High-Resolution Spatial Map of the Human Facial Sebaceous Gland Reveals Marker Genes and Decodes Sebocyte Differentiation

JID Open

Tolga Düz[1,2], Daniel Torocsik[3,4], Annika Simmering[1], Peter Wolf[1], Stefan Gallinat[1], Jan Baumbach[2,5] and Nicholas Holzscheck[1]

Journal of Investigative Dermatology (2026) **146,** 40–54; doi:10.1016/j.jid.2025.04.041

The sebaceous gland (SG) is essential for skin homeostasis by producing sebum to lubricate and protect the skin. Dysfunctions in SG activity are associated with skin disorders such as acne, seborrheic dermatitis, and alopecia. However, its cellular and molecular mechanisms in humans remain poorly understood because most studies have been conducted in mouse models. This study provides a comprehensive molecular analysis of the human SG, focusing on cellular interactions; sebocyte differentiation; and to our knowledge, previously unreported gene markers. By integrating Stereo-seq spatial transcriptomics, single-cell RNA sequencing, and validation by multiplexed error-robust FISH, we identified 4 distinct stages of sebocyte differentiation, each characterized by unique gene signatures. These results reveal that sebocyte differentiation is a dynamic and complex process. Our findings enhance the understanding of SG biology and provide a valuable reference for future research and the development of therapies for SG-related disorders, including acne.

Keywords: Acne, Pilosebaceous unit, scRNA-seq, Single-cell RNA sequencing, Stereo-seq

INTRODUCTION

The skin, our largest organ, serves as a protective barrier for the body. The sebaceous gland (SG), typically associated with hair follicles, combines with arrector pili muscles to form the pilosebaceous unit (PSU) (Figure 1b) (Schneider et al, 2009). Despite major advances, critical gaps remain in the understanding of the cellular and molecular mechanisms of the SG in humans, which are essential to address both their physiological role and dysregulated pathways in skin diseases. Indeed, dysregulation in the SG activity is linked to numerous human-specific skin disorders, such as acne (Dréno, 2017), seborrheic dermatitis (Gupta and Bluhm, 2004), and cicatricial alopecia (Sundberg et al, 2018), which have significant clinical and cosmetic implications (Hazarika and Archana, 2016).

The SG has an acinar structure, with its outermost layer composed of basal cells that undergo radial differentiation toward the SG's interior. Throughout this process, lipids begin to accumulate, and sebocytes near the sebaceous duct (SD), the opening that connects the SG to the infundibulum, ultimately burst (Figure 1b) (Fischer et al, 2017; Schneider and Paus, 2010; Veniaminova et al, 2023). This process, known as holocrine secretion, releases sebum to lubricate and protect the skin (Zouboulis, 2017). Such continuous turnover of sebocytes relies on basal stem cells to provide a steady supply of new cells.

Studying the SG remains a challenge owing to its structural and cellular heterogeneity. Single-cell RNA sequencing (scRNA-seq) (Kolodziejczyk et al, 2015) provides valuable insights into SG biology but is limited by its reliance on intact cells, making sebocytes difficult to analyze due to their large size, lipid-filled cytoplasm, and tendency to rupture during holocrine secretion (Veniaminova et al, 2023). As a result, sebocytes are rarely present or only represented as a small subset in skin scRNA-seq datasets. Furthermore, scRNA-seq does not capture the spatial context, which is crucial for understanding the organization and function of the SG. Spatial transcriptomics offers a transformative approach to overcome these challenges (Rao et al, 2021). Unlike scRNA-seq, this technique preserves the spatial context of cells, allowing for a more comprehensive analysis of sebocyte differentiation. By sequencing spatial spots, this method captures large, lipid-filled sebocytes that are often missing from scRNA-seq.

For this study, we obtained facial skin samples from healthy donors and then leveraged the combined power of spatial transcriptomics and scRNA-seq for a detailed molecular dissection of the human SG (Figure 1). Although sebocyte differentiation has been described in mouse models (Veniaminova et al, 2023), recent findings

[1]*R&D Discovery, Beiersdorf AG, Hamburg, Germany;* [2]*Institute for Computational Systems Biology, University of Hamburg, Hamburg, Germany;* [3]*Department of Dermatology, MTA Centre of Excellence, Faculty of Medicine, University of Debrecen, Debrecen, Hungary;* [4]*HUN-REN-UD Allergology Research Group, University of Debrecen, Debrecen, Hungary; and* [5]*Computational BioMedicine Lab, Institute of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark*

Correspondence: Nicholas Holzscheck, R&D Discovery, Beiersdorf AG, Beiersdorfstraße 1-9, Hamburg 20245, Germany. E-mail: nicholas.holzscheck@beiersdorf.com

Abbreviations: CyPA, cyclophilin A; IFE, interfollicular epidermis; JZ, junctional zone; K, keratin; KC, keratinocyte; MERFISH, multiplexed error-robust FISH; PHATE, Potential of the Heat-diffusion for Affinity-based Transition Embedding; PSU, pilosebaceous unit; scRNA-seq, single-cell RNA sequencing; SD, sebaceous duct; SG, sebaceous gland; ssDNA, single-stranded DNA

Received 2 February 2025; revised 28 March 2025; accepted 15 April 2025; accepted manuscript published online 29 May 2025; corrected proof published online 5 July 2025

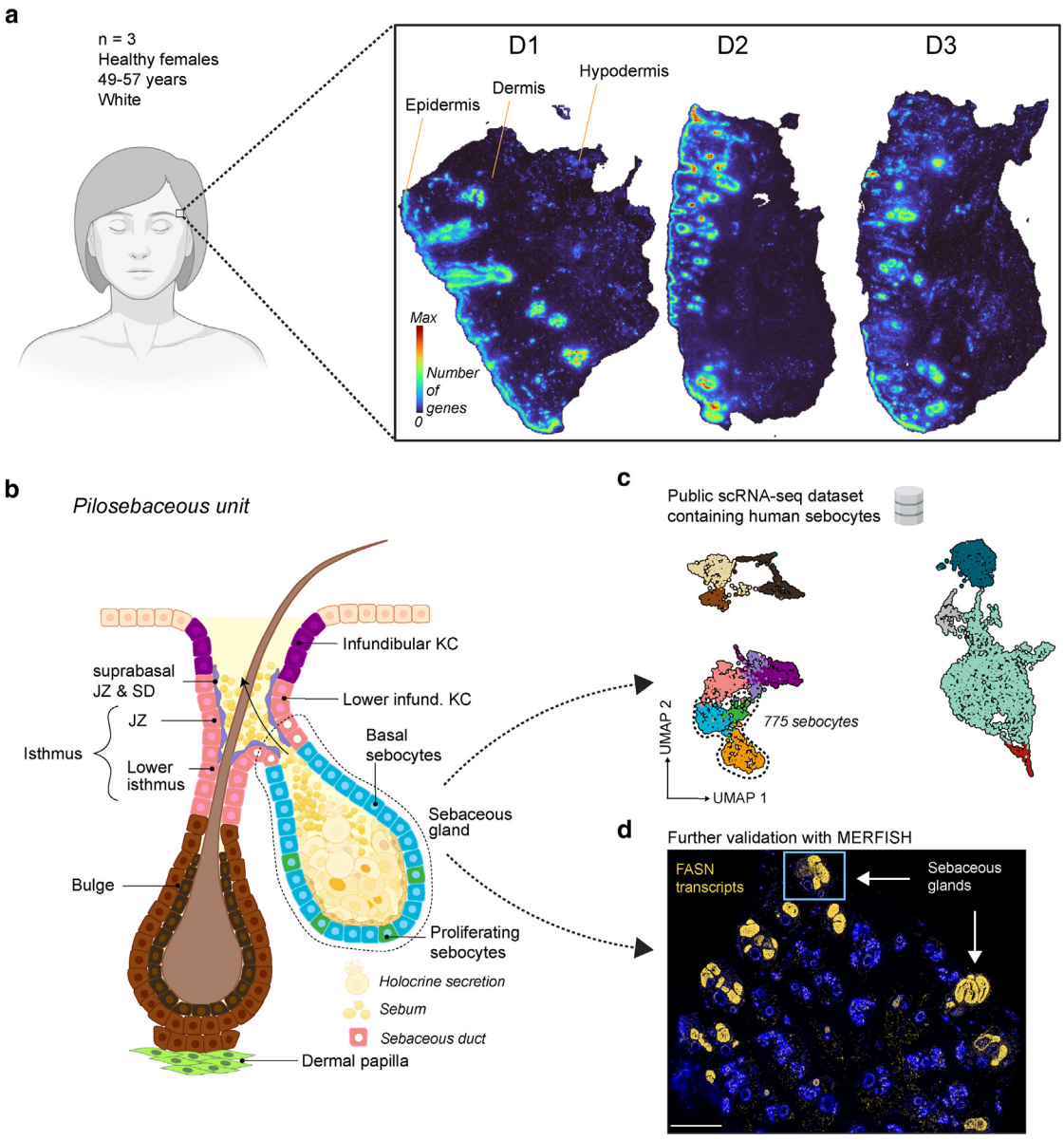


Figure 1. Study design integrating spatial transcriptomics, scRNA-seq, and MERFISH validation. (**a**) Spatial visualization of the number of genes across 3 tissue sections (D1, D2, and D3) from the temporal region of healthy female White donors' facial skin, using a bin size of 20 (~10 μm). (**b**) Schematic representation of the telogen PSU and its associated cell types. The arrector pili muscle is not depicted. The SG differentiates from basal sebocytes (shown in blue) toward the sebaceous duct, where sebocytes undergo holocrine secretion, releasing sebum. (**c**) Public human scRNA-seq dataset containing sebocytes (GSE147424). (**d**) MERFISH section was visualized by FASN transcripts. Bar = 1 mm. MERFISH, multiplexed error-robust FISH; PSU, pilosebaceous unit; scRNA-seq, single-cell RNA sequencing; SG, sebaceous gland; UMAP, Uniform Manifold Approximation and Projection.

(Thalheim and Schneider, 2024) indicate that sebocytes in mice and humans share limited similarities, highlighting the importance of human-specific studies. Among other findings, Thrane et al (2023) reconstructed the PSU by integrating 4 human scRNA-seq datasets and identified 2 SG clusters, which were not further characterized. Another study used 10x Genomics Visium to investigate human sebocyte differentiation, but the low resolution of this technology limited the analysis to hyperplastic SGs (Schmidt et al, 2024). To overcome this limitation, we used Stereo-seq (Chen et al, 2022), which captures whole transcriptional profiles at cellular resolution. It utilizes DNA nanoball chips and in situ sequencing, offering a smaller spot size (~200 nm) and tighter spacing between spots (500 nm) than other methods (Chen et al, 2022). This enabled us to comprehensively analyze SG differentiation, identify key stages, and pinpoint gene markers that may have implications for skin disorders associated with disrupted SG function. We validated the markers from the stages of sebocyte differentiation using multiplexed error-robust FISH (MERFISH) (Xia et al, 2019). In addition, we examined spatially resolved cell–cell communication within the SG, shedding light on how different cell populations interact to regulate SG function.

RESULTS

High-resolution spatial transcriptomics of facial skin reveals transcriptomic architecture of different skin cell types

We analyzed sections from 3 tissue samples of healthy female facial skin from different White donors (aged 49–57 years),

focusing on the temporal region, using Stereo-seq (Chen et al, 2022) (Figure 1a). The single-stranded DNA (ssDNA) staining of the respective sections is shown in Supplementary Figure S1. Stereo-seq uses a binning approach, where spots are grouped into virtual squares representing cells. We explored different binning sizes (Supplementary Figures S2–4), which affect key parameters such as the number of genes detected per cell. A larger binning size increases the number of detected genes but reduces the resolution because multiple cells are grouped together and vice versa. Because the focus of this study is the SG, we benefit from its acinar, isolated structure and the enlargement of sebocytes during differentiation. The former allows similar cell types to cluster together, whereas the latter justifies a binning size that prioritizes higher gene coverage over exact single-cell resolution. Visual evaluation using Uniform Manifold Approximation and Projection suggests that a binning size of 50 results in the most distinct separation of cell types (Supplementary Figure S2a–c). In contrast, a binning size of 20 results in lower gene coverage (Supplementary Figure S3), leading to the absence of certain cell types, such as those from the bulge and dermal papilla, and causing clusters to become significantly blurred owing to insufficient relevant features for clear separation. Conversely, although a binning size of 100 yields higher gene coverage, it compromises spatial resolution and results in mixed cell types. These trends are confirmed by the spatial visualizations (Supplementary Figure S2d–f), dot plots (Supplementary Figure S4), and the differentially expressed gene lists in Supplementary Tables S1–3, which show that more relevant features for the respective cell types can be identified when using a binning size of 50. Consequently, we ultimately chose a binning size of 50 (~25 μm cell diameter) for data analysis, which has also been used in other Stereo-seq studies (Chen et al, 2022; Cheng et al, 2022; Liao et al, 2023[1]; Wu et al, 2023).

After filtering out low-quality cells, a total of 9374 cells were obtained. The average number of genes captured per cell ranged from 1448 in tissue section D1 to 1081 in tissue section D2 and 965 in tissue section D3. Figure 2a shows the integrated Uniform Manifold Approximation and Projection (McInnes et al, 2018[2]) of cell-type clusters derived from samples D1, D2, and D3, whereas Figure 2b–d shows the localization of these clusters in the tissue sections.

These clusters are subdivided into the following groups: granular keratinocytes (KCs), spinous KCs, basal KCs, infundibular KCs, lower infundibular KCs, junctional zone (JZ), suprabasal JZ, SD, bulge, outer root sheath, inner root sheath, dermal papilla, dermal sheath, basal sebocytes, sebocytes, bursted sebocytes, vascular endothelial cells, vascular smooth muscle cells, mast cells, fibroblasts, adipocytes, B cells, eccrine sweat glands, and immune cells. A dot plot with the marker genes for these cell types can be found in Supplementary Figure S4b.

Next, we applied the Hotspot method to functionally characterize the cell-type clusters (DeTomaso and Yosef, 2021). This analysis identified several gene modules, which are groups of coexpressed genes with shared functions (Supplementary Figure S5a) that aligned strongly with the cell-type clusters from Figure 2a. Spatially visualized gene modules (Supplementary Figure S5b) are associated with the interfollicular epidermis (IFE) (M2), infundibular KCs (M8), SG (M1), bulge (M14), outer root sheath (M6), and immune cells (M4). Interestingly, for gene module M8, which spatially overlaps with the infundibular KCs, enrichment was found not only for the term "KC differentiation" but also for "antimicrobial humoral response" and "leukocyte chemotaxis" (Supplementary Figure S5a). In contrast, module M2, which represents features of IFE, showed no enrichment for any immune-related terms. Module M6 is expressed where the outer root sheath is present and correlates with terms related to the hair cycle and hair follicle development. The bulge-associated module M14 shows enrichment for Wnt signaling and neuron-related terms (Supplementary Figure S5a and b). M1 is involved in lipid biosynthesis, and M4 is enriched for immune-regulated pathways.

In the 10 SGs from the 3 sections, basal sebocytes were located at the edge of the SG, whereas differentiating sebocytes were in the center (Figure 2b–d). The yellow cluster likely represents remnants of burst sebocytes because these spots correspond to areas where holocrine secretion occurs, namely near the SD (Fischer et al, 2017). ssDNA staining revealed nearby areas with enlarged sebocytes without nuclei or entirely cell-free regions, suggesting remnants of holocrine secretion (Supplementary Figure S6). The suprabasal cells of the SD and the JZ were characterized by the expression of keratin (K)6A, CST6, and K79 (Supplementary Figures S6 and S7a) (Oulès et al, 2020). LRIG1, the most prominent marker for the JZ, shows no specific clustering for this region in either the scRNA-seq or Stereo-seq data (Supplementary Figure S7b). Although this marker is valid in mice (Jensen et al, 2009), it appears to be imprecise in humans. GATA6 has been described as a specific marker for the lower infundibulum, JZ, the SD, and upper SG (Oulès et al, 2020). LGR6 expression could also be observed at the transition between the JZ and basal sebocytes (Supplementary Figure S7b and c). Overall, the JZ clusters exhibited heterogeneity, encompassing various cell fates, including the SD. Although our analysis captures this complexity, it does not specifically resolve which subpopulation might give rise to sebaceous progenitors.

A notable observation was the location of immune cells, around both the SG and the infundibular KCs. To characterize their composition, we generated spatial feature plots for a variety of immune cell markers (Supplementary Figure S8a) and detected a diverse range of immune cells. Langerhans cells are predominantly located in the upper hair follicle and extend toward the SG, consistent with their established role in antigen presentation and host defense (Nagao et al, 2012). Macrophages and dendritic cells were also present, supporting their involvement in inflammatory responses. In addition, the presence of cytotoxic and regulatory T cells suggested active immune modulation within the PSU, potentially balancing immune surveillance with the immune-

[1] Liao S, Heng Y, Liu W, Xiang J, Ma Y, Chen L, et al. Integrated spatial transcriptomic and proteomic analysis of fresh frozen tissue based on stereo-seq. bioRxiv 2023.
[2] McInnes L, Healy J, Umap MJ. Uniform manifold approximation and projection for dimension reduction. arXiv 2018.

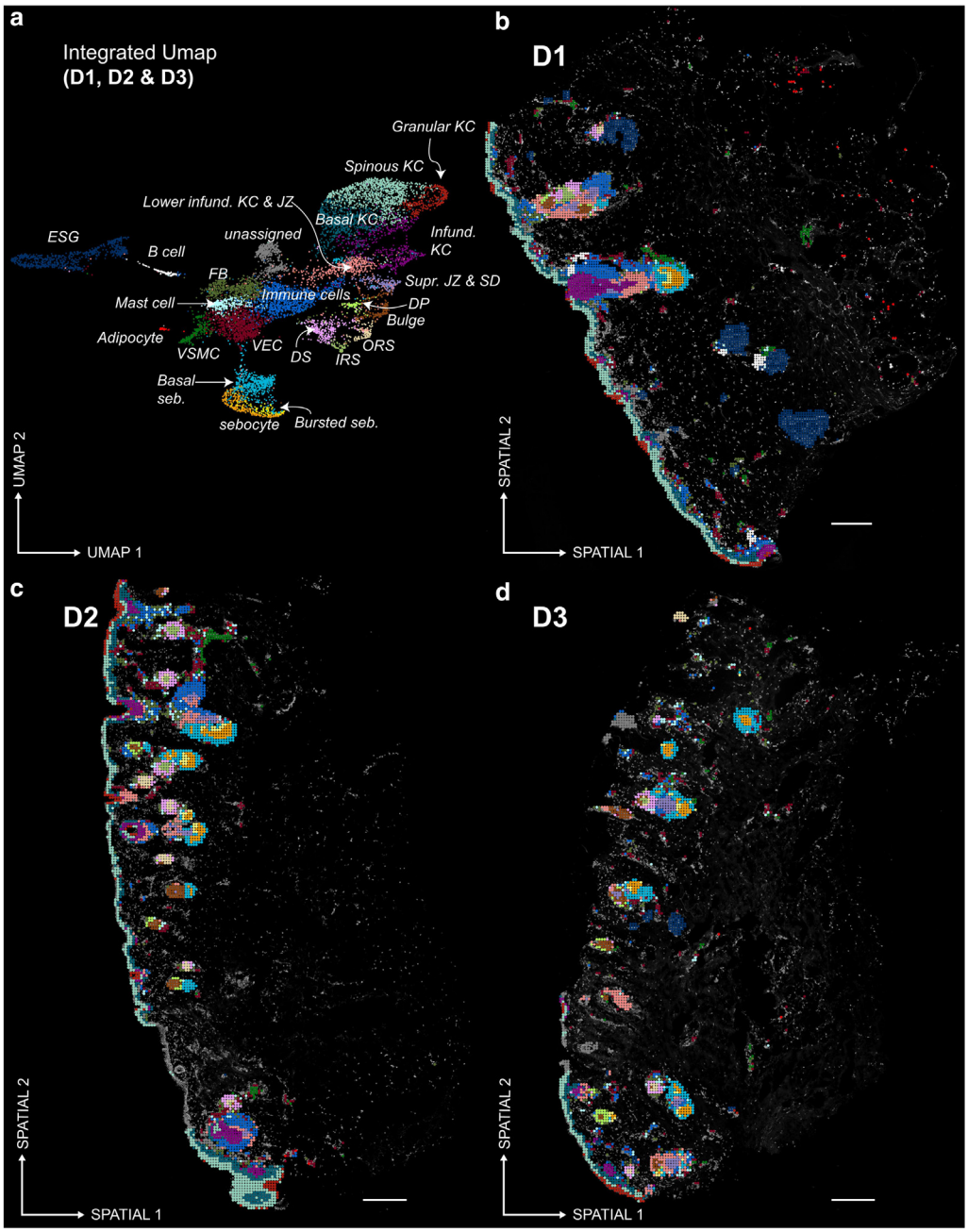


Figure 2. Integrated spatial analysis of human facial skin tissue sections. (**a**) Integrated UMAP of Stereo-seq data from 3 tissue sections, clustered with cell-type annotations, with a bin size set to 50 (~25 μm). seb denotes sebocyte, and Infund. denotes infundibular. (**b**–**d**) Projection of cell types onto ssDNA staining of tissue sections D1, D2, and D3. Bar = 200 μm. DP, dermal papilla; DS, dermal sheath; ESG, eccrine sweat gland; FB, fibroblast; IRS, inner root sheath; JZ, junctional zone; KC, keratinocyte; ORS, outer root sheath; SD, sebaceous duct; ssDNA, single-stranded DNA; Supr. JZ, suprabasal junctional zone; UMAP, Uniform Manifold Approximation and Projection; VEC, vascular endothelial cell; VSMC, vascular smooth muscle cell.

privileged status of the hair follicle. These findings are consistent with previous studies highlighting the hair follicle and SG as key sites of immune interaction (Christoph et al, 2000).

Spatially resolved cell–cell communication analysis identifies signaling pathways in the SG

We applied CellChat to our Stereo-seq data from human facial skin and identified the signaling pathways that are most active in the SG (Figure 3a) (Jin et al, 2024).

The SG showed a unique communication through cyclophilin A (CyPA) and ANGPTL (Dahlhoff et al, 2014). Although CyPA is the most dominant autocrine signaling pathway in the SG, it has not been described in this context.

In addition, CD6 crosstalk, secreted by immune cells and exclusively received by basal sebocytes through ALCAM, represents, to our knowledge, a previously unreported interaction (Figure 3b and c). The infundibular KCs showed strong signaling activity for the macrophage migration inhibitory factor and amyloid precursor protein pathways, with the immune cell cluster being the primary receiver for both (Figure 3d). In general, immune cells are the primary target of macrophage migration inhibitory factor signaling from other cell types within the PSU (Figure 3e).

The expression of the ligands and receptors of the respective signaling pathway is visualized in tissue section D1 (Supplementary Figure S9). The spatial expression of ligands and receptors is in line with expectations based on the analyses. However, the CyPA pathway showed widespread expression of its ligand, PPIA, in almost all cell types, whereas its receptor, BSG, was primarily localized in sebocytes.

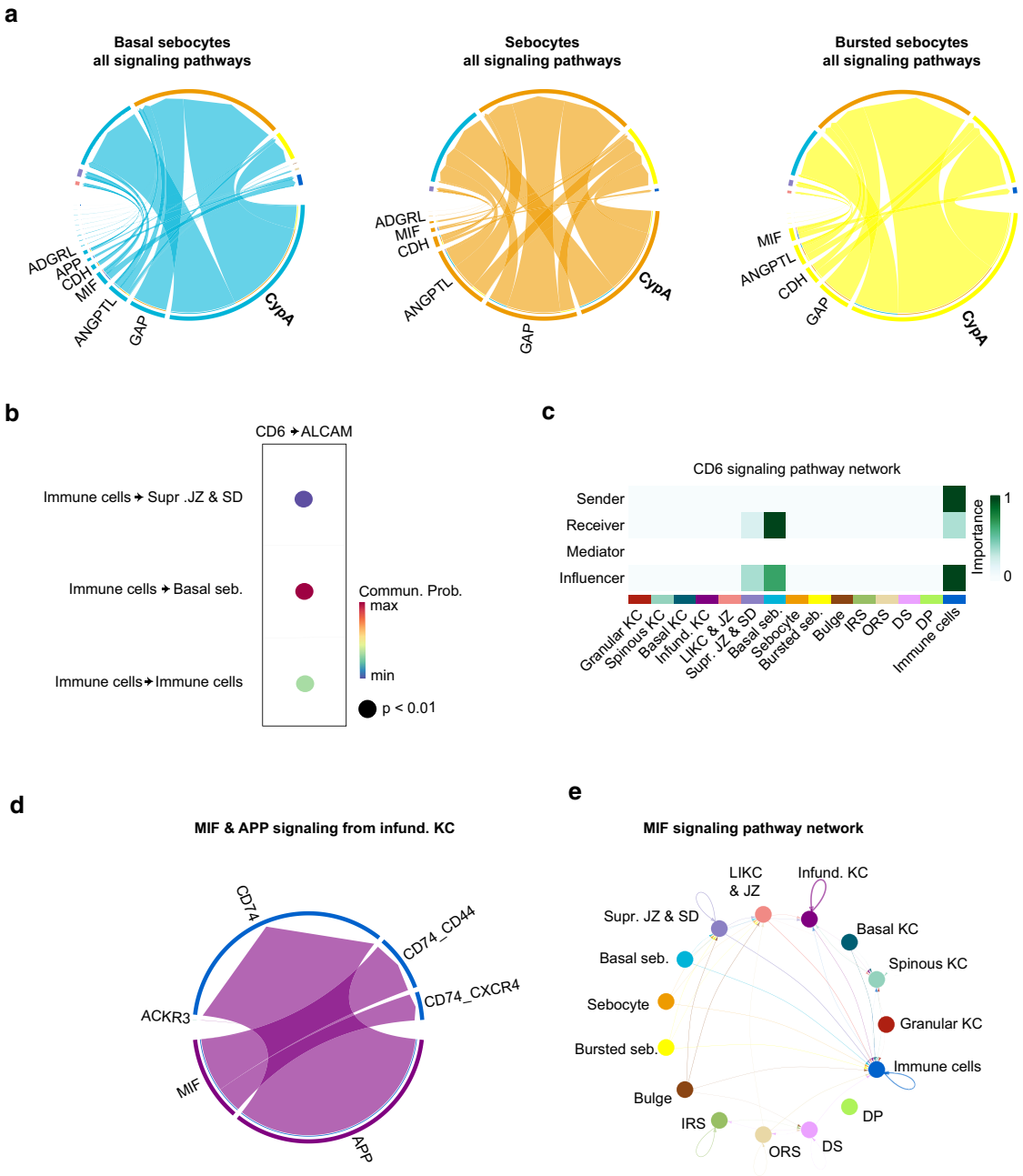


Figure 3. Spatial crosstalk analysis of facial human sebaceous glands. (**a**) Chord diagram of all significant signaling pathways active in the SG. (**b**) Detected communication probability between CD6 ligand and ALCAM receptor. (**c**) Heatmap of CD6 signaling pathway network importance. (**d**) Chord diagram of ligand–receptor signaling interactions of the MIF and APP pathways originating from infundibular keratinocytes, excluding autocrine signaling. (**e**) Circle plot of all cell types involved in the MIF signaling pathway. seb. denotes sebocyte. APP, amyloid precursor protein; DP, dermal papilla; DS, dermal sheath; IRS, inner root sheath; KC, keratinocyte; LIKC, lower infundibular keratinocyte; MIF, macrophage migration inhibitory factor; ORS, outer root sheath; SD, sebaceous duct; SG, sebaceous gland; Supr. JZ, suprabasal junctional zone.

Identification of gene markers for basal sebocytes and infundibular KCs

We reconstructed the entire human PSU with single-cell resolution and high gene coverage using healthy skin from the dataset GSE147424 (He et al, 2020) (Figure 4a and Supplementary Figure S10).

We projected the scRNA-seq dataset onto a subset of the Stereo-seq dataset corresponding to PSU and IFE cell types to assign the most likely cell types to each spot (Figure 4b) (Stuart et al, 2019). The prediction accuracy showed an almost perfect alignment of cell types between datasets (Supplementary Figure S11), demonstrating the robustness of our analyses. When projecting the cluster for infundibular KCs onto the tissue sections, a high prediction accuracy and a focused expression in the infundibulum were observed (Figure 4c). The SG was also accurately predicted (Figure 4c).

For infundibular KCs or upper hair follicles, K79 is a known marker (Joost et al, 2016). However, it performed poorly compared with the markers S100A7, S100A8, S100A9, TYMP, and SERPINB4 shown in the violin plot (Figure 4d).

Known gene markers from the literature are shown in bold, whereas those not in bold are, to the best of our knowledge, newly identified, because no prior references were found.

Basal sebocytes are located at the edge of the SG, for which K7 is a known marker (de Bengy et al, 2019). However, similar to PPARG, K7 does not exclusively separate basal from mature sebocytes (Figure 4e). On the basis of our data, more adequate markers were NNAT, IL-1R2, TINAGL1, and WFDC2. Notably, the only difference we observed between proliferating and nonproliferating basal sebocytes was limited to cell cycle markers

Unravelling the differentiation of stem cell niches within the JZ

Differentiation occurs along a continuum, a characteristic not well-captured by Uniform Manifold Approximation and Projection embeddings, which tend to compress cells into discrete clusters (Figure 3a) (Moon et al, 2019). The Potential of the Heat-diffusion for Affinity-based Transition Embedding (PHATE) algorithm has been shown to be effective in analyzing differentiation, capturing both local and global structures robustly and at scale (Moon et al, 2019). In this study, we applied PHATE to the scRNA-seq dataset (Figure 4a) and the Stereo-seq dataset using the PSU and IFE (Figure 4b) and observed a strikingly similar differentiation trajectory across them, despite the different datasets and profiling methods used.

We applied RNA velocity to predict cell states on the basis of the relative abundance of spliced and unspliced mRNA, providing unbiased insight into the direction and speed of cellular transitions (Bergen et al, 2021). Velocity analyses revealed a bipotent trajectory from the JZ toward both the SG and the infundibulum (Jensen et al, 2009) (Figure 5c and d). Within the infundibulum (Figure 5d, upper panel), cells terminally differentiated inward and desquamate. Interestingly, the differentiation of the SG did not follow a radial pattern from all sides of the basal membrane toward the interior of the gland. Instead, it progressed from the most apical point of the SG toward the SD. The observed anomaly in the velocity vectors within the granular KCs (Figure 5c) was likely due to cornification, a process characterized by reduced transcriptional activity and RNA degradation (Eckhart et al, 2013), which may disrupt the assumptions underlying RNA velocity analysis (Bergen et al, 2021).

Cells within the SGs are grouped according to 4 distinct gene clusters

To understand the genes responsible for differentiation, we performed a pseudotime analysis on both scRNA-seq and Stereo-seq datasets. Pseudotime algorithms arrange cells along a virtual timeline on the basis of their transcriptomic similarity and assign them a pseudotime value, which reflects the differentiation state of a cell relative to that of its progenitor (Cao et al, 2019). For this analysis, the JZ and basal KC clusters were used as starting points. We observed clear pseudotime trajectories in both datasets (Figure 5e and f), consistent with the velocity vectors.

Next, we used CellRank to identify putative sebocyte lineage driver genes from the pseudotime analysis (Lange et al, 2022). This allowed us to identify a total of 5 distinct gene clusters, 4 of which belong to the SG, which we functionally annotated through gene set enrichment analysis (Figure 6a). The first cluster, located in the JZ, exhibited stem cell characteristics, including downregulated cell differentiation and motility. This was followed by a cluster of genes expressed by basal sebocytes (designated as SEB-B), where protein biosynthesis processes are upregulated. The next cluster represented a transitional state (SEB-T) between basal and lipid-filled sebocytes. It was characterized by signals of hormonal sensitivity and differentiation. In SEB-1, we observed lipid biosynthesis and sebum production. SEB-2 was distinguished not only by lipid and steroid biosynthesis but also by catabolic processes. Anticipating that single-cell data may not provide a complete description of sebocyte differentiation owing to the inability to capture large lipid-filled sebocytes or cells that are no longer intact during holocrine secretion, we further analyzed the sebocyte cluster from the Stereo-seq dataset using Gene Ontology enrichment analysis (Figure 6b). The incomplete characterization of SEB-2 by scRNA-seq was also indicated by the gradual increase in gene cluster expression, which had not yet reached its peak. This analysis revealed previously unreported terms, and we hypothesized that SEB-2 represented the final stage of sebocyte differentiation in its entirety. Enriched terms were particularly associated with increased energy metabolism, apoptosis, and autophagy. Interestingly, we also identified terms related to detoxification, the endoplasmic reticulum–associated degradation pathway, gluconeogenesis, COPII-coated vesicle binding, and the GPI anchor biosynthetic process.

Independent validation of sebocyte differentiation signatures using MERFISH

We employed MERFISH, an RNA imaging-based technique using FISH, known for its high sensitivity and robustness against errors to validate the results of the scRNA-seq and Stereo-seq analyses. Instead of the conventional longitudinal sectioning, which typically includes the epidermis, dermis, and hypodermis, we strategically applied a tangential cut to capture as many SGs as possible (Figure 7a) to confirm our postulated sebocyte differentiation stages. The gene panel for our MERFISH analyses included driver genes identified in the scRNA-seq sebocyte differentiation analysis as well as additional markers of interest relevant to SG biology and general skin function. The full gene panel of 278 genes is detailed in Supplementary Table S4. Quality control of the transcripts from the SG lobes is shown in Supplementary Figure S16.

Overall, both Stereo-seq and MERFISH data revealed a remarkable trend of mutually concordant gene expression across the different stages (Figure 7b). NNAT and IL1R2 served as markers for the SEB-B stage, with their expression localized to the basement membrane of the SG (Figure 7b and Supplementary Figure S12). K7, WFDC2, and TBC1D4 characterized transitional sebocytes. HAO2 and ACO1 were expressed by SEB-1 sebocytes and cluster as the first group not at the basement membrane but exclusively within the interior of the SG. The expression of PNPLA3 and PLIN5 in SEB-2 indicated a gradual transition from SEB-1 gene expression further into the SG. Whereas scRNA-seq captured SEB-2 only up to markers such as PNPLA3 and PLIN5, Stereo-seq enabled the identification of the final stage of SEB-2 differentiation, characterized by

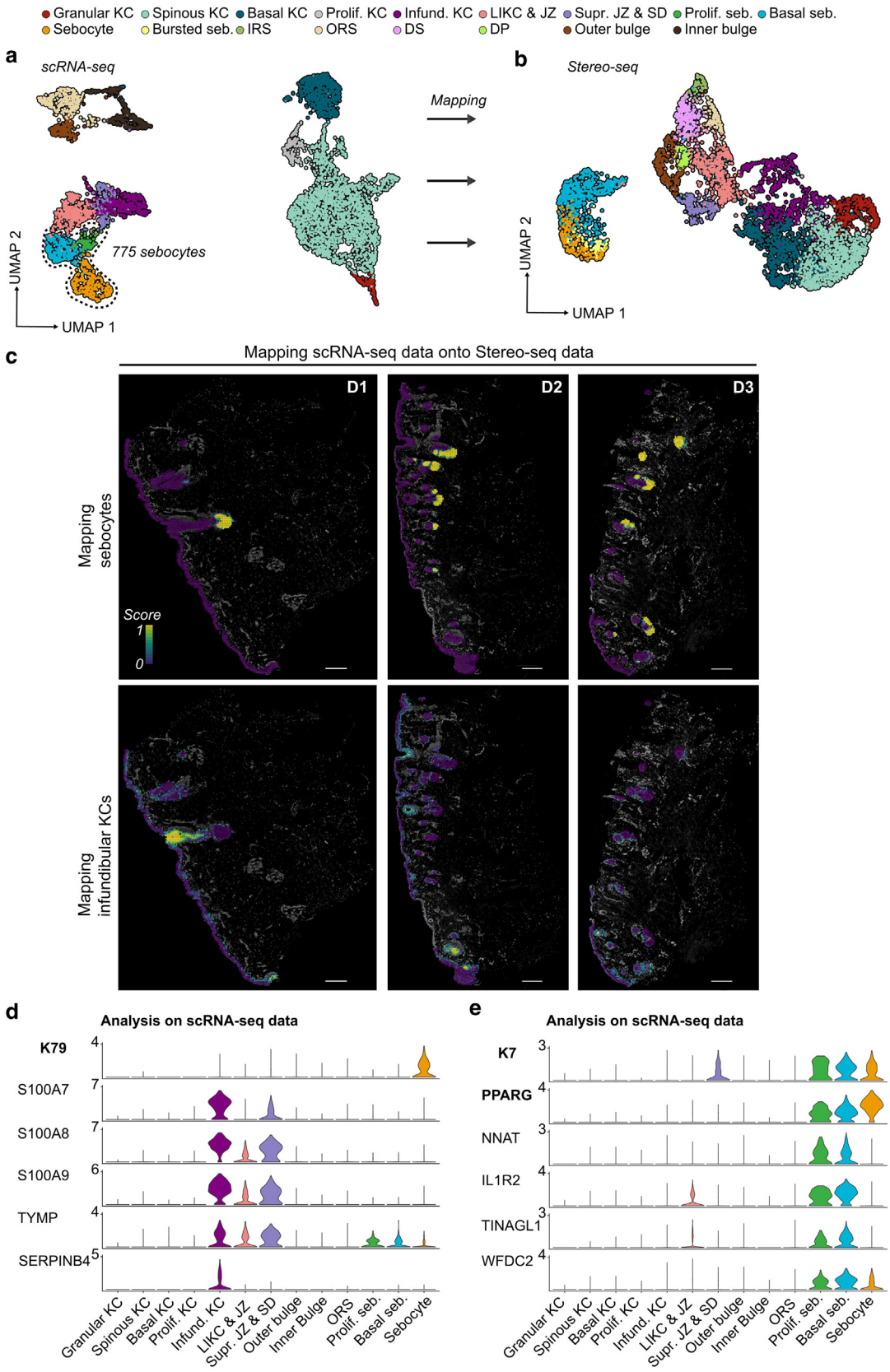


Figure 4. Identification of gene markers for basal sebocytes and infundibular keratinocytes. (**a**) Reclustering of the epidermal cluster (Supplementary Figure S10b) revealed structural regions corresponding to the PSU and IFE. (**b**) Reclustering of the IFE and PSU cell types from Stereo-seq (Figure 2a). (**c**) The probabilistic transfer score for infundibulum KCs shows enrichment in the infundibulum and for sebocytes in the SG. Bar = 200 μm. (**d, e**) Marker gene expression is shown as violin plots, with cell types indicated above and gene markers listed on the left. Bold markers are literature known, whereas others, to our knowledge, are previously unreported. IFE, interfollicular epidermis; KC, keratinocyte; PSU, pilosebaceous unit; SG, sebaceous gland.

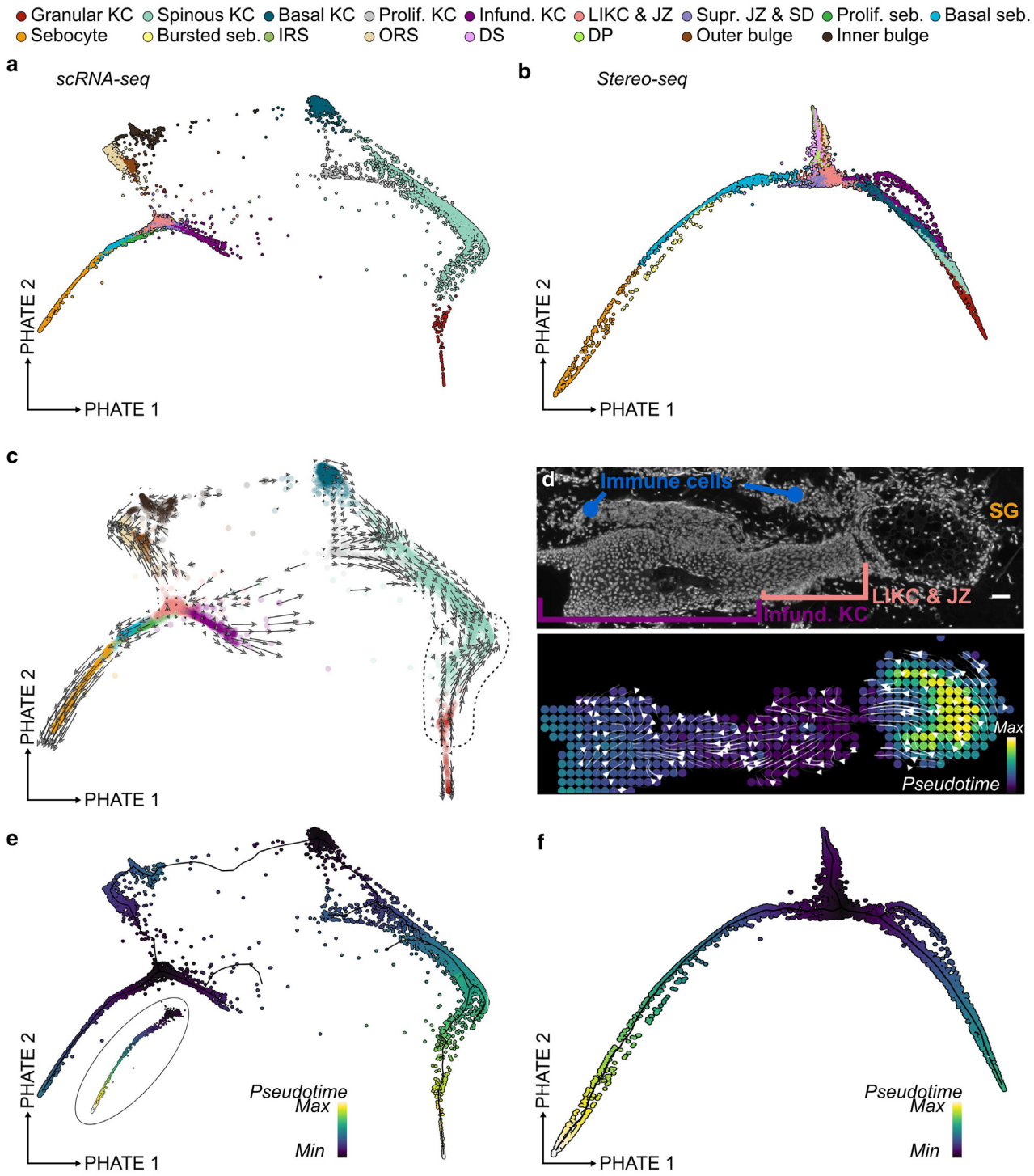


Figure 5. Investigating differentiation in the SG using RNA velocity and pseudotime analysis. (**a**) The PHATE algorithm was applied to the scRNA-seq dataset from Figure 4a, using parameters comparable with the RunUMAP function in Seurat (knn and npca both set to 30). (**b**) PHATE was applied to the Stereo-seq data from Figure 4b. (**c**) RNA velocity analysis was performed using DeepVelo on the same PHATE embedding from **a**. (**d**) Upper panel: ssDNA staining of the SG and upper hair follicle from section D1, annotated with cell-type names. Bar = 25 μm. Lower panel: spatial RNA velocity for the corresponding section from the upper panel, visualized using the dynamo framework and colored on the basis of pseudotime calculated in **f**. (**e**) Pseudotime analysis using Monocle3, with the JZ and basal KCs as designated starting clusters. In the circular plot of the sebocyte lineage, the same pseudotime is shown but without the dominant influence of the longer IFE lineage. Because the uniform scale in the initial plot compresses the sebocyte lineage, its distinct progression is obscured. (**f**) Pseudotime analysis using Monocle3, designating the lower infundibular KCs and JZ cluster as the starting cluster. seb. denotes sebocyte. DP, dermal papilla; DS, dermal sheath; IFE, interfollicular epidermis; IRS, inner root sheath; JZ, junctional zone; KC, keratinocyte; LIKC, lower infundibular keratinocyte; ORS, outer root sheath; PHATE, Potential of the Heat-diffusion for Affinity-based Transition Embedding; scRNA-seq, single-cell RNA sequencing; SD, sebaceous duct; SG, sebaceous gland; ssDNA, single-stranded DNA; Supr. JZ, suprabasal junctional zone.

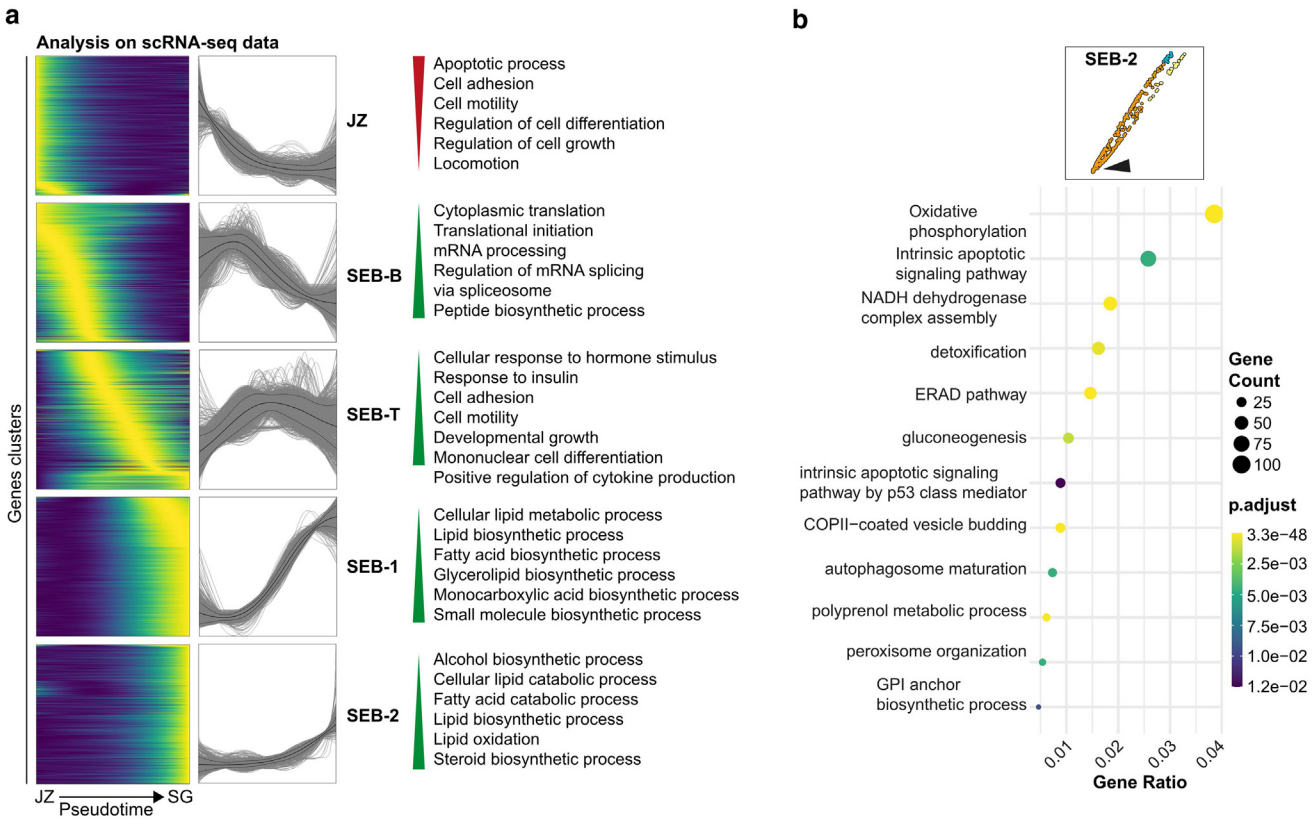


Figure 6. Cells within the SGs are grouped on the basis of 4 distinct gene clusters. (**a**) Heatmap illustrating the gene clusters associated with cell-fate divergence from the JZ to the SG. CellRank was used, and GSEA was conducted for each gene cluster, with the top functional enrichments displayed on the right. (**b**) GO enrichment plot for the sebocyte cluster from the Stereo-seq dataset in Figure 5b. GO, Gene Ontology; GSEA, gene set enrichment analysis; ERAD, endoplasmic reticulum−associated degradation pathway; JZ, junctional zone; SG, sebaceous gland.

the expression of SEC14L6, DNASE1L2 (Fischer et al, 2017), and CRAT in SEB-3 just before the SD. Figure 8 summarizes the identified marker genes for the different stages.

DISCUSSION

Advances in bioinformatic analysis of high-throughput techniques, such as scRNA-seq and spatial transcriptomics (Rao et al, 2021), have significantly enhanced our understanding of complex tissues. In this study, we used the state-of-the-art spatial transcriptomics method Stereo-seq (Chen et al, 2022) to capture spatial gene expression with remarkable resolution. Combined with scRNA-seq, this provided deeper gene coverage for individual cell types. By leveraging both methods along with advanced computational analyses, we were able to unravel the intricate complexities of the SG and thereby improve our understanding of this fascinating structure.

An important finding is the distinction between KCs from the infundibulum and those from the IFE based on transcriptomic signatures. Although both share a primary function, infundibular KCs are strongly involved in inflammatory pathways. An accumulation of immune cells around the infundibulum was also observed. This fits with the concept of "PSU immune privilege" that was first characterized by Paus (Christoph et al, 2000), which is essential to prevent (auto) immune responses against (auto)antigens under physiological conditions. Loss of this immune privilege underlies diseases such as alopecia areata (Bertolini et al, 2020). We also identified various immune cells surrounding infundibular KCs, whereas the chemotaxis of these cells may indicate them as the primary source for signaling molecules released by the PSU that could attract immune cells. Indeed, macrophage migration inhibitory factor signaling is initiated by nonimmune cells such as KCs and attracts immune cells (Hsieh et al, 2014). Owing to the binning size of 50 (~25 μm), it is inevitable that in the peripheral region of the infundibulum, close to the immune cells, some cells are defined that exhibit markers from both infundibular KCs and immune cells. However, the ssDNA section (Supplementary Figure S1a) shows that the infundibulum is not cut off internally, indicating that the mixing effect can be considered relatively insignificant. Furthermore, the scRNA-seq data from the infundibulum showed the same antimicrobial gene markers (Figure 4d).

A recent study (Acosta et al, 2023) showed that viable skin bacteria such as *Cutibacterium acnes* are predominantly found in hair follicles owing to a protective environment. This could trigger a baseline immune activation that increases susceptibility to inflammation upon colonization. However, our analysis of toll-like receptors shows low level of expression (Supplementary Figure S8b) within the PSU. This suggests that immune cells around the infundibulum act as a precaution, ready to respond to infection, whereas toll-like receptor expression may be altered in inflammatory conditions such as acne. A further understanding of this interplay between infundibular KCs, immune cells, and bacteria may

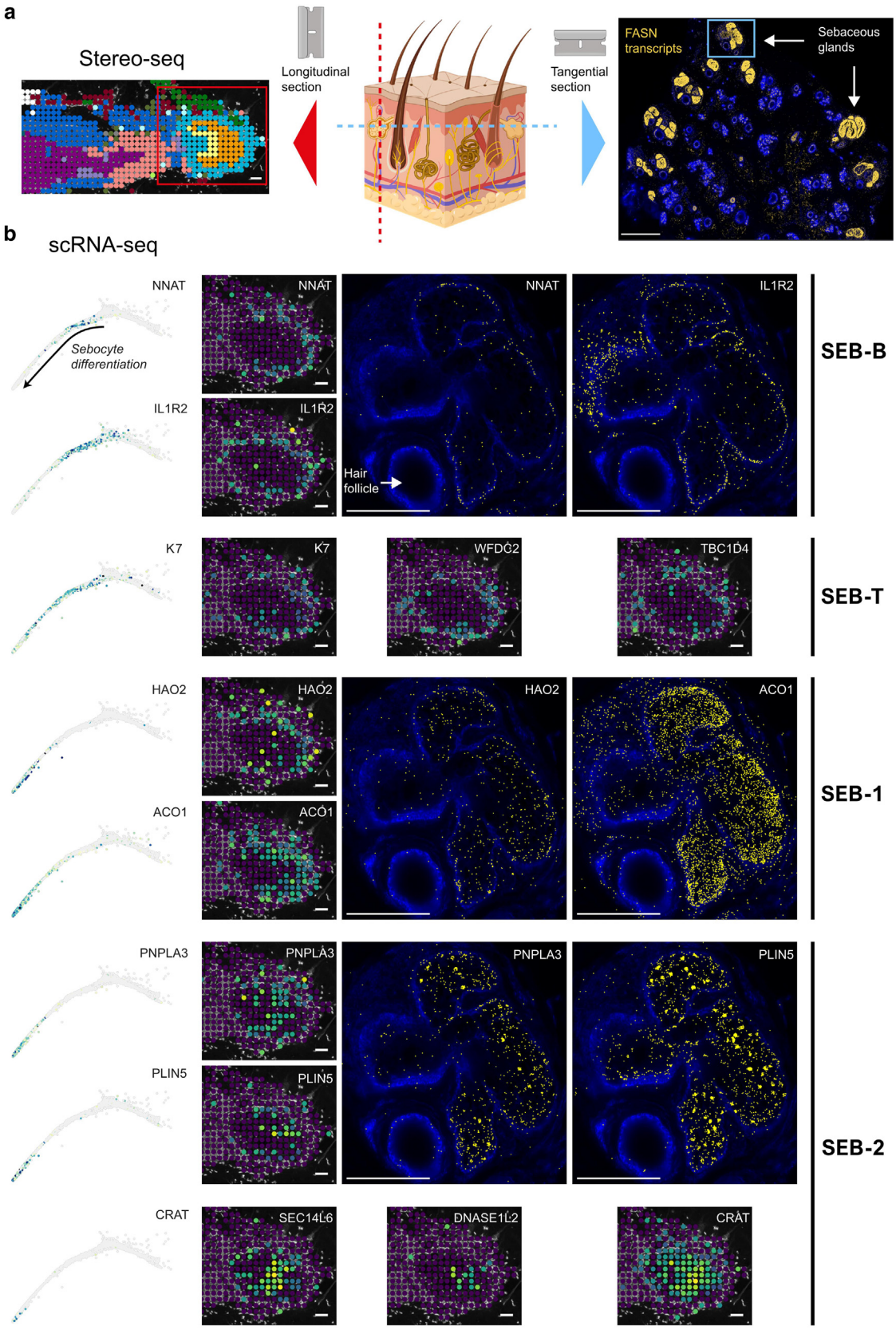


Figure 7. Validating sebocyte differentiation signatures using Stereo-seq and MERFISH. We used Stereo-seq and MERFISH technology to validate the gene signature of sebocyte differentiation stages as determined by scRNA-seq analysis. MERFISH allows the quantification and precise intracellular localization of mRNA transcripts. (**a**) To maximize the number of SGs analyzed, we applied a specialized tangential cutting technique for MERFISH (bar = 1 mm). In contrast, Stereo-seq was performed on a longitudinal section (bar = 25 μm). The red and blue rectangles indicate the areas captured in **b**. (**b**) Our data show mutually exclusive gene expression patterns from the basement membrane to the interior of the SG, culminating in the sebaceous duct. Key markers include NNAT and

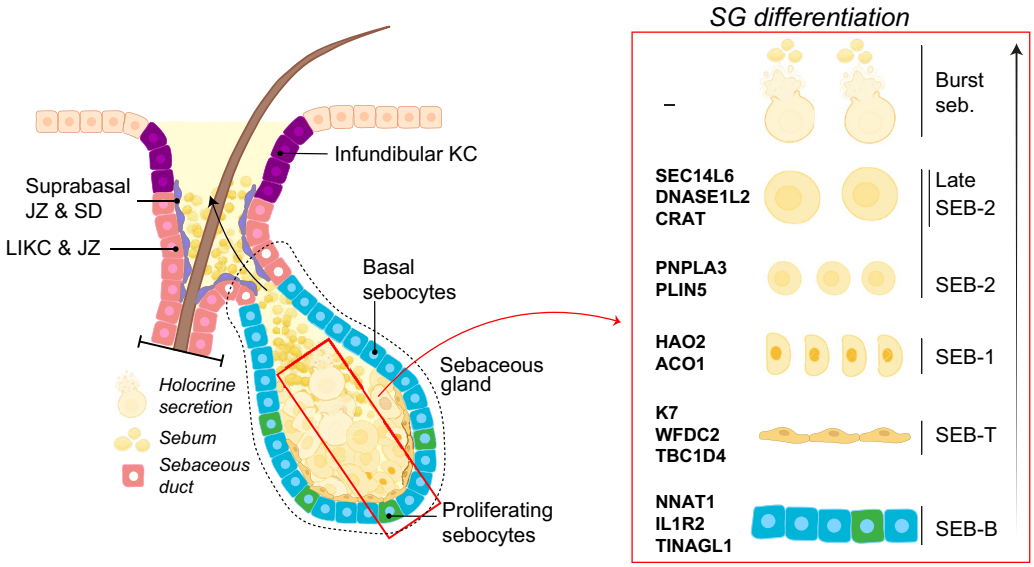


Figure 8. Overview figure of sebaceous gland differentiation stages with corresponding marker genes.

reveal key mechanisms behind the shift from normal immune surveillance to pathological inflammation.

Regarding sebocyte differentiation, our data provide a high-resolution analysis of healthy human SG differentiation using spatial transcriptomics and scRNA-seq, following the analysis of hyperplastic SGs by Schmidt et al (2024) using Visium. To assess how these changes might influence sebocyte differentiation signatures, we visualized the genes identified in their stages within our scRNA-seq and Stereo-seq dataset (Supplementary Figure S13).

TKT and HSPD1, proposed as basal markers by Schmidt et al (2024), showed broad expression in sebocytes in our datasets and were not restricted to basal sebocytes. Similarly, AWAT2 was expressed throughout the healthy SG, suggesting a more general role in sebocyte function rather than stage-specific differentiation. In contrast, ERV3-1 represented a specific marker of early sebocyte differentiation. Although ECHDC1 showed nonspecific expression in scRNA-seq, it appeared to be restricted to a later differentiation stage within the gland in Stereo-seq. GLRX showed consistent alignment in both datasets, supporting its reliability as a sebocyte marker. However, CDSN was not expressed in healthy SGs but rather marked cornifying KCs (Supplementary Figure S14). GSDMA appeared to be associated with terminal sebocytes.

Overall, our comparison suggests that SEB-T provides a refined resolution of transitional sebocytes, and our basal markers cluster with greater specificity. Most of the genes associated with later differentiation stages in hyperplastic SGs align well with those observed in healthy glands, suggesting that despite pathological alterations, the core mechanisms of sebocyte differentiation remain conserved.

In our further results, we identified 4 sebocyte differentiation stages and followed the naming conventions of Schmidt et al (2024) to establish a consistent nomenclature. Differentiation was continuous, but cells changed so profoundly along the trajectory that they can be categorized as distinct types. Because clustering allows for an arbitrary number of clusters, we used spatial transcriptomics to visualize gene expression, providing a guideline for setting the cluster resolution.

The elevated protein biosynthesis in SEB-B suggests that these cells produce a protein machinery essential for differentiation and specialized functions, such as de novo lipogenesis. The SEB-T stage suggests a potential cell layer–selective role for hormone signaling in SG differentiation (Szöllősi et al, 2017). In SEB-1, as sebocytes detach from the basement membrane, lipid biosynthesis begins. During SEB-2, alcohol biosynthesis is significantly enriched, contributing to the production of complex lipids such as wax esters and sphingolipids. The activation of p53 signaling in SEB-2, captured by spatial transcriptomics, indicates the initiation of apoptosis, in line with the natural life cycle of sebocytes, where they drive sebum production through holocrine secretion (Agamia et al, 2023). Mitochondrial activity appears central to this process because the high energy demand in late SEB-2 supports adenosine triphosphate production for de novo lipogenesis. Notably, sebocytes have a high percentage of mitochondrial mRNA, reflecting their metabolic activity rather than poor data quality. Investigating mitochondrial function in acne may provide insights into sebocyte metabolism and inflammatory responses. To avoid loss of sebocytes during quality control, we recommend setting the mitochondrial mRNA cut off between 15 and 25% when analyzing human skin transcriptomic data. A limitation of this analysis is that late SEB-2 could only be examined at the cluster level in the spatial data, whereas earlier stages were additionally assessed along a pseudotime continuum in the scRNA-seq data, resulting in a partially overlapping boundary between SEB-2 and late SEB-2 that remains difficult to define

IL1R2 for the SEB-B stage, predominantly localized to the basement membrane; K7, WFDC2, and TBC1D4 for transitional sebocytes (SEBT); and HAO2 and ACO1 for SEB-1 sebocytes. Expression of PNPLA3 and PLIN5 marks a gradual transition to SEB-2, whereas SEC14L6, DNASE1L2, and CRAT are expressed in late SEB-2 before reaching the sebaceous duct. Bars = 25 μm for Stereo-seq plots and 250 μm for MERFISH plots. K7, keratin 7; MERFISH, multiplexed error-robust FISH; scRNA-seq, single-cell RNA sequencing; SG, sebaceous gland.

precisely. Our results suggest that basal and proliferating sebocytes share the same markers, with proliferation-specific markers being the only distinguishing feature. We hypothesize that basal sebocytes primarily proliferate to expand the SG and create space for differentiated sebocytes, without significantly contributing to differentiation. This stochastic proliferation may explain the amorphous structure of SGs.

In sebocytes, we identified CyPA as a specific autocrine signaling factor (Figure 2b). CyPA, a member of the immunophilin family, is known for its roles in protein folding, trafficking, and immune modulation (Nigro et al, 2013). In particular, it has been shown to regulate key cellular processes, including inflammation and apoptosis (Chen et al, 2024). Notably, inhibition of CyPA has been demonstrated to restore adipocyte size and function, thereby improving insulin sensitivity (Zhang et al, 2015). Research on transplant recipients suggests that cyclosporine, a known CyPA inhibitor, can induce sebaceous hyperplasia, a side effect not observed with other immunosuppressants (Boschnakow et al, 2003). Notably, this condition occurs predominantly on the faces of older men, raising the question of whether it is directly caused by cyclosporine or an early sign of dysplasia in genetically susceptible individuals (Pang and Chau, 2005). Our results reveal dominant CyPA signaling through the BSG receptor in SGs, suggesting that sebaceous hyperplasia may result from cyclosporine inhibition. Further investigations are warranted to elucidate the mechanisms through which CyPA influences sebocytes and its broader implications for skin health and disease.

CD6−ALCAM interactions are crucial for regulating T-cell activation and proliferation (Chalmers et al, 2022). KCs can activate memory T cells, partly through CD6−ALCAM signaling, which influences metabolic reprogramming (Orlik et al, 2022). Inhibition of the CD6−ALCAM axis impairs mTOR activation, fatty acid synthesis, and memory T-cell formation while increasing glucose consumption and lactate production (Orlik et al, 2022). This suggests that CD6 may prevent excessive metabolic activation and maintains immune homeostasis. Disruption of this pathway could affect sebocyte metabolism and inflammatory responses. In acne, hyperactive mTOR signaling contributes to SG dysfunction (Melnik, 2018) and may promote T-cell activation and inflammation.

In conclusion, our study has provided insights into the molecular basis of SG function and its potential role in skin pathologies. We have uncovered gene markers for infundibular KCs and basal sebocytes, providing valuable tools for precise cell-type identification. Because marker genes often have biological roles, they can also reveal novel functional properties, although mRNA-based findings do not always match protein-level functions. We have also successfully analyzed cell−cell communication at spatial resolution, providing insight into the complex interactions within the skin and elucidating the stages of sebocyte differentiation. These findings enhance the knowledge of SG dynamics and may inform therapeutic strategies for skin diseases.

MATERIALS AND METHODS

Experimental design of wet laboratory studies

Subject details. Human facial skin specimens from the temporal region were commercially obtained from written informed and consenting donors undergoing facelift surgery. The donors were healthy, White females aged 49−57 years. Skin samples were processed to remove subcutaneous fat, cut into 5−7 × 5−7 mm pieces, and frozen and stored at −80 °C.

Stereo-seq. The generation of Stereo-seq chips, collision-induced dissociation sequence calling (short barcode sequence for identifying the location of captured molecules), tissue processing, in situ reverse transcription, amplification, library construction, sequencing, probe and section preparation, hybridization, and staining were all conducted by MIRXES], following the protocols detailed in the Stereo-seq paper.

ssDNA staining was performed, and the tissue sections were placed on the chip for fixation and permeabilization. The permeabilization time was specifically optimized for the PSU (Supplementary Figure S15a) to ensure vertical diffusion of mRNA into corresponding spots while minimizing lateral diffusion effects.

MERFISH. We used the differentiation signature derived from our scRNA-seq analysis to design a gene panel, which was subsequently used for MERFISH analysis. The generation of the MERFISH data was performed following the standard methods outlined by Chen et al (2015) except for the tissue clearing procedure. After gel embedding, tissue digestion was conducted at 37 °C for 6 hours. Subsequently, clearing was performed at 47 °C overnight, followed by autofluorescence quenching for 6 hours. Finally, the samples were cleared at 37 °C for 36−48 hours.

Computational analysis

An overview of the computational methods used in this study is provided in Supplementary Table S5.

Stereo-seq raw data processing. FASTQ file generation, barcode mapping, read filtering, read alignment, count matrix generation, and image registration and tissue cutting were performed by MIRXES using the SAW pipeline, as detailed in the Stereo-seq paper (Chen et al, 2022).

In binning, the spots on the chip are grouped into virtual squares, representing cells (Chen et al, 2022). We chose a binning size of 50 for data analysis, which corresponds to a cell diameter of approximately 25 μm, a size that has also been used in several other studies (Chen et al, 2022; Cheng et al, 2022; Wu et al, 2023).

Identification of spatially informative gene modules. We used Hotspot to identify spatially variable gene modules in tissue section D1, using a gene count matrix derived from the top 3000 variable genes. Data were normalized by the total unique molecular identifier count per barcode. The Bernoulli background model was applied. A neighborhood graph of genes was created with the function create_knn_graph with a nondefault n_neighbors parameter of 40. Genes with significant spatial variation (false discovery rate $<$ 0.05) were retained. Hierarchical clustering was applied to define gene modules using the create_modules function, with a nondefault min_gene_threshold parameter set to 15.

Data Processing and integration of Stereo-seq tissue sections using Stereopy and Seurat. We processed the 3 tissue GEF files corresponding to tissue sections D1, D2, and D3, respectively, using Stereopy, with a binning size of 50 for data analysis. In spatial transcriptomics, cells are defined by either segmentation or binning. Segmentation identifies cell boundaries, such as cell walls, and aggregates transcriptomic data within these boundaries. To date, Stereo-seq only integrates ssDNA staining and does not include H&E

staining, meaning that cell walls are not visible for segmentation (Chen et al, 2022). The binning strategy provides an effective alternative to simulate single-cell resolution. The StereoExpData objects were converted to AnnData format using st.io.stereo_to_anndata. The resulting .h5ad files were converted to .rds format using h5ad2rds.R, available from the official Stereopy documentation site.

The generated .rds files were then imported into Seurat by creating a Visium V1 class. Cells with <500 genes and >25% mitochondrial RNA content were excluded. We applied SCTransform, which combines normalization, variable feature selection, and scaling. Principal component analysis (RunPCA), neighbor finding (FindNeighbors), clustering (FindClusters), and Uniform Manifold Approximation and Projection were then performed, all using default parameters according to the official Seurat documentation. To account for batch effects, we used Harmony and defined each tissue section as a batch. Clustering was performed on the batch-corrected Harmony components (resolution = 2.2). We subset the data into clusters expressing markers from the IFE and PSU; reanalyzed the subset (Figure 4b); assigned the cell groups on the basis of known gene markers; and projected these clusters back onto the full dataset containing all cells for D1, D2, and D3 (Figure 2a).

Cell–cell communication analysis with spatial information using CellChat. We followed the standard workflow with default parameters as described in the CellChat-Paper (Jin et al, 2024).

The spatial.factors parameter for creating the CellChat object depends on the spatial transcriptomics method used. It is a data frame that specifies the conversion factor from spatial coordinates (eg, pixels) to micrometers. For example, setting ratio = 0.18 means that 1 pixel equals 0.18 μm. The data frame also contains a tolerance factor (tol), which increases robustness when comparing center-to-center distances with the interaction range. According to the documentation, for Stereo-seq, we used a recommended ratio of 0.5 and tol of 12.5.

We used the extended CellChat database from version 2 but filtered out extracellular matrix–receptor interactions and nonprotein signaling. Extracellular matrix interactions were excluded owing to their high prevalence in our skin samples, where interactions such as collagen, laminin, and tenascin were overwhelmingly dominant and overshadowed other interactions. Nonprotein signaling was excluded on the basis of recommendations for neuronal data because inferring signaling pathways from upregulated key enzymes can be error prone and miss important factors. In computeCommunProb, we calculated the truncated mean and set the trimming parameter to 0.10 to avoid overly stringent filtering of signaling pathways. The contact.range for cell–cell interactions was set to 25 μm, reflecting the average cell diameter after setting the binning size to 50.

Modeling the human PSU with scRNA-seq and characterization of gene markers. We imported each filtered feature-barcode matrix separately for each sample and created Seurat objects using CreateSeuratObject. A filtering criterion of at least 200 features per cell was applied.

To perform quality control, we first calculated additional quality control metrics, including the log-transformed total RNA counts (log1p_total_counts), log-transformed number of genes detected (log1p_n_genes_by_counts), and the percentage of mitochondrial genes (percent_mito) for each sample. Outliers were identified using MAD (median absolute deviation). Outliers were flagged in the log1p_total_counts and log1p_n_genes_by_counts metrics if they were outside 3 MADs from the median and in the percent_mito metric if they were outside 5 MADs from the median. The lenient threshold for mitochondrial RNA was set to avoid over-removal of sebocytes, which naturally have higher levels of mitochondrial RNA, as shown in the Stereo-seq data. Cells were retained if they were not flagged as outliers in any of these quality control metrics. The filtered cells were then subsetted from each sample for further analysis.

The Seurat objects were merged and processed using the standard workflow, including SCTransform, RunUMAP, FindNeighbors, and FindClusters. An amorphous cluster with epithelial cell markers was then subset and reanalyzed to achieve finer resolution. This involved subsetting the Seurat object to include only this amorphous cluster and its constituent clusters, followed by reapplying the standard workflow while correcting for cell cycle effects. We regressed out the difference between G2M and S phase scores using Seurat. This preserved the separation between cycling and noncycling cells while removing uninteresting variation within proliferating cells.

After excluding clusters of low-quality cells, we repeated the standard operations from the previous step and manually annotated cell groups on the basis of known gene markers. For the cell types within the PSU, we performed differential gene expression analysis using PrepSCTFindMarkers and FindMarkers on the entire dataset. In this analysis, ident.1 corresponded to the respective cell types of the PSU.

The markers were visually inspected, and those that best separated the cell types and were, to our knowledge, previously unreported were visualized as violin plots using ScCustomize.

Validation of Stereo-seq cell types using anchor-based integration with scRNA-seq. To validate the cell types identified in Stereo-seq, we used single-cell data and applied the anchor-based integration workflow introduced in Seurat, version 3. This method allowed the probabilistic transfer of cell-type annotations from scRNA-seq to Stereo-seq data. We obtained prediction scores for each spot, which were visualized spatially and as a heatmap using minimum–maximum normalization. Our results showed successful recovery of the cell types in Stereo-seq, validating the data (Supplementary Figure S7).

Projecting sebocyte signatures from scRNA-seq onto Stereo-seq data aligned better with basal than differentiated sebocytes, likely owing to differences in the technologies. scRNA-seq captures intact cells, whereas spatial transcriptomics can also detect ruptured or large lipid-filled cells through spot sequencing (Veniaminova et al, 2023). On the other hand, we analyzed sebocyte differentiation using scRNA-seq data because it allows true single-cell resolution, whereas the uniform binning of spatial transcriptomics struggles with increasing sebocyte size. Because sebocytes grow significantly during differentiation, the fixed binning can never be optimally adjusted to handle such a wide range of cell sizes. In addition, scRNA-seq provides greater gene coverage, allowing for more precise identification of gene markers.

Using PHATE to illustrate differentiation more intuitively and naturally. We used PHATE in R through the reticulate package. PHATE encodes local structure by applying a kernel function to Euclidean distances and encodes global relationships through potential distance. A diffusion process is applied using a t-step random walk, generating so-called diffusion probabilities, which are then contextualized within the global cell interaction network before being embedded for visualization using multidimensional scaling

(Moon et al, 2019). We set the parameters for knn and npca to 30 to ensure comparability with the default parameters of the RunUMAP function. Stereo-seq data were subsetted to include only the PSU cell types and the IFE, reanalyzed, and visualized using PHATE in the same manner.

Velocity analysis. For the scRNA-seq velocity analysis, we used Velocyto with the run10x command and the same reference files as for the Cell Ranger analysis to generate loom files for each sample. The subsetted Seurat object containing the PSU and IFE was augmented with the spliced and unspliced RNA count matrices and preprocessed using the standard scVelo workflow. Unless otherwise noted, all functions were run with default settings. For scvelo.pp.filter_and_normalize, the n_top_genes argument was set to 2000, whereas in scvelo.pp.momentum, both n_npcs and n_neighbors were set to 30. The velocity layer was calculated using DeepVelo, chosen for its superior accuracy to those of other methods. Single-cell velocity vectors were visualized using scvelo.pl.velocity_embedding.

Dynamo was used to calculate the spatial RNA velocity for the PSU region in tissue section D1. The loom file for D1 was generated using generate_loom from Stereopy with a bin size of 50, read into Dynamo using dyn.read, and preprocessed using recipe_monocle with default parameters according to the documentation. We used the stochastic model to learn the high-dimensional velocity vectors for all genes. The Pearson method in dyn.tl.cell_velocities was used to compute the transition matrix, needed to project the high-dimensional data into a low-dimensional space. Highly variable genes were passed to the transition_genes argument.

Pseudotime analysis with Monocle3. Pseudotime algorithms arrange cells along a timeline on the basis of their transcriptomic similarity and assign them a unit called pseudotime, which reflects the differentiation state of a cell relative to that of its progenitor (Luecken and Theis, 2019). Because similarity alone does not imply direction, these algorithms require biological prior knowledge in the form of a defined start or end point. After applying PHATE to the PSU and IFE subsets from scRNA-seq and Stereo-seq, Monocle3 was used with default parameters to analyze the pseudotime. The designated start points for scRNA-seq were a cell from the lower infundibular KCs and JZ and the basal KC cluster. For Stereo-seq, only the lower infundibular KCs and JZ were used owing to the close spatial proximity of basal KCs to the lower infundibular KCs and JZ, which likely resulted from the binning.

Uncovering the putative lineage drivers of sebocyte differentiation. We used CellRank to infer lineage drivers for sebocyte differentiation from the pseudotime computed with Monocle3, following the standard instructions. The estimator was fit using g.fit, specifying n_states = *2* to obtain distinct macrostates: one for the sebocyte lineage and another for the infundibulum lineage.

To identify genes with similar expression patterns in a trajectory-specific manner for the sebocyte lineage, we applied cr.pl.cluster_trends, using only significantly correlated genes to accelerate computation and setting the resolution parameter to 0.18. We then visualized the temporal activation of each gene cluster along the trajectory using the cr.pl.heatmap function.

Putative driver genes were identified by correlating fate probabilities with gene expression using the compute_lineage_drivers() method. These genes were sorted in descending order on the basis of their Pearson correlation values, and gene set enrichment analysis was performed. This was done using the fgsea package.

After performing a differential gene expression analysis on the SG cluster from the Stereo-seq subset data, we performed Gene Ontology enrichment analysis using the enrichGO function. The analysis was performed on the set of differentially expressed genes using the human reference database "org.Hs.eg.db," using gene symbols as identifiers and focusing on biological processes.

ETHICS STATEMENT

This study was conducted in accordance with the Declaration of Helsinki. Face skin explants were provided by Alphenyx and obtained from healthy participants undergoing esthetic surgical resection. Alphenyx obtained all necessary authorizations from the French Ministry of Research, with approval from the French Ethical Committee (numbers AC-2019-3567 and IE-2019-1077). A letter of approval is available upon request. All adult participants provided written, informed consent to participate in this study.

DATA AVAILABILITY STATEMENT

The raw and processed data are deposited in National Center for Biotechnology Information's Gene Expression Omnibus database under the accession codes GSE292156 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE292156) for Stereo-seq and GSE292394 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE292394) for multiplexed error-robust FISH. The processed data objects for downstream analysis are available on Figshare under the following DOI: https://doi.org/10.6084/m9.figshare.27691512.v1. FASTQ files for healthy patient samples from dataset GSE147424 were downloaded from the Sequence Read Archive (https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA614539&o=acc_s%3Aa) using fasterq-dump. These files were aligned to GRCh38 using the standard Cell Ranger pipeline 7.2.0 with default parameters, generating a filtered feature-barcode matrix for each sample. This matrix was then used for downstream analysis. All analyses were performed using publicly available software and packages. Any additional information required to reanalyze the data presented in this study is available from the corresponding author, NH.

ORCIDs

Tolga Düz: http://orcid.org/0009-0002-3179-8206
Nicholas Holzscheck: http://orcid.org/0000-0002-3446-2487

CONFLICT OF INTEREST

TD, AS, PW, SG, and NH are employees of Beiersdorf AG. DT and JB received consultation fees from Beiersdorf AG. The remaining authors state no conflict of interest.

ACKNOWLEDGMENTS

This work was funded by Beiersdorf AG. DT was supported by the Hungarian National Research, Development and Innovation Office (grant FK-132296 and ANN 139589).

AUTHOR CONTRIBUTIONS

Conceptualization: NH, TD; Formal Analysis: TD; Funding Acquisition: SG; Resources: AS, PW; Software: TD; Supervision: NH, DT, JB; Visualization: TD; Writing — Original Draft Preparation: TD, DT; Writing — Review and Editing: TD, DT, NH, JB

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at www.jidonline.org, and at https://doi.org/10.1016/j.jid.2025.04.041.

REFERENCES

Acosta EM, Little KA, Bratton BP, Lopez JG, Mao X, Payne AS, et al. Bacterial DNA on the skin surface overrepresents the viable skin microbiome. eLife 2023;12:RP87192.

Agamia NF, El Mulla KF, Alsayed NM, Ghazala RM, El Maksoud REA, Abdelmeniem IM, et al. Isotretinoin treatment upregulates the expression of p53 in the skin and sebaceous glands of patients with acne vulgaris. Arch Dermatol Res 2023;315:1355–65.

Bergen V, Soldatov RA, Kharchenko PV, Theis FJ. RNA velocity-current challenges and future perspectives. Mol Syst Biol 2021;17:e10282.

Bertolini M, McElwee K, Gilhar A, Bulfone-Paus S, Paus R. Hair follicle immune privilege and its collapse in alopecia areata. Exp Dermatol 2020;29:703–25.

Boschnakow A, May T, Assaf C, Tebbe B, Zouboulis ChC. Ciclosporin A-induced sebaceous gland hyperplasia. Br J Dermatol 2003;149:198–200.

Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, et al. The single-cell transcriptional landscape of mammalian organogenesis. Nature 2019;566:496–502.

Chalmers SA, Ayilam Ramachandran R, Garcia SJ, Der E, Herlitz L, Ampudia J, et al. The CD6/ALCAM pathway promotes lupus nephritis via T cell-mediated responses. J Clin Invest 2022;132:e147334.

Chen A, Liao S, Cheng M, Ma K, Wu L, Lai Y, et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. Cell 2022;185:1777–92.e21.

Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science 2015;348:aaa6090.

Chen L, Zeng Z, Luo H, Xiao H, Zeng Y. The effects of CypA on apoptosis: potential target for the treatment of diseases. Appl Microbiol Biotechnol 2024;108:28.

Cheng M, Wu L, Han L, Huang X, Lai Y, Xu J, et al. A cellular resolution spatial transcriptomic landscape of the medial structures in postnatal mouse brain. Front Cell Dev Biol 2022;10:878346.

Christoph T, Müller-Röver S, Audring H, Tobin DJ, Hermes B, Cotsarelis G, et al. The human hair follicle immune system: cellular composition and immune privilege. Br J Dermatol 2000;142:862–73.

Dahlhoff M, Camera E, Picardo M, Zouboulis CC, Schneider MR. Angiopoietin-like 4, a protein strongly induced during sebocyte differentiation, regulates sebaceous lipogenesis but is dispensable for sebaceous gland function in vivo. J Dermatol Sci 2014;75:148–50.

de Bengy AF, Forraz N, Danoux L, Berthelemy N, Cadau S, Degoul O, et al. Development of new 3D human ex vivo models to study sebaceous gland lipid metabolism and modulations. Cell Prolif 2019;52:e12524.

DeTomaso D, Yosef N. Hotspot identifies informative gene modules across modalities of single-cell genomics. Cell Syst 2021;12:446–56.e9.

Dréno B. What is new in the pathophysiology of acne, an overview. J Eur Acad Dermatol Venereol 2017;31:8–12.

Eckhart L, Lippens S, Tschachler E, Declercq W. Cell death by cornification. Biochim Biophys Acta 2013;1833:3471–80.

Fischer H, Fumicz J, Rossiter H, Napirei M, Buchberger M, Tschachler E, et al. Holocrine secretion of sebum is a unique DNase2-dependent mode of programmed cell death. J Invest Dermatol 2017;137:587–94.

Gupta AK, Bluhm R. Seborrheic dermatitis. J Eur Acad Dermatol Venereol 2004;18:13–26 [quiz: 19–20].

Hazarika N, Archana M. The psychosocial impact of acne vulgaris. Indian J Dermatol 2016;61:515–20.

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.

Hsieh CY, Chen CL, Lin YS, Yeh TM, Tsai TT, Hong MY, et al. Macrophage migration inhibitory factor triggers chemotaxis of CD74+CXCR2+ NKT cells in chemically induced IFN-γ-mediated skin inflammation. J Immunol 2014;193:3693–703.

Jensen KB, Collins CA, Nascimento E, Tan DW, Frye M, Itami S, et al. Lrig1 expression defines a distinct multipotent stem cell population in mammalian epidermis. Cell Stem Cell 2009;4:427–39.

Jin S, Plikus MV, Nie Q. CellChat for systematic analysis of cell-cell communication from single-cell transcriptomics. Nat Protoc 2024:1–40.

Joost S, Zeisel A, Jacob T, Sun X, La Manno G, Lönnerberg P, et al. Single-cell transcriptomics reveals that differentiation and spatial signatures shape epidermal and hair follicle heterogeneity. Cell Syst 2016;3:221–37.e9.

Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA. The technology and biology of single-cell RNA sequencing. Mol Cell 2015;58:610–20.

Lange M, Bergen V, Klein M, Setty M, Reuter B, Bakhti M, et al. CellRank for directed single-cell fate mapping. Nat Methods 2022;19:159–70.

Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol 2019;15:e8746.

Melnik BC. Acne vulgaris: the metabolic syndrome of the pilosebaceous follicle. Clin Dermatol 2018;36:29–40.

Moon KR, Van Dijk D, Wang Z, Gigante S, Burkhardt DB, Chen WS, et al. Visualizing structure and transitions in high-dimensional biological data. Nat Biotechnol 2019;37:1482–92.

Nagao K, Kobayashi T, Moro K, Ohyama M, Adachi T, Kitashima DY, et al. Stress-induced production of chemokines by hair follicles regulates the trafficking of dendritic cells in skin. Nat Immunol 2012;13:744–52.

Nigro P, Pompilio G, Capogrossi MC. Cyclophilin A: a key player for human disease. Cell Death Dis 2013;4:e888.

Orlik C, Berschneider KM, Jahraus B, Niesler B, Balta E, Schäkel K, et al. Keratinocyte-induced costimulation of human T cells through CD6 - but not CD2 - activates mTOR and prevents oxidative stress. Front Immunol 2022;13:1016112.

Oulès B, Philippeos C, Segal J, Tihy M, Vietri Rudan M, Cujba AM, et al. Contribution of GATA6 to homeostasis of the human upper pilosebaceous unit and acne pathogenesis. Nat Commun 2020;11:5067.

Pang SM, Chau YP. Cyclosporin-induced sebaceous hyperplasia in renal transplant patients. Ann Acad Med Singap 2005;34:391–3.

Rao A, Barkley D, França GS, Yanai I. Exploring tissue architecture using spatial transcriptomics. Nature 2021;596:211–20.

Schmidt M, Hansmann F, Loeffler-Wirth H, Zouboulis CC, Binder H, Schneider MR. A spatial portrait of the human sebaceous gland transcriptional program. J Biol Chem 2024;300:107442.

Schneider MR, Paus R. Sebocytes, multifaceted epithelial cells: lipid production and holocrine secretion. Int J Biochem Cell Biol 2010;42:181–5.

Schneider MR, Schmidt-Ullrich R, Paus R. The hair follicle as a dynamic miniorgan. Curr Biol 2009;19:R132–42.

Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, et al. Comprehensive integration of single-cell data. Cell 2019;177:1888–902.e21.

Sundberg JP, Shen T, Fiehn O, Rice RH, Silva KA, Kennedy VE, et al. Sebaceous gland abnormalities in fatty acyl CoA reductase 2 (Far2) null mice result in primary cicatricial alopecia. PLoS One 2018;13:e0205775.

Szöllősi AG, Oláh A, Bíró T, Tóth BI. Recent advances in the endocrinology of the sebaceous gland. Dermatoendocrinol 2017;9:e1361576.

Thalheim T, Schneider MR. Skin single-cell transcriptomics reveals a core of sebaceous gland-relevant genes shared by mice and humans. BMC Genomics 2024;25:137.

Thrane K, Winge MCG, Wang H, Chen L, Guo MG, Andersson A, et al. Single-cell and spatial transcriptomic analysis of human skin delineates intercellular communication and pathogenic cells. J Invest Dermatol 2023;143:2177–92.e13.

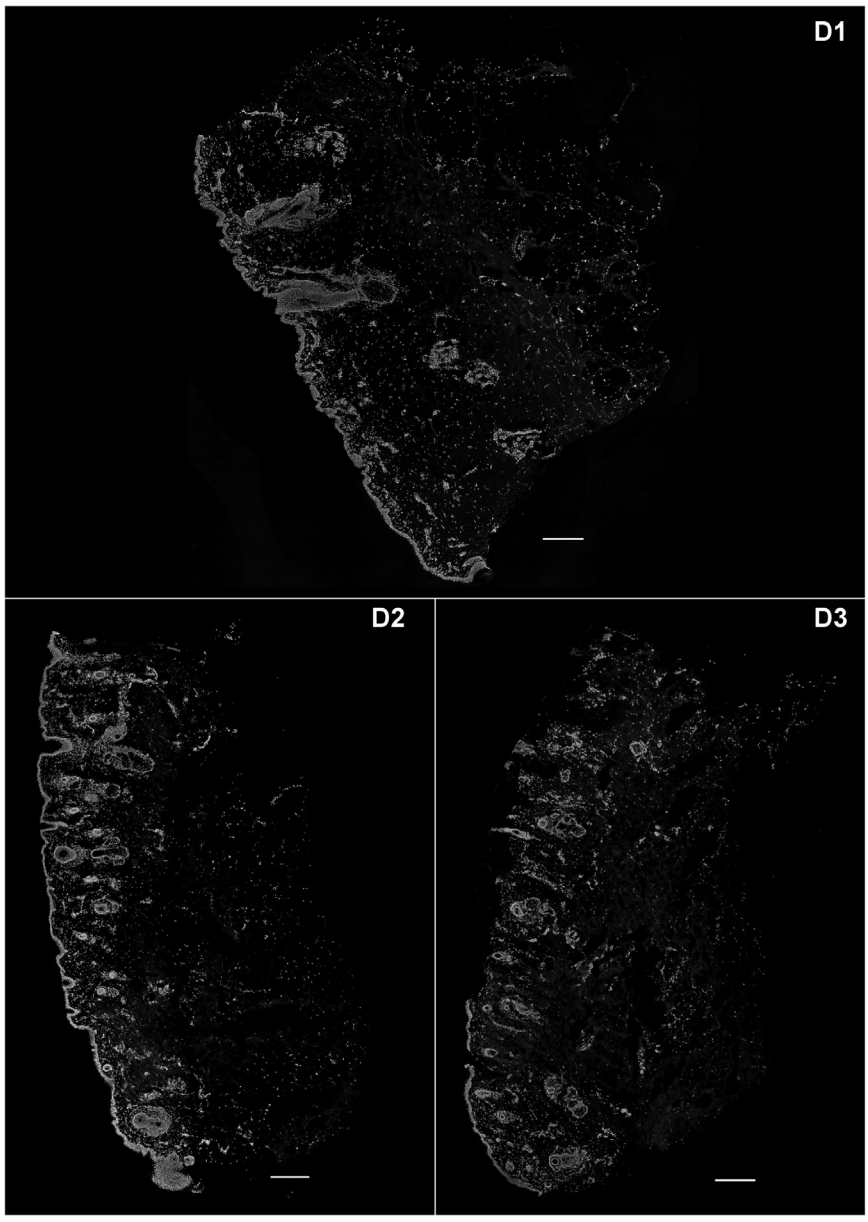
Veniaminova NA, Jia Y, Hartigan AM, Huyge TJ, Tsai SY, Grachtchouk M, et al. Distinct mechanisms for sebaceous gland self-renewal and regeneration provide durability in response to injury. Cell Rep 2023;42:113121.

Wu L, Yan J, Bai Y, Chen F, Zou X, Xu J, et al. An invasive zone in human liver cancer identified by Stereo-seq promotes hepatocyte-tumor cell crosstalk, local immunosuppression and tumor progression. Cell Res 2023;33:585–603.

Xia C, Fan J, Emanuel G, Hao J, Zhuang X. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. Proc Natl Acad Sci USA 2019;116:19490–9.

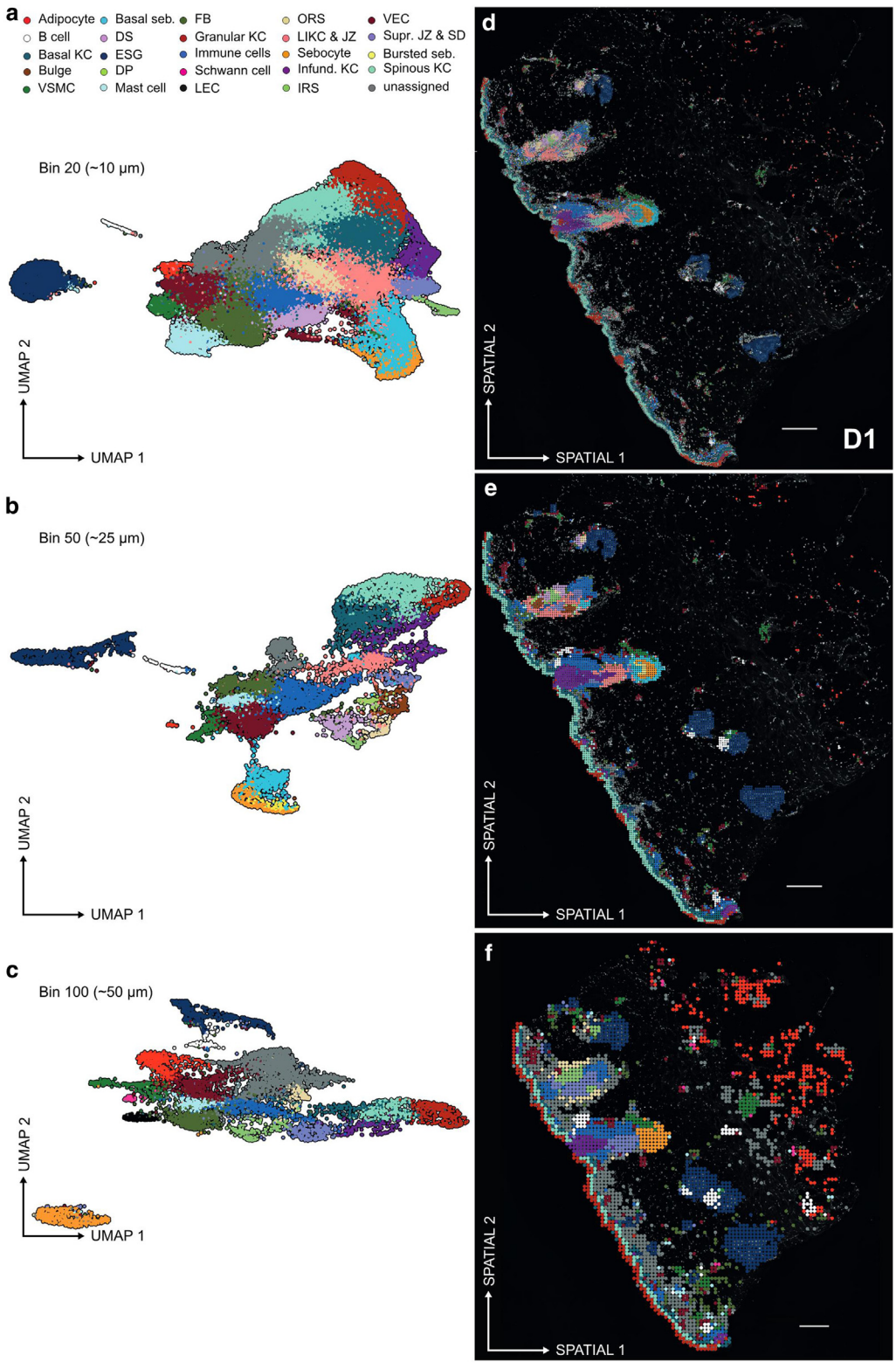
Zhang L, Li Z, Zhang B, He H, Bai Y. PPIA is a novel adipogenic factor implicated in obesity. Obesity (Silver Spring) 2015;23:2093–100.

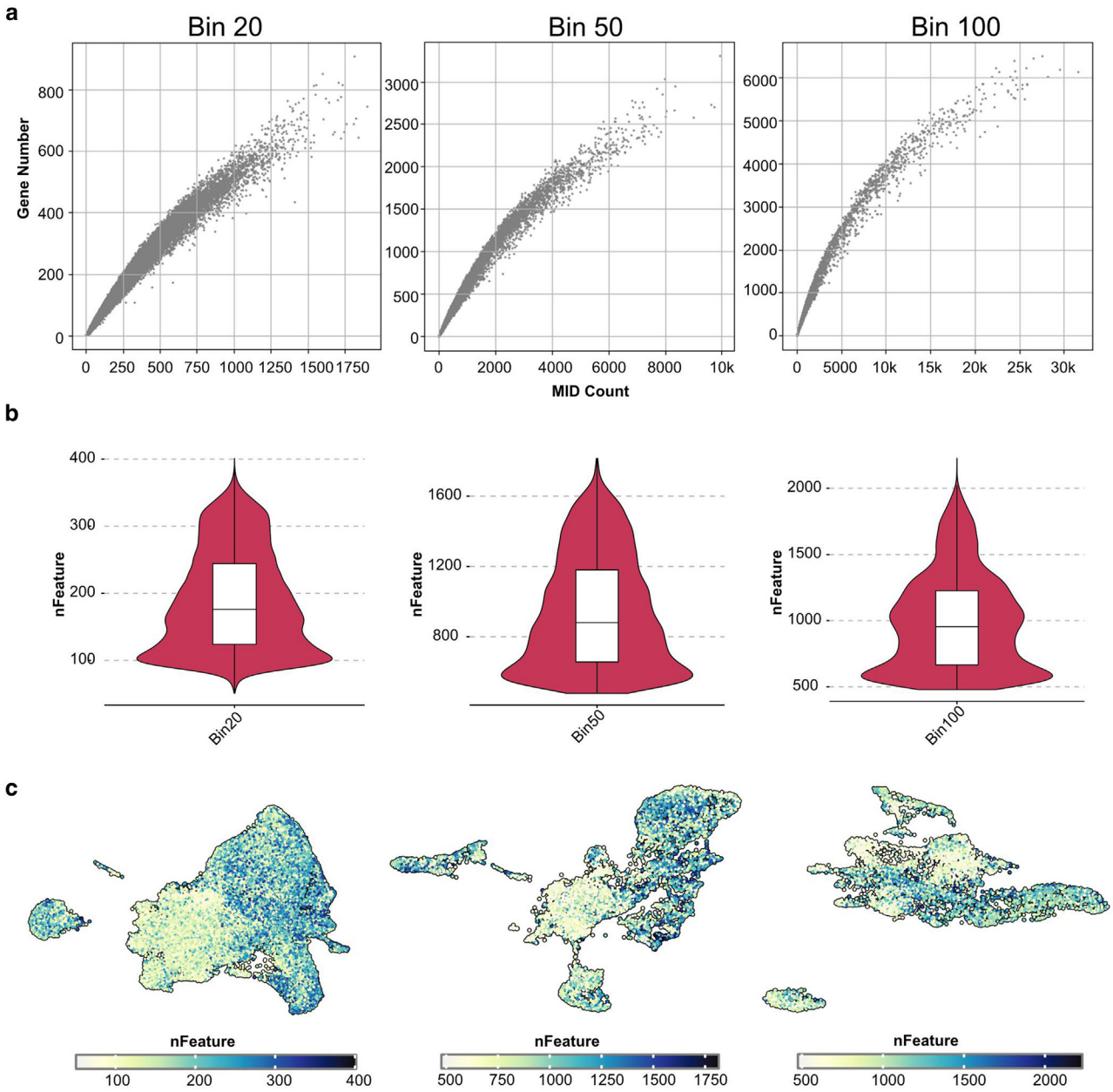
Zouboulis CC. Further evidence of sebaceous differentiation uniqueness: holocrine secretion of sebocytes is a multistep, cell-specific lysosomal DNase2-mediated mode of programmed cell death. J Invest Dermatol 2017;137:537–9.



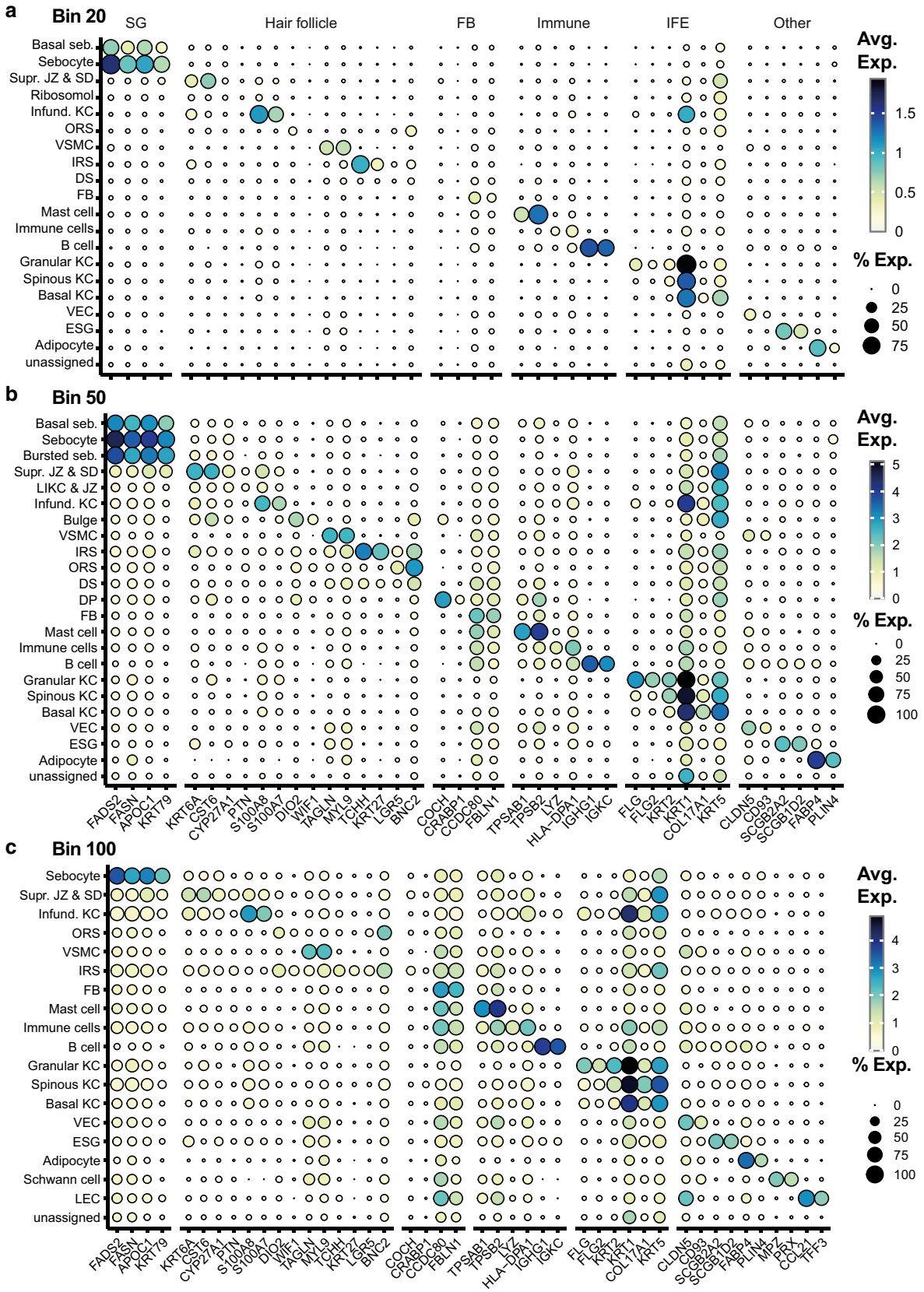
Supplementary Figure S1. Single-stranded DNA staining of tissue sections. Bar = 200 μm.

Supplementary Figure S2. UMAP and spatial visualization of cell types at different binning sizes. (**a**–**c**) UMAPs for binning sizes 20, 50, and 100. (**d**–**f**) Spatial plots of cell types in section D1 for binning sizes 20, 50, and 100. Bar = 200 μm. UMAP, Uniform Manifold Approximation and Projection.

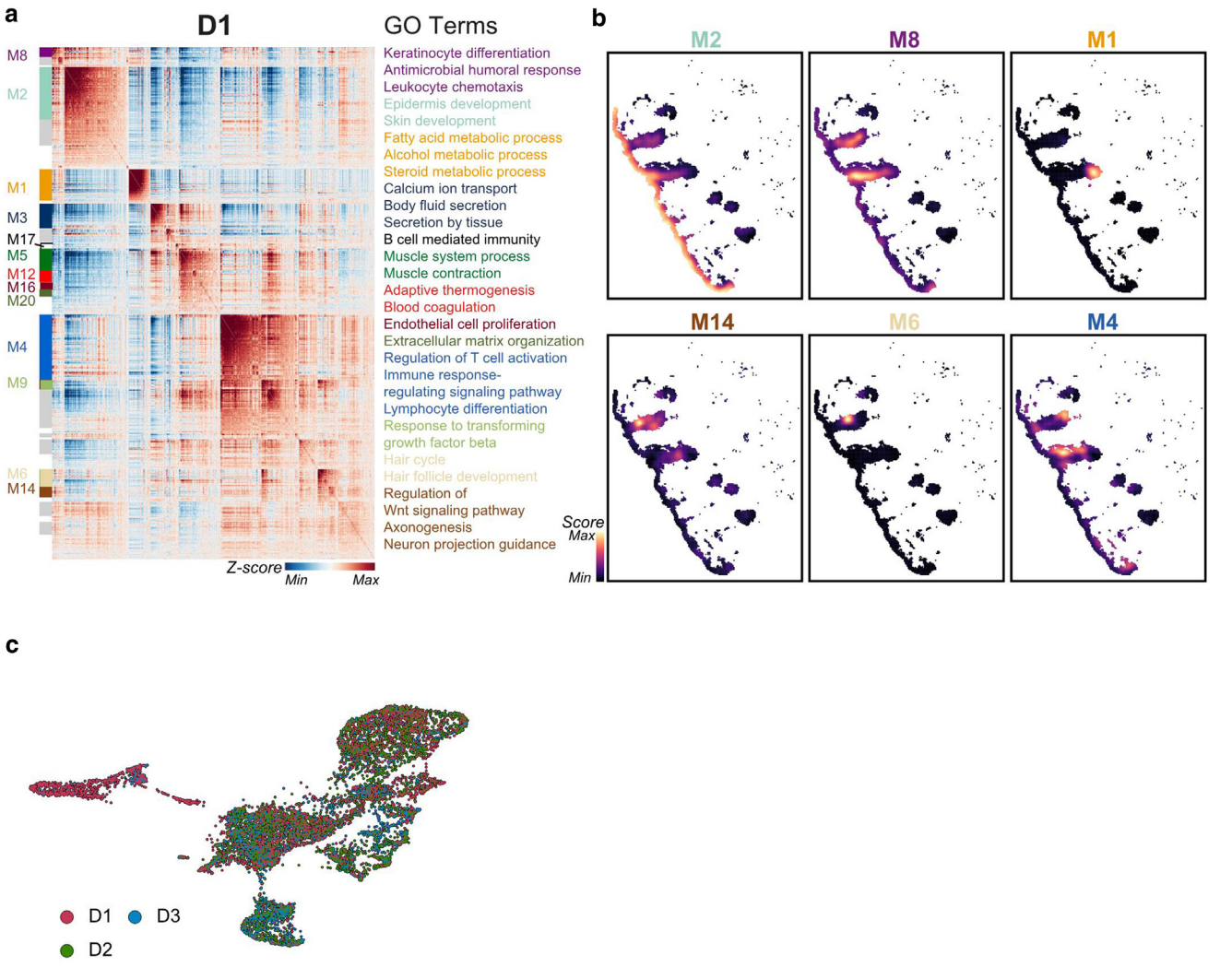




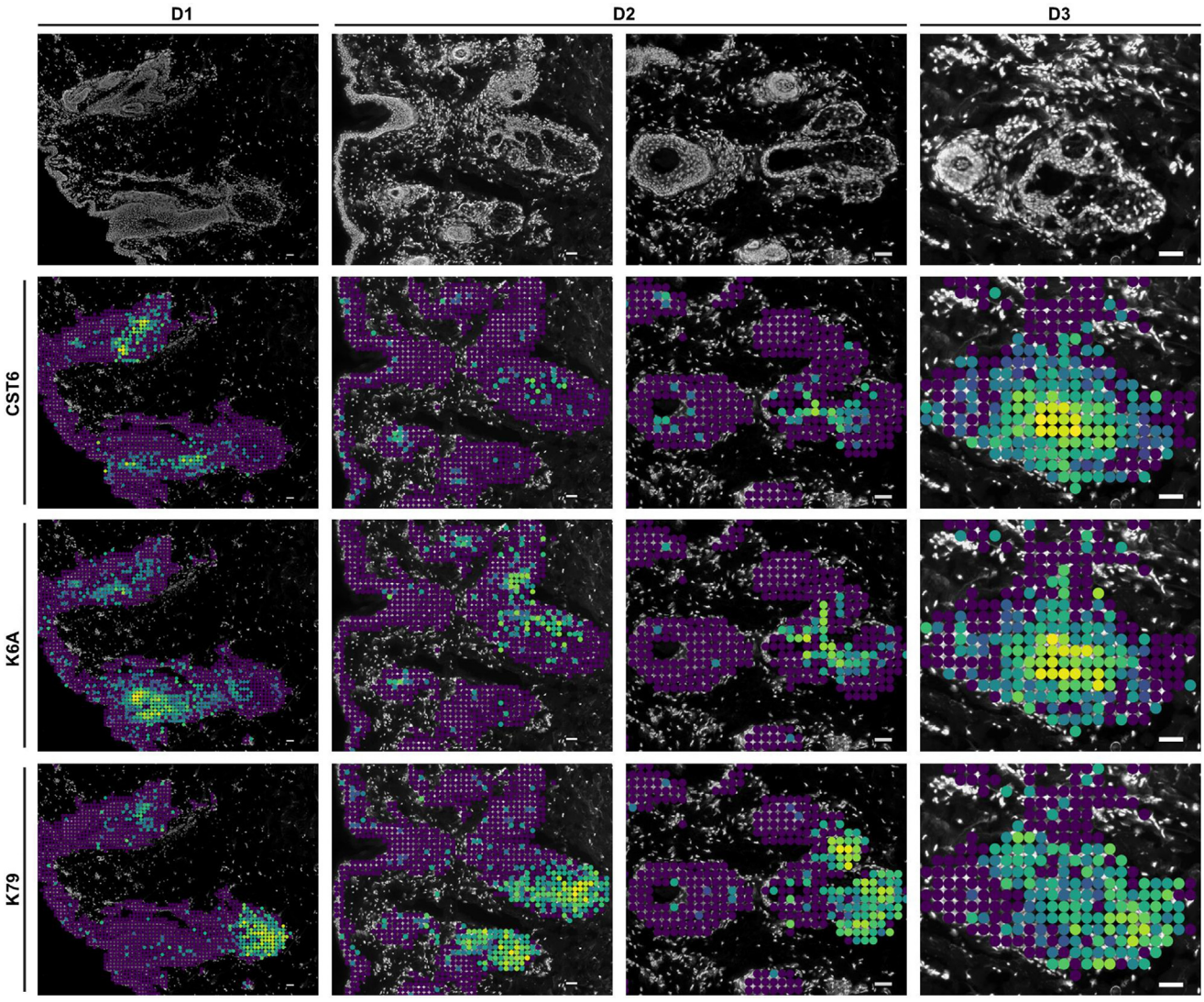
Supplementary Figure S3. Effects of binning sizes on the number of detected features. (**a**) Number of genes plotted against UMI counts for bin sizes of 20 (~10 μm), 50 (~25 μm), and 100 (~50 μm) for tissue section D1. (**b, c**) Violin and feature plots showing the number of detected genes for the binning sizes 20, 50, and 100. UMI, unique molecular identifier.



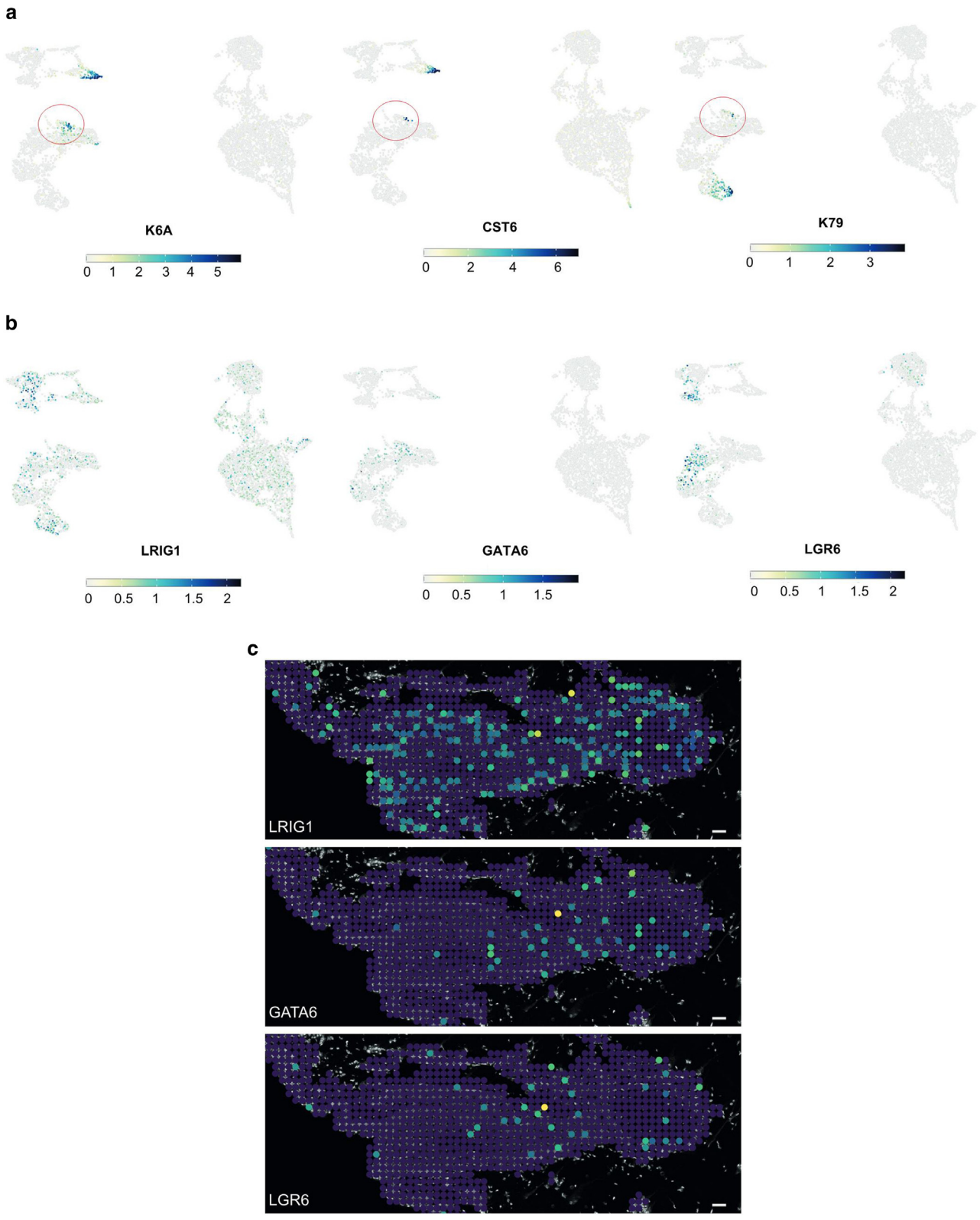
Supplementary Figure S4. Dot plots showing marker genes for cell types in Stereo-seq across different binning resolutions. (**a**–**c**) Binning sizes of 20, 50, and 100, respectively. DP, dermal papilla; DS, dermal sheath; ESG, eccrine sweat gland; FB, fibroblast; IRS, inner root sheet; JZ, junctional zone; KC, keratinocyte; ORS, outer root sheath; SD, sebaceous duct; VEC, vascular endothelial cell; VSMC, vascular smooth muscle cell.



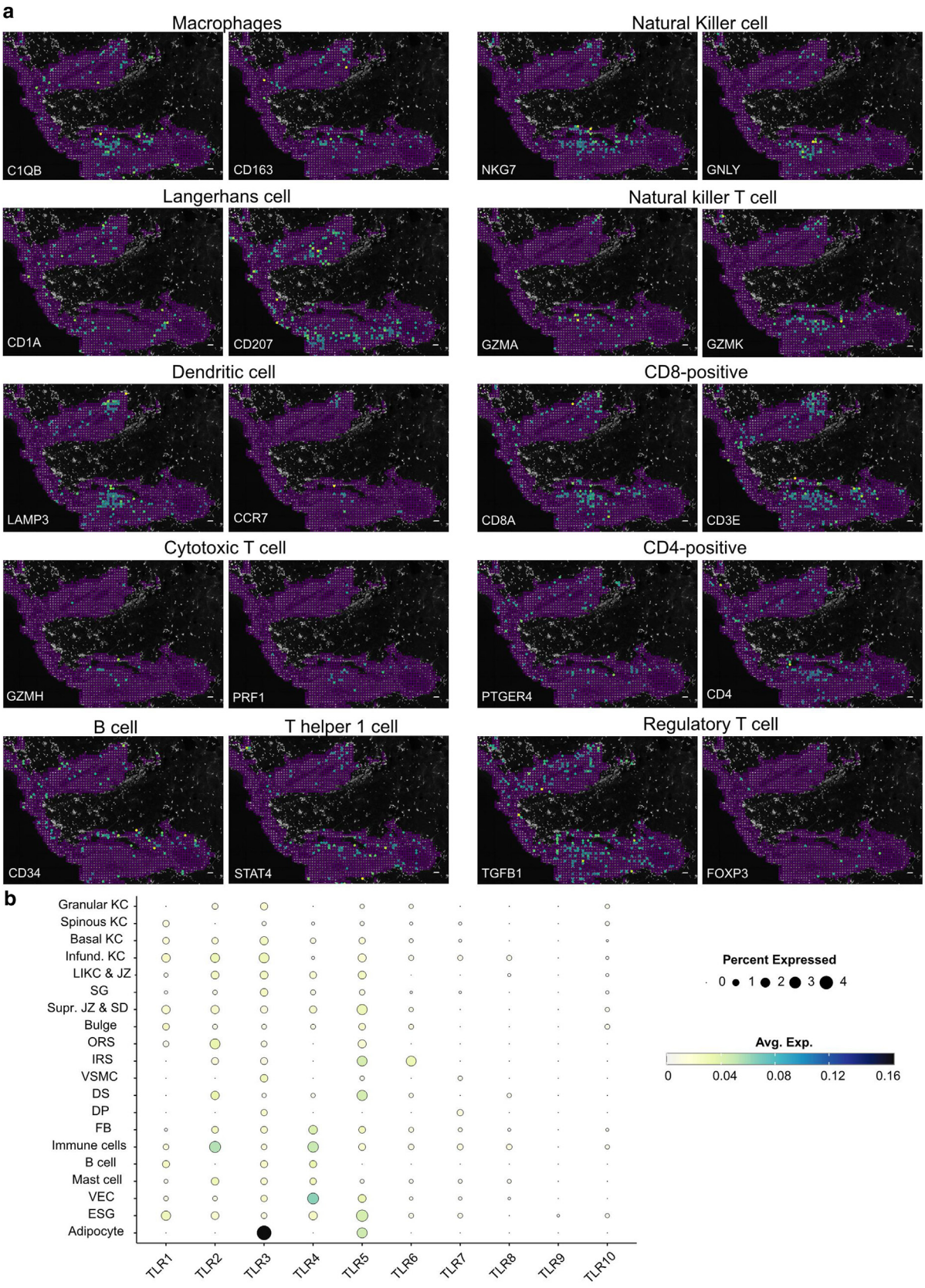
Supplementary Figure S5. Marker genes for cell types in Stereo-seq and functional investigation. (**a**) Heatmap displaying Z-scores of gene module activities detected by Hotspot, with functional annotations through GO analysis. Noisy cell clusters that do not match any cell types are colored in gray. (**b**) Visualization of selected gene modules from **a**: IFE (M2), infundibular keratinocytes (M8), sebaceous gland (M1), outer bulge (M14), ORS (M6), and immune cells (M4). (**c**) Integrated UMAP of the spatial transcriptomics data from the 3 tissue sections, colored by donor, with a bin size set to 50 (~25 μm). GO, Gene Ontology; IFE, interfollicular epidermis; ORS, outer root sheath; UMAP, Uniform Manifold Approximation and Projection.



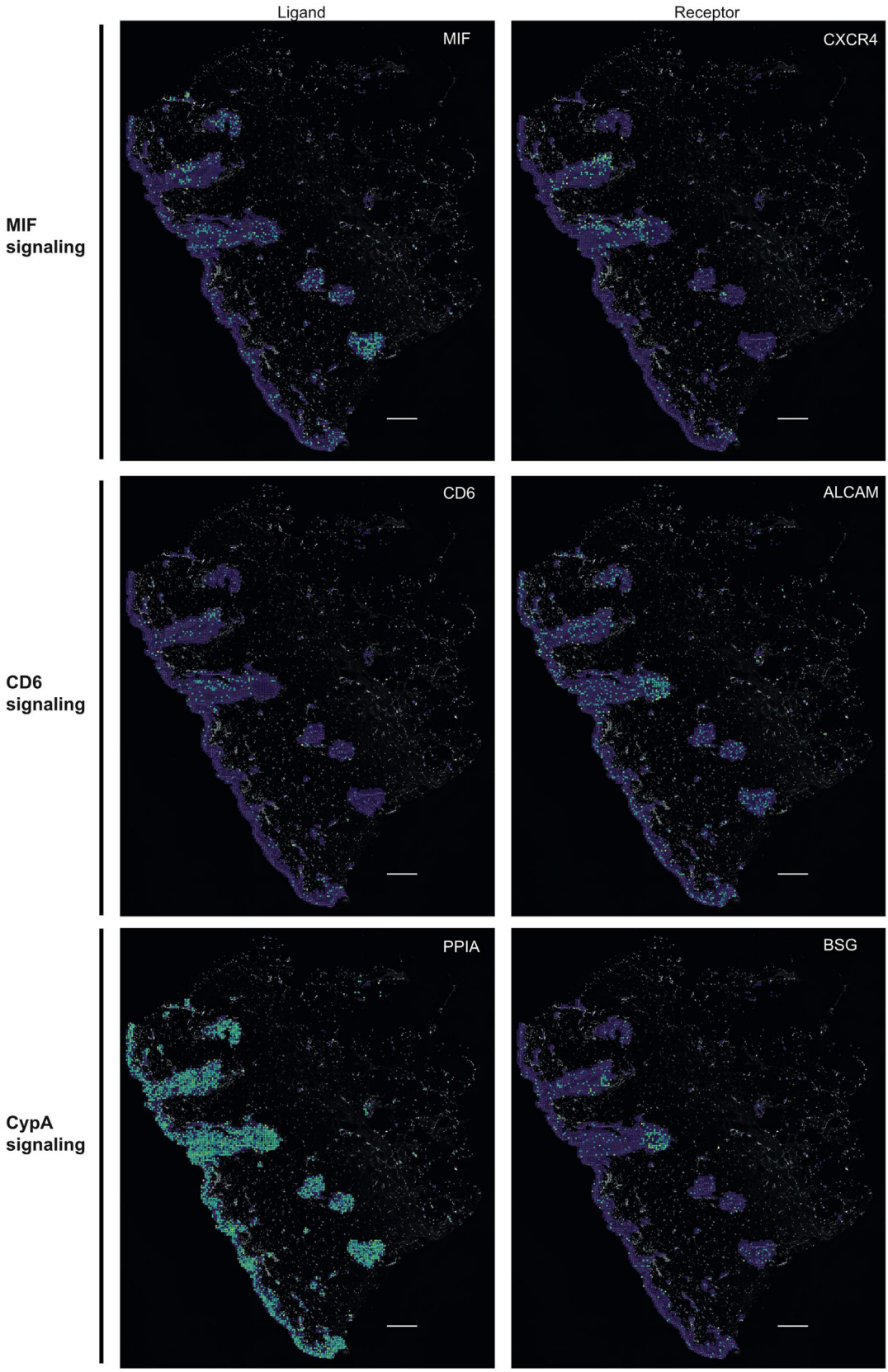
Supplementary Figure S6. Spatial feature plots of genes for the suprabasal junctional zone. The first row shows distinct regions from the tissue sections with sebaceous glands and their connection to the hair follicle through the junctional zone. The rows below display literature markers for the suprabasal JZ and infund. KC, including CST6, K6A, and K79. Bar = 25 μm. infund. KC, infundibulum keratinocyte; K, keratin; JZ, junctional zone; KC, keratinocyte.



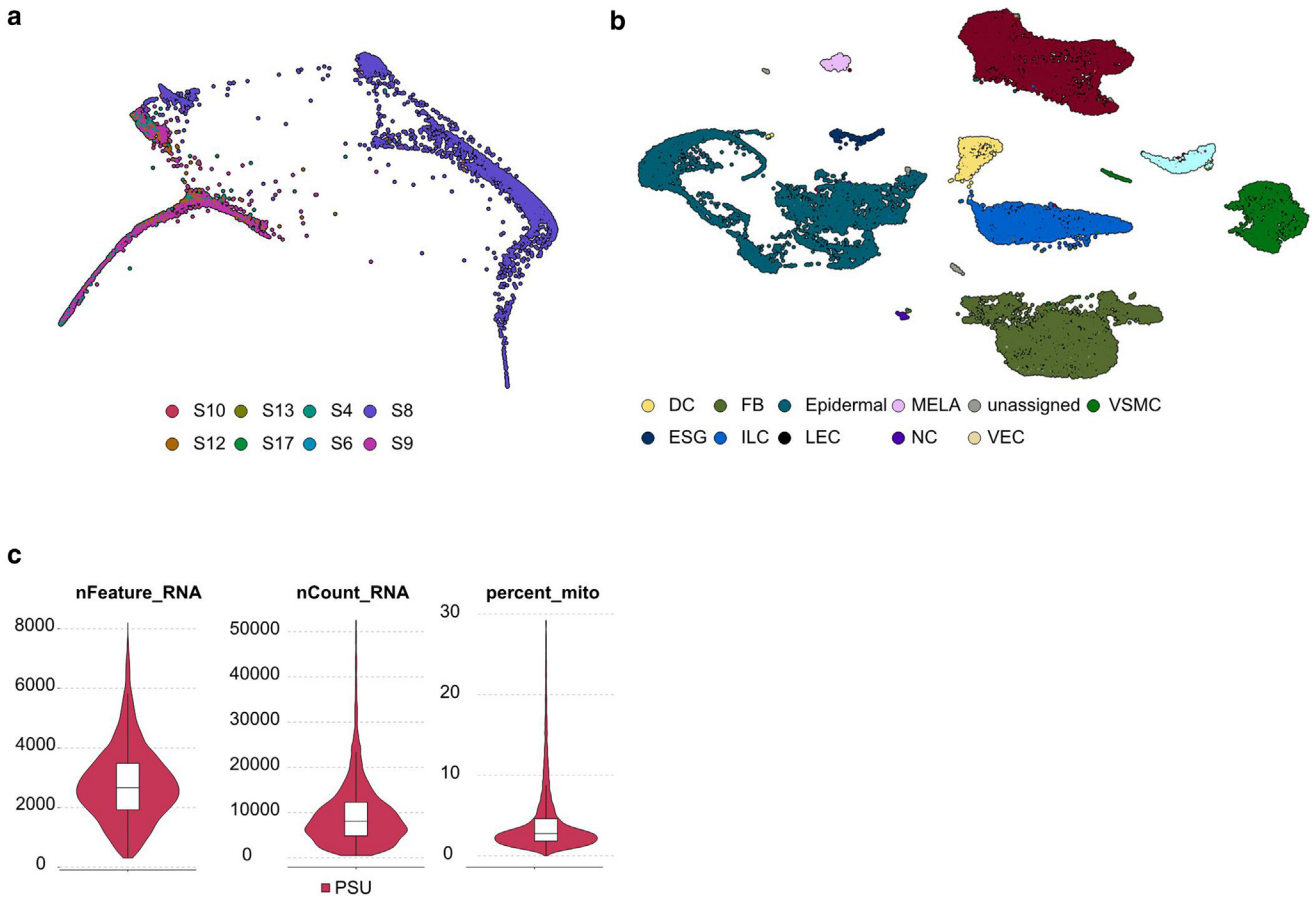
Supplementary Figure S7. Feature plots and spatial visualizations of marker genes from the sebaceous duct and junctional zone. (**a**) Feature plots of K6A, CST6, and K79. The red circle highlights markers of the suprabasal sebaceous duct and junctional zone. (**b**) Feature plots of LRIG1, GATA6, and LGR6. (**c**) Spatial feature plots of LRIG1, GATA6, and LGR6 for the follicle section with the attached sebaceous gland in section D1. Bar = 25 μm. K, keratin.



Supplementary Figure S8. Spatial feature plots of gene markers for various immune cells. (**a**) Gene markers for various immune cells. (**b**) Dot plot of selected toll-like receptors for all cell types in Stereo-seq. Bar = 25 μm. DP, dermal papilla; DS, dermal sheath; ESG, eccrine sweat gland; JZ, junctional zone; KC, keratinocyte; ORS, outer root sheath; TLR, toll-like receptor; VEC, vascular endothelial cell; VSMC, vascular smooth muscle cell.

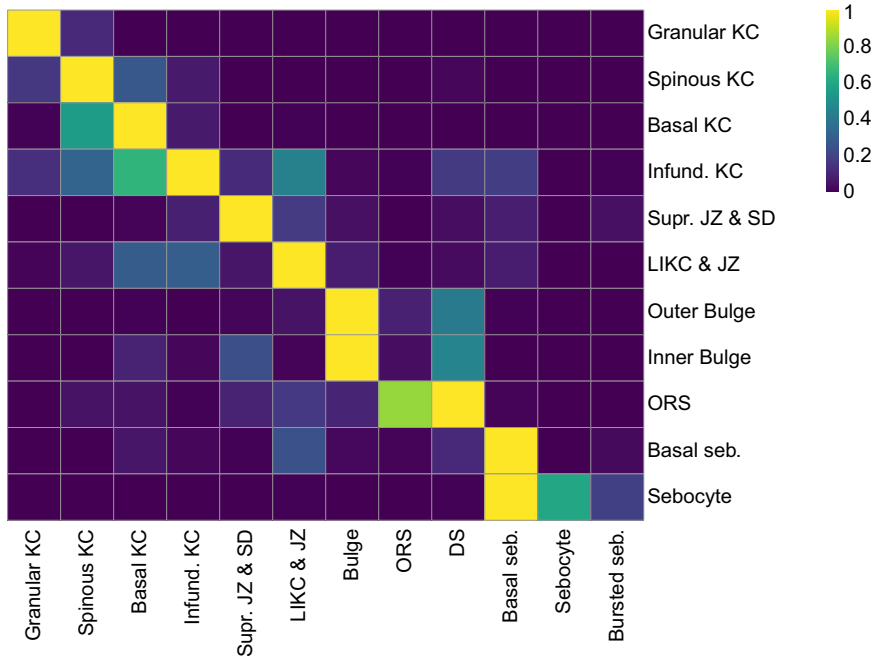


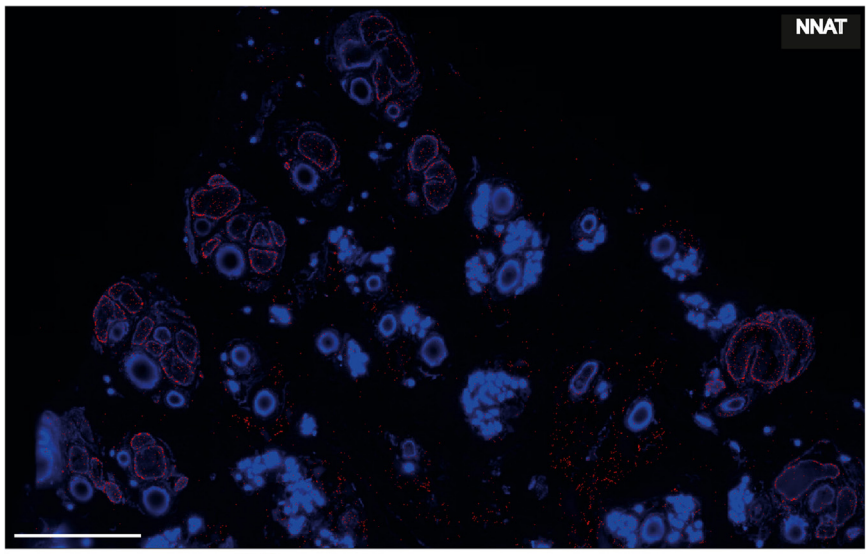
Supplementary Figure S9. Spatial feature plots of ligand and receptor expression for the MIF, CD6, and CypA signaling pathways. Bar = 200 μm. CyPA, cyclophilin A; MIF, macrophage migration inhibitory factor.

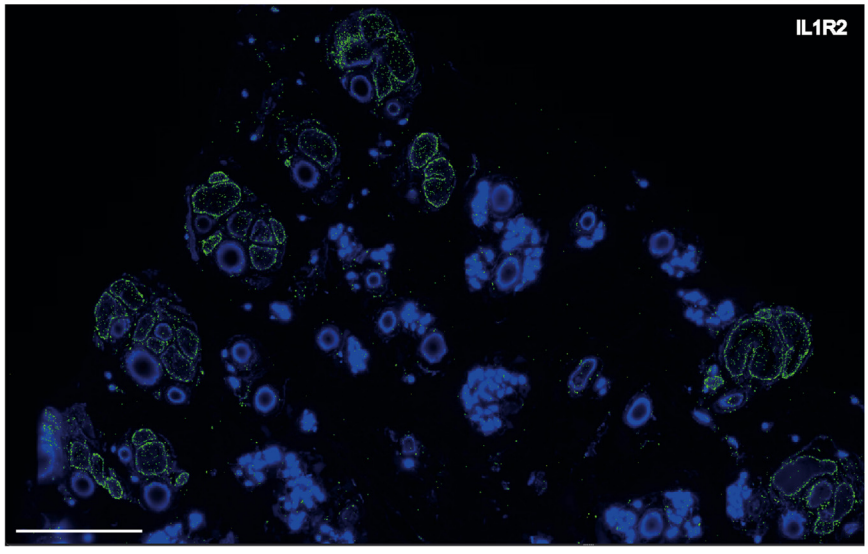


Supplementary Figure S10. Metadata for the scRNA-seq object. (**a**) scRNA-seq object from Figure 5a, colored by patient. (**b**) UMAP embedding of the full scRNA-seq dataset, colored by cell types. (**c**) Violin plots of the scRNA-seq object from Figure 5a based on quality control parameters: number of genes detected (nFeature_RNA), total read counts (nCount_RNA), and percentage of mitochondrial genes (percent_mito). DC, dendritic cell; ESG, eccrine sweat gland; FB, fibroblast; ILC, innate lymphoid cell; KC, keratinocyte; LEC, lymphatic endothelial cell; MELA, melanocyte; NC, neuronal cell; scRNA-seq, single-cell RNA sequencing; SG, sebaceous gland; UMAP, Uniform Manifold Approximation and Projection; VEC, vascular endothelial cell; VSMC, vascular smooth muscle cell.

Supplementary Figure S11. Probabilistic transfer scores from scRNA-seq to Stereo-seq. The correspondence between cell types identified by Stereo-seq (bottom) and cell clusters from scRNA-seq (right) was determined using the anchor-based integration workflow introduced in Seurat, version 3. This method facilitated the probabilistic transfer of cell-type annotations from scRNA-seq to Stereo-seq data. Prediction scores for each spot were calculated and visualized in a heatmap, using Min–Max normalization. DS, dermal sheath; JZ, junctional zone; KC, keratinocyte; LIKC, lower infundibular keratinocyte; Max, maximum; Min, minimum; ORS, outer root sheath; scRNA-seq, single-cell RNA sequencing; SD, sebaceous duct.

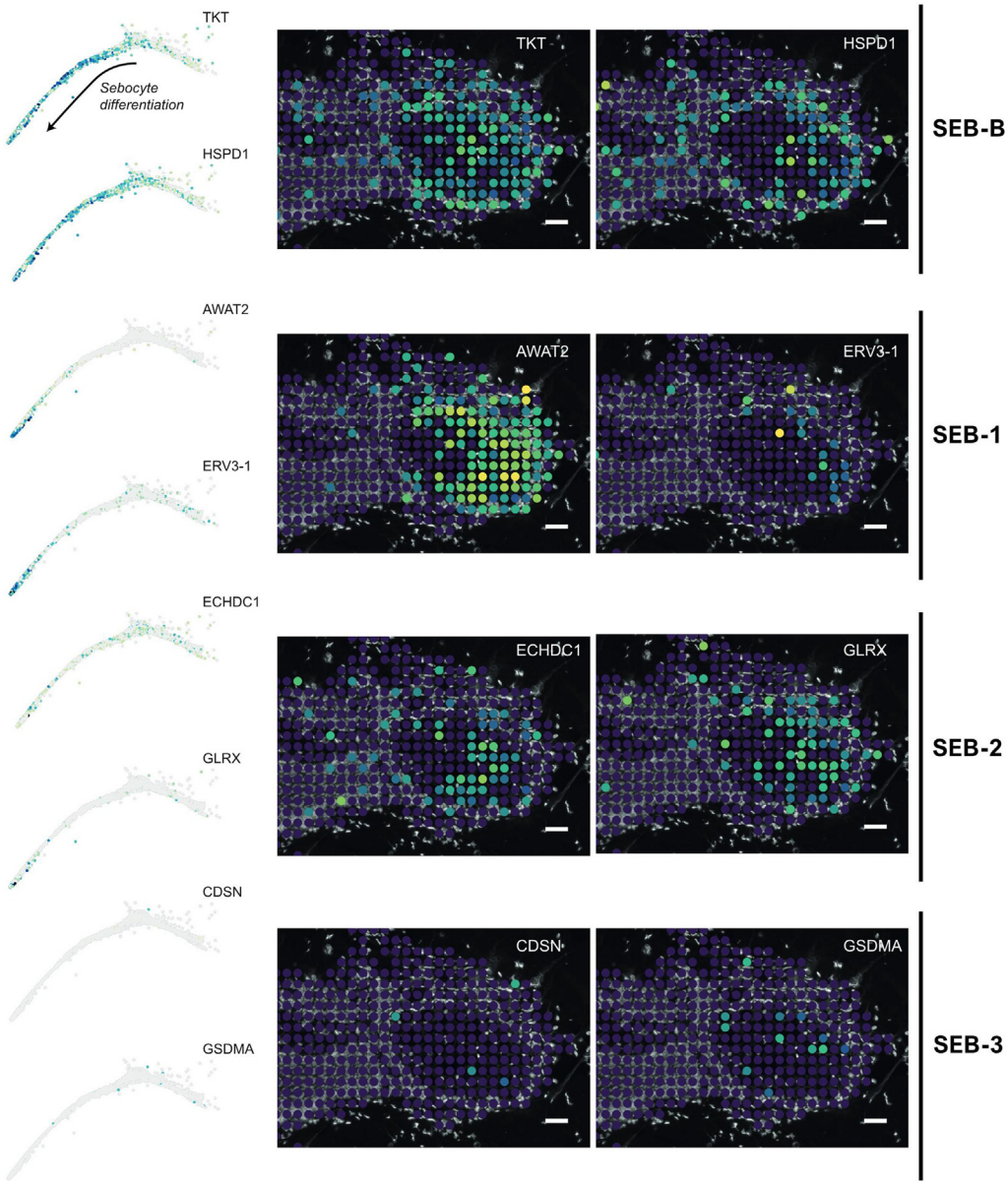


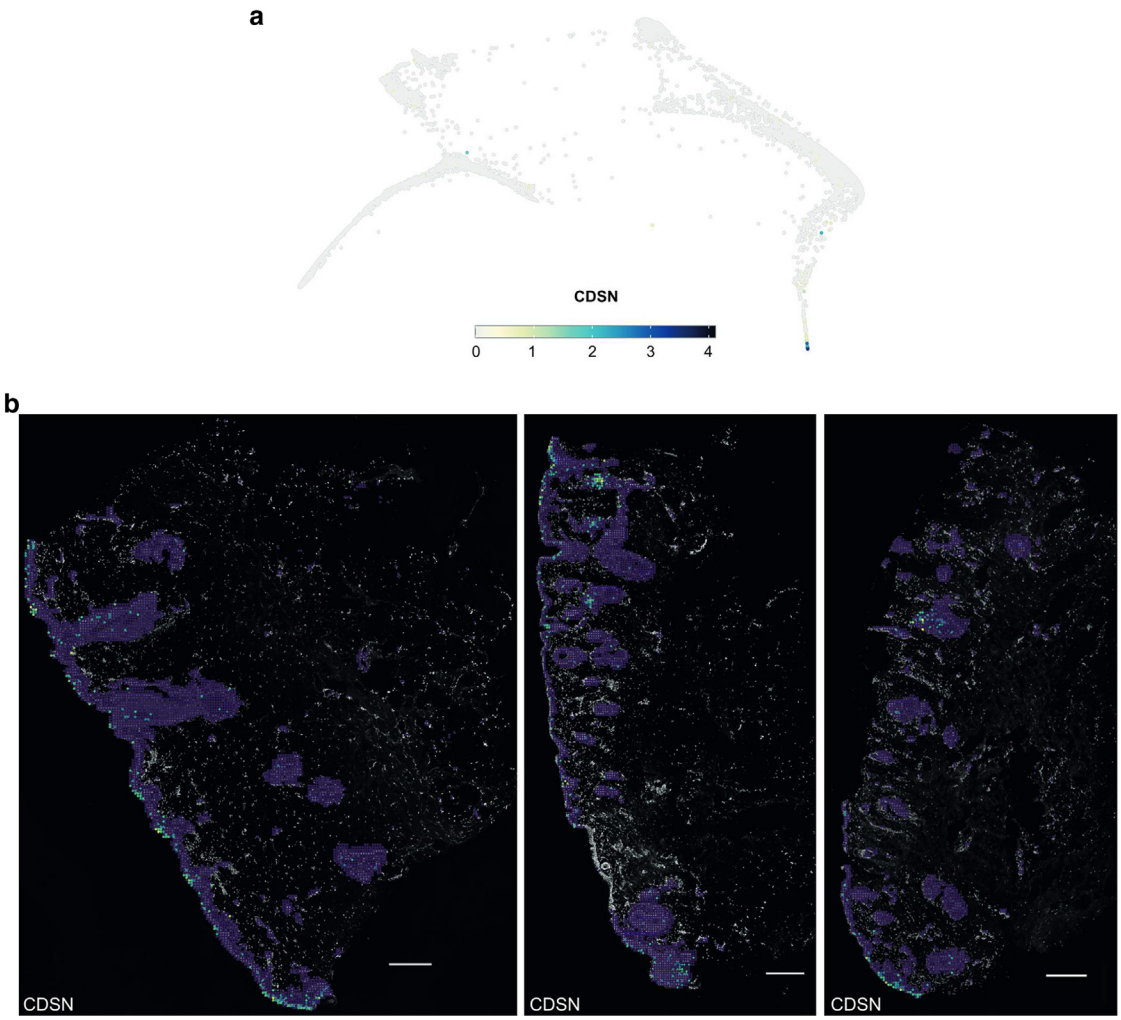




Supplementary Figure S12. MERFISH-based visualization of basal sebocyte gene expression. Bar = 1 mm. MERFISH, multiplexed error-robust FISH.

Supplementary Figure S13. Investigation of sebocyte differentiation signatures from Schmidt et al (2024) using scRNA-seq and Stereo-seq. Bar = 25 μm. scRNA-seq, single-cell RNA sequencing.

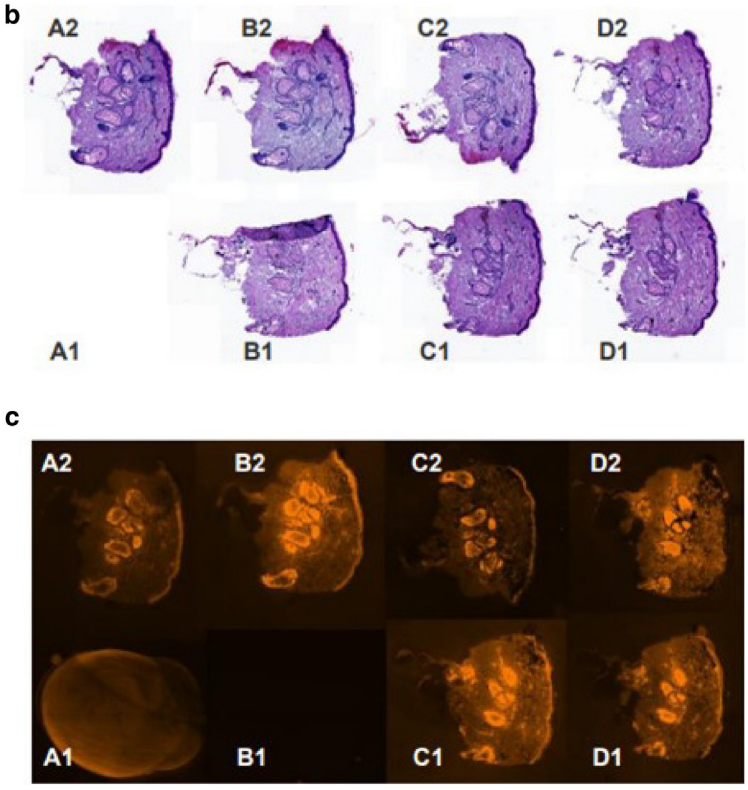




Supplementary Figure S14. Feature plot and spatial feature plot of CDSN expression. (**a**) Feature plot of CDSN expression on the PHATE embedding of the scRNA-seq dataset from Figure 5a. (**b**) Spatial feature plots of CDSN across all tissue sections. Bar = 200 μm. PHATE, Potential of the Heat-diffusion for Affinity-based Transition Embedding; scRNA-seq, single-cell RNA sequencing.

a

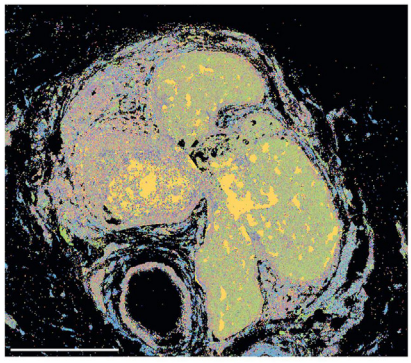
Position	Loading	Permeabilization time (min)
A1	Control RNA	0
B1	Tissue Sections 10 μm each	0
C1		60
D1		45
A2		30
B2		20
C2		10
D2		5



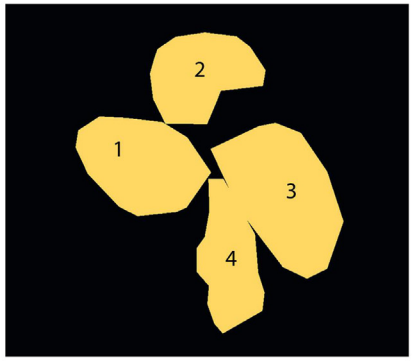
Supplementary Figure S15. Optimization of permeabilization time based on previous tissue sections. (**a**) Optimization of permeabilization time. Tissue sections were cut at intervals of 10 μm and permeabilized for 0, 5, 10, 20, 45, and 60 minutes. (**b**) H&E staining. (**c**) Fluorescence signal, with 10 minutes (C2) showing the least signal diffusion for the SG. SG, sebaceous gland.

Supplementary Figure S16. Quality control of transcript counts across different ROIs in MERFISH analysis of sebaceous gland lobes. (**a**) Visualization of total transcript counts in a sebaceous gland, with ROIs highlighted in yellow. Bar = 250 μm. (**b**) Same visualization as in **a** but without transcript counts, showing only ROIs, numbered 1−4 for each sebaceous gland lobe. (**c**) Table displaying volume, transcript count, and cell count for the different ROIs. MERFISH, multiplexed error-robust FISH; ROI, region of interest.

a



b



c

	1	2	3	4
VOLUME	48595.1	37638.6	63672.2	34061
TRANSCRIPT COUNT	84671	164888	203885	134218
CELLS COUNT	299	130	160	112