Supplementary information for

Dnmt3a deficiency in the skin causes focal, canonical DNA hypomethylation and a cellular proliferation phenotype

David Y. Chen[a], Ian M. Ferguson[a], Krista A. Braun[a,b], Leslie A. Sutton[a], Nichole M. Helton[b], Sai Mukund Ramakrishnan[b], Amanda M. Smith[b], Christopher A. Miller[b], Timothy J. Ley[b]

[a]Division of Dermatology, Department of Medicine, Washington University School of Medicine, Saint Louis MO, 63110
[b]Section of Stem Cell Biology, Division of Oncology, Department of Medicine, Washington University School of Medicine, Saint Louis MO, 63110

Correspondence:

Timothy J. Ley, M.D.
timley@wustl.edu

This appendix includes:

Supplementary methods

Figures S1-S7

Legends for Datasets S1-S3

Other supplementary materials for this manuscript include the following:

Datasets S1-S3

Supplementary materials and methods

Animals. Mice with the conditional *Dnmt3a*^{flox} (60) and *Dnmt3b*^{flox} (61) alleles were obtained from the Mutant Mouse Regional Resource Center (MMRRC, stock number 029884 and 029887). *Dnmt3a*^{flox} mice were maintained in the C57/Bl6 background, while *Dnmt3b*^{flox} were in the Bl6/129 mixed background. Mice with the human keratin 14 promoter and *Cre* recombinase transgene (*Krt14-Cre*, The Jackson Laboratory, Stock 018964) maintained in the C57/Bl6J background. Floxing efficiency primer sequences were obtained from prior studies (68).

Epidermal cell isolation. Mice were analyzed at maturity (greater than 7 weeks of age). For all studies, we harvested single cell suspensions of dorsal epidermis in second telogen (7-10wks of age), or in older mice, when skin is >90% telogen based on pale or pink skin color. Single cell suspensions were generated as previously described (62, 63). Briefly, mice were sacrificed by CO_2 asphyxiation, dorsal skin shaved and dissected from the mouse. Subcutaneous fat and panniculus carnosum were removed and prepared skin was suspended with the dermal side in contact with 0.25% trypsin for 1 hour at 37°C. Epidermal layers were then scraped from the dermis, minced, and disassociated with vigorous trituration. Debris was cleared by passing the suspension through a 70μm filter, followed by washing with analysis buffer (phosphate buffered saline with 2% fetal bovine serum and 0.25mM EDTA). Cell viabilities were routinely >85%. All analyses were performed on fresh samples.

Whole genome bisulfite sequencing. Whole-genome bisulfite-converted sequencing libraries were generated with the AccelNGS Methyl-Seq DNA library kit (Swift Biosciences, #30096). Indexed sequencing was performed on Illumina HiSeq instruments. Data alignment and methylation inference used the pipeline described at https://github.com/genome/analysis-

workflows/blob/b4b23be/definitions/pipelines/bisulfite.cwl. In summary, reads were mapped with Biscuit (version 0.3.8) using default parameters. Duplicates were marked, VCFs were generated using 'biscuit pileup' (params: -q 4) and then converted to CpG bed files. The program 'metilene' (64) was used to analyze the raw methylation ratios at all CpGs to identify differentially methylated regions (DMRs), requiring >10 CpGs and a mean methylation difference between groups of >0.2. DMRs were filtered to retain those with false discovery rate (FDR) <0.05, and adjacent DMRs <50 bp apart were merged. Following these procedures, methylation data from individual CpGs were imported into R as 'bsseq' objects (65). Multiple testing correction was performed using the p.adjust function in R version 3.5, using default parameters.

Single-cell RNA sequencing and analysis. Single-cell suspensions of murine epidermis were freshly prepared as described above. Libraries were generated using the 10X Chromium Controller system with the 5' GEX V2 library kit according to manufacturer's protocols (https://tinyurl.com/y96l7lns) and sequenced on an Illumina NovaSeq 6000 (2 x 150 paired end reads) yielding a mean of 636,759,110 reads per sample. Transcript alignment and counting were performed using the Cell Ranger pipeline (10x Genomics, default settings, Version 3.0.1). Using a nearest neighbor algorithm implemented in R (version 3.5.1). Data analysis was performed using the standard tools available in the Partek Flow software suite (Partek, Inc. Chesterfield, MO). We eliminated cells that contained fewer than 500 or greater than 5,000 expressed genes, less than 1000 or more than 50,000 total reads, or more than 6% mitochondrial transcripts, to remove dead cells and doublets. Expression of each gene was normalized to the sequencing depth of the cell, scaled to a constant depth (10,000), and log-transformed. Dimensionality reduction was initially accomplished by principal component analysis, and visualization was achieved using the UMAP algorithm. Differentially expressed genes were determined using gene specific analysis algorithm and defined as having an FDR of <0.01 and a fold change greater than an absolute value of 1. Between 5,000 and 10,000 cells

per sample were routinely analyzed, with between 32-135,000 reads per cell, and 2300-2900 genes identified per cell. Unbiased clustering was performed using the Louvain ("graph based") clustering model (66) as part of the Partek suite. A population defined by *Krt18* expression comprised 3.05% of the *Dnmt3a*^{KO} sample KO7 and was not present in KO6. Its signature dominated differential gene expression analyses, and for this reason, it was excluded from further analysis.

Immunoblotting. Freshly isolated single cell suspensions of murine epidermis were lysed at 20,000 cells per microliter using RIPA buffer (Cell Signaling, Inc) supplemented with protease inhibitor cocktail (Sigma) and sonicated using a Branson 450 probe for 12 cycles at 30% output with 30% duty cycle. Lysates were cleared by centrifugation and protein content was quantified by BCA assay (ThermoFisher), normalized, and loaded on the capillary-based Simple Western per manufacturer's guidelines. The following antibodies were used to detect indicated antigens; Dnmt3a (D23G1, Cell Signaling Technologies), Krt7 (EPR17078, Abcam), Irx1 (HPA043160, Sigma), Actin (AC-15, Novus Biologicals).

Flow cytometry. Freshly isolated single cell suspensions of murine epidermis were resuspended at a concentration of 10^7 cells per milliliter in staining buffer (phosphate buffered saline with 2% fetal bovine serum and 0.25mM EDTA) with the following antibodies: Cd34-FITC (RAM34, BD Pharmingen), Cd49f-PE (NKI-GoH3, Biorad), Sca1-PerCP-Cyanine5.5 (D7, eBioscience), Epcam-APC (G8.8, Invitrogen), Cd45-APC-eFluor 780 (30-F11, Invitrogen). Cells were stained for 30 minutes at 4°C, then viable cells were determined by dye exclusion with SYTOX-blue (S34857, ThermoFisher). Samples were run on a modified BD FACScan cytometer with 488nm, 647nm, and 407nm laser lines. Data was analyzed using FlowJo version 10 software with gating strategy adapted from published studies (67).

BrdU uptake evaluation. *Dnmt3a*^{WT} and *Dnmt3a*^{KO} mice at 8-9 weeks of age were injected with 100 microliters of 10mg/mL BrdU solution twice (separated by 8 hours), and sacrificed 24 hours after the first injection. Epidermal single cell suspensions were generated as outlined above, and stained with surface antibodies to Cd34, Cd49f, Sca1, and Cd45. Stained cells were then fixed, permeabilized, and stained for BrdU according to the protocol supplied with BrdU flow cytometry kit (BD Pharmingen 552598). The total proliferative population was determined from the Cd45-negative gate for n=7 *Dnmt3a*^{WT} and n=5 *Dnmt3a*^{KO} mice, and statistical significance was determined using an unpaired, two-tailed t test.

Deposition of sequence data

All sequencing data for human patients were deposited to dbGaP (phs000159.v10) and mouse data to the sequence read archive (BioProject PRJNA674614).

Supplemental References

60. M. Kaneda, *et al.*, Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429**, 900–3 (2004).

61. J. E. Dodge, *et al.*, Inactivation of Dnmt3b in Mouse Embryonic Fibroblasts Results in DNA Hypomethylation, Chromosomal Instability, and Spontaneous Immortalization. *Journal of Biological Chemistry* **280**, 17986–17991 (2005).

62. U. Lichti, J. Anders, S. H. Yuspa, Isolation and short-term culture of primary keratinocytes, hair follicle populations and dermal cells from newborn mice and keratinocytes from adult mice for in vitro analysis and for grafting to immunodeficient mice. *Nature Protocols* **3**, 799–810 (2008).

63. K. B. Jensen, R. R. Driskell, F. M. Watt, Assaying proliferation and differentiation capacity of stem cells using disaggregated adult mouse epidermis. *Nature Protocols* **5**, 898–911 (2010).

64. F. Jühling, *et al.*, metilene: fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. *Genome Res* **26**, 256–262 (2016).

65. K. D. Hansen, B. Langmead, R. A. Irizarry, BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol* **13**, R83 (2012).

66. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks. *J Statistical Mech Theory Exp* **2008**, P10008 (2008).

67. T. Adachi, *et al.*, Hair follicle-derived IL-7 and IL-15 mediate skin-resident memory T cell homeostasis and lymphoma. *Nature medicine* **21**, 1272–9 (2015).

68. Y. Tadokoro, H. Ema, M. Okano, E. Li, H. Nakauchi, De novo DNA methyltransferase is essential for self-renewal, but not for differentiation, in hematopoietic stem cells. *The Journal of Experimental Medicine* **204**, 715–722 (2007).

Supplementary figures

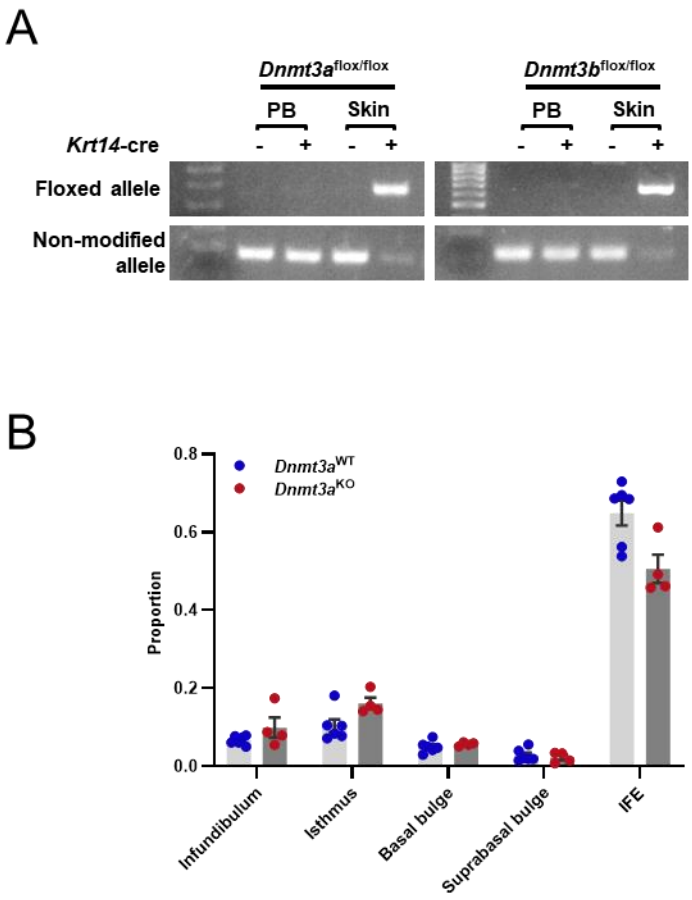

Figure S1. *Dnmt3a* and *Dnmt3b* have high degree of floxing efficiency and *Dnmt3a* deficiency does not change immunophenotype defined populations in the skin

A. PCR floxing efficiency from peripheral blood (PB) and skin, demonstrating the tissue specificity of floxing from Krt14-Cre for the genotypes indicated. Primers and conditions from a prior study (68).

B. Single cell suspensions of murine epidermis from *Dnmt3a*^{WT} and *Dnmt3a*^{KO} were analyzed by flow cytometry to identify the interfollcular epidermis, IFE (Cd45^{neg};Sca1^{pos};Epcam^{neg}), follicular infundibulum (Cd45^{neg}; Sca1^{pos};Epcam^{pos}), isthmus (Cd45^{neg};Sca1^{neg};Epcam^{pos}), suprabasal bulge (Cd45^{neg};Sca1^{neg}; Cd34^{pos}; Cd49f^{neg}), and basal bulge (Cd45^{neg};Sca1^{neg};Cd34^{pos}; Cd49f^{pos}). None of the values shown are significantly different between the WT and KO samples by unpaired t-test with Holm-Sidak multiple comparisons correction.

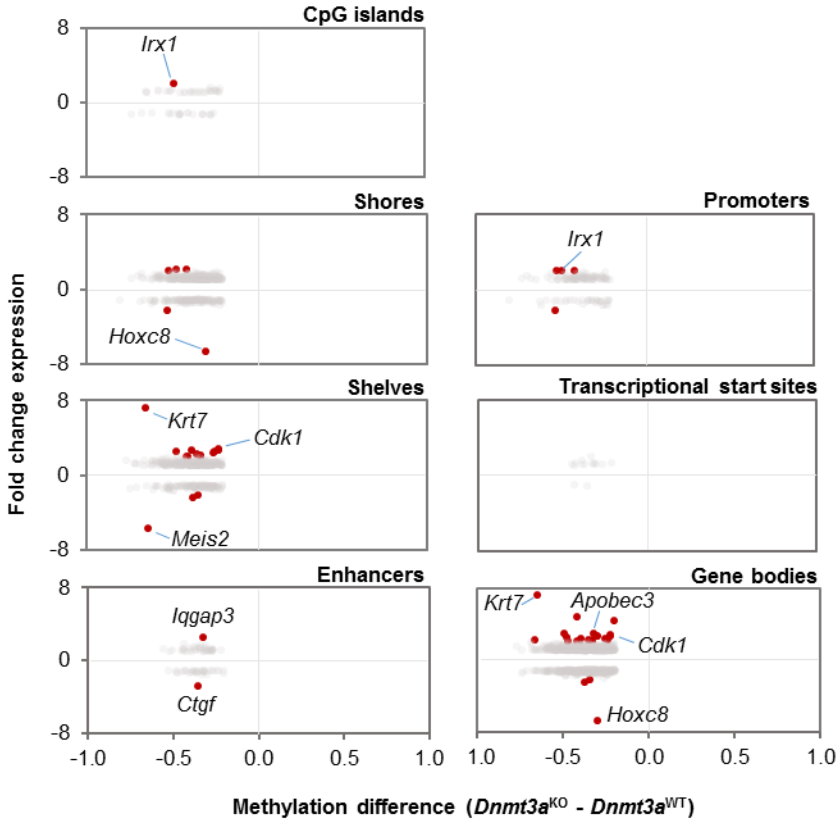

Figure S2. Expression versus methylation in *Dnmt3a*^{KO} epidermal cells

Differentially methylated regions from *Dnmt3a*^{KO} vs *Dnmt3a*^{WT} mice were associated with nearest genes, whose expression is plotted as a fold change from *Dnmt3a*^{KO} vs *Dnmt3a*^{WT}. DMRs were separated by annotated regions in which they reside. Red dots indicate absolute fold change expression > 2, while gray dots represent absolute fold changes between 1 and 2, both with FDR < 0.01.

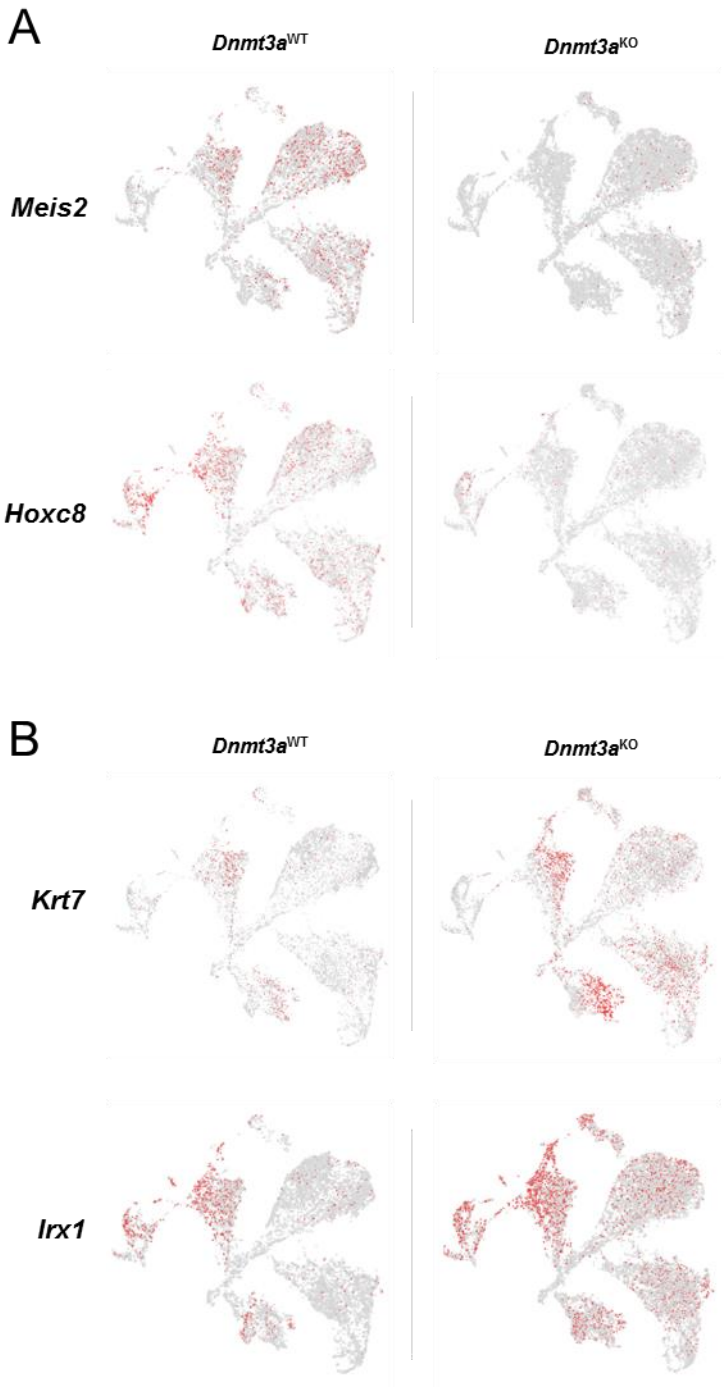

Figure S3. Gene expression changes in the *Dnmt3a*^{KO} epidermis

A. UMAP plots of *Meis2* and *Hoxc8* gene expression by single cell RNA sequencing demonstrates wide expression in the *Dnmt3a*^{WT} epidermis that is lost in the *Dnmt3a*^{KO} epidermis.

B. UMAP plots of *Krt7* and *Irx1* gene expression demonstrate Krt7 expression limited to the upper hair follicle in *Dnmt3a*^{WT} epidermis that is expanded to the bIFE and sbIFE in in the *Dnmt3a*^{KO} epidermis. Likewise, Irx1 expression is initially limited to the HFB and uHF but is broadened to all compartments in the *Dnmt3a*^{KO} epidermis, demonstrating that Dnmt3a limits expression promiscuity for certain genes.

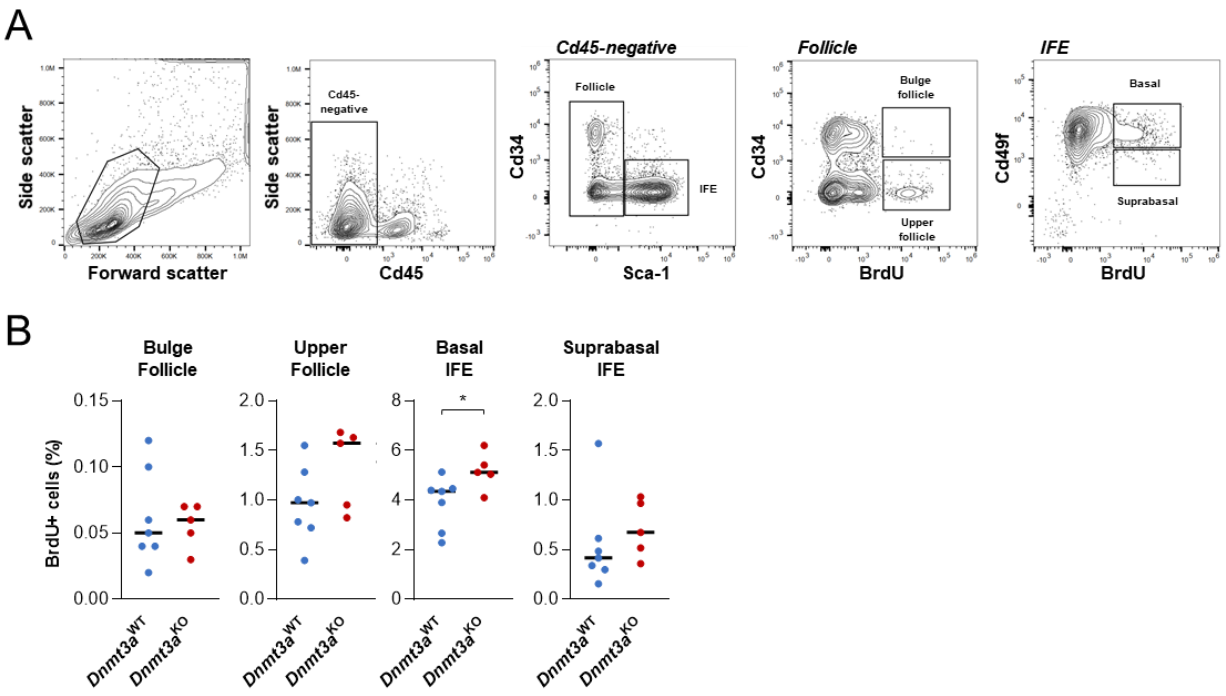

Figure S4. Proliferative phenotype in Dnmt3a-deficient keratinocytes *in vivo*

A. Flow cytometry gating strategy for keratinocyte subpopulations. Cd45-negative cells in the isolated epidermis were partitioned into follicular (Sca-1-negative) or interfollicular epidermis (Sca-1-positive). The follicular population was partitioned to the hair follicle bulge (Cd34-positive) or upper hair follicle (Cd34-negative), while the interfollicular epidermis (IFE) was partitioned into basal (Cd49f-positive) or suprabasal (Cd49f-negative).

B. Quantitation of BrdU positive cells assessed by flow cytometry as a proportion of Cd45-negative epidermal cells. *p<0.05 by unpaired t test.

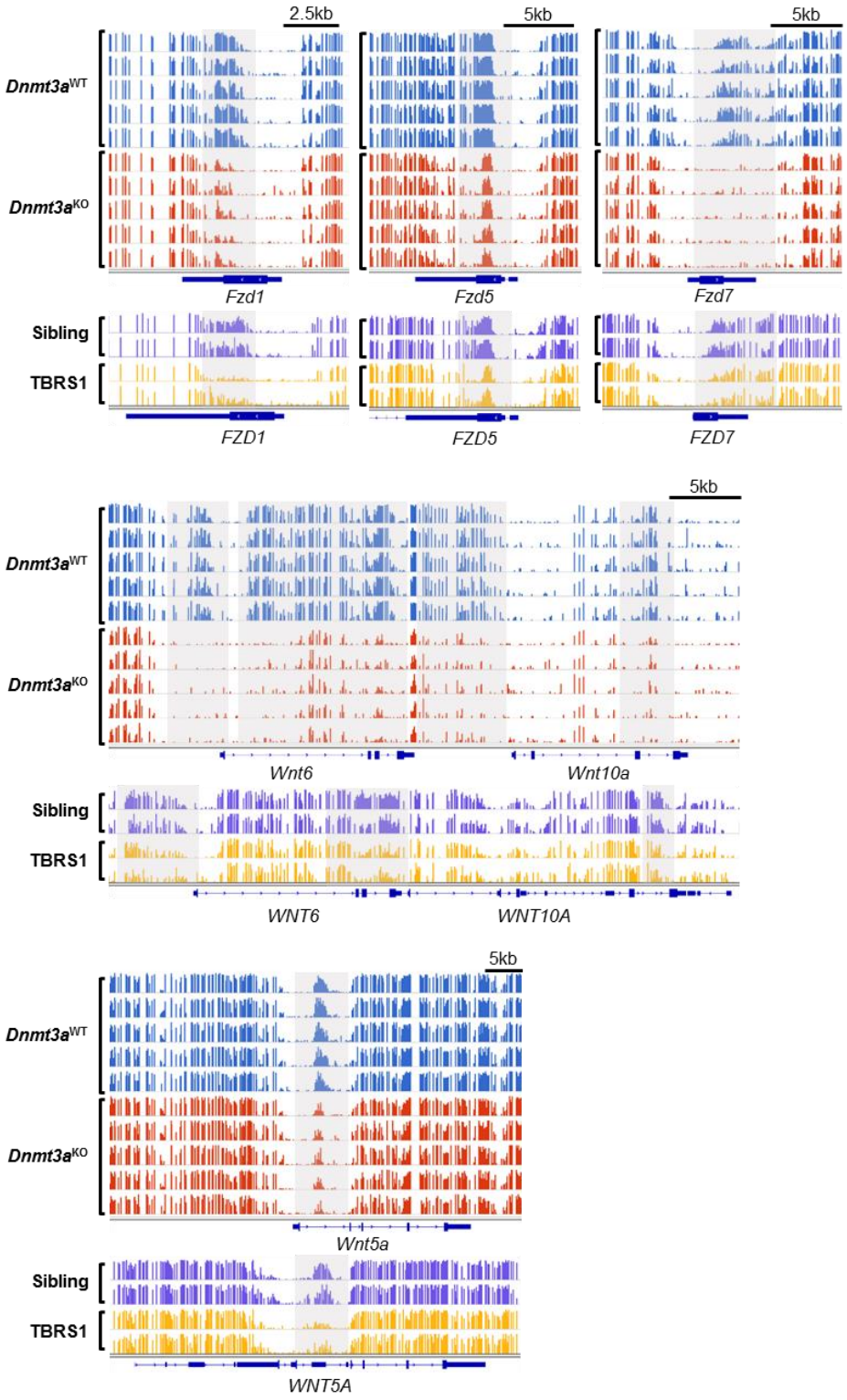

Figure S5. DNA methylation profiles in the *WNT/Wnt* pathway
IGV views of *Dnmt3a*^{WT} and *Dnmt3a*^{KO} (blue and red, respectively) compared with TBRS and unaffected sibling (yellow and purple, respectively) for multiple members of the Wnt signaling pathway. Areas of differential methylation are highlighted in gray.

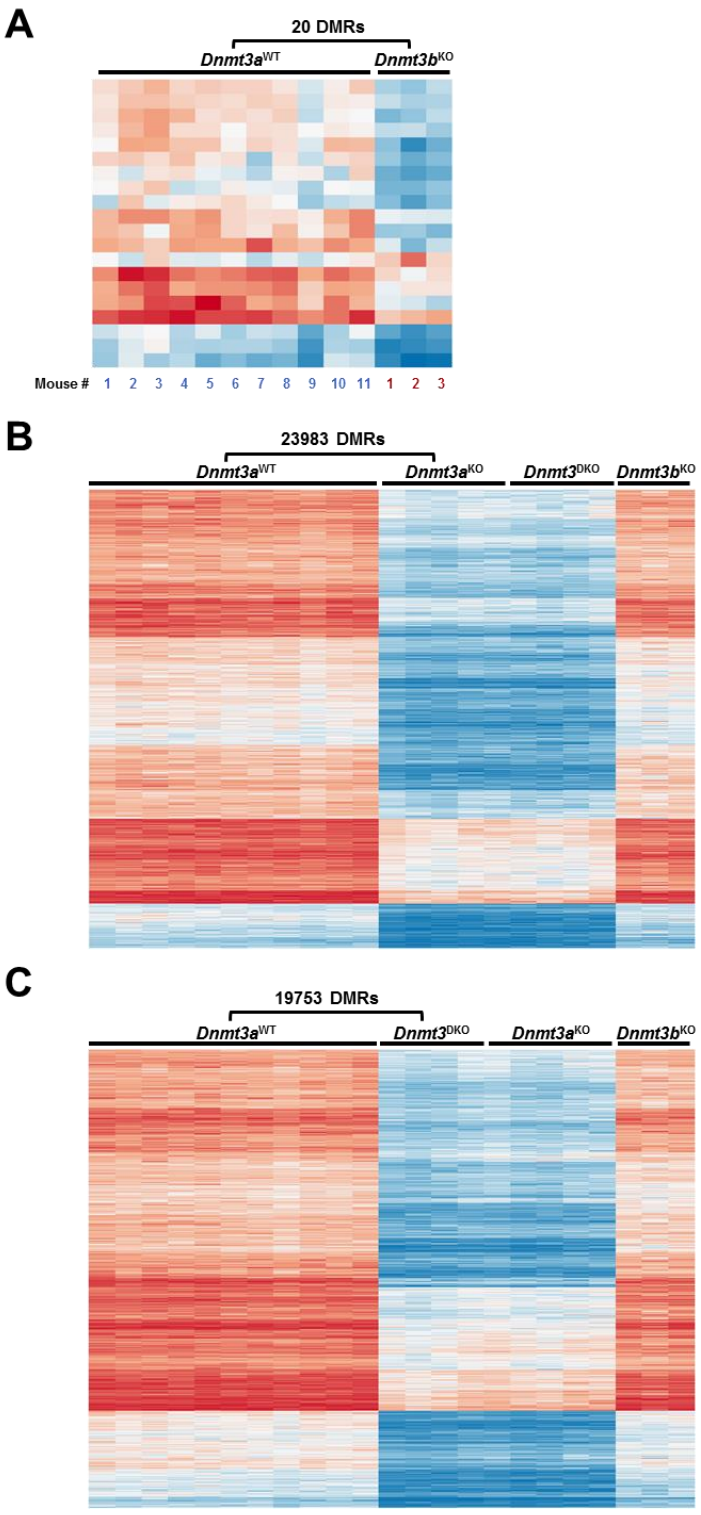

Figure S6. Dnmt3a is the dominant *de novo* DNA methyltransferase in the skin.

A. WGBS analysis of *Dnmt3a*^{WT} murine epidermis compared to targeted deficiency of *Dnmt3b*^{KO} revealed 20 differentially methylated regions. Each column represents an independent animal.

B. WGBS analysis of *Dnmt3a*^{WT} versus *Dnmt3a*^{KO} identifies 23983 DMRs. Methylation values at these *Dnmt3a*^{KO} defined DMRs for *Dnmt3a* and *Dnmt3b* compound deficiency (DKO) and *Dnmt3b*^{KO} are shown.

C. WGBS analysis of *Dnmt3a* and *Dnmt3b* compound deficiency (DKO) identifies 19753 DMRs. Methylation values at these DKO defined DMRs for *Dnmt3a*^{KO} and *Dnmt3b*^{KO} are shown.

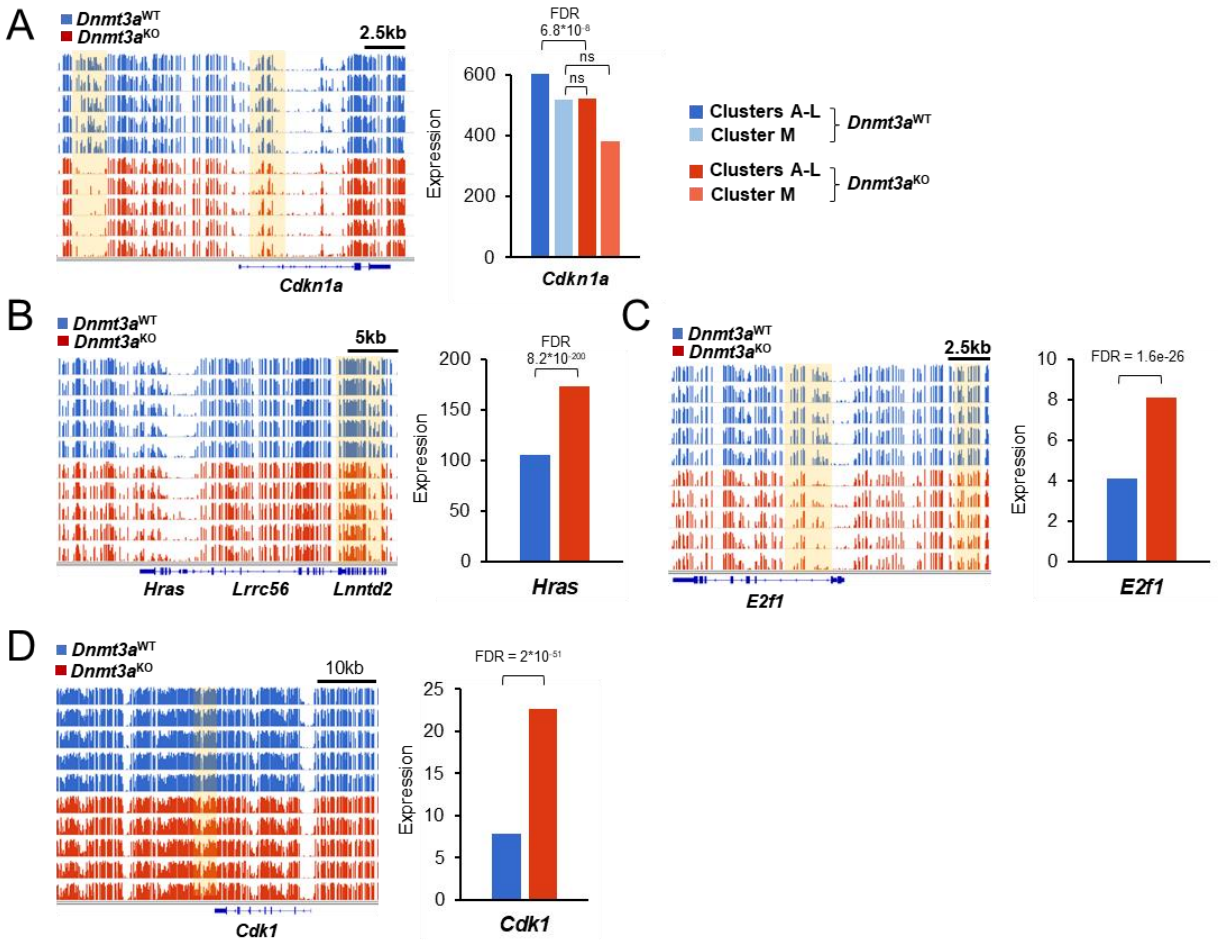

Figure S7. Cancer relevant gene expression changes in *Dnmt3a*^{KO} keratinocytes

A. IGV view of the *Cdkn1a* locus comparing methylation values between *Dnmt3a*^{WT} (blue) and *Dnmt3a*^{KO} (red) epidermal cells. DMRs are highlighted in light yellow. Expression data from single cell RNA sequencing comparing Cdkn1a expression between Cluster A-L and Cluster M for *Dnmt3a*^{WT} (blue) and *Dnmt3a*^{KO} (red) epidermal cells (right panel). Significance is defined as FDR < 0.01.

B. IGV view of the *Hras* locus comparing methylation values between *Dnmt3a*^{WT} (blue) and *Dnmt3a*^{KO} (red) epidermal cells. DMRs are highlighted in light yellow. Expression data from single cell RNA sequencing comparing Hras expression between *Dnmt3a*^{WT} (blue) and *Dnmt3a*^{KO} (red) epidermal cells (right panel).

C. IGV view of the *E2f1* locus comparing methylation values between *Dnmt3a*^{WT} (blue) and *Dnmt3a*^{KO} (red) epidermal cells. DMRs are highlighted in light yellow. Expression data from single cell RNA sequencing comparing *Ef21* expression between *Dnmt3a*^{WT} (blue) and *Dnmt3a*^{KO} (red) epidermal cells (right panel).

D. IGV view of the *Cdk1* locus comparing methylation values between *Dnmt3a*^{WT} (blue) and *Dnmt3a*^{KO} (red) epidermal cells. DMRs are highlighted in light yellow. Expression data from single cell RNA sequencing comparing *Cdk1* expression between *Dnmt3a*^{WT} (blue) and *Dnmt3a*^{KO} (red) epidermal cells (right panel).

Supplemental dataset legends

Dataset S1

Differentially methylated regions called between *Dnmt3a*^{WT} and *Dnmt3a*^{KO} epidermal cells, determined by whole genome bisulfite sequencing

Dataset S2

Differentially expressed genes between *Dnmt3a*^{WT} and *Dnmt3a*^{KO} epidermal cells, determined by single cell RNA sequencing

Dataset S3

Differentially expressed genes defining each graph-based cluster from single cell RNA sequencing. These represent the aggregate of all single cell experiments, including *Dnmt3a*^{WT} and *Dnmt3a*^{KO} epidermal cells.