

Exploring machine learning strategies for single-cell transcriptomic analysis in wound healing

Jianzhou Cui^{1,2,3,*}, Mei Wang^{4,†}, Chenshi Lin^{1,2,3}, Xu Xu⁵, Zhenqing Zhang^{6,*}

¹Immunology Translational Research Program, Yong Loo Lin School of Medicine, National University of Singapore, 28 Medical Drive, Singapore, 117456, Singapore

²Immunology Program, Life Sciences Institute, National University of Singapore, 28 Medical Drive, Singapore, 117456, Singapore

³NUS-Cambridge Immunophenotyping Centre, National University of Singapore, 28 Medical Drive, Singapore, 117456, Singapore

⁴Key Laboratory of Basic Pharmacology of Ministry of Education, Joint International Research Laboratory of Ethnomedicine of Ministry of Education, Zunyi Medical University, 1, Xiaoyuan Road, Zunyi, 563000, China

⁵Shenzhen Key Laboratory of Marine Bioresources and Ecology, College of Life Sciences and Oceanography, Shenzhen University, 1066, Xueyuan Road, Shenzhen, 518060, China

⁶College of Pharmaceutical Sciences and Jiangsu Key Laboratory of Neuropsychiatric Diseases, Soochow University, 199, Renai Road, Suzhou, 215021, China

*Corresponding authors. Jianzhou Cui, Email: jianzhou.cui@u.nus.edu; Zhenqing Zhang, Email: z_zhang@suda.edu.cn

†Jianzhou Cui and Mei Wang contributed equally.

Abstract

Wound healing is a highly orchestrated, multiphase process that involves various cell types and molecular pathways. Recent advances in single-cell transcriptomics and machine learning have provided unprecedented insights into the complexity of this process, enabling the identification of novel cellular subpopulations and molecular mechanisms underlying tissue repair. In particular, single-cell RNA sequencing (scRNA-seq) has revealed significant cellular heterogeneity, especially within fibroblast populations, and has provided valuable information on immune cell dynamics during healing. Machine learning algorithms have enhanced data analysis by improving cell clustering, dimensionality reduction, and trajectory inference, leading to a better understanding of wound healing at the single-cell level. This review synthesizes the latest findings on the application of scRNA-seq and machine learning in wound healing research, with a focus on fibroblast diversity, immune responses, and spatial organization of cells. The integration of these technologies has the potential to revolutionize therapeutic strategies for chronic wounds, fibrosis, and tissue regeneration, offering new opportunities for precision medicine. By combining computational approaches with biological insights, this review highlights the transformative impact of scRNA-seq and machine learning on wound healing research.

Keywords: Deep learning; Single-cell; Machine learning; Fibroblast diversity; Trajectory inference; Cellular plasticity; Immune cell dynamics

Highlights

- Machine learning algorithms, particularly deep learning, have significantly enhanced the analysis of scRNA-seq data in wound healing research, improving critical tasks such as dimensionality reduction, cell clustering, and trajectory inference.
- Single-cell analyses have revealed remarkable fibroblast heterogeneity, identifying distinct subpopulations involved in scar formation and regenerative healing.
- Integration of scRNA-seq with machine learning has provided insights into immune cell function, elucidating the roles of macrophages and other immune cells in orchestrating both inflammatory responses and tissue repair during wound healing.

Background

Wound healing is an essential biological process that restores tissue integrity and function following injury. It involves a tightly regulated sequence of overlapping phases: clot formation, inflammation, tissue formation, and tissue remodeling [1]. Each phase is characterized by the activation of specific cell types and signaling pathways. For example, hemostasis is initiated immediately after injury by platelets, which release growth factors to recruit immune cells and initiate the coagulation cascade [2]. This is followed by an inflammatory phase, where neutrophils and macrophages play a

critical role in clearing debris and pathogens while secreting cytokines that promote tissue repair [3]. The tissue formation phase involves re-epithelialization, angiogenesis, and fibroblast-mediated extracellular matrix (ECM) deposition, which collectively lead to the restoration of tissue architecture [4]. Finally, during the remodeling phase, collagen and other matrix components undergo reorganization to enhance tissue strength and functionality.

Despite decades of research, chronic wounds such as diabetic foot ulcers and hypertrophic scars remain significant clinical challenges. These wounds often fail to

Received: September 6, 2024. Revised: April 27, 2025. Accepted: May 12, 2025

© The Author(s) 2025. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

progress through the normal healing phases, resulting in poor outcomes, including excessive scarring or impaired tissue regeneration. Traditional wound care therapies, while effective in some cases, have limited efficacy in addressing the underlying biological complexities that contribute to chronic wound pathophysiology. Therefore, understanding the cellular and molecular mechanisms that drive wound healing, and identifying the factors that disrupt this process in chronic wounds, is critical for developing more effective therapies [5].

Recent advances in single-cell technologies, particularly single-cell RNA sequencing (scRNA-seq), have revolutionized our understanding of wound healing by providing detailed insights into cellular heterogeneity and dynamic gene expression changes. scRNA-seq allows researchers to profile the transcriptomes of individual cells, enabling the identification of distinct cell populations and the discovery of novel cell types involved in wound healing [6]. For instance, fibroblasts, traditionally viewed as a homogeneous population of cells involved in ECM deposition, have been shown to comprise multiple subtypes with distinct roles in healing, including scar formation and tissue regeneration [7]. This level of detail has opened new avenues for exploring cellular plasticity and the interactions between different cell types during wound repair.

As part of the foundational analysis principle in scRNA-seq data analysis, machine learning, particularly deep learning techniques, has been instrumental in handling the large-scale and high-dimensional nature of single-cell datasets. Machine learning algorithms such as autoencoders, variational autoencoders (VAE), and graph-based models have been applied to scRNA-seq data to enhance processes such as dimensionality reduction, dropout imputation, and batch effect correction [8]. These techniques allow researchers to extract meaningful, low-dimensional representations of scRNA-seq data, enabling better visualization and interpretation of the cellular landscape in wound healing. Machine learning also plays a pivotal role in cell clustering and trajectory inference, which are critical for understanding the dynamic progression of cell states during wound healing. For example, algorithms such as Deep Embedding for Single-cell Clustering (DESC) [9] integrate batch correction with clustering, offering more accurate identification of cell types and subpopulations. In addition, trajectory inference methods like Monocle and Variational Inference for Trajectory by Autoencoder (VITAE) provide insights into the developmental pathways that cells undergo during healing, helping to map the transitions between cell states such as fibroblast-to-myofibroblast differentiation [10, 11]. These insights have significant implications for understanding how certain cells contribute to pathological outcomes, such as excessive scarring or impaired regeneration, in wounds.

This review aims to summarize the recent advances in the application of scRNA-seq and machine learning to wound healing research. We focus on three key areas: (1) the diversity and plasticity of the skin cell population involved in wound healing, (2) the key roles of fibroblast cells and immune cells in driving inflammation and tissue repair, and (3) the spatial organization and interaction of cell populations within the wound microenvironment. This review highlights how scRNA-seq and machine learning can provide new insights into the molecular mechanisms of wound healing, with the potential to inform the development of novel therapeutic strategies for chronic wounds and fibrosis. Through these efforts, we aim to underscore the transformative potential of

these tools in advancing precision medicine and regenerative therapies for wound healing.

Review

Cellular and molecular mechanisms of wound healing

Wound healing is a complex, multifaceted process that orchestrates the coordinated actions of diverse tissues and cell lineages. This intricate biological response aims to restore tissue integrity and function following injury, progressing through overlapping phases of clot formation, inflammation, new tissue formation, and remodeling [1, 12]. Recent advances in molecular biology, high-resolution imaging, and multi-omics approaches have significantly enhanced our understanding of this process, revealing new insights into the roles of various cell types and signaling pathways [5].

Temporal dynamics and molecular mechanisms of wound healing

The wound healing cascade is initiated by rapid hemostasis, where platelets adhere to exposed collagen and release a plethora of growth factors and cytokines. This triggers the coagulation cascade, culminating in the formation of a fibrin-rich clot [2]. The intrinsic pathway, activated when factor XII binds to negatively charged surfaces like collagen, plays a key role in this process [2]. Platelet degranulation releases crucial mediators such as transforming growth factor- β (TGF β), platelet-derived growth factor (PDGF), stromal cell-derived factor 1 (SDF1), vascular endothelial growth factor (VEGF), and endostatin [13]. Intriguingly, while platelets are important contributors, they are not absolutely essential for repair, highlighting the redundancy built into this vital process [14]. The inflammatory phase is characterized by the sequential influx of neutrophils and macrophages. These cells not only clear debris and pathogens but also secrete a range of mediators that modulate the subsequent repair process.

Re-epithelialization, a critical component of the tissue formation phase, involves complex changes in keratinocyte behavior. Studies in mouse models have revealed that the proliferative zone in the epidermis develops several hours post-injury and persists for days, with a spatial organization that varies depending on wound size [15–17]. The regulation of keratinocyte proliferation involves a sophisticated interplay of mechanical cues and growth factor signaling. For instance, the mechanosensitive receptor PIEZO1 plays a role in sensing reduced cell density, leading to Ca²⁺-dependent phosphorylation of extracellular signal-regulated kinase 1 (ERK1) and ERK2, resulting in cyclin B transcription necessary for mitosis [18]. During re-epithelialization, keratinocytes undergo significant changes in their adhesion molecule profile. They switch off hemidesmosomal $\alpha 6 \beta 4$ integrins and upregulate integrins that facilitate migration across the provisional wound matrix [4, 19, 20]. At the same time, these cells upregulate the production of proteases, especially matrix metalloproteinase 1 (MMP1), which likely aids in integrin–matrix adhesion by locally reshaping the extracellular matrix [21].

Cellular heterogeneity and functional diversity in wound repair

Recent single-cell analyses have unveiled an unprecedented level of cellular heterogeneity in the wound environment, particularly evident in fibroblast populations. Studies in mouse

models have revealed that fibroblasts from different dermal layers show varying contributions to wound healing and scarring [6, 22]. This research highlights the varying phenotypes of distinct spatial populations, and possibly lineages, of fibroblasts in the skin, each playing a unique role in wound healing through different mechanisms. These insights may help explain why adult wounds commonly result in scarring and rarely regenerate appendages. These variations likely stem from differences in developmental origins and molecular characteristics. For example, in dorsal back skin, papillary fibroblasts arise from fetal cells expressing LRIG1, while deeper skin layers are populated by fetal cells expressing DLK1 [6].

Recent investigations have identified new markers that help distinguish different fibroblast subtypes based on their roles in wound repair. One significant finding is that fibroblasts expressing engrailed 1 (EN1) are commonly recruited to postnatal wounds and are implicated in the formation of scars [23]. Remarkably, suppressing EN1 activity or introducing EN1-negative fibroblasts into wounds has been associated with reduced scarring and more regenerative healing outcomes [23, 24]. Another emerging area of interest is the role of adipocytes in the wound-healing process. Research suggests that adipocytes near the wound site may undergo substantial phenotypic shifts, possibly dedifferentiating and contributing to the pool of myofibroblasts [25]. It has been proposed that these adipocytes release lipids, which may serve as signals to attract inflammatory cells and influence their transformation into myofibroblasts. There is also evidence suggesting that some adipocytes may develop the ability to move and migrate toward the wound site [26]. However, this concept remains controversial, with conflicting evidence highlighting the need for further research.

Immunological orchestration of wound healing

The immune system, particularly macrophages, plays a pivotal role in coordinating the wound healing response. Macrophages exhibit remarkable functional plasticity, transitioning from pro-inflammatory to pro-resolving phenotypes as healing progresses [27]. During the initial phase of wound healing, macrophages predominantly exhibit a pro-inflammatory (M1) phenotype. This phenotype is driven by damage-associated molecular patterns and inflammatory mediators such as IFN γ and TNF. M1 macrophages are characterized by their cytotoxic and microbicidal functions, as well as their production of pro-inflammatory cytokines, including TNF, IL-1 β , and IL-6 [28]. As healing progresses, macrophages transition to an anti-inflammatory (M2) phenotype, triggered by signals such as IL-4 and IL-13 [29]. This phenotypic switch involves significant metabolic reprogramming, with early pro-inflammatory macrophages relying primarily on glycolysis, while later anti-inflammatory macrophages engage in oxidative phosphorylation [30]. The triggers for this macrophage phenotype transition are multifaceted. Efferocytosis of apoptotic neutrophils appears to be a key signal promoting the shift toward a pro-resolving phenotype [31]. This process not only clears cellular debris but also initiates the production of specialized pro-resolving mediators, such as lipoxins, which further modulate the inflammatory response [32].

Beyond the extensively studied of macrophages, other immune cells play significant roles in wound healing. Mast cells are present throughout the repair process, recruited to

injury sites by complement factors C3a and C5a produced by coagulation factor XII [33]. The nonredundant role of these immune cells in tissue repair remains a subject of debate. While some studies demonstrate that mast cell deficiency markedly impairs skin wound healing, others report only minor effects when mast cell numbers are reduced [34, 35]. Similarly, eosinophils are abundantly present throughout all stages of skin repair, yet their precise function has not been conclusively established, as knockout studies directly targeting eosinophils are lacking [36].

The $\gamma\delta$ T cells are abundantly present in the skin, where they settle during early embryonic stages and are crucial for both maintaining skin equilibrium and aiding in tissue repair [37]. In mouse models, these dendritic epidermal $\gamma\delta$ T cells rapidly alter their structure after injury, transforming from their typical sensory dendritic form to an activated, rounded shape within just a few hours of the wound [38]. These cells secrete growth factors, including FGF7, FGF10, and IGF1, which are instrumental in promoting keratinocyte proliferation and facilitating re-epithelialization during wound healing [39, 40]. Additionally, they produce FGF9, a factor that stimulates hair follicle neogenesis following injury [41]. Although human epidermal $\gamma\delta$ T cells exhibit distinct morphological and functional differences from their murine counterparts, evidence indicates that they similarly contribute to wound healing, primarily through the secretion of IGF1 [42]. In addition, regulatory T cells (Tregs) are present in the unwounded skin of both humans and mice [43]. Their suppressive functions may also modulate scarring, as studies in mice have shown that conditional deletion of GATA3 in Tregs leads to increased dermal fibrosis [44]. In contrast, skin-resident memory T cells (TRM) persist in the skin long term, primarily serving as a defense against viral infections [45]. Indeed, upon activation, TRMs condition the surrounding tissue by secreting inflammatory proteins and regulating gene expression, thereby creating a protective environment that enhances resistance to reinfection.

Human skin harbors both resident and recirculating B-cell populations [46]. While B cells are traditionally recognized for their role in infection prevention via antibody production, emerging evidence suggests that specific B-cell subsets also contribute to wound healing. For instance, in mice, the absence of B1 cells leads to increased neutrophil recruitment and neovascularization, resulting in delayed tissue repair [47]. In a similar manner, mice lacking B cells due to CD19 deficiency exhibit delayed wound healing, while the overexpression of CD19 in B cells has been shown to enhance and speed up the healing process [48].

Innate lymphoid cells (ILCs) are a subset of immune cells that, unlike B and T cells, lack antigen-specific receptors [49]. Among these, ILC2s have been identified as key contributors to skin wound healing. When IL-33-dependent activation of ILC2s is disrupted after skin injury, re-epithelialization is markedly hindered, highlighting their essential role in the tissue repair process [50].

Machine learning in single-cell data analysis

Machine learning general principles

Machine learning, a fundamental aspect of artificial intelligence, involves creating algorithms that allow computers to learn from data and make predictions or decisions (Figure 1). As a subset of AI, machine learning systems are designed to enhance their performance on specific tasks by learning

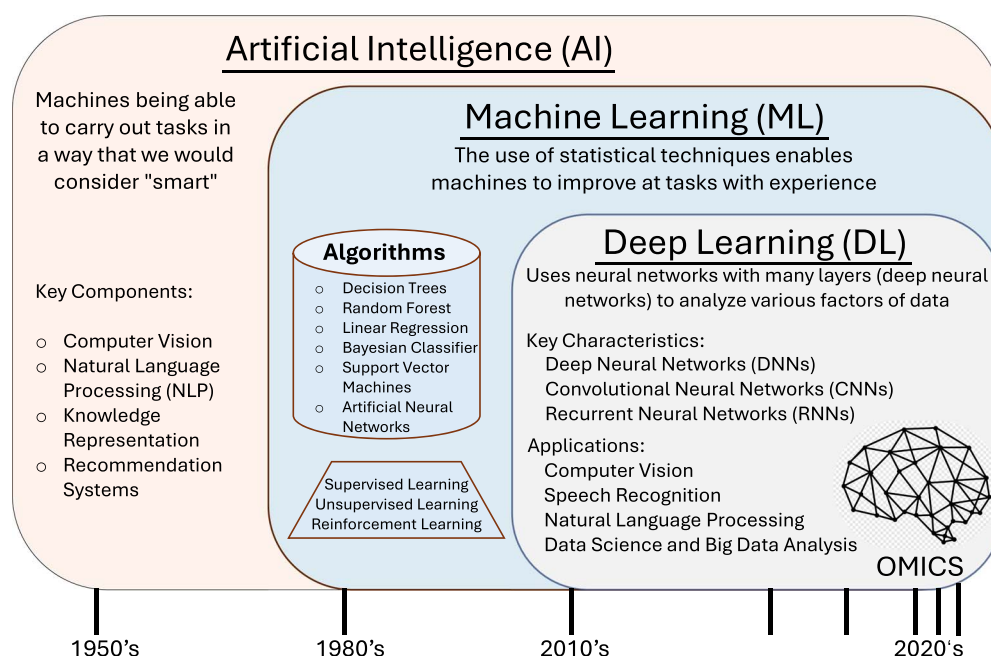


Figure 1. Hierarchical relationship between artificial intelligence (AI), machine learning (ML), and deep learning (DL). AI encompasses tasks requiring human-like intelligence, such as computer vision and natural language processing. ML, a subset of AI, uses statistical methods like decision trees and support vector machines to improve task performance through learning. DL, a further subset of ML, uses deep neural networks to handle complex data, with applications in fields like computer vision and speech recognition. The timeline illustrates the evolution from AI in the 1950s to the rise of DL in the 2010s

from experience, rather than relying on explicit instructions for every possible scenario [51]. This field encompasses several methods, including supervised learning, where models are trained using labeled data; unsupervised learning, which focuses on detecting patterns within unlabeled data; and reinforcement learning, where systems learn through feedback from their actions in an environment. Machine learning algorithms range from simple decision trees to complex support vector machines, each suited to different types of problems and data structures. Deep learning, a subset of machine learning, employs multilayered neural networks to handle and interpret intricate patterns within large datasets, making it particularly effective in tasks such as image and speech recognition [52]. This progression from broad AI concepts to specific machine learning techniques and further to deep learning represents an evolution toward increasingly sophisticated and capable systems for handling complex data analysis and decision-making tasks, especially in fields like OMICS research where the volume and complexity of data continue to grow exponentially. In general, machine learning (ML) algorithms are broadly categorized into two types, supervised learning and unsupervised learning, based on whether the model leverages labeled datasets with known outcomes [53].

The application of machine learning to scRNA-seq analysis follows a canonical working principle that underpins its effectiveness, as illustrated in Figure 2, which integrates a vertical process diagram. This workflow comprises four key steps (Figure 2a), data cleaning and preprocessing, feature extraction, model fitting, and evaluation, each contributing to robust analysis in scRNA analysis in wound healing [52, 54]. Data cleaning and preprocessing address noise and dropout events, leveraging methods like linear regression and SVM to normalize gene expression data, as seen in preprocessing stages of wound healing datasets [55]. Feature extraction

employs techniques such as dimensionality reduction (e.g. PCA, t-SNE, UMAP) to distill biologically relevant patterns, enhancing the identification of fibroblast subpopulations in human wounds [56]. Model fitting, using approaches like clustering (e.g. *k*-means) and neural networks, infers cellular dynamics, such as keratinocyte trajectories during repair [57, 58]. Evaluation validates model performance, ensuring reliability across diverse scRNA-seq datasets [58, 59]. These principles offer significant advantages: noise reduction improves data quality, feature extraction reveals hidden cellular heterogeneity, and model fitting enables dynamic process modeling, collectively advancing therapeutic insights into wound healing as demonstrated by Liu *et al.* [54].

Supervised learning—regression analysis. Supervised learning involves training models on labeled datasets where each input has a corresponding output, allowing the model to learn accurate input–output mappings for tasks like image classification and disease prediction. Regression analysis tasks aim to generate numerical predictions for continuous variables [60]. Linear regression, one of the most straightforward approaches, predicts the value of a dependent variable (*y*) by analyzing its relationship with a single independent variable (*x*) (Figure 2b). Alternatively, classification tasks involve assigning observations to predefined discrete categories, as seen in logistic regression (Figure 2b). In addition, support vector machines (SVMs) are prominent algorithms in classification tasks, designed to find a hyperplane that optimally separates classes in labeled data, facilitating the classification of unseen instances (Figure 2c) [61]. Other classification algorithms, including Naïve Bayes, decision trees, and random forests, similarly aim to assign new data to predefined categories [51].

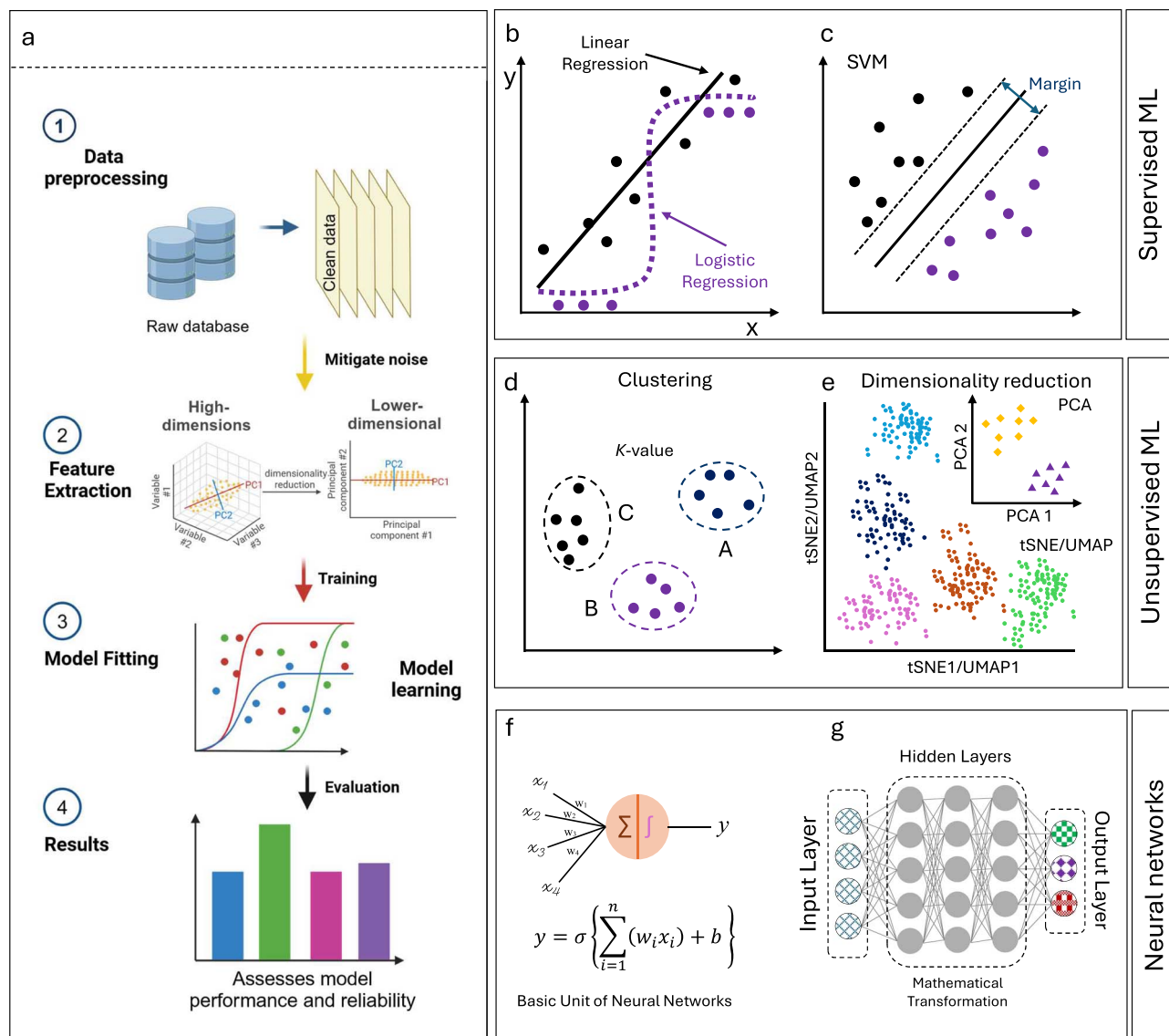


Figure 2. Illustration of machine learning principles and architectures in scRNA-seq analysis. The left workflow (a) outlines the canonical machine learning process: (1) data cleaning and preprocessing normalizes and cleans scRNA-seq data; (2) feature extraction distills high-dimensional features; (3) model fitting identifies patterns; (4) evaluation assesses performance. Right panels depict associated methods: Supervised machine learning (b, c) operates with labeled data, where each input is paired with a known ground truth. This approach includes regression tasks, which predict numeric outcomes based on input variables (e.g. linear regression) (b), and classification tasks, which categorize observations into predefined discrete classes (e.g. logistic regression (b) or support vector machines (SVMs)) (c). (b) Linear regression and (c) SVM support data preprocessing and classification, aiding noise reduction in wound healing datasets. Unsupervised machine learning (d, e) is exploratory, working with unlabeled data. Clustering techniques, such as k-means clustering (d), are employed to identify groups of similar cells—such as fibroblast subpopulations—based on shared features. Dimensionality reduction methods, including principal component analysis (PCA) and non-linear techniques like t-SNE or UMAP (e), are used to visualize these high-dimensional relationships in a lower-dimensional space. Deep learning tasks (f, g) involve artificial neural networks, where input variables are represented as nodes within an input layer. These nodes are connected to an output layer through multiple hidden layers, with various mathematical transformations applied along the way. (g) The function used in neural network. w_i denotes a learnable weight associated with the i -th input, while b represents a learnable bias term. The symbol σ refers to a nonlinear activation function, which operates on a single input and produces a single output. This activation function introduces nonlinearity into the model, enabling the network to capture complex patterns and relationships in the data. (g) Basic design of a multilayer perceptron consists of nodes. Neural networks fit complex models, e.g. cell trajectories in wound healing

Unsupervised learning—clustering and dimensionality reduction analysis. Unsupervised learning works with unlabeled data, focusing on discovering hidden patterns or structures through techniques like clustering (e.g. k -means) or dimensionality reduction (e.g. principal component analysis, PCA), commonly applied in customer segmentation or anomaly detection [62]. Clustering represents the most prevalent form of unsupervised learning, aiming to identify similarities and differences among observations, which are subsequently

categorized into distinct groups (Figure 2d) [63]. In contrast to supervised learning, these groups are not predefined by labeled data but emerge based on the specific context of the ML task and the assumptions embedded in the ML algorithm, often referred to as hyperparameters [64]. A notable example is the k -value in the commonly employed k -means clustering method. The k -value dictates the number of clusters produced, with similar data points grouped together within the same cluster. In addition to clustering, dimensionality

reduction algorithms represent some of the most commonly applied methods in unsupervised learning (Figure 2e). These techniques aim to condense datasets with a large number of variables into a more manageable set, while preserving the intrinsic relationships within the data. Notable examples include PCA, t-distributed stochastic neighbor embedding (t-SNE), and uniform manifold approximation and projection (UMAP) [65–67]. These methods are frequently employed to generate interpretable two-dimensional visualizations of high-dimensional data especially in scRNA data analysis. In wound healing research, unsupervised learning offers a valuable approach for exploring the heterogeneity of skin cell types, enabling the identification and characterization of novel subgroups within complex datasets.

Artificial neural networks—deep learning. Alternatively, machine learning techniques can be categorized based on their reliance on neural networks. ‘Traditional machine learning’ methods operate without the use of neural networks, whereas deep learning utilizes artificial neural networks [68]. As a subset of machine learning, deep learning leverages deep neural networks to process complex data and is applicable in both supervised and unsupervised learning scenarios. The foundation of neural networks lies in the concept of artificial neurons, which are the fundamental building blocks for all neural network models. At its core, an artificial neuron is a mathematical function designed to transform inputs into outputs in a defined manner. The process can be broken down as follows: (1) Input Reception: The artificial neuron receives multiple input values. (2) Mathematical Transformation: It applies a specified mathematical function to these inputs. (3) Output Generation: Based on the transformation, the neuron produces an output value (Figure 2f, g).

Neural network architectures represent diverse strategies for organizing artificial neurons, each designed to address specific types of problems or data structures. These architectures include, but are not limited to, multilayer perceptrons (MLPs), recurrent neural networks (RNNs), convolutional neural networks (CNNs), graph convolutional networks (GCNs), and autoencoders. Each architecture has distinct characteristics: MLPs are versatile and suited for general-purpose tasks; CNNs excel in processing grid-like data, particularly images; RNNs are optimized for sequential data analysis; GCNs specialize in processing graph-structured data; and autoencoders are adept at unsupervised learning and dimensionality reduction [69]. Importantly, these architectures are not mutually exclusive and can be combined to leverage their respective strengths. For instance, CNNs often incorporate fully connected layers, akin to those in MLPs, to produce final classification outputs. This hybridization of architectures allows for the creation of sophisticated models capable of handling complex, multifaceted tasks.

The diverse neural network architectures have significantly advanced biological research by tackling complex datasets and predictive challenges. MLPs, with their fully connected layers, have been effectively applied to tasks like protein secondary structure prediction, leveraging their capacity to model intricate relationships from fixed feature sets [70]. RNNs have excelled in analyzing sequential biological data, such as protein sequences for engineering novel peptides, capturing temporal or positional dependencies critical to understanding molecular functions [71]. CNNs have proven powerful in processing structured biological inputs, notably

in predicting protein residue–residue contacts from sequence co-evolution data, enabling high-accuracy 3D structure modeling [72]. GCNs have been utilized to interpret molecular structures, predicting drug properties by modeling atom–bond relationships as graphs, which is particularly valuable for understanding interaction networks [73]. Autoencoders, meanwhile, have facilitated unsupervised analysis in single-cell RNA sequencing, compressing high-dimensional transcriptomic data into meaningful latent representations to uncover cellular patterns [74]. These applications, thriving on large datasets, demonstrate deep learning’s predictive prowess in biology, yet their opaque nature often complicates mechanistic insights, requiring careful validation. The field of neural network architecture design continues to evolve, with ongoing research exploring novel configurations and combinations to enhance performance across various domains of artificial intelligence and machine learning. For a more comprehensive exploration of these architectures and their applications, Greener *et al.* provides in-depth analysis and examples [64].

Leveraging machine learning in scRNA-seq data analysis

Single-cell transcriptome sequencing, a pivotal technique in single-cell omics research, remains the most widely employed method in the field [75]. In recent years, machine learning has become integral to analyzing scRNA-seq data. Machine learning algorithms are commonly used for tasks such as denoising, dropout imputation, doublet removal, and batch effect correction. Moreover, they are increasingly applied to cell cycle variance annotation, cell clustering, dimensionality reduction, and trajectory inference. In this review, we specifically focus on the application of machine learning in dimensionality reduction, cell clustering/annotation, and trajectory analysis as these areas are central to our investigation into wound healing. For a more comprehensive overview of other machine learning applications in scRNA-seq data analysis, please refer to Brendel *et al.*, Li *et al.*, Madadi *et al.*, Sun *et al.*, and Hughes *et al.* [62, 76–79].

Dimensionality reduction. scRNA-seq yields a high-dimensional dataset that poses challenges for analysis. Dimensionality reduction is crucial for extracting meaningful, low-dimensional representations from scRNA-seq data, facilitating processes such as cell clustering, annotation, and trajectory inference [80].

Early methods of dimensionality reduction in scRNA-seq data focused on identifying highly variable genes [2]. PCA has been widely adopted for this task, particularly in software like Seurat [80]. However, PCA is limited in capturing nonlinear patterns, necessitating advanced techniques like deep learning [81]. Deep learning approaches, such as autoencoders, have shown success in learning nonlinear, informative patterns from scRNA-seq data. For instance, scScope uses autoencoders to improve low-dimensional representations while correcting dropout events [8]. Similarly, variational autoencoders (VAEs), which disentangle latent representations, have demonstrated notable results. VASC, an early VAE-based method, showed improved performance in identifying rare cell populations [82]. Other VAE-based approaches like scVI [83] and scDHA [84] have successfully addressed batch effects and overfitting challenges. Modifications to the prior distribution, such as Gaussian mixture models in scVAE [85] and the von Mises–Fisher distribution

in scSphere [86], further enhance latent representations. These innovations reduce crowding in latent space and improve the temporal resolution of data. Recent models also focus on improving disentanglement by altering loss functions, as seen in DiffVAE [87] and MMD-VAE [88], and by introducing adversarial components, like DR-A and scRAE [89, 90].

Moreover, graph-based autoencoders (GAEs) have also been applied, modeling both gene expression profiles and topological relationships between cells [91]. Techniques such as Graph-DiffVAE and scGAE reconstruct cell graphs based on gene expression, offering more robust low-dimensional representations. Interpretability of deep learning models remains a concern, particularly in biologically meaningful representations. Methods such as sparsely connected autoencoders (SCA) integrate prior biological knowledge into the model structure, enhancing interpretability [92]. Additionally, domain knowledge is embedded into models via Gene Ontology-based approaches, as in GOAE [93], pmVAE [94], and VEGA [95], allowing the models to focus on biologically relevant pathways.

Cell clustering and annotation. A critical goal in scRNA-seq analysis is to cluster heterogeneous cells into subpopulations, representing distinct cell types or states. Established tools like Seurat and Scanpy [80, 96] employ graph-based clustering methods such as Louvain [97] and Leiden [98]. Traditional clustering approaches, such as *k*-means, rely on low-dimensional representations derived from techniques like PCA or autoencoders [99]. Advanced methods like DESC [9] use deep learning for simultaneous batch correction, denoising, and clustering. CarDEC [100] builds upon this by treating highly variable and lowly variable genes differently to improve clustering accuracy. Hybrid architectures, such as scziDesk [101] and GraphSCC [102], combine graph-based methods with deep learning to improve clustering performance. Model interpretability and clustering stability are critical, and recent approaches, such as SAME-clustering [103] and the CellBIC algorithm [104], integrate multiple clustering techniques to enhance results.

Trajectory analysis. Beyond clustering, trajectory inference provides insights into dynamic cellular processes such as differentiation and development. Trajectory inference methods aim to reconstruct the developmental pathways of cells based on gene expression data. Tools such as Monocle [10], CellTree [105], and scTDA [106] have been developed for this purpose. Deep learning has furthered this field, with VITAE providing a probabilistic approach to trajectory inference [11]. By integrating multiple datasets and correcting for confounders, VITAE allows for more accurate trajectory reconstruction in complex biological systems.

Many researchers have also combined cell clustering with trajectory inference to better understand cell types and their dynamic processes. Machine learning models, such as SVM classifiers, have demonstrated strong performance in cell clustering across diverse datasets [107]. Recent developments like DESC have addressed both clustering and batch effect removal, improving clustering outcomes while preserving biological variance [9]. Moreover, trajectory inference methods such as VITAE provide new opportunities to study dynamic cellular processes such as cell differentiation and activation.

Machine learning and deep learning approaches have significantly enhanced dimensionality reduction, cell clustering, and

trajectory inference in scRNA-seq data analysis. These methods not only improve clustering performance but also facilitate a deeper understanding of cellular dynamics, enabling advances in fields such as skin wound healing.

Transformer models in single-cell RNA-sequencing analysis

Transformer models originated in the field of Natural Language Processing (NLP), marked by the seminal paper “Attention Is All You Need” [108]. This architecture signaled a paradigm shift away from sequential processing models like recurrent neural networks (RNNs) and long short-term memory (LSTM) networks [109]. The core innovation of the Transformer is the self-attention mechanism [110]. Recognizing the conceptual parallels between the structure of language and biological sequences (such as DNA, RNA, and protein sequences), the bioinformatics community has increasingly adapted Transformer models for various biological data analysis tasks [111]. Notable applications include protein structure prediction, where attention mechanisms similar to those in Transformers play a key role in models like AlphaFold 2 [112], as well as genome sequence analysis [113] and, more recently, the analysis of gene expression data derived from scRNA-seq experiments.

Core mechanisms: self-attention and positional encoding in the scRNA-seq context. The cornerstone of the Transformer architecture is the self-attention mechanism [114]. In the context of scRNA-seq, self-attention allows the model to compute pairwise interaction scores between input elements. Depending on the specific model design, these elements could represent genes within a single cell or perhaps cells within a dataset. The mechanism calculates attention weights that determine how much ‘focus’ or importance each element should place on every other element when computing its updated representation [115]. This allows the model to learn context-dependent relationships; for instance, the significance of a particular gene’s expression level might depend on the expression levels of other functionally related genes within the same cell. Positional encoding is another key component in standard Transformers, designed to inject information about the order or position of elements within the input sequence, as the self-attention mechanism itself is permutation invariant [116]. In NLP, word order is crucial for meaning. However, the concept of ‘order’ for genes within a cell’s expression profile is less inherent. While genes exist at specific genomic locations, their expression levels in scRNA-seq data are typically presented as a high-dimensional vector or matrix without a universally meaningful intrinsic order. Consequently, the direct application of standard positional encoding might be less relevant or require modification. Some approaches might impose an artificial order, e.g. by ranking genes based on expression levels [117], or use relative positional encoding schemes [118].

A critical challenge lies in transforming the sparse, high-dimensional cell-by-gene expression matrix into a format suitable for Transformer inputs. The sheer number of genes (often exceeding 20 000) poses a significant hurdle for standard Transformer architectures due to the quadratic complexity of self-attention [119]. Various strategies such as Input Tokenization and Embedding and Handling Sparsity are employed to address this. For example, genes are often treated as analogous to ‘tokens’ in NLP [117]. Instead of simple word

embeddings, models might utilize learned gene embeddings that aim to capture functional similarities or relationships between genes [120]. On the other hand, scRNA-seq data are characterized by high sparsity, with a large proportion of zero counts (often >90%) [121]. These zeros can represent true biological absence of transcripts or technical artifacts known as ‘dropout’ events. Pre-training strategies often involve masking certain gene expression values (e.g. randomly masking a percentage of non-zero values) and training the model to predict or reconstruct them, forcing it to learn contextual gene relationships [119]. Therefore, developing effective input representation strategies is a central and ongoing area of research for applying Transformers to scRNA-seq data.

Applications of transformer models across the scRNA-seq analysis pipeline. Transformer-based models are being explored and applied across a wide spectrum of tasks within the standard scRNA-seq analysis workflow, demonstrating the architecture’s versatility. A significant trend is the development of large, pre-trained ‘foundation models’ that learn general representations from massive datasets, which can then be adapted (fine-tuned) for specific downstream applications [119]. This approach mirrors the success of large language models (LLMs) in NLP and aims to leverage the vast amount of publicly available scRNA-seq data.

Cell type annotation and classification. Identifying and labeling cell types is a foundational step in interpreting scRNA-seq data. Transformers are employed to automate or assist this process, often achieving high accuracy. Models like scBERT [122], scGPT [123], scTransSort [124], xTrimGene [119], CellPLM [125], and TOSICA [22] have been developed specifically for this task. Many utilize a pre-training/fine-tuning strategy: the model first learns general gene expression patterns from large unlabeled datasets and is then fine-tuned on smaller labeled datasets to predict specific cell types. These models often capture complex gene signatures more effectively than methods relying solely on canonical marker genes. Some models, like TOSICA, are designed with interpretability in mind, attempting to link predictions to known biological entities or pathways [22]. Intriguingly, even general-purpose LLMs like GPT-4 have shown promise in annotating cell types based on provided marker gene lists, though evaluation remains challenging without a perfect gold standard [126].

Gene expression imputation. The inherent sparsity and dropout events in scRNA-seq data can hinder downstream analyses [121]. Imputation methods aim to predict these missing or zero values, creating a more complete expression matrix. Transformer models leverage their ability to understand context (i.e. the expression of other genes in the cell) to perform imputation. Models like scBERT and scGPT incorporate imputation capabilities, often linked to their pre-training objective of reconstructing masked gene expression values. GexBERT is another example focused on modeling gene co-expression for tasks including imputation [127]. While not always strictly Transformers, related deep learning architectures employing attention mechanisms or diffusion models (which share conceptual links) like scIDPMs are also used for this purpose, aiming to recover biologically meaningful expression values [128].

Batch effect correction and data integration. Combining scRNA-seq datasets from different experiments, laboratories, technologies, or time points is often necessary but complicated by technical variations known as batch effects [129]. Effective integration requires removing these technical artifacts while preserving true biological variation. Transformers are used to learn integrated data representations that align cells across batches or even different data modalities. For example, scGPT includes batch correction capabilities [123]. TOSICA achieves batch-insensitive integration implicitly through its attention mechanism guided by prior biological knowledge [22]. SELF-Former specifically tackles the integration of scRNA-seq with spatial transcriptomics data [130]. ScRAT demonstrates cross-species integration and batch effect removal in plant data [131]. SCARE, using a Mamba architecture (an efficient alternative to Transformers), integrates scRNA-seq and scATAC-seq data [132]. These models aim to create a unified latent space where biological similarities are emphasized over technical differences.

Trajectory inference and pseudotime analysis. Understanding dynamic biological processes like cell differentiation or development often involves ordering cells along a continuous path, known as trajectory inference or pseudotime analysis [133]. While dedicated Transformer models for end-to-end trajectory inference are less commonly highlighted in the provided sources compared to other tasks, the underlying capabilities of Transformers are highly relevant. By learning meaningful, low-dimensional representations (embeddings) of cells that capture subtle state transitions [76], Transformers can provide improved input for existing trajectory inference algorithms (e.g. Slingshot, Monocle) [134]. The ability to model dependencies and capture cellular states suggests potential for more direct application in future models.

Performing machine learning on scRNA data in wound healing

Wound healing is a complex biological process involving intricate interactions between various cell types and molecular pathways. Recent advancements in single-cell technologies and machine learning algorithms have revolutionized our understanding of this process, offering unprecedented insights into cellular heterogeneity, dynamic gene expression, and intercellular communication during tissue repair. In order to fully present how machine learning is applied in single-cell analysis for wound healing research, we downloaded single-cell data on skin wound healing from the DISCO [135] database (GSM4673372) and completed the main process of single-cell analysis. The process demonstrates both supervised and unsupervised machine learning methods, as shown in Figure 3. Essentially, we presented applications in various aspects, including data preparation, clustering, cell type annotation, dimensionality reduction, gene model scoring, gene set analysis, gene ontology enrichment analysis, pseudotime trajectory analysis, and cell-to-cell communication. Our analysis process fully demonstrates the extensive application of various machine learning methods in wound healing research. Subsequently, by reviewing other literature, this section synthesizes recent findings that leverage these cutting-edge approaches to elucidate the molecular mechanisms underlying wound healing, with a focus on integrating scRNA-seq (Figure 4). Figure 4 had been designed to illustrate the relationship between specific machine learning models

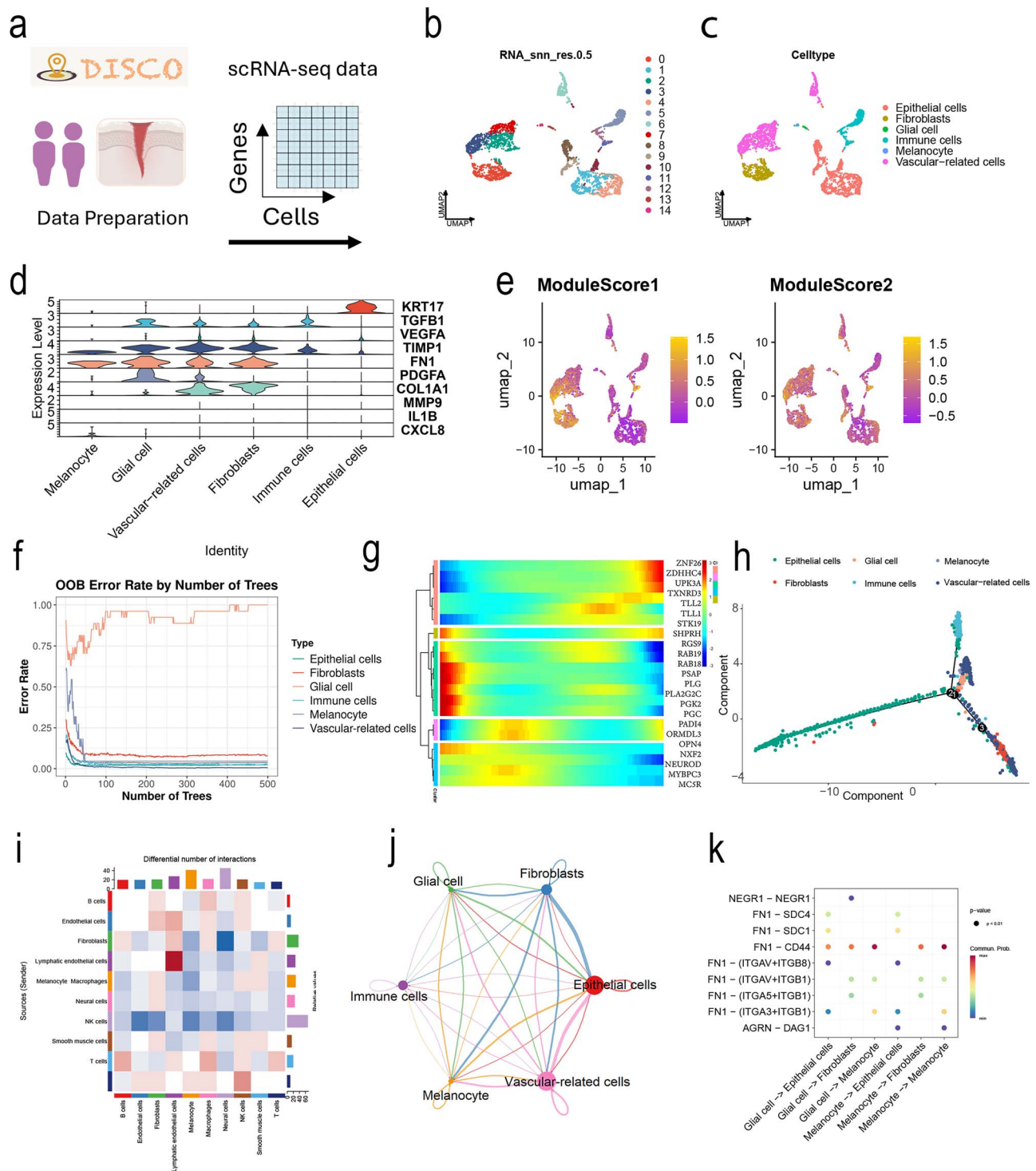


Figure 3. Workflow of single-cell RNA-sequencing analysis with machine learning and deep learning integration in the wound healing dataset. **(a)** Data preparation: The raw scRNA-seq data undergoes preprocessing, where genes and cells are organized into a matrix. This step includes quality control, filtering, and normalization of the data. Various deep learning methods, such as neural networks, may be employed here to impute missing values and correct batch effects. **(b, c)** Clustering analysis: Unsupervised machine learning techniques, such as principal component analysis (PCA) and uniform manifold approximation and projection (UMAP), are used to reduce dimensionality and visualize the data. Cells are then clustered into distinct groups using algorithms like *k*-means or graph-based clustering (Louvain/Leiden). These clusters reveal the inherent cellular heterogeneity in the tissue sample. **(d, e)** Gene features and module scoring: Key gene features and modules are identified using statistical techniques and machine learning-based feature selection methods. Module scoring, often based on deep learning models, quantifies the expression of gene sets in individual cells. Feature importance can be assessed using models like Random Forest or gradient boosting. **(f)** Training a classification model: A supervised machine learning classifier, such as Random Forest, is trained to categorize cells into predefined types (e.g. epithelial cells, immune cells, fibroblasts) based on their gene expression profiles. Cross-validation and out-of-bag (OOB) error rates are used to optimize the model's performance and ensure robust cell-type classification. **(g, h)** Trajectory analysis: Pseudotime and trajectory analysis are performed using machine learning methods, such as Monocle. These methods infer the developmental progression of cells and visualize transitions between cellular states in a low-dimensional space, predicting cell fate changes over time. **(i, j, k)** Cell-to-cell communication: Machine learning techniques, such as the CellChat model, were applied to infer cell-cell communication networks. These models integrate receptor-ligand interaction data and infer how different cell types communicate in the tissue microenvironment, identifying key signaling pathways and intercellular communication dynamics.

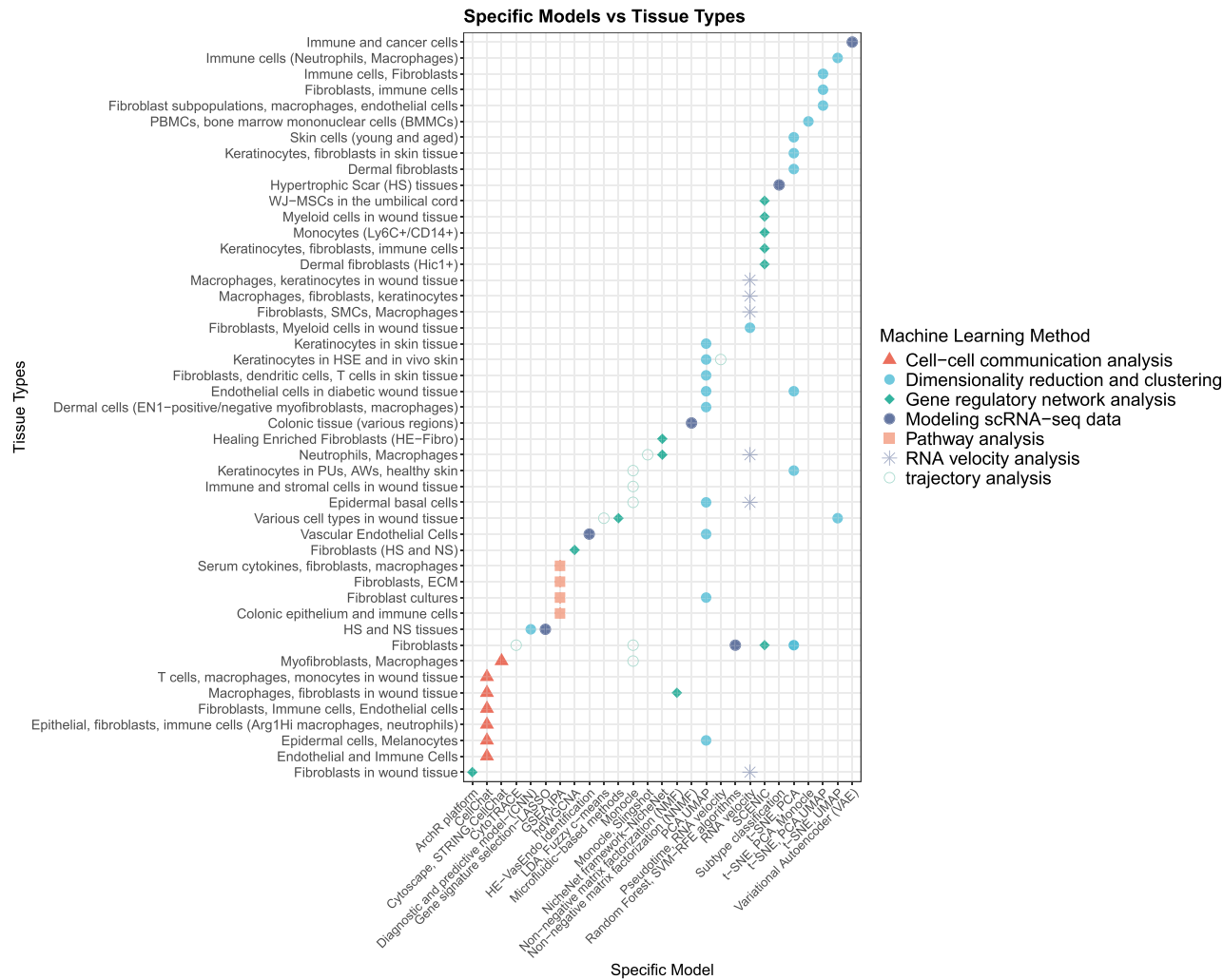


Figure 4. The relationship between specific machine learning models and applicable cell or tissue types on scRNA in wound healing. This figure was developed by synthesizing findings from studies reviewed in section on Performing machine learning on scRNA data in wound healing, mapping each specific machine learning model to the cell types it was applied to in the original research. Each point represents a particular combination of a machine learning model applied to a specific tissue type. Machine learning methods: Represented by different shapes, these include a variety of methods applied for cell–cell communication analysis, dimensionality reduction, gene regulatory network analysis, and modeling of scRNA-seq data. Specific models: The X-axis lists the specific models used in each method, such as *CellChat*, *PCA*, *t-SNE*, and *SCENIC*. Applicable cell or tissue types: The Y-axis indicates the cell types or tissues where the respective models were applied. These include endothelial cells, fibroblasts, macrophages, keratinocytes, and various immune cells involved in wound healing processes

and the cell or tissue types they were applied to in wound healing research, drawing from our synthesis of literature in this Section.

To provide a more comprehensive exploration of machine learning applications in wound healing research, we included detailed descriptions of the methods employed in key studies cited within this review. For instance, Guerrero-Juarez *et al.* utilized pseudotime trajectory analysis and RNA velocity to uncover fibroblast-to-myofibroblast differentiation trajectories in regenerating mouse skin wounds. Specifically, they implemented the Monocle3 algorithm, which constructs a trajectory by ordering cells along a pseudotime axis based on gene expression changes, using a minimum spanning tree approach optimized via reversed graph embedding [7]. This method allowed the identification of three distinct differentiation paths, revealing dynamic transcriptional regulation during repair. Similarly, Theodoridis *et al.* applied the CellChat algorithm to infer cell–cell communication

networks in diabetic foot ulcers, leveraging a database of ligand–receptor interactions to quantify signaling strength between fibroblast subpopulations and immune cells [5]. CellChat employs a mass-action model to estimate interaction probabilities, incorporating normalization steps to account for cell type abundance, which provides insights into the role of Healing Enriched Fibroblasts (HE-Fibro) in promoting repair, as well as in the investigation of how wound fibroblasts engage in dynamic signaling crosstalk with immune cells [56]. Additionally, dimensionality reduction techniques like t-SNE (t-distributed Stochastic Neighbor Embedding) were widely used across studies (e.g. Haensel *et al.*) to visualize high-dimensional scRNA-seq data in two dimensions, relying on a probabilistic approach to preserve local data structure while minimizing Kullback–Leibler divergence [136]. These detailed implementations highlight how machine learning enhances the resolution of cellular dynamics in wound healing, ensuring a consistent depth of discussion throughout this section.

Fibroblast diversity and plasticity

Single-cell RNA sequencing has revealed remarkable heterogeneity within fibroblast populations during wound healing. Guerrero-Juarez *et al.* identified 12 distinct fibroblast subclusters with unique gene expression signatures in regenerating mouse skin wounds [7]. Their study employed droplet-based scRNA-seq to profile 21 819 cells from regenerating mouse skin wounds at 12 days post-wounding. Using pseudotime trajectory analysis and RNA velocity, they uncovered three putative fibroblast-to-myofibroblast differentiation trajectories, highlighting the dynamic nature of these cells during repair [7, 137]. The scEpath method was employed to identify pseudotime-dependent gene expression changes, revealing dynamic regulation of transcription factors and signaling molecules during fibroblast differentiation [138]. Intriguingly, the study identified a rare population of myofibroblasts with myeloid features, suggesting potential cellular plasticity during wound healing. These findings were further validated using full-length scRNA-seq on flow-sorted cells at multiple time points. Foster *et al.* further delineated four key fibroblast subtypes—mechanofibrotic, activated responder, proliferator, and remodeling fibroblasts—each contributing distinct roles in tissue repair. By integrating scRNA-seq with scATAC-seq, they uncovered critical insights into chromatin accessibility and gene expression changes that drive fibroblast activation and differentiation [59]. The role of aging in fibroblast heterogeneity has also been explored. Mahmoudi *et al.* (2019) found that fibroblasts from older mice exhibit increased variability in reprogramming efficiency and wound healing rates, attributed to a shift in fibroblast composition with an increased proportion of ‘activated fibroblasts’ secreting inflammatory cytokines [59]. This variability was linked to distinct cytokine profiles in specific fibroblast subpopulations, suggesting potential targets for improving wound healing in aged individuals. In the context of diabetic foot ulcers (DFUs), Theocharidis *et al.* discovered a unique population of fibroblasts termed Healing Enriched Fibroblasts (HE-Fibro), characterized by the overexpression of matrix remodeling genes (e.g. MMP1, MMP3) and immune/inflammatory genes (e.g. CHI3L1, TNFAIP6) [139]. These HE-Fibro cells were predominantly found in the wound bed of healing ulcers and were shown to play a critical role in promoting wound repair.

Immune cell dynamics and inflammatory responses

The integration of scRNA-seq with machine learning has provided crucial insights into the role of immune cells in wound healing and fibrosis. Chen *et al.* revealed significant transcriptomic shifts in monocytes, particularly within the Ly6C+/CD14+ subpopulations, characterized by the induction of inflammatory pathways and suppression of steady-state immune features [140]. This approach identified two distinct trauma patient subtypes based on monocyte signatures, correlating with different clinical outcomes, including severity of organ dysfunction and infection rates. In chronic wounds, Yakupu *et al.* employed scRNA-seq and the CellChat algorithm to uncover the role of melanocytes in synthesizing and secreting Cathepsin G (CTSG), contributing to the chronic inflammatory environment in pressure ulcers [141]. The CTSG-F2RL1 ligand–receptor pairs mediated by melanocytes were significantly upregulated in pressure ulcers, suggesting a potential therapeutic target for managing chronic wounds. The impact of pathogens on wound healing has

also been elucidated through single-cell approaches. Celik *et al.* used scRNA-seq to uncover the detrimental effects of *Enterococcus faecalis* infection on wound healing, revealing immunosuppressive changes in macrophages and premature epithelial–mesenchymal transition in keratinocytes [142]. Their study generated a detailed single-cell atlas, showing that *E. faecalis* infection triggers significant immunosuppressive changes, particularly in macrophages, which were found to polarize toward an M2-like phenotype, suppressing the pro-inflammatory responses necessary for effective healing.

Epidermal dynamics and regeneration

Single-cell transcriptomics has provided valuable insights into the intricate behavior of epidermal cells during wound healing. In their 2020 study, Haensel *et al.* identified four distinct basal cell states in healthy skin, each with unique metabolic characteristics. These states undergo spatial reorganization during the re-epithelialization phase of wound repair [136]. Their analysis uncovered a quasi-linear differentiation hierarchy in which basal cells transition from a Col17a1Hi/Trp63Hi state to early-response (ER) states before either proliferating or becoming growth arrested prior to differentiation into spinous cells. Li *et al.* identified a specific subpopulation of keratinocytes expressing major histocompatibility complex class II (MHC II) in pressure ulcers, associated with poorer healing outcomes [143]. This study underscores the importance of keratinocyte heterogeneity in chronic wound healing and suggests potential targets for therapeutic intervention. The researchers applied various computational techniques, including t-distributed stochastic neighbor embedding (t-SNE) for dimension reduction and clustering, revealing significant differences in gene expression between fibrotic and regenerative myofibroblasts and macrophages. In the context of diabetic wound healing, Liao *et al.* used scRNA-seq to uncover significant alterations in gene expression profiles across eight distinct cell clusters within the skin tissue of patients with type 2 diabetes mellitus (T2DM). Keratinocytes from T2DM patients showed differential expression of 356 genes, implicating pathways like oxidative phosphorylation, cytokine–cytokine receptor interactions, and prion disease in the pathology of delayed wound healing [144].

Spatial organization and microenvironment in wound healing

The integration of spatial transcriptomics with scRNA-seq has provided a more comprehensive understanding of the spatial organization of gene expression patterns during wound healing. Parigi *et al.* utilized spatial transcriptomics to map the molecular regionalization of colonic tissue during mucosal healing, revealing distinct transcriptional signatures associated with different stages of tissue repair, immune cell activation, and remodeling. This unbiased technology enabled the identification of critical insights into the spatial organization of gene expression patterns during the healing process. Machine learning approaches, such as non-negative matrix factorization (NNMF), were applied to deconvolve the spatial transcriptomics data, revealing critical insights into the spatial organization of gene expression patterns during the healing process [145]. Yang *et al.* identified a distinct ‘pro-regenerative’ immune niche within large wound centers, characterized by the presence of T cells, pro-inflammatory macrophages, and monocytes co-localizing with hair follicle stem cells and papillary fibroblasts. This spatial organization

was found to promote tissue regeneration through specific signaling pathways like IL-6 and ACTIVIN. The study employed CellChat for cell–cell communication analysis, revealing that these signaling pathways are preferentially active in regenerative regions [146].

The application of scRNA-seq coupled with machine learning techniques such as RNA velocity analysis has provided significant insights into the cellular heterogeneity and spatial organization of fibroblasts in wound healing. This study demonstrated that upper wound fibroblasts, characterized by the expression of *Crabp1*, have a distinct differentiation trajectory toward dermal papilla formation, which is crucial for hair follicle regeneration in large wounds. By employing RNA velocity, the research predicted the future states of fibroblasts, revealing the regenerative competence of upper wound fibroblasts, particularly those located at the wound periphery. These findings underscore the critical role of integrating spatial transcriptomics with machine learning to decipher the complex cellular behaviors that underpin regenerative wound healing, offering potential strategies for enhancing tissue repair [147]. The study by Mascharak *et al.* provides a comprehensive multi-omics analysis to delineate the molecular pathways that drive scarring versus regenerative wound healing in the context of verteporfin-induced YAP inhibition. By integrating scRNA-Seq, proteomics, and extracellular matrix (ECM) ultrastructural analysis, the researchers identified distinct molecular trajectories associated with fibrotic and regenerative healing outcomes. The study utilized scRNA-seq to profile the transcriptomes of 2986 cells from mouse wounds across various time points, revealing divergent gene expression patterns between fibrotic and regenerative fibroblasts. Fibrotic wounds were characterized by the activation of mechanical signaling pathways and pro-fibrotic markers such as *Dpp4* and *Fap*, while regenerative wounds exhibited upregulation of Wnt signaling and regenerative markers like *Trps1* and *Rspo1*. The researchers applied Monocle3 for pseudotime trajectory analysis, which revealed bifurcating paths representing fibrotic and regenerative fates [58]. The regenerative trajectory was marked by decreased mechanotransduction and increased Wnt pathway activation, with RNA velocity and CytoTRACE analyses further supporting these findings. ECM ultrastructural analysis, quantified using a machine learning–based image processing algorithm, demonstrated that verteporfin-treated wounds restored ECM architecture to a state resembling unwounded skin, in contrast to the dense, aligned ECM fibers observed in fibrotic scars. Finally, the researchers used CellChat for cell–cell communication analysis, identifying key ligand–receptor interactions that differentiated regenerative from fibrotic healing, particularly in the modulation of fibroblast activity through Wnt and inflammatory signaling pathways. This integrative approach highlights the potential of combining multi-omics techniques with machine learning to uncover the molecular mechanisms underlying wound healing and fibrosis, offering insights into potential therapeutic targets for enhancing regenerative outcomes [58].

Gene regulatory network analysis and trajectory inference in wound healing

The integration of scRNA-seq and spatial transcriptomics with machine learning has provided profound insights into the heterogeneity and therapeutic potential of Wharton's jelly–derived mesenchymal stem cells (WJ-MSCs) in wound

healing. This study identified four key WJ-MSC subpopulations—proliferative, niche-supporting, metabolism-related, and biofunctional-type—each contributing uniquely to tissue repair. Utilizing the SCENIC pipeline, researchers reconstructed gene regulatory networks, identifying critical transcription factors that govern these cells' differentiation and specialized functions. Spatial transcriptomics further revealed that biofunctional-type MSCs, enriched in the fetal segment of the umbilical cord, possess superior wound healing capabilities. This comprehensive approach demonstrates the power of combining single-cell and spatial transcriptomics with machine learning to uncover the cellular and molecular underpinnings of effective regenerative therapies [55].

Guerrero-Juarez *et al.* employed Monocle 2, which uses reversed graph embedding to learn a parsimonious principal graph, to uncover fibroblast-to-myofibroblast differentiation trajectories [7]. Foster *et al.* used RNA velocity, a method that leverages spliced and unspliced mRNA counts to predict the future transcriptional state of individual cells, to predict the dynamic changes in fibroblast populations during wound repair [59]. More recently, deep learning–based trajectory inference methods have emerged, offering improved accuracy and scalability. For instance, scVelo, a stochastic model of transcriptional dynamics, has been applied to wound healing datasets to provide a more nuanced view of cell fate decisions during tissue repair [58]. These advanced methods allow researchers to capture the continuous nature of cellular transitions and identify key decision points in the wound healing process.

Cell–cell communication analysis

Machine learning algorithms have been pivotal in unraveling the complex intercellular communication networks in wound healing. CellChat, a tool that employs a mass-action model and pattern recognition approaches, has been widely used to predict and validate key cell–cell signaling pathways crucial for wound regeneration [148]. By employing non-negative matrix factorization (NMF) and pseudotime analysis, researchers identified multicellular 'movements' and conserved gene programs that span across various cell types, particularly macrophages and fibroblasts, revealing their coordinated roles in tissue repair [149]. The study also utilized the CellChat algorithm to predict and validate key cell–cell signaling pathways, such as the oncostatin M (OSM) pathway between macrophages and fibroblasts, which are critical for wound regeneration. Furthermore, by comparing these wound healing gene programs with those found in tumor microenvironments, the research illustrates how certain wound healing mechanisms are repurposed during cancer progression [150, 151]. This integrative approach highlights the power of combining single-cell transcriptomics, spatial analysis, and machine learning to uncover the complex cellular interactions that drive both wound healing and tumor growth, offering novel insights for therapeutic development [149].

The study conducted by Theodoridis *et al.* integrates skin transcriptomics and serum multiplex assays to uncover novel mechanisms involved in wound healing of diabetic foot ulcers (DFUs). The research combined bulk RNA sequencing of forearm skin biopsies with scRNA-seq of foot skin specimens from diabetic patients, both with and without DFUs, and control subjects. The study identified significant differences in the inflammatory profiles and cellular compositions between healers and non-healers. Single-cell RNA

sequencing revealed the presence of various fibroblast subpopulations, macrophages, and endothelial cells, highlighting the distinct cellular environments in DFUs compared to non-diabetic skin. The study utilized advanced computational tools, such as t-SNE and UMAP for dimensionality reduction, and Ingenuity Pathway Analysis (IPA) for functional analyses. The scRNA-seq data uncovered fibroblast clusters specific to DFUs, characterized by markers like COL7A1, indicating their role in impaired wound healing. Furthermore, the study applied canonical correlation analysis (CCA) to compare cellular states across different patient conditions, identifying distinct fibroblast subtypes that are primed for a wounding phenotype in diabetic skin. The integration of serum cytokine profiles with transcriptomic data using machine learning methods provided insights into systemic factors influencing wound healing outcomes. Markers such as soluble vascular cell adhesion molecule, vascular endothelial growth factor (VEGF), and interferon-gamma (IFN- γ) were found to be associated with better healing outcomes. The application of IPA further revealed dysregulated pathways in non-healers, including the inhibition of IL-13 and IFN- γ , suggesting potential therapeutic targets to enhance DFU healing [57]. Recent advancements in deep learning have led to the development of more sophisticated cell-cell communication inference tools. For example, Tensor-cell and CellPhoneDB, the deep learning framework to model multiway interactions, has been applied to wound healing datasets to capture higher-order interactions between multiple cell types simultaneously [152, 153]. These advanced methods provide a more comprehensive view of the complex signaling networks that orchestrate the wound healing response.

Predictive modeling and biomarker identification

Machine learning approaches have been instrumental in developing predictive models for wound healing outcomes and identifying potential biomarkers. Song *et al.* employed machine learning algorithms, including Random Forest and Support Vector Machine-Recursive Feature Elimination (SVM-RFE), to screen for significant glycosphingolipid metabolism-related genes (GSLMRGs) in keloid development [154]. The research involved the analysis of keloid tissues and normal skin samples to identify key GSLMRGs and their role in fibroblast differentiation and immune cell infiltration. Machine learning algorithms (SVM-RFE) were employed to screen for significant GSLMRGs from a set of 46 genes. Six key genes (ARSA, GBA2, SUMF2, GLTP, GALT, HEXB) were identified as critical markers for keloid diagnosis and prognosis. A diagnostic model based on these genes was constructed using logistic regression, achieving high accuracy as demonstrated by a ROC curve with an area under the curve value of 0.945. The study further applied scRNA-seq to explore the heterogeneity of fibroblasts in keloid tissues. Pseudotime trajectory analysis using the Monocle2 algorithm revealed that increased activity of the GSL metabolism pathway was associated with fibroblast differentiation, particularly in the transition toward myofibroblasts and stromal fibroblasts, which are crucial in keloid formation. The research also included cell-cell communication analysis using the CellChat package, highlighting the interactions between fibroblasts with different levels of GSL metabolism activity and other cell types, including immune cells and endothelial cells. These interactions were mediated through specific ligand-receptor pairs, suggesting that the GSL metabolism

pathway plays a vital role in regulating the cellular microenvironment in keloids. This integrative approach, combining single-cell transcriptomics with machine learning, provides new insights into the pathogenesis of keloids and offers potential targets for therapeutic intervention by modulating the GSL metabolism pathway [154].

In the context of hypertrophic scars (HS), which is a significant challenge in wound healing, these are characterized by excessive extracellular matrix deposition due to fibroblast activation. Recent studies have integrated scRNA-seq with machine learning to unravel the complex molecular mechanisms underpinning HS formation. Using the hdWGCNA method, specific gene modules associated with fibroblasts were identified in HS, followed by non-negative matrix factorization to classify HS subtypes. Various machine learning algorithms, including LASSO and a convolutional neural network (CNN), were employed to pinpoint key gene signatures—THBS2, CTSK, and CTHRC1—that are pivotal in diagnosing and predicting HS. Functional experiments further revealed that THBS2 promotes fibroblast proliferation and migration through the TGF β 1/P-Smad2/3 signaling pathway. These findings underscore the potential of THBS2 as a biomarker and therapeutic target in HS, offering new avenues for improving wound healing outcomes [155].

Deep learning aiding bioinformatics analysis for scRNA-seq in wound healing

scRNA-seq analysis in wound healing and skin repair harnesses major deep learning categories—convolutional neural networks (CNNs), recurrent neural networks (RNNs), graph convolutional networks (GCNs), and autoencoders [52, 64, 76] (Figure 5). CNNs excel in processing spatially resolved gene expression data, enabling the mapping of immune niches and wound bed architecture, as seen in spatial transcriptomics studies of human wounds [59, 147]. RNNs model temporal dependencies, inferring cell state transitions like keratinocyte differentiation trajectories across healing phases, as explored by Liu *et al.* and Guerrero-Juarez *et al.* [7, 54]. GCNs capture cell-cell interactions by representing cellular networks as graphs, elucidating signaling among fibroblasts, immune cells, and keratinocytes during repair, as demonstrated in Theodoridis *et al.* and Liu *et al.* [5, 54]. Autoencoders integrate high-dimensional data and extract features, identifying key cell populations such as fibroblast subtypes, with applications in batch correction and heterogeneity analysis across wound stages [54, 55, 76, 156]. Commonly used tools include scVI (Autoencoders) for data integration [157], scVelo (RNNs) for RNA velocity [158], CellChat (potential GCN adaptations) for communication analysis, and CNN implementations in spatial transcriptomics frameworks. These deep learning approaches enhance bioinformatics analysis, providing predictive and mechanistic insights into wound healing and skin repair processes.

Biological interpretation of machine learning outcomes

The machine learning results derived from scRNA-seq analysis had provided substantial biological insights into the intricate dynamics of wound healing, particularly through trajectory inference. Utilizing scVelo [137, 158], the analysis delineated developmental trajectories that included keratinocyte migration during re-epithelialization and fibroblast differentiation into regenerative subtypes, with a notable focus on the Crabp1+ upper wound fibroblasts identified by Phan *et al.*

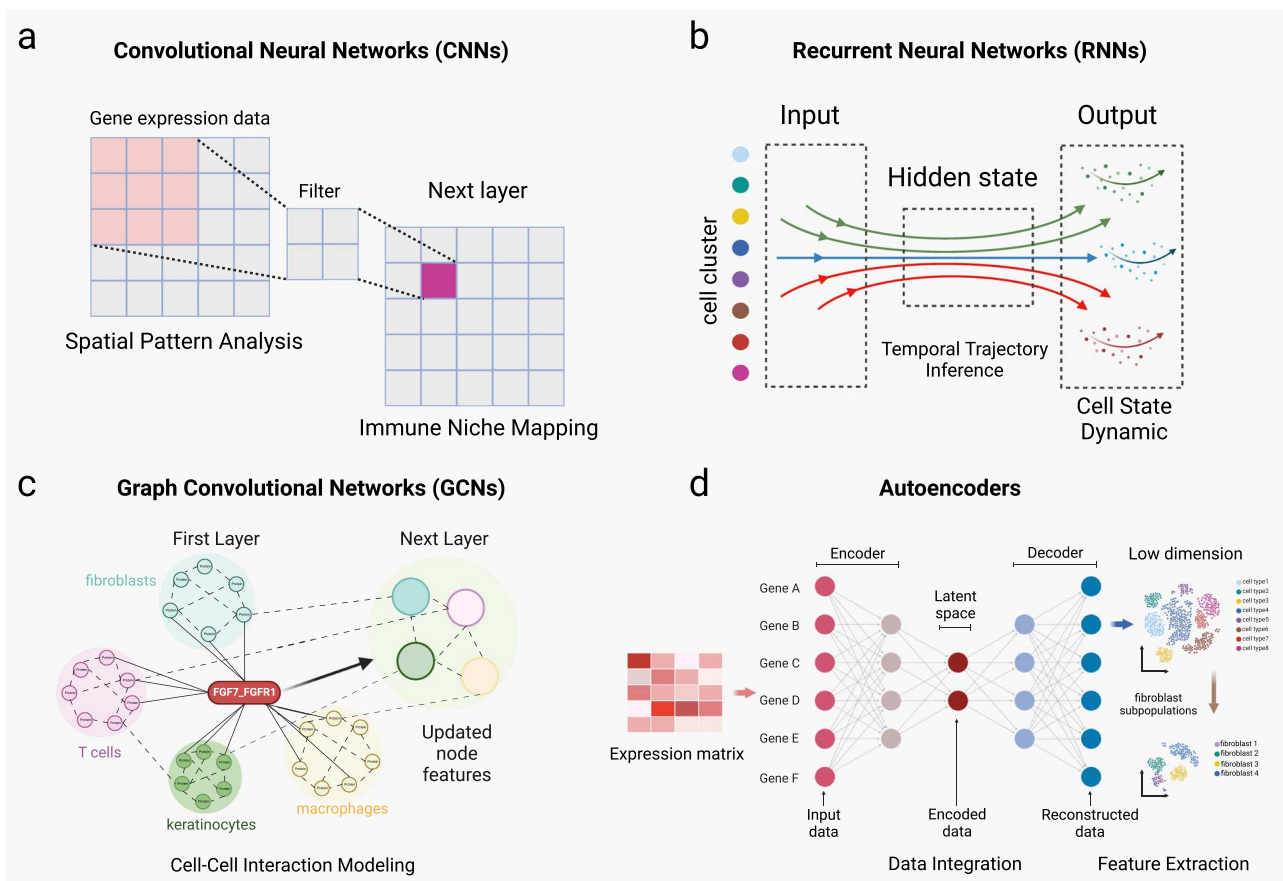


Figure 5. Deep learning categories for scRNA-seq analysis in wound healing. **(a)** Convolutional neural networks (CNNs), analyzing spatially resolved data to map immune niches and wound margin structures. **(b)** Recurrent neural networks (RNNs) infer temporal cell trajectories, tracing keratinocyte dynamics across healing phases. **(c)** Graph convolutional networks (GCNs) model cell–cell interactions, capturing signaling networks (e.g. FGF7–FGFR1) among fibroblasts, T cells, macrophages, and keratinocytes. **(d)** Autoencoders integrate data and extract features, identifying fibroblast subpopulations across wound stages. Commonly used tools include scVI (autoencoders), scVel (RNNs), CellChat (GCN potential), and CNN frameworks in spatial transcriptomics, enhancing insights into skin repair

[147] as critical for hair follicle neogenesis in large murine wounds. These trajectories aligned with human wound healing data, where Theocharidis *et al.* [139] reported upregulated FOSL1 in keratinocytes driving re-epithelialization in diabetic foot ulcers, and Liu *et al.* [54] confirmed COL1A1 expression in fibroblasts supporting extracellular matrix remodeling. Additionally, Graph Convolutional Networks (GCNs) implemented via CellChat [148] mapped macrophage–keratinocyte signaling through CXCL1, corroborating immune modulation mechanisms in chronic wounds as detailed by Theocharidis *et al.* [57]. The higher proportion of upper wound fibroblasts in regenerating wounds [147] further suggested their migration from the periphery to the center, a process potentially modulated by retinoic acid signaling, thereby enhancing the understanding of regenerative potential and offering precise therapeutic targets for wound repair.

Despite these insights, discrepancies between inferred trajectories and established biological knowledge necessitated a rigorous resolution framework. Greener *et al.* [64] emphasized validating machine learning predictions against ground truth data, proposing the use of spatial transcriptomics from Sun *et al.* [79] to verify spatial organization and cell transitions, such as the unexpected timing of fibroblast-to-myofibroblast transitions [147]. To address such conflicts, the approach recommended cross-validation with independent

datasets, including histological assays or spatial mapping, and consultation with domain experts to discern whether deviations reflected novel biology or methodological artifacts, as recommended by Brendel *et al.* [76]. Where data limitations persisted, the strategy intended to enrich models with specific markers, such as Crabp1+ upper wound fibroblasts [147] to track regenerative competence and CD14+ dendritic cells [159] to capture immune interactions. This multilayered strategy aimed to reconcile machine learning outputs with biological evidence, fostering an iterative cycle of validation, refinement, and hypothesis generation tailored to the complexities of wound healing.

Algorithm principles and programming languages in wound healing scRNA-seq analysis

Machine learning applications in single-cell RNA sequencing (scRNA-seq) for wound healing research are predominantly implemented in R due to its extensive bioinformatics ecosystem (Supplemental Table 1). The algorithmic principles underpinning these analyses fall into several key categories. Dimensionality reduction algorithms constitute the foundation, with nonlinear approaches like uniform manifold approximation and projection (UMAP) preserving local neighborhood structures [54, 55, 160] and t-distributed

stochastic neighbor embedding (t-SNE) emphasizing distance-based probability distributions [5, 140]. These are frequently complemented by linear dimensionality reduction via principal component analysis (PCA), which captures maximum variance along orthogonal components [139]. Trajectory inference algorithms represent another critical algorithmic class, with diffusion map-based approaches implemented in Monocle 3 modeling cellular transitions as diffusion processes [58, 161], while principal curve-based methods in Slingshot fit smooth curves through high-dimensional data clouds [54]. The computational analysis of cellular communication leverages probability-based algorithms in CellChat and NicheNetR that quantify ligand-receptor interaction likelihoods [54, 160], while regulatory network inference employs graph-based algorithms in SCENIC and CellOracle to reconstruct transcription factor networks from correlation patterns [162, 163].

The R programming language provides distinctive characteristics that facilitate these complex analyses. R's vectorized operations enable efficient manipulation of high-dimensional expression matrices, while its functional programming paradigm supports the analytical pipeline construction prominent in packages like Seurat (versions 3.0 to 5.2.0) [164]. Statistical testing in R implements nonparametric algorithms including Wilcoxon rank-sum tests [156, 161] for differential expression analysis, accommodating the non-normal distributions typical of scRNA-seq data. Integration of multimodal data represents an algorithmic frontier, with methods like SPOTlight [55] employing transfer learning principles to bridge scRNA-seq with spatial transcriptomics, while Cell2location [54] uses Bayesian algorithms for spatial mapping. Machine learning classification algorithms, exemplified by Random Forest implementation in Haensel *et al.*'s study, leverage ensemble decision trees for cell state prediction [136]. The language's extensibility has facilitated specialized analytical modules for chromatin accessibility analysis through ArchR, pseudotemporal ordering through CytoTRACE [58, 59, 165], and pathway enrichment via statistical overrepresentation tests in GSEA, GSVA, and GO analyses [54, 143]. These computational approaches collectively constitute a sophisticated algorithmic framework that systematically deconstructs the complex cellular dynamics of wound healing, revealing mechanistic insights that bridge fundamental biology with therapeutic potential.

Challenges and future perspectives of single-cell sequencing and artificial intelligence in wound healing research

Limitations of scRNA-seq and machine learning in wound healing research

While single-cell RNA sequencing (scRNA-seq) and machine learning have revolutionized wound healing research, these technologies are not without limitations that can impact the reliability of findings. A primary challenge in scRNA-seq is the presence of high noise levels and dropout events, where lowly expressed genes fail to be detected due to low sequencing depth or technical variability [166, 167]. This can obscure rare cell populations, such as specific fibroblast subtypes critical to regeneration, potentially leading to incomplete clustering or trajectory inference results. In machine learning, challenges include overfitting, particularly in deep learning models like autoencoders, where excessive tuning to training data may produce overly optimistic predictions that fail to generalize across datasets [168, 169]. Additionally, the

lack of interpretability in complex models—often described as a ‘black box’ problem—limits the ability to link computational outputs to biological mechanisms, hindering clinical translation [170, 171]. The dependency on high-quality, well-annotated training data further complicates applications, as wound healing datasets may vary in batch effects or sampling conditions, affecting model performance [56, 172]. These limitations can lead to misidentification of cellular subpopulations (e.g. mistaking noise for a rare cell type) or unreliable therapeutic target predictions (e.g. overestimating a pathway's role due to overfitting). Addressing these issues requires robust preprocessing (e.g. dropout correction with scScope and batch correction with methods like DESC) and validation with orthogonal data (e.g. proteomics) [8], ensuring that findings are both accurate and actionable in wound healing research.

Future horizons: integrating interpretable AI and multi-omics for personalized wound healing

The future of wound healing research lies in the development of interpretable AI models that can provide clear insights into the biological mechanisms underlying model predictions, which is crucial for translating computational findings into actionable clinical insights [173, 174]. As the field progresses, there will be an increasing emphasis on the integration of multi-omics data, such as single-cell transcriptomics, spatial metabolomics, and chromatin accessibility profiling, using advanced machine learning techniques like multi-view learning and tensor factorization [175]. For example, single-cell RNA sequencing reveals fibroblasts as the most transcriptionally dynamic cells in wound healing, transitioning from transient *Prg4*-expressing populations at early wound edges to stable, spatially distinct populations that stratify mature scars. These fibroblast states engage in complex, time-dependent signaling with immune cells, highlighting the spatial and temporal heterogeneity in fibroblast function throughout the repair process [56]. Recent advances in wound healing research highlight the transformative potential of multi-omics approaches that integrate transcriptomic, proteomic, and ultrastructural analyses to decipher the molecular distinctions between regenerative and fibrotic repair pathways [176]. By applying cell barcoding techniques and sophisticated data integration methods, researchers have identified critical genetic regulators like *Trps1* that distinguish regenerative from fibrotic trajectories, providing a comprehensive roadmap for future investigations targeting enhanced tissue regeneration [58, 177]. Additionally, real-time monitoring and prediction of wound healing progress through the integration of wearable sensor data, imaging, and molecular profiling will become more feasible, enabling timely interventions [178]. Personalized treatment strategies based on large-scale, multimodal datasets will allow for tailored therapies that consider individual patient characteristics [179]. Moreover, machine learning approaches, particularly deep learning models, will play a significant role in drug discovery and repurposing, predicting the efficacy of existing drugs and aiding in the design of novel therapeutic compounds [180]. Artificial intelligence offers transformative potential in wound care by providing advanced computer-assisted methods to assess and predict healing across the complex, multifactorial wound healing process involving hemostasis, inflammation, proliferation, and remodeling [173]. Global studies demonstrate AI's promising capabilities in analyzing wound components, enabling more accurate assessment and personalized treatment selection, though significant

challenges remain that must be addressed before widespread clinical implementation.

Conclusions

In this review, we have explored the application of machine learning to various aspects of scRNA-seq data analysis in wound healing, acknowledging both the significant advances and the challenges that remain. The convergence of single-cell transcriptomics, spatial genomics, and advanced machine learning techniques has revolutionized wound healing research, offering unprecedented insights into cellular heterogeneity and molecular mechanisms. These approaches have been instrumental in developing predictive models, identifying novel therapeutic targets, and designing advanced biomaterials. As these technologies continue to evolve and integrate with other omics, they hold the potential to transform wound healing into a more precise and personalized field of regenerative medicine. However, translating these complex datasets and models into clinical practice will require validation in larger patient cohorts, standardized protocols, and careful consideration of ethical and regulatory issues associated with AI-driven healthcare decisions.

Abbreviations

AI, artificial intelligence; ML, machine learning; DL, deep learning; scRNA-seq, single-cell RNA sequencing; PCA, principal component analysis; t-SNE, t-distributed stochastic neighbor embedding; UMAP, uniform manifold approximation and projection; SVM, support vector machine; CNN, convolutional neural network; RNN, recurrent neural network; GCN, graph convolutional network; VAE, variational autoencoder; ECM, extracellular matrix; NMF, non-negative matrix factorization; OOB, out-of-bag; GSEA, gene set enrichment analysis; GO, Gene Ontology; GSVA, gene set variation analysis; HS, hypertrophic scars; GSLMRGs, glycosphingolipid metabolism-related genes; DFU, diabetic foot ulcer; TGF β , transforming growth factor- β ; PDGF, platelet-derived growth factor; SDF1, stromal cell-derived factor 1; VEGF, vascular endothelial growth factor; DESC, deep embedding for single-cell clustering; CCA, canonical correlation analysis; IPA, ingenuity pathway analysis; T2DM, type 2 diabetes mellitus; VITAE, variational inference for trajectory by autoencoder; NNMF, non-negative matrix factorization.

Acknowledgements

We gratefully acknowledge Dr Tan Shuwen for her expert assistance in proofreading and polishing the manuscript.

Author contributions

Jianzhou Cui (Conceptualization [lead], Data curation [equal], Formal analysis [equal], Funding acquisition [equal], Investigation [equal], Methodology [equal], Project administration [equal], Resources [equal], Software [equal], Supervision [equal], Writing—original draft [equal], Writing—review & editing [equal]), Mei Wang (Conceptualization [equal], Data curation [equal], Funding acquisition [equal], Writing—original draft [equal]), Chenshi Lin (Data curation [equal], Formal analysis [equal], Software [equal], Writing—review & editing [equal]), Xu Xu (Conceptualization [equal], Investigation [equal], Project administration [equal], Resources [equal], Writing—review & editing [equal]),

and Zhenqing Zhang (Conceptualization [equal], Funding acquisition [equal], Project administration [equal], Supervision [equal], Writing—review & editing [equal])

Supplementary data

Supplementary data is available at *Burns & Trauma Journal* online.

Consent for publication

All authors agree to publish this review.

Conflict of interest: The authors declare that they have no competing interests.

Funding

This work was funded by the Bioinformatics Core of the Immunology Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore; the National Natural Science Foundation of China (82260788); the Guizhou Provincial Science and Technology Support Plan (General Project) (Qiankehe Support [2023] General 428); and the College Student Innovation and Entrepreneurship Training Program (ZYDC202202299, ZYDC202301087). This work was also supported by the China Scholarship Council (CSC).

Data availability

Not applicable.

References

1. Pena OA, Martin P. Cellular and molecular mechanisms of skin wound healing. *Nat Rev Mol Cell Biol.* 2024;25:599–616. <https://doi.org/10.1038/s41580-024-00715-1>.
2. de Bont CM, Boelens WC, Puijn GJM. Netosis, complement, and coagulation: a triangular relationship. *Cell Mol Immunol.* 2019;16:19–27. <https://doi.org/10.1038/s41423-018-0024-0>.
3. Burzynski LC, Humphry M, Pyrrillou K, Wiggins KA, Chan JN, Figg N, *et al.* The coagulation and immune systems are directly linked through the activation of interleukin-1 α by thrombin. *Immunity.* 2019;50:1033–1042.e6. <https://doi.org/10.1016/j.immuni.2019.03.003>.
4. Grose R, Hutter C, Bloch W, Thorey I, Watt FM, Fassler R, *et al.* A crucial role of beta 1 integrins for keratinocyte migration in vitro and during cutaneous wound repair. *Development.* 2002;129:2303–15. <https://doi.org/10.1242/dev.129.9.2303>.
5. Theocharidis G, Tekkela S, Veves A, McGrath JA, Onoufriadis A. Single-cell transcriptomics in human skin research: available technologies, technical considerations and disease applications. *Exp Dermatol.* 2022;31:655–73. <https://doi.org/10.1111/exd.14547>.
6. Driskell RR, Lichtenberger BM, Hoste E, Kretschmar K, Simons BD, Charalambous M, *et al.* Distinct fibroblast lineages determine dermal architecture in skin development and repair. *Nature.* 2013;504:277–81. <https://doi.org/10.1038/nature12783>.
7. Guerrero-Juarez CF, Dedhia PH, Jin S, Ruiz-Vega R, Ma D, Liu Y, *et al.* Single-cell analysis reveals fibroblast heterogeneity and myeloid-derived adipocyte progenitors in murine skin wounds. *Nat Commun.* 2019;10:650. <https://doi.org/10.1038/s41467-018-08247-x>.
8. Deng Y, Bao F, Dai Q, Wu LF, Altschuler SJ. Scalable analysis of cell-type composition from single-cell transcriptomics using deep recurrent learning. *Nat Methods.* 2019;16:311–4. <https://doi.org/10.1038/s41592-019-0353-7>.
9. Li X, Wang K, Lyu Y, Pan H, Zhang J, Stambolian D, *et al.* Deep learning enables accurate clustering with batch effect removal in single-cell RNA-seq analysis. *Nat Commun.* 2020;11:2338. <https://doi.org/10.1038/s41467-020-15851-3>.

10. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol.* 2014;32:381–6. <https://doi.org/10.1038/nbt.2859>.
11. Du J-H, Chen T, Gao M, Wang J. Joint trajectory inference for single-cell genomics using deep learning with a mixture prior. *Proceedings of the National Academy of Sciences.* 2024;121:e2316256121. <https://doi.org/doi:10.1073/pnas.2316256121>.
12. Guo XF, Gu SS, Wang J, Sun H, Zhang YJ, Yu PF, *et al.* Protective effect of mesenchymal stem cell-derived exosomal treatment of hippocampal neurons against oxygen-glucose deprivation/reperfusion-induced injury. *World J Emerg Med.* 2022;13:46–53. <https://doi.org/10.5847/wjem.j.1920-8642.2022.015>.
13. Golebiewska EM, Poole AW. Platelet secretion: from haemostasis to wound healing and beyond. *Blood Rev.* 2015;29:153–62. <https://doi.org/10.1016/j.blre.2014.10.003>.
14. Szpaderska AM, Egozi EI, Gamelli RL, DiPietro LA. The effect of thrombocytopenia on dermal wound healing. *J Invest Dermatol.* 2003;120:1130–7. [https://doi.org/10.1016/S0022-202X\(18\)32320-0](https://doi.org/10.1016/S0022-202X(18)32320-0).
15. Park S, Gonzalez DG, Guirao B, Boucher JD, Cockburn K, Marsh ED, *et al.* Tissue-scale coordination of cellular behaviour promotes epidermal wound repair in live mice. *Nat Cell Biol.* 2017;19:155–63. <https://doi.org/10.1038/ncb3472>.
16. Aragona M, Dekoninck S, Rulands S, Lenglez S, Mascré G, Simons BD, *et al.* Defining stem cell dynamics and migration during wound healing in mouse skin epidermis. *Nat Commun.* 2017;8:14684. <https://doi.org/10.1038/ncomms14684>.
17. Turley J, Chenchiah IV, Martin P, Liverpool TB, Weavers H. *Deep Learning for Rapid Analysis of Cell Divisions In Vivo during Epithelial Morphogenesis and Repair.* Cambridge, UK: eLife Sciences Publications, Ltd, 2024. <https://doi.org/10.7554/eLife.87949.2>.
18. Gudipaty SA, Lindblom J, Loftus PD, Redd MJ, Edes K, Davey C, *et al.* Mechanical stretch triggers rapid epithelial cell division through piezo1. *Nature.* 2017;543:118–21. <https://doi.org/10.1038/nature21407>.
19. Singh P, Chen C, Pal-Ghosh S, Stepp MA, Sheppard D, Van De Water L. Loss of integrin $\alpha 9 \beta 1$ results in defects in proliferation, causing poor re-epithelialization during cutaneous wound healing. *J Invest Dermatol.* 2009;129:217–28. <https://doi.org/10.1038/jid.2008.201>.
20. McAndrews KM, Miyake T, Ehsanipour EA, Kelly PJ, Becker LM, McGrail DJ, *et al.* Dermal α SMA+ myofibroblasts orchestrate skin wound repair via $\beta 1$ integrin and independent of type I collagen production. *EMBO J.* 2022;41:e109470. <https://doi.org/10.15252/embj.2021109470>.
21. Hertle MD, Kubler MD, Leigh IM, Watt FM. Aberrant integrin expression during epidermal wound healing and in psoriatic epidermis. *J Clin Invest.* 1992;89:1892–901. <https://doi.org/10.1172/JCI115794>.
22. Chen J, Xu H, Tao W, Chen Z, Zhao Y, Han J-DJ. Transformer for one stop interpretable cell type annotation. *Nature. Communications.* 2023;14:223. <https://doi.org/10.1038/s41467-023-35923-4>.
23. Rinkevich Y, Walmsley GG, Hu MS, Maan ZN, Newman AM, Drukker M, *et al.* Skin fibrosis. Identification and isolation of a dermal lineage with intrinsic fibrogenic potential. *Science.* 2015;348:aaa2151. <https://doi.org/10.1126/science.aaa2151>.
24. Jiang D, Correa-Gallegos D, Christ S, Stefanska A, Liu J, Ramesh P, *et al.* Two succeeding fibroblastic lineages drive dermal development and the transition from regeneration to scarring. *Nat Cell Biol.* 2018;20:422–31. <https://doi.org/10.1038/s41556-018-0073-8>.
25. Shook BA, Wasko RR, Mano O, Rutenberg-Schoenberg M, Rudolph MC, Zirak B, *et al.* Dermal adipocyte lipolysis and myofibroblast conversion are required for efficient skin repair. *Cell Stem Cell.* 2020;26:880–895.e6. <https://doi.org/10.1016/j.stem.2020.03.013>.
26. Kalgudde Gopal S, Dai R, Stefanska AM, Ansari M, Zhao J, Ramesh P, *et al.* Wound infiltrating adipocytes are not myofibroblasts. *Nat Commun.* 2023;14:3020. <https://doi.org/10.1038/s41467-023-38591-6>.
27. Murray PJ. Macrophage polarization. *Annu Rev Physiol.* 2017;79:541–66. <https://doi.org/10.1146/annurev-physiol-022516-034339>.
28. Belge KU, Dayyani F, Horelt A, Siedlar M, Frankenberger M, Frankenberger B, *et al.* The proinflammatory CD14+CD16+DR++ monocytes are a major source of TNF. *J Immunol.* 2002;168:3536–42. <https://doi.org/10.4049/jimmunol.168.7.3536>.
29. Stein M, Keshav S, Harris N, Gordon S. Interleukin 4 potently enhances murine macrophage mannose receptor activity: a marker of alternative immunologic macrophage activation. *J Exp Med.* 1992;176:287–92. <https://doi.org/10.1084/jem.176.1.287>.
30. Willenborg S, Sanin DE, Jais A, Ding X, Ulas T, Nuchel J, *et al.* Mitochondrial metabolism coordinates stage-specific repair processes in macrophages during wound healing. *Cell Metab.* 2021;33:2398–2414.e9. <https://doi.org/10.1016/j.cmet.2021.10.004>.
31. Ariel A, Serhan CN. New lives given by cell death: macrophage differentiation following their encounter with apoptotic leukocytes during the resolution of inflammation. *Front Immunol.* 2012;3:4. <https://doi.org/10.3389/fimmu.2012.00004>.
32. Levy BD, Clish CB, Schmidt B, Gronert K, Serhan CN. Lipid mediator class switching during acute inflammation: signals in resolution. *Nat Immunol.* 2001;2:612–9. <https://doi.org/10.1038/89759>.
33. Komi DEA, Khomtchouk K, Santa Maria PL. A review of the contribution of mast cells in wound healing: involved molecular and cellular mechanisms. *Clin Rev Allergy Immunol.* 2020;58:298–312. <https://doi.org/10.1007/s12016-019-08729-w>.
34. Willenborg S, Eckes B, Brinckmann J, Krieg T, Waisman A, Hartmann K, *et al.* Genetic ablation of mast cells redefines the role of mast cells in skin wound healing and bleomycin-induced fibrosis. *J Invest Dermatol.* 2014;134:2005–15. <https://doi.org/10.1038/jid.2014.12>.
35. Antsiferova M, Martin C, Huber M, Feyerabend TB, Forster A, Hartmann K, *et al.* Mast cells are dispensable for normal and activin-promoted wound healing and skin carcinogenesis. *J Immunol.* 2013;191:6147–55. <https://doi.org/10.4049/jimmunol.1301350>.
36. Joshi N, Pohlmeier L, Ben-Yehuda Greenwald M, Haertel E, Hiebert P, Kopf M, *et al.* Comprehensive characterization of myeloid cells during wound healing in healthy and healing-impaired diabetic mice. *Eur J Immunol.* 2020;50:1335–49. <https://doi.org/10.1002/eji.201948438>.
37. Ribot JC, Lopes N, Silva-Santos B. Gammadelta T cells in tissue physiology and surveillance. *Nat Rev Immunol.* 2021;21:221–32. <https://doi.org/10.1038/s41577-020-00452-4>.
38. Jameson J, Ugarte K, Chen N, Yachi P, Fuchs E, Boismenu R, *et al.* A role for skin $\gamma \delta$ T cells in wound repair. *Science.* 2002;296:747–9. <https://doi.org/10.1126/science.1069639>.
39. Toulon A, Breton L, Taylor KR, Tenenhaus M, Bhavsar D, Lanigan C, *et al.* A role for human skin-resident T cells in wound healing. *J Exp Med.* 2009;206:743–50. <https://doi.org/10.1084/jem.20081787>.
40. Sharp LL, Jameson JM, Cauvi G, Havran WL. Dendritic epidermal T cells regulate skin homeostasis through local production of insulin-like growth factor 1. *Nat Immunol.* 2005;6:73–9. <https://doi.org/10.1038/ni1152>.
41. Gay D, Kwon O, Zhang Z, Spata M, Plikus MV, Holler PD, *et al.* Fgf9 from dermal $\gamma \delta$ T cells induces hair follicle neogenesis after wounding. *Nat Med.* 2013;19:916–23. <https://doi.org/10.1038/nm.3181>.

42. Jameson JM, Cauvi G, Sharp LL, Witherden DA, Havran WL. $\Gamma\delta$ T cell-induced hyaluronan production by epithelial cells regulates inflammation. *J Exp Med*. 2005;201:1269–79. <https://doi.org/10.1084/jem.20042057>.
43. Ho AW, Kupper TS. T cells and the skin: from protective immunity to inflammatory skin disorders. *Nat Rev Immunol*. 2019;19:490–502. <https://doi.org/10.1038/s41577-019-0162-3>.
44. Kalekar LA, Cohen JN, Prevel N, Sandoval PM, Mathur AN, Moreau JM, et al. Regulatory T cells in skin are uniquely poised to suppress profibrotic immune responses. *Sci Immunol*. 2019;4:eaaw2910. <https://doi.org/10.1126/sciimmunol.aaw2910>.
45. Szabo PA, Miron M, Farber DL. Location, location, location: tissue resident memory T cells in mice and humans. *Sci Immunol*. 2019;4:eaas9673. <https://doi.org/10.1126/sciimmunol.aas9673>.
46. Debes GF, McGettigan SE. Skin-associated B cells in health and inflammation. *J Immunol*. 2019;202:1659–66. <https://doi.org/10.1049/jimmunol.1801211>.
47. Oliveira HC, Popi AF, Bachi AL, Nonogaki S, Lopes JD, Mariano M. B-1 cells modulate the kinetics of wound-healing process in mice. *Immunobiology*. 2010;215:215–22. <https://doi.org/10.1016/j.imbio.2009.03.009>.
48. Iwata Y, Yoshizaki A, Komura K, Shimizu K, Ogawa F, Hara T, et al. Cd19, a response regulator of B lymphocytes, regulates wound healing through hyaluronan-induced TLR4 signaling. *Am J Pathol*. 2009;175:649–60. <https://doi.org/10.2353/ajpath.2009.080355>.
49. Spits H, Artis D, Colonna M, Diefenbach A, Di Santo JP, Eberl G, et al. Innate lymphoid cells—a proposal for uniform nomenclature. *Nat Rev Immunol*. 2013;13:145–9. <https://doi.org/10.1038/nri3365>.
50. Rak GD, Osborne LC, Siracusa MC, Kim BS, Wang K, Bayat A, et al. IL-33-dependent group 2 innate lymphoid cells promote cutaneous wound healing. *J Invest Dermatol*. 2016;136:487–96. <https://doi.org/10.1038/jid.2015.406>.
51. Sarker IH. Machine learning: algorithms, real-world applications and research directions. *SN Comput Sci*. 2021;2:160. <https://doi.org/10.1007/s42979-021-00592-x>.
52. Angermueller C, Parnamaa T, Parts L, Stegle O. Deep learning for computational biology. *Mol Syst Biol*. 2016;12:878. <https://doi.org/10.15252/msb.20156651>.
53. Sommer C, Gerlich DW. Machine learning in cell biology - teaching computers to recognize phenotypes. *J Cell Sci*. 2013;126:5529–39. <https://doi.org/10.1242/jcs.123604>.
54. Liu Z, Bian X, Luo L, Bjorklund AK, Li L, Zhang L, et al. Spatiotemporal single-cell roadmap of human skin wound healing. *Cell Stem Cell*. 2024;32:479–498.e8. <https://doi.org/10.1016/j.stem.2024.11.013>.
55. Chen P, Tang S, Li M, Wang D, Chen C, Qiu Y, et al. Single-cell and spatial transcriptomics decodes Wharton's jelly-derived mesenchymal stem cells heterogeneity and a subpopulation with wound repair signatures. *Adv Sci (Weinh)*. 2023;10:e2204786. <https://doi.org/10.1002/advs.202204786>.
56. Almet AA, Liu Y, Nie Q, Plikus MV. Integrated single-cell analysis reveals spatially and temporally dynamic heterogeneity in fibroblast states during wound healing. *J Invest Dermatol*. 2025;145:645–659.e25. <https://doi.org/10.1016/j.jid.2024.06.1281>.
57. Theodoridis G, Baltzis D, Roustit M, Tellechea A, Dangwal S, Khetani RS, et al. Integrated skin transcriptomics and serum multiplex assays reveal novel mechanisms of wound healing in diabetic foot ulcers. *Diabetes*. 2020;69:2157–69. <https://doi.org/10.2337/db20-0188>.
58. Mascharak S, Talbott HE, Januszyk M, Griffin M, Chen K, Davitt MF, et al. Multi-omic analysis reveals divergent molecular events in scarring and regenerative wound healing. *Cell Stem Cell*. 2022;29:315–327.e6. <https://doi.org/10.1016/j.stem.2021.12.011>.
59. Foster DS, Januszyk M, Yost KE, Chinta MS, Gulati GS, Nguyen AT, et al. Integrated spatial multiomics reveals fibroblast fate during tissue repair. *Proc Natl Acad Sci USA*. 2021;118:e2110025118. <https://doi.org/10.1073/pnas.2110025118>.
60. Stanton JM. Galton, Pearson, and the peas: a brief history of linear regression for statistics instructors. *J Stat Educ*. 2017;9:1–13. <https://doi.org/10.1080/10691898.2001.11910537>.
61. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20:273–97. <https://doi.org/10.1023/a:1022627411411>.
62. Hughes BK, Wallis R, Bishop CL. Yearning for machine learning: applications for the classification and characterisation of senescence. *Cell Tissue Res*. 2023;394:1–16. <https://doi.org/10.1007/s00441-023-03768-4>.
63. Kingsmore KM, Puglisi CE, Grammer AC, Lipsky PE. An introduction to machine learning and analysis of its use in rheumatic diseases. *Nat Rev Rheumatol*. 2021;17:710–30. <https://doi.org/10.1038/s41584-021-00708-w>.
64. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. *Nat Rev Mol Cell Biol*. 2022;23:40–55. <https://doi.org/10.1038/s41580-021-00407-0>.
65. van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res*. 2008;9:2579–2605.
66. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci*. 2016;374:20150202. <https://doi.org/10.1098/rsta.2015.0202>.
67. McInnes, et al. UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software*, 2018;3:861. <https://doi.org/10.21105/joss.00861>.
68. Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, et al. Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface*. 2018;15:1–47. <https://doi.org/10.1098/rsif.2017.0387>.
69. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–44. <https://doi.org/10.1038/nature14539>.
70. Buchan DWA, Jones DT. The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res*. 2019;47:W402–7. <https://doi.org/10.1093/nar/gkz297>.
71. Muller AT, Hiss JA, Schneider G. Recurrent neural network model for constructive peptide design. *J Chem Inf Model*. 2018;58:472–9. <https://doi.org/10.1021/acs.jcim.7b00414>.
72. Xu J, McPartlon M, Li J. Improved protein structure prediction by deep learning irrespective of co-evolution information. *Nat Mach Intell*. 2021;3:601–9. <https://doi.org/10.1038/s42256-021-00348-5>.
73. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*. 2018;34:i457–66. <https://doi.org/10.1093/bioinformatics/bty294>.
74. Wang J, Ma A, Chang Y, Gong J, Jiang Y, Qi R, et al. scGNN is a novel graph neural network framework for single-cell RNA-Seq analyses. *Nat Commun*. 2021;12:1882. <https://doi.org/10.1038/s41467-021-22197-x>.
75. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods*. 2009;6:377–82. <https://doi.org/10.1038/nmeth.1315>.
76. Brendel M, Su C, Bai Z, Zhang H, Elemento O, Wang F. Application of deep learning on single-cell RNA sequencing data analysis: a review. *Genomics Proteomics Bioinformatics*. 2022;20:814–35. <https://doi.org/10.1016/j.gpb.2022.11.011>.
77. Li R, Li L, Xu Y, Yang J. Machine learning meets omics: applications and perspectives. *Brief Bioinform*. 2022;23:1–22. <https://doi.org/10.1093/bib/bbab460>.
78. Madadi Y, Monavarfeshani A, Chen H, Stamer WD, Williams RW, Yousefi S. Artificial intelligence models for cell type and subtype identification based on single-cell RNA sequencing data in vision science. *IEEE/ACM Trans Comput Biol Bioinform*. 2023;20:2837–52. <https://doi.org/10.1109/TCBB.2023.3284795>.

79. Sun Y, Kong L, Huang J, Deng H, Bian X, Li X, *et al.* A comprehensive survey of dimensionality reduction and clustering methods for single-cell and spatial transcriptomics data. *Brief Funct Genomics*. 2024;23:733–44. <https://doi.org/10.1093/bfgp/ela023>.
80. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, *et al.* Comprehensive integration of single-cell data. *Cell*. 2019;177:1888–1902.e21. <https://doi.org/10.1016/j.cell.2019.05.031>.
81. Li J, Jiang W, Han H, Liu J, Liu B, Wang Y. ScGSLC: an unsupervised graph similarity learning framework for single-cell RNA-seq data clustering. *Comput Biol Chem*. 2021;90:107415. <https://doi.org/10.1016/j.compbiolchem.2020.107415>.
82. Wang D, Gu J. VASC: dimension reduction and visualization of single-cell RNA-seq data by deep variational autoencoder. *Genomics Proteomics Bioinformatics*. 2018;16:320–31. <https://doi.org/10.1016/j.gpb.2018.08.003>.
83. Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. *Nat Methods*. 2018;15:1053–8. <https://doi.org/10.1038/s41592-018-0229-2>.
84. Tran D, Nguyen H, Tran B, La Vecchia C, Luu HN, Nguyen T. Fast and precise single-cell data analysis using a hierarchical autoencoder. *Nat Commun*. 2021;12:1029. <https://doi.org/10.1038/s41467-021-21312-2>.
85. Grønbech CH, Vording MF, Timshel PN, Sønderby CK, Pers TH, Winther O. scVAE: variational auto-encoders for single-cell gene expression data. *Bioinformatics*. 2020;36:4415–22. <https://doi.org/10.1093/bioinformatics/btaa293>.
86. Ding J, Regev A. Deep generative model embedding of single-cell RNA-seq profiles on hyperspheres and hyperbolic spaces. *Nat Commun*. 2021;12:2554. <https://doi.org/10.1038/s41467-021-22851-4>.
87. Bica I, Andres-Terre H, Cvejic A, Lio P. Unsupervised generative and graph representation learning for modelling cell differentiation. *Sci Rep*. 2020;10:9790. <https://doi.org/10.1038/s41598-020-66166-8>.
88. Zhang C. Single-cell data analysis using mmd variational autoencoder for a more informative latent representation. *bioRxiv*. 2019:613414. <https://doi.org/10.1101/613414>.
89. Lin E, Mukherjee S, Kannan S. A deep adversarial variational autoencoder model for dimensionality reduction in single-cell RNA sequencing analysis. *BMC Bioinformatics*. 2020;21:64. <https://doi.org/10.1186/s12859-020-3401-5>.
90. Mondal AK, Asnani H, Singla P, Prathosh A. scRAE: deterministic regularized autoencoders with flexible priors for clustering single-cell gene expression data. *IEEE/ACM Trans Comput Biol Bioinform*. 2021;19:2996–3007. <https://doi.org/10.1109/TCBB.2021.3098394>.
91. Luo Z, Xu C, Zhang Z, Jin W. A topology-preserving dimensionality reduction method for single-cell RNA-seq data using graph autoencoder. *Sci Rep*. 2021;11:20028. <https://doi.org/10.1038/s41598-021-99003-7>.
92. Alessandri L, Cordero F, Beccuti M, Licheri N, Arigoni M, Olivero M, *et al.* Sparsely-connected autoencoder (SCA) for single cell RNAseq data mining. *NPJ Syst Biol Appl*. 2021;7:1. <https://doi.org/10.1038/s41540-020-00162-6>.
93. Peng J, Wang X, Shang X. Combining gene ontology with deep neural networks to enhance the clustering of single cell RNA-seq data. *BMC Bioinformatics*. 2019;20:284. <https://doi.org/10.1186/s12859-019-2769-6>.
94. Gut G, Stark SG, Rätsch G, Davidson NR. Pmvae: Learning interpretable single-cell representations with pathway modules. *bioRxiv*. 2021:2021.01.28.428664. <https://doi.org/10.1101/2021.01.28.428664>.
95. Seninge L, Anastopoulos I, Ding H, Stuart J. Biological network-inspired interpretable variational autoencoder. *bioRxiv*. 2020:2020.12.17.423310. <https://doi.org/10.1101/2020.12.17.423310>.
96. Wolf F, Angerer P, Theis FJ. SCANPY: largescale single-cell gene expression data analysis. *Genome Biol*. 2018;19:15. <https://doi.org/10.1186/s13059-017-1382-0>.
97. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Stat Mech Theory Exp*. 2008;2008:P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>.
98. Traag V, Waltman L, Van Eck N. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep*. 2019;9:5233. <https://doi.org/10.1038/s41598-019-41695-z>.
99. Srinivasan S, Leshchik A, Johnson NT, Korkin D. A hybrid deep clustering approach for robust cell type profiling using single-cell RNA-seq data. *RNA*. 2020;26:1303–19. <https://doi.org/10.1261/rna.074427.119>.
100. Lakkis J, Wang D, Zhang Y, Hu G, Wang K, Pan H, *et al.* A joint deep learning model enables simultaneous batch effect correction, denoising, and clustering in single-cell transcriptomics. *Genome Res*. 2021;31:1753–66. <https://doi.org/10.1101/gr.271874.120>.
101. Chen L, Wang W, Zhai Y, Deng M. Deep soft k-means clustering with self-training for single-cell RNA sequence data. *NAR Genom Bioinform*. 2020;2:lqaa039. <https://doi.org/10.1093/nargab/lqaa039>.
102. Zeng Y, Lin J, Zhou X, Lu Y, Yang Y. Graph convolutional network-based method for clustering single-cell rna-seq data. *bioRxiv*. 2021:2020.09.02.278804. <https://doi.org/10.1101/2020.09.02.278804>.
103. Huh R, Yang Y, Jiang Y, Shen Y, Li Y. SAME-clustering: Single-cell Aggregated Clustering via Mixture Model Ensemble. *Nucleic Acids Res*. 2020;48:86–95. <https://doi.org/10.1093/nar/gkz959>.
104. Kim J, Stanescu DE, Won KJ. CellBIC: bimodality-based top-down clustering of single-cell RNA sequencing data reveals hierarchical structure of the cell type. *Nucleic Acids Res*. 2018;46:e124. <https://doi.org/10.1093/nar/gky698>.
105. duVerle DA, Yotsukura S, Nomura S, Aburatani H, Tsuda K. CellTree: an R/bioconductor package to infer the hierarchical structure of cell populations from single-cell RNA-seq data. *BMC Bioinformatics*. 2016;17:363. <https://doi.org/10.1186/s12859-016-1175-6>.
106. Rizvi AH, Camara PG, Kandror EK, Roberts TJ, Schieren I, Maniatis T, *et al.* Single-cell topological RNA-seq analysis reveals insights into cellular differentiation and development. *Nat Biotechnol*. 2017;35:551–60. <https://doi.org/10.1038/nbt.3854>.
107. Abdelaal T, Michielsen L, Cats D, Hoogduin D, Mei H, Reinders MJT, *et al.* A comparison of automatic cell identification methods for single-cell RNA sequencing data. *Genome Biol*. 2019;20:194. <https://doi.org/10.1186/s13059-019-1795-z>.
108. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, *et al.* Attention is all you need. *Adv Neural Inf Proces Syst*. 2017;30:1–15. <https://doi.org/10.48550/arXiv.1706.03762>.
109. Kamath U, Graham K, Emara W. *Transformers for Machine Learning: A Deep Dive*. Boca Raton, FL: Chapman and Hall/CRC. 2022. <https://doi.org/10.1201/9781003170082>.
110. Hu L, Qiu P, Qin H, Cao L, Jiang W, Zhang Y, *et al.* RegFormer: a single-cell foundation model powered by gene regulatory hierarchies. *bioRxiv*. 2025:2025.01.24.634217. <https://doi.org/10.1101/2025.01.24.634217>.
111. Mellors T, Schneider M. A transformer-based approach to survival outcome prediction. *bioRxiv*. 2024:2024.11.03.621674. <https://doi.org/10.1101/2024.11.03.621674>.
112. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, *et al.* Highly accurate protein structure prediction with alphafold. *nature*. 2021;596:583–89. <https://doi.org/10.1038/s41586-021-03819-2>.
113. Choi SR, Lee M. Transformer architecture and attention mechanisms in genome data analysis: a comprehensive review. *Biology (Basel)*. 2023;12:12. <https://doi.org/10.3390/biology12071033>.

114. Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate arXiv preprint arXiv:1409.0473. 2014. <https://doi.org/10.48550/arXiv.1409.0473>.
115. Erfanian N, Heydari AA, Feriz AM, Iañez P, Derakhshani A, Ghasemigol M, *et al.* Deep learning applications in single-cell genomics and transcriptomics data analysis. *Biomed Pharmacother.* 2023;165:115077. <https://doi.org/10.1016/j.biopha.2023.115077>.
116. Ekman M. *Learning Deep Learning: Theory and Practice of Neural Networks, Computer Vision, Natural Language Processing, and Transformers Using Tensorflow*. Boston, MA: Addison-Wesley Professional, 2021.
117. Rizvi SA, Levine D, Patel A, Zhang S, Wang E, He S, *et al.* Scaling large language models for next-generation single-cell analysis. *bioRxiv.* 2025 2025.04.14.648850. [10.1101/2025.04.14.648850](https://doi.org/10.1101/2025.04.14.648850).
118. Zhang S, Fan R, Liu Y, Chen S, Liu Q, Zeng W. Applications of transformer-based language models in bioinformatics: a survey. *Bioinform Adv.* 2023;3:1–19. <https://doi.org/10.1093/bioadv/vbad001>.
119. Gong J, Hao M, Cheng X, Zeng X, Liu C, Ma J, *et al.* xTrimoGene: an efficient and scalable representation learner for single-cell RNA-seq data. *Adv Neural Inf Proces Syst.* 2023;36:69391–403.
120. Hu H, Zhang S, Choi Y, Malladi VS, Quon G. scIL-MA: a single-cell representation learning framework to leverage prior knowledge from large language models. *bioRxiv.* 2025, 2025.01.28.635153. <https://doi.org/10.1101/2025.01.28.635153>.
121. Zhang Z, Liu L. scIDPMs: single-cell RNA-seq imputation using diffusion probabilistic models. *bioRxiv.* 2024, 2024.02.29.582870. <https://doi.org/10.1101/2024.02.29.582870>.
122. Yang F, Wang W, Wang F, Fang Y, Tang D, Huang J, *et al.* scBERT as a large-scale pretrained deep language model for cell type annotation of single-cell RNA-seq data. *Nat Mach Intell.* 2022;4:852–66. <https://doi.org/10.1038/s42256-022-00534-z>.
123. Cui H, Wang C, Maan H, Pang K, Luo F, Duan N, *et al.* scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nat Methods.* 2024;21:1470–80. <https://doi.org/10.1038/s41592-024-02201-0>.
124. Jiao L, Wang G, Dai H, Li X, Wang S, Song T. scTransSort: transformers for intelligent annotation of cell types by gene embeddings. *Biomolecules.* 2023;13:611. <https://doi.org/10.3390/biom13040611>.
125. Wen H, Tang W, Dai X, Ding J, Jin W, Xie Y, *et al.* CellPLM: pre-training of cell language model beyond single cells. *BioRxiv.* 2023. 2023.10.03.560734.
126. Hou W, Ji Z. Assessing GPT-4 for cell type annotation in single-cell RNA-seq analysis. *bioRxiv.* 2023 2023.04.16.537094. [10.1101/2023.04.16.537094](https://doi.org/10.1101/2023.04.16.537094).
127. Jiang S, Hassanpour S. Transformer-based representation learning for robust gene expression modeling and cancer prognosis.arXiv:2504.09704. <https://doi.org/10.48550/arXiv.2504.09704>.
128. Zhang Z, Liu L. Scidpms: Single-cell rna-seq imputation using diffusion probabilistic models. *IEEE J Biomed Health Inform.* 2025;29:3057–68. <https://doi.org/10.1109/jbhi.2024.3430554>.
129. Cheng C, Chen W, Jin H, Chen X. A review of single-cell RNA-seq annotation, integration, and cell-cell communication. *Cells.* 2023;12:1–39. <https://doi.org/10.3390/cells12151970>.
130. Chen T, Wei X, Xie L, Zhang Y, Liu C, Shen W, *et al.* Self-former: multi-scale gene filtration transformer for single-cell spatial reconstruction. *Brief Bioinform.* 2024;25:1–13. <https://doi.org/10.1093/bib/bbae523>.
131. Mao Y, Lin Y-Y, Wong NKY, Volik S, Sar F, Collins C, *et al.* Phenotype prediction from single-cell RNA-seq data using attention-based neural networks. *Bioinformatics.* 2024;40:1–8. <https://doi.org/10.1093/bioinformatics/btae067>.
132. Liu G, Zhao Y, Zhao Y, Wang T, Cai Q, Wang X, *et al.* SCARF: single cell ATAC-seq and RNA-seq foundation model. *bioRxiv.* 2025 2025.04.07.647689. [10.1101/2025.04.07.647689](https://doi.org/10.1101/2025.04.07.647689).
133. Batut B, Hiltmann S, Bagnacani A, Baker D, Bhardwaj V, Blank C, *et al.* Community-driven data analysis training for biology. *Cell Syst.* 2018;6:752–58.e1. <https://doi.org/10.1016/j.cels.2018.05.012>.
134. Saelens W, Cannoodt R, Todorov H, Saeys Y. A comparison of single-cell trajectory inference methods. *Nature Biotechnology.* 2019;37:547–54. <https://doi.org/10.1038/s41587-019-0071-9>.
135. Li M, Zhang X, Ang KS, Ling J, Sethi R, Lee NYS, *et al.* DISCO: a database of Deeply Integrated human Single-Cell Omics data. *Nucleic Acids Res.* 2022;50:D596–602. <https://doi.org/10.1093/nar/gkab1020>.
136. Haensel D, Jin S, Sun P, Cinco R, Dragan M, Nguyen Q, *et al.* Defining epidermal basal cell states during skin homeostasis and wound healing using single-cell transcriptomics. *Cell Rep.* 2020;30:3932–3947.e6. <https://doi.org/10.1016/j.celrep.2020.02.091>.
137. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, *et al.* RNA velocity of single cells. *Nature.* 2018;560:494–8. <https://doi.org/10.1038/s41586-018-0414-6>.
138. Jin S, MacLean AL, Peng T, Nie Q. scEpath: energy landscape-based inference of transition probabilities and cellular trajectories from single-cell transcriptomic data. *Bioinformatics.* 2018;34:2077–86. <https://doi.org/10.1093/bioinformatics/bty058>.
139. Theocharidis G, Thomas BE, Sarkar D, Mumme HL, Pilcher WJR, Dwivedi B, *et al.* Single cell transcriptomic landscape of diabetic foot ulcers. *Nat Commun.* 2022;13:181. <https://doi.org/10.1038/s41467-021-27801-8>.
140. Chen CJ, Kajita H, Takaya K, Aramaki-Hattori N, Sakai S, Asou T, *et al.* Single-cell RNA-seq analysis reveals cellular functional heterogeneity in dermis between fibrotic and regenerative wound healing fates. *Front Immunol.* 2022;13:875407. <https://doi.org/10.3389/fimmu.2022.875407>.
141. Yakupu A, Zhang D, Guan H, Jiang M, Dong J, Niu Y, *et al.* Single-cell analysis reveals melanocytes may promote inflammation in chronic wounds through cathepsin G. *Front Genet.* 2023;14:1072995. <https://doi.org/10.3389/fgene.2023.1072995>.
142. Celik C, Lee STT, Tanoto FR, Veleba M, Kline K, Thibault G. Decoding the complexity of delayed wound healing following Enterococcus faecalis infection. *Elife.* 2024;13:1–30. <https://doi.org/10.7554/eLife.95113>.
143. Li D, Cheng S, Pei Y, Sommar P, Karner J, Herter EK, *et al.* Single-cell analysis reveals major histocompatibility complex II-expressing keratinocytes in pressure ulcers with worse healing outcomes. *J Invest Dermatol.* 2022;142:705–16. <https://doi.org/10.1016/j.jid.2021.07.176>.
144. Liao B, Ouyang Q, Song H, Wang Z, Ou J, Huang J, *et al.* Characteristic analysis of skin keratinocytes in patients with type 2 diabetes based on the single-cell levels. *Chin Med J.* 2022;135:2461–6. <https://doi.org/10.1097/CM9.0000000000002323>.
145. Parigi SM, Larsson L, Das S, Ramirez Flores RO, Frede A, Tripathi KP, *et al.* The spatial transcriptomic landscape of the healing mouse intestine following damage. *Nat Commun.* 2022;13:828. <https://doi.org/10.1038/s41467-022-28497-0>.
146. Yang Y, Chu C-Y, Man Y, Liu L, Wang C-B, Qu Y-L. Spatial exploration of immune cell heterogeneity in skin functional scarless wound regeneration model at single-cell resolution. *bioRxiv.* 2023 2022.03.24.485669. [10.1101/2022.03.24.485669](https://doi.org/10.1101/2022.03.24.485669).
147. Phan QM, Sinha S, Biernaskie J, Driskell RR. Single-cell transcriptomic analysis of small and large wounds reveals the distinct spatial organization of regenerative fibroblasts. *Exp Dermatol.* 2021;30:92–101. <https://doi.org/10.1111/exd.14244>.
148. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat Commun.* 2021;12:1088. <https://doi.org/10.1038/s41467-021-21246-9>.

149. Hu KH, Kuhn NF, Courau T, Tsui J, Samad B, Ha P, *et al.* Transcriptional space-time mapping identifies concerted immune and stromal cell patterns and gene programs in wound healing and cancer. *Cell Stem Cell.* 2023;30:885–903.e10. <https://doi.org/10.1016/j.stem.2023.05.001>.
150. Guo Z, Saw PE, Jon S. Non-invasive physical stimulation to modulate the tumor microenvironment: unveiling a new frontier in cancer therapy. *BIO Integration.* 2024;5:986. <https://doi.org/10.15212/bioi-2024-0012>.
151. Luo Z, Mei J, Wang X, Wang R, He Z, Geffen Y, *et al.* Voluntary exercise sensitizes cancer immunotherapy via the collagen inhibition-orchestrated inflammatory tumor immune microenvironment. *Cell Rep.* 2024;43:114697. <https://doi.org/10.1016/j.celrep.2024.114697>.
152. Jin S, Ramos R. Computational exploration of cellular communication in skin from emerging single-cell and spatial transcriptomic data. *Biochem Soc Trans.* 2022;50:297–308. <https://doi.org/10.1042/BST20210863>.
153. Abbasi S, Sinha S, Labit E, Rosin NL, Yoon G, Rahmani W, *et al.* Distinct regulatory programs control the latent regenerative potential of dermal fibroblasts during wound healing. *Cell Stem Cell.* 2020;27:396–412.e6. <https://doi.org/10.1016/j.stem.2020.07.008>.
154. Song B, Zheng Y, Chi H, Zhu Y, Cui Z, Chen L, *et al.* Revealing the roles of glycosphingolipid metabolism pathway in the development of keloid: a conjoint analysis of single-cell and machine learning. *Front Immunol.* 2023;14:1139775. <https://doi.org/10.3389/fimmu.2023.1139775>.
155. Song B, Zhu Y, Zhao Y, Wang K, Peng Y, Chen L, *et al.* Machine learning and single-cell transcriptome profiling reveal regulation of fibroblast activation through thbs2/tgfbeta1/p-smad2/3 signalling pathway in hypertrophic scar. *Int Wound J.* 2024;21:e14481. <https://doi.org/10.1111/iwj.14481>.
156. He H, Suryawanshi H, Morozov P, Gay-Mimbrera J, Del Duca E, Kim HJ, *et al.* Single-cell transcriptome analysis of human skin identifies novel fibroblast subpopulation and enrichment of immune subsets in atopic dermatitis. *J Allergy Clin Immunol.* 2020;145:1615–28. <https://doi.org/10.1016/j.jaci.2020.01.042>.
157. Bernstein NJ, Fong NL, Lam I, Roy MA, Hendrickson DG, Kelley DR. Solo: doublet identification in single-cell RNA-seq via semi-supervised deep learning. *Cell Syst.* 2020;11:95–101.e5. <https://doi.org/10.1016/j.cels.2020.05.010>.
158. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol.* 2020;38:1408–14. <https://doi.org/10.1038/s41587-020-0591-3>.
159. Nakamizo S, Dutertre CA, Khalilnezhad A, Zhang XM, Lim S, Lum J, *et al.* Single-cell analysis of human skin identifies CD14+ type 3 dendritic cells co-producing IL1b and IL23a in psoriasis. *J Exp Med.* 2021;218:218. <https://doi.org/10.1084/jem.20202345>.
160. Vu R, Jin S, Sun P, Haensel D, Nguyen QH, Dragan M, *et al.* Wound healing in aged skin exhibits systems-level alterations in cellular composition and cell-cell communication. *Cell Rep.* 2022;40:111155. <https://doi.org/10.1016/j.celrep.2022.111155>.
161. Sole-Boldo L, Raddatz G, Schutz S, Mallm JP, Rippe K, Lonsdorf AS, *et al.* Single-cell transcriptomes of the human skin reveal age-related loss of fibroblast priming. *Commun Biol.* 2020;3:188. <https://doi.org/10.1038/s42003-020-0922-4>.
162. Liu Y, Wang P, Li J, Chen L, Shu B, Wang H, *et al.* Single-cell RNA sequencing reveals the impaired epidermal differentiation and pathological microenvironment in diabetic foot ulcer. *Burns & Trauma.* 2025;13:1–16. <https://doi.org/10.1093/burnst/tkae065>.
163. Magaletta ME, Lobo M, Kernfeld EM, Aliee H, Huey JD, Parsons TJ, *et al.* Integration of single-cell transcriptomes and chromatin landscapes reveals regulatory programs driving pharyngeal organ development. *Nat Commun.* 2022;13:457. <https://doi.org/10.1038/s41467-022-28067-4>.
164. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol.* 2024;42:293–304. <https://doi.org/10.1038/s41587-023-01767-y>.
165. Mascharak S, desJardins-Park HE, Longaker MT. Fibroblast heterogeneity in wound healing: hurdles to clinical translation. *Trends Mol Med.* 2020;26:1101–6. <https://doi.org/10.1016/j.molmed.2020.07.008>.
166. Chen G, Ning B, Shi T. Single-cell RNA-seq technologies and related computational data analysis. *Front Genet.* 2019;10:317. <https://doi.org/10.3389/fgene.2019.00317>.
167. Chu SK, Zhao S, Shyr Y, Liu Q. Comprehensive evaluation of noise reduction methods for single-cell RNA sequencing data. *Brief Bioinform.* 2022;23:1–17. <https://doi.org/10.1093/bib/bbab565>.
168. Gygi JP, Kleinstein SH, Guan L. Predictive overfitting in immunological applications: pitfalls and solutions. *Hum Vaccin Immunother.* 2023;19:2251830. <https://doi.org/10.1080/21645515.2023.2251830>.
169. Lones MA. Avoiding common machine learning pitfalls. *Patterns (N Y).* 2024;5:101046. <https://doi.org/10.1016/j.patter.2024.101046>.
170. Budhkar A, Song Q, Su J, Zhang X. Demystifying the black box: a survey on explainable artificial intelligence (XAI) in bioinformatics. *Comput Struct Biotechnol J.* 2025;27:346–59. <https://doi.org/10.1016/j.csbj.2024.12.027>.
171. Hassija V, Chamola V, Mahapatra A, Singal A, Goel D, Huang K, *et al.* Interpreting black-box models: a review on explainable artificial intelligence. *Cogn Comput.* 2023;16:45–74. <https://doi.org/10.1007/s12559-023-10179-8>.
172. Yu Y, Mai Y, Zheng Y, Shi L. Assessing and mitigating batch effects in large-scale omics studies. *Genome Biol.* 2024;25:254. <https://doi.org/10.1186/s13059-024-03401-9>.
173. Le DTP, Pham TD. Unveiling the role of artificial intelligence for wound assessment and wound healing prediction. *Explor Med.* 2023;4:589–611. <https://doi.org/10.37349/emed.2023.00163>.
174. Zhang H, Lv G, Liu S, Liu D, Wu X. The artificial intelligence watcher predicts cancer risk by facial features. *Tradit Med Res.* 2022;7:1. <https://doi.org/10.53388/TMR20211227255>.
175. Wu Y, Xie L. AI-driven multi-omics integration for multi-scale predictive modeling of genotype-environment-phenotype relationships. *Comput Struct Biotechnol J.* 2025;27:265–77. <https://doi.org/10.1016/j.csbj.2024.12.030>.
176. Zheng Y, Cao T, Han X, Cao P, Zhan Q. Structurally diverse polydopamine-based nanomedicines for cancer therapy. *Acta Mater Med.* 2022;1:427–44. <https://doi.org/10.15212/A-MM-2022-0023>.
177. Ma P, Wang G, Men K, Li C, Gao N, Li L. Advances in clinical application of nanoparticle-based therapy for cancer treatment: a systematic review. *Nano TransMed.* 2024;3:100036. <https://doi.org/10.1016/j.ntm.2024.100036>.
178. Pang Q, Yang F, Jiang Z, Wu K, Hou R, Zhu Y. Smart wound dressing for advanced wound management: real-time monitoring and on-demand treatment. *Mater Des.* 2023;229:111917. <https://doi.org/10.1016/j.matdes.2023.111917>.
179. Luo Z, Zhang T, Chen S. Exercise prescription: pioneering the “third pole” for clinical health management. *Research.* 2023;6:0284. <https://doi.org/10.34133/research.0284>.
180. Rehman AU, Li M, Wu B, Ali Y, Rasheed S, Shaheen S, *et al.* Role of artificial intelligence in revolutionizing drug discovery. *Fundamental Research.* 2024;3:1273–1287. <https://doi.org/10.1016/j.fmre.2024.04.021>.