nature portfolio

Corresponding author(s): Shizhao Ji, Zhaofan Xia, Runzhi Huang, Chenguang Bai

Last updated by author(s): Aug 6, 2025

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our Editorial Policies 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	Data used in the study were downloaded from public database, including GSE163973; SRP137071;GSE188952, GSE7890, and GSE44270.
Data analysis	Seurat R package (Version: 3.2.2) and SCANPY module (Version: 1.7.1) in Python (version 3.6) ; iTALK R package (Version: 0.1.0); the Molecular Signatures Database (MSigDB) (Version 7.1) ; Gene Set Variation Analysis (GSVA) (Version: 1.38.0); the Monocle R package (Version: 2.18.0); "ConsensusClusterPlus" R software; the CIBERSORT database; Single-Sample Gene Set Enrichment Analysis (ssGSEA); the igraph package in R; the Cmap database; R version 4.0.3 software (Institute for Statistics and Mathematics, Vienna, Austria; www.r-project.org), Strawberry Perl version 5.30.0.1 software (https://www.perl.org/), and GraphPad Prism 9 software

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Provide your data availability statement here.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	We used sex in the study, as suggested by the reviewers. Our data were download from public database, and the sequencing data included both male and female. In verification of our keloid classification with protein sequencing and immunohistochemistry, we also collected patient information (n =26 in protein sequencing; n=19 in immunohistochemistry), including sex. The purpose of protein sequencing and immunohistochemistry was to verify the expression of key genes, so no analysis between genders was performed.
Population characteristics	We included patients in three groups: normal skin, hypertrophic scar, and keloid. The included patients were of various ages.
Recruitment	The patients were randomly selected without bias. When obtaining informed consent from patients, patients are recruited.
Ethics oversight	This study was approved by the Ethics Committees of the Shanghai Changhai Hospital (CHEC2023-088).

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

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	This study used data from public databases for analysis. Due to the limitations of public databases, there were relatively few related sequencing studies. The single-cell sequencing data selected had three replicate samples in each group. The selected bulk sequencing had the largest sample size among the existing sequencing data and has temporality, which was more in line with our research purpose. In the verification stage, we did not calculate the sample size. Instead, we collected as many clinical patient samples as possible for verification.
Data exclusions	Patients lacking complete clinical information will be excluded.
Replication	Yes.
Randomization	No. Our grouping information was for patients with a corresponding definite diagnosis. The corresponding samples collected were all uniformly analyzed and there was no allocation process involved.
Blinding	No. We performed protein sequencing or immunohistochemistry on patient samples. The data information was obtained through batch operations in accordance with the protocol. Not conducting blinding does not affect the experimental results.

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 and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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

Antibodies

Antibodies used

Anti-MMP1 antibody [EP1247Y](ab52631; Abcam;1:500), Anti-Kallikrein 6 antibody - C-terminal (ab190924; Abcam;1:1000), Anti-TPSB2 Polyclonal Antibody(K007207P; Solarbio;1:500), and FOXC2 Polyclonal Antibody (PA5-24588; Invitrogen;1:50)

Validation

Anti-MMP1 antibody: suitable for ICC/IF, Flow Cyt (Intra), and IHC-P; reacts with Human
Anti-Kallikrein 6 antibody - C-terminal: suitable for ICC/IF, WB, and IHC-P; reacts with Human
Anti-TPSB2 Polyclonal Antibody: suitable for WB and IHC; reacts with Human
FOXC2 Polyclonal Antibody: suitable for WB, IHC, Flow, and chIP; reacts with Human