Cell Reports, Volume *30*

Supplemental Information

Defining Epidermal Basal Cell States during Skin Homeostasis and Wound Healing Using Single-Cell Transcriptomics

Daniel Haensel, Suoqin Jin, Peng Sun, Rachel Cinco, Morgan Dragan, Quy Nguyen, Zixuan Cang, Yanwen Gong, Remy Vu, Adam L. MacLean, Kai Kessenbrock, Enrico Gratton, Qing Nie, and Xing Dai

SUPPLEMENTARY FIGURES

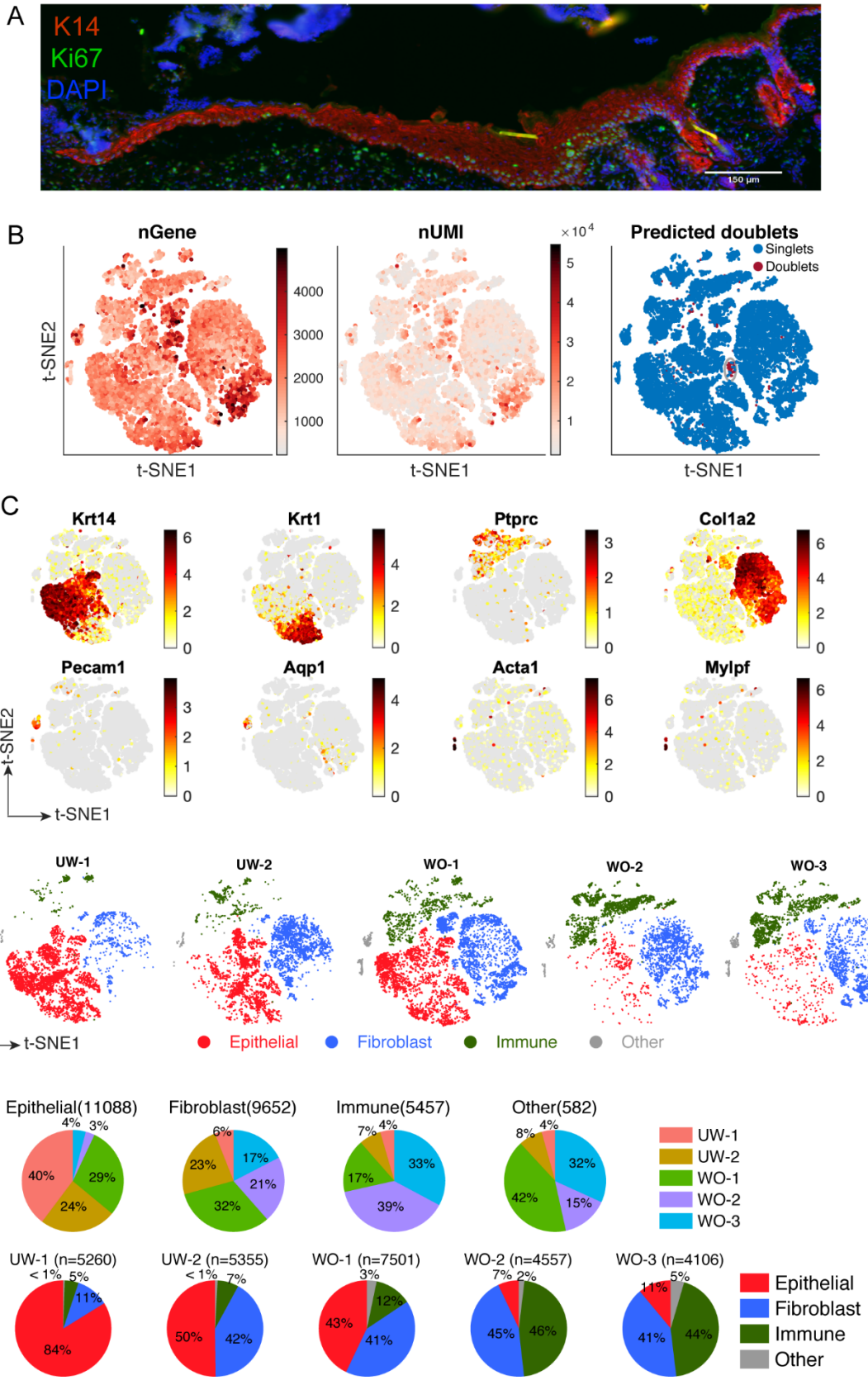

Figure S1: "Combined" analysis of all five (two UW and three WO) samples . Related to Figure 1 and STAR Methods.

A. Ki67 staining for proliferating cells in equivalent sample taken for scRNA-Seq, to enable delineation of the wound proliferative zone for subsequent analysis. Scale bar: 150 μm.
B. Quality control of cells from all five samples. On the tSNE space (same as in Figure 1C), cells are colored by nGene (left), nUMI (middle), or predicted singlet (blue) and doublet (red) identity. The predicted potential doublets (highlighted by a grey oval) are fibroblast-like cells (see Figures S1C and S2E).
C. Feature plots of major cell type/populations from all five samples. Cells are colored based on the normalized expression levels of each gene. Epithelial/epidermal cells are marked by *Krt14* (basal) or *Krt1* (spinous) expression. Fibroblast population is marked by *Col1a2* expression. Immune cell population is marked by *Ptprc* expression. Endothelial population is marked by *Pecam1* and *Aqp1* expression. Skeletal muscle population is marked by *Acta1* and *Mylpf* expression.
D. tSNE plot for each individual replicate (two UW and three WO) with the same tSNE architecture as in Figure 1C.
E. Percent contribution from each individual sample of each major cell type (top) and percent contribution from each major cell type of each individual sample (bottom). The number of cells for each individual sample or major cell type is indicated in parenthesis.

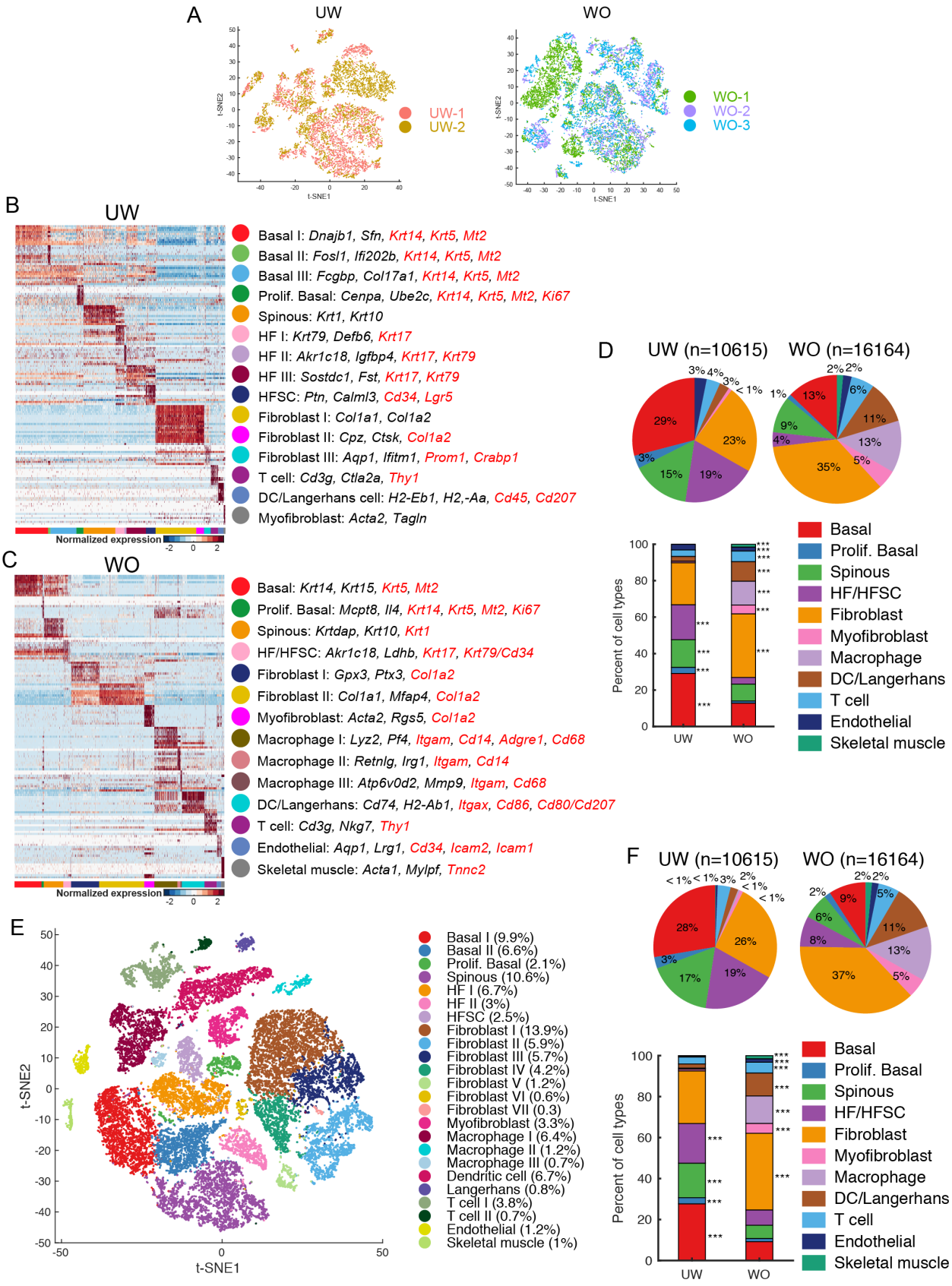
A
UW
WO
t-SNE1
t-SNE2
UW-1
UW-2
WO-1
WO-2
WO-3
B
UW
Basal I: Dnajb1, Sfn, Krt14, Krt5, Mt2
Basal II: Fosl1, Ifi202b, Krt14, Krt5, Mt2
Basal III: Fcgbp, Col17a1, Krt14, Krt5, Mt2
Prolif. Basal: Cenpa, Ube2c, Krt14, Krt5, Mt2, Ki67
Spinous: Krt1, Krt10
HF I: Krt79, Defb6, Krt17
HF II: Akr1c18, Igfbp4, Krt17, Krt79
HF III: Sostdc1, Fst, Krt17, Krt79
HFSC: Ptn, Calml3, Cd34, Lgr5
Fibroblast I: Col1a1, Col1a2
Fibroblast II: Cpz, Ctsk, Col1a2
Fibroblast III: Aqp1, Ifitm1, Prom1, Crabp1
T cell: Cd3g, Ctla2a, Thy1
DC/Langerhans cell: H2-Eb1, H2,-Aa, Cd45, Cd207
Myofibroblast: Acta2, Tagln
Normalized expression
C
WO
Basal: Krt14, Krt15, Krt5, Mt2
Prolif. Basal: Mcpt8, Il4, Krt14, Krt5, Mt2, Ki67
Spinous: Krtdap, Krt10, Krt1
HF/HFSC: Akr1c18, Ldhb, Krt17, Krt79/Cd34
Fibroblast I: Gpx3, Ptx3, Col1a2
Fibroblast II: Col1a1, Mfap4, Col1a2
Myofibroblast: Acta2, Rgs5, Col1a2
Macrophage I: Lyz2, Pf4, Itgam, Cd14, Adgre1, Cd68
Macrophage II: Retnlg, Irg1, Itgam, Cd14
Macrophage III: Atp6v0d2, Mmp9, Itgam, Cd68
DC/Langerhans: Cd74, H2-Ab1, Itgax, Cd86, Cd80/Cd207
T cell: Cd3g, Nkg7, Thy1
Endothelial: Aqp1, Lrg1, Cd34, Icam2, Icam1
Skeletal muscle: Acta1, Mylpf, Tnnc2
Normalized expression
D
UW (n=10615)
WO (n=16164)
Percent of cell types
Basal
Prolif. Basal
Spinous
HF/HFSC
Fibroblast
Myofibroblast
Macrophage
DC/Langerhans
T cell
Endothelial
Skeletal muscle
E
Basal I (9.9%)
Basal II (6.6%)
Prolif. Basal (2.1%)
Spinous (10.6%)
HF I (6.7%)
HF II (3%)
HFSC (2.5%)
Fibroblast I (13.9%)
Fibroblast II (5.9%)
Fibroblast III (5.7%)
Fibroblast IV (4.2%)
Fibroblast V (1.2%)
Fibroblast VI (0.6%)
Fibroblast VII (0.3)
Myofibroblast (3.3%)
Macrophage I (6.4%)
Macrophage II (1.2%)
Macrophage III (0.7%)
Dendritic cell (6.7%)
Langerhans (0.8%)
T cell I (3.8%)
T cell II (0.7%)
Endothelial (1.2%)
Skeletal muscle (1%)
F
UW (n=10615)
WO (n=16164)
Percent of cell types
Basal
Prolif. Basal
Spinous
HF/HFSC
Fibroblast
Myofibroblast
Macrophage
DC/Langerhans
T cell
Endothelial
Skeletal muscle

Figure S2: "Separate" and "combined" analysis of UW and WO samples showing replicate identity, heatmaps and associated marker genes. Related to Figure 1.

A. tSNE plot for the two UW and three WO replicates that were aggregated and then batch corrected using CCA. Each replicate sample is identified by a unique color identifier. There is mixing between the replicates, but some residual batch effects still persisted after CCA correction likely due to individual variations.
B. Heatmap for the top ten genes enriched in each of the unique clusters from UW skin. Genes listed in black represent the top two marker genes for each cluster, whereas those in red represent the additional genes used in the final identification of the cluster cell type. All marker genes of each cluster are listed in Table S1A.
C. Heatmap for the top ten genes enriched in each of the unique clusters from WO skin. See legends in (C). All marker genes of each cluster are listed in Table S1B.
D. Pie charts and bar graph showing the percentage of each major cell type in UW and WO samples from the "separate" analysis in Figures 1F and G. The number of total cells for UW and WO samples is indicated in parenthesis. Chi-squared test was used to determine whether the proportion differences of each cell type between UW and WO samples is statistically significant. *** $p < 0.0005$.
E. tSNE plot containing all five samples from both UW and WO samples. Cells were colored by the annotated cell subpopulations. Markers used for identifying each of the indicated cell types are listed in Table S2.
F. Pie charts and bar graph showing the percentage of each major cell types from the "combined" analysis in panel (E). The number of total cells for UW and WO samples is indicated in parenthesis. Chi-squared test was used to determine whether the proportion differences of each cell type between UW and WO samples is statistically significant. *** $p < 0.0005$.

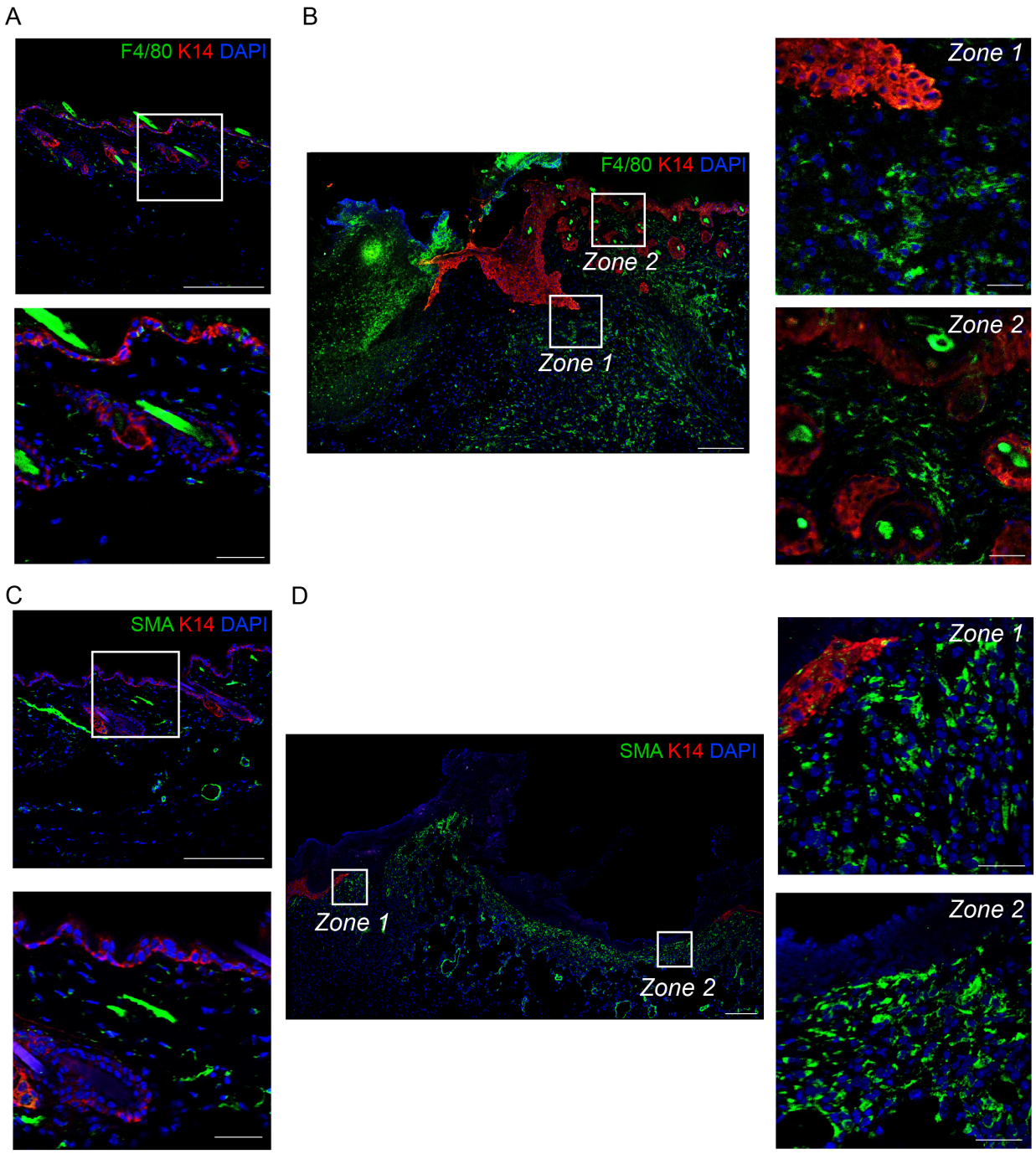

Figure S3: Increased numbers of macrophages and myofibroblasts in wounded skin. Related to Figure 1.

A. Indirect immunofluorescence for F4/80 and K14 in unwounded skin, with boxed area enlarged on the bottom. DAPI stains the nuclei. Scale bars: 200 μm and 20 μm for top and bottom images, respectively.
B. Indirect immunofluorescence for F4/80 and K14 in wounded skin, with boxed areas enlarged on the right. DAPI stains the nuclei. Scale bars: 200 μm and 20 μm for images on the left and right, respectively.
C. Indirect immunofluorescence for SMA and K14 in unwounded skin, with boxed area enlarged on the bottom. DAPI stains the nuclei. Scale bars: 200 μm and 20 μm for top and bottom images, respectively.
D. Indirect immunofluorescence for SMA and K14 in wounded skin, with boxed areas enlarged on the right. DAPI stains the nuclei. Scale bars: 200 μm and 20 μm for images on the left and right, respectively.

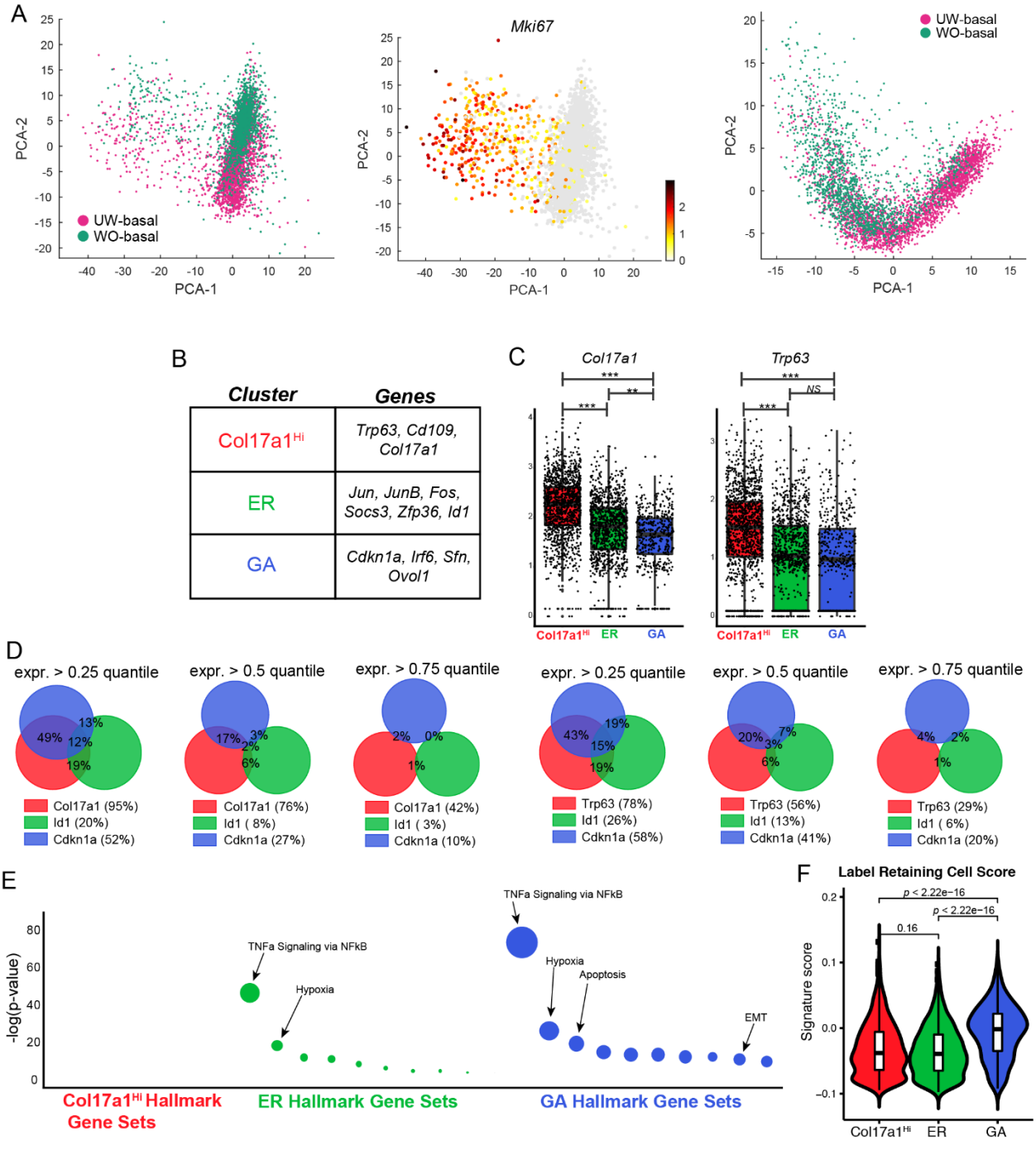
A
PCA-1
PCA-2
UW-basal
WO-basal
Mki67
B
Cluster
Genes
Col17a1Hi
Trp63, Cd109, Col17a1
ER
Jun, JunB, Fos, Socs3, Zfp36, Id1
GA
Cdkn1a, Irf6, Sfn, Ovol1
C
Col17a1
Trp63
NS
Col17a1Hi
ER
GA
D
expr. > 0.25 quantile
expr. > 0.5 quantile
expr. > 0.75 quantile
Col17a1 (95%)
Id1 (20%)
Cdkn1a (52%)
Col17a1 (76%)
Id1 (8%)
Cdkn1a (27%)
Col17a1 (42%)
Id1 (3%)
Cdkn1a (10%)
Trp63 (78%)
Id1 (26%)
Cdkn1a (58%)
Trp63 (56%)
Id1 (13%)
Cdkn1a (41%)
Trp63 (29%)
Id1 (6%)
Cdkn1a (20%)
E
-log(p-value)
TNFa Signaling via NFkB
Hypoxia
Apoptosis
EMT
Col17a1Hi Hallmark Gene Sets
ER Hallmark Gene Sets
GA Hallmark Gene Sets
F
Label Retaining Cell Score
Signature score
p < 2.22e−16
0.16
Col17a1Hi
ER
GA

Figure S4: Basal cell analysis. Related to Figures 3 and 4.
A: Proliferative basal cells have more distinct gene expression patterns compared to non-proliferative basal cells. Related to Figure 3. B-F: Basal cell analysis in UW skin. Related to Figure 4.

A. PCA analysis of total basal cells (proliferative and non-proliferative) from UW and WO samples. Feature plot indicating the expression of proliferative marker *Mki67* in a subset of cells. PCA analysis of non-proliferative basal cells from UW and WO samples after removal of proliferative cells.
B. Key marker genes enriched in the *Col17a1*^{Hi}, ER, and GA basal cell subclusters.
C. Boxplots showing median value of *Col17a1* and *Trp63* expression in each basal cell subcluster. *p* values are from Wilcoxon rank sum tests. NS: not significant (i.e., $p > 0.05$).
D. Overlap analysis among *Col17a1*, *Cdkn1a* and *Id1*, or *Trp63*, *Cdkn1a* and *Id1* for the basal cells expressing these genes in scRNA-seq data. We binarized the expression of each gene by choosing thresholds based on the quantile of all expressed cells. We varied the thresholds using three typical quantiles: 0.25, 0.5 and 0.75 quantiles. We then quantified the percentage of cells expressing one single gene, two genes or three genes. In each panel, the number in parenthesis associated with the indicated gene represents the percentage of cells expressing that particular gene among the total number (2838) of cells from two UW replicates. The percentage of cells expressing two or three genes among the total number (2838) of cells is shown in the overlap region of the circles.
E. GO analysis of the UW basal subclusters.
F. LRC gene score for *Col17a1*^{Hi}, ER, and GA basal cell subclusters. *p* values are from Wilcoxon rank sum tests. The list of LRC gene set used in this analysis is shown in Table S7.

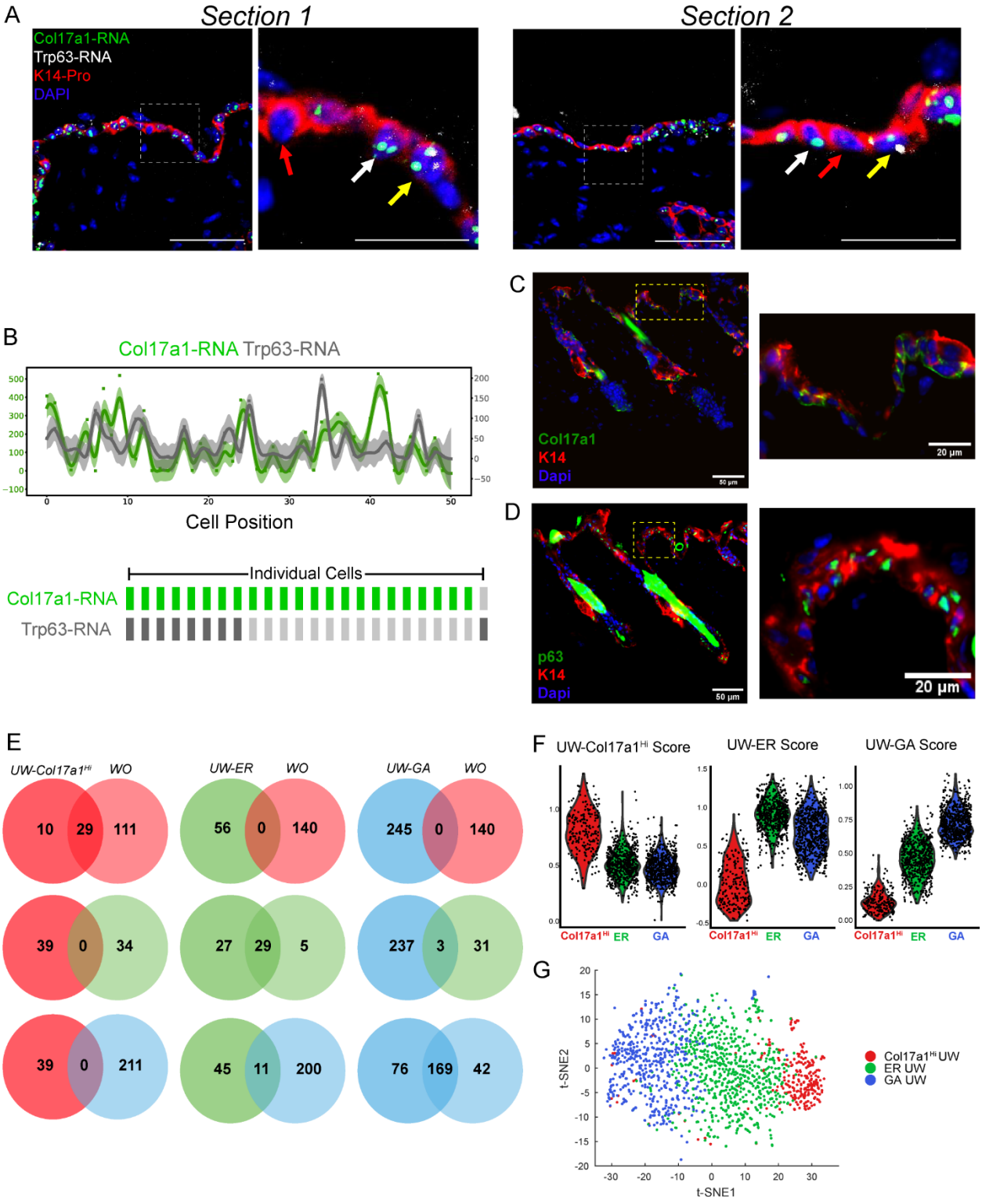
A
Section 1
Section 2
Col17a1-RNA
Trp63-RNA
K14-Pro
DAPI
B
Col17a1-RNA Trp63-RNA
Cell Position
Individual Cells
Col17a1-RNA
Trp63-RNA
C
Col17a1
K14
Dapi
20 µm
D
p63
K14
Dapi
20 µm
E
UW-Col17a1Hi
WO
UW-ER
WO
UW-GA
WO
10 29 111
56 0 140
245 0 140
39 0 34
27 29 5
237 3 31
39 0 211
45 11 200
76 169 42
F
UW-Col17a1Hi Score
UW-ER Score
UW-GA Score
Col17a1Hi ER GA
G
t-SNE1
t-SNE2
Col17a1Hi UW
ER UW
GA UW

Figure S5: Additional analysis of basal cells in UW and WO skin. Related to Figures 4 and 5.

A. RNAScope data of two different sections showing spatial distribution of *Col17a1*, *Trp63* transcripts and K14 protein in UW skin. DAPI stains the nuclei. Enlarged image of the boxed area in left panel is shown on the right. Red arrow indicates no *Col17a1* or *Trp63* expression; yellow arrow indicates both *Col17a1* and *Trp63* expression; white arrow indicates *Col17a1* expression only. Scale bars: 50 μm and 10 μm for images on the left and right, respectively, for each section.

B. Quantification of fluorescence intensity (represented by a color-coded dot) for *Col17a1* and *Trp63* transcripts and in each individual cell from a representative section (top). The curve represents a Gaussian Process Regression (GPR) and a 95% confidence interval is shown as shaded area. Bottom, OncoPrint representation of *Col17a1* and *Trp63* expression in individual cells where a rectangle represents an individual cell. A color-coded rectangle indicates high expression of the corresponding marker gene. The cells are sorted based on the on/off state of the markers to show mutually exclusive expression pattern.

C. Indirect immunofluorescence for Col17a1 and K14 proteins in unwounded skin. Enlarged image of the boxed area in left panel is shown on the right. Scale bar: 50 μm and 20 μm as indicated.

D. Indirect immunofluorescence for p63 and K14 proteins in unwounded skin. Enlarged image of the boxed area in left panel is shown on the right. Scale bar: 50 μm and 20 μm as indicated.

E. Marker gene overlap between the *Col17a1*^{Hi} (red), ER (green), and GA (blue) basal cell subclusters from the UW (left) and UW (right) samples. Note that the same-colored UW and WO clusters show the largest overlap. A complete list of marker genes for each subpopulation in UW and WO samples is provided in Table S5 and S6, respectively.

F. Scoring different basal cell subclusters from WO sample using marker genes of basal cell subclusters from UW samples. A complete list of marker genes for each UW subpopulation is provided in Table S5.

G. Random forest classification of basal cells from WO sample using basal cells from UW sample as training class.

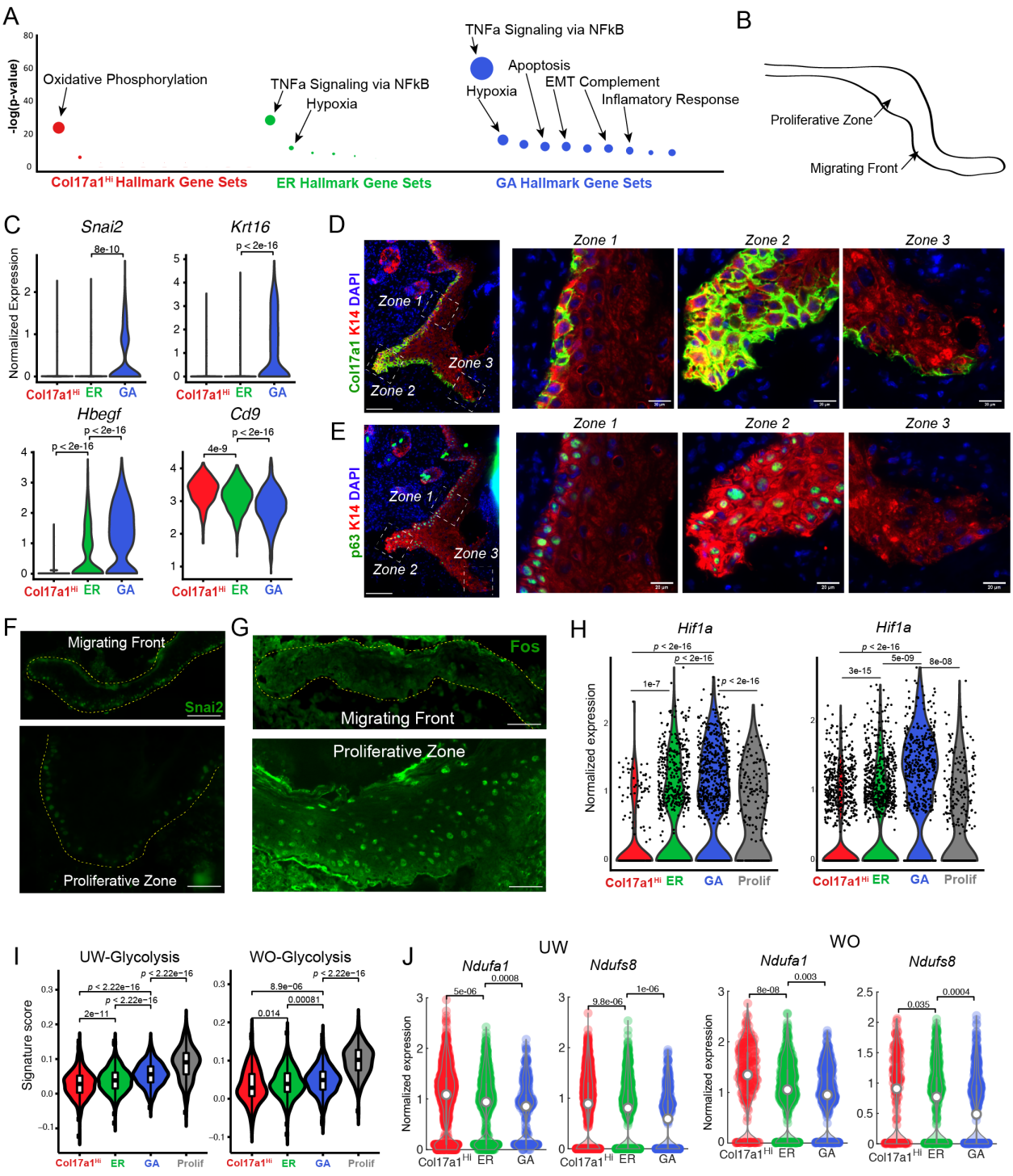
A
-log(p-value)
Oxidative Phosphorylation
TNFa Signaling via NFkB
Hypoxia
TNFa Signaling via NFkB
Apoptosis
Hypoxia
EMT Complement
Inflamatory Response
Col17a1Hi Hallmark Gene Sets
ER Hallmark Gene Sets
GA Hallmark Gene Sets
B
Proliferative Zone
Migrating Front
C
Snai2
Krt16
Hbegf
Cd9
Normalized Expression
Col17a1Hi ER GA
D
Col17a1 K14 DAPI
Zone 1
Zone 2
Zone 3
E
p63 K14 DAPI
F
Migrating Front
Snai2
Proliferative Zone
G
Fos
Migrating Front
Proliferative Zone
H
Hif1a
Normalized expression
Col17a1Hi ER GA Prolif
I
UW-Glycolysis
WO-Glycolysis
Signature score
J
UW
WO
Ndufa1
Ndufs8
Normalized expression

Figure S6: Basal cell analysis in WO skin. Related to Figures 5 and 6.

A. GO analysis of WO basal cell subclusters.
B. Diagram of wounded skin with migrating front and proliferative zone indicated.
C. Expression of known wound marker genes in WO basal cell subpopulations.
D. Indirect immunofluorescence for Col17a1 and K14 proteins in wounded skin. Enlarged images of the boxed areas in the left panel are shown on the right as Zone 1 (distal from the wound), Zone 2 (hyperproliferative), and Zone 3 (migrating front). DAPI stains the nuclei. Scale bar: 100 μm for left panel and 20 μm for Zone 1-3 images.
E. Indirect immunofluorescence for p63 and K14 proteins in wounded skin. See (D) legends for zone definition and scale bar information.
F. Indirect immunofluorescence for Snai2 protein in different regions of the wound. The yellow dashed line in the top and bottom panels outlines the entire migrating front and the basement membrane, respectively. Scale bar: 50 μm.
G. Indirect immunofluorescence for Fos protein in different regions of the wound. The yellow dashed line in the top panel outlines the entire migrating front. Scale bar: 50 μm.
H. Violin plot for *Hif1a* expression in basal cells from UW sample (left) and WO sample (right). *p* values are from Wilcoxon rank sum tests.
I. Gene scoring analysis of all four basal cell subclusters from UW and WO samples using a glycolysis signature. *p* values are from Wilcoxon rank sum tests. The list of gene set used in this analysis is shown in Table S7.
J. Expression of the indicated oxidative phosphorylation-associated genes in UW and WO basal cell subpopulations.

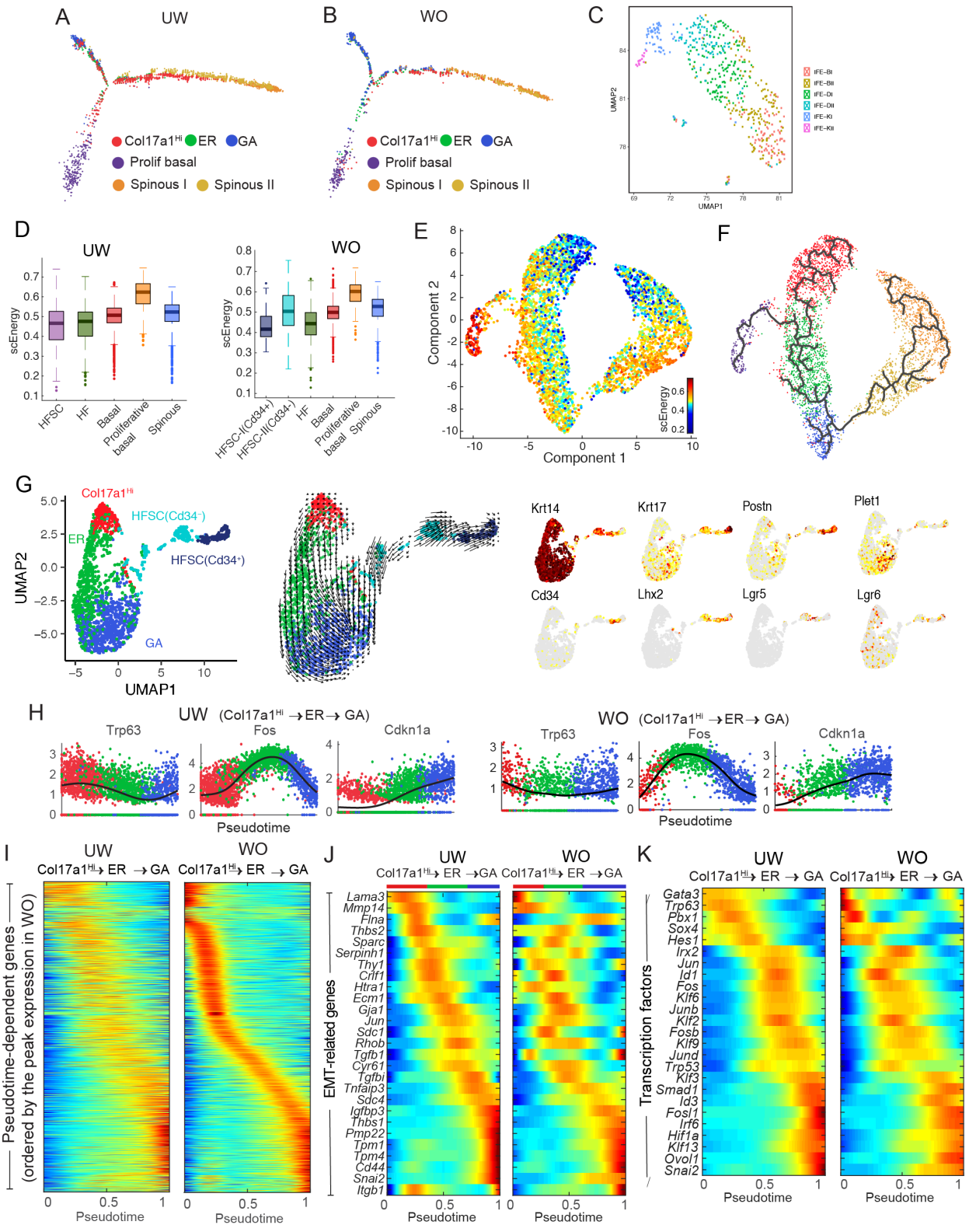
A
UW
Col17a1Hi ER GA
Prolif basal
Spinous I Spinous II
B
WO
Col17a1Hi ER GA
Prolif basal
Spinous I Spinous II
C
UMAP1
UMAP2
iFE-BI
iFE-BII
iFE-DI
iFE-DII
iFE-KI
iFE-KII
D
UW
WO
scEnergy
HFSC
HF
Basal
Proliferative basal
Spinous
HFSC-I(Cd34+)
HFSC-II(Cd34-)
E
Component 1
Component 2
scEnergy
F
G
Col17a1Hi
HFSC(Cd34-)
ER
HFSC(Cd34+)
GA
UMAP1
UMAP2
Krt14
Krt17
Postn
Plet1
Cd34
Lhx2
Lgr5
Lgr6
H
UW (Col17a1Hi →ER→ GA)
WO (Col17a1Hi →ER→ GA)
Trp63
Fos
Cdkn1a
Pseudotime
I
UW
WO
Col17a1Hi→ ER →GA
Pseudotime-dependent genes
(ordered by the peak expression in WO)
Pseudotime
J
EMT-related genes
Lama3
Mmp14
Flna
Thbs2
Sparc
Serpinh1
Thy1
Cnfl1
Htra1
Ecm1
Gja1
Jun
Sdc1
Rhob
Tgfb1
Cyr61
Tgfbi
Tnfaip3
Sdc4
Igfbp3
Thbs1
Pmp22
Tpm1
Tpm4
Cd44
Snai2
Itgb1
K
Transcription factors
Gata3
Trp63
Pbx1
Sox4
Hes1
Irx2
Jun
Id1
Fos
Klf6
Junb
Klf2
Fosb
Klf9
Jund
Trp53
Klf3
Smad1
Id3
Fosl1
Irf6
Hif1a
Klf13
Ovol1
Snai2

Figure S7: Lineage analysis in UW and WO skin. Related to Figure 7.

A. Lineage analysis of basal cells (proliferative and non-proliferative) and spinous cells in UW sample using Monocle 2.
B. Lineage analysis of basal cells (proliferative and non-proliferative) and spinous cells in WO sample using Monocle 2.
C. UMAP dimensional reduction for 536 interfollicular epidermal cells from Joost et al., 2016. Cell color corresponds to the exact annotations found in the original paper.
D. Boxplot indicating scEnergy values for different cell type/states from UW and WO samples containing epithelial cells from Figures 2A and 2B, respectively. The quiescent Bu-HFSC population ($CD34^{+}$) displayed lowest scEnergy values.
E. UMAP dimensional reduction for UW samples with scEnergy indicated.
F. Monocle 3 analysis for the UW sample with UMAP representation as in Figure 7A. The black lines indicate the structure of the inferred trajectory, showing how the cells transition from one local state to another.
G. Left panel, UMAP dimensional reduction for basal cells and HFSCs from WO sample. Middle panel, RNA velocity analysis using a nonlinear model with vectors projected onto the UMAP space. Right panel, feature plots highlighting basal-, HF-, and HFSC-specific genes. Normalized expression of each cell is color-coded.
H. Dynamic changes in gene expression of individual genes *Trp63*, *Fos*, and *Cdkn1a* over pseudotime from the *Col17a1*^{Hi}-to-GA path.
I. Pseudotemporal dynamics of the identified WO 3,128 pseudotime-dependent genes along the *Col17a1*^{Hi}-to-GA path in UW and WO samples. Cell state identity (*Col17a1*^{Hi}, ER or GA) is indicated on the top of each heatmap generated by the smoothed, normalized gene expression. Blue and red colors within the heatmap indicate low and high expression, respectively. Each row/gene was normalized to its peak value along the pseudotime.
J. Subsets of "EMT-related" genes (from the 3,699 UW pseudotime-dependent genes) with their corresponding pseudotemporal dynamics in both UW and WO basal cells along the *Col17a1*^{Hi}-to-GA path.
K. Subsets of transcription factors (from the 3,699 UW pseudotime-dependent genes) with their corresponding pseudotemporal dynamics in both UW and WO basal cells along the *Col17a1*^{Hi}-to-GA path.