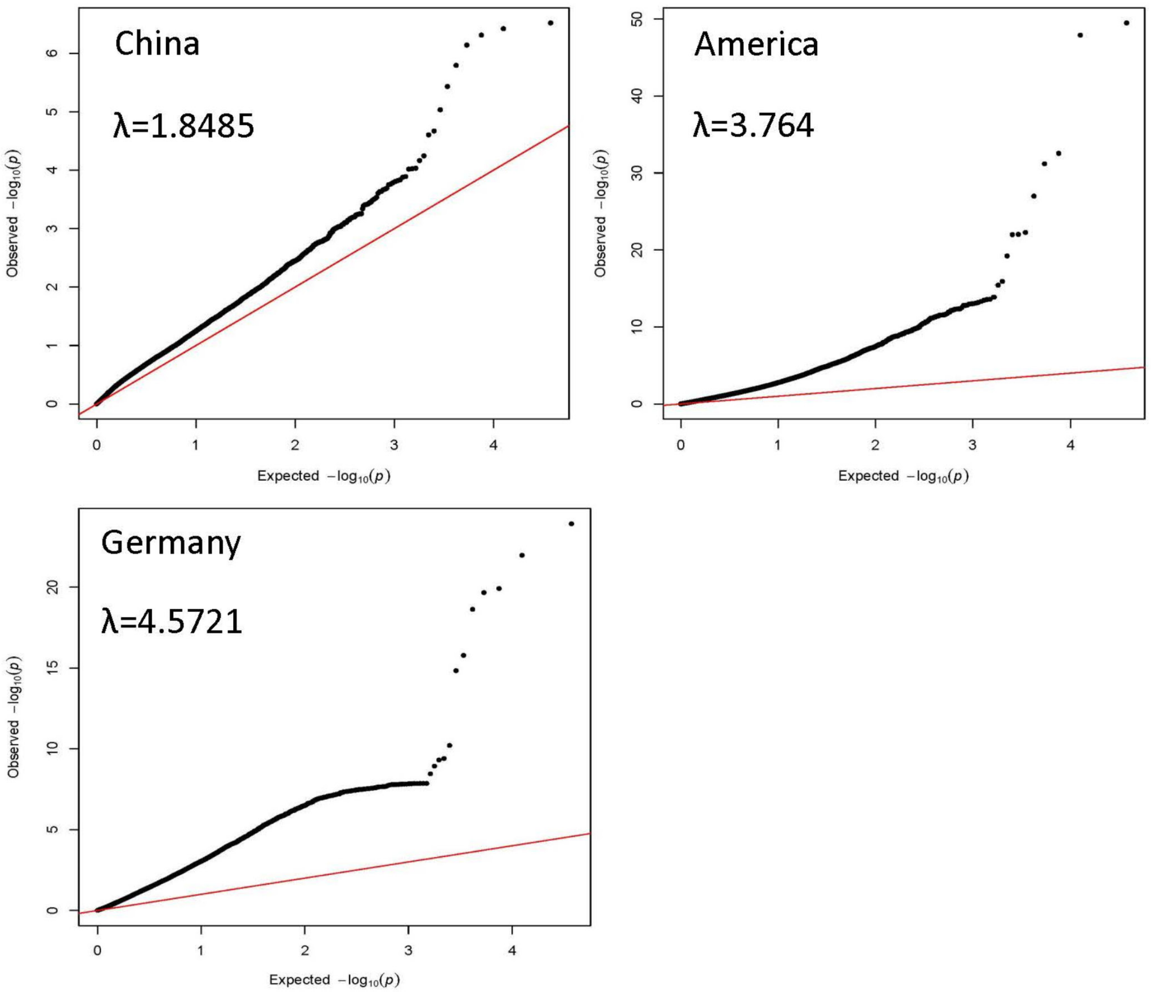
[35] Widgerow AD, Fabi SG, Palestine RF, Rivkin A, Ortiz A, Bucay VW, Chiu A, Naga L, Emer J and Chasan PE. Extracellular matrix modulation: optimizing skin care and rejuvenation procedures. J Drugs Dermatol 2016; 15 Suppl: s63-71.

[36] Diaz-Canestro C, Puspitasari YM, Liberale L, Guzik TJ, Flammer AJ, Bonetti NR, Wust P, Costantino S, Paneni F, Akhmedov A, Varga Z, Ministrini S, Beer JH, Ruschitzka F, Hermann M, Luscher TF, Sudano I and Camici GG. MMP-2 knockdown blunts age-dependent carotid stiffness by decreasing elastin degradation and augmenting eNOS activation. Cardiovasc Res 2022; 118: 2385-2396.

[37] Jensen SA and Handford PA. New insights into the structure, assembly and biological roles of 10-12 nm connective tissue microfibrils from fibrillin-1 studies. Biochem J 2016; 473: 827-838.



Supplementary Figure 1. Manhattan plots demonstrating the correlation results between other gene expression levels and collagen III in three different populations.



Supplementary Figure 2. Quantile-quantile plots of the distribution of the correlation results in three different populations.

Supplementary File 1. log2_tpm_plus1_before_correction.

Supplementary File 2. pm_corrected_after_combat.