A Roadmap for a Consensus Human Skin Cell Atlas and Single-Cell Data Standardization

JID Open

Axel A. Almet[1,2], Hao Yuan[3], Karl Annusver[3], Raul Ramos[2,4,5], Yingzi Liu[4,5], Julie Wiedemann[4,6], Dara H. Sorkin[7,8], Ning Xu Landén[9,10,11], Enikö Sonkoly[9,10,12], Muzlifah Haniffa[13,14,15], Qing Nie[1,2,4], Beate M. Lichtenberger[16], Malte D. Luecken[17,18], Bogi Andersen[2,5,8,19], Lam C. Tsoi[20,21,22,23], Fiona M. Watt[24,25], Johann E. Gudjonsson[20], Maksim V. Plikus[2,4,5] and Maria Kasper[3]

Single-cell technologies have become essential to driving discovery in both basic and translational investigative dermatology. Despite the multitude of available datasets, a central reference atlas of normal human skin, which can serve as a reference resource for skin cell types, cell states, and their molecular signatures, is still lacking. For any such atlas to receive broad acceptance, participation by many investigators during atlas construction is an essential prerequisite. As part of the Human Cell Atlas project, we have assembled a Skin Biological Network to build a consensus Human Skin Cell Atlas and outline a roadmap toward that goal. We define the drivers of skin diversity to be considered when selecting sequencing datasets for the atlas and list practical hurdles during skin sampling that can result in data gaps and impede comprehensive representation and technical considerations for tissue processing and computational analysis, the accounting for which should minimize biases in cell type enrichments and exclusions and decrease batch effects. By outlining our goals for Atlas 1.0, we discuss how it will uncover new aspects of skin biology.

Journal of Investigative Dermatology (2023) **143,** 1667–1677; doi:10.1016/j.jid.2023.03.1679

WHY A SKIN CELL ATLAS?

The skin contains diverse cell lineages, including epithelial cells of the epidermis and ectodermal appendages—hair follicles, nails, sebaceous glands, and sweat glands—which exist in close association with mesenchymal cell lineages, including smooth muscle cells, adipocytes, and fibroblasts. The latter produces extracellular matrix for mechanical support and signals that guide immune and epithelial cell behavior across both spatial dimensions of the skin (e.g., epidermal differentiation at the surface) and time (e.g., cyclic growth of hair follicles). In addition to the principal skin cell types, there are less abundant cell types essential for skin function, including pigment-producing melanocytes, innate and adaptive immune cells, vascular and perivascular cells, and cells of neuroendocrine origin. Working together, these cell populations form a barrier organ—so large that no individual dataset can sample the entire skin—that plays mechanoprotective, UV-shielding, antimicrobial, thermoregulatory functions and more (Alexander et al., 2015; Donati et al., 2017; Gurtner et al., 2008; Takeo et al., 2015; Watt, 2014). To support these functions, different skin compartments are richly populated by stem cells that respond to insults by mounting reparative responses. As the skin heals, such as after wounding, it restores anatomical integrity and functions by forming a scar containing new stable cell states that are distinct

[1]*Department of Mathematics, University of California, Irvine, Irvine, California, USA;* [2]*NSF-Simons Center for Multiscale Cell Fate Research, University of California, Irvine, Irvine, California, USA;* [3]*Department of Cell and Molecular Biology, Karolinska Institute, Stockholm, Sweden;* [4]*Department of Developmental and Cell Biology, School of Biological Sciences, University of California, Irvine, Irvine, California, USA;* [5]*Sue and Bill Gross Stem Cell Research Center, University of California, Irvine, Irvine, California, USA;* [6]*Mathematical, Computational & Systems Biology, Department of Medicine, University of California, Irvine, Irvine, California, USA;* [7]*Institute for Clinical & Translational Science, University of California, Irvine, Irvine, California, USA;* [8]*Department of Medicine, School of Medicine, University of California, Irvine, Irvine, California, USA;* [9]*Dermatology and Venereology Division, Department of Medicine, Solna, Karolinska Institute, Stockholm, Sweden;* [10]*Center for Molecular Medicine, Karolinska Institute, Stockholm, Sweden;* [11]*Ming Wai Lau Centre for Reparative Medicine, Karolinska Institute, Stockholm, Sweden;* [12]*Dermatology and Venereology, Department of Medical Sciences, Uppsala University, Uppsala, Sweden;* [13]*Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom;* [14]*Biosciences Institute, Newcastle University, Newcastle Upon Tyne, United Kingdom;* [15]*Department of Dermatology and NIHR Newcastle Biomedical Research Centre, Newcastle Hospitals NHS Foundation Trust, Newcastle Upon Tyne, United Kingdom;* [16]*Skin & Endothelium Research Division (SERD), Department of Dermatology, Medical University of Vienna, Vienna, Austria;* [17]*Institute of Computational Biology, Helmholtz Munich, Neuherberg, Germany;* [18]*Institute of Lung Health and Immunity, Helmholtz Munich, Member of the German Center for Lung Research (DZL), Munich, Germany;* [19]*Department of Biological Chemistry, School of Medicine, University of California, Irvine, Irvine, California, USA;* [20]*Department of Dermatology, University of Michigan, Ann Arbor, Michigan, USA;* [21]*Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA;* [22]*Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA;* [23]*Center for Statistical Genetics, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA;* [24]*Centre for Gene Therapy & Regenerative Medicine, Faculty of Life Sciences & Medicine, School of Basic & Medical Biosciences, King's College London, London, United Kingdom; and* [25]*Directors' Research Unit, European Molecular Biology Laboratory, Heidelberg, Germany*

Correspondence: Maksim V. Plikus, Department of Developmental and Cell Biology, University of California, Irvine, 845 Health Sciences Road, Room 3018, Irvine, California 92697, USA. E-mail: plikus@uci.edu or Maria Kasper, Department of Cell and Molecular Biology, Karolinska Institute, Stockholm, Sweden. E-mail: maria.kasper@ki.se

Abbreviations: DNA-seq, DNA sequencing; HCA, Human Cell Atlas; HSCA, Human Skin Cell Atlas; scATAC-seq, single-cell assay for transposase-accessible chromatin sequencing; scRNA-seq, single-cell RNA sequencing; snRNA-seq, single-nucleus RNA-sequencing

Received 24 March 2023; revised 24 March 2023; accepted 29 March 2023

from unwounded cell states (Donati et al., 2017; Gurtner et al., 2008; Sun et al., 2022; Takeo et al., 2015).

Several hundred clinically distinct disorders, both monogenic and multifactorial, affect human skin (Feramisco et al., 2009). Although many skin diseases are well-characterized clinically, histologically, genetically, and by bulk biochemical assays, deep mechanistic understanding remains obscured in part by a lack of characterization at single-cell resolution. Moreover, numerous distinct diseases have near-identical clinical manifestations (Feramisco et al., 2009; Lamartine, 2003), challenging correct diagnosis and resulting in ineffective therapies. Certain diseases, such as psoriasis, are spatially predisposed to appear in certain body regions (Dhabale and Nagpure, 2022). Other diseases, such as facial acne, disproportionately occur at certain ages (Williams et al., 2012), and other diseases occur more commonly in females than in males, such as scleroderma (Andersen and Davis, 2016). Diseases such as keloids are more common in certain ancestries (Chike-Obi et al., 2009) and require tissue injury to manifest (Tuan and Nichter, 1998). Developing a greater understanding of how different skin regions, age, gender, ancestry, and wound response affect skin cell types and cell states at single-cell resolution will enable further understanding of skin disease mechanisms through the comparison of pathological states with normal skin at multiple scales.

The Skin Biological Network set a goal to build a consensus Human Skin Cell Atlas (HSCA), as part of the Human Cell Atlas (HCA) (humancellatlas.org) (Regev et al., 2017). This effort builds on previous skin cell atlas work under the auspices of the HCA and newer Chan Zuckerberg Initiative (CZI)-supported efforts on pediatric and ancestral skin as well as the work of individual laboratories that have been working independently to profile normal human skin using single-cell RNA sequencing (scRNA-seq) to develop a comprehensive atlas of diverse skin cell types across several scales (Figure 1): (i) space, capturing different body sites (Figure 1a); (ii) time, capturing maturation states of skin at the same body site (e.g., facial skin during prenatal development, childhood, adolescence, adulthood, and advanced age) (Figure 1b); (iii) gender scale,

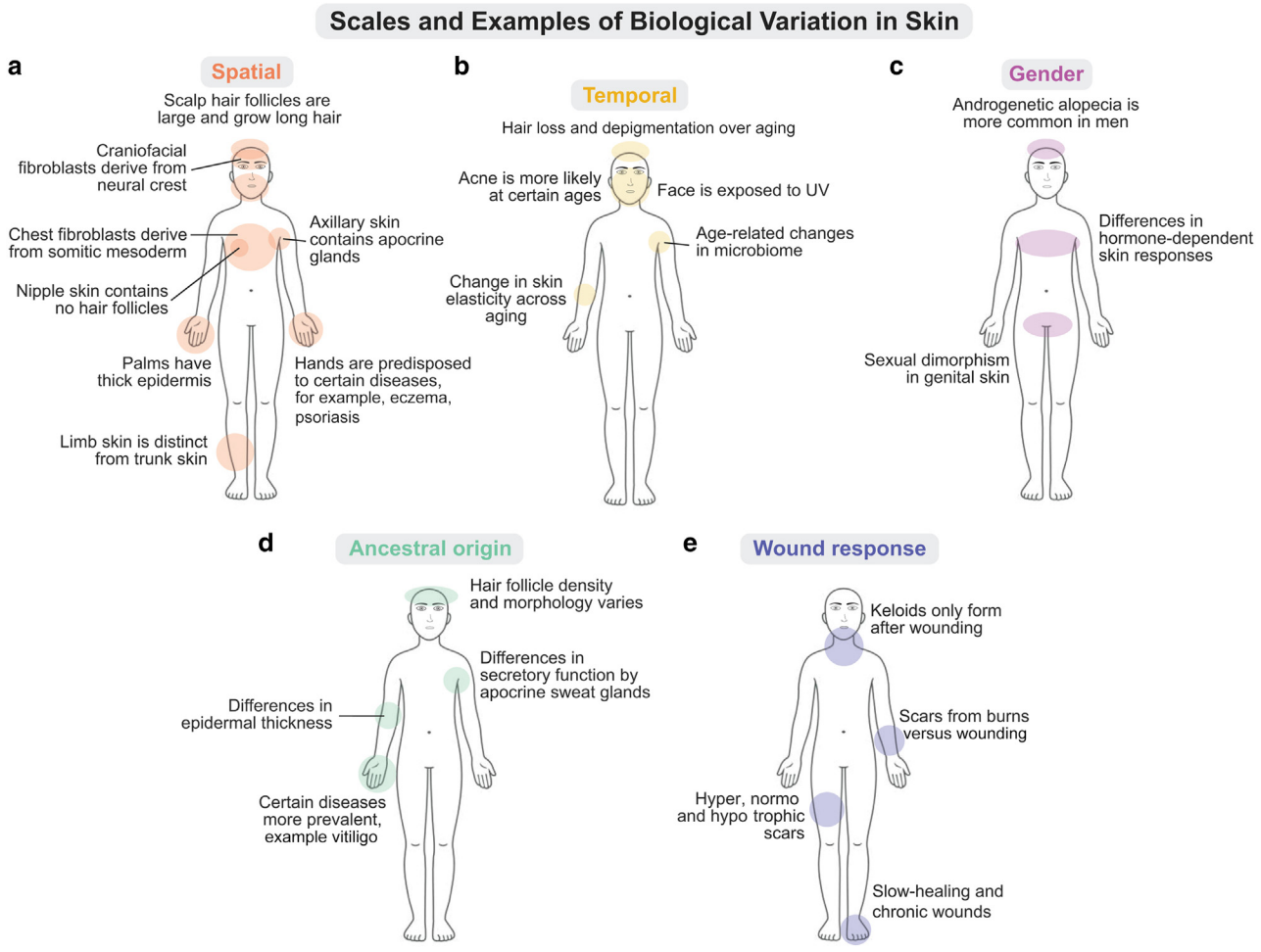

Figure 1. The many scales of biological variation in the skin that we predict will have a significant effect on transcriptional heterogeneity. (**a**) The spatial scale represents how skin composition and characteristics vary across anatomical regions. (**b**) The temporal scale represents changes to skin across the human lifespan. (**c**) The gender scale represents gender-specific differences between skin sites and function. (**d**) The ancestral origin scale, which affects skin characteristics and proclivity to disease. (**e**) The wound response scale, where unwounded skin is distinct from the skin that is permanently altered after the innate, acute wound repair program. For each scale, we include notable examples that illustrate why these factors need to be considered when generating a high-quality single-cell skin atlas.

capturing major sexually dimorphic skin regions and changes associated with puberty (Figure 1c); (iv) ancestry, capturing cutaneous anatomical features that prominently vary across ancestries, such as skin pigmentation, sweat gland, and hair follicle differences (Figure 1d); and (v) wound response scale, capturing new cell states of healed skin, that is, scars versus unwounded skin (Figure 1e).

WHAT ARE THE KEY CONSIDERATIONS FOR THE HSCA?

Aspects contributing to physiological diversity of the skin

To represent the full anatomical and functional diversity of skin at a single-cell level, several factors must be considered. First, skin from different body sites can have different embryonic origins. For example, dermal fibroblasts in frontal scalp skin arise from the craniofacial mesoderm, whereas fibroblasts in the chest skin originate from the somitic mesoderm (Thulabandu et al., 2018). Second, different anatomical skin sites have different dominant features. For example, eyelids have a thin epidermis, whereas palms and feet have significantly thicker epidermis (Sandby-Møller et al., 2003). Skin sites also vary by hair follicle size, density, and growth cycle parameters: scalp skin has large (i.e., terminal) hair follicles that produce long hairs continuously over several years, whereas adjacent forehead skin features diminutive (i.e., vellus) hair follicles that grow short, thin hairs (Vogt et al., 2007), and areola, lip skin, and palmoplantar skin are largely devoid of hair follicles (Stone and Wheeler, 2015; Tsai et al., 2022). Third, the skin microbiome profile varies significantly across distinct body sites and correlates with each site's physical properties (Byrd et al., 2018), that is, whether the skin is moist (e.g., axillary or popliteal fossa skin), dry (e.g., volar forearm skin), or oily (e.g., forehead skin). Fourth, UV exposure has a prominent physiological effect on the skin; its extent and history naturally differ across anatomical space and are further impacted by social dress codes and skin color (Matsumura and Ananthaswamy, 2004). For instance, sun-exposed forehead skin commonly experiences high UV exposure compared with typically clothed buttock skin. Fifth, many skin sites, particularly facial, scalp, axillary, chest, and pubic skin, undergo prominent age-dependent changes (Farage et al., 2013; Haydont et al., 2019) and differ between genders (Dao and Kazin, 2007). Finally, given the inherent anatomical and physiological diversity across human populations, single-cell datasets from genetically diverse ancestries should be included to generate a comprehensive atlas. Priority should be placed on ancestral groups that are underrepresented in biomedical research, including but not limited to African, Asian, Hispanic, and Middle Eastern populations (Hirano et al., 2012; Ma et al., 2021a). A further priority should be skin sites with prominent morphological differences across ancestral groups, such as scalp skin, where there is variation in hair follicle density and hair morphology, or axillary skin, where there are differences in apocrine sweat gland function (Luther et al., 2012) and propensity to disease (Kilgour et al., 2021).

Practical considerations for sample collection

Sample acquisition is easier for certain body sites and states (e.g., aged vs. younger skin), and unless proactively countered, knowledge gaps in skin-wide data will persist (Figure 2). Fresh skin samples can be acquired in two main ways. The first is from discarded tissue that can be collected during certain routine surgical procedures, including (i) blepharoplasty (upper eyelid skin), (ii) facelifts (temporal, frontal, and parotid facial skin), (iii) abdominoplasty (hypogastric abdominal skin), (iv) mammoplasty (chest and areola skin), and (v) hair transplantation (occipital scalp skin). Because plastic surgery procedures commonly remove old scars (e.g., C-section scars during abdominoplasty), scar tissue can also be readily obtained. Second, normal skin samples can be obtained through biopsies from healthy volunteers. For certain skin sites, including trunk skin (e.g., from the upper buttocks region) and the extremities (e.g., from thighs or forearms), obtaining biopsies will typically result in minimal morbidity and is well-tolerated. Other skin sites, such as genital skin and axillary skin, can rarely be sampled without resulting in significant morbidity and will require specialized arrangements.

DISTINCT CHALLENGES

Sampling skin to saturation

To date, predominantly fresh skin tissue has resulted in high-quality single-cell skin transcriptomes. For the reasons discussed earlier, many anatomical skin sites will be generally precluded from sampling, leading to underrepresentation of certain skin regions and their distinct anatomies and functions in Atlas 1.0. Specific efforts will need to be undertaken to minimize anticipated data gaps. One possible solution could be greater use of postmortem tissues, albeit with the additional challenge that these tissues must be as fresh as possible to obtain high-quality data.

Global representation of human skin heterogeneity

A representative skin cell atlas must reflect the entire human population. However, the desire to include diverse populations and those typically underrepresented in clinical research is met with multiple challenges (Mapes et al., 2020; Swartz et al., 2019) that should be addressed with directed community-engagement approaches (Borthwick et al., 2023; Holzer et al., 2014). First, low income is a barrier to participation. Indeed, low-income individuals are often excluded from clinical research, partly because individuals seen at locations that serve low-income patients are often not informed about potential opportunities for clinical trial participation, or they may not be able to afford surgery. Second, language and literacy can be a barrier to participation (Nicholson et al., 2015). Low reading literacy can make reliance on written recruitment materials ineffective. This can be partly aided through translators or by including native speakers on the research team, but cultural and language tailoring must go deeper. One should seek help from cultural and language-concordant research facilitators who can support bidirectional communication and knowledge transmission between researchers, participants, and their communities. Third, mistrust is a barrier to participation. A long history of medical and scientific exploitation has targeted and adversely affected diverse groups of people, creating perceptions of mistrust toward biomedical research.

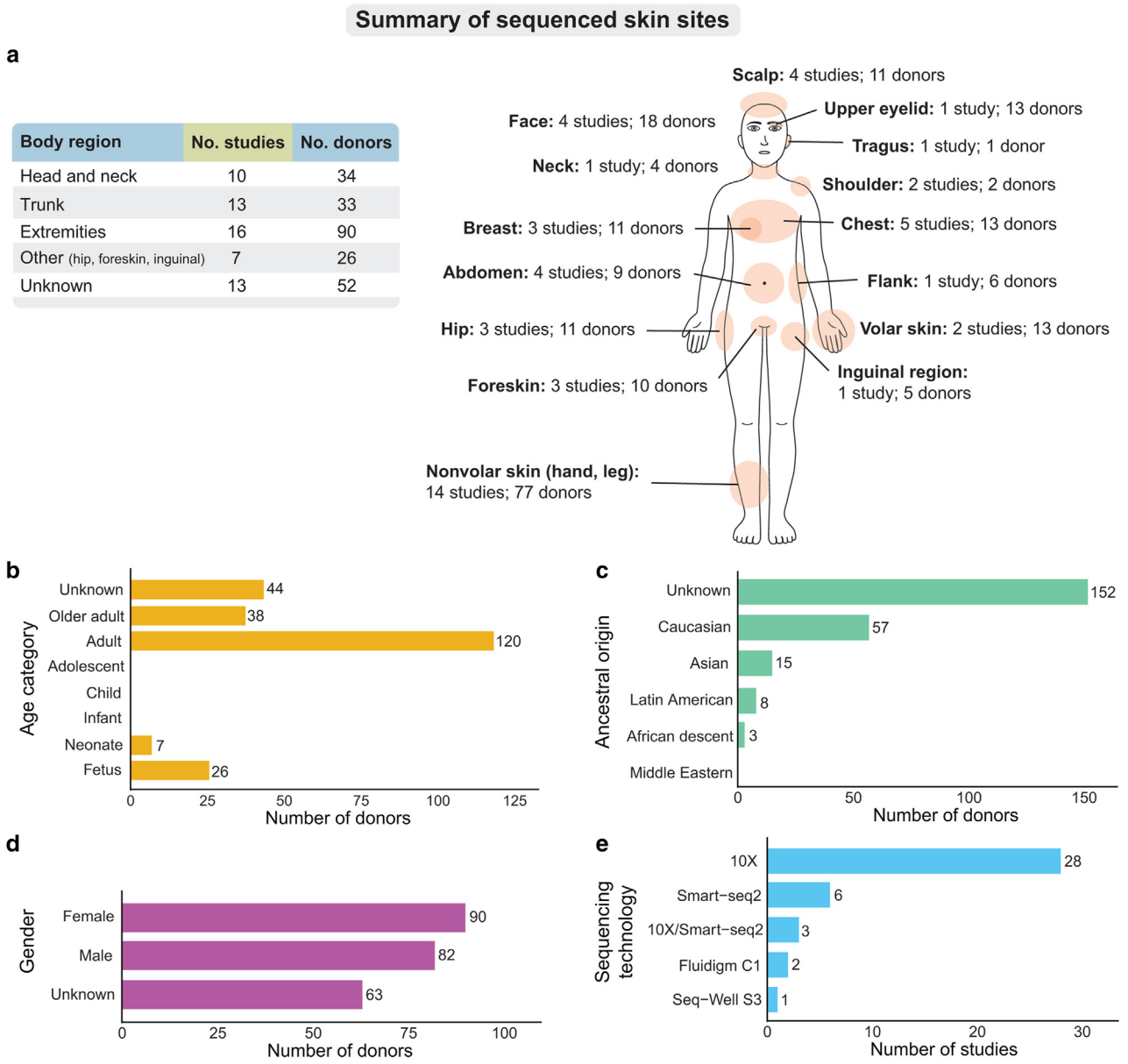

Figure 2. Summary of current skin scRNA-seq datasets in the literature. (**a**) A summary of the number of studies and skin donors per major body region. Head and neck sites include the scalp, face, upper eyelid, neck, and tragus. Trunk sites include the shoulder, chest, abdomen, breast, and flank. Extremity sites include both volar skin (palm and sole) and nonvolar (hand and leg) skin. (**b**–**e**) The number of skin donors summarized with respect to (**b**) age category, (**c**) ancestral origin, (**d**) gender, and (**e**) sequencing technology. There are numerous gaps in the literature that can be filled when constructing new data for Atlas 1.0. No., number; scRNA-req, single-cell RNA sequencing.

One critical way to combat mistrust is through clear communication and transparency by making a conscious effort to ensure that research materials (i.e., informed consent documents, study results, etc.) are designed in ways that promote a clear understanding of the research questions, study design, participant protections, and potential community benefit (Day et al., 2020).

We note that this section specifically refers to collecting skin samples from underrepresented groups in the United States, but more broadly, researchers worldwide should be recruited to obtain these underrepresented samples. We propose to recruit participants through deep engagement with the local community when the benefits of participation in such research are clearly explained and all questions are answered beforehand. One must work closely with community partners to understand not only why people may be interested in participating in skin research but also to ensure that all recruitment materials and consenting documents are easy to understand and facilitate individuals making informed consent.

Participants of these studies should certainly receive monetary compensation for their time and effort and for study-related expenses. For example, researchers must recognize that being given time off from work (e.g., sick or vacation) is a benefit that not every employer provides, so if under-resourced people are to be included in the research, appropriate compensation, but not financial inducement, must be provided.

Comprehensive cell type analysis

When collecting representative skin data, both tissue handling and digestion protocols significantly impact the efficiency of cell isolation and viability, with uneven effects across cell types. For example, trimming adipose tissue before cell isolation leads to the underrepresentation of cell types residing in the dermal adipose layer. Adipose minimization artifacts and loss of full-depth anagen hair follicle cell coverage also happen when skin is not biopsied at full thickness. Different skin microstructures demand different enzymatic dissociation protocols, making balanced isolation of viable cells from the epidermis, dermis, and appendages challenging. For example, dermal papilla fibroblasts from anagen scalp hair follicles are difficult to isolate because they are encased in a compact extracellular matrix and further enveloped by the epithelial matrix. Whereas some dissociation protocols preferentially yield epithelial cells, others are better optimized for cells from the dermis. Other factors affecting cell type coverage include (i) whether the dataset was sequenced from fresh versus frozen tissue, (ii) using a live-cell or nuclei-based protocol, (iii) whether there was cell type enrichment before profiling, or (iv) whether the platform is plate or droplet based. Thus, each dataset included in the atlas should be accompanied by a reference and/or detailed documentation, including biological and experimental metadata and the used cell isolation method.

Balanced batch effect removal

A representative skin cell atlas requires the integration of samples from different body sites that exhibit both commonalities and substantial differences in microanatomies, functions, and gene expression signatures. For example, skin also shares many essential anatomical and functional similarities across different sites, such as the presence of an outer epidermal barrier, an elastic and tear-resistant dermis, a dynamic immune niche, or hair growth. In contrast, axillary and breast-associated skin have hormone-rich signaling environments that may result in body site–specific cell subclusters. Unlike laboratory mouse models, scRNA-seq data sampled from human individuals exhibit significant batch effects, even when sampled from the same anatomical site. Variation in cell composition and gene expression can occur owing to differences in individual factors, such as gender, age, ancestral origin, and lifestyle factors (e.g., UV exposure, environmental pollution, and nutritional preferences) but also owing to differences in technical factors, such as sampling method, dissociation protocol, or library preparation protocol. These batch effects must be accounted for and, if possible, corrected to ensure that they do not mask genuine biological differences. The central challenge in constructing a skin cell atlas is that differences in gene expression patterns between samples may consist of both batch effects and meaningful biological differences. To integrate these datasets, we must remove batch effects from the data while retaining biological variation, that is, to preserve both transcriptional similarities of common cell lineages from different body sites and significant axes of biological variation within a lineage, which may reflect cell type adaptation to a particular spatial niche. For example, hair follicle keratinocytes from the scalp are expected to be more like hair follicle keratinocytes from the trunk owing to their hair-producing functions rather than like barrier-forming epidermal keratinocytes from the scalp. Therefore, the HSCA requires computational methodologies that strike the right balance between the two important yet diametrically opposed technical challenges of biological conservation and batch correction (Luecken et al., 2022). In addition, it is important to integrate multiple datasets from each major axis of biological variation that we aim to represent in the atlas. For example, to show biological variation across different body sites, we must have at least two donors per site. This allows for validation that batch effects were removed (e.g., cells from different datasets per site are well-mixed), whereas biological differences are preserved (e.g., cells sampled from different locations do not mix).

Despite the plethora of methods developed to integrate multiple scRNA-seq datasets, there is yet no universal method (Chazarra-Gil et al., 2021; Luecken et al., 2022; Thi et al., 2020). Some methods, for example, fastMNN (Haghverdi et al., 2018), scANVI (Xu et al., 2021), and scGen (Lotfollahi et al., 2019), are more effective at ensuring biological conservation, at the risk of retaining some batch effects, whereas others, for example, BBKNN (Polański et al., 2020), Scanorama (Hie et al., 2019), and trVAE (Lotfollahi et al., 2020), are more effective at batch correction at the risk of obscuring biological variation (Luecken et al., 2022). What is the best way to identify the most suitable method for a given dataset and intention? One strategy has been to test several methods and assess the most appropriate one using metrics that measure biological conservation and batch correction efficacy (Chazarra-Gil et al., 2021; Luecken et al., 2022), which was the adopted strategy used to construct the provisional Human Lung Cell Atlas (Sikkema et al., 2022[1]). To confirm the robustness of cell populations identified after integration, it is important to map novel cell types and cell states that were identified in individual datasets back onto the tissue.

Constructing the atlas will require several levels of clustering to identify all major cell types and cell states: first, to identify the cell lineages present across all skin samples; second, to identify all cell types in a lineage; and third, to identify subtypes and cell states within each cell type. For example, the first round of clustering would identify broad immune cell populations. The second round would identify conventionally recognizable immune cell types such as macrophages, T cells, or B cells. The third round of clustering, applied only to T cells, would further discriminate between major subtypes, such as $CD4^+$, $CD8^+$, and NK T cells, and possibly characterize new T-cell states. Given the major variations in cell state plasticity across skin lineages, it is important to consider these different cell type resolutions. Some lineages may require new integrations at certain levels to better represent cell type heterogeneity. This multistep integration and clustering analysis may require different batch correction methods at different levels, as has been done in previous integration studies of skin (Reynolds et al., 2021; Solé-Boldo et al., 2020; Zou et al., 2021)

[1] Sikkema L, Strobl D, Zappia L, Madissoon E, Markov NS, Zaragosi L, et al. An integrated cell atlas of the human lung in health and disease. bioRxiv 2022.

TOWARD PROVISIONAL SKIN CELL ATLAS

Dataset selection

The major goal set by the Skin Biological Network for the provisional cell atlas, version 1.0, denoted as Atlas 1.0, is to generate a consensus nomenclature of cells in healthy human skin, differentiating cell types, their subtypes, and their states. The atlas seeks to achieve (i) high coverage of all major cell types (keratinocytes, fibroblasts, neural crest–derived cells, vessel-associated cells, adipose-associated cells, muscle-associated cells, immune cells), (ii) coverage of all major skin microstructures (pilosebaceous units, sweat glands, touch domes), and (iii) inclusion of men and women. Datasets that form the basis for Atlas 1.0 will naturally vary across other scales, such as age, body site origin, microbiota association, and ancestry. However, these variables will not be the primary focus because limited sample size will not permit their comprehensive coverage (Figure 2).

Rather than generating scRNA-seq data de novo, the Skin Biological Network will construct a provisional skin cell atlas using previously reported and publicly available high-quality datasets. To date, we have identified 40 human skin scRNA-seq datasets (Figure 2a and Supplementary Table S1), which provide a sufficient basis to generate Atlas 1.0 and achieve the set Atlas 1.0 goals with reasonable certainty. Integrating data from these studies will produce over half a million skin cells. Several candidate datasets were selected on the basis of their high coverage of certain cell types, such as melanocytes (Belote et al., 2021), fibroblasts (Solé-Boldo et al., 2020; Tabib et al., 2018; Wiedemann et al., 2023), keratinocytes (Cheng et al., 2018; Wang et al., 2019; Wiedemann et al., 2023), immune cells (Reynolds et al., 2021), and hair follicle cells (Takahashi et al., 2020), whereas all cells (including underrepresented cell types) of the individual datasets will add cumulative value to the aggregated data. In addition, integration will be performed on datasets that were generated using various cell isolation protocols and sequencing technologies, such as 10X (10X Genomics, Pleasanton, CA), Drop-seq (Macosko et al., 2015), Smart-seq2 (Picelli et al., 2013), and Smart-seq3 (Hagemann-Jensen et al., 2020), ensuring that outputs will be robust to major technical variables.

Integration strategy

Atlas 1.0 will be far from exhaustive. New scRNA-seq studies will continue to emerge, and other biological data modalities, such as epigenetic (Buenrostro et al., 2015, 2013), proteomic (Stoeckius et al., 2017), and spatial (Ståhl et al., 2016) transcriptomics, are becoming more readily available. It will become increasingly difficult to continually reintegrate and reannotate all datasets from scratch. Instead, newer computational approaches based on a reference and query strategy of integration have been proposed (Gao et al., 2021; Hao et al., 2021; Lotfollahi et al., 2022). Methods such as Azimuth (Hao et al., 2021), integrative non-negative matrix factorization (Gao et al., 2021), and scArches (Lotfollahi et al., 2022) perform integration in an iterative fashion using the latent representation of an initial reference atlas, which may be constructed using a simultaneous integration approach. The result is an updated latent representation, which can then serve as the new reference atlas. This iterative strategy is particularly effective when the query datasets represent perturbations owing to, for example, disease, and for instance, was employed to construct the Human Lung Cell Atlas (Sikkema et al., 2022[1]).

There are several benefits to a sequential integration approach over conventional simultaneous approaches. First, overall computation time is reduced because integration does not need to be repeated from scratch. Second, with an initial reference atlas, subsequent integration enables more rapid identification of novel cell states present in the query dataset (in the context of the reference data), focusing on biological investigation. Third, novel computational approaches have been developed that are now able to project data from other data modalities, such as single-cell assay for transposase-accessible chromatin sequencing (scATAC-seq) (Buenrostro et al., 2013, 2015), to the transcriptomic atlas (Cao and Gao, 2022). Such projection methods require using a trained atlas reference model and thus enable building a multimodal atlas even with only a few multimodal datasets. However, we note that eventually there will be such significant additions addressing underrepresented or missing body sites, ancestral origins, or age categories that a simultaneous integration approach will be needed to construct Atlas 2.0 and later versions.

With any integration task, certain axes of variation are inevitably prioritized when constructing the integrated latent embedding. Because other underemphasized axes contain valuable information for certain research questions and laboratories, we will ameliorate this subjectivity of integration in two ways. First, we will make the individual gene expression matrices and accompanying metadata used for integration available for local analysis through, for example, Zenodo or the CZI's CELLxGENE repository. Second, we will supplement the integrated embedding with specialized embeddings that emphasize specific axes, for example, skin sites or cell lineages.

Annotation strategy

Annotation efforts by the Skin Biological Network aim to achieve consensus between community experts. We will follow the two-step strategy employed by the Human Lung Cell Atlas (Sikkema et al., 2022[1]). First, we will curate original cell annotations from published datasets and generate a hierarchical reference framework that spans all appropriate annotation levels, from the broadest cell lineage classification (e.g., epithelial, mesenchymal, immune, vascular) to specific within-lineage cell types and possibly skin site–specific cell states. The framework will be a reference to annotate datasets containing unlabeled cells, help harmonize labels between datasets to guide the selection of an optimal integration method, and improve data integration through semisupervised learning. To properly benchmark integration and make full use of previous annotations, the original cell type annotations from published datasets must also be made consistent when constructing the reference framework. Second, we will draw upon a network of experts to generate consensus annotation and identify cell types that are prone to ambiguous or conflicting classification. This will occur in three stages: (i) providing experts with the lineage-specific integrated objects (e.g., through CELLxGENE or the Cell

Annotation Platform) to preannotate generated clusters, (ii) computationally harmonizing expert annotations and identifying disagreements, and (iii) hosting an annotation jamboree to resolve these disagreements.

To facilitate reproducibility, we will use tools such as the Cell Annotation Platform (celltype.info) to share both our cell annotation metadata and also our rationale behind each annotation choice, for example, marker gene expression or label transfer, and tools such as protocols.io (protocos.io) to share sequencing protocols.

Metadata standards

For standardized and meaningful analysis of scRNA-seq data, for example, to analyze interindividual or body site-to-site variations in cell states, it is important to annotate metadata with standard nomenclature. Current metadata accompanying published skin scRNA-seq datasets lack standardization and contain gaps (Figure 2 and Supplementary Table S1). For example, of the currently identified 235 individual skin scRNA-seq datasets sampled from the 40 studies, 52 datasets do not have information about the sampled skin site (Figure 2a). Data sampled from infants, children, and adolescents are also a notable age data gap (Figure 2b). One concerning data gap is that over half of the datasets (n = 152) are lacking ancestry information (Figure 2c), whereas almost one quarter of the datasets lacks gender information (Figure 2d). Future studies should aim to address these data gaps because they have unique skin biology and disease profiles. Furthermore, most current studies use 10X-based sequencing technology (Figure 2e). Finally, the lack of detailed and standardized annotation of anatomic sites is a significant limitation of many studies. Supplementary Table S2 details the metadata variables that the Skin Biological Network suggests to document for each individual skin dataset (as far as regional ethical and consent policies allow). Please note that metadata recommendation updates will be released at (skincommunity.org). Biological variations include age, ancestral origin, gender, time of day, sun exposure, smoking history, and systemic disease condition(s), whereas technical variations include cell enrichment protocols (if any), sequencing library chemistry, sequencing platform, and sequencing technology. To mitigate risk to privacy, strategies such as metadata aggregation will be used.

Data visualization and web portals

Intuitive and informative data visualization is necessary to convey scRNA-seq findings meaningfully. However, there are significant differences and limitations across current visualization methods: the linear PCA (Principal Component Analysis) method can be confounded by technical factors, the nonlinear t-SNE (t-distributed Stochastic Neighbor Embedding) method prioritizes local structure over global structure, whereas the nonlinear UMAP (Uniform Manifold Approximation and Projection) method can be distorted by cell composition. Moreover, these visualizations can only be interpreted qualitatively. For quantitative interpretation, direct visualizations of gene expression and cell state composition, which can take the form of box plots, violin plots, heatmaps, and general line graphs, are required. Tools such as web-based portals are important to make visualization user friendly.

We will make Atlas 1.0 accessible for easy data exploration and as a reference to project new disease states and annotations. Atlas 1.0 will be available on several web portals for exploration. Current web portals for skin include the Development Cell Atlas (developmental.cellatlas.io) and SkinGenes (skingenes.net). To facilitate new data annotation and projection onto the reference atlas, we will host Atlas 1.0 on other portals, such as Azimuth (azimuth.hubmapconsortium.org) or CellTypist (celltypist.org). For local analysis, the atlas will be uploaded to platforms such as Zenodo. All related tools and Skin Biological Network updates will be available on the skin community landing platform (skincommunity.org).

BEYOND ATLAS 1.0

Sample procurement

All datasets considered for such collections come with logistical and ethical limitations. For a more comprehensive atlas, sampling strategies should be expanded to include (i) postmortem tissues to enable a dramatic expansion of body site coverage and sampling skin sites that are otherwise challenging to procure from voluntary donors (e.g., lip skin, nail fold skin) and (ii) frozen tissues to enable collection when or where fresh tissue cannot be immediately processed (e.g., hard-to-reach geographical locations or unexpected sampling times).

Closing data gaps

Constructing Atlas 1.0 will lead to the identification of gaps that will set goals for Atlas 2.0. These may include (i) anatomically distinct body sites for which there are no scRNA-seq data, (ii) different axes of variation with missing data such as age groups or ancestral groups, (iii) missing or underrepresented cell lineages, or (iv) lack of within-donor datasets. Several members of the Skin Biological Network are generating body map datasets by sampling from multiple distinct body sites per donor.

Human skin cells have been investigated using nearly all available scRNA-seq technologies (Figure 2). Although scarce in number, there are studies that have employed single-nucleus RNA-sequencing (snRNA-seq) technology, such as the study by Satpathy et al. (2019), who performed an snRNA-seq and scATAC-seq study of cells from human basal cell carcinoma. Using snRNA-seq–based approaches is important for certain cell types such as lipid-containing adipocytes, for which droplet-based technologies are not feasible owing to the large size and high buoyancy of adipocytes. In addition, because most common scRNA-seq methods are based on short-read sequencing of the ends of RNAs only, sequencing data will be generated using technologies such as FLASH-seq (Hahaut et al., 2022) and Smart-seq3xpress (Hagemann-Jensen et al., 2022) to capture full-length transcript isoform information and enable splice variant discovery of skin-specific differences between, for example, subjects of different ancestral origins (Cechova and Miga, 2023). Data with high sequencing depth and sensitivity are also needed to better capture stem cell markers and effector transcription factors that are naturally expressed at low levels or to better represent the biology of epidermal lineages and hair follicles that undergo gradual changes in

gene expression (Cockburn et al., 2022). Furthermore, other types of sequencing, such as DNA sequencing (DNA-seq), may help provide information about somatic mutations and clonal organization in, for example, sun-exposed skin, which would be particularly relevant for same-site, different ancestral origin studies. Although current DNA-seq methods cannot be reliably used for single-cell studies, we expect newer technologies to emerge that can perform DNA-seq at the single-cell level.

Other data modalities, including scATAC-seq (Buenrostro et al., 2015a, 2013a) and CITE-seq (cellular indexing of transcriptomes and epitopes sequencing) (Stoeckius et al., 2017), can be integrated into Atlas 2.0 to better understand cell states. This multiomic approach is particularly important for lineages such as skin fibroblasts, for which inferring true cell states from scRNA-seq data alone has been challenging owing to their significant transcriptional state plasticity and their exhibiting characteristics of both progenitors (high proliferative potential) and specialized differentiated cells (high expression levels of specialized extracellular matrix genes). Although no such multiomic study of human skin currently exists in the literature, Thompson et al. (2022) profiled neonatal fibroblasts in mice using parallel scRNA-seq and scATAC-seq. The chromatin accessibility landscape revealed that despite distinct fibroblast lineages already present in neonatal mouse skin, the inferred epigenetic landscape suggested a degree of state plasticity and capacity for state transition greater than what transcriptomic data alone suggests.

Another rapidly emerging technology is spatial transcriptomics (Burgess, 2019), which can capture gene expression information while retaining spatial information in tissues, albeit, at present, at the cost of true single-cell resolution. Because skin function depends on the interplay of various cell types to maintain tissue renewal and specific functions within spatially distinct niches, integrating scRNA-seq data with spatially resolved methods is highly important. There are several technologies to generate spatial data that vary by spatial resolution (single-cell vs. multicellular spot), tissue coverage, and transcriptome coverage. These technologies can be further divided on the basis of RNA detection approach: some methods use in situ hybridization to profile native RNA species, capturing single-cell and even subcellular information at the cost of limited and biased gene coverage (several hundred selected genes at present), whereas other methods use next-generation sequencing to profile the entire transcriptome (tens of thousands of genes) at the cost of losing single-cell resolution or wide tissue coverage (Moses and Pachter, 2022). Currently, the most prominent commercially available platform is 10X Visium, which uses next-generation sequencing. Spatial information in the first-generation Visium platform is captured using microprinted barcode-type primer spots of 55 μm in diameter with a center-to-center distance of 100 μm. Visium has already been used in both mouse skin (Foster et al., 2021; Konieczny et al., 2022) and human skin (Ji et al., 2020; Ma et al., 2021b; Schäbitz et al., 2022; Shim et al., 2022) studies. Beyond spatial transcriptomics, scRNA-seq data can be integrated with previous imaging of skin, including single-cell resolution multiphoton microscopy and optical coherence tomography.

UTILITY OF THE ATLAS

A high-quality reference skin atlas will enable numerous analyses and potential insights. First, it will allow better identification of the transcriptional heterogeneity of key cell lineages, that is, what are the invariant gene markers of cell types and what is the spectrum of within-lineage cell states and their corresponding markers? Second, consensus cell type annotation can serve as a reference map for emerging newer studies, particularly those that consider perturbations (for instance, due to disease), increasing confidence in the identification of novel cell states and markers (Hao et al., 2021; Lotfollahi et al., 2022). Third, the atlas can be used to deconvolve cell type annotation to spatial transcriptomic spot data, helping to circumvent the current issue of spatial studies having reduced statistical power in analyses owing to the lower number of samples (Li et al., 2022). Fourth, with comprehensive metadata and sufficient sample sizes of each skin site and cell type, well-informed statistical models can be constructed using one of the many appropriate statistical frameworks (Soneson and Robinson, 2018; Squair et al., 2021) to perform rigorous hypothesis testing. These parametric models allow for covariate specification to reduce confounding effects in statistical analyses and enable more accurate differential expression analysis of important biological factors, such as skin site, gender, ancestral origin, or health status or even between individual donors. Other downstream applications of the atlas include characterizing active signaling pathways (Almet et al., 2021; Lewis et al., 2021); inferring potential gene regulatory networks of transcription factors and downstream targets that regulate epigenetic states (Pratapa et al., 2020), which will be further validated by scATAC-seq; and functional interpretation of the atlas using predefined gene sets (Buettner et al., 2017; Lotfollahi et al., 2022; Seninge et al., 2021; Zhao et al., 2021).

Atlas 1.0 can serve as a launching pad for many new scientific inquiries into human skin biology. Using statistical models, one could (i) investigate whether there is an association between developmental origin and microanatomy of a given skin site and its functional specialization and environmental exposure or why certain skin diseases are preferentially localized to specific body sites; (ii) examine the dependence between skin site function and cellular composition, for example, whether there are distinct mesenchymal populations that enable the elasticity and stretchability of eyelid skin or the mechanical rigidity of palmar and plantar skin; (iii) describe how stable molecular coordinates at a given body region (e.g., developmentally assigned *HOX* gene expression patterns) impact the transcriptional states of dermal cell lineages and in turn affect epithelial patterning (Chang et al., 2002; Rinn et al., 2008); (iv) analyze whether there are identifiable single-cell bases for microanatomical tissue differences across ancestries; (v) determine the effect of long-term exposure to UV with respect to changes in cell lineage populations or changes in gene expression (secretome or DNA damage response); or (vi) establish how skin microbiome composition may relate to single-cell states in host tissue.

ORCIDs
Axel A. Almet: http://orcid.org/0000-0001-9173-8278
Hao Yuan: http://orcid.org/0000-0002-4200-3600

Karl Annusver: http://orcid.org/0000-0002-9515-7216
Raul Ramos: http://orcid.org/0000-0003-0853-1028
Yingzi Liu: http://orcid.org/0000-0002-8704-4756
Julie Wiedemann: http://orcid.org/0000-0002-5293-031X
Dara H. Sorkin: http://orcid.org/0000-0003-0742-9240
Ning Xu Landén: http://orcid.org/0000-0003-4868-3798
Enikö Sonkoly: http://orcid.org/0000-0002-4909-5413
Muzlifah Haniffa: http://orcid.org/0000-0002-3927-2084
Qing Nie: http://orcid.org/0000-0002-8804-3368
Beate M. Lichtenberger: http://orcid.org/0000-0001-6882-0257
Malte D. Luecken: http://orcid.org/0000-0001-7464-7921
Bogi Andersen: http://orcid.org/0000-0001-7181-2768
Lam C. Tsoi: http://orcid.org/0000-0003-1627-5722
Fiona M. Watt: http://orcid.org/0000-0001-9151-5154
Johann E. Gudjonsson: http://orcid.org/0000-0002-0080-0812
Maksim V. Plikus: http://orcid.org/0000-0002-8845-2559
Maria Kasper: http://orcid.org/0000-0002-6117-2717

CONFLICT OF INTEREST

MDL contracted for the Chan Zuckerberg Initiative and received speaker fees from Pfizer and Janssen Pharmaceuticals. The remaining authors state no conflict of interest.

ACKNOWLEDGMENTS

MVP, BA, DHS and QN are supported by the Chan Zuckerberg Initiative grant AN-0000000062. MVP and QN are also supported by National Institutes of Health grants U01-AR073159 and R01-AR079150. MVP is also supported by LEO Foundation grants LF-AW-RAM-19-400008 and LF-OC-20-000611, W.M. Keck Foundation grant WMKF-5634988, National Science Foundation grant DMS1951144, and National Institutes of Health grants R01-AR079470 and R21-AR078939. BA is also supported by the National Institutes of Health grant R01-AR044882. MK is supported by LEO Foundation grant LF-OC-19-000225, Swedish Research Council grant 2022-01059, Swedish Cancer Society grant 21 1821 Pj, and Karolinska Institutet Consolidator grant 2-2111/2019. NXL is supported by LEO Foundation grant LF-OC-22-001035. MDL is supported by the Human Cell Atlas Data Ecosystem grant CZIF2022-007488 from Chan Zuckerberg Initiative. BML is supported by Austrian Science Fund P 35307-B and the LEO Foundation award 2021 LF-AW_EMEA[1]21-400116. MH is supported by Wellcome Trust. FMW is supported by Wellcome Trust grant 211276/E/18/Z. YL is supported by a clinical fellowship from California Institute for Regenerative Medicine training grant EDUC4-12822. KA is supported by a Karolinska Institutet KID grant. HY is supported by European Union's Horizon 2020 Research and Innovation Programme, Marie Skłodowska-Curie (Cancerprev, 859860). Additional support comes from National Institutes of Health grants P30-AR075047 and P30-AR075043, National Science Foundation grant DMS1763272 and Simons Foundation grant (594598, QN), and UC Irvine Institute for Clinical and Translational Science grant UL1 TR0001414. MVP is the corresponding author gaurentor for this article.

AUTHOR CONTRIBUTIONS

Conceptualization: MVP, MK; Writing - Original Draft Preparation: AAA, HY, KA, RR, YL, DHS, MVP, MK; Writing - Review and Editing: AAA, HY, KA, RR, YL, JW, NXL, ES, MH, BML, MDL, QN, BA, LCT, FMW, JEG, MVP, MK

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.2023.03.1679.

REFERENCES

Alexander CM, Kasza I, Yen CLE, Reeder SB, Hernando D, Gallo RL, et al. Dermal white adipose tissue: a new component of the thermogenic response. J Lipid Res 2015;56:2061–9.

Almet AA, Cang Z, Jin S, Nie Q. The landscape of cell–cell communication through single-cell transcriptomics. Curr Opin Syst Biol 2021;26:12–23.

Andersen LK, Davis MDP. Sex differences in the incidence of skin and skin-related diseases in Olmsted County, Minnesota, United States, and a comparison with other rates published worldwide. Int J Dermatol 2016;55:939–55.

Belote RL, Le D, Maynard A, Lang UE, Sinclair A, Lohman BK, et al. Human melanocyte development and melanoma dedifferentiation at single-cell resolution. Nat Cell Biol 2021;23:1035–47.

Borthwick J, Evertsz N, Pratt B. How should communities be meaningfully engaged (if at all) when setting priorities for biomedical research? Perspectives from the biomedical research community. BMC Med Ethics 2023;24:6.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 2013;10:1213–8.

Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. Nature 2015;523:486–90.

Buettner F, Pratanwanich N, McCarthy DJ, Marioni JC, Stegle O. f-scLVM: scalable and versatile factor analysis for single-cell RNA-seq. Genome Biol 2017;18:212.

Burgess DJ. Spatial transcriptomics coming of age. Nat Rev Genet 2019;20:317.

Byrd AL, Belkaid Y, Segre JA. The human skin microbiome. Nat Rev Microbiol 2018;16:143–55.

Cao ZJ, Gao G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. Nat Biotechnol 2022;40:1458–66.

Cechova M, Miga KH. Comprehensive variant discovery in the era of complete human reference genomes. Nat Methods 2023;20:17–9.

Chang HY, Chi JT, Dudoit S, Bondre C, Van De Rijn M, Botstein D, et al. Diversity, topographic differentiation, and positional memory in human fibroblasts. Proc Natl Acad Sci USA 2002;99:12877–82.

Chazarra-Gil R, Van Dongen S, Kiselev VY, Hemberg M. Flexible comparison of batch correction methods for single-cell RNA-seq using BatchBench. Nucleic Acids Res 2021;49:e42.

Cheng JB, Sedgewick AJ, Finnegan AI, Harirchian P, Lee J, Kwon S, et al. Transcriptional programming of normal and inflamed human epidermis at single-cell resolution. Cell Rep 2018;25:871–83.

Chike-Obi CJ, Cole PD, Brissett AE. Keloids: pathogenesis, clinical features, and management. Semin Plast Surg 2009;23:178–84.

Cockburn K, Annusver K, Gonzalez D, Ganesan S, May D, Kawaguchi K, et al. Gradual differentiation uncoupled from cell cycle exit generates heterogeneity in the epidermal stem cell layer. Nat Cell Biol 2022;24:1692.

Dao H, Kazin RA. Gender differences in skin: a review of the literature. Gend Med 2007;4:308–28.

Day S, Mathews A, Blumberg M, Vu T, Rennie S, Tucker JD. Broadening community engagement in clinical research: designing and assessing a pilot crowdsourcing project to obtain community feedback on an HIV clinical trial. Clin Trials 2020;17:306–13.

Dhabale A, Nagpure S. Types of psoriasis and their effects on the immune system. Cureus 2022;14:e29536.

Donati G, Rognoni E, Hiratsuka T, Liakath-Ali K, Hoste E, Kar G, et al. Wounding induces dedifferentiation of epidermal Gata6+ cells and acquisition of stem cell properties. Nat Cell Biol 2017;19:603–13.

Farage MA, Miller KW, Elsner P, Maibach HI. Characteristics of the aging skin. Adv Wound Care (New Rochelle) 2013;2:5–10.

Feramisco JD, Sadreyev RI, Murray ML, Grishin NV, Tsao H. Phenotypic and genotypic analyses of genetic skin disease through the Online Mendelian Inheritance in Man (OMIM) database. J Invest Dermatol 2009;129:2628–36.

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.

Gao C, Liu J, Kriebel AR, Preissl S, Luo C, Castanon R, et al. Iterative single-cell multi-omic integration using online learning. Nat Biotechnol 2021;39:1000–7.

Gurtner GC, Werner S, Barrandon Y, Longaker MT. Wound repair and regeneration. Nature 2008;453:314–21.

Hagemann-Jensen M, Ziegenhain C, Chen P, Ramsköld D, Hendriks GJ, Larsson AJM, et al. Single-cell RNA counting at allele and isoform resolution using Smart-seq3. Nat Biotechnol 2020;38:708–14.

Hagemann-Jensen M, Ziegenhain C, Sandberg R. Scalable single-cell RNA sequencing from full transcripts with Smart-seq3xpress. Nat Biotechnol 2022;40:1452–7.

Haghverdi L, Lun ATL, Morgan MD, Marioni JC. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat Biotechnol 2018;36:421–7.

Hahaut V, Pavlinic D, Carbone W, Schuierer S, Balmer P, Quinodoz M, et al. Fast and highly sensitive full-length single-cell RNA sequencing using FLASH-seq. Nat Biotechnol 2022;40:1447–51.

Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell 2021;184:3573–87.e29.

Haydont V, Bernard BA, Fortunel NO. Age-related evolutions of the dermis: clinical signs, fibroblast and extracellular matrix dynamics. Mech Ageing Dev 2019;177:150–6.

Hie B, Bryson B, Berger B. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. Nat Biotechnol 2019;37:685–91.

Hirano SA, Murray SB, Harvey VM. Reporting, representation, and subgroup analysis of race and ethnicity in published clinical trials of atopic dermatitis in the United States between 2000 and 2009. Pediatr Dermatol 2012;29:749–55.

Holzer JK, Ellis L, Merritt MW. Why we need community engagement in medical research. J Investig Med 2014;62:851–5.

Ji AL, Rubin AJ, Thrane K, Jiang S, Reynolds DL, Meyers RM, et al. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma [published correction appears in Cell 2020;182:1661–2] Cell 2020;182:497–514.e22.

Kilgour JM, Li S, Sarin KY. Hidradenitis suppurativa in patients of color is associated with increased disease severity and healthcare utilization: a retrospective analysis of 2 U.S. cohorts [published correction appears in JAAD Int 2021;3:88] JAAD Int 2021;3:42–52.

Konieczny P, Xing Y, Sidhu I, Subudhi I, Mansfield KP, Hsieh B, et al. Interleukin-17 governs hypoxic adaptation of injured epithelium. Science 2022;377:eabg9302.

Lamartine J. Towards a new classification of ectodermal dysplasias. Clin Exp Dermatol 2003;28:351–5.

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

Li B, Zhang W, Guo C, Xu H, Li L, Fang M, et al. Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. Nat Methods 2022;19:662–70.

Lotfollahi M, Naghipourfar M, Luecken MD, Khajavi M, Büttner M, Wagenstetter M, et al. Mapping single-cell data to reference atlases by transfer learning. Nat Biotechnol 2022;40:121–30.

Lotfollahi M, Naghipourfar M, Theis FJ, Wolf FA. Conditional out-of-distribution generation for unpaired data using transfer VAE. Bioinformatics 2020;36:i610–7.

Lotfollahi M, Wolf FA, Theis FJ. scGen predicts single-cell perturbation responses. Nat Methods 2019;16:715–21.

Luecken MD, Büttner M, Chaichoompu K, Danese A, Interlandi M, Mueller MF, et al. Benchmarking atlas-level data integration in single-cell genomics. Nat Methods 2022;19:41–50.

Luther N, Darvin ME, Sterry W, Lademann J, Patzelt A. Ethnic differences in skin physiology, hair follicle morphology and follicular penetration. Skin Pharmacol Physiol 2012;25:182–91.

Ma F, Hughes TK, Teles RMB, Andrade PR, de Andrade Silva BJ, Plazyo O, et al. The cellular architecture of the antimicrobial response network in human leprosy granulomas. Nat Immunol 2021b;22:839–50.

Ma MA, Gutiérrez DE, Frausto JM, Al-Delaimy WK. Minority representation in clinical trials in the United States: trends over the past 25 years. Mayo Clin Proc 2021a;96:264–6.

Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 2015;161:1202–14.

Mapes BM, Foster CS, Kusnoor SV, Epelbaum MI, AuYoung M, Jenkins G, et al. Diversity and inclusion for the All of Us research program: a scoping review. PLoS One 2020;15:e0234962.

Matsumura Y, Ananthaswamy HN. Toxic effects of ultraviolet radiation on the skin. Toxicol Appl Pharmacol 2004;195:298–308.

Moses L, Pachter L. Museum of spatial transcriptomics. Nat Methods 2022;19:534–46.

Nicholson LM, Schwirian PM, Groner JA. Recruitment and retention strategies in clinical studies with low-income and minority populations: progress from 2004–2014. Contemp Clin Trials 2015;45:34–40.

Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods 2013;10:1096–8.

Polański K, Young MD, Miao Z, Meyer KB, Teichmann SA, Park JE. BBKNN: fast batch alignment of single cell transcriptomes. Bioinformatics 2020;36:964–5.

Pratapa A, Jalihal AP, Law JN, Bharadwaj A, Murali TM. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. Nat Methods 2020;17:147–54.

Regev A, Teichmann SA, Lander ES. The human cell atlas. ELife 2017;6:e27041.

Reynolds G, Vegh P, Fletcher J, Poyner EFM, Stephenson E, Goh I, et al. Developmental cell programs are co-opted in inflammatory skin disease. Science 2021;371:eaba6500.

Rinn JL, Wang JK, Allen N, Brugmann SA, Mikels AJ, Liu H, et al. A dermal HOX transcriptional program regulates site-specific epidermal fate. Genes Dev 2008;22:303–7.

Sandby-Møller J, Poulsen T, Wulf HC. Epidermal thickness at different body sites: relationship to age, gender, pigmentation, blood content, skin type and smoking habits. Acta Derm Venereol 2003;83:410–3.

Satpathy AT, Granja JM, Yost KE, Qi Y, Meschi F, McDermott GP, et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. Nat Biotechnol 2019;37:925–36.

Schäbitz A, Hillig C, Mubarak M, Jargosch M, Farnoud A, Scala E, et al. Spatial transcriptomics landscape of lesions from non-communicable inflammatory skin diseases. Nat Commun 2022;13:7729.

Seninge L, Anastopoulos I, Ding H, Stuart J. Vega is an interpretable generative model for inferring biological network activity in single-cell transcriptomics. Nat Commun 2021;12:5684.

Shim J, Oh SJ, Yeo E, Park JH, Bae JH, Kim SH, et al. Integrated analysis of single-cell and spatial transcriptomics in keloids: highlights on fibrovascular interactions in keloid pathogenesis. J Invest Dermatol 2022;142:2128–39.e11.

Solé-Boldo L, Raddatz G, Schütz 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.

Soneson C, Robinson MD. Bias, robustness and scalability in single-cell differential expression analysis. Nat Methods 2018;15:255–61.

Squair JW, Gautier M, Kathe C, Anderson MA, James ND, Hutson TH, et al. Confronting false discoveries in single-cell differential expression. Nat Commun 2021;12:5692.

Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science 2016;353:78–82.

Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, et al. Simultaneous epitope and transcriptome measurement in single cells. Nat Methods 2017;14:865–8.

Stone K, Wheeler A. A review of anatomy, physiology, and benign pathology of the nipple. Ann Surg Oncol 2015;22:3236–40.

Sun X, Joost S, Kasper M. Plasticity of epithelial cells during skin wound healing [epub ahead of print]. Cold Spring Harb Perspect Biol 2022. https://doi.org/10.1101/cshperspect.a041232 (accessed March 17, 2023).

Swartz TH, Palermo AS, Masur SK, Aberg JA. The science and value of diversity: closing the gaps in our understanding of inclusion and diversity. J Infect Dis 2019;220:S33–41.

Tabib T, Morse C, Wang T, Chen W, Lafyatis R. SFRP2/DPP4 and FMO1/LSP1 define major fibroblast populations in human skin. J Invest Dermatol 2018;138:802–10.

Takahashi R, Grzenda A, Allison TF, Rawnsley J, Balin SJ, Sabri S, et al. Defining transcriptional signatures of human hair follicle cell states. J Invest Dermatol 2020;140:764–73.e4.

Takeo M, Lee W, Ito M. Wound healing and skin regeneration. Cold Spring Harb Perspect Med 2015;5:a023267.

Thi H, Tran N, Ang KS, Chevrier M, Zhang X, Yee N, et al. A benchmark of batch-effect correction methods for single-cell RNA sequencing data. Genome Biol 2020;21:12.

Thompson SM, Phan QM, Winuthayanon S, Driskell IM, Driskell RR. Parallel single-cell multiomics analysis of neonatal skin reveals the transitional fibroblast states that restrict differentiation into distinct fates. J Invest Dermatol 2022;142:1812–23.e3.

Thulabandu V, Chen D, Atit RP. Dermal fibroblast in cutaneous development and healing. Wiley Interdiscip Rev Dev Biol 2018;7:1–13.

Tsai J, Rostom M, Garza LA. Understanding and harnessing epithelial–mesenchymal interactions in the development of palmoplantar identity. J Invest Dermatol 2022;142:282–4.

Tuan TL, Nichter LS. The molecular basis of keloid and hypertrophic scar formation. Mol Med Today 1998;4:19–24.

Vogt A, Hadam S, Heiderhoff M, Audring H, Lademann J, Sterry W, et al. Morphometry of human terminal and vellus hair follicles. Exp Dermatol 2007;16:946–50.

Wang S, Drummond M, Guerrero-Juarez C, Tarapore E, MacLean A, Stabell A, et al. Single cell transcriptomics of human epidermis reveals basal stem cell transition states. Nat Commun 2019;11:4239.

Watt FM. Mammalian skin cell biology: at the interface between laboratory and clinic. Science 2014;346:937–40.

Wiedemann J, Billi AC, Bocci F, Kashgari G, Xing E, Tsoi LC, et al. Differential cell composition and split epidermal differentiation in human palm, sole, and hip skin. Cell Rep 2023;42:111994.

Williams HC, Dellavalle RP, Garner S. Acne vulgaris. Lancet 2012;379:361–72.

Xu C, Lopez R, Mehlman E, Regier J, Jordan MI, Yosef N. Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. Mol Syst Biol 2021;17:e9620.

Zhao Y, Cai H, Zhang Z, Tang J, Li Y. Learning interpretable cellular and gene signature embeddings from single-cell transcriptomic data. Nat Commun 2021;12:5261.

Zou Z, Long X, Zhao Q, Zheng Y, Song M, Ma S, et al. A single-cell transcriptomic atlas of human skin aging. Dev Cell 2021;56:383–97.e8.