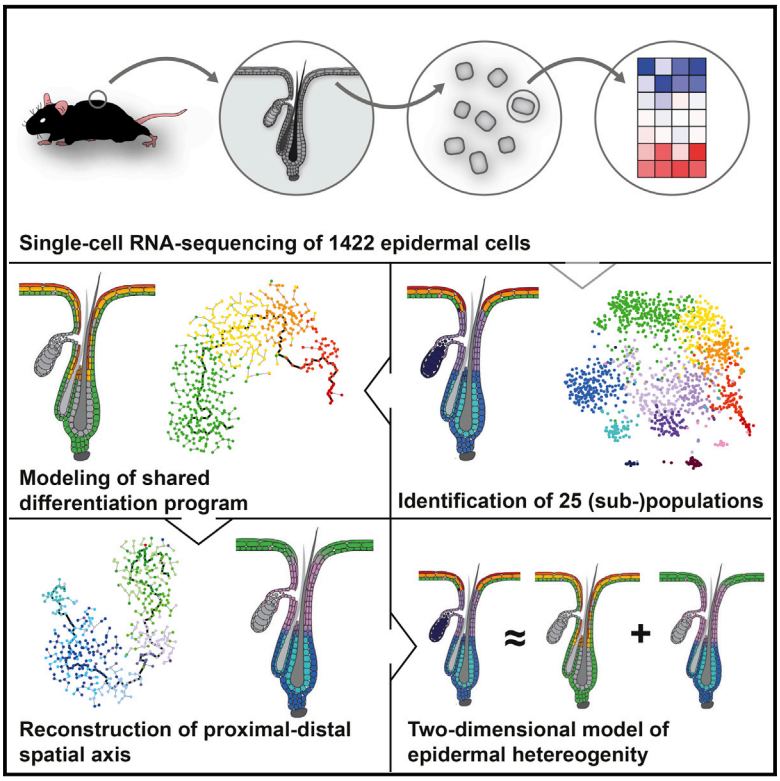
Article

Cell Systems

Single-Cell Transcriptomics Reveals that Differentiation and Spatial Signatures Shape Epidermal and Hair Follicle Heterogeneity

Graphical Abstract



Authors

Simon Joost, Amit Zeisel, Tina Jacob, ..., Peter Lönnerberg, Sten Linnarsson, Maria Kasper

Correspondence

sten.linnarsson@ki.se (S.L.), maria.kasper@ki.se (M.K.)

In Brief

Joost et al. use high-throughput single-cell RNA-seq to describe gene expression in mouse epidermis and hair follicles at unprecedented detail and explain epidermal heterogeneity as the interplay of differentiation-related and spatial gene expression signatures.

Highlights

- Single-cell RNA-seq analysis identifies 25 populations of epidermal cells
- Differentiation and spatial gene expression signatures can be defined
- Interplay of differentiation and spatial signatures explains most heterogeneity
- Stem cell populations are divided by spatial signatures and only share basal identity

Data Resources

GSE67602

Joost et al., 2016, Cell Systems 3, 221–237
September 28, 2016
http://dx.doi.org/10.1016/j.cels.2016.08.010

CellPress

Cell Systems
Article

Single-Cell Transcriptomics Reveals that Differentiation and Spatial Signatures Shape Epidermal and Hair Follicle Heterogeneity

Simon Joost,[1] Amit Zeisel,[2] Tina Jacob,[1] Xiaoyan Sun,[1] Gioele La Manno,[2] Peter Lönnerberg,[2] Sten Linnarsson,[2,*] and Maria Kasper[1,3,*]
[1]Department of Biosciences and Nutrition and Center for Innovative Medicine, Karolinska Institutet, Novum, 141 83 Huddinge, Sweden
[2]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Scheeles väg 2, 171 77 Stockholm, Sweden
[3]Lead Contact
*Correspondence: sten.linnarsson@ki.se (S.L.), maria.kasper@ki.se (M.K.)
http://dx.doi.org/10.1016/j.cels.2016.08.010

SUMMARY

The murine epidermis with its hair follicles represents an invaluable model system for tissue regeneration and stem cell research. Here we used single-cell RNA-sequencing to reveal how cellular heterogeneity of murine telogen epidermis is tuned at the transcriptional level. Unbiased clustering of 1,422 single-cell transcriptomes revealed 25 distinct populations of interfollicular and follicular epidermal cells. Our data allowed the reconstruction of gene expression programs during epidermal differentiation and along the proximal-distal axis of the hair follicle at unprecedented resolution. Moreover, transcriptional heterogeneity of the epidermis can essentially be explained along these two axes, and we show that heterogeneity in stem cell compartments generally reflects this model: stem cell populations are segregated by spatial signatures but share a common basal-epidermal gene module. This study provides an unbiased and systematic view of transcriptional organization of adult epidermis and highlights how cellular heterogeneity can be orchestrated in vivo to assure tissue homeostasis.

INTRODUCTION

The epidermis and its appendages form the outer layer of the mammalian skin and shield the body from external harm (Fuchs, 2007). Its regenerative capacity along with its accessibility and compartmentalized microanatomy has made the epidermis one of the most important model systems for stem cell biology (Hsu et al., 2014; Schepeler et al., 2014), and many paradigms of tissue maintenance and regeneration have been established or validated in the murine epidermis (Rompolas and Greco, 2014).

In mice, the epidermis consists of two main compartments with distinct physiological functions: the interfollicular epidermis (IFE), and the hair follicle (HF) including the sebaceous gland (SG) (Niemann and Watt, 2002). Cells of the IFE constitute the majority of epidermal cells and form a squamous, stratified, multilayered epithelium that plays the key role in securing the skin barrier function (Fuchs, 1990). In contrast, the main role of HFs lies in producing the hair shaft to maintain the murine fur. While the cells of IFE and SG are constantly replaced, the HF is subjected to cycles of rest (telogen), growth (anagen), and degeneration (catagen). The telogen HF exhibits a characteristic microanatomy including the bulge and hair germ fuelling hair growth, the isthmus and junctional zone encompassing the opening of the SG, and the infundibulum connecting the HF to the IFE (Figure 1B). The lower part of the HF closest to the hair-growth inductive dermal papilla is often referred to as the proximal part, and consequently the upper HF as distal (Müller-Röver et al., 2001).

The cellular composition of the epidermis has been extensively studied during the last decades. It has been shown that the keratinocytes of the IFE can be morphologically, molecularly, and functionally divided into basal cells, suprabasal spinous, and granular layer cells, which each play distinct roles in producing and maintaining the skin barrier (Fuchs, 1990). In a similar fashion, it has been established how SG cells differentiate to fulfill glandular functions or how HF keratinocytes maintain the hair shaft (Niemann and Horsley, 2012). More recently, reporter constructs and lineage tracing studies have characterized stem cell and progenitor populations in the IFE, the SG, and sub-compartments of the HF (Alcolea and Jones, 2014; Kretzschmar and Watt, 2014; Petersson and Niemann, 2012). The molecular relationship between the different stem and progenitor populations and "non-stem cell" populations is, however, still insufficiently addressed.

A large number of studies have investigated the transcriptomes of cell populations in the human and murine epidermis in vivo and in vitro. While a few pioneering studies were performed at single-cell resolution but were limited by low sensitivity or small numbers of analyzed genes (Jensen and Watt, 2006; Tan et al., 2013), most of the studies relied on bulk-sampling techniques and cell enrichment using pre-defined markers (Blanpain et al., 2004; Brownell et al., 2011; Füllgrabe et al., 2015; Greco et al., 2009; Jaks et al., 2008; Janich et al., 2011; Mascré et al., 2012; Page et al., 2013; Snippert et al., 2010; Tumbar et al., 2004). As nearly all of these studies were restricted to certain subpopulations or compartments of the epidermis, it has been difficult to directly compare results

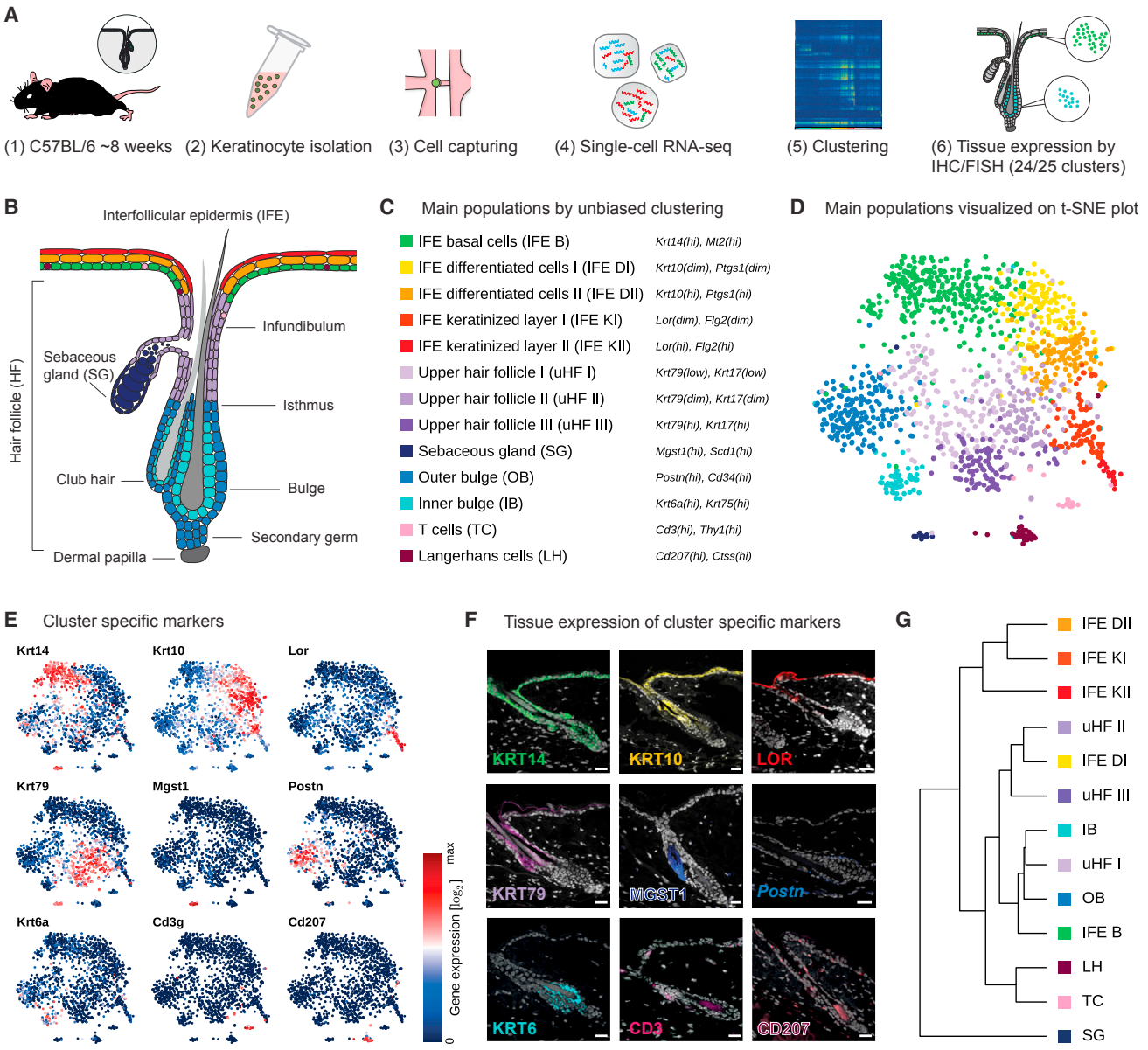


Figure 1. Defining the Main Epidermal Cell Populations

(A) Overview of the experimental workflow.

(B) Illustrated microanatomy and compartmentalization of the murine epidermis including HF and SG, colored according to main populations (C).

(C) Identity and marker genes of cell populations defined during first-level clustering.

(D) Epidermal cell transcriptomes (n = 1,422) visualized with t-distributed stochastic neighbor embedding (t-SNE), colored according to unsupervised (first level) clustering (C).

(E) Expression of group-specific marker genes projected onto the t-SNE map.

(F) Immunostaining or single-molecule FISH for group-specific genes. Protein or mRNA (symbols italics) expression is pseudocolored corresponding to groups shown in (C). Cell nuclei are shown in white. Scale bars, 20 μm. See also Figure S2J.

(G) Hierarchical clustering (Ward's linkage) of gene expression data averaged over each group.

across studies and to analyze epidermal heterogeneity in a systematic fashion. In contrast, recent advances in single-cell RNA-sequencing (RNA-seq) technologies have made it possible to profile large numbers of cells in parallel (Hashimshony et al., 2012; Islam et al., 2014; Picelli et al., 2013) in order to comprehensively dissect the cellular composition of complex tissues (Sandberg, 2014). In addition to unveiling novel epidermal cell populations, high-throughput single-cell transcriptomics of the epidermis may also reveal heterogeneity within previously described populations in the murine skin (Jaks et al., 2010; Kretzschmar and Watt, 2014). However, such studies are lacking so far.

Here, we used quantitative single-cell RNA-seq to sequence 1,422 cells from the murine telogen epidermis to systematically dissect the cellular heterogeneity of epidermal cells during tissue homeostasis. We provide a high-resolution transcriptome map that is available online, present potential novel transcriptional regulators along the differentiation and spatial axes, and model the impact of each axis on transcriptional heterogeneity.

RESULTS

Single-Cell Transcriptome Analysis of Mouse Epidermis

To study the transcriptional heterogeneity of the telogen epidermis, we isolated epidermal cells from dorsal skin of C57BL/6 wild-type mice during second telogen at around 8 weeks (Figures 1A, S1A, and S1B). The isolated cells of individual mice (n = 19 biological replicates) were, after one HF cell enrichment step, directly loaded into 96-well microfluidic C1 chips (Fluidigm) and randomly captured for sequencing. Because we expected higher cellular heterogeneity within HFs compared to IFE (Figure 1B), we used SCA-1 microbeads to enrich for HF cells and sampled HF (SCA-1^{-}) and IFE/infundibulum (SCA-1^{+}) cell numbers in a 2:1 ratio (Figures S1C–S1E). Although single-cell capturing in C1 chips showed a minor bias for larger cells, the whole size range of both cell fractions was represented in the dataset (Figure S1F). Through imaging of the C1 chips, chambers containing more than one cell were excluded. Next, we prepared and sequenced single-cell cDNA libraries using a quantitative single-cell RNA-seq protocol (Islam et al., 2014). Sequencing yield and quality was comparable to our previous studies (Figures S1G–S1N) (Zeisel et al., 2015). Single cells with <2,000 unique detected molecules failed to reach quality-control standards and were excluded, leaving 1,422 single-cell transcriptomes in the final dataset (Figure S1K).

Unbiased Clustering Confirms Known Epidermal Cell Populations

First, we dissected the global structure of the dataset through unsupervised clustering with affinity propagation (Frey and Dueck, 2007) based on the expression of high variance genes (Figure S2A). Importantly, all clusters (representing distinct groups of cells) were derived without considering a priori knowledge from the literature. We robustly identified 13 highly distinct main groups of epidermal cells, which we visualized in two-dimensional space using t-distributed stochastic neighbor embedding (t-SNE) (Van der Maaten and Hinton, 2008) (Figures 1C, 1D, and S2B–S2F): SG cells marked by *Scd1*/*Mgst1*, inner and outer bulge keratinocytes characterized by expression of *Krt6a*/*Krt75* and *Cd34*/*Postn*, respectively, predominantly IFE-derived basal cells with high expression levels of *Krt14*/*Mt2*, two stages of differentiated cells marked by *Krt10*/*Ptgs1* and two stages of terminally differentiated keratinized layer cells expressing *Lor*/*Flg2*, three distinct groups of upper HF cells marked by different levels of *Krt79*/*Krt17*, and two immune cell populations Langerhans cells (*Cd207*$^{+}$/*Ctss*$^{+}$) and resident T cells (*Cd3*$^{+}$/*Thy1*$^{+}$). We subsequently used a negative binomial Bayesian regression model to identify group-specific gene expression signatures, and, as expected, each group of cells expressed a distinct set of genes (Figures 1C, 1E, and S2G–S2I; Table S1).

To confirm the existence of these cell populations with a sequencing-independent method, we selected known and newly derived marker genes and subsequently stained telogen skin tissue sections using immunohistochemistry (IHC) and/or single-molecule mRNA fluorescence in situ hybridization (FISH) (STAR Methods). This also allowed us to map the defined populations to their spatial location in the telogen epidermis (Figures 1F and S2J). Interestingly, comparing transcriptional similarity among the 13 epidermal groups revealed that the cell populations did not always cluster based on their physical location, raising the question whether similar cellular functions render cells more similar than location (Figure 1G). Overall, even though the first round (first level) of clustering did not reveal novel populations of cells, an outcome that is not unexpected given that the murine epidermis is one of the best studied mammalian organ systems (Fuchs, 2007; Niemann and Watt, 2002; Schepeler et al., 2014), it robustly recapitulated the expected main epidermal structures and cell populations.

Subclustering of Main Populations Reveals New Subpopulations

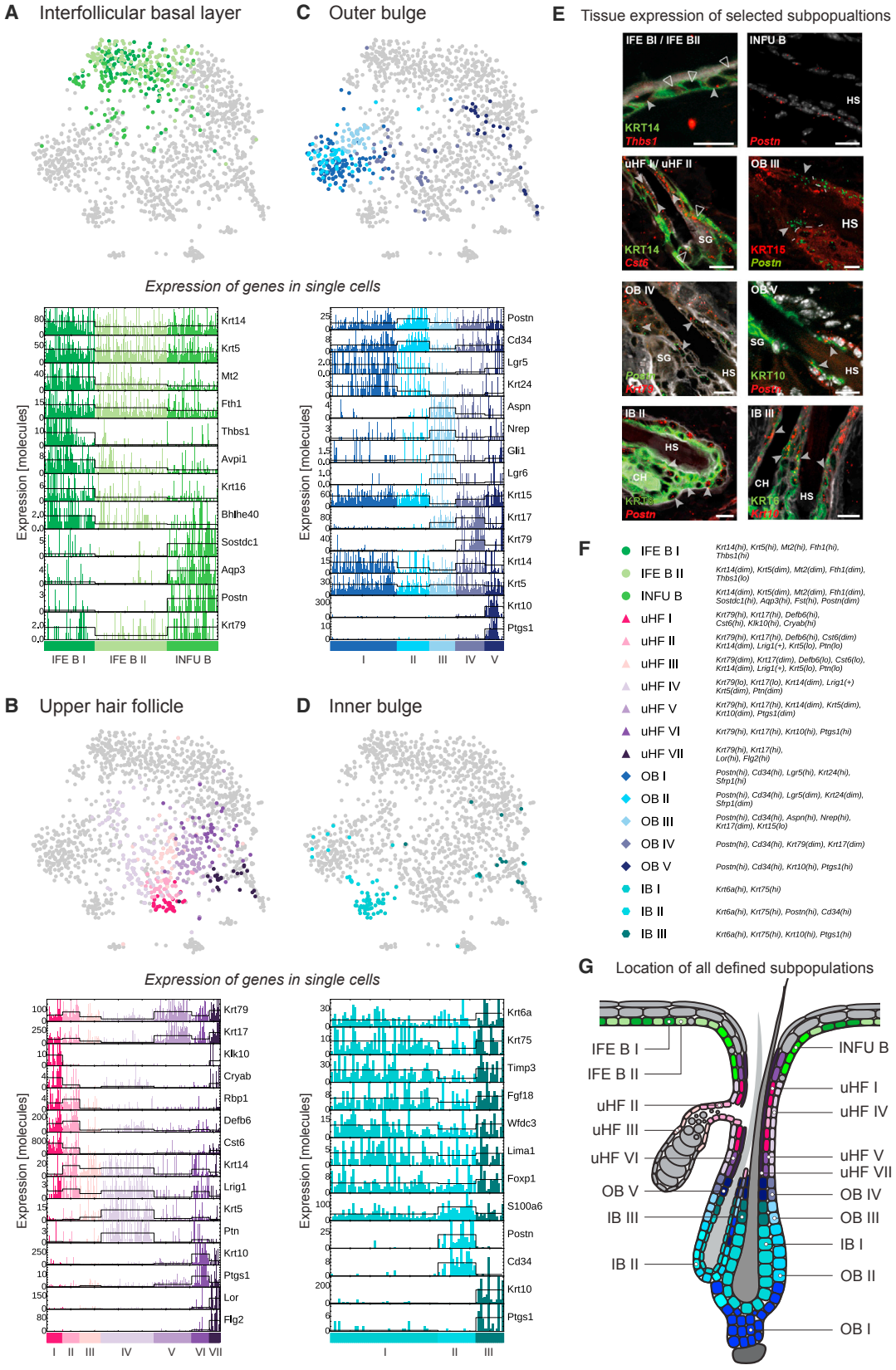
To further resolve cellular heterogeneity of HF and IFE cells, we selected all cells that were in the first-level clustering defined as having an outer bulge, inner bulge, upper HF, and basal IFE signature, respectively, and subjected them to a second round (second level) of unsupervised clustering (Figures S3A and S3B). We divided the upper HF into seven, the outer bulge into five, and the inner bulge as well as the basal IFE into three subpopulations, respectively (Figures 2A–2G and S3C–S3L; Table S2). To exclude that any population was merely the result of biological (e.g., variability between mice) or technical artifacts (e.g., variability in cell isolation, or cell doublets [Macosko et al., 2015]), we used three different validation strategies (STAR Methods): (1) verification that each cluster was formed by an adequate number of biological replicates, (2) resampling approach to test robustness of each cell cluster, (3) systematic staining of all populations by IHC and/or FISH. The results show that cells of at least eight different mice formed each cluster, the majority of clusters were highly robust (Figures S3G–S3J), and all populations could be identified by IHC and/or FISH staining.

Upper HF

The cells of the upper HF could be separated into four known (uHF IV–VII), one indistinct (uHF III), and two new cell populations (uHF I and uHF II) (Figures 2B, 2E, 2G, S3D, and S3L). The new populations were located around the SG opening and could be distinguished by *Rbp1* expression as well as high levels of *Defb6* and *Cst6*. While uHF I cells showed additional expression of unique markers such as *Klk10* and could be located to two suprabasal rings of cells around the SG opening, uHF II cells expressed a small subset of typical basal genes such as *Krt14* (but not *Krt5*) and could be linked to the SG duct. The other subpopulations of uHF cells (uHF IV–VII) showed a typical uHF signature (high levels of *Krt17*, *Krt79*, *Cd44*, *Cd200*, and *Lrig1* in the more basal cells) combined with expression of gene signatures linked to the basal (*Krt5*, *Krt14*), suprabasal (*Krt10*, *Ptgs1*), and keratinized layer (*Flg2*, *Lor*) of the IFE.

Outer Bulge

The outer bulge is the most well-investigated HF compartment and is characterized by high expression of *Cd34*, *Krt15*, and



(legend on next page)

Lgr5 (Blanpain et al., 2004; Cotsarelis et al., 1990; Jaks et al., 2008; Morris et al., 2004). The degree of transcriptional heterogeneity within the outer bulge cells is, however, only partly explored (Blanpain et al., 2004; Janich et al., 2011; Tumbar et al., 2004). Subclustering cells with outer bulge signature revealed five subpopulations (Figures 2C and S3E). Most of the cells of the outer bulge belonged to either a *Cd34*^{hi}, *Postn*^{hi}, *Lgr5*^{hi}, *Krt24*^{hi} population (OB I) located in the proximal part of the outer bulge and the hair germ or a *Cd34*^{hi}, *Postn*^{hi}, *Lgr5*^{dim}, *Krt24*^{dim} population (OBII) that was mapped to the central part of the outer bulge (Figures 2G and S3L). The three additional OB-cell populations (OB III, IV, and V) were demarcated at the distal end of the bulge area and at the lower isthmus (Figures 2E, 2G, and S3L). OB III was characterized by a unique signature of genes including *Aspn*, *Nrep*, and *Robo2* (Figures 2C and S3E), and, interestingly, this population also showed the strongest expression of *Gli1* and *Lgr6* in the HF indicating that this cluster includes cells from both the Gli1^{+} population defined by Brownell et al. and the Lgr6^{+} population described by Snippert et al. (Brownell et al., 2011; Snippert et al., 2010). In contrast to OB III, the cells of OB IV located distal to OB III did not express unique genes; instead, they were marked by an overlapping outer bulge (including *Postn* and *Cd34*) and upper HF signature (including *Krt79*, *Krt17*, *Lrig1*, and *Cd44*) (Figure 2E). OB V is a population of suprabasal cells, which expressed both an outer bulge signature and differentiation markers such as *Krt10* and *Ptgs1* (Figure 2E).

Inner Bulge

The majority of inner bulge cells belonged to a population (IB I) solely expressing the typical inner bulge signature (e.g., high levels of *Krt6a*, *Krt75*, *Timp3*, *Fgf18*). The second population (IB II) consisted of cells expressing both inner bulge and outer bulge markers and could be mapped to the outer bulge (Figure 2E). The third population (IB III) co-expressed an inner bulge and a differentiation signature (e.g., *Krt10*, *Ptgs1*) and was mapped to the distal end of the inner bulge compartment (Figure 2E).

Overall, we were able to resolve 16 distinct subpopulations of HF cells, of which many have not been previously described (Table S3). Intriguingly, only three of those subpopulations—the Gli1^{+} upper bulge population (OB III) and the upper HF populations located around the SG (uHF I and uHF II)—were defined by unique genetic signatures. In contrast, most heterogeneity in the HF seemed to result from the combination of recurring genetic signatures (Figures 2A–2D, S3C–S3F, and S3K; Table S2), suggesting that the vast complexity of cellular identities found in the HF might be the consequence of the coordinated interplay of just a few classes of genetic signatures. As a consequence, dividing lines (i.e., borders) between some populations (Figure S5E) became less distinct, exemplified by the overlap of genetic signatures in OB IV (upper HF and outer bulge signatures) and IB II (inner bulge and outer bulge signatures). Importantly, these observations were not limited to cells of the HF.

Basal IFE

While subclustering IFE basal cells, we found a subpopulation that expressed low levels of upper HF markers such as *Krt79*, the bulge marker *Postn*, and pan-HF markers like *Sostdc1*, *Aqp3*, and *Fst* in addition to the IFE basal signature (Figures 2A, 2E, and S3C). This unique combination of signatures turned out to mark basal cells of the infundibulum, the structure that connects the HF to the IFE, which was never transcriptionally resolved before. Moreover, we found two distinct basal IFE populations (IFE BI and II; Figure 2E) both expressing high levels of *Krt14* and *Krt5*, and IFE BI additionally expressed high levels of *Avpi1*, *Krt16*, *Thbs1*, and the transcription factor *Bhlhe40*. Interestingly, Thrombospondin 1 (THBS1) was reported to inhibit angiogenesis and to modulate cell adhesion, motility, and growth (Guo et al., 1997), and BHLHE40 has been suggested to take part in the control of the circadian rhythm and counteract cell differentiation (Bi et al., 2015; Honma et al., 2002; Sato et al., 2004).

In summary, the observation that overlapping gene signatures frequently determine subpopulations justified the question whether the cellular heterogeneity in the epidermis was best represented as a set of distinct, clearly delineated clusters, or can be explained better by another model. Thus, we next sought to identify and characterize the biological processes that may give rise to HF and IFE keratinocyte heterogeneity.

Reconstruction of IFE Cell Differentiation by Pseudotemporal Ordering of Single-Cell Transcriptomes

Since the IFE is constantly renewed, it contains the whole range of basal to terminally differentiated keratinocytes (Fuchs, 1990; Toufighi et al., 2015). An advantage of sequencing single cells is that cells can be ordered along a path according to their transcriptional profile using a network-based approach (Trapnell et al., 2014). This allowed us to reconstruct the differentiation processes by ordering IFE cells along a pseudotemporal differentiation trajectory (Figures 3A and S4A). Increasing cell diameters with differentiation (data not shown), and expression levels of the well-known markers *Krt14* (basal), *Krt10* (mature), and *Lor* (terminally differentiated) along the defined pseudotime axis confirmed that our cell alignment was correct and in accordance with epidermal stratification (Fuchs, 1990). *Mt4* marked a transitory stage, which we resolved in this study (Figure 3B).

We identified 1,627 genes with statistically significant variation in expression levels along the differentiation trajectory

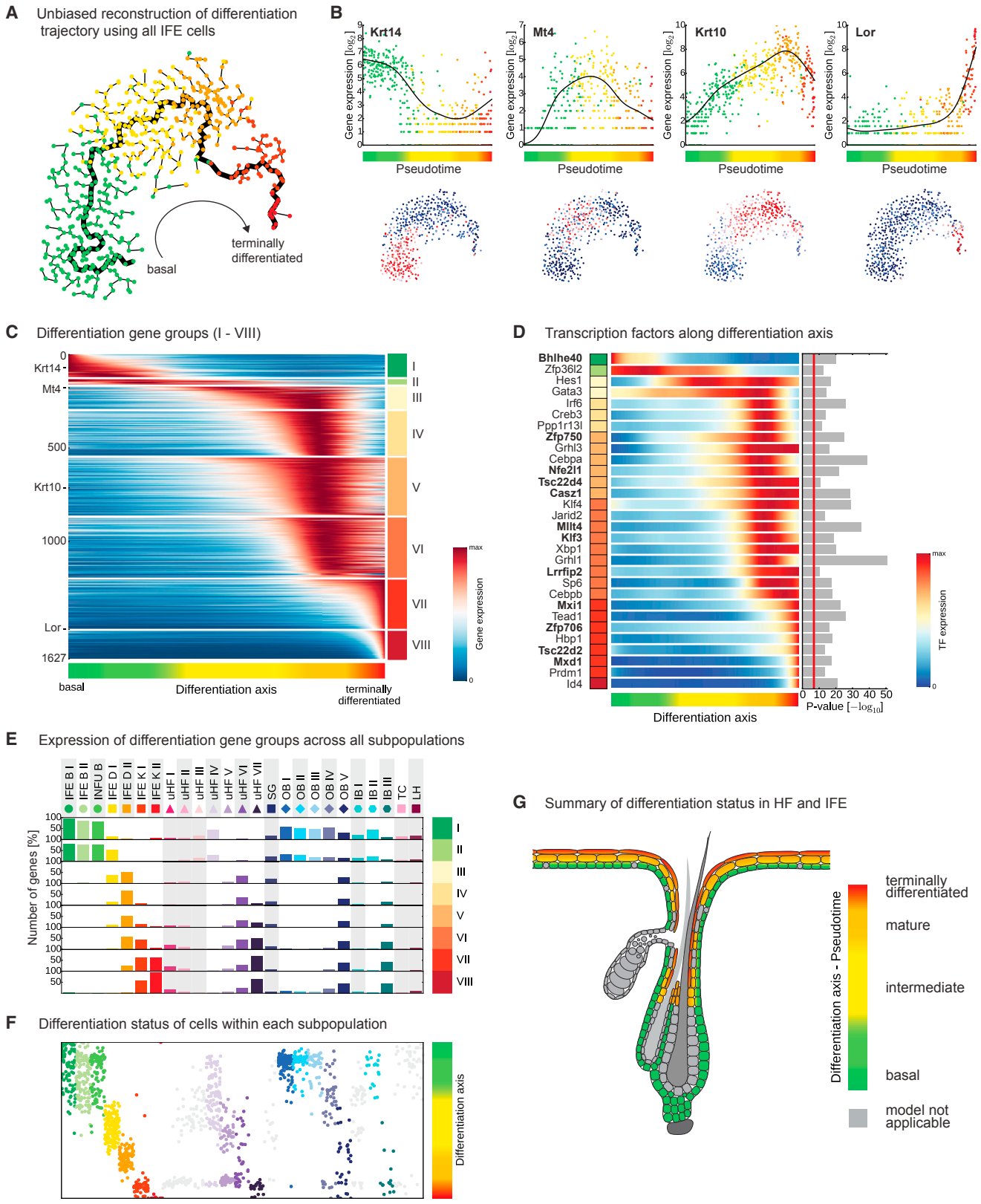
Figure 2. Subclustering of Epidermal Cell Populations

(A–D) Subclustering (second-level clustering) of epidermal cells from the IFE basal (A), upper HF (B), outer bulge (C), and inner bulge (D) compartments. Upper panel: projection of subpopulations onto the t-SNE map of the full dataset introduced in Figure 1D. Lower panel: barplots showing the expression of marker genes per subpopulation. Each bar represents a single cell, and the black line indicates the average expression over each subpopulation.

(E) Selection of immuno- and single-molecule FISH (symbols italics) stainings to visualize subpopulation localization within the tissue. Arrowheads highlight the position of the populations: IFE BI (filled arrowhead)/BII (empty arrowhead); uHF I (filled arrowhead)/II (empty arrowhead); OB III (filled arrowhead; dashed line marks lower end of KRT15 gap). HS, hair shaft. SG, sebaceous gland. CH, club hair. Scale bars, 10 μm. See also Figure S3L.

(F) Identity and marker genes of cell populations defined during second-level clustering.

(G) Summary of the approximate location of each defined subpopulation in the IFE, SG, and HF.



(legend on next page)

(pseudotime-dependent genes, Figure S4B), and these genes clustered into eight groups according to their expression pattern during the differentiation process (Figures 3C and S4C), which also were linked to distinct functional terms (Figure S4D). Basal cells (group I) were defined by a low number of genes primarily involved in extracellular matrix deposition and interaction, cell proliferation, and tissue development. After a transitional stage (II), in which the basal signature was slowly reduced while ribosomal genes peak (III), we saw a first wave of genes linked to epidermal maturation, fatty acid metabolism and cholesterol synthesis, cell-cell junction formation, and protein transport (IV–VI). Toward the end of the cell's life cycle, a second wave of genes involved in cornified envelope formation, ceramide synthesis, and proteolysis became active (VII and VIII) (Table S4). To gain insight into the molecular regulation of epidermal differentiation, we selected the 30 most pseudotime-dependent transcription factors (TFs) and analyzed their expression patterns during the differentiation process (Figures 3D and S4D). While only a few TFs (e.g., *Bhlhe40*, *Zfp36l2*) could be linked to the basal and intermediate signatures, we found a high number of new (e.g., *Casz1, Klf3*, *Lrrfip2*, *Mllt4*) and previously described (*Gata3*, *Grhl1*, *Hes1*, and *Prdm1*) (Kaufman et al., 2003; Kretzschmar et al., 2014; Mlacki et al., 2014; Wang et al., 2008) TFs that could play a role in the regulation of epidermal maturation and terminal differentiation (Figure 3D). In sum, our single-cell resolution data enabled the reconstruction of genetic programs during IFE differentiation in unprecedented detail.

A Majority of HF Subpopulations Express Large Sets of Pseudotime-Dependent Genes

Having defined the genetic program of differentiation in the IFE, we next asked to what degree this differentiation program was applicable to other epidermal cell populations. Interestingly, we observed that the vast majority of epidermal cell populations expressed large numbers of pseudotime-dependent genes in accordance with distinct stages in the differentiation process (Figures 3E, 3F, S4E, and S4F). For instance, most outer bulge subpopulations (OB I–OB V) robustly expressed a large subset of basal genes, while the cells of the upper HF seemed to traverse the complete differentiation program from basal (uHF IV) over intermediate (uHF V) to mature (uHF VI) and terminally differentiated (uHF VII). In order to further demonstrate that IFE and HF cells share core differentiation gene signatures, we identified and modeled the differentiation program independently in the upper HF and found large congruency with IFE differentiation (Figure S4G). The few cell populations (TC, LH, SG, uHF I–III, and IB I) that could not be robustly linked to a particular stage in the differentiation program (Figures 3E, 3F, S4E, and S4F), exhibited immune- and SG-related cellular functions, or underwent an entirely distinct differentiation path like the inner bulge cells (Hsu et al., 2011). Overall, the differentiation program that was identified from analyses of IFE cells seemed universal for most epidermal keratinocytes, summarized in Figure 3G, and accounted for one of the largest sources of cellular heterogeneity throughout the epidermis.

Identification of Spatial Gene Signatures along the Proximal-Distal HF Axis

To further dissect sources of cellular heterogeneity in the HF that are independent of the differentiation signature, we selected all basal IFE and basal HF cells and projected them into t-SNE space. Cells with IFE, uHF, OB, and IB signatures separated into four overlapping clusters positioned along a path, which was used to model a pseudospatial axis similar to the pseudotemporal ordering of the differentiation trajectory (Figures 4A and S5A). Intriguingly, this pseudospatial ordering robustly reproduced the spatial localization of basal subpopulations (Figure 2G) along the proximal-distal axis of the HF (Figures 4B and 4E).

We identified 547 significantly pseudospace-dependent genes and grouped these into eight spatial signatures (Figures 4C and S5B–S5D). A first group of pan-basal genes with peaked expression in the IFE (I), a group of genes most highly expressed in IFE basal (II), a group of genes shared by IFE and uHF basal cells (III), an exclusive uHF signature (IV), a group of genes linked to the Gli1^{+} population in the distal bulge region (V), an outer bulge signature (VI), a pan-bulge signature (VII), and an exclusive inner bulge signature (VIII) (Table S5). Screening for pseudospace-dependent TFs revealed that only a small number of TFs were linked to IFE and uHF basal signatures (e.g., *Ahr*, *Ets2*, *Gata6*, *Tsc22d1*) (Figures 4D and S5D). In contrast, TFs were overrepresented in bulge signature genes that can be roughly classified into three groups: TFs most strongly linked to upper bulge signatures (e.g., *Gli1*, *Runx1*), the outer bulge

Figure 3. Reconstruction of the Epidermal Differentiation Process

(A) Pseudotemporal ordering of IFE cells (n = 536) in t-SNE space, using a minimum spanning tree. The longest path through the graph is highlighted and cells are colored according to first-level clustering.

(B) Validation of pseudotemporal ordering of IFE cells using the known basal (*Krt14*), mature (*Krt10*), and terminally differentiated (*Lor*) cell stage markers and *Mt4*, a transient marker defined in this study. Upper panel: gene expression in IFE cells plotted along pseudotime and fitted with a cubic smoothing spline (black line). Lower panel: gene expression projected onto the t-SNE map shown in (A).

(C) "Rolling wave" plot showing the spline-smoothed expression pattern of pseudotime-dependent genes (n = 1,627) clustered into eight groups (I–VIII) and ordered according to their peak expression.

(D) "Rolling wave" plot showing the spline-smoothed expression pattern of the 30 most significantly differentiation-related transcription factors (TFs). TFs were ordered according to group membership (left) and peak expression as shown in (C). P-values for pseudotime dependency are shown on the right. Red line marks Bonferroni-corrected significance threshold of 0.001. TFs marked in bold have not been previously described as relevant for epidermal stratification.

(E) Expression of differentiation-related genes in all epidermal subpopulations defined by either first- or second-level clustering. Bars show the percentage of genes expressed over baseline with 95% posterior probability (negative binomial regression model) in each of the populations for every differentiation group (I–VIII). Populations where the pseudotime model is not applicable are shaded gray.

(F) Position of epidermal cells from each subpopulation plotted on the differentiation axis (defined by highest Pearson correlation). Populations where the pseudotime model is not applicable are colored light gray.

(G) Summary illustrating the differentiation status of cells in the HF and IFE.

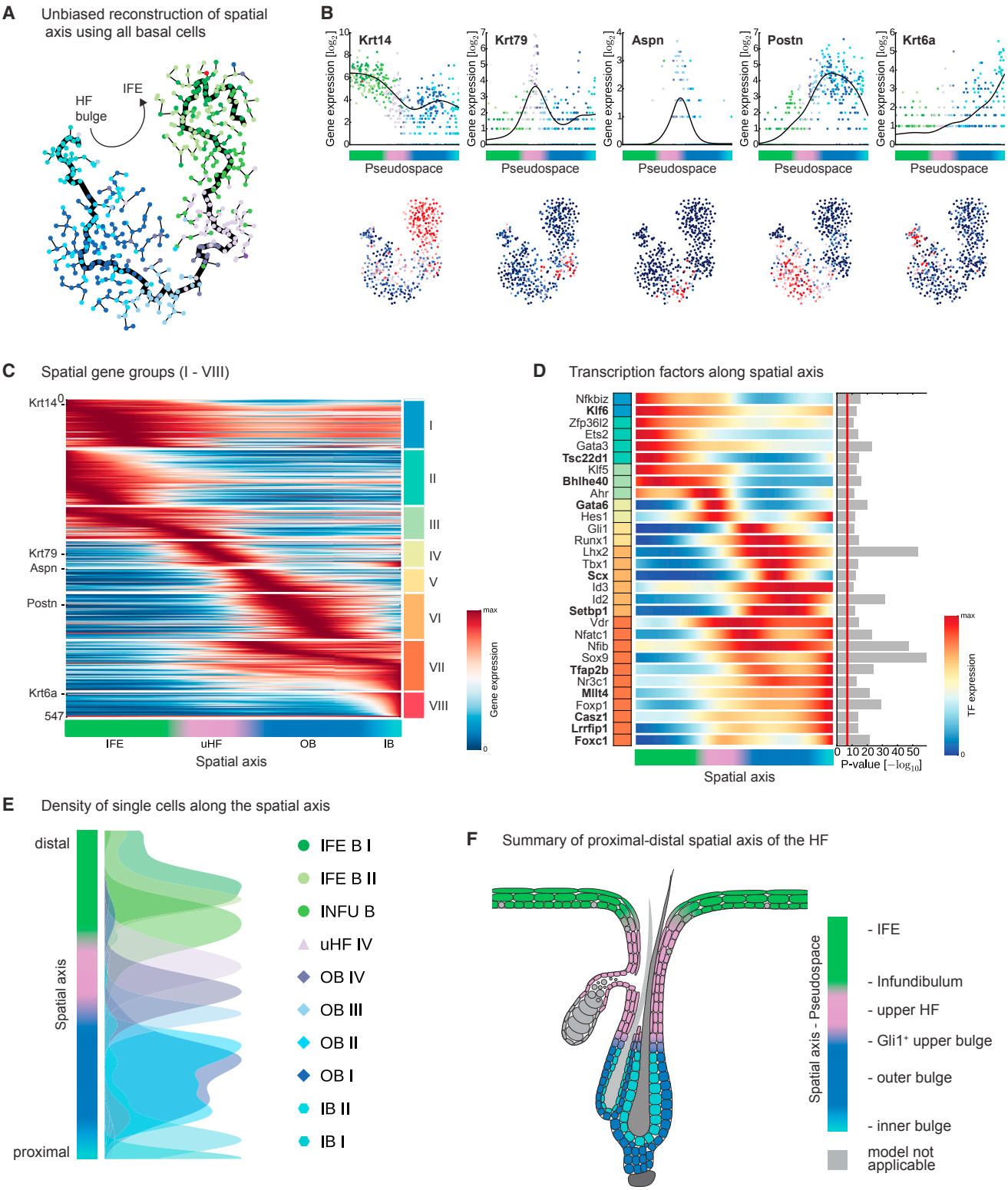


Figure 4. Defining Spatial Gene Expression Signatures

(A) Pseudospatial ordering of basal cells (n = 486) in t-SNE space, using a minimum spanning tree. The longest path through the graph is highlighted and cells are colored according to second-level clustering.

(B) Validation of pseudospatial ordering of basal cells using known and **new** IFE basal (*Krt14*), upper HF *(Krt79*), Gli1^{+} outer bulge (***Aspn***), general outer bulge (***Postn***), and inner bulge (*Krt6a*) markers. Upper panel: gene expression in basal cells plotted along the pseudospace trajectory and fitted with a cubic smoothing spline (black line). Lower panel: gene expression projected onto the t-SNE map shown in (A).

(legend continued on next page)

(e.g., *Tbx1*, *Lhx2*), and pan-bulge or pan-HF TFs (e.g., *Foxp1*, *Sox9*, *Tfap2b*). Overall, we identified well-known TFs in the HF and a variety of putatively new regulatory factors in the HF and IFE (Figures 4D and S5D). The fact that the proximal-distal axis spanning from the inner HF bulge to the IFE could be robustly recapitulated (Figures 4E and 4F) suggests that spatial cues generate gradient responses in keratinocyte populations along the proximal-distal axis (Figure S5E). Moreover, most spatial signatures in the HF were expressed independently of the differentiation state (Figures S5F–S5I). In sum, this analysis demonstrated that spatial gene signatures have a large influence on the overall cellular heterogeneity.

The Differentiation and Spatial Signatures Explain Most Epidermal Heterogeneity

To quantitatively assess to what extent differentiation and spatial gene signatures could explain the observed cellular heterogeneity in the epidermis, we modeled the gene expression profile of each cell as a combination of differentiation and spatial signatures, and five additional types of signatures (two SG signatures and three immune cell related signatures) (Figure 5). We first explored the positions of cells along the pseudotime- and pseudospace-axis (pseudospacetime model, Figures 5A and S6A), and most epidermal subpopulations were located in specific regions in pseudospacetime (Figure 5B). We divided the pseudospacetime model into 15 equally sized bins along each axis and used bin-membership of cells as predictors in a negative binomial regression model (STAR Methods). For each predictor, we were able to define distinct gene sets, which were expressed over the model baseline (i.e., the background expression found in all cells of the data) (Figure 5A, upper and left-hand side panel, and Figure 5C). To evaluate how well the model explained the observed single-cell data, we compared the in silico transcriptomes generated from the model for each cell with the experimentally observed number of molecules. We computed the numbers of molecules that were in agreement (explained molecules), and the numbers of molecules in excess (overexplained molecules) or lacking (underexplained molecules) in the modeled compared to the observed transcriptomes per cell (Figures S6B and S6C). In parallel, we used the same modeling strategy but binned cells based on the first- or second-level clustering. Intriguingly, the pseudospacetime model had an equally high ''explanatory performance'' as the first- and second-level clustering data (Figures 5D and S6D), suggesting that the differentiation and spatial signatures effectively covered all heterogeneity identified across the main populations (first-level clustering) and sub-populations (second-level clustering). The baseline signature explained around 50% of molecules in the dataset (Figure 5E), and we next investigated the additional ''explanatory power'' of the respective signatures. The differentiation signature could resolve additional 25%, and, together with the spatial signatures, more than 95% of transcriptome molecules could be explained. The remaining signatures had minor roles, as they were only important for certain cells such as immune cells (Figure 5E). When analyzed from a cell population perspective, the spatial signatures played larger roles in explaining gene expression in basal cells, and the differentiation signatures accounted for most of the non-baseline molecules in suprabasal cells (Figure S6E). We conclude that the gene expression programs associated with differentiation and the proximal-distal spatial axis explain most transcriptional heterogeneity within the epidermis.

Stem Cells Share a Basal Transcriptional Signature

In the last two decades, numerous studies have described and transcriptionally profiled distinct murine epidermal cell populations in the HF and the IFE with long-term self-renewal capabilities (Blanpain et al., 2004; Brownell et al., 2011; Füllgrabe et al., 2015; Greco et al., 2009; Jaks et al., 2008; Mascré et al., 2012; Page et al., 2013; Snippert et al., 2010). These studies have identified important gene signatures, but they were inherently limited to measuring averages across cell populations due to predefined marker-based sorting strategies. Therefore, it is still unknown what distinguishes cells that express stem cell and progenitor markers (SCMs) from cells that do not. To this end, we selected cells expressing the established SCMs *Cd34*, *Lgr5*, *Lgr6*, *Gli1*, *Lrig1*, or high levels of *Krt14* (*Krt14*^{hi}). As expected, we found that most of the SCM^{+} cells exhibited a basal phenotype (Figure 6A). We next selected all basal cells (STAR Methods), projected them into t-SNE space (Figures 6B and S7B), and marked *Cd34*, *Lgr5*, *Lgr6*, *Gli1*, *Lrig1*, or *Krt14*^{hi} cells on this t-SNE map to display their location (Figures 6B and 6C). As a control, pre-sorted *Lgr5*-EGFP^{+} keratinocytes (Jaks et al., 2008) were processed in the same way as the 1,422 cells in this study and found to occupy the same locations in the t-SNE plot as *Lgr5*-expressing cells did in Figure 6C (data not shown). Interestingly, we observed that, although showing clear peaks in distinct compartments, the expression of most SCMs was scattered over several basal compartments (Figures 6B, 6C, S7A, and S7B), and SCM expression alone was not sufficient to clearly delineate basal cell populations in our dataset. It needs to be determined whether or not these observations could have implications when using SCM-promoter-based lineage tracing (Kretzschmar and Watt, 2014). However, when analyzing each heterogeneous SCM^{+} population for shared gene expression, we identified robust SCM-linked signatures that were independent of differentiation stages (Figures S7C–S7F; Table S6), underlining the strong impact of niches on gene expression.

As most of the SCMs were predominantly expressed in basal cells (Figure 6A), we asked whether basal cells that expressed

(C) ''Rolling wave'' plot showing the spline-smoothed expression pattern of pseudospace-dependent genes (n = 547) clustered into eight groups (I-VIII) and ordered according to their peak expression.
(D) ''Rolling wave'' plot showing the spline-smoothed expression pattern of the 30 most significant spatially expressed TFs. TFs were ordered according to group membership and peak expression as shown in (C). P-values for pseudospace dependency are shown on the right. Red line marks Bonferroni-corrected significance threshold of 0.001. TFs marked in bold have not been previously described as relevant for cellular heterogeneity along the proximal-distal axis.
(E) Peak positions of basal cell populations and IB I (defined in second-level clustering) on the spatial axis visualized by kernel density estimation. The organization of the cell populations confirms their spatial positioning in IFE and HF along the proximal-distal axis.
(F) Summary illustrating spatial signatures in epidermal cell populations.

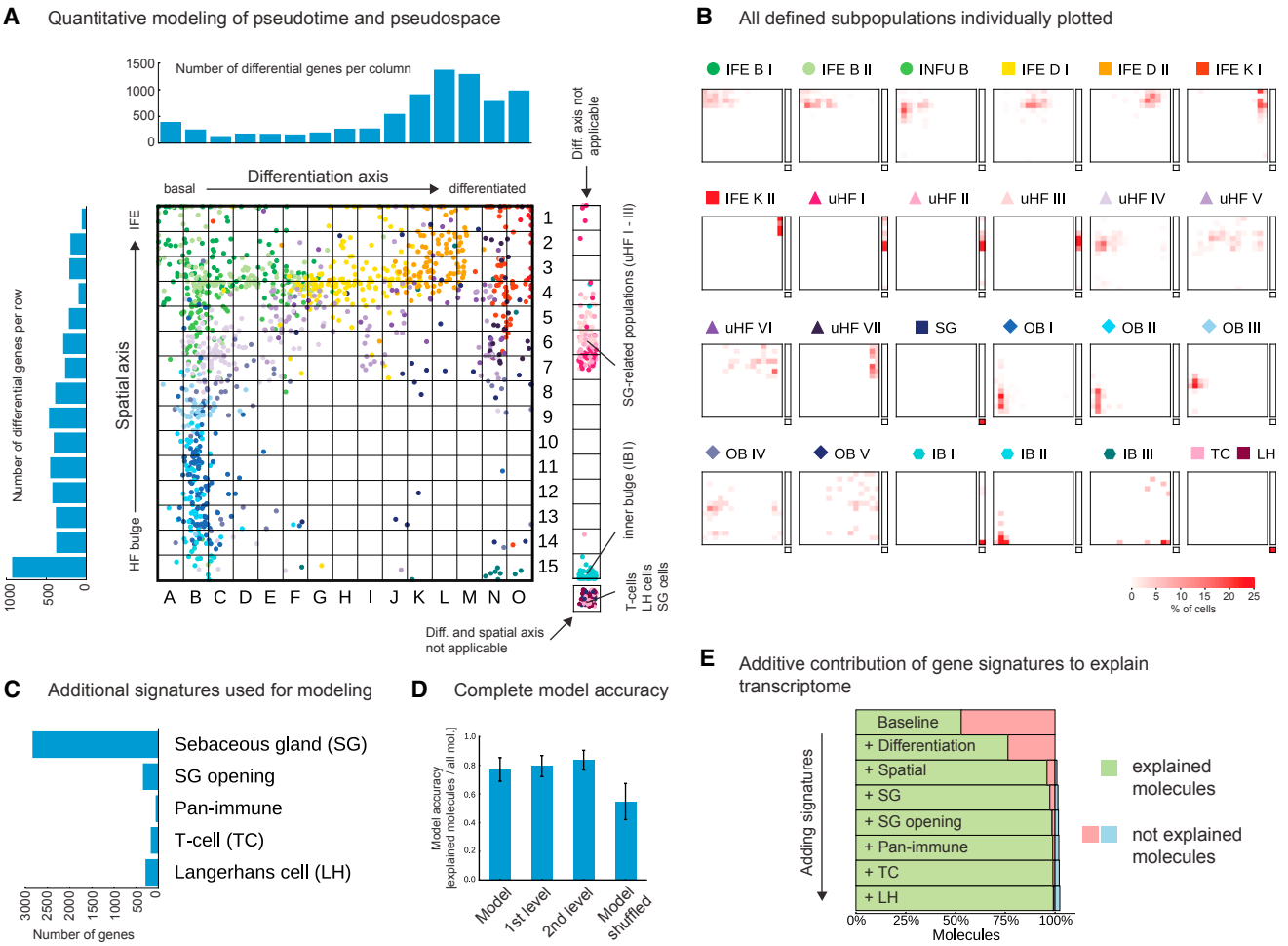


Figure 5. Modeling Transcriptional Heterogeneity Using Space and Time Signatures

(A) Pseudospacetime: matrix showing each cell's (dots) identity along the differentiation- and spatial-axis, in which both axes were divided into 15 equally sized bins. The numbers of genes expressed over baseline (95% posterior probability, negative binomial regression model) for each bin are shown in barplots (upper and left panels). Cells with expression patterns that could not be placed along the differentiation- and spatial-axes are presented in a separated bar to the right.

(B) The pseudospacetime positions of cells from each cell population defined by either first- or second-level clustering, visualized as percentage of cells per bin.

(C) The number of genes expressed over baseline (95% posterior probability) for the additional signatures used for modeling the transcriptomes of all cells (including SG-related and immune populations).

(D) Model accuracy for the model (including all signature model predictors) in comparison with model accuracy based on either grouping cells according to the first- or second-level clustering or after shuffling the model-predictor matrix (negative control). The model accuracy was computed as the ratio of explained molecules (present in both the simulated and observed) to the sum of explained and unexplained molecules. For each model, the mean and SD of the model accuracy over each group are shown. See Figure S6D for results of each individual cell population.

(E) Percentage of molecules (averaged over all cells) explained by models of increasing complexity. The explained molecules are indicated in green, under-explained in red, and overexplained in blue.

SCMs (73% of basal cells, Figure 6D) had distinct transcriptional programs in comparison to basal cells without SCM expression. SCM$^-$ basal cells were in general "less basal" than those cells expressing SCMs, as evident from projecting these two groups of cells onto the differentiation axis (Figure 6E) and were enriched in the IFE and upper HF compartments (Figure 6F). Using negative binomial regression, we obtained a set of genes that was higher expressed in SCM$^+$ compared to the SCM$^-$ cells. Interestingly, the SCM$^+$-enriched genes did not constitute a "unique stem cell signature" and were instead mostly part of a pan-basal gene expression program including components that are involved in the extracellular matrix (ECM) and basement membrane formation, and cell adhesion (Figures 6G and S7G–S7J; Table S6). Some of these genes have been found to be expressed in SCM$^+$ cell populations (Blanpain et al., 2004; Greco et al., 2009; Tumbar et al., 2004), and the recently reported importance of COL17A1 for counteracting HF stem cell aging underpins our findings (Matsumura et al., 2016).

Altogether, we did not observe a clearly delineated transcriptional state (i.e., a set of genes uniquely expressed in stem cells) that set SCM$^+$ and SCM$^-$ basal cells apart. What was shared between all SCM$^+$ basal cells was a stronger pan-basal signature. Moreover, the gene expression signatures separating

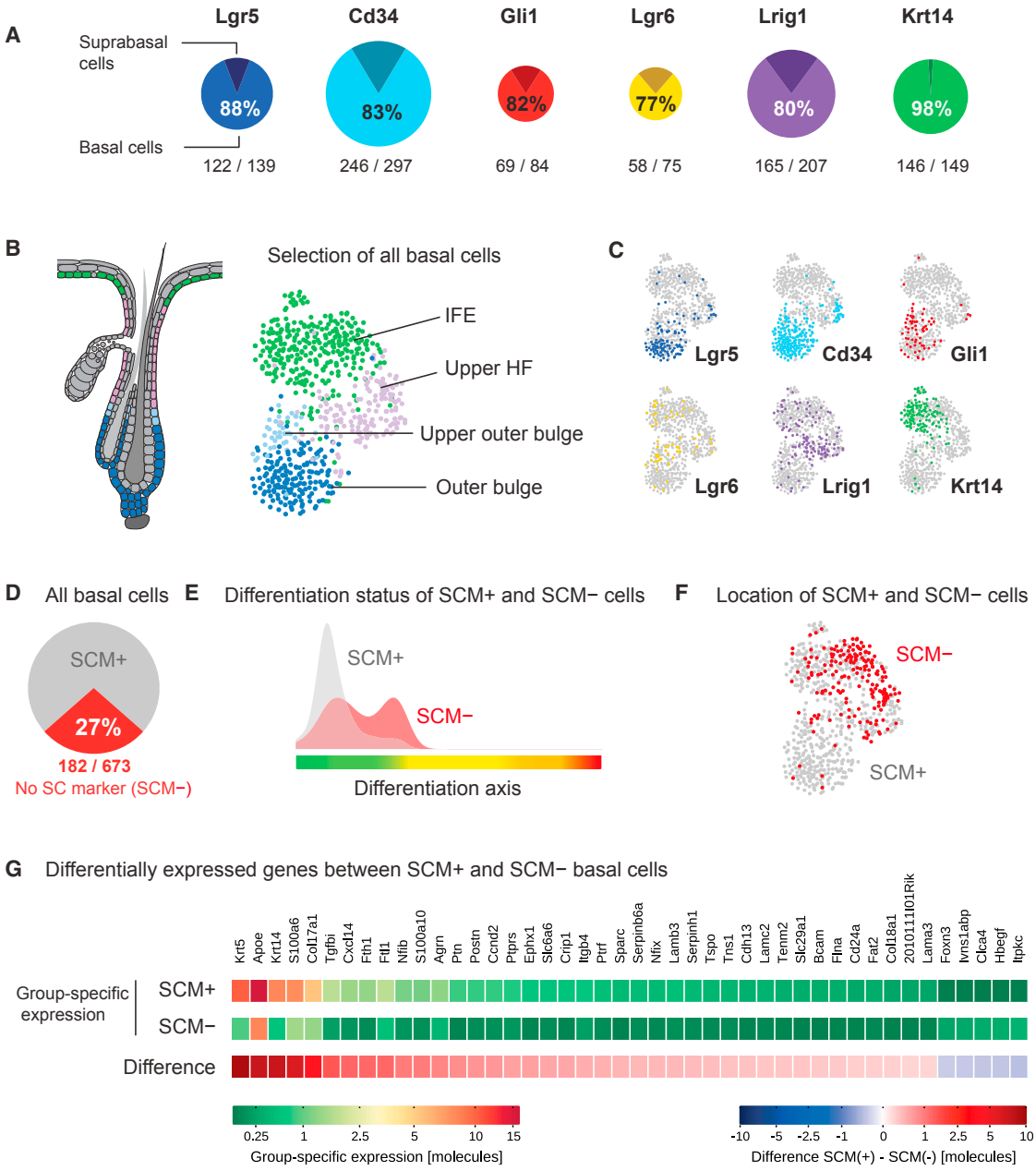


Figure 6. Single-Cell Analyses of Epidermal Stem Cell Populations

(A) Percentage of basal (pseudotime ≤300) and non-basal cells, in each population of cells expressing *Lgr5*, *Cd34*, *Gli1*, *Lgr6*, *Lrig1*, or *Krt14*, respectively. For basal cells, the percentage and the number of cells per total cells are given.

(B) Selection of all basal cells. Right panel: projection of all basal cells (pseudotime ≤300; with and without SCM expression) onto t-SNE space, colored according to the defined cell compartments (first- and second-level clustering). Left panel: illustration summarizing the location of the compartments.

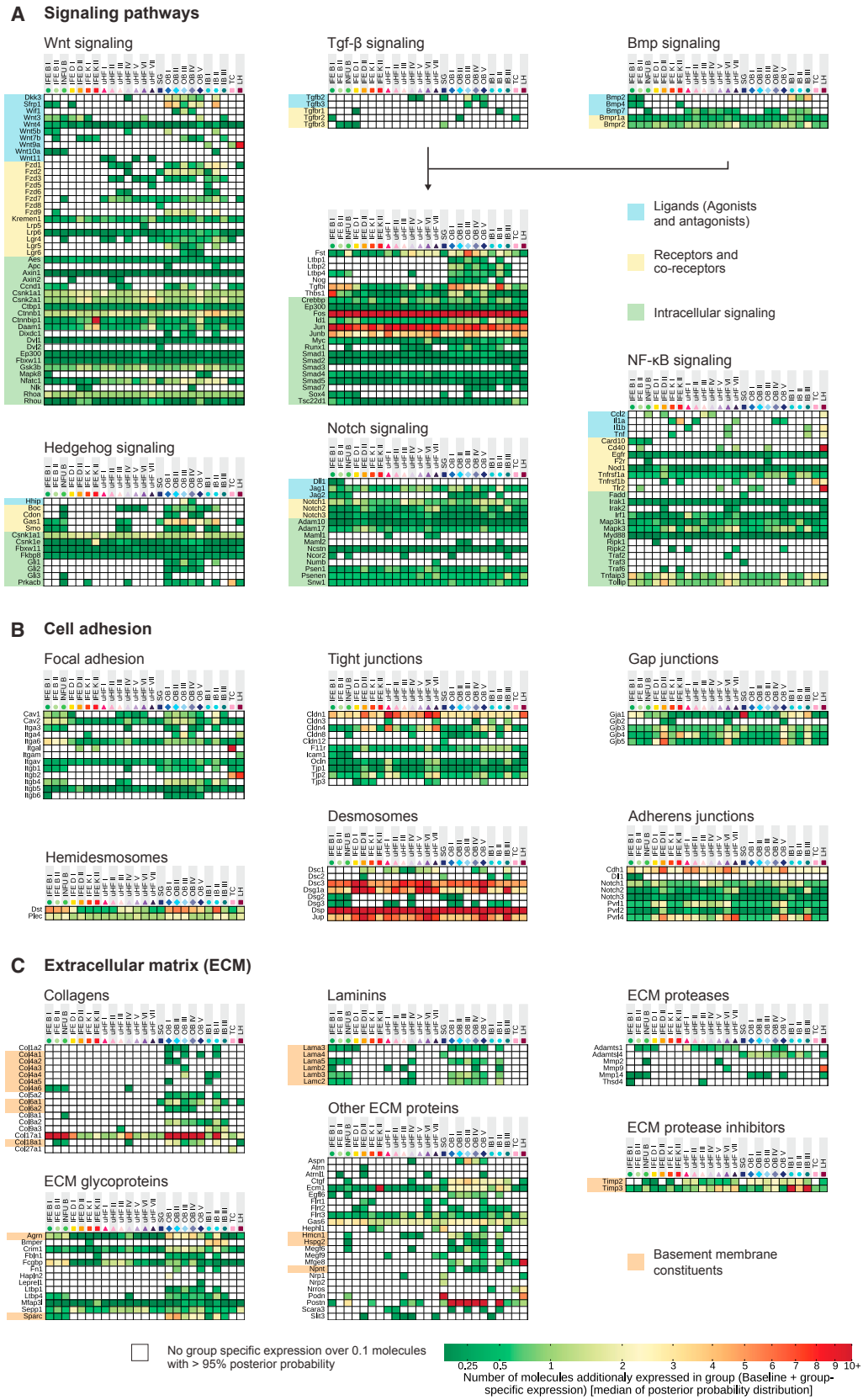
(C) Mapping of basal cells to the t-SNE map defined in (B) according to the expression of SCMs, for each marker gene respectively.

(D) Percentage of basal cells that do not express any of the SCMs *Lgr5*, *Cd34*, *Gli1*, *Lgr6*, *Lrig1*, or *Krt14* (in red).

(E) Density of basal cells with (gray) and without (red) SCM expression along the pseudotime axis.

(F) Projection of the basal cells that did not express any SCMs (red) onto the t-SNE map defined in (B).

(G) Heatmap of 44 genes that are differentially expressed between SCM^+ and SCM^- basal cells. Negative binomial regression was used to define specific SCM^+ and SCM^- gene expression signatures (i.e., the additional number of molecules expressed for each gene if a cell belongs to the SCM^+ or SCM^- group). For each gene, the group-specific expression in SCM^+ and SCM^- cells as well as the difference between both groups is shown (median number of molecules).



(legend on next page)

established SCM$^+$ populations are mostly linked to the spatial axis (Figure S7K).

Comparison of Signaling Pathway, Cell Adhesion, and ECM Components across All Epidermal Subpopulations

The identification of 25 distinct (sub-) populations in telogen epidermis enabled direct comparisons of gene expression patterns across all these cell populations. For epidermal homeostasis, firm regulation of signaling pathway activation, niche-component expression, and epigenetic mechanisms are critically important (Hsu et al., 2014; Mesa et al., 2015; Rompolas and Greco, 2014; Botchkarev et al., 2012; Botchkarev and Flores, 2014). Thus, we focused the comparison between subpopulations on six epidermal key pathways (Wnt, Hedgehog [Hh], NF-kB, Notch, Bmp, and Tgf-b), cell adhesion and ECM components (Figures 7A–7C), and components of the epigenetic machinery (data not shown). Unlike the expression of signaling pathway and ECM-related genes, the analysis of epigenetic components did not reveal distinctive expression patterns and these genes were generally expressed at relatively low levels throughout the epidermis.

Markedly, in the Wnt, Hh, Bmp, and Tgf-b signaling pathways we observed most heterogeneity in the expression of ligands, receptors, and their corresponding modulators, whereas their intracellular pathway components were expressed relatively evenly across all subpopulations with a few exceptions such as *Gli1* expression indicating active Hedgehog signaling in outer bulge subpopulations (Brownell et al., 2011). Notch pathway components were generally expressed in all subpopulations, with exception of *Jag2*, which was detected over baseline only in the most basal layers of the IFE and the bulge. Interestingly, there seemed to be a trend of a receptor-ligand division between IFE and HF, most evident in the Wnt and Tgf-b pathways. Wnt ligands for example showed higher expression in the IFE basal layer while Wnt receptors were predominantly expressed in HF populations.

While the expression of signaling pathway genes diverged primarily along the spatial axis, genes linked to different types of cell-cell and cell-ECM junctions showed a strong heterogeneity along the differentiation axis. As expected, genes linked to focal adhesion and hemidesmosome formation were highest expressed in basal populations irrespective of location, while the formation of tight junctions, adherens junctions, gap junctions, and desmosomes was increased in all suprabasal populations.

Among ECM genes, we observed functional division between gene sets linked to a pan-basal state and niche/location related gene signatures. While collagen *Col17a1*, a subset of glycoproteins (*Agrn, Fcgbp*) and most laminins (*Lama3, Lama5, Lamb2, Lamc2*) were expressed at equally high levels across all basal keratinocytes, the majority of ECM genes exhibited a spatial expression corresponding to the pseudospace-related expression patterns identified in Figure 4C.

Overall, these comparisons demonstrated the utility of the transcriptional data of murine epidermis generated within this study, and with the accompanying online tool (http://kasperlab.org/tools or http://linnarssonlab.org/epidermis/) we hope to inspire and enable additional studies in skin biology by using this in-depth single-cell resource.

DISCUSSION

We generated a large resource of single-cell gene expression profiles from murine keratinocytes and used it to dissect epidermal heterogeneity. Four major novelties and highlights of this study are discussed in the following sections.

Identification of Previously Unidentified Epidermal Subpopulations in IFE and the HF

Two cycles of unsupervised clustering, using all cells or subsets of cells, revealed an apparent transcriptional hierarchy between populations (main clusters) and their subpopulations in the epidermis. The 13 main clusters reflected the major IFE differentiation stages and three broad spatial compartments of the HF (upper HF, outer bulge, and inner bulge) and were grouped according to their compartments and functions supporting compartmentalized HF maintenance (Schepeler et al., 2014). Surprisingly, our unbiased clustering (first and second level) failed to demarcate several previously described cell populations, such as Gli1$^+$ or Lgr5$^+$ cells in the lower bulge, Lgr6$^+$ cells of the isthmus, and the Lrig1$^+$ cells in the infundibulum (Table S3) (Brownell et al., 2011; Füllgrabe et al., 2015; Jaks et al., 2008; Jensen et al., 2009; Snippert et al., 2010). Instead, we found that each of these marker-based populations encompassed several subpopulations that were defined in this study. In consequence, although expression of these marker genes has been very useful as genetic tools to study general cell and lineage dynamics during HF maintenance (Jaks et al., 2010; Kretzschmar and Watt, 2014), these markers are not well suited for defining transcriptionally homogenous populations.

Many of the subpopulations we identified have been previously described using immunostaining, lineage tracing or cell-sorting based transcriptional profiling (e.g., Blanpain et al., 2004; Brownell et al., 2011; Füllgrabe et al., 2015; Jaks et al., 2008; Jensen et al., 2009; Snippert et al., 2010; Veniaminova et al., 2013). However, the clustered single-cell transcriptomes of this study yielded more ''pure'' transcriptional signatures compared to marker-based sorting strategies and thus allowed for a more precise molecular characterization of subpopulations. In addition, we describe several populations that have not been previously identified, have not been described in molecular terms or were only assumed to exist (Table S3). For example, we found two basal subpopulations in the IFE that neither represented the previously described Ivl$^+$ or Lgr6$^+$ populations (Füllgrabe et al., 2015; Mascré et al., 2012). Future studies are needed to resolve whether these two IFE populations represent coexisting cell populations of closed lineages or reflect certain stromal microenvironments or different differentiation stages. Moreover, we found a group of cells in the HF with simultaneous

Figure 7. Functional Signatures Expressed in Epidermal Subpopulations

(A–C) Expression of genes linked to signaling pathways (A), cell adhesion (B), and extracellular matrix and basement membrane constituents (C) in each epidermal population (defined in either first- or second-level clustering). Shown is the median number of molecules expressed in each cell population (negative binomial regression model).

expression of outer bulge (OB) and inner bulge (IB) signatures, which could be placed in the OB. IB cells have the important role to keep OB cells quiescent, until inductive hair growth signals from the dermal papilla stimulate proliferation of lower bulge and hair germ cells in a gradient fashion (Greco et al., 2009; Hsu et al., 2011). Given that in principle all OB cells are competent to enter cell cycle upon damage (Hsu et al., 2011) yet only a subset does during homeostatic hair growth, some cells may have an extra safety mechanism to counteract cell-cycle entry during early anagen by autocrine expression of inhibitory IB signals such as *Fgf18*.

We also identified two populations lining the opening of the SG with a remarkably high expression of the defensin *Defb6*. Defensins are small cysteine-rich cationic proteins and function as host defense peptides (Gallo and Nakatsuji, 2011 and references therein). The strategic placement of these two populations at the SG opening, where sebum is released to grease the entire epidermis, indicates DEFB6 as critical in protecting the HF bulge against microorganisms (Chronnell et al., 2001). Elucidating the function of these cells in the context of epidermal physiology will be an interesting topic for future studies.

Transcriptional Resolution of the Differentiation and Proximal-Distal Axis

While our reconstruction of IFE differentiation did not challenge the accepted three-tier model, which postulates a differentiation trajectory from the basal layer over maturation in the spinous layer toward terminal differentiation in the granular layer, we found transient cell states, which are nearly unresolvable with bulk cell methods. Intriguingly, we observed a dramatic transcriptional change along the differentiation axis between gene groups I and III (Figure 3C). It is tempting to speculate whether this change indicates a point of no return along the differentiation trajectory, so that all basal cells—before reaching this point—are to some extent plastic and can provide long-term renewal capacity, although their likelihood to give rise to a long-term surviving clone declines as they move further along the differentiation axis.

Most of the HF subpopulations expressed large sets of genes associated with a distinct differentiation stage and could be positioned along the IFE differentiation axis. To what extent HF and IFE subpopulations share differentiation programs needs further analysis, but these results are indicative of a general pan-differentiation program for keratinocytes with only a few exceptions: SG-related cells and one inner bulge cell cluster (IB I). Most interesting in this regard are the IB I cells. These cells originate from one of the outer bulge populations, relocate during anagen to the lower part of the growing HF, and home back to the bulge in the following catagen-telogen transition to function as proliferation-inhibitory bulge-niche cells (Hsu et al., 2011). The fact that IB I cells could not be placed along the axis of the pan-differentiation program raises the question of whether anagen growth uses an entirely different differentiation program compared to keratinocytes of the non-cycling part of the HF.

Applying a similar strategy as for the reconstruction of the differentiation trajectory (Trapnell et al., 2014), we observed that the basal cells can be aligned along a continuous trajectory reflecting the proximal-distal HF axis. Recent lineage-tracing studies suggest compartmentalized maintenance of the HF, implying that "invisible" borders keep cells within their compartments and compartments separated (Schepeler et al., 2014). The reconstruction of a continuous profile along the spatial axis, however, requires that cells have gradually overlapping sets of genes along the entire HF axis. Thus, it is tempting to speculate whether this feature is important for the extraordinary plasticity of HF cells, reflected in their ability to replace each other upon damage, and take over the role and functions of the replaced cells (Donati and Watt, 2015). For example, isthmus as well as hair germ cells can directly repair bulge cell damage (Rompolas et al., 2013). During wound repair, HF cells are recruited to the IFE and can even convert to permanent progenitors of the IFE epidermis (Ito et al., 2005; Kasper et al., 2011; Levy et al., 2005; Page et al., 2013), but contribution in the opposite direction to damaged existing HFs has, to our best knowledge, never been reported. In concordance, all HF cells expressed typical IFE signature genes, but IFE cells did not express HF-specific genes. The overlapping expression signatures along the spatial axis do not exclude the existence of compartmental borders during homeostasis, established, for example, by a few critical proteins, but may explain the rapid cellular adaptability of epidermal cells upon damage (Rompolas and Greco, 2014; Takeo et al., 2015), because only a small number of additional genes is necessary for a cell to adjust to a new environment.

A Quantitative Model to Explain Tissue Heterogeneity

The transcriptional differences between most subpopulations of keratinocytes could be quantitatively modeled and reconstructed using only the differentiation and spatial signatures. The only exceptions were *Defb6*$^{+}$ cells around the SG opening (uHF I and uHF II), which exhibited a unique signature and gene expression patterns of their spatial niche but no pattern of pan-keratinocyte differentiation, and mature SG cells, T cells, and Langerhans cells that only expressed cell-type-associated gene expression signatures. That keratinocyte populations and cellular heterogeneity can effectively be modeled using only two continuous signatures represents unique quantitative insights into cellular heterogeneity, and it will be interesting to investigate the universality of this model for other cell types in other tissues.

Comparison of Epidermal Stem Cell Populations

Finally, we compared basal cells with and without expression of reported stem and progenitor cell markers in an effort to identify a "stemness" gene expression signature. Interestingly, no unique gene expression signature was found in cells expressing these markers. Instead, our results suggest that long-term self-renewing cells in the IFE and the HF do not have a distinct stemness signature other than having a strong basal signature in common, whereas they differ in expression of spatial signatures relating to their location. Altogether, the capacity for long-term self-renewal in the IFE and HF might not require a stemness gene expression signature (Clevers, 2015), but stem cell function might rather coincide with the ability of cells to maintain or occupy certain spatial positions within a tissue and the ability to attach to the basement membrane.

In summary, our reference atlas of transcriptionally distinct cells in the murine epidermis and online tools for custom data

visualization and querying will enable deeper inquiries into the physiology of the skin.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mice
- METHOD DETAILS
  - Cell Isolation
  - Cell Capturing, Quality Control, and Single-Cell cDNA Synthesis
  - Tagmentation and Isolation of 5′ fragments
  - Illumina High-Throughput Sequencing and Processing of Sequencing Reads
  - Yield and Quality of Sequencing
  - Systematic Staining of All Populations by Immunohistochemistry and Single Molecule FISH
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Analysis and Visualization of Processed Sequencing Data
  - Implementation
  - Unsupervised Clustering Using Affinity Propagation
  - Nonlinear Dimensionality Reduction with t-Distributed Stochastic Neighbor Embedding
  - Negative Binomial Regression of Gene Expression
  - Pseudotemporal/-Spatial Ordering of Cells
  - Constructing Gene-Gene Neighbor Networks
  - Gene Set Enrichment Analysis
  - Data Analysis Process
- DATA AND SOFTWARE AVAILABILITY
  - Software
  - Data Resources
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.cels.2016.08.010.

AUTHOR CONTRIBUTIONS

S.J., S.L., and M.K. conceived and designed the study. S.J., A.Z., G.L.M., and P.L. performed sequencing experiments and computational analyses. S.J., T.J., and X.S. performed immunostaining experiments and microscopy analyses. S.J., A.Z., T.J., S.L., and M.K. interpreted data. S.J. and M.K. wrote the manuscript with input from all authors.

ACKNOWLEDGMENTS

We thank Alexandra Are, Karl Annusver, and Åsa Bergström for technical help with immunohistochemistry and mice and Anna Juréus for help with RNA sequencing. We are grateful to Rickard Sandberg and Rune Toftgård for feedback and discussion on the manuscript. This work was supported by grants from the Swedish Cancer Society, Swedish Research Council (STARGET), Swedish Foundation for Strategic Research, Center for Innovative Medicine, and Ragnar Söderberg Foundation to M.K., European Research Council (261063, BRAINCELL), and Swedish Research Council (STARGET) to S.L., Human Frontier Science Program to A.Z., and Karolinska Institutet KID funding to S.J. and T.J. Parts of this study were performed at the Live Cell Imaging facility/Nikon Center of Excellence, Department of Biosciences and Nutrition, Karolinska Institutet, supported by grants from the Knut and Alice Wallenberg Foundation, the Swedish Research Council, the Center for Innovative Medicine, and the Jonasson donation to the School of Technology and Health, Royal Institute of Technology, Sweden.

Received: February 4, 2016
Revised: May 11, 2016
Accepted: August 11, 2016
Published: September 15, 2016

SUPPORTING CITATIONS

The following references appear in the Supplemental Information: Collette et al., 2013; Fujiwara et al., 2011; Horsley et al., 2006; Magwene et al., 2003; Nijhof et al., 2006; Zeeuwen et al., 2002.

REFERENCES

Alcolea, M.P., and Jones, P.H. (2014). Lineage analysis of epidermal stem cells. Cold Spring Harb. Perspect. Med. *4*, a015206.

Bi, H., Li, S., Qu, X., Wang, M., Bai, X., Xu, Z., Ao, X., Jia, Z., Jiang, X., Yang, Y., and Wu, H. (2015). DEC1 regulates breast cancer cell proliferation by stabilizing cyclin E protein and delays the progression of cell cycle S phase. Cell Death Dis. *6*, e1891.

Blanpain, C., Lowry, W.E., Geoghegan, A., Polak, L., and Fuchs, E. (2004). Self-renewal, multipotency, and the existence of two cell populations within an epithelial stem cell niche. Cell *118*, 635–648.

Botchkarev, V.A., and Flores, E.R. (2014). p53/p63/p73 in the epidermis in health and disease. Cold Spring Harb. Perspect. Med. *4*, a015248–a015248.

Botchkarev, V.A., Gdula, M.R., Mardaryev, A.N., Sharov, A.A., and Fessing, M.Y. (2012). Epigenetic regulation of gene expression in keratinocytes. J. Invest. Dermatol. *132*, 2505–2521.

Brownell, I., Guevara, E., Bai, C.B., Loomis, C.A., and Joyner, A.L. (2011). Nerve-derived sonic hedgehog defines a niche for hair follicle stem cells capable of becoming epidermal stem cells. Cell Stem Cell *8*, 552–565.

Chronnell, C.M., Ghali, L.R., Ali, R.S., Quinn, A.G., Holland, D.B., Bull, J.J., Cunliffe, W.J., McKay, I.A., Philpott, M.P., and Müller-Röver, S. (2001). Human beta defensin-1 and -2 expression in human pilosebaceous units: Upregulation in acne vulgaris lesions. J. Invest. Dermatol. *117*, 1120–1125.

Clevers, H. (2015). STEM CELLS. What is an adult stem cell? Science *350*, 1319–1320.

Collette, N.M., Yee, C.S., Murugesh, D., Sebastian, A., Taher, L., Gale, N.W., Economides, A.N., Harland, R.M., and Loots, G.G. (2013). Sost and its paralog Sostdc1 coordinate digit number in a Gli3-dependent manner. Dev. Biol. *383*, 90–105.

Cotsarelis, G., Sun, T.T., and Lavker, R.M. (1990). Label-retaining cells reside in the bulge area of pilosebaceous unit: Implications for follicular stem cells, hair cycle, and skin carcinogenesis. Cell *61*, 1329–1337.

Donati, G., and Watt, F.M. (2015). Stem cell heterogeneity and plasticity in epithelia. Cell Stem Cell *16*, 465–476.

Edelstein, A.D., Tsuchida, M.A., Amodaj, N., Pinkard, H., Vale, R.D., and Stuurman, N. (2014). Advanced methods of microscope control using μManager software. J. Biol. Methods *1* (2), e10.

Faith, J.J., Hayete, B., Thaden, J.T., Mogno, I., Wierzbowski, J., Cottarel, G., Kasif, S., Collins, J.J., and Gardner, T.S. (2007). Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. PLoS Biol. *5*, e8.

Frey, B.J., and Dueck, D. (2007). Clustering by passing messages between data points. Science *315*, 972–976.

Fuchs, E. (1990). Epidermal differentiation: The bare essentials. J. Cell Biol. *111*, 2807–2814.

Fuchs, E. (2007). Scratching the surface of skin development. Nature *445*, 834–842.

Fujiwara, H., Ferreira, M., Donati, G., Marciano, D.K., Linton, J.M., Sato, Y., Hartner, A., Sekiguchi, K., Reichardt, L.F., and Watt, F.M. (2011). The basement membrane of hair follicle stem cells is a muscle cell niche. Cell *144*, 577–589.

Füllgrabe, A., Joost, S., Are, A., Jacob, T., Sivan, U., Haegebarth, A., Linnarsson, S., Simons, B.D., Clevers, H., Toftgård, R., and Kasper, M. (2015). Dynamics of Lgr6$^+$ progenitor cells in the hair follicle, sebaceous gland, and interfollicular epidermis. Stem Cell Reports *5*, 843–855.

Gallo, R.L., and Nakatsuji, T. (2011). Microbial symbiosis with the innate immune defense system of the skin. J. Invest. Dermatol. *131*, 1974–1980.

Greco, V., Chen, T., Rendl, M., Schober, M., Pasolli, H.A., Stokes, N., Dela Cruz-Racelis, J., and Fuchs, E. (2009). A two-step mechanism for stem cell activation during hair regeneration. Cell Stem Cell *4*, 155–169.

Guo, N., Krutzsch, H.C., Inman, J.K., and Roberts, D.D. (1997). Thrombospondin 1 and type I repeat peptides of thrombospondin 1 specifically induce apoptosis of endothelial cells. Cancer Res. *57*, 1735–1742.

Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-Seq: Single-cell RNA-Seq by multiplexed linear amplification. Cell Rep. *2*, 666–673.

Honma, S., Kawamoto, T., Takagi, Y., Fujimoto, K., Sato, F., Noshiro, M., Kato, Y., and Honma, K. (2002). Dec1 and Dec2 are regulators of the mammalian molecular clock. Nature *419*, 841–844.

Horsley, V., O'Carroll, D., Tooze, R., Ohinata, Y., Saitou, M., Obukhanych, T., Nussenzweig, M., Tarakhovsky, A., and Fuchs, E. (2006). Blimp1 defines a progenitor population that governs cellular input to the sebaceous gland. Cell *126*, 597–609.

Hsu, Y.-C., Pasolli, H.A., and Fuchs, E. (2011). Dynamics between stem cells, niche, and progeny in the hair follicle. Cell *144*, 92–105.

Hsu, Y.-C., Li, L., and Fuchs, E. (2014). Emerging interactions between skin stem cells and their niches. Nat. Med. *20*, 847–856.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnerberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. Nat. Methods *11*, 163–166.

Ito, M., Liu, Y., Yang, Z., Nguyen, J., Liang, F., Morris, R.J., and Cotsarelis, G. (2005). Stem cells in the hair follicle bulge contribute to wound repair but not to homeostasis of the epidermis. Nat. Med. *11*, 1351–1354.

Jaks, V., Barker, N., Kasper, M., van Es, J.H., Snippert, H.J., Clevers, H., and Toftgård, R. (2008). Lgr5 marks cycling, yet long-lived, hair follicle stem cells. Nat. Genet. *40*, 1291–1299.

Jaks, V., Kasper, M., and Toftgård, R. (2010). The hair follicle—a stem cell zoo. Exp. Cell Res. *316*, 1422–1428.

Janich, P., Pascual, G., Merlos-Suárez, A., Batlle, E., Ripperger, J., Albrecht, U., Cheng, H.-Y.M., Obrietan, K., Di Croce, L., and Benitah, S.A. (2011). The circadian molecular clock creates epidermal stem cell heterogeneity. Nature *480*, 209–214.

Jensen, K.B., and Watt, F.M. (2006). Single-cell expression profiling of human epidermal stem and transit-amplifying cells: Lrig1 is a regulator of stem cell quiescence. Proc. Natl. Acad. Sci. USA *103*, 11958–11963.

Jensen, K.B., Collins, C.A., Nascimento, E., Tan, D.W., Frye, M., Itami, S., and Watt, F.M. (2009). Lrig1 expression defines a distinct multipotent stem cell population in mammalian epidermis. Cell Stem Cell *4*, 427–439.

Kasper, M., Jaks, V., Are, A., Bergström, Å., Schwäger, A., Svärd, J., Teglund, S., Barker, N., and Toftgård, R. (2011). Wounding enhances epidermal tumorigenesis by recruiting hair follicle keratinocytes. Proc. Natl. Acad. Sci. USA *108*, 4099–4104.

Kaufman, C.K., Zhou, P., Pasolli, H.A., Rendl, M., Bolotin, D., Lim, K.-C., Dai, X., Alegre, M.-L., and Fuchs, E. (2003). GATA-3: An unexpected regulator of cell lineage determination in skin. Genes Dev. *17*, 2108–2122.

Kretzschmar, K., and Watt, F.M. (2014). Markers of epidermal stem cell subpopulations in adult mammalian skin. Cold Spring Harb. Perspect. Med. *4*, a013631.

Kretzschmar, K., Cottle, D.L., Donati, G., Chiang, M.-F., Quist, S.R., Gollnick, H.P., Natsuga, K., Lin, K.-I., and Watt, F.M. (2014). BLIMP1 is required for postnatal epidermal homeostasis but does not define a sebaceous gland progenitor under steady-state conditions. Stem Cell Reports *3*, 620–633.

Levy, V., Lindon, C., Harfe, B.D., and Morgan, B.A. (2005). Distinct stem cell populations regenerate the follicle and interfollicular epidermis. Dev. Cell *9*, 855–861.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell *161*, 1202–1214.

Magwene, P.M., Lizardi, P., and Kim, J. (2003). Reconstructing the temporal ordering of biological samples using microarray data. Bioinformatics *19*, 842–850.

Mascré, G., Dekoninck, S., Drogat, B., Youssef, K.K., Broheé, S., Sotiropoulou, P.A., Simons, B.D., and Blanpain, C. (2012). Distinct contribution of stem and progenitor cells to epidermal maintenance. Nature *489*, 257–262.

Matsumura, H., Mohri, Y., Binh, N.T., Morinaga, H., Fukuda, M., Ito, M., Kurata, S., Hoeijmakers, J., and Nishimura, E.K. (2016). Hair follicle aging is driven by transepidermal elimination of stem cells via COL17A1 proteolysis. Science *351*, aad4395–aad4395.

Mesa, K.R., Rompolas, P., Zito, G., Myung, P., Sun, T.Y., Brown, S., Gonzalez, D.G., Blagoev, K.B., Haberman, A.M., and Greco, V. (2015). Niche-induced cell death and epithelial phagocytosis regulate hair follicle stem cell pool. Nature *522*, 94–97.

Mlacki, M., Darido, C., Jane, S.M., and Wilanowski, T. (2014). Loss of Grainy head-like 1 is associated with disruption of the epidermal barrier and squamous cell carcinoma of the skin. PLoS ONE *9*, e89247.

Morris, R.J., Liu, Y., Marles, L., Yang, Z., Trempus, C., Li, S., Lin, J.S., Sawicki, J.A., and Cotsarelis, G. (2004). Capturing and profiling adult hair follicle stem cells. Nat. Biotechnol. *22*, 411–417.

Müller-Röver, S., Handjiski, B., van der Veen, C., Eichmüller, S., Foitzik, K., McKay, I.A., Stenn, K.S., and Paus, R. (2001). A comprehensive guide for the accurate classification of murine hair follicles in distinct hair cycle stages. J. Invest. Dermatol. *117*, 3–15.

Niemann, C., and Horsley, V. (2012). Development and homeostasis of the sebaceous gland. Semin. Cell Dev. Biol. *23*, 928–936.

Niemann, C., and Watt, F.M. (2002). Designer skin: Lineage commitment in postnatal epidermis. Trends Cell Biol. *12*, 185–192.

Nijhof, J.G.W., Braun, K.M., Giangreco, A., van Pelt, C., Kawamoto, H., Boyd, R.L., Willemze, R., Mullenders, L.H., Watt, F.M., de Gruijl, F.R., and van Ewijk, W. (2006). The cell-surface marker MTS24 identifies a novel population of follicular keratinocytes with characteristics of progenitor cells. Development *133*, 3027–3037.

Page, M.E., Lombard, P., Ng, F., Göttgens, B., and Jensen, K.B. (2013). The epidermis comprises autonomous compartments maintained by distinct stem cell populations. Cell Stem Cell *13*, 471–482.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

Petersson, M., and Niemann, C. (2012). Stem cell dynamics and heterogeneity: Implications for epidermal regeneration and skin cancer. Curr. Med. Chem. *19*, 5984–5992.

Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods *10*, 1096–1098.

Rompolas, P., and Greco, V. (2014). Stem cell dynamics in the hair follicle niche. Semin. Cell Dev. Biol. *25-26*, 34–42.

Rompolas, P., Mesa, K.R., and Greco, V. (2013). Spatial organization within a niche as a determinant of stem-cell fate. Nature *502*, 513–518.

Sandberg, R. (2014). Entering the era of single-cell transcriptomics in biology and medicine. Nat. Methods *11*, 22–24.

Sato, F., Kawamoto, T., Fujimoto, K., Noshiro, M., Honda, K.K., Honma, S., Honma, K., and Kato, Y. (2004). Functional analysis of the basic helix-loop-helix transcription factor DEC1 in circadian regulation. Interaction with BMAL1. Eur. J. Biochem. *271*, 4409–4419.

Schepeler, T., Page, M.E., and Jensen, K.B. (2014). Heterogeneity and plasticity of epidermal stem cells. Development *141*, 2559–2567.

Schult, D.A., and Swart, P.J. (2008). Exploring network structure, dynamics, and function using NetworkX. Proceedings of the 7th Python in Science Conference (SciPy2008).

Snippert, H.J., Haegebarth, A., Kasper, M., Jaks, V., van Es, J.H., Barker, N., van de Wetering, M., van den Born, M., Begthel, H., Vries, R.G., et al. (2010). Lgr6 marks stem cells in the hair follicle that generate all cell lineages of the skin. Science *327*, 1385–1389.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

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

Tan, D.W.M., Jensen, K.B., Trotter, M.W.B., Connelly, J.T., Broad, S., and Watt, F.M. (2013). Single-cell gene expression profiling reveals functional heterogeneity of undifferentiated human epidermal cells. Development *140*, 1433–1444.

Toufighi, K., Yang, J.-S., Luis, N.M., Aznar Benitah, S., Lehner, B., Serrano, L., and Kiel, C. (2015). Dissecting the calcium-induced differentiation of human primary keratinocytes stem cells by integrative and structural network analyses. PLoS Comput. Biol. *11*, e1004256.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.

Tumbar, T., Guasch, G., Greco, V., Blanpain, C., Lowry, W.E., Rendl, M., and Fuchs, E. (2004). Defining the epithelial stem cell niche in skin. Science *303*, 359–363.

Van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. J. Mach. Learn. Res. *9*, 2579–2605.

Veniaminova, N.A., Vagnozzi, A.N., Kopinke, D., Do, T.T., Murtaugh, L.C., Maillard, I., Dlugosz, A.A., Reiter, J.F., and Wong, S.Y. (2013). Keratin 79 identifies a novel population of migratory epithelial cells that initiates hair canal morphogenesis and regeneration. Development *140*, 4870–4880.

Wang, X., Pasolli, H.A., Williams, T., and Fuchs, E. (2008). AP-2 factors act in concert with Notch to orchestrate terminal differentiation in skin epidermis. J. Cell Biol. *183*, 37–48.

Yee, T.W. (2010). The VGAM package for categorical data analysis. J. Stat. Softw. *32*, 1–34.

Zeeuwen, P.L.J.M., van Vlijmen-Willems, I.M.J.J., Hendriks, W., Merkx, G.F.M., and Schalkwijk, J. (2002). A null mutation in the cystatin M/E gene of ichq mice causes juvenile lethality and defects in epidermal cornification. Hum. Mol. Genet. *11*, 2867–2875.

Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science *347*, 1138–1142.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rat monoclonal anti-CD3	BioLegend	Cat#100201
Rat monoclonal anti-CD34	eBioscience	Cat#14-0341
Rat monoclonal anti-CD207	eBioscience	Cat#14-2073
Goat polyclonal anti-COX-1 (PTGS1)	Santa Cruz	Cat#sc-1754; RRID: AB_2245319
Rabbit polyclonal anti-EGFP	Thermo Fisher	Cat#A-11122; RRID: AB_2576216
Rabbit polyclonal anti-KI67	Novocastra	Cat#NCL-Ki67p
Goat polyclonal anti-KLK10	Santa Cruz	Cat#sc-20386
Rabbit polyclonal anti-KRT6	Covance	Cat#PRB-169P; RRID: AB_10063923
Rabbit polyclonal anti-KRT10	Covance	Cat#PRB-159P; RRID: AB_291580
Rabbit polyclonal anti-KRT14	Covance	Cat#PRB-155P; RRID: AB_292096
Mouse monoclonal anti-KRT15	Abcam	Cat#ab2414
Rabbit monoclonal anti-KRT17	Cell Signaling	Cat#4543
Goat polyclonal anti-KRT79	Santa Cruz	Cat#sc-243156
Rabbit polyclonal anti-LOR	Covance	Cat#PRB-145P
Goat polyclonal anti-MGST1	Santa Cruz	Cat#sc-17003; RRID: AB_2143472
FISH probes		
Cd34	Advanced Cell Diagnostics	Cat#319161-C2
Cst6	Advanced Cell Diagnostics	Cat#436181
Flg2	Advanced Cell Diagnostics	Cat#430131
Gli1	Advanced Cell Diagnostics	Cat#311001
Krt10	Advanced Cell Diagnostics	Cat#457901
Krt79	Advanced Cell Diagnostics	Cat#436201-C2
Lgr5	Advanced Cell Diagnostics	Cat#312171-C2
Lgr6	Advanced Cell Diagnostics	Cat#404961 / Cat#404961-C2
Lrig1	Advanced Cell Diagnostics	Cat#310521
Thbs1	Advanced Cell Diagnostics	Cat#457891
Postn	Advanced Cell Diagnostics	Cat#418581
Chemicals, Peptides, and Recombinant Proteins		
Agencourt AMPure XP	Beckman Coulter	Cat#A63880
Defined Keratinocyte-SFM (1X)	Thermo Fisher	Cat#10744019
DNase I Solution (1 mg/ml)	Stem Cell Technologies	Cat#07900
Dynabeads MyOne Streptavidin C1	Thermo Fisher	Cat#65001
Minimum Essential Medium Eagle -Spinner modification	Sigma-Aldrich	Cat#M8167
PvuI-HF	NEB	Cat#R3150S
Qiaquick Buffer PB	QIAGEN	Cat#19066
Trypsin solution from porcine pancreas	Sigma-Aldrich	Cat#T4424
Critical Commercial Assays		
Anti-Sca-1 MicroBead Kit (FITC), mouse	Miltenyi Biotec	Cat#130-092-529
C1 Single-Cell Auto Prep IFC for mRNA Seq (10 – 17 μm)	Fluidigm	Cat#100-6041
KAPA Library Quantification Kit	KAPA Biosystems	Cat#07960140001
RNAscope Fluorescent Multiplex Reagent Kit	Advanced Cell Diagnostics	Cat#320850

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Raw data files for RNA sequencing	NCBI GEO	GSE67602
Scripts and computational analysis workflow	Kasper Lab	https://github.com/kasperlab
Online tool for visualization of single-cell data	Kasper Lab Linnarsson Lab	http://kasperlab.org/tools http://linnarssonlab.org/epidermis/
Systematic staining catalog	Kasper Lab	http://kasperlab.org/data
Experimental Models: Organisms/Strains		
Mouse: C57BL/6J	Charles River	JAX: 000664
Mouse: *Lgr5-EGFP-Ires-CreERT2*	Jackson Laboratory	JAX: 008875
Software and Algorithms		
MSigDB	Subramanian et al., 2005	http://www.broadinstitute.org/gsea/msigdb/index.jsp
NetworkX	Schult and Swart, 2008	https://networkx.github.io/
scikit-learn	Pedregosa et al., 2011	http://scikit-learn.org/
VGAM	Yee, 2010	https://cran.r-project.org/web/packages/VGAM/index.html
μManager	Edelstein et al., 2014	http://micro-manager.org/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents or computational resources may be directed to, and will be fulfilled by the corresponding author Maria Kasper (maria.kasper@ki.se).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice

All experiments were performed on female C57BL/6 mice. The mice were fed ad libitum, and handled and housed under standard conditions in the animal facility of Karolinska University Hospital Huddinge. All mouse experiments were performed in accordance to Swedish legislation and approved by the Stockholm South Animal Ethics Committee. Mice were sacrificed in the second telogen and hair cycle stages were determined by staining dorsal skin sections for Ki67 as described previously (Greco et al., 2009; Müller-Röver et al., 2001). Mice that showed signs of early anagen were excluded from this analysis. Cells from n = 19 mice were included in the final dataset.

METHOD DETAILS

Cell Isolation

Full epidermal cells were isolated as described previously (Jaks et al., 2008). In brief, clipped and disinfected dorsal skin was isolated, dermal and adipose tissue was removed, and stripes of skin were floated on trypsin for 2 hr at 32°C. Epidermal tissue was subsequently scraped into S-MEM / 1% BSA and single cells were isolated by magnetic stirring at 120 rpm for 25 min / RT. The resulting cell suspension was filtered through 70 μm and 40 μm cell strainers, resuspended in Defined Keratinocyte Serum-free Medium without supplement (DK-SFM), and SCA-1+ and SCA-1− cells were separated using Anti-SCA-1-FITC magnetic beads according to the manufacturer's instructions. Cells were stored on ice in DK-SFM with 0.1 mg/ml DNase I until capturing. Before capturing, the cell suspension was carefully resuspended and two times passed through a 20 μm cell strainer.

From each experimental mouse, mid-dorsal skin pieces (ca. 0.5 × 0.5 cm) were paraffin-embedded for hair cycle staging and remapping of marker genes.

Cell Capturing, Quality Control, and Single-Cell cDNA Synthesis

Epidermal cells were captured on a medium microfluidic chip (designed for cells from 10 μm – 17 μm) using the Fluidigm C1 Autoprep System. 14 μl filtered cell suspension (~750 cells / μl in DK-SFM with DNase I) was mixed with 6 μl C1 Suspension Reagent and 14 μl were loaded onto the chip. Single-cells were then captured for 30 min at 4°C using the *"Cell Load (1772x/1773x)"* script. Capturing efficiency was evaluated on a Nikon TE2000E automated microscope and both bright field and SCA1-FITC images of every capturing position were taken using μManager. Before proceeding with the tagmentation step, each capture site was manually inspected and only capture sites containing single, healthy cells were processed.

Following the image acquisition, STRT-C1 Lysis, RT and PCR mix was added as previously described (Islam et al., 2014), and the *"RT + AMP (1772x/1773x)"* script was executed. After the cDNA synthesis had been finished (~8.5 h), the amplified cDNA was harvested with 13 μl Harvest Reagent and cDNA quality was measured on an Agilent BioAnalyzer.

Tagmentation and Isolation of 5′ fragments

The amplified cDNA was fragmented and barcoded using Tn5 DNA transposase ('tagmentation') as described previously (Islam et al., 2014). 100 μl Dynabeads MyOne Streptavidin C1 beads were washed in 2x BWT, resuspended in 2 ml 2x BWT, and 20 μl washed beads were added to each well. After 15 min incubation at room temperature, all wells were pooled, the beads were immobilized on a magnet, and the supernatant (containing all internal cDNA fragments) was removed. The beads were resuspended in 100 μl Tris-NaCl-Tween (TNT), washed once in 100 μl Qiaquick PB, and then washed twice in 100 μl TNT. The beads were subsequently incubated in 100 μl restriction mix (1x NEB CutSmart, 0.4 U/μl PvuI-HF enzyme) for 1 hr at 37°C to cleave 3′ fragments which carry a PvuI recognition site. Afterward, the beads were washed three times in TNT, then resuspended in 30 μl ddH2O and incubated for 10 min at 70°C to elute the DNA. To remove short fragments, AMPure beads were used at 1.8 x volume and eluted in 30 μl.

Illumina High-Throughput Sequencing and Processing of Sequencing Reads

The molar concentrations of the libraries were quantified with KAPA Library Quant qPCR and fragment lengths were determined using a reamplified (12 cycles) sample on a BioAnalyzer. Sequencing was performed on an Illumina HiSeq 2000 with C1-P1-PCR2 as read 1 primer and C1-TN5-U as index read primer. Reads of 50 bp as well as 8 bp index reads corresponding to the cell-specific barcodes were generated. Each read was expected to start with a 6 bp unique molecular identifier (UMI), followed by 3-5 guanines and the 5′ end of the mRNA. Reads were processed as described previously (Islam et al., 2014) except that we removed any mRNA molecule (i.e., UMI) supported by only a single read.

Yield and Quality of Sequencing

Sequencing yielded around 25 million mapped reads per C1 chip (793 million mapped reads and 26 million sequenced molecules in total) and around 0.55 million mapped reads per cell after quality control (Figures S1G – S1I). Each unique mRNA molecule was detected 18 times on average during the sequencing indicating sufficient sequencing depth (Figures S1J – S1K). Measurement of RNA spike-in standards indicates strong uniformity between experiments and a sequencing efficiency of 20 - 30 % (Figures S1L – S1N).

Systematic Staining of All Populations by Immunohistochemistry and Single Molecule FISH

The existence and spatial location of the 25 populations and subpopulations defined during 1st and 2nd level clustering were confirmed and determined by antibody staining and/or single-molecule mRNA FISH (FISH) (see Table S7). One subpopulation (uHF III) could not be shown via positive marker staining because this population did not express unique genes in comparison to the other populations, but it formed its own cluster due to the lack of genes. Since all other 24 clusters of cells could be verified, we expect that this population represents a true population and is likely positioned in the SG canal (placed by staining exclusion). The following antibody dilutions were used: CD3 (1:100), CD34 (1:50), CD207 (1:50), COX-1 (PTGS1) (1:50), EGFP (1:500), Ki67 (1:2000), KLK10 (1:50), KRT6 (1:250), KRT10 (1:250), KRT14 (1:250), KRT15 (1:50), KRT17 (1:100), KRT79 (1:50), LOR (1:200), MGST1 (1:50). *Cd34, Cst6, Flg2, Gli1, Krt10, Krt79, Lgr5, Lgr6, Lrig1, Thbs1*, and *Postn* mRNA were visualized by FISH using the RNAscope Fluorescent Multiplex Kit (Advanced Cell Diagnostics, Inc.) according to the manufacturers instructions. Please note that the used FISH protocol was in our hands less sensitive compared to our single-cell RNA-seq data and thus for lower expressed genes only few dots can be expected. According to our negative controls, and the manufacturers description, approx. one false positive signal can occur in one out of 10 cells.

Both, antibody and FISH stainings were performed on formalin-fixed, paraffin-embedded (FFPE) sections of dorsal skin isolated from the same animals that were used for the single-cell sequencing. The only exception was staining for anti-EGFP, which was performed on dorsal skin of 8 week old *Lgr5-EGFP-Ires-CreERT2* mice using horizontal whole mount staining (Füllgrabe et al., 2015). Images were acquired on either a LSM710-NLO confocal microscope (Zeiss) or a Nikon A1R confocal microscope.

QUANTIFICATION AND STATISTICAL ANALYSIS

Analysis and Visualization of Processed Sequencing Data

The following section describes the data analysis approach employed in this study both in general terms (1-7) and with specific details referring to distinct steps in the analysis process (8). To ensure complete transparency and facilitate reproduction, the complete code used in this study is available online (see Key Resources Table).

(1) Implementation

Analysis and visualization of data were performed in a Python environment built on the NumPy, SciPy, matplotlib, and pandas libraries. Affinity propagation and t-SNE used implementations available in the scikit-learn package (Pedregosa et al., 2011). Graphs were drawn using the NetworkX package (Schult and Swart, 2008). Cubic spline smoothing and likelihood ratio tests were performed using the VGAM package (Yee, 2010), which was accessed via Rpy2. The custom made scripts used for this analysis are available online (see Key Resources Table).

(2) Unsupervised Clustering Using Affinity Propagation

(a) Feature Selection

To filter out genes before affinity propagation (AP) clustering, all genes with an average expression below a specified cut-off and/or those with less than five highly correlated neighbors were excluded. Two genes were defined as highly correlated if their correlation value (Pearson r) was within the top 5% of all gene-gene correlation values within the whole dataset. The remaining genes were used to fit a noise model as

$$log_2(CV) = log_2(mean^{\alpha} + k),$$

where *CV* is a gene's coefficient of variation and *mean* its average. The 2,500 genes that showed the largest difference between observed CV and CV as predicted by the noise model were used as features for AP clustering.

(b) Affinity Propagation Clustering

Cell populations were defined using AP, a recently introduced approach for unsupervised clustering (Frey and Dueck, 2007). To ensure robustness toward differences in total gene expression between cells, Pearson correlation of log2-transformed data was used as distance metric for the clustering. To facilitate the visualization of clustered data as heatmaps and barplots, the cells / genes within the AP-defined clusters were brought into one-dimensional order based on Ward's linkage. While mathematical aspects such as the highest possible reduction of variance within clusters were taken into consideration when selecting the clustering parameters *preference* and *damping*, parameter choice was mainly based on subjective measures of clustering performance.

(c) Evaluation of Clustering Robustness

To evaluate robustness of AP clustering, a resampling approach was used, where 25% of cells were removed from the dataset at random. The remaining cells were reclustered using the same parameters as for the main clustering and the percentage of cells in each defined group that remain clustered together was determined. In order to measure the background distribution (i.e., the percentage of cells which remain together by pure chance), the group labels were randomly permutated. Both the resampling and the label permutation were repeated 100 times.

(3) Nonlinear Dimensionality Reduction with t-Distributed Stochastic Neighbor Embedding

Dimensionality reduction to two dimensions for visualization purposes and as input for pseudotemporal/-spatial ordering was performed using t-distributed stochastic neighbor embedding (t-SNE) (Van der Maaten and Hinton, 2008). In most cases, a *perplexity* value between 20 and 25, an *early exaggeration* value of 2.0 – 3.0 and a *learning rate* of 1,000 were used.

(4) Negative Binomial Regression of Gene Expression

(a) Model Description

To assign expression of a gene to a cell population, a Bayesian general linear model (GLM) was used as described elsewhere (Zeisel et al., 2015). In such a model, it is assumed that the outcome (i.e., the measured expression of a gene in a population) is sampled from a distribution whose mean is determined by a linear combination of K predictors x_i with coefficients β_i. Therefore,

$$\mu = \sum_{k=1}^{K} \beta_k x_k (k \in [1, K])$$

For each cell, the outcome and predictors are known and we aim to determine the values of the coefficients.

As predictors, we use a *Baseline* predictor and a binary *Cell Type* predictor. As we expect every gene to have a baseline expression proportional to the total number of expressed molecules within a particular cell, the *Baseline* predictor value is set as a cell's molecule count normalized to the average molecule count of all cells. Meanwhile, the *Cell Type* predictor is set to 1 if a cell is included in a particular cell population cluster or a pseudospace / pseudotime bin. In consequence, the coefficient β_k for a *Cell Type* predictor x_k represents the additional number of molecules of a particular gene that are present if a cell is member of a particular cell type.

As real count data is usually overdispersed when compared to an ideal Poisson distribution, we used a negative binomial distribution, which can be represented as a Gamma distribution of Poisson distributions, for our model. Therefore, if y is the observed count,

$$y \sim Poisson(\lambda)$$

$$\lambda \sim Gamma(a, b)$$

with mean $\mu = ab$ and standard deviation $\sigma = \sqrt{(ab/1+b)}(1+b)$.

As the standard deviation roughly scales as the square root of the mean, it can be described as $\sigma = r\sqrt{\mu}$ with overdispersion factor r. Hence,

$$a = \frac{\mu}{r^2 - 1}$$

$$b = r^2 - 1.$$

By attaching prior distributions to the overdispersion factor r and the coefficients β_k, we acquire a full Bayesian negative binomial regression model, with

$$\mu = \sum_{k=1}^{K} \beta_k x_k$$

$$y \mid \lambda \sim Poisson(\lambda)$$

$$\lambda \mid \mu, r \sim Gamma\left(\frac{\mu}{r^2 - 1}, r^2 - 1\right)$$

$$r \sim Cauchy(0, 1)$$

$$\beta_k = Pareto(0, 1.5).$$

The model was implemented in STAN. A more detailed explanation of the model is provided elsewhere (Zeisel et al., 2015).

(b) Calling Genes That Are Specifically or Uniquely Expressed in Groups / Predictors

To define whether a gene can be considered *specifically expressed* in a particular cell population, we compared the posterior probability distributions of the *Baseline* coefficient and the *Cell Type* coefficient. A gene was considered activated in a cell population if its class-specific coefficient exceeded the *Baseline* coefficient with a specified posterior probability. In order to be defined as *uniquely expressed* in a particular cell population, a gene's *Cell Type* coefficient had to exceed all other *Cell Type* coefficients as well as the *Baseline* coefficient with a specified posterior probability. The posterior probability cut-off at which genes were considered specifically or uniquely expressed was set at 99.9% for the regression model of the 1st level clustering and to 95% for all other regression models.

(c) Evaluating the Exploratory Quality of Regression Models

In order to evaluate how well a regression model explains the data, a simulated dataset was sampled from the model and compared to the observed data. In particular, for every gene and predictor x_k in the model, values were randomly sampled one hundred times from the posterior probability distribution of each coefficient β_k and subsequently multiplied with the predictor matrix used as input for the model. The resulting dataset contains the simulated expression data of g genes in m cells over K predictors. These data were subsequently summarized including either all or a subset of predictors and compared to the observed data. For each gene, the number of 'explained' (molecules both found in the observed and the simulated data), 'underexplained' (molecules found in the observed but not the simulated data) and 'overexplained' (molecules found in the simulated but not the observed data) molecules was determined. Data-model comparison occurred either on a single-cell level, a group level (for each gene, the number of molecules in the observed and simulated data were pooled between all cells within a group, thus averaging in-group noise) or a whole-dataset level (for each gene, the number of molecules in the observed and simulated data were pooled between all cells in the dataset).

(5) Pseudotemporal/-Spatial Ordering of Cells

(a) Bringing Cells into Pseudotemporal/-Spatial Order

Spatial and temporal ordering is based on the same analytical method and only distinguished by the input of cells (differentiating cells of the IFE for pseudotime; basal cells of HF and IFE for pseudospace). The pseudotemporal/-spatial ordering of IFE/basal cells is following a graph-based approach that was recently introduced by Magwene et al., 2003 and Trapnell et al., 2014. In brief, a minimum spanning tree (MST) is constructed between cells, which are defined by their position in – dimensionality-reduced – space. The longest path through the MST, called the diameter path, is subsequently defined and a PQ tree encoding all paths through the graph (or orderings of cells) under the constraints of the diameter path is constructed. The PQ tree is subsequently screened for orderings of cells that minimize the total traveling distance. While we generally follow the approach introduced by Trapnell et al., 2014 we diverge in several points. Since linear dimensionality reduction approaches such as PCA or ICA were insufficient to resolve and visualize the differentiation and spatial trajectories in the dataset, we used the nonlinear t-SNE method for dimensionality reduction and construction of the MST. Due to the high number of single cells included in our analysis (536 IFE cells and 486 basal cells) and due to a relative high level of noise, we furthermore did not consider all permutation emitted from the PQ. Instead, we restricted the number of orderings based on local optima derived from subsets of the graph.

(b) Testing the Robustness of Pseudotemporal or Pseudospatial Ordering

To test the robustness of the pseudotemporal/-spatial ordering, we (1) compared the results to orderings gained without any dimensionality reduction and (2) employed a resampling approach. During the resampling, we either compared the results of one hundred orderings gained from different initial t-SNE plots to our initial results to evaluate robustness against randomness in the dimensionality reduction or we randomly discarded 25% of cells from the dataset for one hundred times and compared the resulting ordering to our initial results to test for robustness against small changes in composition of the dataset. As negative control, we randomly shuffled cell labels.

(c) Modeling Gene Expression over Pseudospace/-Time and Calling Pseudospace/-Time-Dependent Genes

To model gene expression changes in dependency of pseudotime or pseudospace, a cubic smoothing spline with five effective degrees of freedom was fitted to the ordered expression data of all genes in the IFE or basal dataset which showed an average

expression > 0.1 molecules. Pseudospace/-time dependency of gene expression was subsequently tested by comparing the spline-smoothed model to a pseudospace/-time-independent restricted model using the approximate likelihood ratio test. We considered all genes with a p-value below the Bonferroni-corrected significance level $\alpha = 0.001$ to be pseudotime- or pseudospace-dependent. To visualize the expression patterns of all pseudotime- or pseudospace dependent genes and to perform gene set enrichment analysis, spline smoothed gene expression data was clustered using AP as described above. Genes within each cluster were ordered according to expression peak or onset of induction (defined as point in pseudospace/pseudotime where the expression of a gene exceeds 50% of the peak expression).

(d) Positioning Cells in Pseudospace/-Time

To link single cells not included in the model to a specific place in pseudotime or pseudospace, the expression data of g pseudospace/-time dependent genes in a particular cell M is correlated to all points in the fitted model (which contains the spline-fitted expression data of g pseudospace/-timespace-dependent genes over t points in pseudospace/-time) and the point with the highest Pearson r is returned.

To evaluate how well a particular cell or group of cells fits a pseudospace/-time model, we used several qualitative and quantitative approaches: on the one hand, we analyzed how many pseudospace/-time-dependent genes are expressed in a particular group of cells. We reasoned that a group of cells which exhibits e.g., features of a certain differentiation stage will express a high number of genes linked to this particular stage. On the other hand, we consider the p-value of the best fitting cell-to-point correlation a quantitative measure of fit. Furthermore, we employed a resampling approach to test the robustness of the correlation. In this approach, we randomly removed 75% of pseudotime- or pseudospace-dependent genes from the dataset for one hundred times and subsequently correlated each single cell to a specific point on the axis as described above. We then measured the average distance of the correlation points yielded from the reduced dataset to the correlation gained with the full dataset. We reasoned that cells which have a strong pseudotime-/pseudospace signature will be more robust against the resampling of the dataset and will thus show a narrower spread of correlation points.

(6) Constructing Gene-Gene Neighbor Networks

To construct networks of pseudotime- and pseudospace-dependent genes, we used a shared nearest neighbor approach in combination with the previously described context likelihood of relatedness (CLR) algorithm (Faith et al., 2007). Specifically, we initially generated a gene-gene correlation matrix between all selected genes and subsequently used CLR to transform the correlation values based on their network context. For each gene, we then selected the n nearest neighbors. We considered two genes to be linked within the neighbor context if they shared a number $\geq k$ of nearest neighbors. Graphs were drawn using a force-directed spring layout with each node representing a gene and each edge connecting two interlinked genes.

In the pseudotime- and pseudospace-gene networks, two genes were considered linked if they shared at least 5 of 25 nearest neighbors. In the basal gene network, two genes were considered linked if they shared 10 or more of 25 nearest neighbors.

(7) Gene Set Enrichment Analysis

To link gene lists – for instance pseudotime- or pseudospace-dependent genes at particular stages – to potential biological roles, we queried the Molecular Signatures Database MSigDB using the 'Investigate Gene Sets' function (Subramanian et al., 2005). We only considered gene sets included in the *CP*, *CP:BIOCARTA*, *CP:KEGG*, *CP:REACTOME*, and *BP* categories of the dataset and excluded all matches with an FDR q-value ≥ 0.05. To avoid redundancies, the usually five reported gene sets were selected among the 20 most significant matches.

(8) Data Analysis Process

(a) Selection of Cells

Cells with less than 2,000 unique molecules were removed from the dataset, leaving 1,422 cells passing the quality criteria.

(b) 1st Level Clustering – AP Clustering

For the 1st level clustering, 2,500 features were selected as described in (2) using a mean expression cut-off of 0.05 molecules over the whole dataset (1,422 cells). Gene-gene and cell-cell Pearson distances were subsequently calculated and used as input for AP clustering. To achieve a better resolution of cell populations, gene clusters linked to ribosomal, housekeeping and intermediate early genes (IEGs) were removed after an initial round of clustering along the gene axis. In summary, 13 distinct cell populations could be defined during 1st level clustering. Clustering robustness was evaluated as described in (2). Additionally, the AP clustering approach was compared with unsupervised clustering by backSPIN (Zeisel et al., 2015) with good agreement. A t-SNE representation of the whole dataset was generated with the same features as used for the AP clustering.

(c) 1st Level Clustering – Negative Binomial Regression

A negative binomial regression model was generated as described in (4) using the 1st level clusters as predictors. The regression was performed on all genes with an average molecule count ≥ 0.25 over either the whole dataset or within at least one cluster (9,016 genes). Group-specific or –unique genes were called using a 99.9% posterior probability cut-off.

(d) 2nd Level Clustering – Cell Selection

2nd level clustering was performed separately on subsets of cells showing inner bulge (IB), outer bulge (OB), upper HF (uHF), or IFE basal (IFE B) signatures. Signature genes were identified from the 1st level clustering negative regression model: (1) as genes, which are only expressed over *Baseline* in either the IB, OB, uHF, or IFE B cluster(s), or (2) as genes, whose expression in one of these

clusters exceeds the expression in all other clusters with 99.9% posterior probability. Following the identification of signature genes, the cumulative expression of the four different signatures was calculated for every cell in the dataset and cut-offs defining whether or not a single cell expresses a certain signature were specified. To avoid duplication of cells with more than one signature, cells were assigned to the four groups in the following order of primacy: IB > OB > uHF > IFE B. In this way, 87 IB, 273 OB, 364 uHF and 322 IFE B cells (from 630 IFE cells) were defined.

(e) 2nd Level Clustering – AP Clustering

From each of the four subsets of the data, features were selected as described in (2) using a mean expression cut-off of 0.1 molecules and genes linked to ribosomal, housekeeping and IEG clusters in the 1st level clustering were removed. Due to the considerably lower signal-to-noise ratios expected in the subpopulations, the selected genes were subjected to a first round of AP clustering and only clusters of genes that exhibited a strong and coordinated differential expression pattern were used as features for the final clustering of cells. Using this approach, three, seven, five, and three subclusters of cells were identified in the IB, uHF, OB, and IFE B data respectively. Clustering robustness was measured as described in (2).

(f) 2nd Level Clustering – Negative Binomial Regression

To perform negative binomial regression on the 2nd level clustering data while still considering the whole dataset, each cell assigned to the IB, OB, uHF or IFE B subset of the data was grouped according to its 2nd level cluster identity. All remaining cells (e.g., the immune cells or the cells of the IFE differentiation process which did not show an IFE B or IB/OB/uHF signature) were grouped according to 1st level cluster membership. The combination of the 2nd and 1st level clustering data allowed regression with 25 *Cell Type* predictors. The regression was performed on all genes with an average molecule count ≥ 0.25 over either the whole dataset or within at least one cluster (9,784 genes). Group-specific or –unique genes were called using a 95% posterior probability cut-off.

(g) 1st and 2nd Level Clustering – Robustness towards Replication

To ensure that none of the cell populations defined during 1st and 2nd level clustering is the mere result of an experimental or technical artifact, the robustness of each cluster toward biological replication was analyzed. To this end, the number of cells in each cluster, the ratio of cells from SCA-1+ and SCA-1− fractions and the number of experimental mice from which the cells in each cluster were derived was calculated and compared to the number of mice expected by pure chance. To acquire the expected value of mice for a cell population, n_{SCA1+} / n_{SCA1-} cells corresponding to the number of SCA-1+ and SCA-1− cells in the population were randomly sampled from the SCA-1+ and SCA-1− dataset and the total number of mice from which the sampled cells were derived was subsequently calculated. For each population, this sampling was repeated 10,000 times and a p-value was returned.

Population	SCA-1+ Fraction	Number of Cells	Number of Mice	Number of Mice if Random	p-value
IFE B I	91.5 %	94 / 1422	10 / 19	13.26	0.0048
IFE B II	85.8 %	134 / 1422	14 / 19	16.19	0.0703
INFU B	48.9 %	94 / 1422	18 / 19	18.38	0.4925
IFE D I	45.0 %	140 / 1422	19 / 19	18.83	1
IFE D II	30.9 %	97 / 1422	19 / 19	18.65	1
IFE K I	21.1 %	57 / 1422	15 / 19	17.63	0.0249
IFE K II	35.7 %	14 / 1422	11 / 19	9.91	0.9014
uHF I	9.1 %	33 / 1422	13 / 19	14.98	0.1343
uHF II	11.1 %	36 / 1422	15 / 19	15.50	0.4892
uHF III	13.3 %	45 / 1422	14 / 19	16.60	0.0438
uHF IV	23.4 %	111 / 1422	19 / 19	18.76	1
uHF V	15.2 %	79 / 1422	18 / 19	18.23	0.5875
uHF VI	10.8 %	37 / 1422	13 / 19	15.63	0.053
uHF VII	13.0 %	23 / 1422	11 / 19	13.01	0.1333
SG	5.3 %	19 / 1422	8 / 19	11.54	0.0127
OB I	10.5 %	105 / 1422	17 / 19	18.47	0.0583
OB II	9.8 %	51 / 1422	16 / 19	16.91	0.3339
OB III	4.9 %	41 / 1422	17 / 19	15.82	0.9194
OB IV	6.5 %	46 / 1422	16 / 19	16.37	0.5234
OB V	6.7 %	30 / 1422	15 / 19	14.34	0.7982
IB I	7.4 %	54 / 1422	17 / 19	17.03	0.6533
IB II	15.8 %	19 / 1422	9 / 19	11.84	0.0414
IB III	0.0 %	14 / 1422	9 / 19	9.49	0.5027
TC	5.6 %	18 / 1422	9 / 19	11.23	0.0952
LH	9.7 %	31 / 1422	14 / 19	14.66	0.445

(h) Modeling of IFE Differentiation

To model IFE differentiation, all cells belonging to the non-infundibulum IFE basal clusters (IFE BI and IFE BII) or the remaining IFE cells identified in the 1st level clustering were considered (536 cells). Features were selected as described in (2) using a mean expression cut-off of 0.1 molecules and genes linked to ribosomal, housekeeping and IEG clusters in the 1st level clustering were removed. The remaining features were used as input for t-SNE (*perplexity* = 25, *early exaggeration* = 2.0) and the cells were brought into pseudotemporal order as described in (5). Cubic splines were fitted to the expression of 7,354 genes (mean expression $\geq$0.1 molecules), 1,627 significantly pseudotime-dependent genes were identified and subsequently AP clustered into eight subgroups. All cells from the dataset were correlated to the differentiation trajectory and the robustness of the pseudotemporal ordering and the correlation was evaluated as described above.

(i) Modeling of uHF Differentiation

To test whether the differentiation process follows similar lines in different compartments of the epidermis, pseudotemporal ordering of uHF cells was performed. For this, all non-SG (opening) uHF cells (uHF IV – VII, 250 cells) were used. Features were selected as described in (g). In contrast to (g), an initial round of dimensionality reduction (TruncatedSVD, 5 dimensions) was necessary to get a good t-SNE representation of the data (*perplexity* = 100, *early exaggeration* = 2.0). After pseudotemporal ordering and cubic spline fitting, 1,068 significantly pseudotime-dependent genes could be defined.

(j) Modeling of gene Expression Changes Along the Proximal-Distal Spatial Axis

In order to model spatial gene expression changes along the proximal-distal axis without interference from differentiation signatures, only cells from IFE and HF which show a clear basal signature were selected. Cells from the HF (uHF IV – VII, OB I – V, IB I – III*) were considered basal if they were linked to a pseudotime position $\leq$300. Due to the early onset of differentiation in the IFE basal compartment, IFE cells were selected with a more stringent cut-off ($\leq$150). In sum, 486 cells were classified as basal. Features were selected as described in (2) using a mean expression cut-off of 0.1 molecules and genes linked to ribosomal, housekeeping and IEG clusters in the 1st level clustering were removed. To make sure that no differentiation related modules of genes are included in the dataset, the genes were subjected to one round of AP clustering and only clusters not containing typical differentiation markers (e.g., *Mt4* or *Krt10*) were included. Only the genes that passed this additional cycle of quality control were used as input for t-SNE (*perplexity* = 20, *early exaggeration* = 3.0) and the basal cells were subsequently brought into pseudospatial order as described in (5). Cubic splines were fitted to the expression of 6,788 genes (mean expression $\geq$0.1 molecules), 547 significantly pseudospace-dependent genes were identified and subsequently AP clustered into eight subgroups. All cells from the dataset were correlated to the spatial axis and the robustness of the pseudospatial ordering and of the correlation was evaluated as described above.

* Although the cells of the inner bulge population IB I do not seem to show any distinct differentiation signatures, cells from IB I were considered in this model if under the set cut-off.

(k) Pseudospacetime – Creation

To link every cell to its position in two-dimensional space along the differentiation and spatial axes without interference from ambiguous genes, only genes, which were either uniquely pseudotime- (1,409 genes) or pseudospace-dependent (329 genes), were considered and correlation of all cells to both axes was recalculated using only the selected genes. Cells and cell populations which do not seem to fit to any position on either the pseudospace-, the pseudotime- or both axes (e.g., the immune or sebaceous gland cells, see (5)) were subsequently (partially) removed from the pseudospace.

(l) Pseudospacetime – Negative Binomial Regression

To perform negative binomial regression of the data under the constraints of the pseudospacetime model, both the pseudospace- and pseudotime-axis were divided into 15 equally sized bins and each pseudospace-/pseudotime-bin was considered a predictor in the regression model. Furthermore, additional predictors (sebaceous gland, sebaceous gland opening, pan-immune, T-cell and Langerhans cell) were generated for genetic signatures that cannot be explained by the pseudospacetime model. Regression was performed on the same set of 9,784 genes as selected in (f). Predictor-specific or –unique genes were called using a 95% posterior probability cut-off.

As a negative control, predictor identity was randomly shuffled between cells and the regression was performed as described above.

(m) Pseudospacetime – Model Comparison

To evaluate the explanatory quality of the pseudospacetime model, a simulated dataset was sampled from the traces of the negative regression model as described in (4) and subsequently compared to the observed data. To ensure comparability of the pseudospacetime model with the 1st and 2nd level clustering, only genes used consistently in the pseudospacetime, the 1st level, and the 2nd level regression were considered (6,949 genes).

(n) Stem Cell Analysis – Cell Selection

To select cells, which express the stem cell/progenitor markers *Lgr5*, *Cd34*, *Gli1*, *Lgr6*, *Lrig1*, and *Krt14* above *Baseline*, the following cut-offs were chosen:

Marker	Cut-off Selection	Cut-Off Value (Molecules)	Number of Positive Cells
Lgr5	Maximal *Baseline* value predicted during 2nd level clustering regression multiplied by 10	0.34	138
Cd34	Maximal *Baseline* value predicted during 2nd level clustering regression multiplied by 10	0.85	297
Gli1	Maximal *Baseline* value predicted during 2nd level clustering regression multiplied by 10	0.23	84
Lgr6	Maximal *Baseline* value predicted during 2nd level clustering regression multiplied by 10	0.26	75
Lrig1	Maximal *Baseline* value predicted during 2nd level clustering regression multiplied by 5	1.95	207
Krt14	Maximal *Baseline* value predicted during 2nd level clustering regression multiplied by 10	35.18	149

(o) Stem Cell Analysis – AP Clustering

AP clustering was performed separately on all cells expressing a certain stem cell marker using the same approach as described for the 2nd level clustering in (e).

(p) Stem Cell Analysis – Basal Cell Clustering and t-SNE

To compare basal stem cells to each other and to basal cells, which do not express stem cell markers, all cells from IFE, uHF (uHF IV – VII) and OB with a pseudotime position $\leq$300 were selected. In contrast to the cell selection described in (j), IFE cells were selected less stringently, inner bulge cells were not considered and ambiguous genes were removed before the pseudotime correlation (see (k)). In sum, 673 cells were considered as basal cells. Basal cells were subclustered into 7 groups using the same approach as described for the 2nd level clustering. The same features selected for the final clustering were used to generate a t-SNE representation of the basal dataset (*perplexity* = 20, *early exaggeration* = 2.0).

(q) Stem Cell Analysis – Negative Binomial Regression

To model genetic signatures which are either unique for each stem cell population or shared by all basal SCM+ or SCM− cells, we created two negative binomial regression models. (1) In the first model, gene expression in stem cells was modeled as a combination of *Baseline* expression, specific signatures unique to each stem cell population (e.g., all *Lgr5*+ cells) and signatures shared by SCM+ and SCM− cells. This model was used to determine stem cell population-specific gene expression signatures, which were called using a 95% posterior probability cut-off against *Baseline*. (2) The second approach modeled gene expression in stem cells as a combination of *Baseline* expression, two common signatures shared by all basal SCM+ and SCM− cells, and specific signatures unique to each compartment (IFE, uHF, upper OB and OB). As the second approach performed better in modeling SCM+ and SCM− signatures, it was used to define the SCM+ signature (90% posterior probability against *Baseline*; see Figure S7F) and to compare SCM+ to SCM− signatures. A gene was considered differentially expressed in SCM+ compared to SCM− cells (or vice versa) if it was represented with at least 0.25 molecules (median) in the SCM+ signature and if its SCM+ signature exceeds the SCM− signature with 90% posterior probability.

DATA AND SOFTWARE AVAILABILITY

Software

The computational analysis workflow and the scripts are available at https://github.com/kasperlab.

Data Resources

The accession number for the sequencing data reported in this paper is NCBI GEO: GSE67602.

ADDITIONAL RESOURCES

An online tool for the visualization of the single-cell dataset is available at http://kasperlab.org/tools or http://linnarssonlab.org/epidermis/.

A systematic staining catalog is provided at: http://kasperlab.org/data.

Cell Systems, Volume *3*

Supplemental Information

Single-Cell Transcriptomics Reveals that Differentiation and Spatial Signatures Shape Epidermal and Hair Follicle Heterogeneity

Simon Joost, Amit Zeisel, Tina Jacob, Xiaoyan Sun, Gioele La Manno, Peter Lönnerberg, Sten Linnarsson, and Maria Kasper

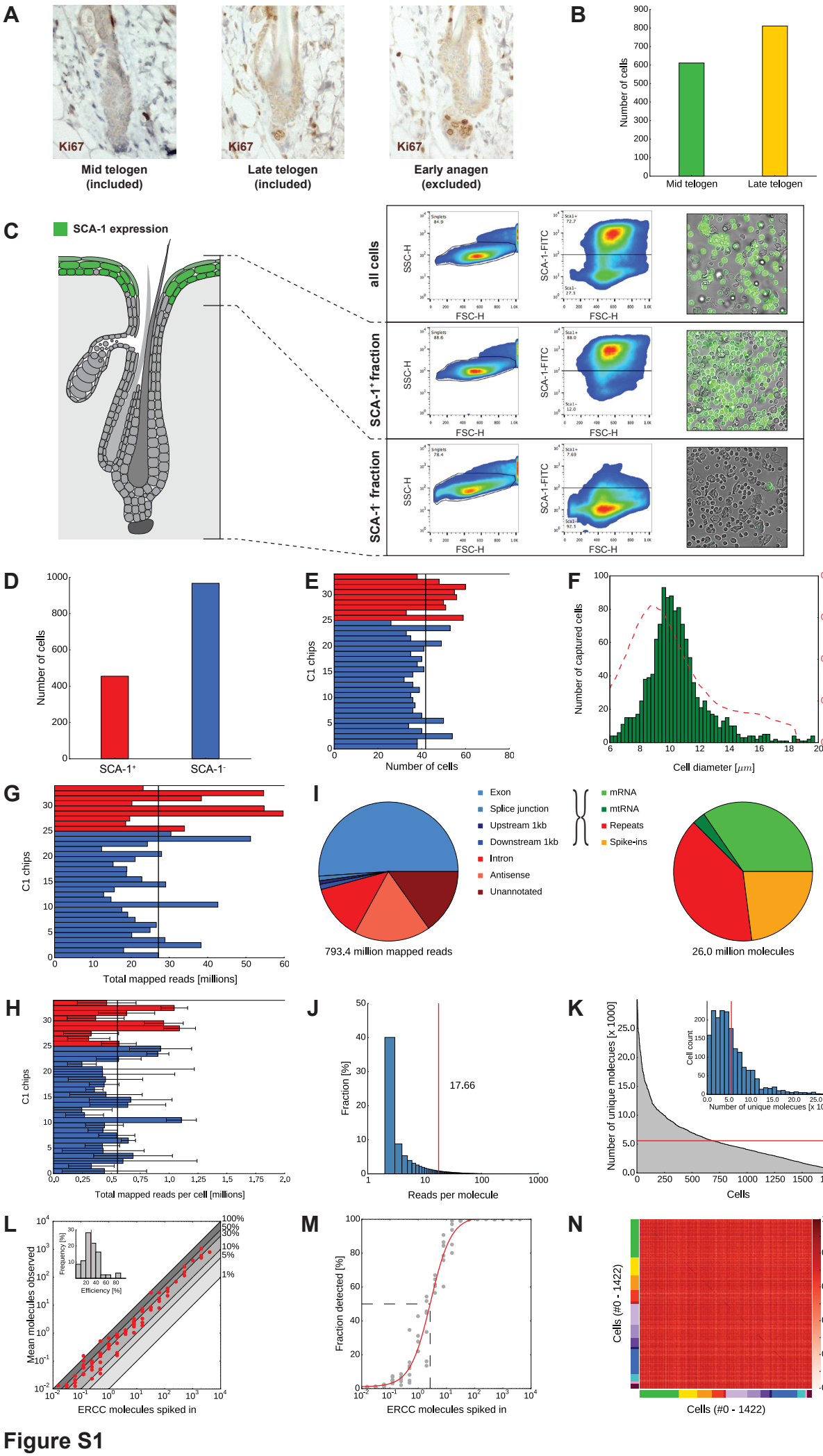
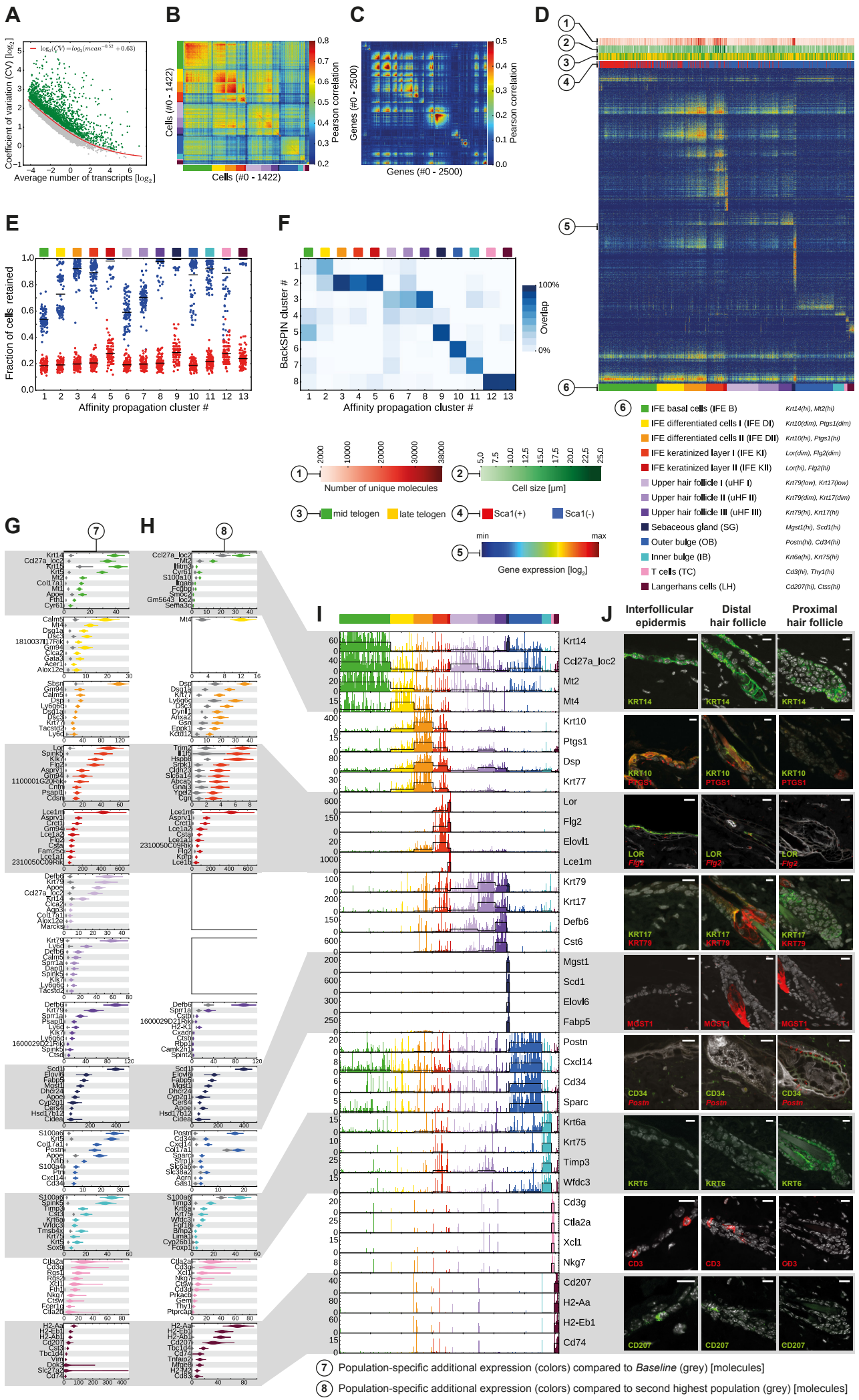

A
Ki67
Mid telogen (included)
Late telogen (included)
Early anagen (excluded)
B
Number of cells
Mid telogen
Late telogen
C
SCA-1 expression
all cells
SCA-1+ fraction
SCA-1- fraction
SSC-H
SCA-1-FITC
FSC-H
D
Number of cells
SCA-1+
SCA-1-
E
C1 chips
Number of cells
F
Number of captured cells
Size distribution of cell input
Cell diameter [μm]
G
C1 chips
Total mapped reads [millions]
I
Exon
Splice junction
Upstream 1kb
Downstream 1kb
Intron
Antisense
Unannotated
mRNA
mtRNA
Repeats
Spike-ins
793.4 million mapped reads
26.0 million molecules
H
C1 chips
Total mapped reads per cell [millions]
J
Fraction [%]
17.66
Reads per molecule
K
Number of unique molecules [x 1000]
Cell count
Number of unique molecules [x 1000]
Cells
L
Mean molecules observed
Frequency [%]
Efficiency [%]
ERCC molecules spiked in
M
Fraction detected [%]
ERCC molecules spiked in
N
Cells (#0 - 1422)
Pearson correlation
Cells (#0 - 1422)

Figure S1

Figure S1. Cell isolation and technical performance. Related to STAR Methods.

(**A**) Representative pictures of Ki67 immunostainings used to stage the dorsal epidermis of all experimental mice. Mice in mid or late telogen (1st and 2nd panel) were included in the study, while mice in anagen (3rd panel) were excluded.
(**B**) Number of captured cells from experimental mice in mid (green bar) and late telogen (yellow bar) that were included in the dataset.
(**C**) Separation of SCA-1+ and SCA-1− cells using microbeads. Left side: illustration of murine telogen epidermis including HF and SG showing the expected expression pattern of SCA-1. Right side: flow cytometry and immunomicroscopy images of epidermal cell suspensions prior to ('all cells') and after ('SCA-1+ fraction', 'SCA-1− fraction') cell separation with SCA-1 microbeads.
(**D**) Number of SCA-1+ (red bar) and SCA-1− cells (blue bar) included in the dataset.
(**E**) Number of cells passing the quality control per C1 chip. 34 C1 chips were sequenced for this study, and in total 1,422 single-cell transcriptomes passed the quality control. Red and blue bars signify chips loaded with either SCA-1+ or SCA-1− cells. Black line: mean over all C1 chips.
(**F**) Size distribution of captured cells included in the dataset (green bars) compared to input cell suspension (red line). Size of captured cells was determined based on the cell area in the microphotographs of cells in the C1 chip. Size distribution of input cell suspensions was measured using a Millipore Scepter Cell Counter and averaged over all experiments. While the single-cell capturing exhibits a minor bias towards larger cells, single cells in the dataset represent the whole size range of the cell input.
(**G–H**) Total mapped reads (G) and total mapped reads per cell (H) for cells passing quality criteria in each sequenced C1 chip. Red and blue bars signify chips loaded with either SCA-1+ or SCA-1− cells. Black lines denote the mean over all C1 chips; error bars in (H) show the standard deviation between individual cells.
(**I**) Census of all reads included in the data based on their alignment to the genome (left side) and census of all unique mRNA molecules based on their class (right side).
(**J**) Number of sequenced reads per molecule (unique molecular identifiers [UMI]). Red line: mean value.
(**K**) Number of unique mRNA molecules (RNA spike-ins and repeats excluded) sequenced from every cell of the initial dataset. Cells were ordered according to number of unique mRNA molecules from highest (38,000 unique molecules) to lowest and cells with less than 2,000 unique molecules were excluded (leaving 1,422 cells in the final dataset). Inset: histogram of cells according to mRNA yield. Red lines represent the average number of unique mRNA molecules over all cells.
(**L**) Efficiency of RNA spike-in detection. For each ERCC spike-in species, the number of molecules added to the reaction is plotted against the average number of molecules detected over all 1,422 cells included in the dataset. The diagonal lines demarcate the efficiency boundaries. Inset: histogram aggregating the detection efficiency of each ERCC spike-in species. Red line: median value.
(**M**) RNA spike-in detection limit. For each ERCC spike-in species, the number of molecules added to the reaction is plotted against the fraction of cells in the dataset in which the molecule was detected at least once. The red line represents a logistic curve fitted to the data. The dotted line marks the inferred minimal number of spike-in molecules necessary for detection in 50% of cases.
(**N**) Uniformity of single-cell cDNA synthesis reaction environment. All cells included in the dataset (ordered based on 1st level clustering) were correlated according to their ERCC spike-in data.



(7) Population-specific additional expression (colors) compared to *Baseline* (grey) [molecules]

(8) Population-specific additional expression (colors) compared to second highest population (grey) [molecules]

Figure S2

Figure S2. First-level clustering of epidermal cells. Related to Figure 1.

(**A**) Selection of genes with high variance for unsupervised clustering using a negative binomial noise model. Scatter represents average number of transcripts (log_2) versus coefficient of variation (CV; log_2) for all genes with average number of transcripts > 0.05 and at least five highly correlated neighbors (7,123 genes). The red line denotes the expected CV as a function of transcript mean according to our fitted noise model $log_2(CV) = log_2(mean^{alpha} + k)$ with alpha = -0.52 and k = 0.63. The 2,500 genes with the largest difference between expected and observed CV are colored in green and were used for 1st level clustering.
(**B–C**) Unsupervised clustering of cells (B) and genes (C). Pearson correlation of cells or genes was used as input for clustering with affinity propagation (AP) and Ward's linkage was subsequently used to define in-cluster order (see STAR Methods). Ordered Pearson correlation matrices of 1,422 cells (B) and 2,500 genes (C) with different color scale cut-offs (right panels) are shown. Group membership for cells in (B**)** is marked in left- and bottom-panels.
(**D**) Heatmap showing the expression of 2,500 genes (rows) in 1,422 single cells (columns). Cells and genes are clustered as in (B) and (C), respectively. Group membership of cells is color-coded in the bottom panel and explained in Figure 1C. The four top panels show cell-specific metadata: number of unique mRNA molecules and size of every cell are visualized in the first and second panel while telogen stage and SCA-1 expression are categorized in the third and fourth panel.
(**E**) Robustness of 1st level clustering was evaluated by resampling (100 iterations) of the dataset and randomly excluding 25% of all cells per iteration. Each subset was reclustered, and the percentage of cells from each cell population that were assigned to the same group was determined (blue dots). The red dots represent the percentage of cells from each group that end up together by pure chance after permutation of cell labels. The black lines show the group means.
(**F**) Comparison of affinity propagation (AP) clustering and unsupervised clustering with backSPIN. Shown is the relative distribution of cells within each AP cluster over all clusters defined by backSPIN.
(**G**) Identification of genes that are most highly expressed over *Baseline* in each population based on negative binomial regression of 1st level clustering data. For each population, the ten genes whose population-specific expression coefficient exceeds the *Baseline* coefficient with 99.9% posterior probability and who show the largest gap to the *Baseline* (difference between the 25th percentile of the population-specific coefficient and the 75th percentile of the *Baseline*) are reported. The gray and colored violin plots show the posterior probability distribution of the *Baseline* and population-specific coefficients respectively (scale in molecules).
(**H**) Identification of genes that are most highly uniquely expressed in each population based on negative binomial regression of 1st level clustering data. For each population, the ten genes whose population-specific expression coefficient exceeds the *Baseline* and all other populations-specific coefficients with 99.9% posterior probability and who show the largest gap to the second highest coefficient (difference between the 25th percentile of the population-specific coefficient and the 75th percentile of the second highest coefficient) are reported. The gray and colored violin plots show the posterior probability distribution of the second highest and population-specific coefficients respectively (scale in molecules).
(**I**) Barplots showing the absolute expression of selected marker genes in each cell. Cells are ordered into groups according to the clustering in (D). Group membership is color-coded in the upper panel and explained in Figure 1C. Black lines show the average expression over each group.
(**J**) Remapping of cell populations according to marker gene expression. For every group, either one or two marker genes were selected and antibodies and single molecule FISH probes (gene names in italics) were used to determine their spatial expression pattern in telogen skin. Counterstaining displayed in gray: DAPI (nuclei) or WGA (cell membranes). Scale bars, 10μm.

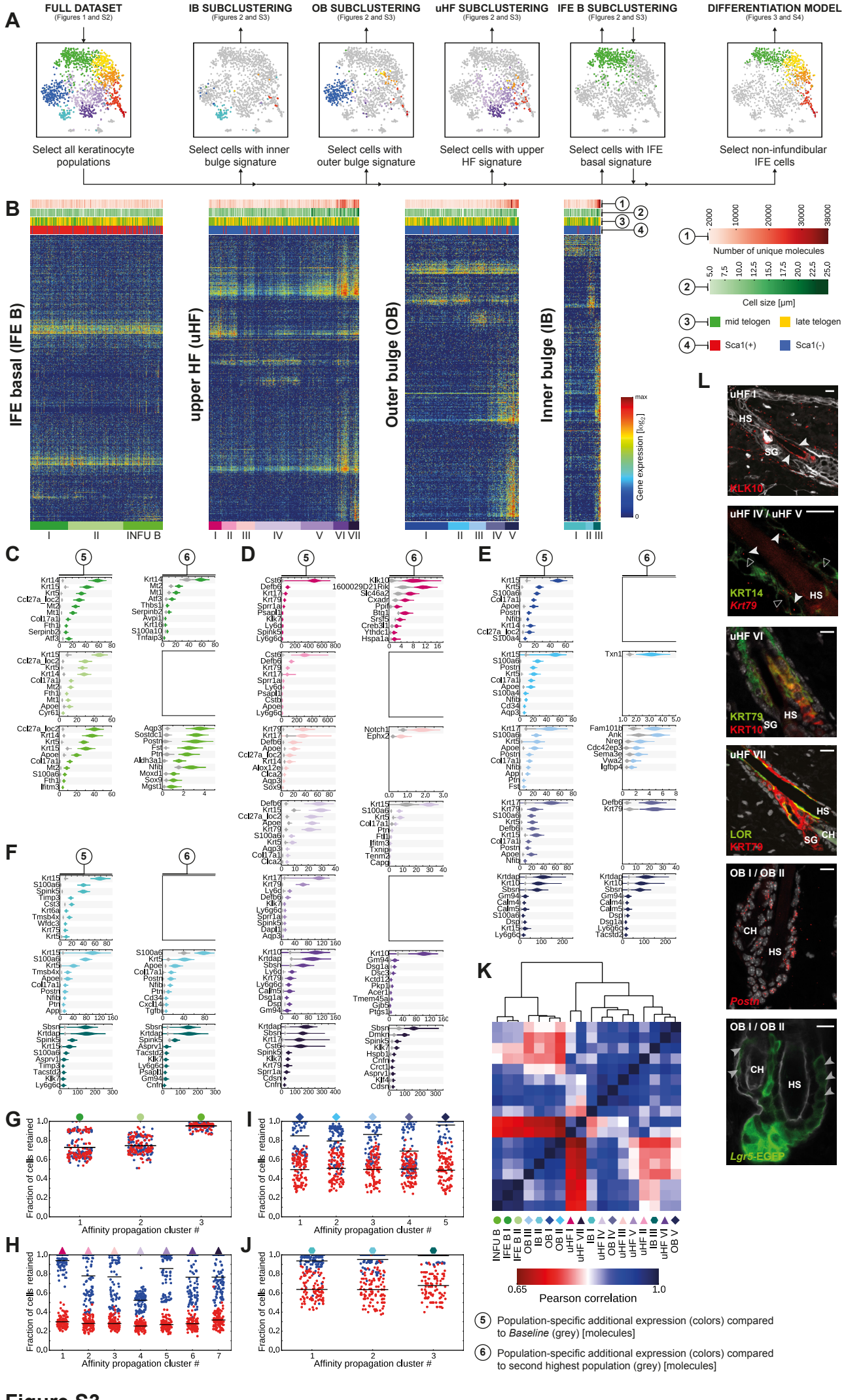
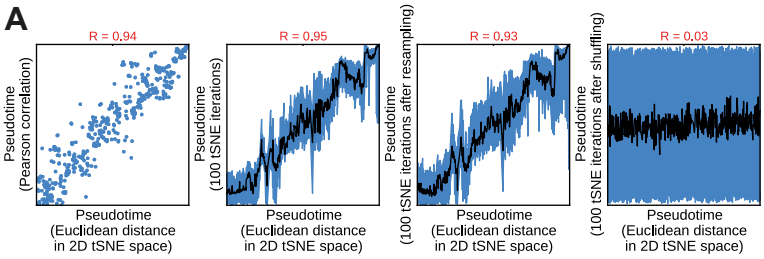
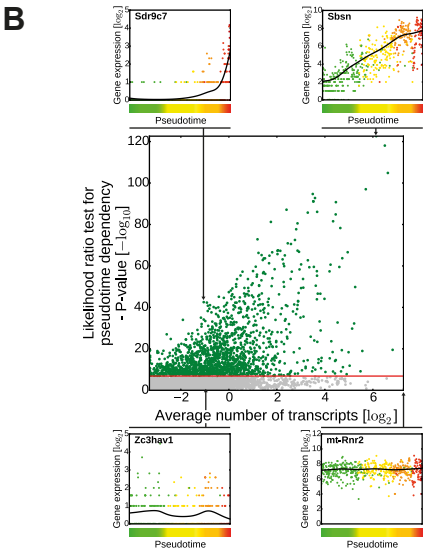

A
FULL DATASET
(Figures 1 and S2)
IB SUBCLUSTERING
(Figures 2 and S3)
OB SUBCLUSTERING
(Figures 2 and S3)
uHF SUBCLUSTERING
(Figures 2 and S3)
IFE B SUBCLUSTERING
(Figures 2 and S3)
DIFFERENTIATION MODEL
(Figures 3 and S4)
Select all keratinocyte populations
Select cells with inner bulge signature
Select cells with outer bulge signature
Select cells with upper HF signature
Select cells with IFE basal signature
Select non-infundibular IFE cells
B
IFE basal (IFE B)
upper HF (uHF)
Outer bulge (OB)
Inner bulge (IB)
I II INFU B
I II III IV V VI VII
I II III IV V
I II III
Gene expression [log2]
max
0
Number of unique molecules
Cell size [µm]
mid telogen
late telogen
Sca1(+)
Sca1(-)
C
D
E
F
G
H
I
J
Fraction of cells retained
Affinity propagation cluster #
K
INFU B
IFE B I
IFE B II
OB III
IB II
OB I
OB II
uHF I
uHF VII
IB I
uHF IV
OB IV
uHF III
uHF V
IB III
uHF VI
OB V
0.65
1.0
Pearson correlation
L
uHF I
HS
SG
XLK10
uHF IV / uHF V
KRT14
Krt79
HS
uHF VI
KRT79
KRT10
HS
SG
uHF VII
LOR
KRT79
HS
SG
CH
OB I / OB II
CH
HS
Postn
OB I / OB II
CH
HS
Lgr5-EGFP
5 Population-specific additional expression (colors) compared to Baseline (grey) [molecules]
6 Population-specific additional expression (colors) compared to second highest population (grey) [molecules]

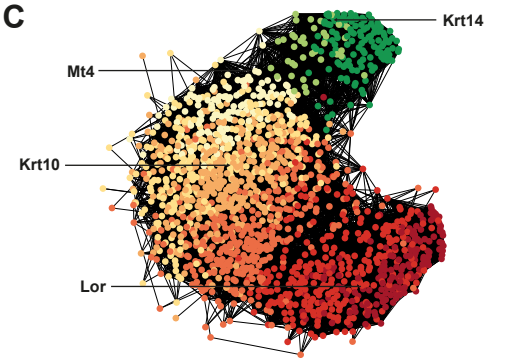
Figure S3

Figure S3. Second-level clustering of epidermal cells. Related to Figure 2.

(**A**) Summary of the cell reselection approach for 2nd level clustering. All cells from the dataset, excluding the sebaceous gland and immune cells, were selected and cells with inner bulge, outer bulge, upper HF and IFE basal signatures were chosen in order of primacy. For pseudotemporal ordering, IFE basal cells, excluding infundibular cells, were combined with the remaining IFE cells (intermediate, mature and terminally differentiated).
(**B**) Heatmaps showing the subclustering of IFE basal, upper HF, outer, and inner bulge cells. Group membership of cells is color-coded in the bottom panel and explained in Figures 2F–2G. The four top panels show cell-specific metadata: number of unique mRNA molecules and size of every cell are visualized in the first and second panel while telogen stage and SCA-1 expression are categorized in the third and fourth panel.
(**C–F**) Identification of genes that are most highly expressed over *Baseline* and most highly uniquely expressed in each IFE basal (C), upper HF (D), outer bulge (E) and inner bulge (F) subpopulation based on negative binomial regression of 2nd level clustering data. Left panels: for each subpopulation, the ten genes whose population-specific expression coefficient exceeds the *Baseline* coefficient with 95% posterior probability and who show the largest gap to the *Baseline* (difference between the 25th percentile of the population-specific coefficient and the 75th percentile of the *Baseline*) are reported. The gray and colored violin plots show the posterior probability distribution of the *Baseline* and population-specific coefficients respectively (scale in molecules). Right panels: for each subpopulation, the ten genes whose subpopulation-specific expression coefficient exceeds the *Baseline* and all other subpopulations-specific coefficients (limited to either the subpopulations of the IFE basal, uHF, OB or IB) with 95% posterior probability and who show the largest gap to the second highest coefficient (difference between the 25th percentile of the subpopulation-specific coefficient and the 75th percentile of the second highest coefficient) are reported. The gray and colored violin plots show the posterior probability distribution of the second highest and subpopulation-specific coefficients respectively (scale in molecules).
(**G–J**) Robustness of IFE basal (G), upper HF (H), outer bulge (I) and inner bulge (J) clustering was evaluated by resampling (100 iterations) of the dataset and randomly excluding 25% of all cells per iteration. Each subset was reclustered, and the percentage of cells from each cell population that were assigned to the same group was determined (blue dots). The red dots represent the percentage of cells from each group that end up together by pure chance after permutation of cell labels. The black lines show the group means.
(**K**) Transcriptomic similarity of 2nd level subclusters visualized by Ward's linkage hierarchical clustering of single-cell gene expression data averaged over each group.
(**L**) Remapping of subpopulations to their spatial location in the epidermis by immunostaining or single molecule FISH (gene symbols in italics). A summary of the populations' spatial localization can be found in Figure 2G. Arrowheads highlight the positions of the populations. uHF IV (empty arrowhead) / uHF V (filled arrowhead). OB II (arrowhead marks Lgr5(dim) cells in *Lgr5-EGFP-Ires-CreERT2* mice using anti-EGFP staining). HS, hair shaft. SG, sebaceous gland. CH, club hair. Scale bars, 10μm.

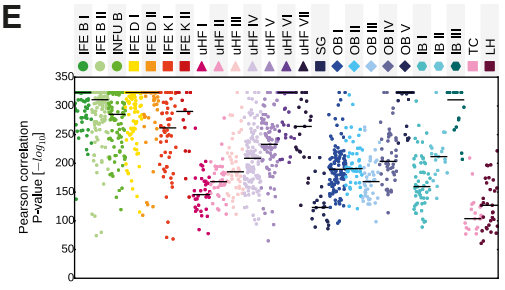


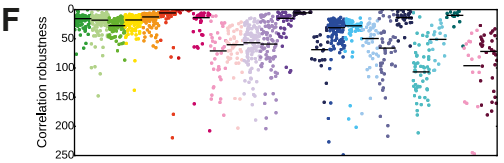




D

	Pseudotime	Network	TFs	Mol. signatures
I	Pseudotime		Bhlhe40, Snai2, Tsc22d1, Klf9	• Extracellular matrix • Focal adhesion • ECM-receptor interaction • Tissue development
II	Pseudotime		Sox6, Trp63, Zfp36l2	• Cellular component assembly • Protein transport • Chloride channel activity
III	Pseudotime		Hes1, Gata3, Pou3f1, Tsc22d3, Ikzf2	• Lipid metabolic process • Ribosome • Cell development • Apoptosis
IV	Pseudotime		Phb, Aff4, Tet3, Zmat3, Irf6, Srebf2, Nfatc2, Creb3, Ppp1r13l	• Steroid / cholesterol biosynthesis • Immune system • Adherens junction • Tight junction • Nectin pathway
V	Pseudotime		Cbfb, Pou2f3, Tcf25, Kdm5b, Zfp750, Mta3, Grhl3, Cebpa, Zbtb7a, Nfe2l1, Tsc22d4, Zfp91, Casz1, Ovol1, Zc3h15'	• Protein transport • Secretory pathway • Wnt signaling • Biopolymer metabolic process • Immune system
VI	Pseudotime		Klf4, Mafk, Nfat5, Bach1, Rora, Jarid2, Mllt4, Klf3, Xbp1, Grhl1, Hivep1, Lrrfip2, Rel, Egr2, Hopx, Sp6, Maff, Csrnp2, Cebpb	• Protein metabolic process / Proteolysis • Signal transduction • Lipid metabolism • Tissue development • Protein transport
VII	Pseudotime		Mxi1, Tead1, Rorc, Zfp706, Max, Hbp1, Tsc22d2, Mxd1, Prdm1	• Proteolysis • Endocytosis • MAPK signaling pathway • Epidermis development • Met pathway
VIII	Pseudotime		Id4	• Sphingolipid metabolism • MAPK signaling pathway • Cytokine / cytokine receptor interaction





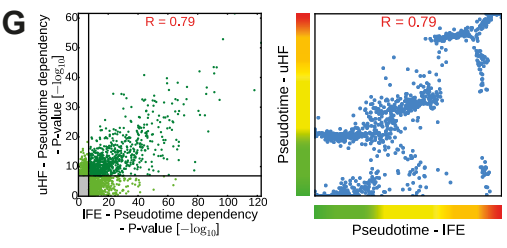


Figure S4

Figure S4. Modeling of the epidermal differentiation process. Related to Figure 3.

(**A**) Robustness of pseudotemporal ordering. Far left panel: comparison of the pseudotemporal ordering selected for Figures 3 and S4 (x-axis) to a pseudotemporal ordering acquired without dimensional reduction through t-SNE (y-axis). Center left panel: comparison of the selected pseudotemporal ordering (x-axis) to one hundred randomly acquired alternative orderings based on different initial t-SNE plots (y-axis). Center right panel: comparison of the selected pseudotemporal ordering (x-axis) to one hundred alternative orderings after randomly removing 25% of cells (y-axis). Far right panel: comparison of the selected pseudotemporal ordering (x-axis) to one hundred alternative orderings after shuffling cell labels (y-axis). The black line in the last three plots shows the median position of each cell over all one hundred iterations, while the blue areas cover the range between the 5^{th} and 95^{th} percentile.
(**B**) Center: average expression of 7,345 genes expressed during IFE differentiation plotted against pseudotime-dependency. Pseudotime-dependency was tested against a pseudotime-independent restricted model using an approximate likelihood ratio test (see STAR Methods) and the p-values are reported. The 1,627 genes (green) with p-values below a Bonferroni-corrected significance threshold of 0.001 (red line) were used for further analysis. Upper/lower panels: examples of low and high expressed genes with or without pseudotime-dependency, respectively.
(**C**) Shared nearest neighbor network of 1,627 pseudotime-dependent genes. Genes are colored according to group membership as established in Figure 3C.
(**D**) Characteristics of different subgroups of differentiation-related genes as defined in Figure 3C. Pseudotime: averaged expression of subgroup-specific genes over pseudotime. The red line shows the median while the gray lines demarcate the 25^{th} and 75^{th} percentile. Network: position of subgroup-specific genes in the shared nearest neighbor network established in (C). TFs: transcription factors included in each subgroup of genes. Mol. signatures: molecular and functional signatures linked to each subgroup of genes.
(**E**) P-values corresponding to the best correlation (highest correlation coefficient) of each cell to the pseudotime model as shown in Figure 3F. Cells are grouped according to (sub) population membership as defined by 2^{nd} level clustering. Black lines denote the median over each group.
(**F**) Robustness of each cell's correlation to the pseudotime model. To measure robustness of correlation, cells were re-correlated to the pseudotime model for one hundred times after randomly removing 75% of pseudotime-dependent genes. Shown is the average distance between a cell's pseudotime position in the full model and its position in the re-correlations. It is assumed that a small average distance is indicative of a more robust link to a particular stage in the pseudotime model.
(**G**) Comparison of pseudotemporal ordering of cells in the IFE and the uHF. Left panel: comparison of differentiation-dependency of genes involved in IFE and uHF differentiation. Right panel: comparison of the pseudotime-positions of epidermal cells derived from an IFE- and uHF-based model of differentiation.

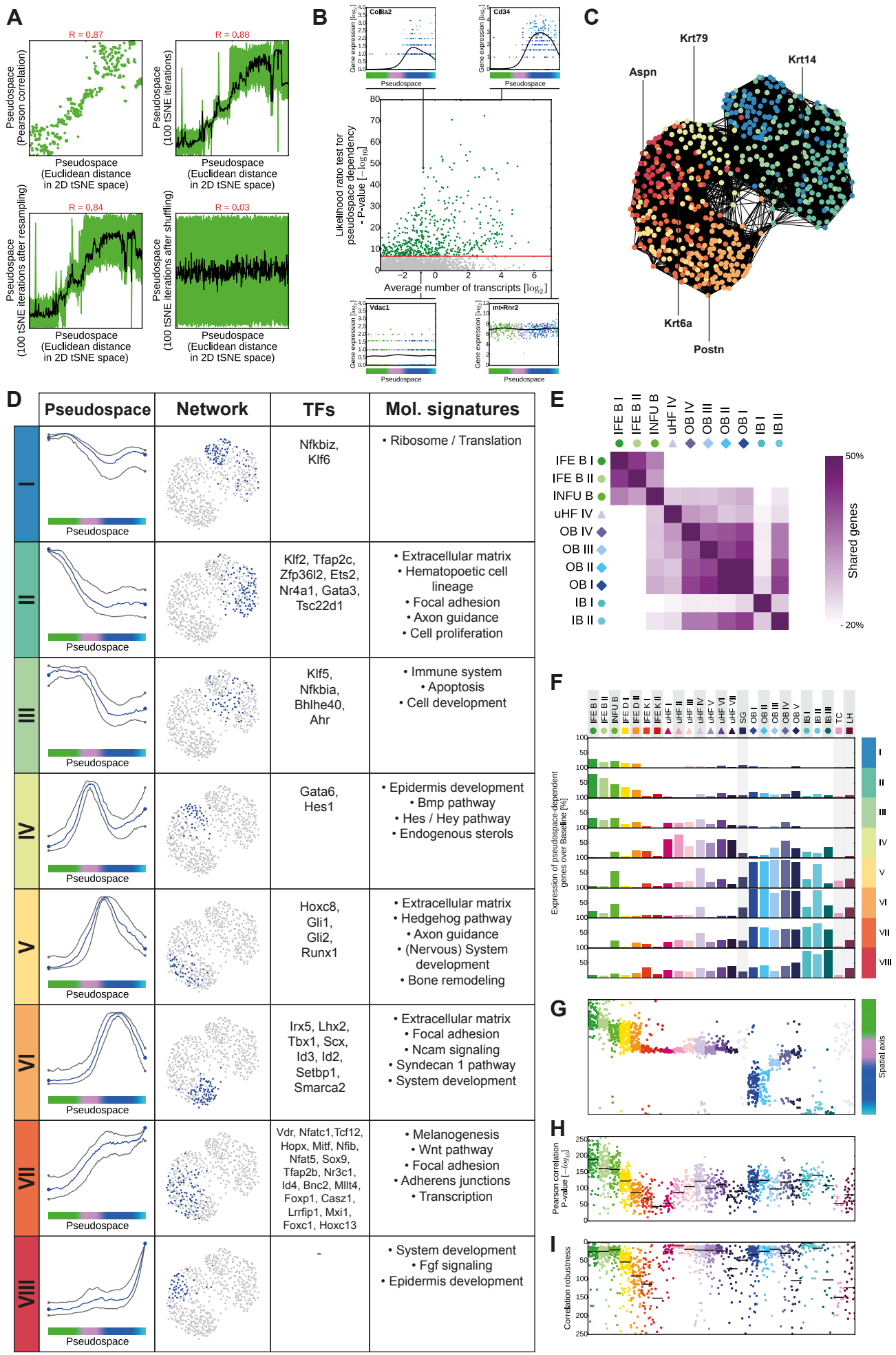

A
R = 0.87
R = 0.88
R = 0.84
R = 0.03
Pseudospace (Pearson correlation)
Pseudospace (100 tSNE iterations)
Pseudospace (100 tSNE iterations after resampling)
Pseudospace (100 tSNE iterations after shuffling)
Pseudospace (Euclidean distance in 2D tSNE space)
B
Col8a2
Cd34
Vdac1
mt-Rnr2
Gene expression [log2]
Pseudospace
Likelihood ratio test for pseudospace dependency - P-value [−log10]
Average number of transcripts [log2]
C
Krt79
Krt14
Aspn
Krt6a
Postn
D
Pseudospace
Network
TFs
Mol. signatures
I
Nfkbiz, Klf6
• Ribosome / Translation
II
Klf2, Tfap2c, Zfp36l2, Ets2, Nr4a1, Gata3, Tsc22d1
• Extracellular matrix
• Hematopoetic cell lineage
• Focal adhesion
• Axon guidance
• Cell proliferation
III
Klf5, Nfkbia, Bhlhe40, Ahr
• Immune system
• Apoptosis
• Cell development
IV
Gata6, Hes1
• Epidermis development
• Bmp pathway
• Hes / Hey pathway
• Endogenous sterols
V
Hoxc8, Gli1, Gli2, Runx1
• Extracellular matrix
• Hedgehog pathway
• Axon guidance
• (Nervous) System development
• Bone remodeling
VI
Irx5, Lhx2, Tbx1, Scx, Id3, Id2, Setbp1, Smarca2
• Extracellular matrix
• Focal adhesion
• Ncam signaling
• Syndecan 1 pathway
• System development
VII
Vdr, Nfatc1,Tcf12, Hopx, Mitf, Nfib, Nfat5, Sox9, Tfap2b, Nr3c1, Id4, Bnc2, Mllt4, Foxp1, Casz1, Lrrfip1, Mxi1, Foxc1, Hoxc13
• Melanogenesis
• Wnt pathway
• Focal adhesion
• Adherens junctions
• Transcription
VIII
-
• System development
• Fgf signaling
• Epidermis development
E
IFE B I
IFE B II
INFU B
uHF IV
OB IV
OB III
OB II
OB I
IB I
IB II
Shared genes
50%
20%
F
Expression of pseudospace-dependent genes over Baseline [%]
G
Spatial axis
H
Pearson correlation P-value [−log10]
I
Correlation robustness

Figure S5

Figure S5. Modeling of spatial gene expression signatures. Related to Figure 4.

(**A**) Robustness of pseudospatial ordering. Upper left panel: comparison of the pseudospatial ordering selected for Figures 4 and S5 (x-axis) to a pseudospatial ordering acquired without dimensional reduction through t-SNE (y-axis). Upper right panel: comparison of the selected pseudospatial ordering (x-axis) to one hundred randomly acquired alternative orderings based on different initial t-SNE plots (y-axis). Lower left panel: comparison of the selected pseudospatial ordering (x-axis) to one hundred alternative orderings after randomly removing 25% of cells (y-axis). Lower right panel: comparison of the selected pseudospatial ordering (x-axis) to one hundred alternative orderings after shuffling cell labels (y-axis). The black line in the last three plots shows the median position of each cell over all one hundred iterations, while the green areas cover the range between the 5^{th} and 95^{th} percentile.
(**B**) Center: average expression of 6,788 genes expressed in basal cells plotted against pseudospace-dependency. Pseudospace-dependency was tested against a restricted model using an approximate likelihood ratio test (see STAR Methods) and the p-values are reported. The 547 genes (green) with p-values below a Bonferroni-corrected significance threshold of 0.001 (red line) were used for further analysis. Upper/lower panels: examples of low and high expressed genes with or without pseudospace-dependency, respectively.
(**C**) Shared nearest neighbor network of 547 pseudospace-dependent genes. Genes are colored according to group membership as established in Figure 4C.
(**D**) Characteristics of different subgroups of spatial genes as defined in Figure 4C. Pseudospace: averaged expression of subgroup-specific genes over pseudospace. The blue line shows the median while the gray lines demarcate the 25^{th} and 75^{th} percentile. Network: position of subgroup-specific genes in the shared nearest-neighbor network established in (C). TFs: transcription factors included in each subgroup of genes. Mol. signatures: molecular and functional signatures linked to each subgroup of genes.
(**E**) Overlap of genes expressed over *Baseline* in basal-cell populations (and IB I). Genes were called from the negative binomial regression model of 2^{nd} level clustering if the population-specific regression coefficient exceeded *Baseline* with 95% posterior probability.
(**F**) Expression of spatial genes in all epidermal (sub) populations defined by either 1^{st} or 2^{nd} level clustering. A gene was considered expressed in a population if its population-specific coefficient in the negative binomial regression model exceeded *Baseline* with 95% posterior probability. Genes are ordered according to group membership introduced in Figure 4C. The shaded populations exhibit gene expression inconsistent with any distinct spatial signature.
(**G**) Position of epidermal cells from each population on the spatial axis as determined by highest Pearson correlation. The shaded cells belong to populations that show gene expression inconsistent with any spatial signature.
(**H**) P-values corresponding to the best correlation of each cell to the pseudospace model as shown in (G). Black lines denote the median over each group.
(**I**) Robustness of each cell's correlation to the pseudospace model. To measure robustness of correlation, cells were re-correlated to the spatial axis for one hundred times after randomly removing 75% of pseudospace-dependent genes. Shown is the average distance between a cell's pseudospace position in the full model and its position in the re-correlations. It is assumed that a small average distance is indicative of a more robust link to a particular stage in the pseudospace model.

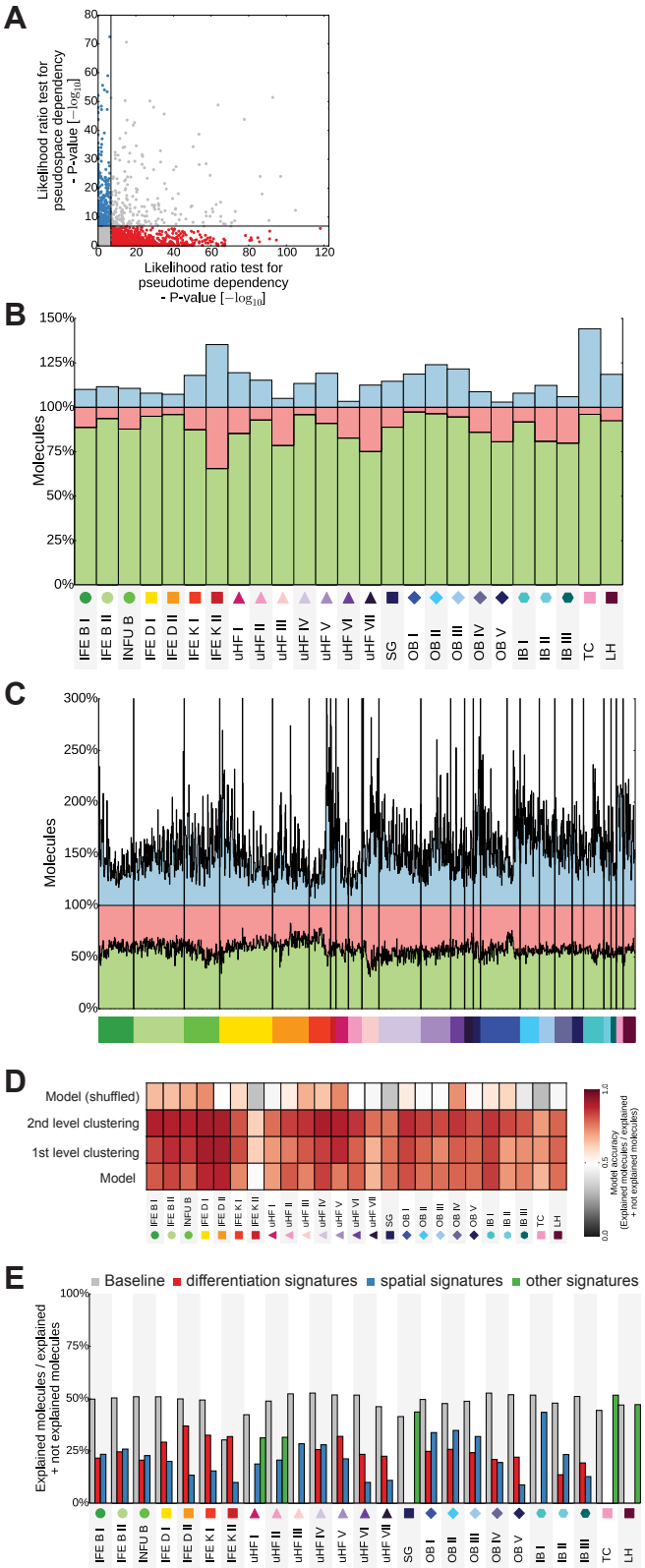

A
Likelihood ratio test for pseudospace dependency - P-value [−log10]
Likelihood ratio test for pseudotime dependency - P-value [−log10]
B
Molecules
C
Molecules
D
Model (shuffled)
2nd level clustering
1st level clustering
Model
Model accuracy (Explained molecules / explained + not explained molecules)
E
Baseline
differentiation signatures
spatial signatures
other signatures
Explained molecules / explained + not explained molecules
IFE B I
IFE B II
INFU B
IFE D I
IFE D II
IFE K I
IFE K II
uHF I
uHF II
uHF III
uHF IV
uHF V
uHF VI
uHF VII
SG
OB I
OB II
OB III
OB IV
OB V
IB I
IB II
IB III
TC
LH

Figure S6

Figure S6. Explaining cellular heterogeneity using differentiation and spatial signatures. Related to Figure 5.

(**A**) Pseudotime- and pseudospace-dependency of genes. Black lines mark the Bonferroni-corrected significance threshold of 0.001. Of 7,893 genes, 1,409 were uniquely pseudotime-, 329 uniquely pseudospace- and 218 both pseudotime- and pseudospace-dependent. Only the uniquely pseudotime- or pseudospace-dependent genes were considered in the pseudospacetime model.
(**B**) Percentage of molecules explained (green), underexplained (red) or overexplained (blue) by the pseudospacetime model. Molecules were pooled across the cells per population.
(**C**) Percentage of molecules explained (green), underexplained (red) or overexplained (blue) by the pseudospacetime model in each single cell per population.
(**D**) Accuracy of pseudospacetime, 1^{st} level clustering, 2^{nd} level clustering and shuffled pseudospacetime model stratified according to cell populations defined in either 1^{st} or 2^{nd} level clustering.
(**E**) Fraction of explained molecules contributed by *Baseline*, differentiation axis, spatial axis and other (sebaceous gland, immune) signatures for each cell population.

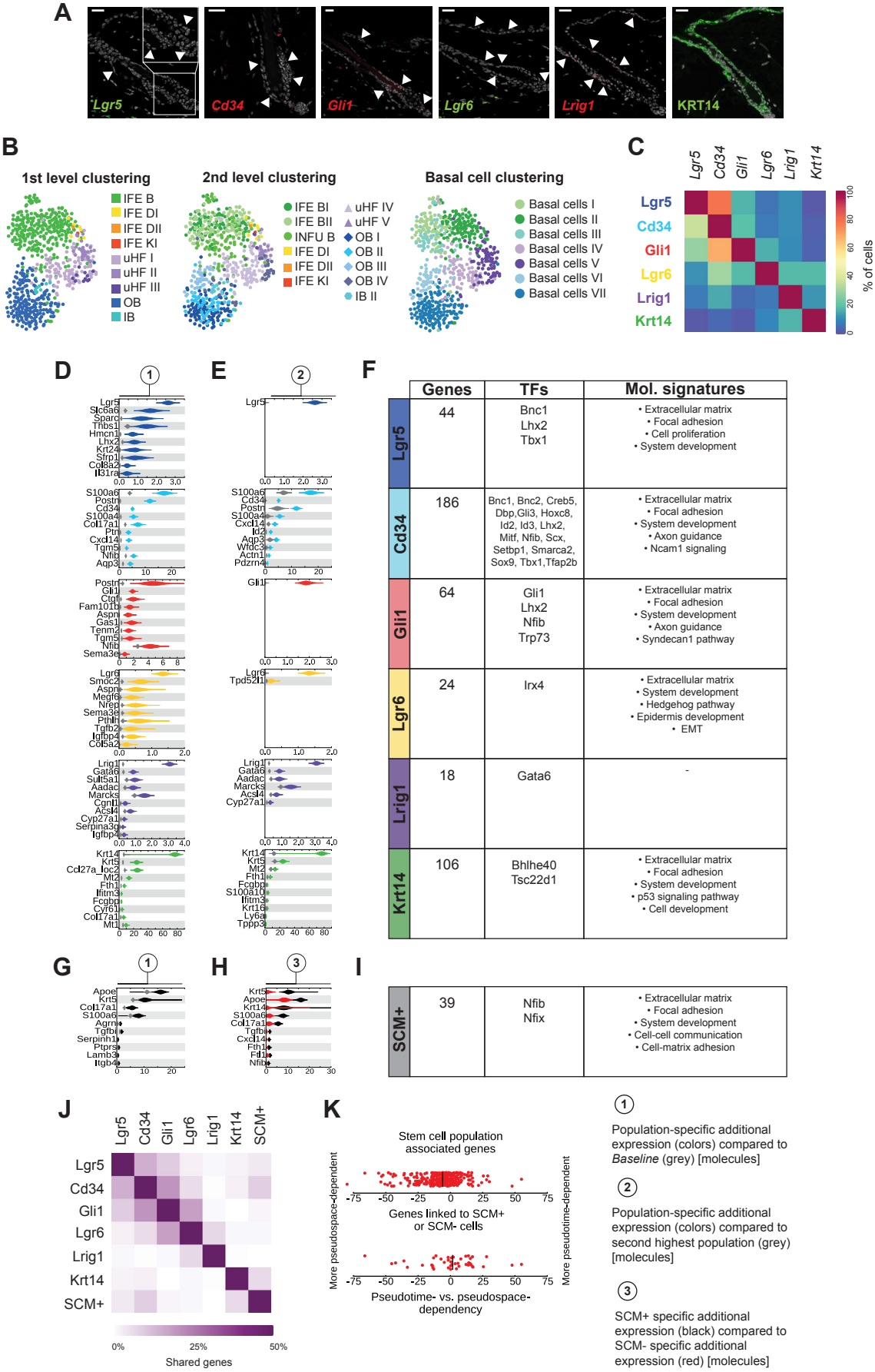

A
Lgr5
Cd34
Gli1
Lgr6
Lrig1
KRT14
B
1st level clustering
IFE B
IFE DI
IFE DII
IFE KI
uHF I
uHF II
uHF III
OB
IB
2nd level clustering
IFE BI
IFE BII
INFU B
IFE DI
IFE DII
IFE KI
uHF IV
uHF V
OB I
OB II
OB III
OB IV
IB II
Basal cell clustering
Basal cells I
Basal cells II
Basal cells III
Basal cells IV
Basal cells V
Basal cells VI
Basal cells VII
C
Lgr5
Cd34
Gli1
Lgr6
Lrig1
Krt14
% of cells
D
E
F
Genes
TFs
Mol. signatures
Lgr5
44
Bnc1
Lhx2
Tbx1
• Extracellular matrix
• Focal adhesion
• Cell proliferation
• System development
Cd34
186
Bnc1, Bnc2, Creb5, Dbp,Gli3, Hoxc8, Id2, Id3, Lhx2, Mitf, Nfib, Scx, Setbp1, Smarca2, Sox9, Tbx1,Tfap2b
• Extracellular matrix
• Focal adhesion
• System development
• Axon guidance
• Ncam1 signaling
Gli1
64
Gli1
Lhx2
Nfib
Trp73
• Extracellular matrix
• Focal adhesion
• System development
• Axon guidance
• Syndecan1 pathway
Lgr6
24
Irx4
• Extracellular matrix
• System development
• Hedgehog pathway
• Epidermis development
• EMT
Lrig1
18
Gata6
-
Krt14
106
Bhlhe40
Tsc22d1
• Extracellular matrix
• Focal adhesion
• System development
• p53 signaling pathway
• Cell development
G
H
I
SCM+
39
Nfib
Nfix
• Extracellular matrix
• Focal adhesion
• System development
• Cell-cell communication
• Cell-matrix adhesion
J
Shared genes
K
Stem cell population associated genes
Genes linked to SCM+ or SCM- cells
Pseudotime- vs. pseudospace-dependency
More pseudospace-dependent
More pseudotime-dependent
1
Population-specific additional expression (colors) compared to Baseline (grey) [molecules]
2
Population-specific additional expression (colors) compared to second highest population (grey) [molecules]
3
SCM+ specific additional expression (black) compared to SCM- specific additional expression (red) [molecules]

Figure S7

Figure S7. Cellular heterogeneity of stem cell populations. Related to Figure 6.

(**A**) Immunostaining and single molecule FISH (gene symbols in italics) of SCMs in epidermis. Note that most markers are expressed in several epidermal compartments. Scale bars, 20μm.
(**B**) Cells projected onto the t-SNE map of basal cells (see Figure 6B), colored according to 1st level, 2nd level and selective clustering of basal cells.
(**C**) Matrix showing the overlap in SCM expression. Percentage of cells expressing each SCM *Lgr5*, *Cd34*, *Gli1*, *Lgr6*, *Lrig1,* or *Krt14* (rows) co-expressing additional SCM genes (columns).
(**D–F**) Analyses of gene expression signatures for SCM-expressing populations. (D) Identification of the top ten genes that were most highly expressed over *Baseline* in each stem cell population based on negative binomial regression. For each population, the ten genes whose population-specific expression coefficient exceeds the *Baseline* coefficient with 95% posterior probability and which show the largest gap to the *Baseline* (difference between the 25th percentile of the population-specific coefficient and the 75th percentile of the *Baseline*) are reported. The gray and colored violin plots show the posterior probability distribution of the *Baseline* and population-specific coefficients, respectively (scale in molecules). (E) Identification of genes that are most highly and uniquely expressed in each stem cell population based on negative binomial regression. For each population, the ten genes whose population-specific expression coefficient exceeds the *Baseline* and all other populations-specific coefficients with 95% posterior probability and that show the largest gap to the second highest coefficient (difference between the 25th percentile of the population-specific coefficient and the 75th percentile of the second highest coefficient) are reported. The gray and colored violin plots show the posterior probability distribution of the second highest and population-specific coefficients, respectively (scale in molecules). (F) Characteristics of different stem cell populations. Genes: number of genes expressed over *Baseline* with 95% posterior probability. TFs: transcription factors included in each set of genes. Mol. signatures: molecular and functional signatures linked to each subgroup of genes.
(**G–I**) Analyses of shared gene expression signatures in SCM+ cells. (G) Identification of the top ten genes that were most highly expressed over *Baseline* among all SCM+ cells. In contrast to (D), a 90% posterior probability cut-off was chosen. (H) Identification of the top ten genes expressed in all SCM+ cells that were most highly expressed compared to SCM− cells. A 90% posterior probability cut-off was chosen and only genes whose SCM+ cell-specific expression exceeded 0.25 molecules were chosen. The black and red violin plots show the posterior probability distribution of the SCM+ and SCM− specific coefficients respectively. (I) Characteristics of the SCM+ population. Genes: number of genes with higher expression than in the SCM− population based on a 90% posterior probability cut-off. TFs: transcription factors. Mol. signatures: molecular and functional signatures.
(**J**) Percentage of shared genes between the specific signatures of *Lgr5*+, *Cd34*+, *Gli1*+, *Lgr6*+, *Lrig1*+, *Krt14*^{hi}, and SCM+ cells. The specific signatures were defined as specified in (D), (F), (G) and (I).
(**K**) Pseudotime- vs. pseudospace-dependency of stem cell-specific genes. Plotted is the difference between the $-\log_{10}$ transformed p-value of pseudotime- and pseudospace-dependency. "Stem cell population associated genes" include all genes, which are expressed over *Baseline* in at least one stem/progenitor population. "Genes linked to SCM+ or SCM– cells" are the 44 genes that are expressed differently between SCM+ and SCM– cells as specified in Figures 6G and S7G – S7I. The black lines denote the median.

SUPPLEMENTAL TABLES

Table S1. Marker genes – 1st level clustering. Related to Figure 1.

Lists of genes that are most highly expressed over *Baseline* (vs. *Baseline,* left column) or the second highest population (vs. other groups, right column). Identification of genes is based on negative binomial regression of 1st level clustering data. For each population, genes whose population-specific expression coefficient exceeded the *Baseline* coefficient (left column) or all other populations-specific coefficients (right column) with 99.9% posterior probability and which show the largest gap to the *Baseline* / the second highest population (difference between the 25th percentile of the population-specific coefficient and the 75th percentile of the *Baseline* / the second highest population) are listed.

(Supplied as Excel file: Joost Table S1.xlsx)

Table S2. Marker genes – 2nd level clustering. Related to Figure 2.

Lists of genes that are most highly expressed over *Baseline* (vs. *Baseline,* left column) or the second highest population (vs. other groups, right column). Identification of genes is based on negative binomial regression of 2nd level clustering data. For each population, genes whose population-specific expression coefficient exceeded the *Baseline* coefficient (left column) or all other populations-specific coefficients (right column) with 95% posterior probability and which show the largest gap to the *Baseline* / the second highest population (difference between the 25th percentile of the population-specific coefficient and the 75th percentile of the *Baseline* / the second highest population) are listed.

(Supplied as Excel file: Joost Table S2.xlsx)

Table S3. Description of cell populations and comparison to literature. Related to Figures 1, 2 and 6.

(**A**) Description of cell populations defined during 1st and 2nd level clustering. Described are the molecular and spatial characteristics of each population defined in this study and their relation to previous work.
(**B**) Previously defined murine epidermal (stem) cell populations and their relation to populations defined in this study.

A. Description of cell populations from 1st and 2nd level clustering

Populations	Molecular and spatial description in Joost et al.	Previous descriptions
Interfollicular basal I (IFE B I)	IFE basal cell population marked by *Thbs1* expression and higher than average expression of IFE basal genes such as *Krt14*, *Mt1* and *Mt2*. IFE B I cells are interspersed with IFE B II cells in the IFE basal layer.	A THBS1-positive subpopulation of the IFE has not been previously described. This population is not congruent with the *Ivl*+ population or *Lgr6*+ population described by Mascré et al., 2012 and Füllgrabe et al., 2015.
Interfollicular basal II (IFE B II)	IFE basal cell population marked by absence of *Thbs1* and lower than average expression of IFE basal genes such as *Krt14*, *Mt1*, *Mt2*. IFE B II cells are interspersed with IFE B I cells in the IFE basal layer.	A distinct THBS1-negative subpopulation of the IFE has not been previously described. This population is not congruent with the *Ivl*+ population or *Lgr6*+ population described by Mascré et al., 2012 and Füllgrabe et al., 2015.
Infundibular basal (INFU B)	A population of cells dominated by typical IFE basal markers such as *Krt14*, *Mt1* and *Mt2*, which additionally expresses pan and upper HF markers such as *Sostdc1*, *Aqp3*, *Postn* and *Krt79*. Located in the infundibular region close to the HF opening. The specific spatial position of this population is clearly identifiable by *Postn* expression.	While it has been shown that pan HF markers such as *Sostdc1* reach into the infundibular region (Collette et al., 2013), the particularly IFE basal character of this population in combination with low-level gradual expression of HF genes has not been clarified before.
IFE differentiated cells I (IFE D I)	Transient population of cells marked by low level expression of both (IFE) basal (*Krt14*, *Mt2*) and (IFE) suprabasal (*Krt10*, *Sbsn*) markers. Expresses *Mt4* as one of a few specific markers.	Although the existence of this basal-to-suprabasal transient population is expected according to the accepted model of epidermal differentiation, it has never been resolved at a transcriptional level.
IFE differentiated cells II (IFE D II)	Mature, suprabasal population expressing high levels of well-established spinous layer markers such as *Krt10* and *Sbsn*.	The cells of the spinous layer are well described and characterized. See for instance Fuchs, 1990.
IFE keratinized layer I (IFE K I)	A transient population of cells which is marked by decreasing levels of spinous layer markers such as *Krt10* and *Sbsn* (when compared to IFE D II) and increasing levels of granular layer markers including *Lor* and *Flg2*.	Although the existence of this suprabasal/spinous-to-terminal/granular transient population is expected according to the accepted model of epidermal differentiation, it has never been resolved at a transcriptional level.
IFE keratinized layer II (IFE K II)	A population of flat, keratinized cells expressing high levels of well-established granular layer markers such as *Lor* and *Flg2*.	The cells of the granular layer are well described and characterized. See for instance Fuchs, 1990.
Upper HF I (uHF I)	A population of cells marked by a typical uHF signature (*Krt79*, *Krt17*, *Cd44*) in combination with expression of a gene module distinguished by high *Rbp1*, *Defb6* and *Cst6* expression. Additionally epidermis- and compartment-unique markers such as *Klk10* and *Cryab* are expressed. Could be mapped to two rings of suprabasal cells above and below the SG opening based on highest CST6 expression. Accordingly, KLK10 is mostly found secreted into the hair canal at the corresponding positions.	Has not been previously described in molecular detail. Although Zeeuwen et al., 2002 and Veniaminova et al. 2013 described CST6 expression in the upper HF, they were unable to differentiate a set of uHF populations with strong CST6 expression and a set of uHF populations with weak or absent CST6 expression or subdivide those populations further. It is not clear whether this population contains cells of BLIMP1 / *Prdm1* population described by Horsley et al., 2006.
Upper HF II (uHF II)	A population of cells marked by a typical uHF signature (*Krt79*, *Krt17*, *Cd44*) in combination with high expression of a gene module distinguished by high *Rbp1*, *Defb6* and *Cst6* expression. This population is additionally distinguished by a *Krt14* / *Lrig1* signature while the *Krt5* / *Ptn* module is absent. Can be mapped to the SG opening based on the location of highest *Defb6* expression in the epidermis. Additionally, the absence of *Krt5* / *Ptn* expression in the population does not allow its location in the basal layer of the upper HF / junctional zone while the presence of *Krt14* argues against suprabasal localization (e.g. adjacent to uHF I).	Has not been previously described. Although it has been shown previously that *Lrig1*+ cells are located in different compartments of the skin including the uHF and the SG (Page et al., 2013), and KRT79 and CST6 are expressed in the uHF (Veniaminova et al., 2013; Zeeuwen et al., 2002), these populations were never dissected on a molecular / transcriptional level.

Upper HF III (uHF III)	Marked by an uHF signature (*Krt79*, *Krt17*, *Cd44*) and a *Krt14* / *Lrig1* module. However, in contrast to basal cells in the upper HF / junctional zone (uHF IV), *Krt5* / *Ptn* are absent. Can only be mapped based on exclusion. While the presence of *Krt14* / *Lrig1* indicated basal location, the absence of *Krt5* / *Ptn* did not allow mapping to the *Krt5*(hi) uHF / junctional zone basal layer. Absence of high *Defb6* / *Cst6* levels excluded localization in the SG opening while non-expression of typical SG markers such as *Scd1* / *Mgst1* excluded mapping to the proximal half of the SG. The population is in consequence most likely located in the distal half of SG.	See uHF II.
Upper HF IV (uHF IV)	This population expressed low levels of uHF markers (*Krt79*, *Krt17*, *Cd44*) in combination with both a *Krt14* / *Lrig1* and a *Krt5* / *Ptn* basal signature. Its low level expression of *Krt79* and *Krt17*, which could both be found expressed in small numbers in the basal layer of the uHF and SG, and its positivity for both *Krt14* and *Krt5* (which excludes the basal cells of the SG) linked this population to the basal layer of the uHF including isthmus.	See uHF II. This population most likely corresponds to the main population of Lrig1+ cells described by Jensen et al., 2009 and Page et al., 2013.
Upper HF V (uHF V)	A population with a strong uHF signature (*Krt79*, *Krt17*, *Cd44*), a fading basal signature (*Krt14* / *Lrig1* / *Krt5* / *Ptn*) and low-level expression of suprabasal markers such as *Krt10* / *Sbsn*. Most likely contains cells during their transition from uHF basal (uHF IV) to uHF suprabasal (uHF VI).	Differentiation in the uHF and the presence of cells expressing both uHF markers (e.g. *Krt79*, *Krt17*) and spinous cell (*Krt10*, *Sbsn*) or granular cell markers (*Lor*, *Flg2*) has been previously described (e.g. by Veniaminova et al., 2013)
Upper HF VI (uHF VI)	Cells distinguished by an uHF signature (*Krt79*, *Krt17*, *Cd44*) in combination with a high expression of suprabasal markers (*Krt10* / *Sbsn*). Located in the 2nd cell layer of the uHF excluding the suprabasal cells which express a *Cst6* / *Defb6* / *Rbp1* signature.	See uHF IV.
Upper HF VII (uHF VII)	Cells distinguished by an uHF signature (*Krt79*, *Krt17*, *Cd44*) in combination with a high expression of terminal markers (*Lor* / *Flg2*). Could be mapped to a 3rd layer of flat, keratinized cells which line the hair canal in the upper HF.	See uHF IV.
Sebaceous gland (SG)	A population of cells, which is very distinct from all other keratinocyte populations, and distinguished by a large signature of genes mainly involved in lipid metabolism. Marked by well-established SG markers such as *Scd1*. MGST1 staining and a relative large heterogeneity in cell size suggested that this population includes both cells from the proximal basal layer of the SG and the inner parts of the gland.	Well described.
Outer bulge I (OB I)	A cell population, which is dominated by an outer bulge signature (*Postn*, *Cd34*). In contrast to OB II, it expressed slightly higher levels of *Lgr5*, *Krt24*, lower levels of *Cd34* and contained more *Gli1* expressing cells. This suggests that the cells of this population tend to be located more proximally than the cells of OB II. Furthermore, OB I cells are limited to the bulge and the cell layer between club hair and bulge.	*Lgr5*^{hi} cells located in the proximal bulge area including the hair germ are well established (Jaks et al., 2008). Most intriguingly, our single-cell transcriptional data indicate that the differences between the proximal and distal bulge populations are small and that the transition between both populations is "fluid".
Outer bulge II (OB II)	The counterpart to OB I. The cells are likely to be located more distally in the outer bulge than the cells of OB I. Cells are spread over the bulge, the cell layer between club hair and bulge as well as the club hair.	See OB I.
Outer bulge III (OB III)	A population, which, in addition to a pan outer bulge signature (*Cd34*, *Postn*), was negative for more proximal outer bulge markers (*Lgr5*, *Krt24*) and instead expressed a unique signature including *Aspn* and *Nrep*. The expression of *Gli1*, *Lgr6* and *Krt17* and the low levels of *Krt15* link these cells to a particular position at the upper edge of the outer bulge region.	Most likely corresponding to the *Gli1*+ population in the isthmus described by Brownell et al. 2011.
Outer bulge IV (OB IV)	Transient population of cells, which expressed both outer bulge (*Cd34*, *Postn*) and upper HF (*Krt79* / *Krt17*) markers. Located at the interface of bulge and upper HF compartment distal of OB III and proximal of uHF IV. Interestingly, a subset of cells co-expressed *Lgr6*.	Although spatially congruent with the isthmus *Lgr6*+ population (Snippert et al., 2010), not previously described as a transient population exhibiting both outer bulge and uHF features.
Outer bulge V (OB V)	Outer bulge population (*Postn*, *Cd34*) additionally expressing differentiation markers including *Krt10* and *Sbsn*, that could be linked to a group of suprabasal cells at the level of OB III and OB IV.	Most likely the CD34+ / ITGA6- population identified by Blainpain et al., 2004.
Inner bulge I (IB I)	Population of cells that are solely distinguished by an inner bulge signature (*Krt6a*, *Krt75*, *Timp3*, *Fgf18*). Represent most of the cells located in the inner bulge.	Well-described epidermal population (Hsu et al., 2011).
Inner bulge II (IB II)	Cells, which expressed an outer bulge signature (*Postn*, Cd34) in combination with an inner bulge signature (*Krt6a*, *Krt75*, *Timp3*, *Fgf18*). This population could be mapped to a group of basal, KRT6^{hi} and *Postn*+ cells in the outer bulge region.	The separation of the club hair bulge into two populations has not been previously described.
Inner bulge III (IB III)	Cells expressing both an inner bulge signature (*Krt6a*, *Krt75*, *Timp3*, *Fgf18*) and differentiation markers. Could be located in the upper edge of the inner bulge.	Not previously described.
T cells (TC)	Population of immune cells. Their specific gene expression (*Cd3g*, *Cd3d*) distinguished them as γδ T cells.	Well described.
Langerhans cells (LH)	Immune cells that expressed typical Langerhans cell markers (*Cd207*, *Cd74*).	Well described.

B. Previously defined murine epidermal (stem) cell populations

References	Representation in Joost et al. dataset
Krt14+/Ivl- and Krt14+/Ivl+ IFE basal cells (Mascré et al., 2012)	While the infundibular (INFU B) and upper HF (uHF IV) basal cells showed a higher *Ivl* baseline expression, *Ivl* expression was rarely found in the IFE basal populations (IFE B I – II). Instead, *Ivl* expression in the IFE was predominantly detected in the differentiating and terminally differentiated populations. It is possible that Mascré et al., 2012 targeted cells, which are in the process of transition from basal to suprabasal and show both basal and suprabasal characteristics.
Spinous layer cells (Miscellaneous)	Cells that show a spinous layer signature could be found in the IFE (IFE DII), the upper HF (uHF VI), the outer bulge (OB V) and the inner bulge compartment (IB III).
Granular layer cells (Miscellaneous)	Cells distinguished by a granular layer signature are present in the IFE (IFE K I and II) and the upper HF compartment (uHF VII).
Lrig1+ cells located in the upper HF (Jensen et al., 2009, Page et al., 2013)	*Lrig1* was expressed in a variety of populations primarily located in the upper HF and SG including uHF I, uHF II, uHF III, and uHF IV. Although most highly expressed in the basal layer, *Lrig1* molecules were also sporadically detected in differentiated uHF cells. Furthermore, low-level sporadic *Lrig1* expression was found in the IFE basal layer and cells of the outer bulge.
Mts24+ cells located in the isthmus (Nijhof, 2006)	Expression of MTS24 (*1600029D21Rik*) could be detected in all populations of the upper HF and in terminally differentiated cells of the IFE. In the upper HF, *1600029D21Rik* expression peaked in the terminally differentiated populations (uHF VI and uHF VII) and in the *Defb6*^{hi}/*Cst6*^{hi} populations uHF I / uHF II.
Lgr6+ cells located in the isthmus and IFE (Snippert et al., 2010, Füllgrabe et al., 2015)	*Lgr6* was expressed only sporadically over the whole dataset. The highest *Lgr6* expression was found in the Gli1+ cells of the outer bulge (OB III). Lgr6 expressing cells were also found in the outer bulge / upper HF transitional population OB IV and the upper HF basal population uHF IV. A distinct *Lgr6*^{hi} isthmus (sub) population could not be resolved. Neither was it possible to clearly demarcate *Lgr6*+ and *Lgr6*- populations in the IFE basal layer.
Gli1+ cells in the upper bulge (Brownell et al., 2011)	Could be identified (OB III) as cells with an outer bulge signature and a unique set of co-expressed genes (*Gli1*, *Aspn*, *Nrep*).
Krt15+ / Cd34+ mid bulge cells (Cotsarelis et al., 1990, Morris et al., 2004)	CD34 is the most prominent marker of the bulge and in our dataset we confirmed *Cd34* as a pan outer bulge marker found in all outer bulge populations. Cells of the mid bulge (CD34+ / *Lgr5*-) were most likely included in OB I, and OB II. However, it was not possible to clearly delineate the CD34+ / *Lgr5*+ and CD34+ / *Lgr5*- populations by unsupervised clustering of our data.
Krt15+ / Cd34+ / Lgr5+ lower bulge cells (Jaks et al., 2008)	See above.
P-cadherin+ cells of the hair germ (Greco et al., 2009)	Although some P-cadherin (*Cdh3)*-expressing cells could be found in OB I, OB II and OB III, it was not possible to resolve those cells as a distinct population.
Suprabasal Cd34+ / Itga6- cells (Blanpain et al., 2004)	Most likely represented by the population of cells with both an outer bulge and a spinous layer signature (OB V).
Egfl6+ bulge population which provides attachment to the arrector pili muscle (Fujiwara et al., 2011)	*Egfl6* is expressed consistently over all outer bulge populations and sporadically in the IFE. No distinct *Egfl6*^{hi} population could be resolved.
Krt6+ inner bulge cells (Miscellaneous)	Inner bulge cells formed a highly distinct 1st level cluster (Krt6+) which could be further divided into three subpopulations (IB I – III)

Table S4. Differentiation-related genes. Related to Figure 3.

List of significantly pseudotime-dependent genes. Genes are grouped according to clustering shown in Figure 3C. Within each cluster, genes are ordered from lowest to highest p-value.

I	Tgfbi, Krt14, Krt5, Mt2, Fth1, Dst, Smoc2, Itga6, Cbr2, Ifitm3, Tspo, Ccnd2, Fcgbp, S100a10, Cav1, Ftl1, Phgdh, Antxr1, Il33, Capg, Itga3, Serpinh1, Serpinb10, Scin, Itgb4, Lamb3, Slc6a6, Ly6c1, Bhlhe40, Slco2a1, Oat, Cdh13, Ly6a, Paics, Sema3c, Fbxo32, Agrn, Cxcl14, Ltbp4, Serpinb6a, Csrp1, Cyr61, Tnfrsf18, S100a6, Gja1, Jag2, Ptrf, Itgb1, Col4a6, Slc27a3, Col23a1, Ackr3, Slc2a12, Igfbp3, Crlf1, Pdlim1, Ier3, 9530053A07Rik, Col7a1, Sdcbp, Dsg2, S100a13, Ctnnal1, Sult5a1, Fgfr2, Col16a1, Ccnd1, Rassf1, H2-Q9, Rbbp8, Icam1, Dll1, Ephx1, Thbs1, Bmp4, Atf3, Gga2, Tagln2, Dsg3, Mgll, Limch1, Lama5, Slc38a2, Adamts14, Tubb5, Lamc2, Slc12a4, Phldb2, Marveld1, Arhgdib, H3f3b, Htra1, 4930523C07Rik, Gapdh, Pmp22, Frem2, Slc2a1, Efemp1, 1810011O10Rik, Fat2, Ccdc3, Snai2, Klf9, Slc4a8, Tsc22d1, Guca2a, Card10, Hmgn2, Fpgs, Sat1, Cyp4b1, Il20ra, Gm13305_loc1, Socs3, Lmna, Wls, mt-Tc, Sh3pxd2b, mt-Tf, Rab30, Ptprs, Ackr4, Man1c1, Mapkbp1, Cisd3, Anxa5
II	Ccl27a_loc2, Col17a1, Mt1, Krt15, Apoe, Clca2, Cav2, Gpx1, Gas5, Snhg1, 2410006H16Rik, Zfp36l2, Tns4, Clca4, Tmsb4x, Gpt, Cd59a, Clca1, Slc7a6, Trp63, Plp2, Myof, Gm5643_loc2, Sox6, Rnase4, Gm11974, Snrpg, Capns2, Rps12, Cyb5r3, Hmgb1, Ggta1, 1110038B12Rik, 1500012F01Rik, Slc22a4, Npm1, Acp5, Atpif1, Prdx6, Oxct1
III	Acer1, Fabp5, Gsn, Dsc3, Lgals3, Pkp1, Alox12e, Sdr16c5, Micu1, Sdc1, Tmprss4, Mt4, Dbi, Pycard, Tesc, Slco3a1, Rab25, Anxa8, 1810037I17Rik, Camsap3, Pvrl1, Tecr, 4833423E24Rik, Galnt1, Slc12a2, Crip2, Ephb6, Rab3d, Gsta2, Lgals7, Ly6d, Serpinb12, Iqgap1, Degs1, Rdh12, Clu, Eef2k, Cd109, Krt1, Tmem254b_loc2, Aldh3b2, Alox12, Smim5, Hes1, Paqr5, Man2a1, Gcat, Tmem254b_loc3, Tmem254b_loc1, Rhov, Vdac1, Pkp3, Cd9, Tacc2, Hpgds, Sfn, Mir703, Kctd15, Gata3, Atp6v1a, Anxa1, Cyb5, Mif, Npl, Serbp1, Sepp1, Pla2g4f, Atp2b4, Ap1b1, Cat, F2rl1, Bag1, Il17rc, Dock6, Fxyd3, Ell3, Lrp4, Rps2, Epha4, Tm9sf3, Rplp0, S100a11, Epn2, Ctnnd1, Pgrmc1, Bmpr2, Acad9, Atp1b3, Anxa7, Syngr2, Ikzf2, Pdcd4, Urah, Clca5, Ghr, Tsc22d3, Ache, Cst3, Acaa2, Ufsp1, Srd5a1, Zfp185, Ppia, Cyp2b10, Rassf3, Lamp2, Pou3f1, Prdx1, Pkib, Apoc1, Notch3, Serpinb3b, Tmem208, Rell1, Nod1, Ywhaq, Eef1a1, Agpat3, Mbnl2, Rpl10, Sh3bgrl3, Cyb561, 5430435G22Rik, Ralbp1, Ndufv3, Rpl14, Rps26, Nat8l, Dync1li2, Memo1, Hs3st6
IV	Itpr3, Hba-a2_loc2, Sqle, Hba-a2_loc1, Ankrd35, Sptlc3, Tmem154, Ptprf, Gjb5, Gdpd2, Fdps, Sema3d, Crabp2, Dsg1c, Itgb5, Chl1, Xpnpep1, Ttk, Cdh1, Tpm4, Mvk, Acer3, Myh14, Wdr47, Atp2b1, Cyp51, BC064078, Tmem123, Irf6, Add3, Emp1, Cdc42, Aacs, Hmgcs1, Arhgap24, Actr2, Idi1, Sc5d, Dhcr24, Lad1, Snrnp25, Pcyt1a, Tnik, Canx, Lamp1, Cyp2j6, Daglb, Ywhaz, Niacr1, Flrt1, Hspa1b, Flnb, Ldlr, Ahcyl2, Sult2b1, Actg1, Pdlim5, Mvd, Cnn2, Ermp1, Scd2, Adtrp, Magt1, Hn1, Abhd6, Slc6a19, Kif5b, Rab1, Dbnl, Gng5, Eif2s2, Eif1ad, Atp2c2, Jak1, Sar1b, Rnase1, Smagp, Hsd17b7, Pcdhgb4, Fam20b, Gm7334, Hsdl2, Gldc, Hsp90ab1, B4galnt2, Lss, Fut8, Ept1, Anxa2, Serpinb2, P4hb, H2-D1, Creb3, Prom2, 2310022B05Rik, Erbb2, Taf13, Cldn1, Mapkapk3, Map7d1, Osbpl5, Sltm, Slc16a6, Klc3, Tpm3, 2810025M15Rik, Wwp1, Palm, Actn4, Rangap1, BC031181, Cltc, Akirin1, Selt, Atp6ap1, Fam210b, Pgs1, N4bp3, Sh3d19, Rab6a, Slc9a9, Efna5, Fermt3, Ppp1r13l, Fasn, Dusp7, Ctnna1, Prpf18, Zdhhc5, Clic1, Rmnd5b, Llgl2, 2210404O07Rik, F11r, Spry1, Sypl, Plekha6, Rpn2, Tm7sf2, Adam17, Arf4, Timm23, Arpc2, Rai14, Tubb2b, Mtch1, Hsd17b12, Neurl3, Nans, Nsun2, Neu3, Atf4, Fam111a, Gnb1, Slc5a1, Gnb2, Atp11b, Pvrl2, Tjp1, Naa15, Gm15987, Eif3c, Manf, Fnbp1, Ktn1, Ralgapa1, Rdh11, Cyfip1, Sbno1, Eif3a, Shisa5, Tmod3, Ubxn4, Wwc2, Flrt2, Rhbdf1, Rb1, Sec24a, Chit1, Ptma, Phactr2, Prob1, Herc4, Aff4, Prodh, 1110008P14Rik, Fam63b, Cept1, Zmat3, Fkbp1a, Il15, Hsp90b1, Nkd2, Slc39a6, Eif6, BC016579, Pon3, Kdm1b, Vimp, Sumf2, 2510039O18Rik, Gapvd1, Ywhae, Dlg2, Coro1c, Tram1, Mapre1, Slc24a3, Ddx5, Vps41, Srebf2, Itgav, Pla2g4a, Snx3, Bcap31, Psma2, Prrc2b, Tmem35, Scrib, Cand1, Kpna4, Cnbp, Psmb3, Tmem41a, Tet3, Aktip, Emc10, Psmd3, Axl, Nedd8, Slain2, Gab1, Sacm1l, Cdc14b, Ier3ip1, Gnas, C77080, Fgfbp1, Atl2, Nfatc2, Gipc2, Gfod1, Phb, Akr1a1, Vps37a, Ccdc25, Sec31a, Tmem33
V	Sbsn, Krt10, Krtdap, Dmkn, Dsg1a, Calm4, Dsp, Ly6g6c, Gm94, Elovl4, Pvrl4, Susd2, Kctd12, Krt77, Fam25c, Pdzk1ip1, Skint5, Emp2, Ptgs1, Lypd3, Prdx5, Calm5, Gltp, Dapl1, Tmem45a, Ly6g6e, Acsl1, Skint9, Krt78, Them5, Scd1, Perp, Serpinb5, Eppk1, Clip4, Dynll1, Dstn, Kif21a, Skint6, Kif1c, Gng12, Ahnak, Evpl, Pla2g16, Ptgr1, Scel, Cdkn1a, Prnp, Jup, Hebp2, Cebpa, Ablim1, Mfsd6, Cmtm6, Nupr1, Lsr, Sgpp1, Sptssa, Gpr115, Itm2b, Cyp4f39, Skint4, Dsg1b, Me1, Ggh, App, Arpc1a, Pak3, Rab11a, Casz1, Tanc1, Trim29, Ptbp3, Gjb3, Mpzl2, Scamp1, Acly, Ace2, Chmp5, Rcn1, Sort1, Zfp750, Pon2, Rab10, Fuca1, Kank1, Pgrmc2, Pea15a, Nrarp, Metrnl, Nfe2l1, Rab18, Mocs2, Dynlt3, Fut2, Ezr, Tuba4a, Micu2, Ank3, Lrrc58, Tmbim6, Tmed2, Pim3, Ormdl2, Sppl3, Rhbg, Al314180, Plekhn1, mt-Co3, Tollip, Ppp1cb, Sptan1, Arf1, Eif4g2, Notch2, Tom1, Cs, Elovl7, Pglyrp4, Slc6a8, Kif1b, Creg1, Abhd5, Polb, Erlin1, Kremen1, Nhp2, Ptplb, Psen1, Ss18, Golgb1, Ppfia3, Grhl3, Esrp1, Acap2, Scamp2, Hdac5, Rnh1, Cltb, Fam213b, Tuba1a, Nlrc5, Map4k4, Rfwd2, Sptbn2, Atp6v0b, mt-Nd4, Psmd7, Mpp7, H2afy, Tmem79, Wbp1l, Ptpra, Ppp2ca, Nagk, Atp6v0e, Spint2, Hr, Mfsd5, Zfp266, Mob4, Eif5a, Sun2, Fut1, Gpr125, Capn5, Tprgl, Cbx4, Fam57a, Klk11, Baz2a, Ppp1r2, Csnk1a1, Wdfy3, Itfg3, Rnasek, Rtn3, Oas1f, Chmp2b, Rab14, Il20rb, Sphk2, Slc31a2, Ywhab, Stard5, Laptm4a, Acbd3, Dusp22, Atp6v1g1, Jhdm1d, mt-Rnr1, Wnk1, Pdlim2, Ppp2r1a, Ubqln1, Acsbg1, Psmd2, Atp5d, Arf5, Cyth1, mt-Atp6, Ppp2r5e, Rhbdf2, Atp6v1e1, Srgap2, 0610009D07Rik, Abhd12, Tsc22d4, Cope, Adh1, Gp1bb, Ocln, Ovol1, Acaca, 6430548M08Rik, Rab2a, Ddx3x, Tmem134, Aftph, Arap2, Atp7a, Sptlc1, Synrg, Tmbim1, Ttc39c, Mycbp, Kdm5b, S1pr5, Ostf1, Nsdhl, Usp20, Spop, Mta3, Rer1, Usp32, Tspan9, Zmiz1, Lars2, Erp29, Git1, Ztbt7a, Cldn25, Cox7c, Tmem66, Npepps, Clint1, mt-Co2, Pcyox1, Ugt1a2, Smpdl3a, Tmem159, Car13, Ist1, Arpc5, Larp1, Tec, BC048507, Pou2f3, Dctn1, Psmc2, Tmem109, Gbf1, Fuca2, Fbxo6, Usp4, Fbxl3, Aadacl2, Dync1i2, Zmiz2, mt-Ti, Rap2c, Kdsr, Napg, Chmp1a, mt-Nd3, Tmem141, Tiam1, Slc9a3r1, Znrf2, Tmem254c_loc1, 1110057K04Rik, S100a16, Bnipl, Rapgefl1, Pex3, Yif1a, Psmc1, Nsmaf, Rnpepl1, Hdlbp, Zfp91, Tmem254c_loc3, Mtmr1, Ube2v1, Uchl3, Tpd52, Zc3h15, Pls3, Ppp2cb, Dynlrb1, Mcur1, Agfg1, Chmp1b, Tcf25, Mcu, Arpc5l, Wwp2, Cbfb, Ghitm, Emc3, Ppp3r1, Ndufc1, Gnptg, Herc2, Ctage5, Crb3, Lta4h, Celsr2, Irak2, Ppp2r2a
VI	Spink5, Psapl1, Ssfa2, Orm1, Clic3, Hal, Dgat2, Calm1, Mboat2, Skint10, Homer2, Pla2g2f, Ank, Ivl, Agpat4, Grhl1, Ephx3, Ppl, Skint3, Dkkl1, Klk8, Ano9, Tspan8, Plxdc2, Hiatl1, Hspb1, Elovl6, Fam129b, Gjb4, Ybx3, Lipm, Dnase1l3, Tubb2a, Trim2, Klk5, Gdpd3, Blmh, Ccdc64b, Gsdma, Adipor2, Csnk2a2, Rgs2, Myo18a, Calm2, Mllt4, Defb1, Ift20, Cd82, Sdr16c6, Ptpn21, Macf1, Emb, Dsc1, Tmem43, Csnk2a1, Faah, 'Sept8', Rarres1, Abtb2, Klf4, Ccdc50, Sc4mol, Adrbk1, Lgalsl, Pmvk, Pepd, Rhou, Cldn4, Sec11c, Ly6e, Adipor1, Fam107b, Skint11, Ano10, Setd8, Tuft1, Mal2, Hmgcr, Sgms1, Tprg, Sprr1a, Tacstd2, Ndel1, AA986860, Nudt17, Tmed5, Map2k4, Ifitm6, Desi1, Usp7, Cpeb2, Gm11992, Gdi2, Areg, Prdx2, Cox17, Rab27b, Mall, Fdft1, Rdh9, Xbp1, Sorbs1, Stx19, Cast, Spint1, Stk40, 4833439L19Rik, Ceacam19, Klf3, Chmp4b, Inpp5b, Liph, Paip2, Adh6a, Actb, Reep3, Ifrd1, Ppfibp1, Dcun1d1, Pdzd11, Cebpb, Otop3, Spag1, Pld3, Gsk3b, Ptdss1, Arf6, Arhgef10l, Gba, Sp6, Eif1, Rdh1, Atp6v0d1, Sft2d2, Atp6v1d, Stard4, Nt5e, Slc22a23, Bicd2, Gal3st1, Ttc7, C130079G13Rik, Arsa, Tmppe, Gstm5, Scnn1a, Csde1, Golga4, Dpp3, Ttc39b, Defb6, Tgm5, Iffo2, Cab39, H2-T9, Prrc1, Gltpd1, Cep170b, Fam188a, Pdcd6ip, Rhod, Otud1, Kifc3, Cdr2, Chmp2a, Plxnb2, Ndrg2, Il1rn, Ube2b, Jarid2, Atp6v1b2, D15Ertd621e, Grina, Gng13, H2-K1, Eea1, Slc11a2, Tjp3, Clcn3, Atp2a2, Eif4g1, Riok3, Erp44, Sh3bgrl2, Ndfip2, Psmc5, Tmem65, Tmed3, Atp6v0c-ps2, Mknk2, Cstb, Ankrd27, Sphk1, Rnf39, Ptges3, Vcp, Cmpk1, Atp9a, Lyst, Ppfibp2, Baiap2, Usp25, Icmt, Clip1, Slc6a9, Atp6v0a4, Pnpla8, Tyro3, Hspa5, Slk, Ndrg1, Gpr111, Cmtm8, Ssh3, Fam135a, Abcd3, Arl4d, Clasp1, Cyb5r4, Tspan5, Sqstm1, Lrrfip2, Plekha5, St14, Krcc1, Suds3, Usp54, Pum2, H2-Ab1, Pfdn1, Mpzl3, Suco, Syngr1, Stk17b, Tsg101, Rnf13, Ppp1r12a, Golga2, Ykt6, Nipbl, Csrnp2, Cdkn2b, Itpkb, Akap13, Mapre2, Dohh, Tmem50a, Camk2d, Cxcl16, Abi1, Pcdh1, Ilvbl, Mrgprb3, Arl5b, Tmem170b, Bsg, Nfat5, Egr2, Cryl1, Lrrc8e, Maff, Tob2, Slc39a2, Tmem9b, Vamp8, BC100530, Gch1, Pgd, Mad2l2, Gm1821_loc1, Hivep1, H2-Eb1, Mir22hg, Ccdc6, Rab32, Tmed10, Nras, H2-Aa, Rasal2, Leprel4, Selk, Rbx1, Napa, Fbrs, Lig3, Psmc3, Derl1, Rgs1, Hopx, Gns, Golt1b, Srgap1, Il18r1, Guk1, Frmd4b, Gal3st4, Capza1, Mafk, Rec8, Tmed9, Arhgap17, Irf2bp2, Senp2, Pitpna, Glrx2, Man1a2, Emc4, Ell2, Uba1, Vamp2, Mcf2l, Slc30a7, Macc1, Rel, Tmem167, Mtmr6, Mgea5, Atxn7l3b, Mroh6, Hyou1, Errfi1, Lrrc8a, Leprotl1, Kbtbd4, Bach1, Wsb2, Cyth3, Map2k2, Arhgef12, Uso1, Ogdh, Rora
VII	Asprv1, Flg2, Lor, Elovl1, Lce1m, Cst6, Klk7, Pof1b, Crct1, Krt23, 2310002J15Rik, Ide, 2310050C09Rik, Gabarapl2, Nrd1, Ndufa4, Serpinb7, C2cd2, Bpifc, Il1f5, Cnn3, Cnfn, Cdsn, Hrnr, Tgm1, Agpat5, Abca12, Ggct, Eps8l1, Nlrp10, Ugcg, Map2, Ctsd, 2310042E22Rik, Alox12b, Kprp, Tmem54, Arhgef5, Ereg, Sdcbp2, Sec62, Nipal4, Laptm4b, Ctnnbip1, Pcsk6, Osbpl1a, Gnai3, Trex2, Wfdc5, Pvrl3, Serpinb8, Rab5a, Cers3, Vgll3, D8Ertd738e, Rab11fip1, Ubl3, Amotl2, Slc46a2, Ctsh, Atp6v1c2, Tex264, Txnrd1, Mprip, Calml3, Aloxe3, Tmprss13, Cyp2b19, Srpk1, Daam1, Zdhhc9, Cpm, Tead1, Wipf2, 'March7', Rnf11, Nccrp1, Cgn, Hspb8, Abca5, Cldn23, Dhcr7, Il22ra1, Zfand6, Kctd11, Myo6, Mapk3, Slc6a14, Pcmtd1, Dhrs1, 'March8', Serinc3, Mxi1, Map1lc3a, Stk4, Gan, Tmem40, Ypel2, Fcho1, F3, Yod1, Sh3kbp1, Cobl, Fam89a, Rab8b, Por, Ctsa, Epcam, Wdr26, Nipal1, Ifi30, Smpd1, Cln8, Cul3, Tob1, Hook3, Pdzd2, Psmd8, Appl2, Peli1, Tuba8, B2m, Arhgap23, Pnpla1, Fndc9, Dlgap4, Hectd1, 4933426M11Rik, Hbp1, Slc30a4, AI661453, Mxd1, Taok1, Rps6ka6, Hip1r, Zswim4, Ptrh1, Pim1, Hdgf, Cds1, Blnk, Gramd3, Pard3, Eif5, Card14, Zfp706, Slc15a1, Abca3, Chp1, Arg1, Ypel3, Hars, Kctd5, Shroom3, Ube2d3, Atp6v0a2, Asah2, Sidt2, Capn9, Sertad1, Rmnd5a, Endou, Ccdc120, Ier5, Fis1, Ppp1cc, Fhdc1, Tma7, Map3k9, Sucla2, Casp14, Ppp1r13b, Slc27a4, Mid1ip1, Faf1, Tsc22d2, Pnpla3, Dusp16, Prdm1, Itsn2, BC021767, Stx12, Eps8l2, Tpm1, Lrp10, Chmp3, Lmo7, Cited4, Slc19a2, Ube2k, Snx2, Ttpal, Pdxdc1, Huwe1, Vasn, Myo5b, Krt79, Xkrx, Cited2, Aim1, Midn, Tmcc3, Fam84a, Ranbp9, Pink1, Gdpd1, Cyfip2, Sppl2a, Gpr87, Tab2, Aim1l, Acpp, Mreg, Abhd12b, Usp6nl, Slmo2, Tnfaip8, Spag9, Endod1, Zfand3, Fam73b, Golga7, Dnaja4, 1600029D21Rik, Usp16, Epn3, Abhd17c, Ankrd44, Prkch, Paf1, Rtfdc1, Eps15, Tjp2, Jund, Ypel5, Tnfaip1, Shfm1, Litaf, Ammecr1, Pacsin2, Mtss1, Dnaja1, Arhgap29, Mlf2, Ehd1, Dab2ip, Sdr42e1, Dusp14, Sdf2, Amotl1, Arrb1, Bin3, Sash1, Trp53inp2, Wipi2, Eif4h, Rnf208, Mbnl1, Neat1, Max, Prkx, Lypd6b, S100a14, Pip5k1a, Eif4e, Tpra1, Sec14l1, Hmgcl, Ube2n, Sh3gl1, Med4, Wasf2, Timp2, Rorc
VIII	Csta, Ctsl, 1100001G20Rik, A030009H04Rik, Anxa9, Sdr9c7, Lypd5, Ube2r2, Serpina12, Elmod1, Nt5c3, Aif1l, 2310014L17Rik, Krt17, Hsbp1l1, Il18, Mapk13, 2010109I03Rik, Lce1c, Spns2, Btc, Lce1a2, Slurp1, Lipk, Lce1b, Smpd3, Ecm1, Dusp18, Lce6a, Lipn, Bnip3l, Hmox1, 2200002D01Rik, Lce1a1, Il1f9, Cpe, Tmprss11f, Gas7, Lrrc28, Minpp1, Msrb1, Id4, Lphn2, Glrx, Hexb, Map2k3, Slc35e4, Il1r2, Malat1, Fchsd2, Usp2, Rnf222, Gm14137, Sod3, Tgfa, Plk3, Cpa4, Spsb3, Stk10, Myzap, Atg9b, Scgb1a1, Smap2, Gng4, Prss8, Tmem229a, Nck2, Fmo2, Ttc9, Lce1l, Ano8, Psors1c2, Scpep1, Actn1, Slc38a3, Crlf3, Rassf5, Robo1, Tnfrsf11b, Pqlc1, Map4, Tmem177, Klk9, Esyt3, Acot7, Kctd4, Rab3b, Nkpd1, Heca, Wdr45, Lce1f, Dnase1l2, Calcb, Ankrd23, Ralgds, Oser1, Piezo1, Il1rap, Tmem56, Ubtd1, Rapgef2, Lce1d, Ckb, Krtap3-2, Cyp26b1, Gm2a, Rnf103, Thop1, Gramd1c, Dennd1b, Ccl20, Lce1e, Cyp4v3, Glcci1, Lce1k, Lce1i, Ormdl1, Tmem120a, Lce1j, Pcdh7, Lce1g, Clec2g, Ppif, Abcg1, Strip1, Chic2, Gpld1, 3300005D01Rik, Gpsm1, Fam214a, Gba2, Myo1b, Myo10, Lrch1, Btbd3, Plcxd2, Fam110a, Map1lc3b, Mdfic, Blzf1, Asap3, Lnx1, Plcxd1, Map3k8, Ak1, Bnip3, Lpin3, Krt80, Fmn1, Tnfrsf21, Cnst, Arhgap32, Rassf7, Ubash3b, Trim16

Table S5. Spatial axis related genes. Related to Figure 4.

List of significantly pseudospace-dependent genes. Genes are grouped according to clustering shown in Figure 4C. Within each cluster, genes are ordered from lowest to highest p-value.

I	Krt14, Mt2, Chit1, Mt1, Ccl27a_loc2, Rplp1, Rps12, Tnfrsf19, Rps16, Gnb2l1, Rps20, Rps15a-ps4, S100a11, Rps28, Gm6654, Rpl39, Rpl12, Rps3a1, Arhgdib, Rpsa, Txndc17, Rps24, Rpl29, Gpx1, Rpl4, Rpl23, Rplp2, Fau, Rpl18a, Rps3, Gm6402, Tnfrsf18, Rps29, Rpl5, Nme2, Nfkbiz, Rps6, Gm5643_loc2, Rpl8, Rps5, Gm13139_loc1, Hmgcs1, Rpl13a, Eef1b2, Rpl32, Slco2a1, Gm13826, Klf6, Rpl35, Eef2, Gm13139_loc2, Rpl31-ps12, Rpl24, 2410006H16Rik, Rps25, Htra1, Rps13, Cd55, Rps18, Fgfbp1, Dapl1, Calm5, Cox4i1, Gpt, Jun, Rpl36al, Limch1, Naca, Npm1, 1810037I17Rik, Dstn, Hspa5, Atox1, Pkp4, Emp1, Psme2, Gm5643_loc1, Ptger4, 1500012F01Rik, Atp5h, Litaf, Mif, LOC100861978_loc4, Btf3, Gm15421
II	Serpinb2, H3f3b, Il1r2, Avpi1, Fth1, Il33, Anxa2, Ly6a, Tnfaip3, Sat1, Wnt3, Thbs1, Gata3, Ly6c1, Itga6, Gja1, Ifngr1, Slc22a4, Adamts14, Actg1, Gdpd2, Pak6, Slc2a12, Krt16, Serpina3h, Tsc22d1, Fbxo32, Atf3, Itm2b, Adrb2, Ets2, Antxr1, Cyr61, Gm4832, Sfn, Sema3c, Mgll, Krt5, Pde4b, Sema3d, Chl1, Ipmk, Col23a1, Tubb4b, Wdr65, S100a10, Il22ra2, Wee1, Cbr2, Ackr3, 1810011O10Rik, Il20ra, Zfp36l2, Dst, Rnase4, Ifi27, Tppp3, Cpxm2, Phgdh, Igfbp3, Bzw1, Hmgn1, Cks2, Man1c1, Fam162a, Nr4a1, Bmp4, Ifi202b, Higd1a, Slc27a3, Ppp2r2c, Abcb1b, Ubc, Itga3, Crlf1, Cd59a, Efemp1, Prdx6, Gnai1, Tfap2c, Serpinb10, Fam25c, Eif1a, Atp5l, Fam213b, Il6ra, Rhoa, Atf4, Psat1, Dusp1, Klk11, Ptges, Klf2, Snhg1, Dnajb1, Wnk2, S100a13, Epgn
III	B2m, H2-K1, Ifitm3, Tacstd2, H2-Q9, Ly6d, Clca2, Fcgbp, Smoc2, H2-L, H2-D1, Scin, Clca1, Clca4, Aldh3a1, Bhlhe40, Ly6e, Atpif1, Ccnd2, Anxa1, Klf5, Calm2, Cstb, Ahr, Ptma, Oat, Ahnak, Rps9, Ivns1abp, Oas1f, Ptpn14, Nfkbia, Xist, Hmgb2, Lrig1, H2-Q6, Fam134b, Plbd1, 9530053A07Rik, Pkp1, Acsbg1, Tap1, Ifrd1, Psmb9, Slc2a1, Gm15987, Slco2b1, Tnfaip8, Car12, Cnn2, Neurl1b, Nlrc5, H2afz, Cdh1, Card10, Sik1, Pkib, Itpr2, H3f3a
IV	Krt17, Cst6, Defb6, Krt79, Fst, Aadac, Gsn, Ly6g6c, Gstm5, Sostdc1, Gata6, Epcam, Psapl1, Marcks, Efnb2, Pdzk1ip1, Tm4sf1, Pthlh, Skint4, Emp2, Skint3, Bmp7, Alox12e, Cyp1b1, Krtdap, Hes1, Tmem45a, Cxadr, Lphn2, Cyp27a1, Sprr1a, Apoe, 1600029D21Rik, Acsl4, Lmo7, Serpina3g, Klk7, Rbp1, Cwh43, Defb1, Susd2, Aldh3a2, Lgals7, Camk2n1, Tle1, Klhl8, Ttc39c, Pof1b
V	Ank, Fam101b, Fgfr1, Alcam, Gas1, Sema3e, Vwa2, Grem1, Aspn, Cspg4, Cd200, Nrep, Ltbp1, Moxd1, Robo2, Igfbp4, Wif1, Steap4, Egfl6, Runx1, Col5a2, Crim1, Nrbp2, Nudt4, Tgfb2, Tnfrsf11b, Cdc42ep3, Cald1, Gli1, Megf6, Cgnl1, Bdnf, Hoxc8, Boc, Hk2, Rbms3, Dab2, Mbnl1, Ndufa11, Crispld1, Gli2, Gpr125, Fam83d
VI	Cd34, Postn, Tgm5, S100a4, Lhx2, Sfrp1, Dkk3, Sparc, Col8a2, Fzd2, Shisa2, Col6a1, Lgr5, Slc6a6, Ctgf, Id2, Pdzrn4, Il31ra, Cxcl14, Col17a1, Krt24, Crip1, Adamtsl4, Gm973, Duoxa1, Ecm1, Igfbp5, Hmcn1, Trpv4, Ltbp2, Cadm1, Trabd2b, Ism1, Igfbp7, Fbln1, Kcnk2, Tns1, Col18a1, Ltbp3, Tbx1, Fgf1, Itm2a, Chat, Kcnma1, Scx, Setbp1, Slc38a2, Ppap2a, Id3, Fstl1, Smarca2, Col6a2, Cited2, Tgfbi, Duox1, Sardh, Gfra1, Mfge8, Cck, Nog, Col12a1, Golim4, Agrn, Sh3rf1, Dpy19l1, Prss23, Pfn2, Angptl2, Gfra2, Flrt2, Tnc, Il11ra1, Gcat, Serpinh1, Myoc, Sncg, Lrrn3, Emb, Angptl7, Cables1, Irx5
VII	Sox9, Aqp3, Wfdc3, Timp2, S100a6, Calml3, App, Sbsn, Nfib, Flnb, Actn1, Ptn, Nt5e, Serpinb11, Foxp1, Thsd1, Gpc6, Plxna2, Fzd1, Dmkn, Tfap2b, Nfatc1, Mllt4, Foxc1, Prlr, Aplp2, Hr, Wwp2, Fam132a, Pdzrn3, Tmtc1, Fzd7, Macf1, Tacc2, Tenm2, Zmiz1, Vdr, Casz1, Txn1, Lrrfip1, Ptprk, Efna5, Pi15, Nr3c1, Ptprf, Capn2, Atp13a2, Fzd3, Ppap2b, Sdc1, Lgr4, Hoxc13, Myo9a, Sdhb, Dapk2, Ctnnb1, Pdlim3, Mgst1, Rnf152, Ndrg2, Cbx6, Pcsk6, Vwa1, Dsp, Bnc2, Camsap3, Pik3r1, Tcf12, Klhl29, Myh14, Filip1l, Igdcc4, Hopx, Ftl1, Cd47, Mitf, Acot1, Grip1, Itgb5, Trim2, mt-Tp, Nfat5, Ssfa2, Arhgap44, Mxi1, Ssbp3, Pkd1, Id4, Itpr3
VIII	Spink5, Timp3, Krt75, Cst3, Cyp26b1, Fam167a, Fgf18, Krt6a, Cryab, Bmp2, Adcy1, Arg1, S100a1, Cdsn, Fam84a, Tspan2, Endod1, Cd24a, Perp, Krt15, Dap, Lima1, Ptgds, Fxyd6, Pvrl4, Sh3kbp1, Lypd3, Hspa2, Slc45a3, Fam26e, Dlgap4, Atp6v0e2, Dclk1, St3gal4, Sgk1, Fmn1, Clic3, Them5, Homer2, F3, Gm2a, Sulf2, Tmsb4x, Fgfr3

Table S6. Marker genes – Stem cell analysis. Related to Figure 6.

Lists of genes that are most highly expressed over *Baseline* (vs. *Baseline,* left column) or the second highest population (vs. other groups, right column). Identification of genes is based on negative binomial regression of stem/progenitor marker expressing populations. In the case of *Lgr5*, *Cd34*, *Gli1*, *Lgr6*, *Lrig1* and *Krt14* expressing cells, each population was compared to either a shared *Baseline* or all other populations. For each population, genes whose population-specific expression coefficient exceeded the *Baseline* coefficient (left column) or all other populations-specific coefficients (right column) with 95% posterior probability and which show the largest gap to the *Baseline* / the second highest population (difference between the 25th percentile of the population-specific coefficient and the 75th percentile of the *Baseline* / the second highest population) are listed.
The SCM+ basal cells were compared to SCM− basal cells or a *Baseline* shared by both populations. Genes which exceeded the *Baseline* coefficient (left column) or the SCM− basal cell coefficient (right column) with 90% posterior probability and which show the largest gap to the *Baseline* / the SCM− basal cells (difference between the 25th percentile of the population-specific coefficient and the 75th percentile of the *Baseline* / the SCM− basal cells) are listed.

(Supplied as Excel file: Joost Table S6.xlsx)

Table S7. Immunohistochemistry and single molecule FISH stainings. Related to STAR Methods.

Listed are the markers used for the validation and localization of the defined cell populations, the respective number of mice and analyzed images, and the corresponding figures in the manuscript.

Population	Clustering level	Staining to identify populations based on our sequencing data	Number of mice	Number of images taken (HF / HF+IFE / IFE)	Number of images showing the respective population	Corresponding figure in manuscript
IFE BI	2	KRT14(hi)/*Thbs1*(hi)	3	1 / 15 / 11	25 out of 27	2E
IFE BII	2	KRT14(dim)/*Thbs1*(lo)	3	1 / 15 / 11	25 out of 27	2E
INFU B	2	*Postn*(dim)	4	25 / 78 / 0	61 out of 103	2E
IFE DI	1	KRT10(dim)/PTGS1(dim)	2	0 / 7 / 0	7 out of 7	S2J
IFE DII	1	KRT10(hi)/PTGS1(hi)	2	0 / 7 / 0	7 out of 7	S2J
IFE KI	1	LOR(dim)/*Flg2*(dim)	3	0 / 11 / 5	16 out of 16	S2J
IFE KII	1	LOR(hi)/*Flg2*(hi)	3	0 / 11 / 5	16 out of 16	S2J
uHF I	1	KRT17(lo)/KRT79(lo)	3	0 / 17 / 0	16 out of 17	S2J
uHF II	1	KRT17(dim)/KRT79(dim)	3	0 / 17 / 0	16 out of 17	S2J
uHF III	1	KRT17(hi)/KRT79(hi)	3	0 / 17 / 0	16 out of 17	S2J
uHF I	2	KRT14(lo)/*Cst6*(hi) and KLK10 to localize	3 for Krt14/ Cst6; 3 for Klk10	0 / 16 / 0 for Krt14/Cst6 0 / 15 / 0 for Klk10	16 out of 16 for Krt14/Cst6; 14 out of 15 for Klk10	2E, S3L
uHF II	2	KRT14(hi)/*Cst6*(hi/dim)	3 for Krt14/ Cst6	0 / 16 / 0 for Krt14/Cst6	16 out of 16 for Krt14/Cst6	2E
uHF III	2	can't be stained*		not applicable	not applicable	
uHF IV	2	KRT14(dim)/*Krt79*(lo)	3	3 /12/ 0	15 out of 15	S3L
uHF V	2	KRT14(dim)/*Krt79*(hi)	3	3 /12/ 0	15 out of 15	S3L
uHF VI	2	KRT10(hi)/KRT79(hi)	3	0 / 14 / 0	11 out of 14	S3L
uHF VII	2	LOR(hi)/KRT79(hi)	3	0 / 20 / 0	17 out of 20	S3L
SG	1	MGST1(pos)	3	0 / 18 / 0	18 out of 18	1F, S2J
OB	1	CD34(hi)/*Postn*(hi)	2	7 / 7 / 0	14 out of 14	S2J
OB I	2	*Lgr5*-EGFP(hi)/*Postn*(hi)	4 for Postn; 2 for Lgr5**	25 / 78 / 0 for Postn; 43 / 0 / 0 for Lgr5	95 out of 103 for Postn; 39 out of 43 for Lgr5	S3L
OB II	2	*Lgr5*-EGFP(dim)/*Postn*(hi)	4 for Postn; 2 for Lgr5**	25 / 78 / 0 for Postn; 43 / 0 / 0 for Lgr5	95 out of 103 for Postn; 39 out of 43 for Lgr5	S3L
OB III	2	KRT15(lo)/*Postn*(hi)	3	0 / 16 / 0	9 out of 16	2E
OB IV	2	*Postn*(hi)/*Krt79*(dim)	4	5 / 20 / 0	15 out of 25	2E
OB V	2	KRT10(hi)/*Postn*(hi)	3	1 / 21 / 0	18 out of 22	2E
IB I	2	KRT6(hi)	4	6 / 24 / 0	30 out of 30	1F, S2J
IB II	2	KRT6(hi)/*Postn*(hi)	3	5 / 20 / 0	19 out of 25	2E
IB III	2	KRT6(hi)/*Krt10*(hi)	2	3 / 17 / 1	14 out of 20	2E
TC	1	CD3(pos)	3	0 / 25 / 0	24 out of 25	1F, S2J
LH	1	CD207(pos)	2	0 / 10 / 0	10 out of 10	1F, S2J

Legend

hi = high; lo = low; pos = positive

IHC / RNAscope

Populations stained together

* was located via exclusion of positive stainings

** costaining was technically not possible