cellSight: Characterizing dynamics of cells using single-cell RNA-sequencing

Authors: Ranojoy Chatterjee[1], Chiraag Gohel[1], Brett A Shook[2], Ali Rahnavard[1,*]

[1]Computational Biology Institute, Department of Biostatistics and Bioinformatics, Milken Institute School of Public Health, The George Washington University, Washington, DC 20052

[2]Department of Biochemistry and Molecular Medicine, School of Medicine and Health Sciences, The George Washington University, Washington, DC 20052

*Correspondence to rahnavard@gwu.edu

SUMMARY

Single-cell analysis has transformed our understanding of cellular diversity, offering insights into complex biological systems. Yet, manual data processing in single-cell studies poses challenges, including inefficiency, human error, and limited scalability. To address these issues, we propose the automated workflow *cellSight*, which integrates high-throughput sequencing in a user-friendly platform. By automating tasks like cell type clustering, feature extraction, and data normalization, *cellSight* reduces researcher workload, promoting focus on data interpretation and hypothesis generation. Its standardized analysis pipelines and quality control metrics enhance reproducibility, enabling collaboration across studies. Moreover, *cellSight*'s adaptability supports integration with emerging technologies, keeping pace with advancements in single-cell genomics. *cellSight* accelerates discoveries in single-cell biology, driving impactful insights and clinical translation. It is available with documentation and tutorials at https://github.com/omicsEye/cellSight.

KEYWORDS

Single-cell analysis; Intercellular communication; Cell interaction; Software; Bioinformatics

INTRODUCTION

The advent of high-throughput single-cell technologies has generated unprecedented volumes of biological data, presenting significant computational challenges in processing, analyzing, and interpreting cellular heterogeneity and dynamics[1–4]. The field of single-cell genomics emphasizes the growing need for sophisticated computational tools[5,6] capable of handling the increasing complexity of data analysis, particularly in understanding cellular heterogeneity[7–9] and their temporal dynamics. The computational demands are further amplified by recent technological advances in single-cell methodologies, such as droplet-based[10,11] and plate-based assays[12,13], which can generate data from thousands to millions of individual cells in a single experiment.

The computational complexity of single-cell analysis extends beyond mere data volume[14,15]. Current analytical pipelines must address multiple challenges, including batch effect correction, dimensionality reduction, feature selection, and cell type identification[16,17]. Traditional computational approaches often rely on manual quality control(QC) steps and differential expression analyses, which are computationally intensive and prone to reproducibility issues and human error[18–20]. These challenges are compounded by the high-

dimensional nature of single-cell data, where each cell is characterized by the expression levels of tens of thousands of genes[21,22].

To address these computational challenges, we introduce *cellSight*, an innovative automated computational framework specifically designed to handle the complexity and scale of single-cell data analysis[23,24]. Our solution integrates state-of-the-art computational methods for data processing, visualization, and interpretation. Central to *cellSight's* architecture is its robust QC pipeline integrated with statistical modeling and differential expression analyses by considering the usual linear and generalized linear models, emphasizing the role of zero-inflation in the interval data. The results are used in the pipeline to analyze intercellular communication networks during homeostasis.

cellSight represents a significant advance in computational biology, providing an automated, end-to-end solution that addresses key computational bottlenecks in single-cell analysis[25]. The framework incorporates parallel processing capabilities and optimized algorithms for handling large-scale datasets while maintaining statistical rigor and biological relevance. By automating complex computational tasks, *cellSight* not only enhances the efficiency of single-cell analysis but also improves reproducibility and reduces computational overhead[26].

To validate our computational framework, we applied *cellSight* to two distinct datasets: a mouse skin injury model and a previously published skin aging study[27]. Our automated pipeline successfully reproduced the findings from the skin aging study while generating novel insights into the role of fibroblasts during the healing process in the mouse injury model. These results demonstrate the robustness and versatility of our computational approach in handling diverse biological contexts while maintaining computational efficiency.

RESULTS

DYNAMICS OF CELLS IN INJURY-INDUCED SKIN INFLAMMATION

An epithelium covers the outer layer of skin on all animals, protecting against damage, water loss, and infection. When the epithelial layer is damaged, the body swiftly induces a plethora of molecular and cellular changes to repair the tissue. The process of healing has been characterized as three major overlapping phases[28–30]. The first stage is inflammation, which is prompted as the tissue is damaged, involves the recruitment and activation of pro-inflammatory immune cells. The second phase involves the creation of granular tissue and a new epithelium over the injured area. The final stage is the remodeling of the newly generated tissue, particularly the extracellular matrix. The healing process is a time-intensive process that can take months.

Four distinct datasets were meticulously collected from a comprehensive study involving two groups of mice: two datasets from injured skin and two from uninjured skin. This meticulous selection of datasets aimed to unravel the intricate molecular roles and dynamics occurring during the process of injury. Each dataset represents a snapshot of the cellular landscape at a specific point in time, capturing the unique gene expression profiles and cellular responses associated with injury and contrasting them with those from uninjured samples. Including injured and uninjured samples provides a valuable comparative framework, allowing researchers to pinpoint differentially expressed genes and pathways specific to the injury context. This approach enables a thorough exploration of the molecular mechanisms

underlying the response to injury, potentially revealing critical insights into the cellular processes, signaling pathways, and dynamic changes in gene expression that contribute to tissue repair or exacerbate damage. The carefully curated datasets form the foundation for a detailed and nuanced analysis, fostering a deeper understanding of the roles played by specific genes and cellular components in the context of injury and recovery in the mouse model.

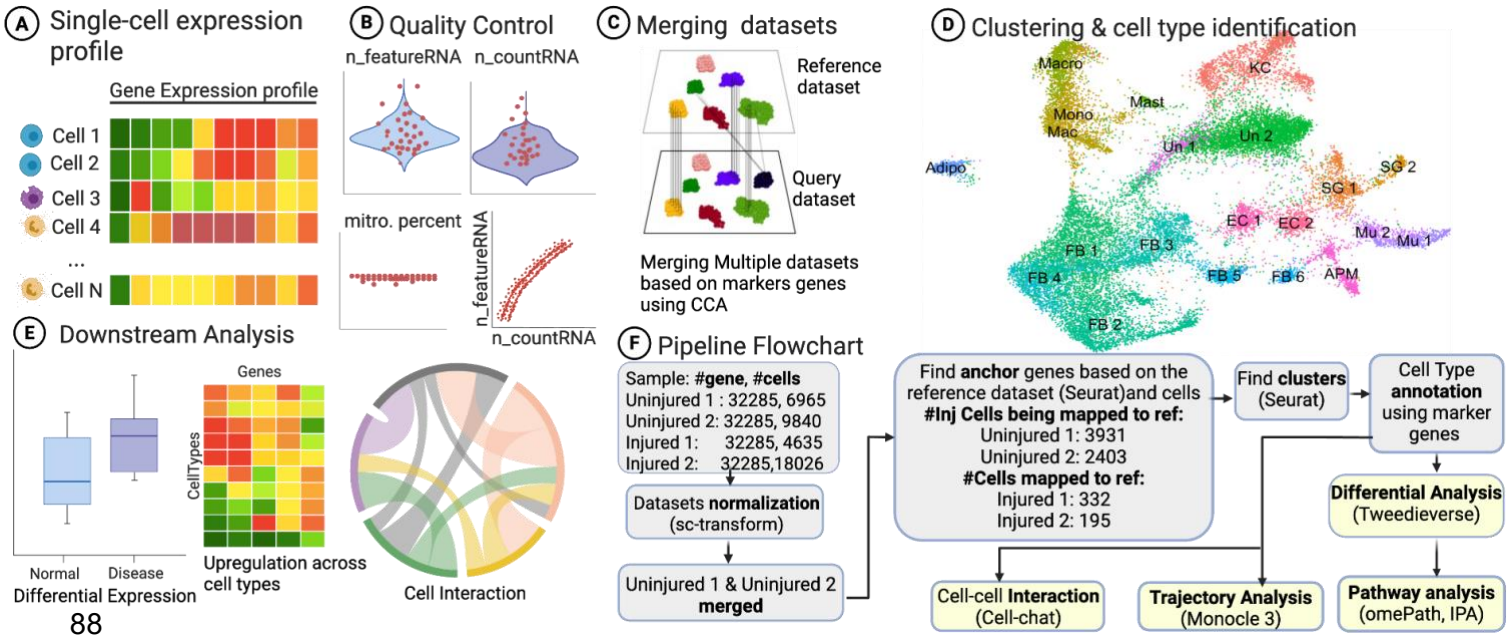

Figure 1. Single-cell Analysis Pipeline for Transcriptional Profiling. (**A**) Single-nucleus RNA expression heatmap displaying gene expression patterns across individual cells (Cell 1, Cell 2,..., Cell N). (**B**) Quality control is based on principal components to distinguish viable cells from unreliable data points. (**C**) Dataset integration through canonical correlation analysis (CCA), merging reference and query datasets to identify common cell populations. (**D**) High-dimensional clustering visualization showing distinct cell type populations with annotated labels (e.g., Endothelial Cells, Fibroblasts) created in the previous step. (**E**) Downstream analyses featuring circular segmented plots of cell-cell interactions and the significance of the disease states visualized by box plots and heatmaps. (**F**) cellSight workflow: Data normalization and integration, cluster annotation, differential expression (Tweedieverse), trajectory analysis (Monocle3), pathway analysis (omePath), and cell-cell interaction mapping (CellChat).

cellSight's pipeline streamlined the complex single-cell data analysis into a straightforward and efficient process, as demonstrated in our study of skin injury response. Through automated quality control, normalization, and clustering, we successfully detected differentially expressed genes between injured and naive samples, revealing the direct influence of fibroblast subtypes on the wound healing process. The integration module enabled an overarching visualization of the differences in biological states across sample types, while cell-cell interaction analysis systematically mapped the multifaceted communication networks governing the coordinated healing response. This case study demonstrates the ability of researchers to rapidly extract biologically meaningful insights from complex single-cell data without requiring exhaustive computational expertise.

To illustrate the end-to-end automation functionalities of cellSight, we ran our developed pipeline on single-cell datasets from a mouse skin injury model. The automated pipeline successfully processed four datasets consisting of 39,466 cells over 32,285 genes, thereby removing the need for manual data manipulation and format conversion, which typically

consumes much time for researchers. cellSight's quality control module automatically applied optimized filtering thresholds (RNA count >200, unique genes >2500, total RNA molecules >8000, mitochondria percentage = 0) without the need for manual parameter tuning or iterative tuning cycles. The pipeline then successfully performed canonical correlation analysis for dataset integration—a computationally demanding step that typically necessitates expert expertise—and carried out unsupervised clustering that revealed 21 cell populations. Moreover, cellSight automates the creation of publication-ready visual representations in the form of violin plots, feature plots, and dimensional reduction embeddings, thereby greatly speeding up the process of cell type annotation. Although the pipeline does not automatically apply cell type labels, it produces standardized expression matrices and interactive visualizations that enhance the efficiency of expert annotation. Through these automated visualizations, we observed that fibroblast populations expressed *Pdgfrα*, Col1a1, and *Dcn*[31–33], while immune populations were demarcated by markers including *Adgre1* (macrophages) and *Ly6c2* (monocytes) cell types[34–36]. This automation-assisted strategy retains the important role of expert biological knowledge in the identification of cell types while reducing annotation times from days to hours. Figure 2A illustrates how cellSight automatically renders extensive visualizations of quality control metrics without manual formatting, while (**Fig. 2B**) depicts successful multi-sample integration through dimensional reduction visualization—an effort that would otherwise entail extensive coding proficiency. Subsequently, cellSight's differential expression module automatically performs statistically rigorous comparisons between experimental conditions using the Tweedie model, specifically designed to handle the zero-inflation properties inherent in single-cell data. By automating such intricate analytical processes—from preprocessing and normalization to visualization and statistical evaluation—cellSight removes many of the error-prone, time-consuming manual steps commonly involved in single-cell workflows. This end-to-end automation makes cutting-edge transcriptomic analysis accessible to researchers lacking specialized computational training, thereby democratizing access to the insights of single-cell biology within the biomedical research community.

After the naming of each cluster was completed, differential expression(D.E.) analysis was performed on the 4 samples to find out the genes that have changed in their expression after injury using the Tweedie model. We found several cytokines and growth factors expressed by keratinocytes and monocytes alongside the fibroblast clusters. In addition, we also found some cytokines and growth factors that were uniquely expressed by the fibroblast clusters only. The results clearly show the importance of *Ccl* in injury since *Ccl*(*Ccl2*) is employed to recruit monocytes and macrophages to the wounded area, facilitating healing. Cell-cell interaction network analysis (**Fig. 2C**) identified activated signaling pathways between keratinocytes, immune cells, and fibroblasts during wound response. The results highlight the potential importance of Ccl in injury, with our analysis showing elevated Ccl2 expression, which is known to be involved in recruiting monocytes and macrophages to wounded areas[37,38]. Cell-cell interaction network analysis (**Fig. 2C**) predicted activated signaling pathways between keratinocytes, immune cells, and fibroblasts during wound response. Our analysis suggests that increased *Ccl2* expression could facilitate monocyte/macrophage recruitment[39], while pro-inflammatory cytokines (*IL-1*, *IL-6*, *IL-8*) may initiate immune response and pathogen clearance[40,41]. Based on known pathway functions, the observed *IL-4/IL-13* expression patterns are predicted to promote fibroblast proliferation and collagen synthesis[42], with IL-8 potentially stimulating endothelial cell migration and angiogenesis[43].

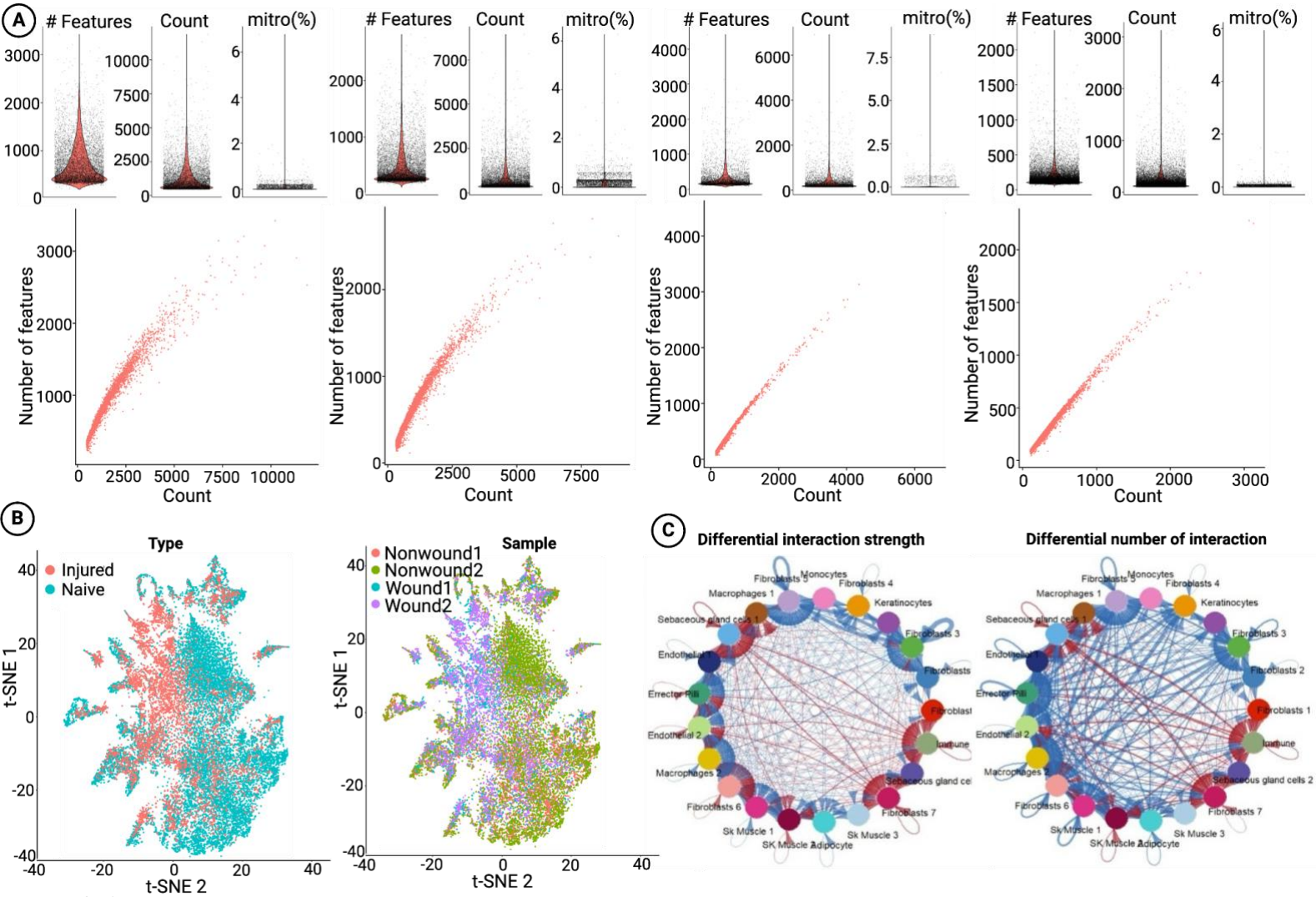

Figure 2. Comprehensive computational analysis of injury-induced transcriptional changes in mouse skin tissue. (**A**) Quality control metrics assessment, including gene detection rates, RNA molecule quantification, and mitochondrial content analysis. (**B**) Multi-sample integration validation through dimensional reduction visualization. (**C**) Cell-cell interaction network analysis highlights the activated communication pathways between keratinocytes, immune cells, and fibroblasts during wound response.

ROLE OF FIBROBLASTS IN SKIN AGING

Historically, knowledge about skin cell components has been largely derived from mouse studies, utilizing reporter constructs, lineage tracing, fluorescence-activated cell sorting (FACS), and immunohistochemistry (IHC). These studies identified distinct fibroblast subtypes in different dermal layers, suggesting functional diversity. However, to bolster our understanding of the mechanism in skin tissue, scRNA-seq data were utilized to capture the underlying workings.

In this study [27], the researchers conducted a comprehensive analysis of human dermal fibroblasts at the single-cell level, focusing on a defined, sun-protected area from both young (25 and 27 years old) and old (53–70 years old) male Caucasian donors. By analyzing over 15,000 cells, the researchers identified four main fibroblast subpopulations with distinct spatial localizations and characteristic functional annotations.

cellSight automated quality control module generates comprehensive multi-parameter metrics visualizations (**Fig. 3A**) with validated data quality across young (25-27 years) and old (53-70 years) male Caucasian donors without manual parameter tuning. The integration analysis module of the pipeline performs automated batch effect correction and provides dimensionality reduction plots showing effective data harmonization (**Fig. 3B**). With its adaptive clustering algorithms, cellSight computes unsupervised clustering visualizations at multiple resolutions (**Fig. 3C**), thereby allowing researchers to effortlessly discern discrete cell populations across all donors without necessitating sophisticated computational skills. The automatically generated high-quality visualizations illustrate how cellSight accelerates the analytical process, allowing researchers to devote more time to biological interpretations and less to computational ones.

The study provided insights into the differential 'priming' of fibroblasts, resulting in functionally heterogeneous subgroups. *cellSight's* built-in visualization modules automatically produced high-resolution UMAPs, feature plots, and violin plots that reveal this cellular heterogeneity, allowing researchers to identify the age-associated loss of identity among fibroblast subpopulations without any manual parameter optimization. The pipeline also creates standardized differential expression heatmaps that graphically display the decrease in unique features of every subgroup, and its cell-cell interaction visualization tools can generate network diagrams automatically displaying how aged fibroblast subpopulations, which express specific skin aging-associated secreted proteins (SAASP), have predicted decreases in interactions with other skin cell types. The figures present how cellSight converts intricate single-cell datasets into easy-to-interpret visualizations that address fundamental transcriptomic inquiries regarding cellular identity, state transitions, and cell-to-cell interactions achieved through an automated pipeline without requiring advanced computational expertise or the creation of custom scripts.

In our second case study, we wanted to verify the proper working of the *cellSight* pipeline. To validate the proper working, we ran the automated pipeline on a previously published study that sampled old and young skin tissue to observe the change in fibroblast priming in old people. Our automated pipeline shows the same results as they obtained from their analysis. In the initial analysis, we integrated cells from all five samples, resulting in a UMAP plot displaying 17 clusters with distinct expression profiles. Notably, each cluster contained cells from all donors. Expression profiling of canonical markers revealed molecular signatures of inflammatory, secretory, and mesenchymal states across fibroblast subpopulations (**Fig. 3C**). Initial analysis integrated cells from all five samples, yielding 17 clusters with distinct expression profiles. Each cluster contained cells from all donors, with fibroblasts, marked by *LUM*, *DCN*, *VIM*, *PDGFRA*, and *COL1A2*[44–46], constituted the most abundant skin cell type, represented by four clusters. Analyzing each sample individually produced a similar number of clusters and identified the same major cell types(**Fig. 4A**).

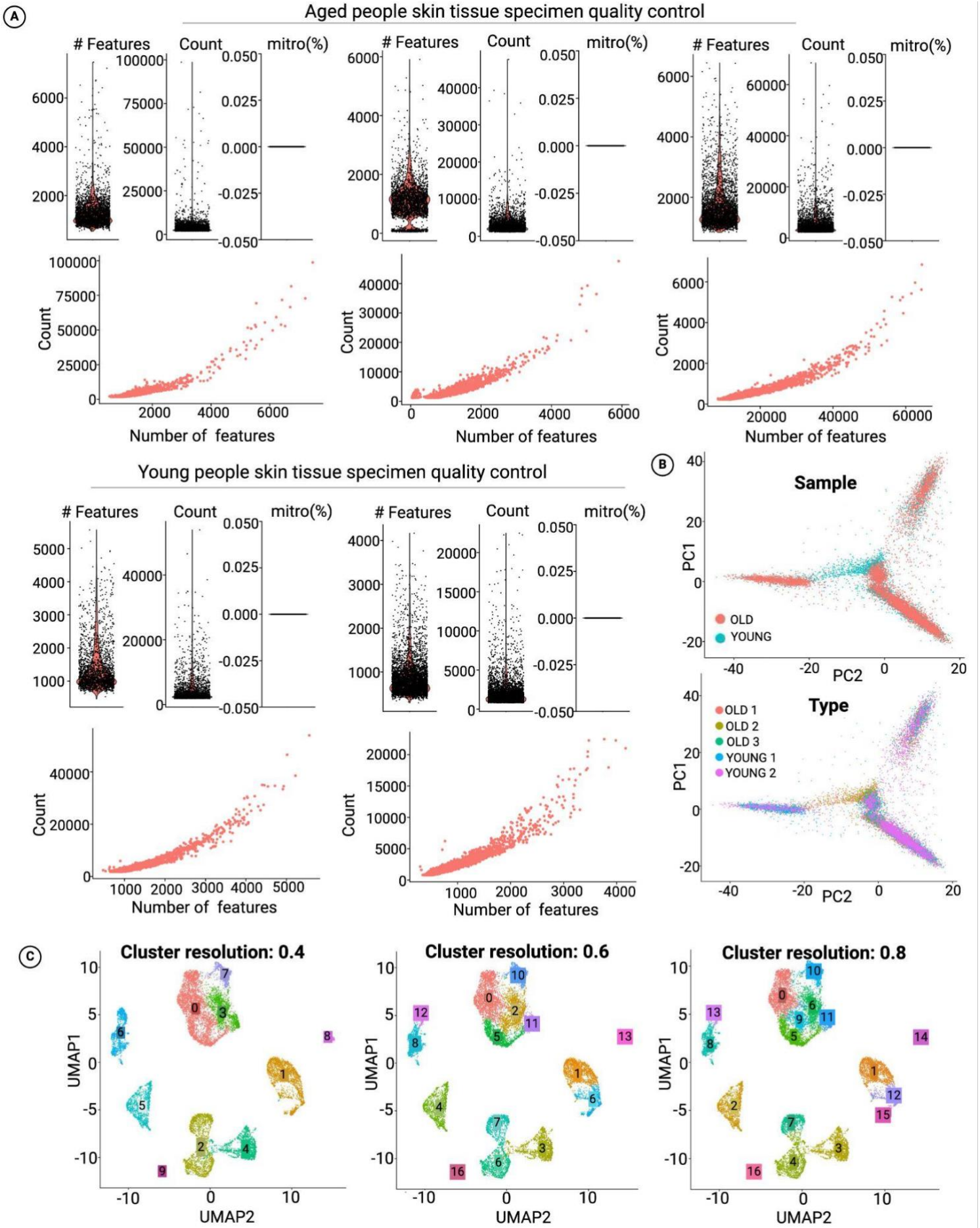
A
Aged people skin tissue specimen quality control
Features
Count
mitro(%)
Count
Number of features
Young people skin tissue specimen quality control
B
Sample
OLD
YOUNG
PC1
PC2
Type
OLD 1
OLD 2
OLD 3
YOUNG 1
YOUNG 2
C
Cluster resolution: 0.4
Cluster resolution: 0.6
Cluster resolution: 0.8
UMAP1
UMAP2

Figure 3. Single-cell transcriptomic analysis of age-dependent changes in human skin fibroblasts. (**A**) Multi-parameter quality control metrics demonstrating data quality across samples. (**B**) Sample integration analysis showing batch effect correction and data harmonization. (**C**) Unsupervised clustering analysis at varying resolutions reveals distinct cell populations.

Intercellular communication network analysis (**Fig. 4C**) revealed age-associated changes in cell-type-specific signaling patterns. Key findings included differential 'priming' of fibroblasts into functionally distinct subgroups and age-related loss of identity across all fibroblast subpopulations (**Fig. 4B**). Notably, older fibroblast subpopulations expressed specific skin aging-associated secreted proteins (SAASP) and showed decreased interactions with other skin cell types.

Individual sample analysis produced consistent clustering patterns and cell type identification, validating the robustness of our analytical approach. This comprehensive characterization provides insights into age-related changes in fibroblast function and intercellular communication networks.

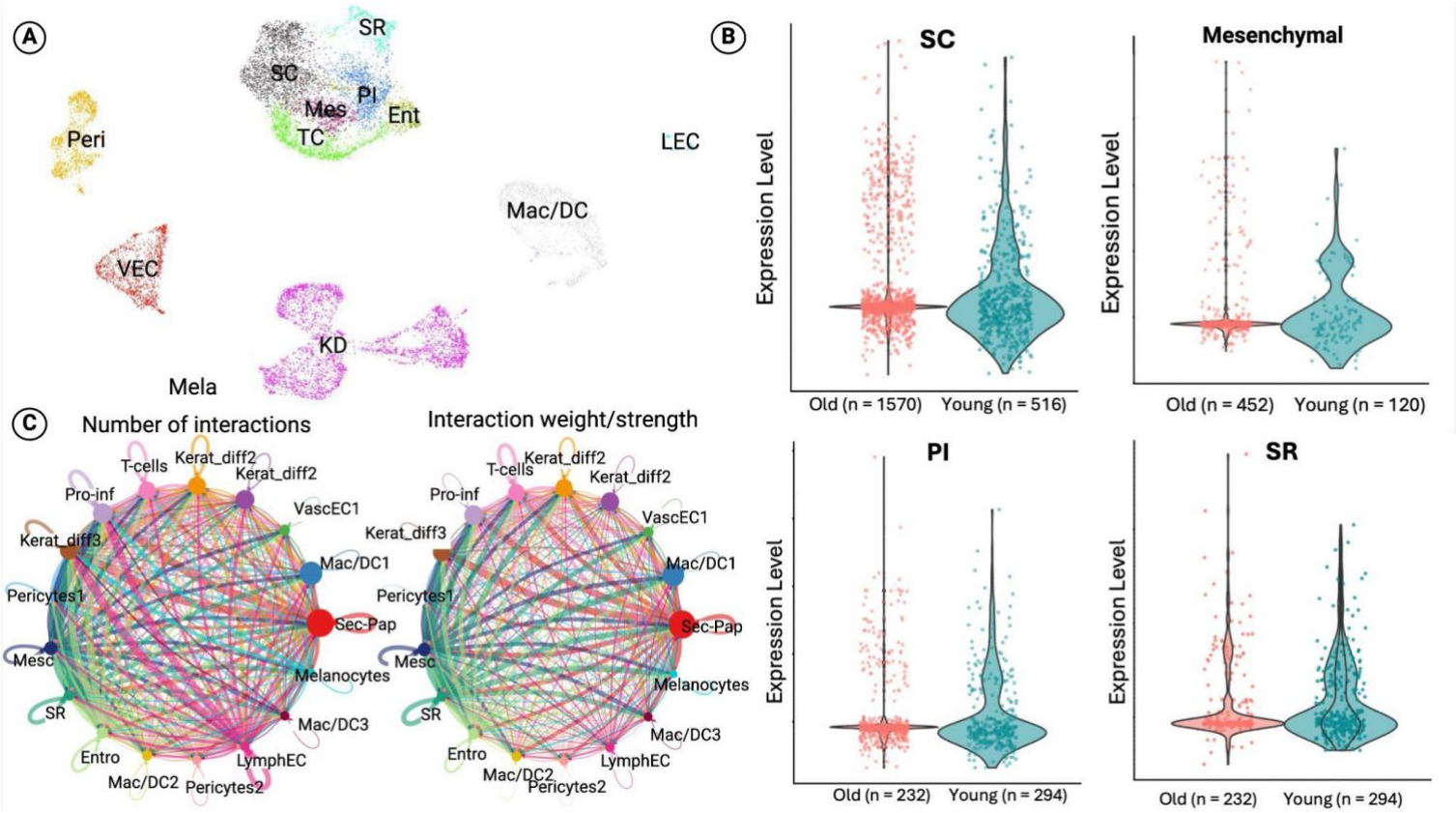

Figure 4. Expression analysis of human skin fibroblasts during aging. (A) Cell type identification revealed diverse populations in the skin tissue, including multiple fibroblast subtypes, epithelial cells, immune cells, and vascular components. (B) Expression profiling of canonical markers across identified fibroblast subpopulations demonstrates molecular signatures of inflammatory, secretory, and mesenchymal states. (C) Intercellular communication network analysis reveals age-associated changes in cell-type specific signaling patterns.

As shown in the step-by-step timeline (**Fig. 5**), *cellSight* condenses ten years of analytic software development (2015-2024) that integrates approaches from early platforms like Seurat to the latest methods in spatial transcriptomics and intercellular communication. The end-to-end pipeline addresses the main bottlenecks in single-cell analysis through automated quality control, normalization, dimensionality reduction, clustering, and biological

interpretation. *cellSight* is unique in that it integrates Tweedieverse for differential expression analysis and CellChat for intercellular communication networks— benchmarked with fibroblast heterogeneity characterization in wound healing and aging models. The pipeline overcomes key analytical hurdles such as batch effect correction, parameter optimization, and cell type annotation without compromising statistical stringency and biological relevance. By reducing computational skill demands at the cost of analysis trade-offs, *cellSight* provides workflow standardization for use across varied tissue sources and experimental setups. With a scalable design tuned for managing datasets ranging from thousands to millions of cells, *cellSight* facilitates reproducibility and enables advanced single-cell analytics for the wider research community.

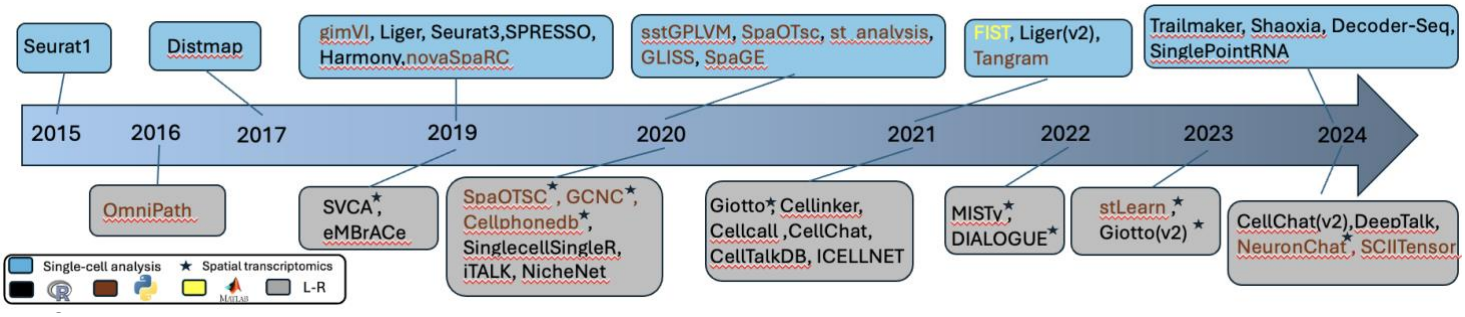

Figure 5. Timeline of single-cell analytical framework development (2015-2024). A comprehensive overview of key computational tools that have shaped single-cell and spatial transcriptomics analysis over the past decade. The timeline illustrates the progression from foundational platforms (blue boxes, top row) like Seurat1 (2015) to specialized spatial transcriptomics tools (gray boxes, bottom row) marked with asterisks. Color-coded text indicates tools focused on specific analytical domains: black for general analysis platforms, blue for single-cell tools, brown for spatial transcriptomics, yellow for integration methods, red for intercellular communication, and gray for ligand-receptor interaction analysis. The evolution demonstrates the field's trajectory from basic dimensionality reduction and clustering approaches toward sophisticated integrated frameworks culminating in advanced communication network analysis.

DISCUSSION

The *cellSight* platform is an important development in single-cell RNA sequencing analysis that provides investigators with an extensible and automated pipeline incorporating a wide range of analytical methods into one platform, as demonstrated in **Figure 6**. The pipeline initiates with stringent quality control and preprocessing procedures prior to creating insightful violin plots (**Fig. 6A)** that reveal the distribution of gene expression in the detected clusters. These meticulously crafted plots simultaneously depict expression prevalence and intensity, where the width precisely captures the proportion of cells expressing particular markers at specific levels while the height corresponds to expression magnitude. This dual visualization enables researchers to immediately discern cluster-specific expression patterns across multiple cell types, revealing both subtle and pronounced differences that define cellular identities without requiring labor-intensive manual annotation procedures.

Complementing this approach, feature plots (**Fig. 6B**) project expression patterns onto two-dimensional UMAP coordinates with remarkable precision. The gradient-based color intensity systematically highlights expression hotspots across the cellular landscape, creating intuitive spatial maps that reveal both localized expression domains and transitional zones between cell states. These visualizations transform complex multidimensional expression data into

interpretable spatial relationships that facilitate the identification of functional territories and developmental trajectories within heterogeneous tissues.

The UMAP visualization module (**Fig. 6C**) demonstrates *cellSight's* sophisticated dimensionality reduction and clustering algorithms by depicting how transcriptionally similar cells aggregate within high-dimensional space. The left panel presents a comprehensive overview of the cellular ecosystem with precisely labeled populations reflecting their distinctive transcriptional signatures. The right panel showcases the pipeline's re-clustering functionality, which addresses a critical challenge in single-cell analysis, the detection of rare or transitional cell states that often remain hidden within broader populations. This re-clustering approach enables targeted analysis of specific cellular compartments (illustrated by the circled vascular region) to reveal cryptic subpopulations with distinct molecular signatures. Such hierarchical subpopulation discovery is particularly valuable for identifying stem cell niches, differentiation intermediates, and disease-associated cell states that might represent only a fraction of the broader population yet play pivotal roles in tissue function. The ability to iteratively re-cluster specific regions of interest allows researchers to progressively refine cellular taxonomies, uncovering biological heterogeneity at multiple resolutions that more accurately reflect the complex organizational hierarchies present in living tissues. This capability proves especially valuable in disease contexts where pathological cell states may emerge as subtle variants of normal populations, enabling more precise identification of therapeutic targets and biomarkers.

A particularly innovative component of *cellSight* is the integrated module for cell-cell interaction analysis (**Fig. 6D)**, generating paired heatmaps that quantitatively assess both the frequency and strength of predicted molecular communications between distinct cellular populations. These comprehensive interaction matrices, with sender populations arrayed vertically and receiver populations horizontally, systematically capture the elaborate intercellular signaling networks that orchestrate tissue function. The complementary metrics of interaction frequency and strength provide multidimensional insights into communication dynamics, highlighting predominant signaling axes and revealing potential regulatory mechanisms underlying tissue homeostasis or disease progression. This systematic approach to mapping cellular crosstalk uncovers critical signaling nodes and potential therapeutic targets that would remain undetected through conventional gene expression analysis alone.

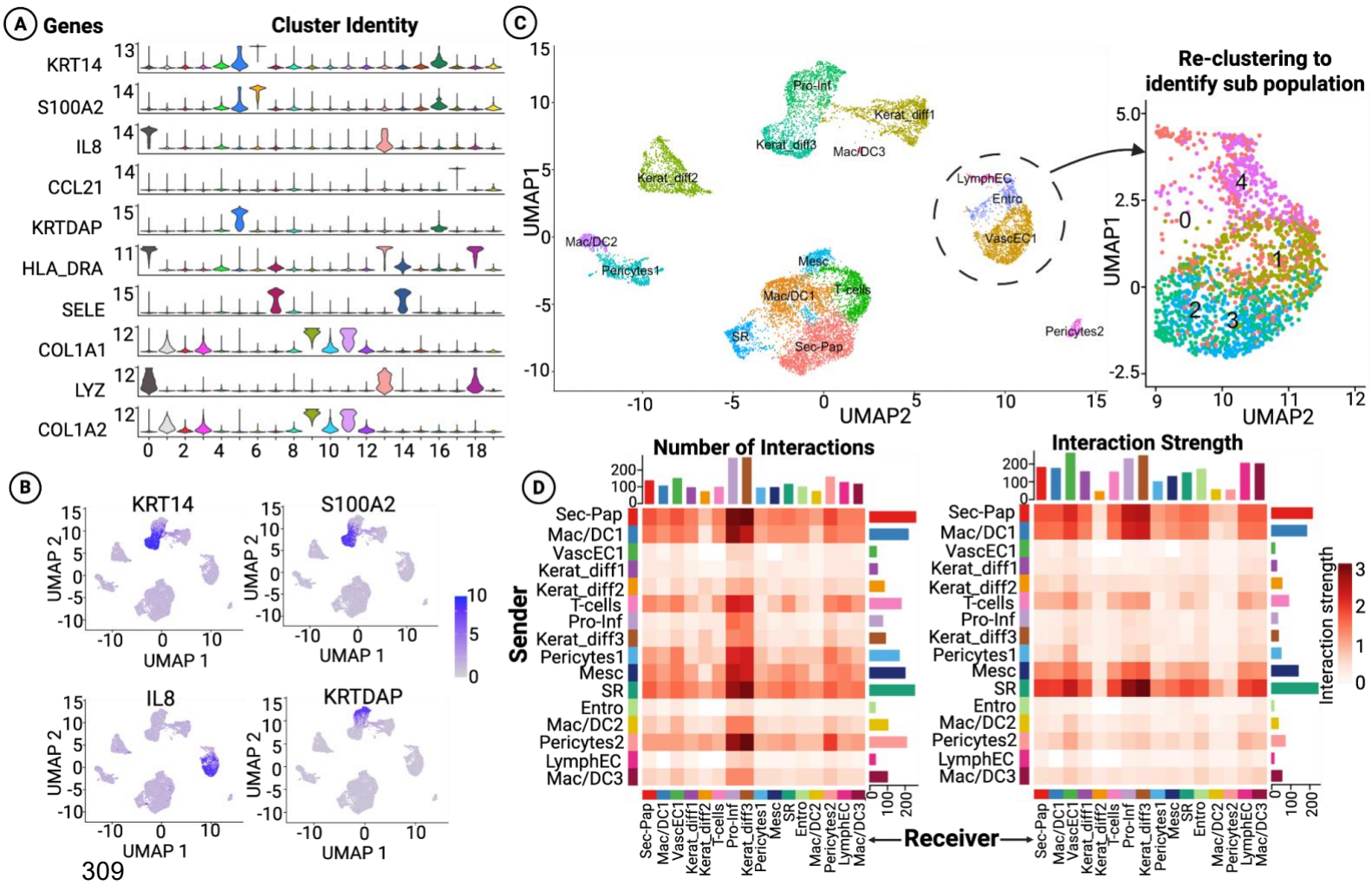

Figure 6. Multi-dimensional analysis of cellular heterogeneity and intercellular interactions within tissue microenvironment. (**A**) Violin plots showing marker gene expression distributions across cell clusters. (**B**) Feature plots mapping marker expression (blue intensity) on UMAP coordinates. (**C**) UMAP visualizations of cell clustering - left panel shows major cell populations, right panel shows five subclusters (0-4) within a specific region. This hierarchical approach is essential for identifying rare cell states, differentiation intermediates, and disease-associated variants that play critical functional roles despite representing small fractions of the tissue. (**D**) Cell-cell interaction analysis through paired heatmaps quantifying communication frequency (left) and strength (right) between sender (y-axis) and receiver (x-axis) populations. These matrices reveal the intercellular signaling networks orchestrating tissue function, highlighting key communication channels and regulatory hubs that coordinate multicellular responses in development, homeostasis, and disease.

In the mouse injury study, the analysis of wounded and naive skin tissue samples revealed significant insights into the role of fibroblasts and other cell types in wound healing. Our discovery of seven distinct fibroblast subtypes (FB1-FB7) emphasizes the functional heterogeneity within this population. This suggests that fibroblasts play diverse roles in different stages of tissue repair, potentially contributing to processes such as extracellular matrix remodeling[47], cytokine secretion[48,49], and cellular signaling[50] during healing.

Detecting differentially expressed genes between the injured and naive samples, particularly those encoding cytokines and growth factors, further supports the pivotal role of fibroblasts and other immune cells (monocytes and macrophages) in modulating the inflammatory response. The upregulation of *Ccl2* is of particular interest as it reinforces the importance of chemokine-mediated recruitment of immune cells to the wound site. Similarly, the observed

increase in pro-inflammatory interleukins (*IL-1, IL-6, IL-8*)[51] aligns with their well-established roles in initiating the early phases of inflammation, immune cell recruitment, and tissue repair. These findings align with current knowledge of wound healing, where an orchestrated sequence of cellular communication is essential for efficient repair. Identifying specific fibroblast-expressed cytokines[52] and growth factors[53,54] unique to this cell population opens new avenues for understanding how fibroblasts regulate their local environment and contribute to effective tissue repair.

In the skin aging study, the focus shifted to understanding how fibroblast identity and functionality change with age. By validating the cellSight pipeline on a previously published dataset, we confirmed the age-related decline in the distinct characteristics of fibroblast subtypes. Notably, older fibroblast populations exhibited reduced functional specialization, coupled with the expression of aging-associated secreted proteins (SAASP)[55,56], consistent with the senescence-related changes in cellular communication observed in aged tissues[57]. The loss of fibroblast identity observed in aged donors is concerning as it suggests that fibroblasts lose their ability to efficiently participate in tissue homeostasis and repair over time, potentially contributing to age-related tissue dysfunction.

Our findings are consistent with previous studies that have linked fibroblast dysfunction with impaired wound healing and tissue regeneration in older individuals. The decreased interactions between fibroblasts and other skin cell types in aged samples, as predicted by the cell-cell communication analysis, may explain the diminished regenerative capacity of aging skin. Moreover, the presence of SAASP in old fibroblasts likely contributes to the chronic low-grade inflammation associated with skin aging, further impairing tissue repair mechanisms.

Taken together, these two studies underscore the critical roles that fibroblasts play in both wound healing and skin aging. In the context of injury, fibroblasts facilitate repair by promoting immune cell recruitment, extracellular matrix deposition, and tissue remodeling. However, during aging, fibroblasts progressively lose their functional identity, contributing to impaired healing and increased tissue fragility. These insights have important implications for developing therapeutic strategies aimed at enhancing wound healing and mitigating the effects of aging on skin repair.

Alongside biological insights, *cellSight's* modular architecture is expressly configured to facilitate the incorporation of novel computational methodologies. The flexible framework is planned to accommodate spatial transcriptomics, improved trajectory inference, and improved functionality for cell-cell communication analysis. Our open-source strategy invites contributions from the research community, with researchers able to craft bespoke modules while ensuring compatibility with the primary pipeline. This flexibility means that *cellSight* improves in tandem with methodological innovation in the field, allowing researchers to access state-of-the-art analysis without requiring extensive computational know-how.

METHODS

Single-cell RNA sequencing (scRNA-seq) has transformed our ability to study cellular diversity by enabling the analysis of gene expression at the individual cell level[3,4]. However, the complexity of scRNA-seq data demands sophisticated analytical tools to derive meaningful insights[5,6]. To mitigate this challenge, *cellSight* integrates several key steps in single-cell genomics analysis. The process began with QC and normalization[19], followed by dimension reduction and clustering to identify distinct cell populations[17]. We conducted differential expression analysis to identify gene expression changes between conditions[58] and used cell-cell communication analysis to investigate how different cell types interact[59]. Canonical correlation analysis (CCA) was applied to integrate multiple datasets[22], allowing for a comprehensive identification of clusters and biological states. *cellSight* also incorporates methodologies from *Seurat*[60], enhancing its precision and versatility.

DATA PREPROCESSING AND QC

scRNA-seq data undergoes stringent QC and preprocessing steps before analysis. Raw count matrices are loaded into *cellSight*, where standard QC procedures are implemented. These procedures include identifying and removing low-quality cells[19] based on criteria such as mitochondrial content[20], library size[61], and unique molecular identifier (UMI) counts[21]. Additionally, *cellSight* incorporates widely accepted QC metrics like mitochondrial percentage, number of detected features per cell, and total RNA molecule count per cell from the *Seurat*[22] package, ensuring compatibility with established practices in single-cell analysis.

NORMALIZATION

Normalization is crucial in ensuring the comparability of expression profiles across individual cells[62]. In *cellSight*, normalization is performed using the *sctransform* method[63], which accounts for both technical noise and biological variability. In over-dispersed data, where gene expression variability exceeds the expectations of basic models like the Poisson distribution[64], the increased variability stems from both biological differences between cells and technical factors, such as dropout events and batch effects[65,66], making the data more challenging to model. In scRNA-seq, normalization typically involves adjusting the signal with a dropout parameter to account for these complexities[67]. This strategy has a major flaw of overfitting since different groups of genes are being normalized by the same factor[68]. *Sctransform* uses a generalized linear model to estimate each gene UMI count and then pools similar genes with identical gene expression together to regularize the parameter estimate to produce error models[69]. This process of coupling various genes into groups and then estimating the error preserves the underlying biological dissimilarity by applying *sctransform*[63], *cellSight* enhances the accuracy of downstream analyses, enabling robust identification of differentially expressed genes[70].

DIMENSION REDUCTION AND VISUALIZATION

Dimension reduction techniques are used to uncover the latent structure within single-cell data. Principal Component Analysis (PCA) is a key method integrated into *cellSight*, which maps the high-dimensional data to its principal components. The resulting components are utilized to visualize the data in a reduced-dimensional space, providing insights into the underlying cellular heterogeneity. A k-nearest neighbor (KNN) graph[71] is then constructed

410 based on similarities between cells in this reduced space; the edges in the graph represent
411 connections between cells that share similar gene expression profiles. The Louvain
412 algorithm[72] is applied to this graph to identify clusters, representing groups of cells with similar
413 expression patterns. These clusters can then be visualized and further analyzed to discover
414 biological insights, such as identifying cell types or states. The resolution parameter controls
415 the granularity of clustering, allowing users to define broader or finer groupings of cells.

416 DIFFERENTIAL EXPRESSION ANALYSIS

417 Differential expression [73,74] analysis is a cornerstone of single-cell studies, enabling the
418 identification of genes that exhibit significant expression changes across different cell
419 populations. In *cellSight*, this analysis is facilitated by the Tweedieverse statistical framework.
420 Tweedieverse [58] offers a flexible and robust tool for differential expression analysis,
421 accommodating the unique characteristics of single-cell data and providing reliable results for
422 the identification of genes driving cellular diversity. Tweedieverse uses the Tweedie model[75]
423 to discover differentially expressed genes. scRNA-seq is riddled with zero-inflated values,
424 where the estimated coefficients are adjusted by considering the compound Poisson model,
425 which is a part of the generalized Tweedie model.

426 While DESeq2[76] has been widely adopted for differential expression analysis in RNA
427 sequencing data, single-cell RNA sequencing (scRNA-seq) presents unique analytical
428 challenges that necessitate more specialized statistical approaches [13,77]. Our implementation
429 of Tweedieverse [58]in cellSight offers several key advantages over DESeq2, particularly in
430 handling the characteristic zero inflation of scRNA-seq data, where genes may show no
431 expression in many cells due to both technical dropouts and true biological absence [67,78].
432 Unlike DESeq2's negative binomial model, Tweedieverse implements a list of models,
433 including the normal, Poisson, negative binomial, zero-inflated negative binomial, and
434 Tweedie compound Poisson distributions that naturally accommodate excess zeros while
435 maintaining appropriate variance modeling, resulting in an improvement in sensitivity for
436 detecting differentially expressed genes. Furthermore, Tweedieverse's flexible parametric
437 framework better captures cell-to-cell heterogeneity within defined populations [9], and its
438 specialized normalization approaches are better suited for single-cell expression profiles
439 compared to DESeq2's size factor normalization[63,66]. Importantly, Tweedieverse's
440 computational efficiency scales better with increasing cell numbers, maintaining consistent
441 performance across datasets ranging from 5,000 to 50,000 cells, while DESeq2 shows
442 substantial increases in computational overhead[79], making it particularly well-suited for
443 integration into cellSight's automated workflow.

444 CELL-CELL INTERACTION ANALYSIS

445 Understanding cellular communication is crucial for deciphering complex biological
446 processes[80,81]. To address this, *cellSight* [59] incorporates the *Cellchat* package, allowing for the
447 analysis of cell-cell interactions based on ligand-receptor [82,83] pairs. *cellSight* provides a
448 comprehensive view of the communication network within a cellular population by integrating
449 information on ligand-receptor interactions[84,85].

450 VISUALIZATION AND INTERPRETATION

451 The results obtained from the aforementioned analyses are visualized using various plotting
452 techniques within the *cellSight* environment. These include t-distributed Stochastic Neighbor

Emulation (t-SNE) and Uniform Manifold Approximation and Projection (UMAP), which offer intuitive representations of cellular relationships and structures.

INTEGRATION WITH EXISTING TOOLS

cellSight seamlessly integrates with established tools such as *Seurat*, ensuring compatibility with widely adopted workflows in the single-cell analysis community. This integration allows users to leverage the strengths of both *cellSight* and *Seurat*, expanding the analytical capabilities and providing a more comprehensive toolkit for researchers.

In summary, *cellSight* amalgamates state-of-the-art methods for QC, normalization, dimension reduction, differential expression analysis, and cell-cell interaction analysis, offering a robust and comprehensive solution for unraveling the complexities of single-cell genomics.

Finally, to assess the performance of *cellSight*, we conducted thorough evaluations on two independent skin-related studies. The first dataset, derived from mouse skin tissue 24 hrs after injury, focused on how the different cell type dynamics change to adapt and heal the wound. The second dataset, originating from human skin, centered around the importance of fibroblasts' role in skin aging. Performance metrics, including QC plots, integration plots, and clustering, were calculated to validate the tool's effectiveness in accurately capturing cellular heterogeneity and differential gene expression patterns within the skin-related contexts.

RESOURCE AVAILABILITY

LEAD CONTACT

Further information and requests for resources and analysis should be directed to and will be fulfilled by the lead contact, Ali Rahnavard (rahnavard@gwu.edu).

DATA AND CODE AVAILABILITY

All original code has been deposited at https://zenodo.org/records/10041147 and is publicly available at https://github.com/omicsEye/cellSight.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

The methodology development of this work was supported by the National Science Foundation grant DEB-2109688 to A.R. and application investigation as part of effort supported by an NIH grant to B.A.S. and A.R. from NIAMS (AR082417).

GEO accession GSE147744 and GSE265996.

AUTHOR CONTRIBUTIONS

Conceptualization, A.R. and B.A.S.; methodology, A.R., R.C, and C.G.; investigation, A.R., C.G., A.R., and B.A.S.; writing, original draft, R.C., C.G., and A.R.; writing, review & editing, R.C., A.R., C.G., and B.A.S.; funding acquisition, B.A.S., and A.R.; supervision, A.R., and B.A.S.

DECLARATION OF INTERESTS

The authors have no conflict of interest.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES

The author(s) utilized ChatGPT to help refine sentence structure during the preparation of this work. Following the use of this tool, the author(s) thoroughly reviewed and edited the content as necessary, retaining full responsibility for the final content of the publication.

References

1. Kim, K. (2023). Deep learning approaches for analyzing single-cell transcriptomic data. Nature Methods *19*, 1181–1190.

2. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15.

3. Kharchenko, P.V. (2021). Publisher Correction: The triumphs and limitations of computational methods for scRNA-seq. Nat. Methods *18*, 835.

4. Chen, G., Ning, B., and Shi, T. (2019). Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. Front. Genet. *10*, 317.

504 5. Zappia, L., Phipson, B., and Oshlack, A. (2018). Exploring the single-cell RNA-seq
505 analysis landscape with the scRNA-tools database. PLoS Comput. Biol. *14*, e1006245.

506 6. Jovic, D., Liang, X., Zeng, H., Lin, L., Xu, F., and Luo, Y. (2022). Single-cell RNA
507 sequencing technologies and applications: A brief overview. Clin. Transl. Med. *12*, e694.

508 7. Li, C., Wu, H., Guo, L., Liu, D., Yang, S., Li, S., and Hua, K. (2022). Single-cell
509 transcriptomics reveals cellular heterogeneity and molecular stratification of cervical
510 cancer. Commun Biol *5*, 1208.

511 8. Choi, Y.H., and Kim, J.K. (2019). Dissecting cellular heterogeneity using single-cell RNA
512 sequencing. Mol. Cells *42*, 189–199.

513 9. Wang, Y., Tian, X., and Ai, D. (2021). Cell Heterogeneity Analysis in Single-Cell RNA-seq
514 Data Using Mixture Exponential Graph and Markov Random Field Model. Biomed Res.
515 Int. *2021*, 9919080.

516 10. Svensson, V. (2020). Droplet scRNA-seq is not zero-inflated. Nat. Biotechnol. *38*, 147–
517 150.

518 11. Jiang, Z., Shi, H., Tang, X., and Qin, J. (2023). Recent advances in droplet microfluidics
519 for single-cell analysis. Trends Analyt. Chem. *159*, 116932.

520 12. Strzelecka, P.M., Ranzoni, A.M., and Cvejic, A. (2021). Single-Cell Transcriptomic
521 Analysis of Hematopoietic Cells. In Leukemia Stem Cells: Methods and Protocols, C.
522 Cobaleda and I. Sánchez-García, eds. (Springer US), pp. 135–158.

523 13. Lähnemann, D., Köster, J., Szczurek, E., McCarthy, D.J., Hicks, S.C., Robinson, M.D.,
524 Vallejos, C.A., Campbell, K.R., Beerenwinkel, N., Mahfouz, A., et al. (2020). Eleven grand
525 challenges in single-cell data science. Genome Biol. *21*, 31.

526 14. Luecken, M.D., and Theis, F.J. (2019). Current best practices in single-cell RNA-seq
527 analysis: a tutorial. Mol. Syst. Biol. *15*, e8746.

528 15. Osorio, D., and Cai, J.J. (2021). Systematic determination of the mitochondrial proportion
529 in human and mice tissues for single-cell RNA-sequencing data quality control.
530 Bioinformatics *37*, 963–967.

531 16. Galow, A.-M., Kussauer, S., Wolfien, M., Brunner, R.M., Goldammer, T., David, R., and
532 Hoeflich, A. (2021). Quality control in scRNA-Seq can discriminate pacemaker cells: the
533 mtRNA bias. Cell. Mol. Life Sci. *78*, 6585–6592.

534 17. Zhang, Z., Cui, F., Wang, C., Zhao, L., and Zou, Q. (2021). Goals and approaches for
535 each processing step for single-cell RNA sequencing data. Brief. Bioinform. *22*.
536 https://doi.org/10.1093/bib/bbaa314.

537 18. Soneson, C., Srivastava, A., Patro, R., and Stadler, M.B. (2021). Preprocessing choices
538 affect RNA velocity results for droplet scRNA-seq data. PLoS Comput. Biol. *17*,
539 e1008585.

540 19. Li, X., Nair, A., Wang, S., and Wang, L. (2015). Quality control of RNA-seq experiments.
541 Methods Mol. Biol. *1269*, 137–146.

542 20. Omidsalar, A.A., McCullough, C.G., Xu, L., Boedijono, S., Gerke, D., Webb, M.G.,
543 Manojlovic, Z., Sequeira, A., Lew, M.F., Santorelli, M., et al. (2024). Common
544 mitochondrial deletions in RNA-Seq: evaluation of bulk, single-cell, and spatial
545 transcriptomic datasets. Commun. Biol. *7*, 200.

546 21. You, Y., Tian, L., Su, S., Dong, X., Jabbari, J.S., Hickey, P.F., and Ritchie, M.E. (2021).
547 Benchmarking UMI-based single-cell RNA-seq preprocessing workflows. Genome Biol.
548 *22*, 339.

549 22. Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-
550 cell transcriptomic data across different conditions, technologies, and species. Nat.
551 Biotechnol. *36*, 411–420.

552 23. Andrews, T.S., and Hemberg, M. (2018). False signals induced by single-cell imputation.
553 F1000Res. *7*, 1740.

554 24. Kiselev, V.Y., Andrews, T.S., and Hemberg, M. (2019). Publisher Correction: Challenges
555 in unsupervised clustering of single-cell RNA-seq data. Nat. Rev. Genet. *20*, 310.

556 25. Amezquita, R.A., Lun, A.T.L., Becht, E., Carey, V.J., Carpp, L.N., Geistlinger, L., Marini,
557 F., Rue-Albrecht, K., Risso, D., Soneson, C., et al. (2020). Publisher Correction:
558 Orchestrating single-cell analysis with Bioconductor. Nat. Methods *17*, 242.

559 26. Abdelaal, T., Michielsen, L., Cats, D., Hoogduin, D., Mei, H., Reinders, M.J.T., and
560 Mahfouz, A. (2019). A comparison of automatic cell identification methods for single-cell
561 RNA sequencing data. Genome Biol. *20*, 194.

562 27. Solé-Boldo, L., Raddatz, G., Schütz, S., Mallm, J.-P., Rippe, K., Lonsdorf, A.S.,
563 Rodríguez-Paredes, M., and Lyko, F. (2020). Single-cell transcriptomes of the human skin
564 reveal age-related loss of fibroblast priming. Communications Biology *3*, 1–12.

565 28. Rodrigues, M., Kosaric, N., Bonham, C.A., and Gurtner, G.C. (2019). Wound healing: A
566 cellular perspective. Physiol. Rev. *99*, 665–706.

567 29. Guo, S., and Dipietro, L.A. (2010). Factors affecting wound healing. J. Dent. Res. *89*,
568 219–229.

569 30. Sorg, H., and Sorg, C.G.G. (2023). Skin wound healing: Of players, patterns, and
570 processes. Eur. Surg. Res. *64*, 141–157.

571 31. Driskell, R.R., Lichtenberger, B.M., Hoste, E., Kretzschmar, K., Simons, B.D.,
572 Charalambous, M., Ferron, S.R., Herault, Y., Pavlovic, G., Ferguson-Smith, A.C., et al.
573 (2013). Distinct fibroblast lineages determine dermal architecture in skin development and
574 repair. Nature *504*, 277–281.

575 32. Lynch, M.D., and Watt, F.M. (2018). Fibroblast heterogeneity: implications for human
576 disease. J. Clin. Invest. *128*, 26–35.

577 33. Tabib, T., Morse, C., Wang, T., Chen, W., and Lafyatis, R. (2018). SFRP2/DPP4 and
578 FMO1/LSP1 define major fibroblast populations in human skin. J. Invest. Dermatol. *138*,
579 802–810.

580 34. Gordon, S., and Taylor, P.R. (2005). Monocyte and macrophage heterogeneity. Nat. Rev.
581 Immunol. *5*, 953–964.

582 35. Davies, L.C., Jenkins, S.J., Allen, J.E., and Taylor, P.R. (2013). Tissue-resident
583 macrophages. Nat. Immunol. *14*, 986–995.

584 36. Ginhoux, F., and Jung, S. (2014). Monocytes and macrophages: developmental
585 pathways and tissue homeostasis. Nat. Rev. Immunol. *14*, 392–404.

586 37. Deshmane, S.L., Kremlev, S., Amini, S., and Sawaya, B.E. (2009). Monocyte

587 chemoattractant protein-1 (MCP-1): An overview. J. Interferon Cytokine Res. *29*, 313–
588 326.

589 38. Yadav, A., Saini, V., and Arora, S. (2010). MCP-1: chemoattractant with a role beyond
590 immunity: a review. Clin. Chim. Acta *411*, 1570–1579.

591 39. Willenborg, S., Lucas, T., van Loo, G., Knipper, J.A., Krieg, T., Haase, I., Brachvogel, B.,
592 Hammerschmidt, M., Nagy, A., Ferrara, N., et al. (2012). CCR2 recruits an inflammatory
593 macrophage subpopulation critical for angiogenesis in tissue repair. Blood *120*, 613–625.

594 40. Werner, S., and Grose, R. (2003). Regulation of Wound Healing by Growth Factors and
595 cytokines. Physiol. Rev. *83*, 835–870.

596 41. Barrientos, S., Brem, H., Stojadinovic, O., and Tomic-Canic, M. (2014). Clinical
597 application of growth factors and cytokines in wound healing. Wound Repair Regen. *22*,
598 569–578.

599 42. Salmon-Ehr, V., Ramont, L., Godeau, G., Birembaut, P., Guenounou, M., Bernard, P.,
600 and Maquart, F.X. (2000). Implication of interleukin-4 in wound healing. Lab. Invest. *80*,
601 1337–1343.

602 43. Li, A., Dubey, S., Varney, M.L., Dave, B.J., and Singh, R.K. (2003). IL-8 directly enhanced
603 endothelial cell survival, proliferation, and matrix metalloproteinases production and
604 regulated angiogenesis. J. Immunol. *170*, 3369–3376.

605 44. Reynolds, G., Vegh, P., Fletcher, J., Poyner, E.F.M., Stephenson, E., Goh, I., Botting,
606 R.A., Huang, N., Olabi, B., Dubois, A., et al. (2021). Developmental cell programs are co-
607 opted in inflammatory skin disease. Science *371*, eaba6500.

608 45. Philippeos, C., Telerman, S.B., Oulès, B., Pisco, A.O., Shaw, T.J., Elgueta, R., Lombardi,
609 G., Driskell, R.R., Soldin, M., Lynch, M.D., et al. (2018). Spatial and single-cell
610 transcriptional profiling identifies functionally distinct human dermal fibroblast
611 subpopulations. J. Invest. Dermatol. *138*, 811–825.

612 46. Talbott, H.E., Mascharak, S., Griffin, M., Wan, D.C., and Longaker, M.T. (2022). Wound
613 healing, fibroblast heterogeneity, and fibrosis. Cell Stem Cell *29*, 1161–1180.

614 47. Tracy, L.E., Minasian, R.A., and Caterson, E.J. (2016). Extracellular matrix and dermal
615 fibroblast function in the healing wound. Adv. Wound Care *5*, 119–136.

616 48. Cavagnero, K.J., and Gallo, R.L. (2022). Essential immune functions of fibroblasts in
617 innate host defense. Front. Immunol. *13*, 1058862.

618 49. Boraldi, F., Lofaro, F.D., Bonacorsi, S., Mazzilli, A., Garcia-Fernandez, M., and Quaglino,
619 D. (2024). The role of fibroblasts in skin homeostasis and repair. Biomedicines *12*, 1586.

620 50. Gomes, R.N., Manuel, F., and Nascimento, D.S. (2021). The bright side of fibroblasts:
621 molecular signature and regenerative cues in major organs. NPJ Regen. Med. *6*, 43.

622 51. Cialdai, F., Risaliti, C., and Monici, M. (2022). Role of fibroblasts in wound healing and
623 tissue remodeling on Earth and in space. Front Bioeng Biotechnol *10*, 958381.

624 52. Kang, S.-G., Lee, S.-E., Choi, M.-J., Chang, J.-Y., Kim, J.-T., Zhang, B.-Y., Kang, Y.-E.,
625 Lee, J.-H., Yi, H.-S., and Shong, M. (2021). Th2 cytokines increase the expression of
626 fibroblast growth factor 21 in the liver. Cells *10*, 1298.

627 53. Van Linthout, S., Miteva, K., and Tschöpe, C. (2014). Crosstalk between fibroblasts and

628 inflammatory cells. Cardiovasc. Res. *102*, 258–269.

629 54. Farooq, M., Khan, A.W., Kim, M.S., and Choi, S. (2021). The role of fibroblast growth
630 factor (FGF) signaling in tissue repair and regeneration. Cells *10*, 3242.

631 55. Waldera Lupa, D.M., Kalfalah, F., Safferling, K., Boukamp, P., Poschmann, G., Volpi, E.,
632 Götz-Rösch, C., Bernerd, F., Haag, L., Huebenthal, U., et al. (2015). Characterization of
633 skin aging-associated secreted proteins (SAASP) produced by dermal fibroblasts isolated
634 from intrinsically aged human skin. J. Invest. Dermatol. *135*, 1954–1968.

635 56. Boismal, F., Peltier, S., Ly Ka So, S., Chevreux, G., Blondel, L., Serror, K., Setterblab, N.,
636 Zuelgaray, E., Boccara, D., Mimoun, M., et al. (2024). Proteomic and secretomic
637 comparison of young and aged dermal fibroblasts highlights cytoskeleton as a key
638 component during aging. Aging *16*, 11776–11795.

639 57. Di Micco, R., Krizhanovsky, V., Baker, D., and d'Adda di Fagagna, F. (2021). Cellular
640 senescence in ageing: from mechanisms to therapeutic opportunities. Nat. Rev. Mol. Cell
641 Biol. *22*, 75–95.

642 58. Mallick, H., Chatterjee, S., Chowdhury, S., Chatterjee, S., Rahnavard, A., and Hicks, S.C.
643 (2021). Differential expression of single-cell RNA-seq data using Tweedie models.
644 bioRxiv, 2021.03.28.437378. https://doi.org/10.1101/2021.03.28.437378.

645 59. Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Ramos, R., Kuan, C.-H., Myung, P.,
646 Plikus, M.V., and Nie, Q. (2021). Inference and analysis of cell-cell communication using
647 CellChat. Nat. Commun. *12*, 1088.

648 60. Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial
649 reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502.

650 61. Lamarre, S., Frasse, P., Zouine, M., Labourdette, D., Sainderichin, E., Hu, G., Le Berre-
651 Anton, V., Bouzayen, M., and Maza, E. (2018). Optimization of an RNA-Seq differential
652 gene expression analysis depending on biological replicate number and library size.
653 Front. Plant Sci. *9*, 108.

654 62. Vallejos, C.A., Risso, D., Scialdone, A., Dudoit, S., and Marioni, J.C. (2017). Normalizing
655 single-cell RNA sequencing data: challenges and opportunities. Nat. Methods *14*, 565–
656 571.

657 63. Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-
658 cell RNA-seq data using regularized negative binomial regression. bioRxiv.
659 https://doi.org/10.1101/576827.

660 64. Grün, D., Kester, L., and van Oudenaarden, A. (2014). Validation of noise models for
661 single-cell transcriptomics. Nat. Methods *11*, 637–640.

662 65. Hicks, S.C., Townes, F.W., Teng, M., and Irizarry, R.A. (2018). Missing data and technical
663 variability in single-cell RNA-sequencing experiments. Biostatistics *19*, 562–578.

664 66. Lun, A.T.L., McCarthy, D.J., and Marioni, J.C. (2016). A step-by-step workflow for low-
665 level analysis of single-cell RNA-seq data with Bioconductor. F1000Res. *5*, 2122.

666 67. Kharchenko, P.V., Silberstein, L., and Scadden, D.T. (2014). Bayesian approach to
667 single-cell differential expression analysis. Nat. Methods *11*, 740–742.

668 68. Cole, M.B., Risso, D., Wagner, A., DeTomaso, D., Ngai, J., Purdom, E., Dudoit, S., and
669 Yosef, N. (2019). Performance assessment and selection of normalization procedures for

single-cell RNA-seq. Cell Syst. *8*, 315–328.e8.

69. Lun, A.T.L., Riesenfeld, S., Andrews, T., Dao, T.P., Gomes, T., participants in the 1st Human Cell Atlas Jamboree, and Marioni, J.C. (2019). EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. Genome Biol. *20*, 63.

70. Oneson, C., and Robinson, M.D. (2018). Bias, robustness and scalability in single-cell differential expression analysis. Nat Methods.

71. Qi, R., Ma, A., Ma, Q., and Zou, Q. (2020). Clustering and classification methods for single-cell RNA-sequencing data. Brief. Bioinform. *21*, 1196–1208.

72. Zhu, X., Zhang, J., Xu, Y., Wang, J., Peng, X., and Li, H.-D. (2020). Single-cell clustering based on shared nearest neighbor and graph partitioning. Interdiscip. Sci. *12*, 117–130.

73. Zhao, J., Jaffe, A., Li, H., Lindenbaum, O., Sefik, E., Jackson, R., Cheng, X., Flavell, R.A., and Kluger, Y. (2021). Detection of differentially abundant cell subpopulations in scRNA-seq data. Proc. Natl. Acad. Sci. U. S. A. *118*, e2100293118.

74. Miao, Z., and Zhang, X. (2016). Differential expression analyses for single-cell RNA-Seq: old questions on new data. Quant. Biol. *4*, 243–260.

75. Zhou, H., Qian, W., and Yang, Y. (2022). Tweedie gradient boosting for extremely unbalanced zero-inflated data. Commun. Stat. Simul. Comput. *51*, 5507–5529.

76. Love, M.I. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. Genome Biology.

77. Chen, W. (2019). Statistical methods for single-cell RNA-seq: from quality control to differential expression analysis. Bioinformatics.

78. Svensson, V., and Zhou, H. (2020). Tweedie gradient boosting for extremely unbalanced zero-inflated data. Communications in Statistics.

79. Soneson, C., and Robinson, M.D. (2018). Bias, robustness and scalability in single-cell differential expression analysis. Nat. Methods *15*, 255–261.

80. Armingol, E., Officer, A., Harismendy, O., and Lewis, N.E. (2021). Deciphering cell-cell interactions and communication from gene expression. Nat. Rev. Genet. *22*, 71–88.

81. Zanoni, M. (2019). Cellular complexity: new tools, new challenges, and new ways of thinking.

82. Vento-Tormo, R., Efremova, M., Botting, R.A., Turco, M.Y., Vento-Tormo, M., Meyer, K.B., Park, J.-E., Stephenson, E., Polański, K., Goncalves, A., et al. (2018). Single-cell reconstruction of the early maternal-fetal interface in humans. Nature *563*, 347–353.

83. Shao, X., Lu, X., Liao, J., Chen, H., and Fan, X. (2020). New avenues for systematically inferring cell-cell communication: through single-cell transcriptomics data. Protein Cell *11*, 866–880.

84. Browaeys, R., Saelens, W., and Saeys, Y. (2020). NicheNet: modeling intercellular communication by linking ligands to target genes. Nat. Methods *17*, 159–162.

85. Kumar, M.P. (2018). Analysis of cell-cell interactions using ligand-receptor expression patterns. Nat Methods.