UC Irvine
UC Irvine Previously Published Works

Title

The Msi1-mTOR pathway drives the pathogenesis of mammary and extramammary Paget's disease.

Permalink

https://escholarship.org/uc/item/4kd1d4zd

Journal

Cell research, 30(10)

ISSN

1001-0602

Authors

Song, Yongli
Guerrero-Juarez, Christian F
Chen, Zhongjian
et al.

Publication Date

2020-10-01

DOI

10.1038/s41422-020-0334-5

Peer reviewed

ARTICLE

The Msi1-mTOR pathway drives the pathogenesis of mammary and extramammary Paget's disease

Yongli Song[1,2], Christian F. Guerrero-Juarez [3,4], Zhongjian Chen[5], Yichen Tang[5], Xianghui Ma[1], Cong Lv[1], Xueyun Bi[1], Min Deng[1], Lina Bu[1], Yuhua Tian[1], Ruiqi Liu[1], Ran Zhao[1], Jiuzhi Xu[1], Xiaole Sheng[1], Sujuan Du[1], Yeqiang Liu[5], Yunlu Zhu[5], Shi-jun Shan[6], Hong-duo Chen[7], Yiqiang Zhao[1], Guangbiao Zhou [8], Jianwei Shuai[9], Fazheng Ren[10], Lixiang Xue[11], Zhaoxia Ying[12], Xing Dai[13], Christopher J. Lengner[14], Bogi Andersen[15], Maksim V. Plikus [4], Qing Nie[3,4] and Zhengquan Yu [1]

Mammary and extramammary Paget's Diseases (PD) are a malignant skin cancer characterized by the appearance of Paget cells. Although easily diagnosed, its pathogenesis remains unknown. Here, single-cell RNA-sequencing identified distinct cellular states, novel biomarkers, and signaling pathways — including mTOR, associated with extramammary PD. Interestingly, we identified MSI1 ectopic overexpression in basal epithelial cells of human PD skin, and show that Msi1 overexpression in the epidermal basal layer of mice phenocopies human PD at histopathological, single-cell and molecular levels. Using this mouse model, we identified novel biomarkers of Paget-like cells that translated to human Paget cells. Furthermore, single-cell trajectory, RNA velocity and lineage-tracing analyses revealed a putative keratinocyte-to-Paget-like cell conversion, supporting the in situ transformation theory of disease pathogenesis. Mechanistically, the Msi1-mTOR pathway drives keratinocyte-Paget-like cell conversion, and suppression of mTOR signaling with Rapamycin significantly rescued the Paget-like phenotype in Msi1-overexpressing transgenic mice. Topical Rapamycin treatment improved extramammary PD-associated symptoms in humans, suggesting mTOR inhibition as a novel therapeutic treatment in PD.

INTRODUCTION

Mammary and extramammary Paget's Diseases (PD) are a rare, malignant intra-epidermal adenocarcinoma with high recurrence rate after surgical excision.[1,2] Mammary Paget's Disease (MPD) localizes within the epidermis of the nipple and/or areola of the breast, and in more than 90% of cases it is associated with mammary ductal carcinoma.[3] Extramammary Paget's Disease (EMPD) primarily localizes to the vulvar, perianal, scrotal, penile and axillary regions.[4] In most cases, EMPD initially occurs within the epidermis, whereas 17%–30% of cases are associated with underlying cancer.[4] MPD and EMPD share many common clinical and pathomorphological features. The primary pathognomonic feature of PD is the presence of Paget cells. Paget cells are large cells with a clear cytoplasm and a peripherally positioned, hyperchromic nucleus. Within diseased tissue, Paget cells are most abundant in lower epidermal layers where they occasionally aggregate and form gland-like structures with a central lumen (i.e., reminiscent of Paget cell nests). Paget cells share characteristics with apocrine sweat gland epithelial cells, and stain positive for Periodic Acid-Schiff (PAS) staining, a dye that marks neutral polysaccharides. Paget cells also express simple gland-type epithelial cytokeratins, including KRT7[5] and KRT8;[6] the glandular cell adhesion molecule Cam5.2;[7] the tumor markers carcinoembryonic antigen (CEA) and Mucin1;[6] the oncoproteins Her2,[8] Cyclin D1[9] and p21;[10] as well as other markers such as androgen receptor (AR),[11] epithelial membrane antigen (EMA), gross cystic disease fluid protein (GCDFP-15), GATA3,[12] and p16.[13] Although PD can be diagnosed histologically,[14] its pathogenesis remains largely unknown. Consequently, no drug-based therapies exist for PD, and surgical resection, chemo- or radio-therapies, and laser

ablation remain the standard of care for PD. Due to its high recurrence rate, PD patients usually undergo repeated surgery every 2–3 years. Therefore, it is highly significant and of wide interest to develop effective therapeutic drugs for PD.

Two theories have been proposed to explain the origin of Paget cells.[3] The Epidermotropic theory remains in favor, postulating that Paget cells originate from underlying adenocarcinoma cells that migrate into the epidermis from skin-associated glands. This theory is supported by the histopathological features and HER2 reactivity shared by PD and adenocarcinoma,[15,16] as well as by the presence of underlying breast carcinoma in the majority of MPD cases.[15] The transformation theory, however, suggests that Paget cells originate in situ upon malignant transformation of basal keratinocytes, rather than from migrating adenocarcinoma cells. Supporting this theory, most EMPD and some MPD lesions are initially limited to the skin and are not accompanied by an underlying adenocarcinoma. However, even under the transformation theory, it remains unresolved if Paget cells originate from histologically similar yet benign epidermal Toker cells,[17] or from adjacent apocrine sweat gland sebocytes.[18] The lack of direct evidence, stemming from the lack of a mouse model of PD and consequent lineage tracing data precludes the definitive determination of the origin of Paget cells in PD.

First identified in *Drosophila*,[19] Msi1 is an evolutionarily conserved RNA-binding protein with two functional orthologues in mammals, Musashi1 (MSI1/Msi1)[20] and Musashi2 (MSI2/Msi2).[21] Msi1 has been identified as a marker and critical regulator of intestinal stem cells,[22,23] and it functions as an important driver of oncogenic transformation in the intestine.[23,24] It remains unknown whether Msi1 regulates tumorigenesis in other fast-renewing organs, such as skin.

Here, single-cell RNA-sequencing (scRNA-seq) revealed a large degree of epithelial cell heterogeneity in EMPD epidermis, and identified Paget cells as well as their unique molecular features and the signaling pathways potentially critical for the pathogenesis of PD. Interestingly, we observed MSI1 overexpression in basal epithelial cells of PD skin, and ectopic overexpression of Msi1 in the epidermal basal layer of mice leads to a phenotype that resembles human EMPD at both histopathological and molecular levels. We also provide compelling new evidence in support of the in situ transformation theory in both human EMPD and the EMPD-like mouse model, postulating that keratinocytes can give rise to Paget cells. Furthermore, our findings demonstrated the Msi1-mTOR signaling pathway is critical for EMPD pathogenesis. Topical Rapamycin treatment dramatically improved the appearance and clinical symptoms of EMPD in a small cohort of human patients. These results suggest that targeting mTOR signaling may be a promising avenue for the treatment of PD.

RESULTS

Characterization of the epithelial cellular taxonomy of human extramammary Paget's Disease at single-cell resolution

To begin to understand the cellular features associated with human EMPD, we isolated single cells from human scrotal EMPD skin epithelium and subjected them to droplet-enabled 3′-end scRNA-seq (Fig. 1a; Supplementary information, Fig. S1a and Table S1).[25] We processed a total of 23,511 cells (Normal, $n = 9031$ cells; EMPD, $n = 14,480$ cells) and performed quality control analysis on individual libraries (Supplementary information, Fig. S2). Post-quality control,[26] 12,070 "viable" cells were used for downstream analyses. We used Seurat's Anchoring algorithm[27] to integrate normal ($n = 6315$ cells) and EMPD ($n = 5755$ cells) data sets and visualized them using Uniform Manifold Approximation and Projection (UMAP) (Fig. 1b). Based on known marker genes, we identified thirteen distinct cellular communities, including cycling epithelial cells, $KRT14^+$ basal epithelial cells, $KRT10^+$ suprabasal epithelial cells, granular IVL^+ and LOR^+

epithelial cells, appendage-associated epithelial cells, channel-associated epithelial cells, melanocytic cells, immune cells, endothelial and lymphatic endothelial cells, fibroblasts, and putative Paget cells (Fig. 1c; Supplementary information, Fig. S3a). When split by conditions, these cell communities were dramatically altered both in epithelial and immune cells in EMPD skin, in which Paget cells appear specific to EMPD condition, and immune cells remarkably increased in EMPD condition (Fig. 1c; demarcated ovals).

To study the epithelial diversity of EMPD, we subclustered epithelial cells, and identified seventeen hierarchically distinct cell clusters including $PTTG1^+$, $RRM2^+$, Transitional$^{S\ Phase}$, $ASS1^+$, $POSTN^+$, $CHI3L1^+$, Transitional$^{Basal-I}$, Transitional$^{Basal-II}$, $KRT10^+$ type-I, $KRT10^+$ type-II, IVL^+, LOR^+, $KRT6C^+$ and $GJB2^+$ channel keratinocytes, Paget cells, as well as two clusters of putative intra/inter-follicular keratinocytes (Fig. 1d, e). Among them, $ASS1^+$, $POSTN^+$, $CHI3L1^+$ clusters only express a basal keratin marker KRT14, and $KRT10^+$ type-I, $KRT10^+$ type-II, IVL^+, LOR^+ clusters only express the suprabasal keratin marker KRT10, while the others express both KRT14 and KRT10, suggesting a transition state from basal to suprabasal keratinocytes (Supplementary information, Fig. S3b, c). Interestingly, Paget cells and $KRT6C^+$ keratinocytes appear specific to EMPD condition (Fig. 1d; Supplementary information, Fig. S3). It is worth noting that KRT6C is associated with poor prognosis and worse cancer outcome.[28] In comparison, two clusters of putative follicular keratinocytes appear specific to normal skin condition (Supplementary information, Fig. S4a). Along with RRM2, ASS1 and PTTG1 mark putative basal epidermal stem cells in neonatal human skin.[29] We observed that $ASS1^+$, $PTTG1^+$ and $RRM2^+$ keratinocytes are located and spatially segregated at both basal and expanded basal layers in EMPD skin, with $ASS1^+$ keratinocytes forming rosette-like structures around Paget cells and $PTTG1^+$ and $RRM2^+$ keratinocytes along Rete ridges (Fig. 1f–h). Next, we also found that epithelial cells in EMPD epidermis displayed higher scores for inflammatory genes, similar to psoriatic and scalp skin (Supplementary information, Fig. S4b). Taken together, this scRNA-seq data indicates that Paget cells and $KRT6C^+$ keratinocytes are identified as unique cell clusters in EMPD skin, exhibiting basal-like features and gene programs associated with cancer and inflammation.

We unequivocally identified Paget cells by implementation of a "Paget cell aggregate biomarker score" defined by a core set of known PD-associated biomarkers. Scores were projected onto UMAP space and quantified using violin plots. Scored cells localized to a cluster unique to EMPD condition, further corroborating this cluster represents Paget cells (Fig. 1i). Cell cycle state analysis[30] showed that ~52% of Paget cells were in G1 phase, while ~19% and ~29% of Paget cells were in G2/M and S phase, respectively. These data suggested that Paget cells are present in all phases of the cell cycle[31] (Supplementary information, Fig. S5a, c). Interestingly, when spatially resolved, $MKI67^+$ proliferative keratinocytes were present through the whole epidermis in EMPD, spanning basal and suprabasal cells, while they were strictly localized to the basal compartment in normal skin (Supplementary information, Fig. S5b, d).

Next, we performed differential gene expression analysis to identify genes uniquely expressed in Paget cells. Functional classification of these uniquely expressed genes with PANTHER[32] grouped them into distinct gene classes, including Nucleic Acid Binding, Signaling, Enzyme Modulation, Transcription Factors, Cytoskeleton, Transferases, Oxidoreductases, Intermediate Filament, and Receptor/Membrane ($P < 0.05$; Fig. 1j). High-density plots further demonstrated unique expression of intermediate filament genes including $KRT19$, $KRT8$, $KRT18$, $KRT7$, $MUC1$ and $MUCL1$, all previously established marker genes for Paget cells. Interestingly, previously unidentified transcription factors ($TFAP2B$, $FOXA1$, $FOXA3$, $SMYD3$, and $SPDEF$) and receptor/membrane ($CD163L1$ and $ALCAM$) genes were also uniquely expressed in Paget cells (Fig. 1k), suggesting the identification of novel marker
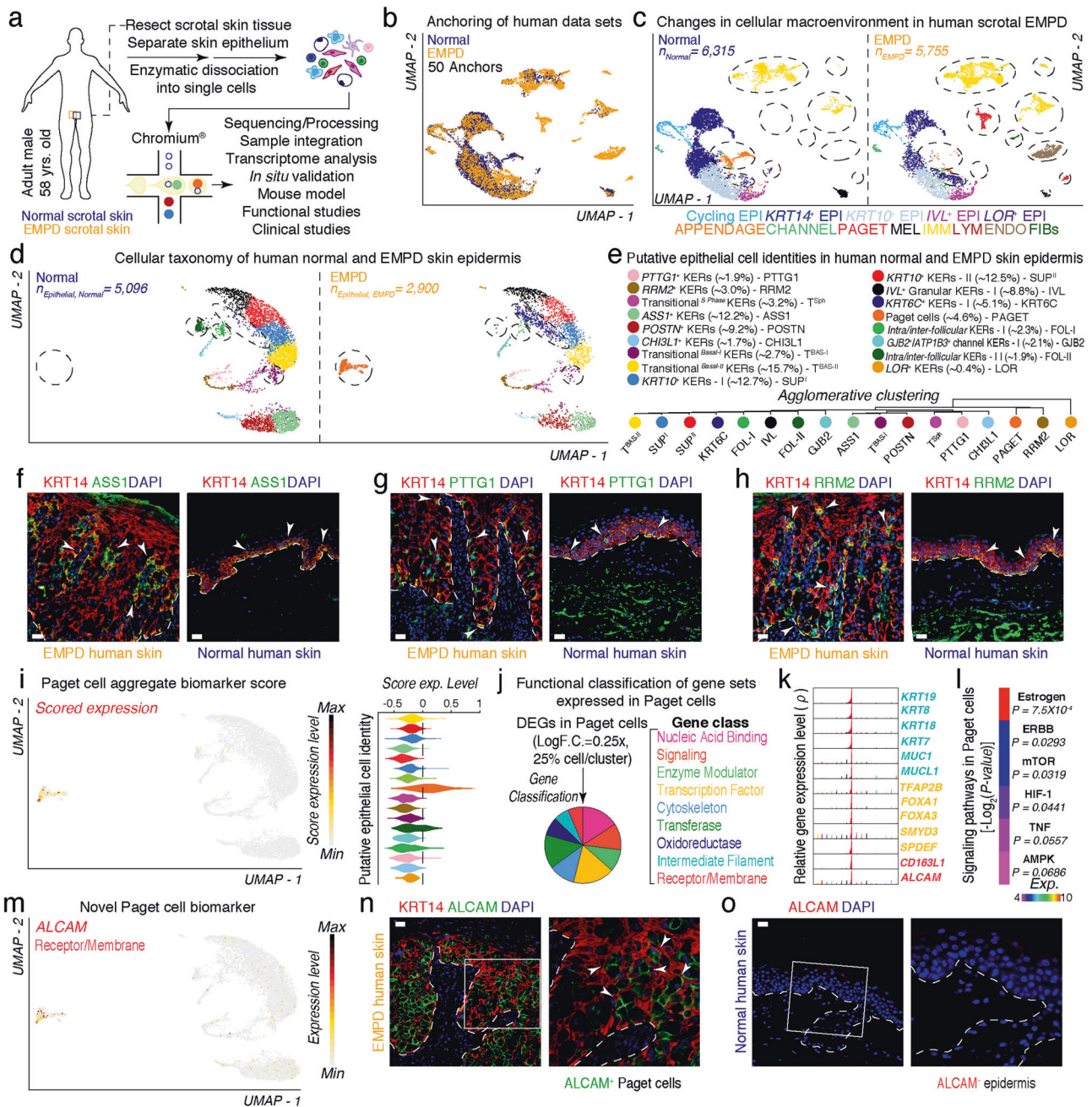
Fig. 1 Characterization of the epithelial macroenvironment of EMPD skin using 3′-end scRNA-seq. a Schematic of patient anatomy, cell isolation, processing, and capture of single cells by droplet-based device, 3′-end scRNA-seq, and downstream analyses. **b** Anchoring (50 anchors) of normal ($n = 6315$ viable cells) and EMPD ($n = 5755$ viable cells) data sets visualized in UMAP space. **c** Clustering and neighbor identification of anchored data sets split by condition. Thirteen putative cell communities were identified and color coded. Putative community identity identified by bona fide biomarker expression is defined, shown in Supplementary information, Fig. S3. **d** Epithelial cellular taxonomy of human normal ($n = 5096$ epithelial cells) and EMPD ($n = 2900$ epithelial cells) skin epidermis split by condition. Epithelial cells that appear to be condition-specific are demarcated with broken oval shapes. **e** Putative epithelial cell identity identified by bona fide marker expression with their respective proportions is defined. Agglomerative clustering demonstrates hierarchical relationships between epithelial cells. **f–h** Immunofluorescence for ASS1/KRT14, PTTG1/KRT14 and RRM2/KRT14 in EMPD and normal skin. White arrows point to ASS1+, PTTG1+, RRM2+ cells. **i** Paget cells were identified by implementation of a 'Paget cell aggregate biomarker score' defined by a core set 10 PD-associated biomarkers. These PD biomarkers included: *KRT8, AR, PDPN, GATA3, MYD88, MUC1, KRT7, VIM, MMP9,* and *PIP*. Scored cells were projected on UMAP, evaluated as feature plot and quantified using violin plot. Gene expression levels are color coded. **j** Functional classification of select gene sets expressed in Paget cells (Log F.C. = 0.25×, 25% cell/cluster, Wilcox test, $P < 0.05$). Selected gene sets were plotted using a pie chart and gene classes were color coded. **k** High-density plots show distribution of select Paget cell-specific transcripts including known and novel biomarkers. **l** Heatmap showing signaling pathways activated in Paget cells. These pathways included Estrogen ($P = 4.75 \times 10^{-3}$), ERBB ($P = 0.0293$), mTOR ($P = 0.0319$), HIF-1 ($P = 0.0441$), TNF ($P = 0.0557$), AMPK ($P = 0.0686$). **m** Distribution of *ALCAM* expression overlaid on feature plot shows high and unique expression in Paget cells. Immunofluorescence of KRT14 and ALCAM in EMPD skin (**n**) and human normal skin (**o**). Insets represent magnified areas. Representative images are shown. Epidermis and dermis are demarcated with broken line. Scale bars, 25 μm (**f–h, n, o**).

genes for Paget cells. Among them, activated leukocyte cell adhesion molecule (ALCAM, i.e., CD166) was selected for further validation. ALCAM partially co-localized with CAM5.2, a known biomarker of Paget cells, and was uniquely expressed in Paget cells in human EMPD ($n = 4$) and MPD ($n = 4$), but not in normal skin ($n = 4$) (Fig. 1m–o; Supplementary information, Fig. S6). We therefore identify ALCAM as a novel biomarker of Paget cells in human PD. Given that ALCAM is a membrane protein, it warrants the possibility to physically sort Paget cells from human EMPD and MPD skin for downstream genomic and epigenetic profiling. Further interrogation of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways[33] using differentially expressed genes (DEGs) identified enriched signaling pathways overrepresented in Paget cells, including Estrogen, ERBB, mTOR, HIF-1, TNF and AMPK (Fig. 1l). These results suggest that these pathways might be important for regulating a Paget cell state in human EMPD.

Characterization of the immune cellular taxonomy of human extramammary Paget's Disease at single-cell resolution

Next, we subclustered immune cells based on expression of pan hematopoietic marker PTPRC (i.e., CD45) (Supplementary information, Fig. S7a). Analysis of immune cells identified eleven distinct cell clusters, including $CD4^+$ T cells, natural killer T (NKT) cells, $LYZ2^+$ myeloid cells, $CD207^+$ Langerhans cells, $LAMP3^+$ Langerhans cells, B cells, Mast cells, plasmacytoid dendritic cells and cycling T cells (Supplementary information, Fig. S7a, b). We observed a higher proportion of immune cells — of all types, in EMPD skin compared to normal skin (Supplementary information, Fig. S7c). Particularly, mast cells, $LAMP3^+$ Langerhans cells, $LYZ2^+$ myeloid cells, B cells, plasmacytoid dendritic cells and cycling T cells account for over 85% of cell cluster identity. These differences in immune cell repertoire and their proportions may account for the inflammatory phenotype observed in EMPD skin and also represent an adaptive immune response in EMPD.

We then analyzed T cell states in EMPD. To do this, we implemented a "cytotoxic-, exhaustive-, and naïve-T cell aggregate biomarker" score defined by a core set of known T cell state-associated biomarkers.[34] We quantified these T cells and demonstrate that $CD8^+$ T cells display a cytotoxic profile in EMPD. Specifically, NKT cells — a subset of $CD8^+$ T cells, express genes coding for cytolytic molecules, including GZMK (Supplementary information, Fig. S7d). These results are suggestive that EMPD-infiltrating $CD8^+$ T cells display cytotoxic activity. Although EMPD-infiltrating $CD8^+$ T cells display higher cytotoxicity activity than normal skin $CD8^+$ T cells, a previous report shows that EMPD-infiltrating $CD8^+$ T cells have impaired cytotoxic activity compared to $CD8^+$ T cells in PBMCs, suggesting that although EMPD-infiltrating $CD8^+$ T cells in our data set display a cytotoxic phenotype, this may not be sufficient to drive a strong adaptive immune response against Paget cells.[35] We also observe exhausted $CD4^+$ T cells in EMPD skin, suggesting impaired cytotoxicity. Furthermore, these cells have an absent cytotoxicity profile (Supplementary information, Fig. S7e, f). These results suggest that cytotoxic activity is potentially impaired in EMPD-infiltrating $CD8^+$ T cells and that a hyporesponsive state might exist in $CD4^+$ T cells in the EMPD microenvironment (Supplementary information, Fig. S7e–g).

Ectopic Msi1 overexpression in mouse epithelium drives a Paget-like phenotype

RNA-binding protein MSI1 acts as a driver of oncogenic transformation in the intestine.[23,24] Interestingly, we observed that MSI1 is highly overexpressed in EMPD basal epithelium, but not in basal epithelial cells in normal skin. Count density distribution and mRNA expression extracted from scRNA-seq data demonstrated that MSI1 is overexpressed in two distinct basal cell types in EMPD skin, including $POSTN^+$, and $CHI3L1^+$ cell populations, and in a transitional $GJB2^+$ cell population (Fig. 2a). Bulk gene expression

profiling of human PD skin further revealed prominent MSI1 mRNA upregulation in EMPD relative to normal skin (Fig. 2b). MSI1 upregulation in EMPD skin was further confirmed at the protein level (Supplementary information, Fig. S8a). In normal skin, MSI1 is largely restricted to the suprabasal layers of the epidermis; however, in 14 out of 20 human EMPD skin samples analyzed, MSI1 was found to be ectopically expressed in Cytokeratin 14-positive ($KRT14^+$) basal epidermal cells (Fig. 2c), further reinforcing our observations from scRNA-seq data. Similarly, we also detected ectopic MSI1 overexpression in basal epidermal cells in 4 out of 5 human MPD skin samples (Fig. 2c). Together, these data reveal that ectopic MSI1 overexpression in the basal epidermal layer of skin is a feature for a subpopulation of human EMPD and MPD.

To investigate the functional consequences of ectopic Msi1 overexpression in the basal epidermal layer of skin in vivo, we generated Krt14-rtTA;TRE-Msi1 double transgenic (DTG) mice, in which Msi1 is overexpressed in $Krt14^+$ basal keratinocytes upon continuous Doxycycline administration (Supplementary information, Fig. S8b). Doxycycline-mediated Msi1 overexpression was validated in DTG mouse skin (Supplementary information, Fig. S8c). Normally, and similar to human skin, low levels of Msi1 expression are detected in the suprabasal epidermal compartment of wild-type (WT) mice, while Msi1 is specifically induced in the basal epidermal layer of DTG mice upon continuous Doxycycline administration (Supplementary information, Fig. S8d). After 72 h of continuous Doxycycline administration, control mice ($n = 8$) were viable and did not display any apparent gross phenotypes, while DTG littermates exhibited significant reductions in body weight and temperature ($n = 5$, Supplementary information, Fig. S8e). Grossly, Doxycycline-induced mice ($n = 12$) were hunched and displayed signs of dehydration and lethargy (Fig. 2d). DTG mice ($n = 12$) rapidly succumbed within 4–7 days of initial Doxycycline administration (Fig. 2e). Accordingly, the mean rate of trans-epidermal water loss (TEWL) was significantly higher in DTG mice ($n = 6$) than their littermate controls ($n = 6$) (Fig. 2f), mimicking TEWL in human EMPD patients. In line with these observations, the mean rate of TEWL was significantly higher in EMPD skin ($n = 3$) compared to normal, healthy skin ($n = 3$) (Fig. 2g). These results suggest a defective skin barrier, which may account for the weight loss and dehydration phenotypes observed in DTG mice, and that TEWL is an unappreciated feature of EMPD human skin.

To evaluate the histopathology of skin associated with Msi1 overexpression, we histologically assessed DTG mouse skin isolated at various time points after continuous Doxycycline administration. After 24 h of Doxycycline administration, the epidermis of DTG mice became prominently thickened; additionally, the basal layer became disorganized with many narrow and vertically elongated cells (Supplementary information, Fig. S9). After 36 h, Paget-like cells start to appear in the epidermis of DTG mice (Supplementary information, Fig. S9). After 48 h, Paget-like cells became more prominent and apparent (Fig. 2h, i). These cells were large and had a clear cytoplasm and eccentric, hyperchromic nucleus — a known characteristic of human Paget cells (Fig. 2j). Collectively, our results suggest that ectopic Msi1 overexpression in basal keratinocytes induced a morphologically relevant Paget-like phenotype in mouse epidermis.

To further characterize the Paget-like phenotype, we evaluate the gene expression pattern of Paget-like cells. These cells expressed many of the established biomarkers for human PD, including Mucin1, Ceacam1, Krt8, Cam5.2, Ar, Podoplanin, Gata3, Myd88, Krt7, Vimentin, Cd31, Mmp9 and Ema (Fig. 2k–n; Supplementary information, Fig. S10a–j), while they were not detected in control epidermis in mice. We also validated and corroborated the expression of KRT14/KRT10, KRT7, KRT8, CAM5.2, EMA, Vimentin GATA3, Podoplanin and MMP9, as well as Periodic Acid-Schiff (PAS) staining in human MPD and EMPD tissue samples (Supplementary information, Fig. S11). PAS staining is a key marker of human PD.
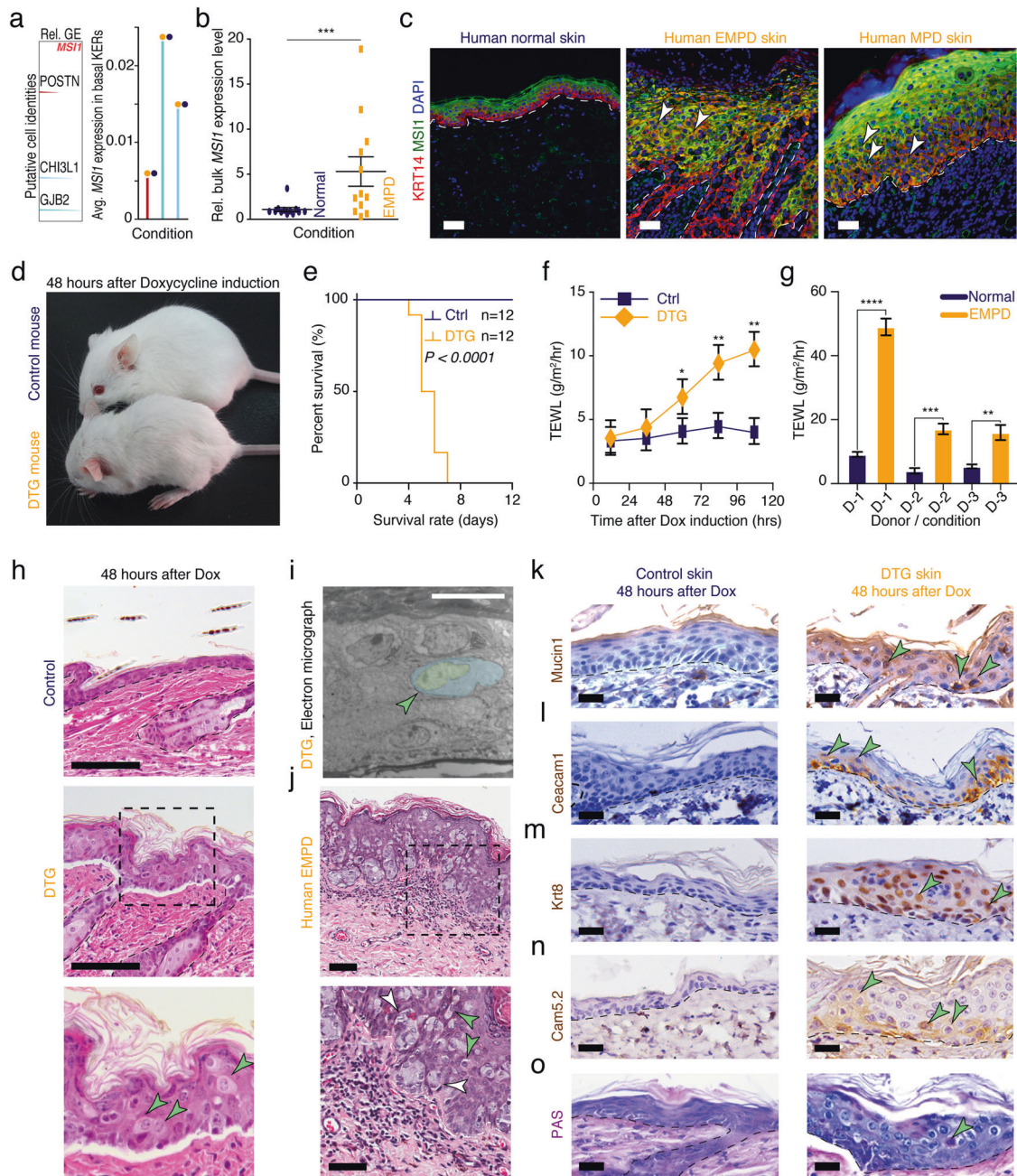
Fig. 2 Ectopic Msi1 overexpression results in Paget-like phenotype in murine skin. a *MSI1* expression is absent in basal epithelial cells in normal human skin but highly expressed in three distinct basal cell types in EMPD human skin. High-density plots showing distribution of *MSI1* transcript in distinct human epithelial cell clusters (left). Average mRNA expression of *MSI1* transcript in distinct human epithelial cell clusters was shown (right). **b** qRT-PCR analysis of *MSI1* in normal and EMPD human skin tissues. Relative mRNA expression values were normalized to *GAPDH*. ***$P < 0.001$. **c** Representative images of immunofluorescence staining against MSI1 and KRT14 in normal ($n = 5$), EMPD ($n = 20$), and MPD ($n = 5$) human skin tissues. White arrows point to MSI1/KRT14 double-positive keratinocytes. **d** Gross images of 6-week-old littermate control and *K14-rtTA;TRE-Msi1* DTG mice 48 h after continuous Doxycycline treatment. DTG mice show signs of lethargy and dehydration. **e** Kaplan-Meir plots demonstrating percent survival rates of control ($n = 12$) and DTG ($n = 14$) mice after 7 days of continuous Doxycycline treatment. $P < 0.0001$. **f** TEWL in skin from littermate control ($n = 6$) and DTG ($n = 6$) mice 0, 24, 48, 72 and 96 h after continuous Doxycycline treatment. *$P < 0.05$; **$P < 0.01$. **g** TEWL in normal ($n = 3$) and diseased ($n = 3$) skin from human EMPD patients. ****$P < 0.0001$; ***$P = 0.003$; **$P = 0.0017$. **h** Histology of control and DTG skin 48 h after continuous Doxycycline treatment. Black dashed inset indicates high magnification images on the bottom. Arrows point to Paget-like cells. **i** False-colored electron micrograph demonstrating characteristics of Paget-like cell in DTG mice. **j** Histopathology of EMPD human skin. Black dashed inset indicates high magnification image on the bottom. White arrows point to a group of Paget cells and green arrows point to individual Paget cells. **k–n** Immunohistochemical staining against Mucin1 (Muc1), Carcinoembryonic Antigen Related Cell Adhesion Molecule 1 (Ceacam1), Cytokeratin 8 (Krt8) and Cam5.2 in control and DTG mouse dorsal skin 48 h after continuous Doxycycline treatment. **o** Periodic Acid-Schiff (PAS) staining in control and DTG mouse dorsal skin 48 h after continuous Doxycycline treatment. Epidermis and dermis are demarcated with broken line. Representative images are shown. Epidermis and dermis are demarcated with broken line. Values in the graphs represent means ± SD. Student's *t*-test was used for calculating *P* values in **b**, **e–g**. Scale bars, 50 μm (**c**, **h**); 100 μm (**f**); 10 μm (**i**); 25 μm (**j–o**).

PAS$^+$ cells were also detected in the epidermis of DTG mice, but not in littermate control epidermis (Fig. 2o; Supplementary information, Fig. S11h). Moreover, Msi1 induction caused expansion of Krt14$^+$ basal cells into the suprabasal layers, with concomitant loss of Krt10$^+$ suprabasal cells (Supplementary information, Fig. S10k), further resembling human EMPD and MPD (Supplementary information, Fig. S11i). Together, these findings reveal that ectopic *Msi1* overexpression in the mouse epidermal basal layer leads to the acquisition of many morphological and immunohistochemical features of Paget cells in human PD.

scRNA-seq analysis identifies and characterizes Paget-like cells in Msi1-overexpressing mouse epidermis

To further characterize the Msi1-induced Paget-like phenotype, we isolated viable, single cells from control and DTG newborn (~P0.5) mouse epidermis induced with Doxycycline in utero and resolved their individual cell transcriptomes using scRNA-seq as before (Fig. 3a; Supplementary information, Fig. S12). It is worth noting that DTG newborn mouse epidermis exhibited milder, yet similar Paget-like phenotype to adult DTG mice (Supplementary information, Fig. S13a). After collective quality control analyses,[26] 18,872 valid cells remained (Control, $n = 10,843$ vs DTG, $n = 8029$) and were used for downstream analyses (Supplementary information, Fig. S14). We anchored and clustered control and DTG cells using Seurat as before (Fig. 3b). We then assigned cell communities to their putative identities based on expression of pan biomarkers (Fig. 3c). These communities included *Msi1*$^+$ basal keratinocytes, suprabasal *Krt10*$^+$ keratinocytes, *Krt79*$^+$ upper hair follicle keratinocytes, Paget-like cells, Merkel cells, melanocytic cells, sebaceous gland cells, and inter-/intra-follicular keratinocytes (Fig. 3c). *Msi1* overexpression was confirmed at RNA and protein level (Fig. 3d; Supplementary information, Fig. S13b). We overlaid *Msi1* expression on UMAP and quantified it using violin plots. *Msi1* expression was specific to basal keratinocytes as well as Paget-like cells. When split by condition, we observed higher expression of *Msi1* in cells in these clusters in DTG compared to control[36] (Fig. 3d, e). Paget-like cells expressed Paget cell-specific biomarker genes (Fig. 3f). We then interrogated the KEGG pathways of DEGs and identified enriched signaling pathways overrepresented in Paget-like cells. These pathways were similar to those in Paget cells in human and included Estrogen ($P = 4.9 \times 10^{-5}$), Hif-1 ($P = 0.00012$), mTOR ($P = 0.0056$), Tnf ($P = 0.0067$), ErbB ($P = 0.0088$), Ampk ($P = 0.0404$) (Fig. 3g). These results suggest that these pathways may also be important for driving and maintaining a Paget-like cell state in DTG mouse epidermis. To determine the level of similarity between Paget and Paget-like cells at the transcriptome level, we identified shared biomarkers between Paget and Paget-like cells by selecting all DEGs in each cluster (LogF.C. = 0.25×, 25% cell/cluster, Wilcox test, $P < 0.05$). We identified a total of 63 genes that were shared between both cell types in human and mouse (Fig. 3h; Supplementary information, Fig. S15).

We performed integrated analysis on our aligned data sets to identify novel molecular biomarkers associated with Paget-like cells in DTG mice that translated to human Paget cells. To do this, we split and visualized control and DTG conditions in tandem and focused on features that were found exclusively in DTG and specifically restricted to Paget-like cell cluster. Using this approach, we identified Paget-like cell-specific gene expression profiles, which included nuclear and cytoplasmic genes such as *S100a9*, *Tslp* and *Shroom3* (Fig. 3i, k, m). S100a9 and Tslp have been reported to be important for epidermal barrier formation,[37–39] while the function of Shroom3 has not been defined in epidermis.[40] The specific expression of all identified markers in Paget-like cells in DTG epidermis was validated by immunofluorescence (Fig. 3j, l, n). We next wondered whether these newly identified biomarkers were also presented in human EMPD Paget cells. Interestingly, S100A9 appeared to be expressed primarily in Paget cells present in upper strata of epidermis, similar to that in DTG mouse (Fig. 3j).

In contrast, TSLP and SHROOM3 specifically mark the majority of Paget cells in basal and suprabasal layers (Fig. 3l, n). In sum, we identified Paget-like cell-specific gene expression patterns that translated to human EMPD Paget cells, further suggesting that DTG mice can not only recapitulate the human PD phenotype, but also be used to identify associated molecular biomarkers of Paget cells in human.

Pseudotime and RNA velocity analyses support the keratinocyte-Paget cell in situ transformation theory in PD

Two competing models have been proposed to explain the origin of Paget cells in PD: the Epidermotropic and the in situ transformation theories.[3] Progress in understanding the origin of Paget cells in PD is hampered by the lack of mouse models that can recapitulate the condition and PD-associated phenotypes. To validate the utility of our mouse model and further corroborate our findings from human EMPD, we inferred a putative *Msi1*$^+$ keratinocyte-Paget-like cell transformation trajectory in pseudotime. This trajectory enabled the reconstruction of a major path, defined as M-PATH 1, where *Msi1*$^+$ basal keratinocytes were placed at the beginning and Paget-like cells at the trajectory terminus (Fig. 4a). To assign directionality to M-PATH 1, we performed RNA velocity analysis. We estimated RNA velocity vectors on Monocle2-ordered *Msi1*$^+$ keratinocytes and Paget-like cells and projected them along M-PATH 1 ($\vec{V}_{M\text{-}PATH1}$). Velocity vectors progressively point toward M-PATH 1 terminus, an area with increased density of Paget-like cells (Fig. 4b), suggesting a transition from *Msi1*$^+$ KERs to Paget-like cells. We then performed scEpath analysis on Monocle2-ordered cells and identified a total of 657 pseudotime-dependent DEGs along M-PATH 1 ($a = 0.05$), which were divided into four distinct segments (mpC1-mpC4) (Fig. 4c, top panel). Furthermore, we identified pseudotime-dependent transcription factors (TFs) that could be important for the development of Paget-like cells (Fig. 4c, bottom panel); these included *Hmgb2*, *Nfib*, *Sox4*, *Hoxa7*, *Gata3*, *Hoxcd8*, *Id3*, *Nfia*, *Tsc22d1*, *Id1*, *Atf3*, *Klf4*, *Zbtb16*, *Cepba*, *Grhl3*, *Rora*, *Ovol1*, *Vdr*, and *Barx2*. Together, our data revealed a putative *Msi1*$^+$ keratinocyte-to-Paget-like cell transformation trajectory in DTG mice.

Our pseudotime and RNA velocity analyses suggest that basal keratinocytes may be able to adopt a Paget/Paget-like cell fate. To further corroborate these observations and determine the cellular origin of Paget-like cells in vivo, we generated inducible *Krt14-rtTA; Tre-Msi1;Krt14-CreERT2;mTmG* quadrupled transgenic (QTG) mice (Supplementary information, Fig. S16), in which Cre recombinase is induced in basal keratinocytes upon tamoxifen treatment, enabling tracing the progeny lineages of basal keratinocytes. Intriguingly, the ventral skin gave rise to more severe Paget-like phenotypes (Supplementary information, Fig. S17). Examination of the lineage identity of Paget-like cells in dorsal and ventral skin of QTG mice revealed that Paget-like cells are derived from *Msi1*-overexpressing basal keratinocytes (Fig. 4d, e; Supplementary information, Fig. S17). These results lend further evidence in support of the in situ transformation theory in PD and suggest that basal keratinocytes may adopt a Paget/Paget-like cell fate.

Msi1-induced Paget-like mouse skin shares signaling pathway profiles with human EMPD skin

To gain insights into the molecular mechanisms driving the pathogenesis of PD, we compared whole dorsal skin transcriptomes between DTG ($n = 4$) and control mice ($n = 4$) 24 h after continuous Doxycycline induction and human EMPD ($n = 4$) and normal ($n = 3$) whole skin tissues. Two- and three-dimensional principal component analysis revealed significant separation between mouse and human samples, yet close clustering between biological replicates — independent of tissue origin or condition (Fig. 5a). We identified 547 down- and 680 up-regulated genes in DTG compared to control skin (Fig. 5b). In contrast, there were 1782 down- and 2591 up-regulated genes in human EMPD compared to normal skin (Fig. 5b). Selected DEGs were confirmed
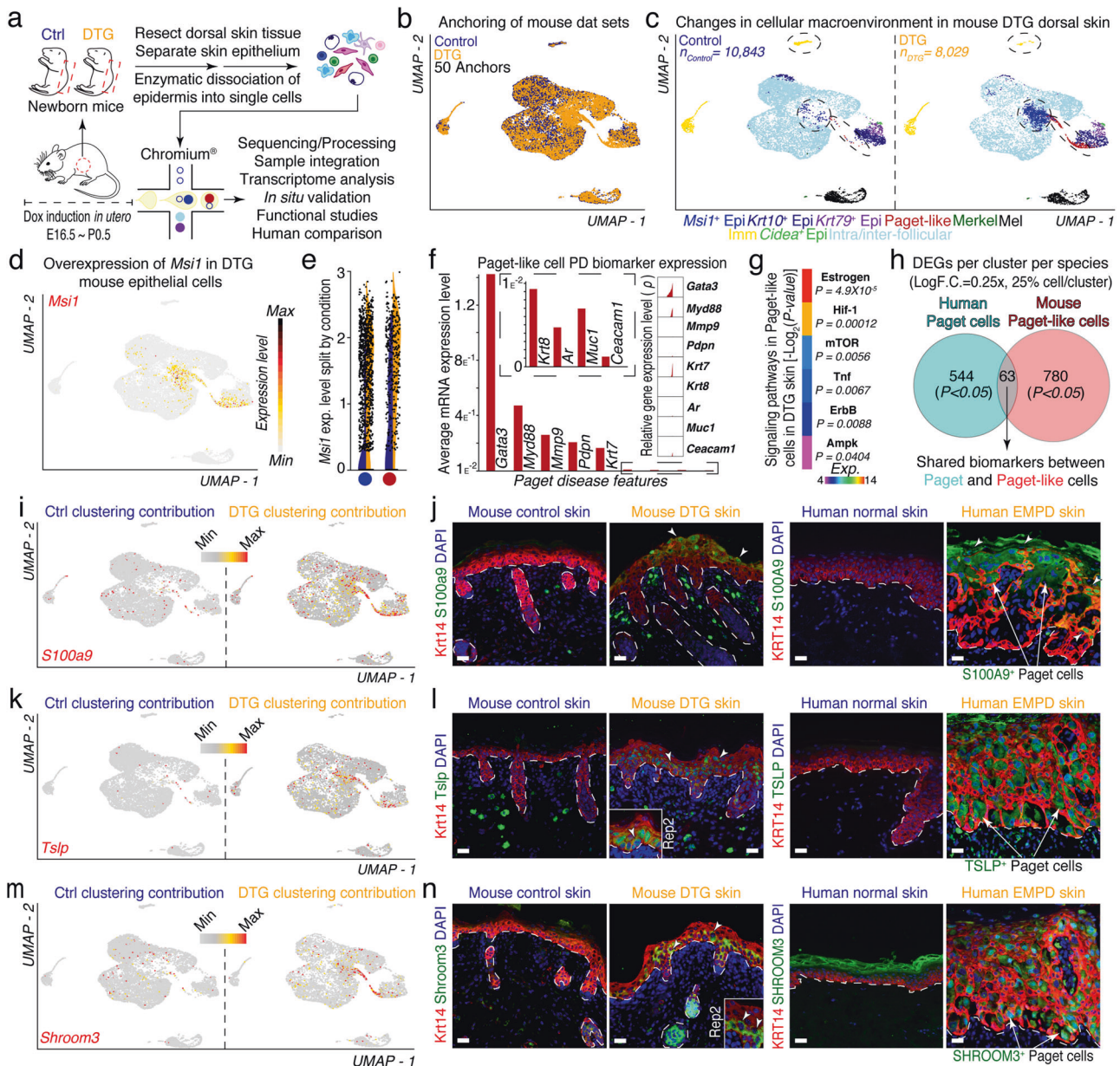
Fig. 3 Characterization of Paget-like cells in DTG murine skin epidermis using 3'-end scRNA-seq. a Schematic of Paget-like phenotype induction, cell isolation, processing, and capture by droplet-based device, 3'-end scRNA-seq, and downstream analyses. **b** Anchoring (50 anchors) of control ($n = 10,843$ viable cells) and DTG ($n = 8029$ viable cells) data sets visualized in UMAP space. **c** Clustering and neighbor identification of anchored data sets. Nine putative cell communities were identified and color coded. Putative cluster identity is defined on the right. **d** Distribution of *Msi1* expression overlaid on feature plot shows high expression in basal and Paget-like cells. **e** Violin plots shows distinct *Msi1* expression in basal and Paget-like cells in control and DTG condition. **f** High-density plots and average expression of selected Paget disease biomarkers expressed in the putative Paget-like cell cluster. **g** Heatmap showing signaling pathways activated in Paget-like cells. These pathways included Estrogen ($P = 4.9 \times 10^{-5}$), Hif-1 ($P = 0.00012$), mTOR ($P = 0.0056$), Tnf ($P = 0.0067$), ErbB ($P = 0.0088$), Ampk ($P = 0.0404$). **h** Venn diagram showing shared biomarkers between human Paget cells and mouse Paget-like cells. Shared biomarkers between Paget and Paget-like cells were identified by selecting all DEGs in each cluster (LogF.C. = 0.25×, 25% cell/cluster, Wilcox test, $P < 0.05$) and then comparing them with each other. A total of 63 genes were shared between both cell types. **i, k, m** Feature plots split by condition identified Paget-like cell-specific gene profiles, including *S100a9*, *Tslp*, and *Shroom3*. Expression level is restricted to Paget-like cell cluster and specific to DTG condition (Ctrl clustering contribution vs DTG clustering contribution). **j, l, n** Left: Immunofluorescence of novel biomarkers on control and DTG skin identified enriched localization of Krt14 with S100a9, Tslp, or Shroom3 in Paget-like cells (white arrows). Insets represent magnified replicate stains. Right: Staining of novel biomarkers on normal and EMPD human skin identified enriched localization of KRT14 with S100A9, TSLP, or SHROOM3 in Paget cells in EMPD human skin (white arrows). Representative images are shown. Epidermis and dermis are demarcated with broken line. Scale bars, 25 μm (**j, l, n**).

by qRT-PCR analysis (Supplementary information, Fig. S18a). 86 down- and 169 up-regulated gene orthologs were shared between mouse and human datasets (LogF.C. > 1.2×) (Fig. 5c). We then interrogated Gene Ontologies and KEGG pathways of DEGs

overrepresented in DTG mouse and human EMPD skin. Strikingly, 33 out of 46 enriched signaling pathways in DTG mouse skin were shared with the human EMPD profile (Supplementary information, Figs. S18b, S19, S20). These included ribosome, PI3K-AKT, MAPK,
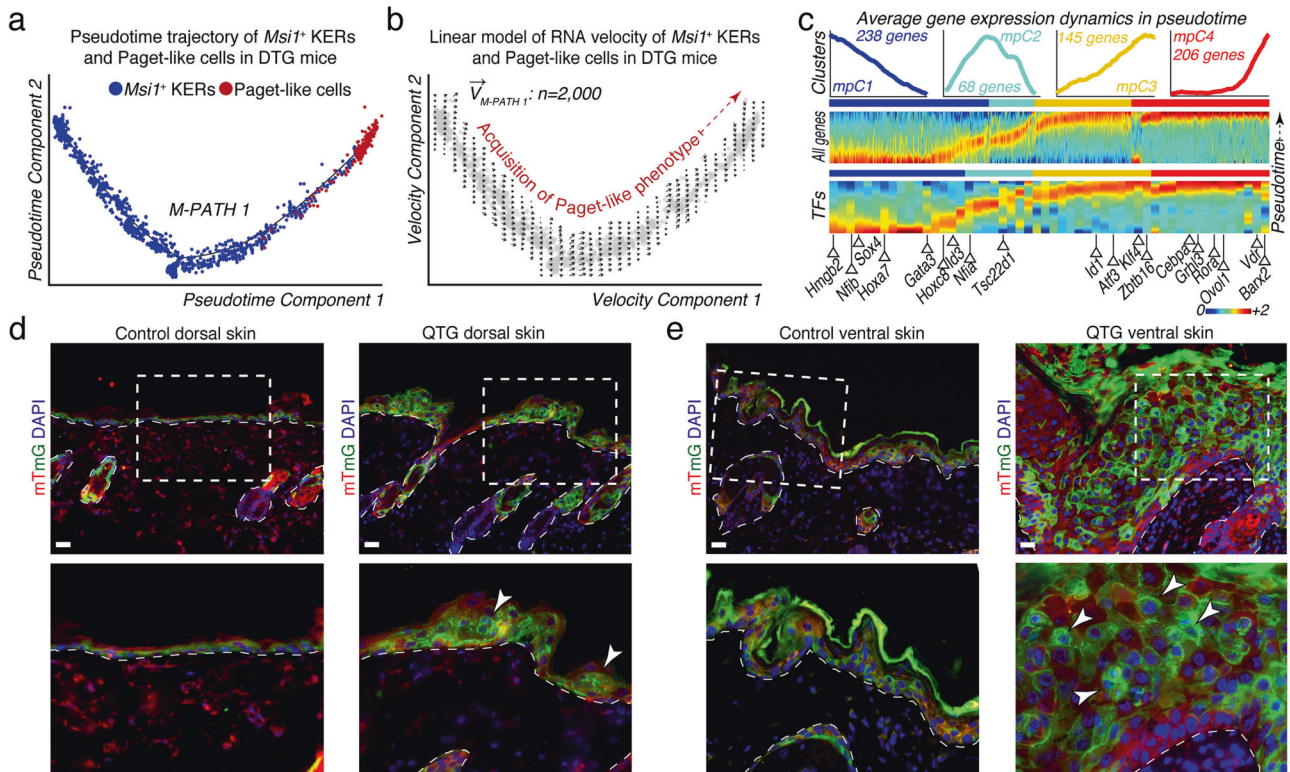
Fig. 4 RNA dynamics and lineage tracing analyses favor the basal keratinocyte-Paget cell in situ transformation theory. a Pseudotime analysis of mouse *Msi1*+ basal keratinocytes and Paget-like cells reconstructed a major path (M-PATH 1). *Msi1*+ basal KERs were placed at the beginning and Paget-like cells at the trajectory terminus. All cells on the pseudotime are color-coded to match the colors in Fig. 3c. **b** Deterministic model of RNA velocity distinguished one set of velocity vectors across pseudotime space pointing toward Paget-like cells. **c** Top: scEpath analysis performed on pseudotime-ordered cells identified 657 pseudotime-dependent genes arranged into four gene clusters ($a = 0.05$) (mpC1 through mpC4). Rolling wave plot shows expression levels for all pseudotime-dependent genes. Bottom: Rolling wave plots for selected transcription factors (TFs) identified as pseudotime-dependent. Selected TFs are labeled. TFs previously implicated in epidermal homeostasis and differentiation are labeled blue. **d, e** Lineage tracing of Paget-like cells in dorsal (**d**) and ventral skin (**e**) from *Krt14-rtTA;TRE-Msi1;Krt14-CreER;mTmG* quadrupled transgenic (QTG) mice (*n* = 4). Control *Krt14-rtTA;Krt14-CreER;mTmG* mice (*n* = 4) do not form Paget-like cells. Lineage tracing studies demonstrate that Paget-like cells originate from *Msi1*+ basal keratinocytes (white arrows). Insets show higher magnification areas on the bottom panels. Representative images are shown. Epidermis and dermis are demarcated with broken line. Scale bars, 25 μm (**d, e**).

ubiquitin-mediated proteolysis, RNA transport, and RAS signaling pathways (Supplementary information, Fig. S18b). The uniquely altered pathways in the DTG mouse skin include small cell lung cancer-associated, pyrimidine synthesis, proteasome-associated and TNF signaling pathways (Supplementary information, Figs. S18b, S19). The uniquely altered pathways in the human EMPD skin samples include adhesion molecule-associated, lysozyme-associated, measles-associated, herpes simplex infection-associated and cAMP signaling pathways (Supplementary information, Figs. S18b, S20). Of interest, the PI3K-AKT, mTOR, WNT, RAS and ERBB signaling pathways, all of which are important for tumorigenesis in many tissues, were significantly overrepresented in both DTG mouse and human EMPD skin (Fig. 5d). Indeed, mTOR and ERBB signaling pathway-associated genes were upregulated in DTG mouse and human EMPD skin tissues (Fig. 5e, f). Consistently, Gene Set Enrichment Analysis (GSEA) showed that the majority of significantly enriched gene sets in DTG mouse skin (15 out of 19) were also enriched in human EMPD skin, including those related to the TNF-α/NFκB signaling, Inflammatory response, p53, Androgen response, DNA repair, and MYC pathways (Supplementary information, Fig. S21). These common gene sets all positively correlated with gene programs both in the DTG and EMPD profiles. The uniquely enriched gene sets in DTG mouse skin include ROS, TGF-β, KRAS and coagulation signaling pathways, whereas the uniquely enriched gene sets in human EMPD skin included pathways associated with mitotic spindle, and G_2M checkpoint

(Supplementary information, Fig. S21). Taken together, our bulk microarray analysis suggests that Msi1-overexpressing skin epidermis shares common signaling pathways with human PD.

To identify gene expression differences and learn keratinocyte-specific responses associated with *Msi1* upregulation, we performed comparative analyses on aligned control and DTG keratinocytes. We identified 84 significantly upregulated genes in DTG *Msi1*-overexpressing mouse keratinocytes — these genes included *Krt6a* (> 1.25×), *Cdkn1a* (> 0.67×), *Nfkbia* (> 1.05×), *Fst* (> 0.99×), *Areg* (> 0.75×), *Slc2a1* (> 0.71×), *Dusp6* (> 0.55×), and *Hbegf* (> 0.79×) (Fig. 5g). Signaling pathways overrepresented in DTG *Msi1*-overexpressing mouse keratinocytes included ErbB, TGF-β, Tnf, Jak-Stat, Mapk, and Pi3k-Akt-mTOR signaling pathways (Fig. 5h), further corroborating our findings from bulk microarray profiling. We identified 116 significantly upregulated genes in EMPD keratinocytes compared to normal keratinocytes (F.C. > 0.50×) (Fig. 5i). Similar to *Msi1*-overexpressing mouse keratinocytes, PI3K-AKT and mTOR pathways are also overrepresented in human basal keratinocytes (Fig. 5j). On analysis, we found upregulated mTORC1 and HER2 pathway genes in Paget cells in human EMPD, including *ERBB2*, and *EIF4EBP1* (Fig. 5e, f). Taken together, our scRNA-seq comparative analysis suggests that Msi1-overexpressing basal keratinocytes share common signaling pathways with human EMPD basal keratinocytes, including mTOR and HER2, suggesting that these pathways might be important for the acquisition of PD and PD-like phenotypes.
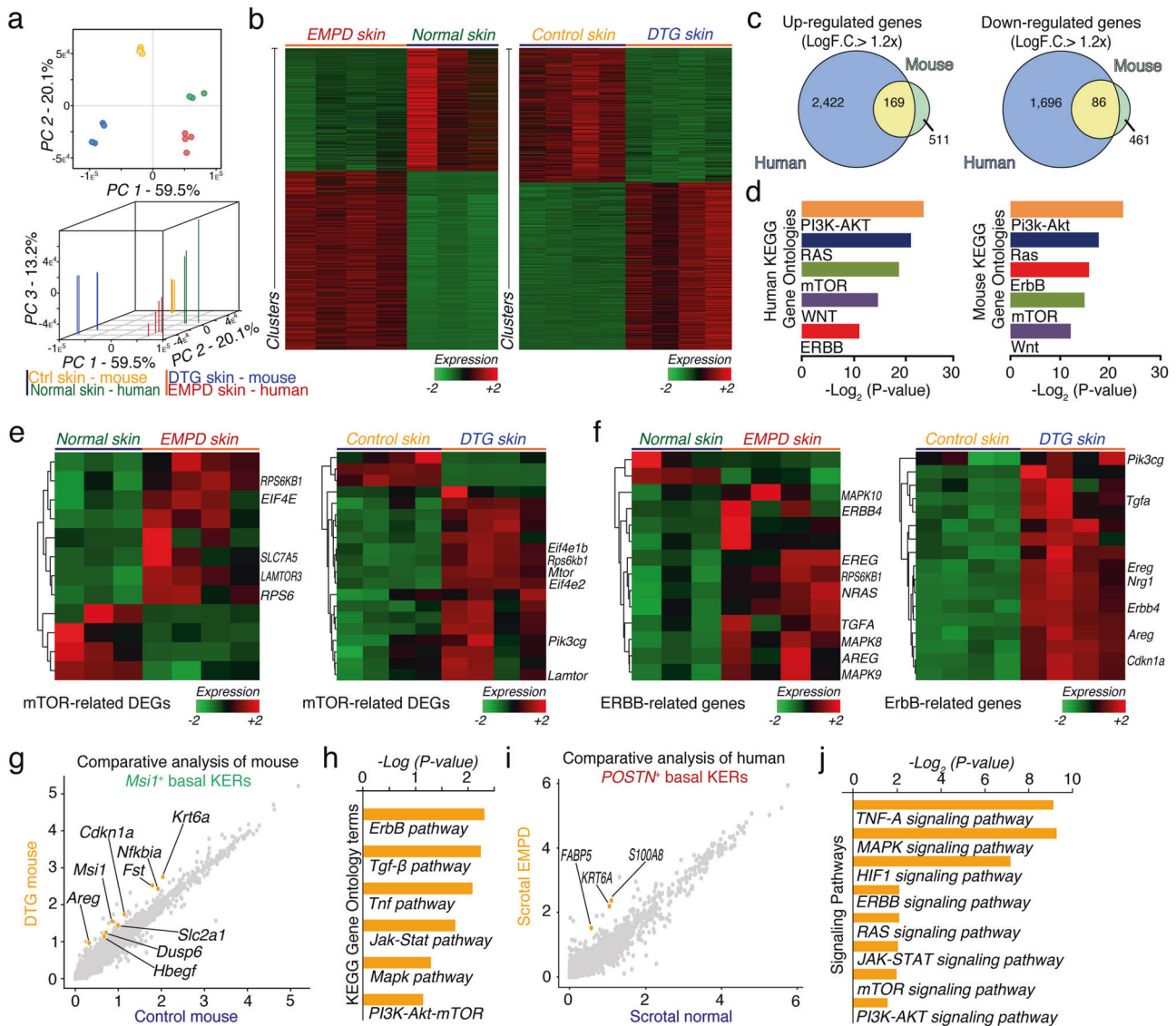
Fig. 5 Murine Msi1-induced Paget-like skin and EMPD human skin share common molecular signaling pathways. a Two- and three-dimensional Principal Component Analysis of mouse and human bulk microarray samples. **b** Heatmaps of DEGs between normal ($n = 3$) and EMPD ($n = 4$) human skin tissues; and control ($n = 4$) and DTG ($n = 4$) mouse skin tissues. **c** Venn diagrams showing overlapping down- and up-regulated genes from human EMPD and mouse comparisons (LogF.C. > 1.2×; $P < 0.05$). A total of 86 down-regulated and 169 up-regulated genes overlap between human and mouse. **d** Gene Ontology (GO) and KEGG pathway analyses of DEGs between human and mouse. PI3K-AKT, RAS, mTOR, WNT and ERBB signaling pathways are selectively represented. **e** Heatmaps of mTOR-related DEGs between normal ($n = 3$) and EMPD ($n = 4$) human skin; and control ($n = 4$) and DTG ($n = 4$) mouse skin tissues. Selected genes are labeled. **f** Heatmaps of ERBB-related DEGs between normal ($n = 3$) and human EMPD ($n = 4$) skin; and control ($n = 4$) and DTG mouse skin ($n = 4$). Select genes are labeled. **g** Comparative analysis on aligned control and DTG keratinocytes identifies gene expression differences and basal keratinocyte-specific responses associated with upregulation of Msi1. Approximately 84 genes were differentially overexpressed between both conditions (F.C. > 0.5×). **h** KEGG GO terms upregulated in $Msi1^+$ keratinocytes. ErbB, Tgf-β, Tnf, Jak-Stat, Mapk, and Pi3k-Akt-mTOR signaling pathways are selectively represented. **i** Comparative analysis on aligned normal and EMPD keratinocytes identifies gene expression differences and KER-specific responses associated with EMPD condition. Approximately 116 genes were differentially expressed between both conditions (F.C. > 0.5×). **j** KEGG GO terms upregulated in $Msi1^+$ keratinocytes. mTOR, HIF1, MAPK, RAS, and PI3K-AKT signaling pathways are selectively represented. Student's t-test was used for calculating P values in **c**.

Msi1 activates the Her2 signaling pathway by directly suppressing Cmtm5

Previous studies have reported that the oncoprotein HER2, the gene product of *ERBB2*, is ectopically expressed in human EMPD and MPD tissues.[8,41] We validated these findings in silico using scRNA-seq, where *ERBB2* expression is observed in Paget cells (Fig. 1l). *ERBB2* expression was also validated in whole EMPD skin tissues at mRNA and protein levels (Fig. 6a, b). We also found Her2

to be overexpressed in DTG mouse epidermis at both the RNA and protein levels (Fig. 6c). Her2 overexpression in the epidermis of DTG mice was primarily localized to the suprabasal layer (Fig. 6d). Consistently, 28 genes related to Her2 signaling were shared between DTG mouse and human EMPD samples on microarray analysis (Fig. 6e). To understand the molecular mechanisms underlying Her2 pathway activation, we analyzed the Msi1-binding sites in transcripts that function as inhibitors of the
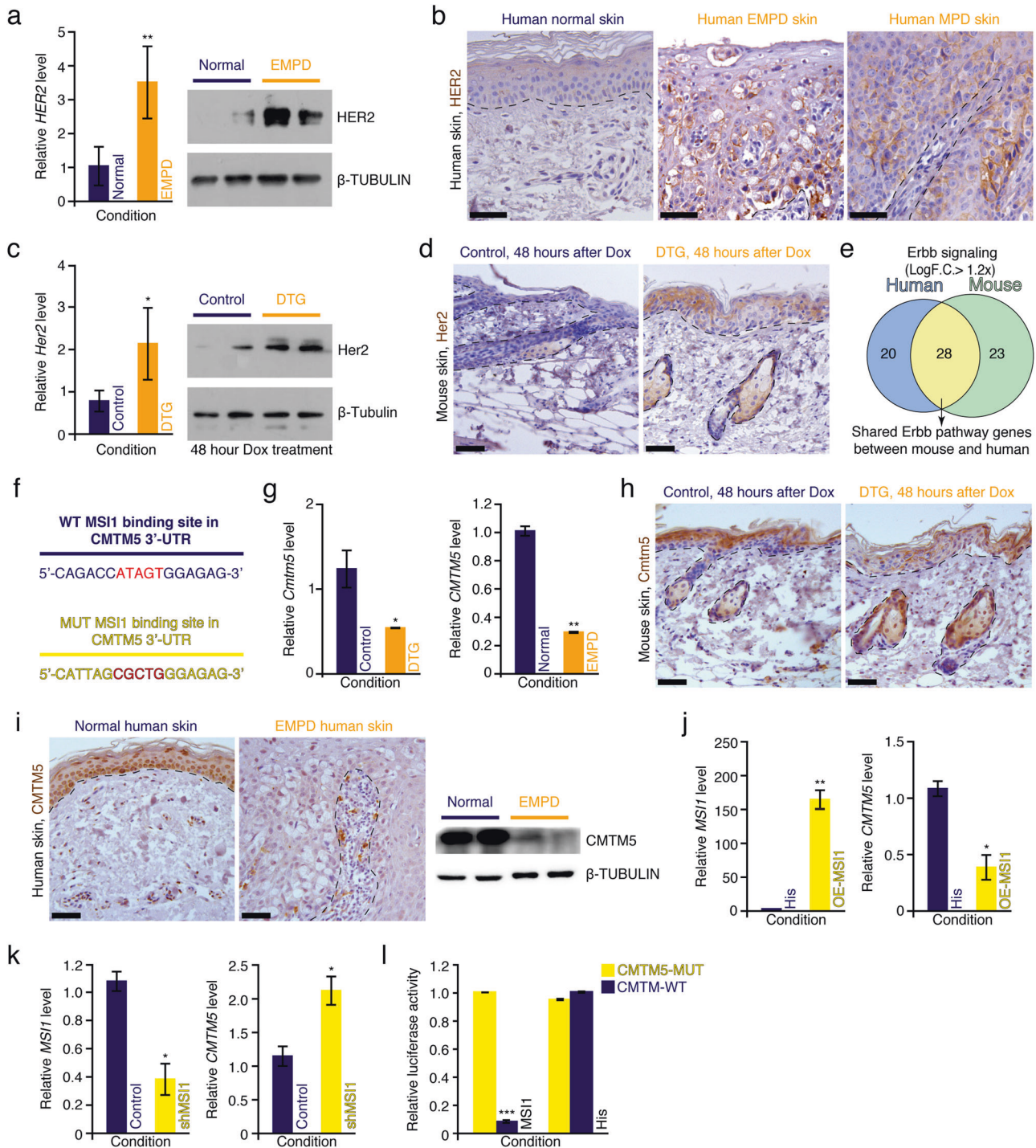
Fig. 6 Msi1 activates Her2 signaling pathway. a qRT-PCR and western blotting analyses of HER2 in normal ($n = 3$) and EMPD ($n = 3$) human skin. Relative mRNA expression values were normalized to GAPDH. β-TUBULIN was used as loading control. **$P < 0.01$. **b** Immunohistochemical staining against HER2 in normal, MPD and EMPD human skin. **c** qRT-PCR and western blotting analyses of Her2 in control ($n = 4$) and DTG ($n = 4$) mice 48 h after continuous Doxycycline treatment. β-Tubulin was used as loading control, which is identical to Supplementary information, Figs. S10j, S22a. *$P < 0.05$. **d** Immunohistochemical staining against Her2 in control ($n = 4$) and DTG ($n = 4$) mice 48 h after continuous Doxycycline treatment. **e** Venn diagram showing overlapping Her2-associated genes in EMPD human and DTG mouse skin. **f** Diagram of WT and mutant (MUT) Msi1 consensus binding sites in *Cmtm5* 3′-UTR. **g** qRT-PCR analysis of *Cmtm5* in control and DTG mice 48 h after Doxycycline treatment (left). qRT-PCR analysis of *CMTM5* in normal and EMPD human skin (right). *$P < 0.05$; **$P < 0.01$. **h** Immunohistochemical staining against Cmtm5 in control ($n = 4$) and DTG ($n = 4$) mice 48 h after continuous Doxycycline treatment. **i** Immunohistochemical and western blotting analyses of CMTM5 in normal ($n = 4$) and EMPD ($n = 4$) human skin. β-TUBULIN was used as loading control. **j** qRT-PCR analysis on *MSI1* and *CMTM5* in 293T cells transfected with PCDNA6.0-hMSI1. *$P < 0.05$; **$P < 0.01$. **k** qRT-PCR analysis of *MSI1* and *CMTM5* in 293T cells transfected with Msi1 shRNA. *$P < 0.05$. **l** Luciferase reporter activity of *CMTM5* WT and MUT 3′-UTR constructs in response to MSI1 binding. ***$P < 0.001$. Values in the graphs represent means ± SD. Student's *t*-test was used for calculating P values in **a**, **c**, **g**, **j**–**l**. Representative images are shown. Epidermis and dermis are demarcated with broken line. Scale bars, 50 μm (**b**, **d**, **h**, **i**).

Her2 signaling pathway. Interestingly, we found that Cmtm5, an inhibitor of Her2 signaling,[42] contains a putative Msi1-binding site in its 3′-UTR (Fig. 6f). Cmtm5 was downregulated both in the DTG mouse epidermis and human EMPD skin both at the mRNA and protein levels (Fig. 6g–i). In vitro, MSI1 overexpression inhibited CMTM5 expression in human HaCaT cells (Fig. 6j), while short hairpin RNA-mediated downregulation of MSI1 upregulated CMTM5 gene expression (Fig. 6k), suggesting that CMTM5 is a direct target of MSI1. Consistently, 3′-UTR luciferase reporter assays showed that MSI1 directly suppresses CMTM5 mRNA expression, and site-directed mutagenesis of MSI1 3′-UTR binding site in the luciferase reporter abrogated this suppression (Fig. 6l). Together, these findings demonstrate that Msi1 can activate the Her2 signaling pathway by directly suppressing Cmtm5. Considering that Her2 can also activate the PI3K-mTOR signaling pathway,[43] we conclude that Msi1-induced activation of mTOR is likely mediated through the Msi1-Cmtm5-Her2-mTOR signaling cascade.

Msi1-mediated activation of the Akt-mTOR signaling pathway contributes to Paget-like phenotype

Our scRNA-seq and bulk microarray profiling analyses revealed that the PI3K-Akt-mTOR signaling pathway was activated both in the Msi1-overexpressing mouse skin and human EMPD samples (Fig. 7a, b). Because our previous work showed that Msi1 contributes to oncogenic intestinal transformation through activation of the Akt-mTOR signaling pathway,[23] we asked whether the Msi1-mediated formation of Paget-like cells in the epidermis also relied on mTOR signaling. To this end, we first examined the level of p-Akt and the status of mTOR pathway activation using its downstream effectors pS6 and pS6K. All these three markers showed extensive activation of mTOR throughout the epidermal layers in DTG mice, while in the control littermates, they were restricted to the suprabasal granular layer (Fig. 7c; Supplementary information, Fig. S22a). Hif1α, a transcription factor that mediates hypoxia, has been identified as a downstream target of mTOR signaling in multiple cell lines.[44,45] Consistently and unlike control mice, the epidermis of DTG mice was Hif1α+ (Fig. 7c). In vitro assays using the HaCaT cell line showed that Msi1 overexpression activates mTOR signaling in a cell-autonomous manner (Supplementary information, Fig. S22b). To test whether mTOR signaling is associated with the pathogenesis of PD in patients, we examined mTOR activity in human EMPD and MPD tissues. Expression of mTOR marker pS6 was limited to the suprabasal epidermal layers in normal human skin, yet it was prominently ectopically activated in the basal epidermal layer of EMPD and MPD tissues (EMPD, $n = 12$; MPD, $n = 5$; Supplementary information, Fig. S22c–e). Thus, the pattern of mTOR activity in DTG mouse skin mimics that of human PD skin. We previously showed that, in the intestine, Msi1 activates mTOR signaling by directly repressing Pten.[23] We next wondered whether a similar mechanism operated in DTG mice. To test this, we examined the status of Pten in mouse skin and found that its protein levels were reduced upon Msi1 overexpression (Fig. 7d). To test whether MSI1 directly targets PTEN in human epidermis, we performed PTEN 3′-UTR luciferase assays in human keratinocytes in vitro. Overexpression of MSI1 significantly suppressed the activity of WT PTEN 3′-UTR reporter but had no effect on the reporter with mutant MSI1-binding site (Fig. 7e, f). These data suggest that MSI1 directly targets PTEN in keratinocytes. Furthermore, CLIP-PCR assays revealed that Pten mRNA was enriched in Msi1 antibody immunoprecipitants from mouse primary skin keratinocytes (Fig. 7g), further confirming that Msi1 physically binds to the 3′-UTR of Pten mRNA and suppresses its expression in skin.

To test whether Msi1-mediated mTORC1 activation functionally contributes to the Paget-like phenotype in DTG mice, we blocked mTORC1 signaling in DTG mice by intraperitoneal administration of Rapamycin for three consecutive days prior to Doxycycline induction. Phosphorylation of S6 was markedly reduced in the
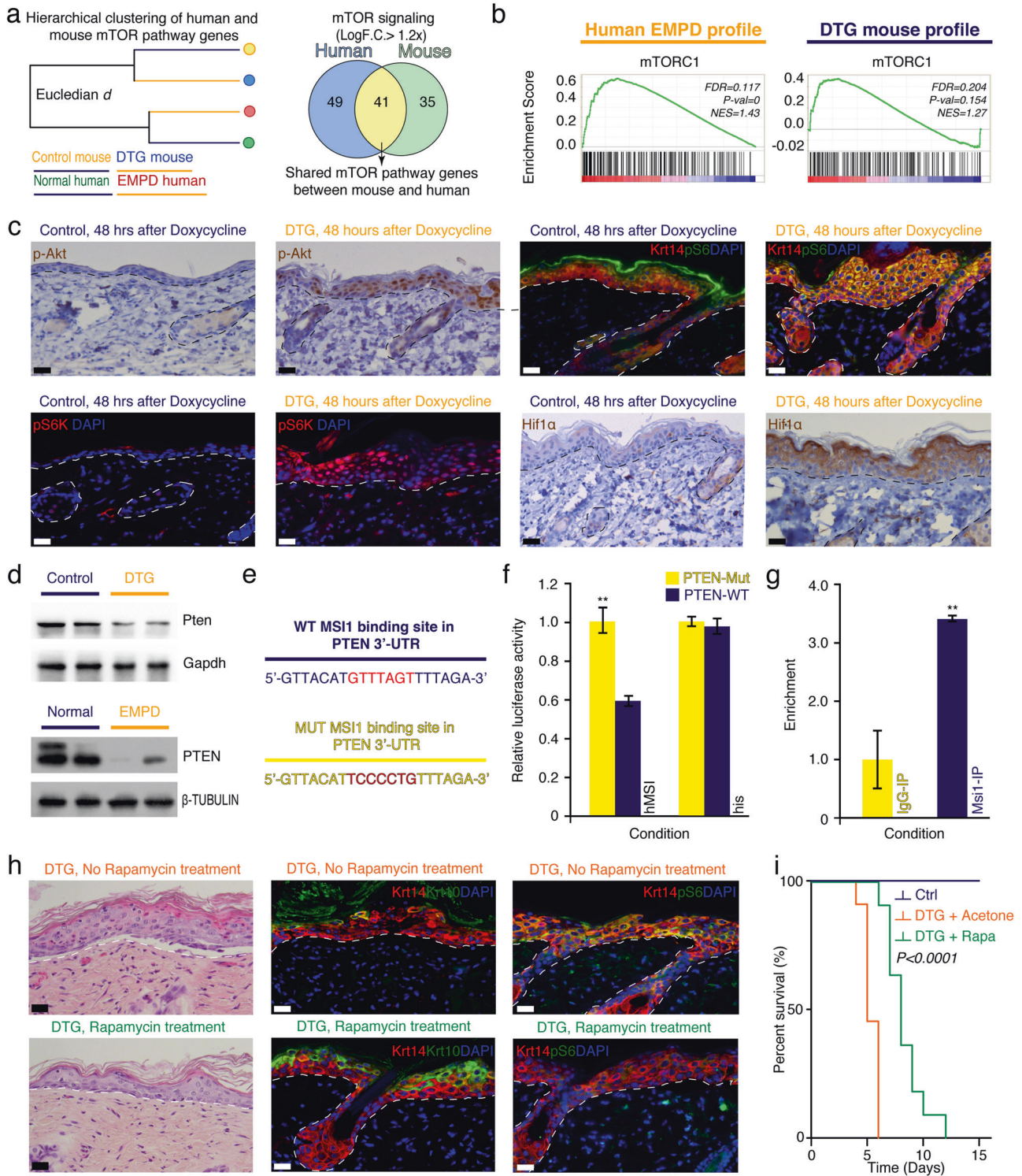
Rapamycin-treated DTG mice 48 h after continuous Doxycycline induction (Fig. 7h; Supplementary information, Fig. S23a), confirming suppression of mTORC1 activity. Strikingly, Rapamycin treatment markedly extended the survival rate of DTG mice (Fig. 7i). Histologically, we observed fewer Paget-like cells, a thinner and more organized epidermis, and normalized expression of Krt14 and Krt10 markers in Rapamycin-treated DTG mice, suggesting substantial amelioration of the Paget-like phenotype (Fig. 7h). Furthermore, Rapamycin treatment reduced the expression of several EMPD markers (Supplementary information, Fig. S23b, c). Taken together, our combined mouse and human data strongly indicate that Msi1-mediated mTORC1 activation is critical for the acquisition of epidermal PD phenotype and that blocking mTOR signaling alleviates clinical features associated with the disease in mice.

Topical Rapamycin alleviates PD-associated symptoms in a cohort of EMPD patients

Our data demonstrates the importance of the mTOR signaling pathway in the pathogenesis of PD. Because Rapamycin has been approved by the U.S. Food and Drug Administration (FDA) for the treatment of facial angiofibromas in tuberous sclerosis complex,[46] we tested its efficacy in alleviating PD-associated histopathological symptoms in a cohort of EMPD patients. We obtained consent from four patients who were pathologically diagnosed with EMPD. No any other treatment was undertaken by these patients before and during topical Rapamycin treatment. EMPD skin lesions were topically treated with 0.4% FDA-approved Rapamycin ointment two times per day for two consecutive weeks. All four patients achieved a clinically positive response to topical Rapamycin treatment (Fig. 8; Supplementary information, Figs. S24–S26). Among the patients was a 68-year-old male with scrotal EMPD (Fig. 8). This patient had a clinical history of EMPD recurrence after three surgical excisions between 2012 and 2015. The skin lesion recurred 3 years after the last excision, and physical examination showed an erythematous skin lesion around the scrotum in 2018 (Fig. 8a). Histological analysis showed that the skin lesion contained numerous Paget cells staining positive for PAS, CEA and Cam5.2 (Fig. 8b–f). Strikingly, after two weeks of 0.4% FDA-approved topical Rapamycin ointment treatment, mTOR signaling activity was reduced, and only occasional PAS+, CEA+ and Cam5.2+ Paget cells could be identified in situ (Fig. 8b–f). Reflectance confocal microscopy showed that prior to topical Rapamycin treatment, the epidermis of the lesioned skin was disorganized. However, after topical Rapamycin treatment, the morphology of the skin epidermis returned to normal (Fig. 8g). Ultrasound imaging further showed that the skin thickness was markedly reduced after topical Rapamycin treatment (Fig. 8h). Similar therapeutic effects of topical Rapamycin treatment were observed in the other three EMPD cases analyzed (Fig. 8i; Supplementary information, Figs. S24–S26). In summary, our pilot clinical study demonstrated the efficacy of topical Rapamycin treatment in EMPD patients.

DISCUSSION

EMPD and MPD are intra-epidermal malignant adenocarcinomas characterized by distinct gland-like Paget cells in the epidermis.[1,2] While both EMPD and MPD are readily diagnosed clinically and histologically, the pathogenesis of PD is poorly understood. In this work, scRNA-seq analysis demonstrated a large degree of heterogeneity of epithelial cells in EMPD epidermis that differ in their relative gene expression program and spatial distribution in healthy and EMPD adult scrotal epidermis. A feature of specific epithelial subpopulations in EMPD skin is an elevated aggregate gene scores for inflammation, accompanied with an increase in number and types of immune cells. In agreement with our findings, it has been reported that exhausted CD8+ T cells are

a Hierarchical clustering of human and mouse mTOR pathway genes

Eucledian *d*

Control mouse   DTG mouse
Normal human   EMPD human

mTOR signaling
(LogF.C.> 1.2x)
Human   Mouse

49   41   35

Shared mTOR pathway genes
between mouse and human

b

Human EMPD profile   DTG mouse profile

mTORC1   mTORC1

FDR=0.117   FDR=0.204
P-val=0   P-val=0.154
NES=1.43   NES=1.27

c

Control, 48 hrs after Doxycycline — p-Akt
DTG, 48 hours after Doxycycline — p-Akt
Control, 48 hrs after Doxycycline — Krt14 pS6 DAPI
DTG, 48 hours after Doxycycline — Krt14 pS6 DAPI

Control, 48 hrs after Doxycycline — pS6K DAPI
DTG, 48 hours after Doxycycline — pS6K DAPI
Control, 48 hrs after Doxycycline — Hif1α
DTG, 48 hours after Doxycycline — Hif1α

d

Control   DTG
Pten
Gapdh

Normal   EMPD
PTEN
β-TUBULIN

e

WT MSI1 binding site in
PTEN 3'-UTR

5'-GTTACATGTTTAGTTTTAGA-3'

MUT MSI1 binding site in
PTEN 3'-UTR

5'-GTTACATTCCCCTGTTTAGA-3'

f

PTEN-Mut
PTEN-WT

Relative luciferase activity
Condition
hMSI   his

g

Enrichment
Condition
IgG-IP   Msi1-IP

h

DTG, No Rapamycin treatment
DTG, No Rapamycin treatment — Krt14 Krt10 DAPI
DTG, No Rapamycin treatment — Krt14 pS6 DAPI

DTG, Rapamycin treatment
DTG, Rapamycin treatment — Krt14 Krt10 DAPI
DTG, Rapamycin treatment — Krt14 pS6 DAPI

i

Ctrl
DTG + Acetone
DTG + Rapa

P<0.0001

Percent survival (%)
Time (Days)

dramatically increased in EMPD skin[35] and that inflammation-related signaling pathways are activated in EMPD skin, including STAT3 and NFκB signaling pathways.[47,48] Thus, this suggests that EMPD is an inflammation-related disease. PD is considered as a rare and malignant intra-epidermal adenocarcinoma.[49] Supporting this idea, scRNA-seq analysis demonstrated that a subset of Paget cells are characterized by a rapidly cycling state. In line with this observation, KRT6C+ keratinocytes are specific for human EMPD skin. Indeed, KRT6C overexpression is associated with poor

diagnosis in cancer.[28] Paget cells are easily identified by their characteristic morphology — which have a clear cytoplasm and a peripherally positioned, hyperchromic nucleus, or by expression of an array of known biomarker genes.[5,7] Despite this, scRNA-seq allowed us to identify a number of genes uniquely expressed in Paget cells. On this note, we report the addition of ALCAM as a validated, novel biomarker for Paget cells. Indeed, ALCAM has been previously used to physically sort actively cycling Lgr5+ intestinal stem cells or cardiac progenitor.[50,51] Therefore, we posit

Fig. 7 Msi1 activates mTOR signaling pathway. a Agglomerative hierarchical clustering of mTOR pathway in mouse and human orthologs and Venn diagram showing overlapping mTOR-associated genes in human EMPD and DTG mouse skin. **b** GSEA profile in 4 control and 4 DTG mouse samples, as well as 3 normal human skin and 4 EMPD human skin samples showing common gene enrichment set for mTOR signaling pathway. **c** Immunostaining against phosphorylated Akt (p-Akt), phosphorylated-S6 Ribosomal protein (pS6), phosphorylated-70 S6 (pS6k), and Hif1α in dorsal skin of control ($n = 3$) and DTG ($n = 3$) mice 48 h after continuous Doxycycline treatment. pS6 were co-stained with Krt14. **d** Western blotting analysis of Pten in control and DTG mouse skin 48 h after continuous Doxycycline treatment. Gapdh was used as loading control. Western blotting against PTEN in normal and EMPD human skin tissues. β-TUBULIN was used as loading control, which is identical to Supplementary information, Figs. S8a, S11j and S22e. **e** Diagram of WT and mutant (MUT) Msi1 consensus binding sites in *Pten* 3′-UTR. **f** Luciferase reporter activity of Pten WT and MUT 3′-UTR constructs in response to Msi1 binding. **$P < 0.01$. **g** CLIP-qPCR assay for Pten in Msi1- and IgG-immunoprecipitates from mouse primary keratinocytes. IgG represents negative control. **$P < 0.01$. **h** Histological assessment of DTG mouse dorsal skin with and without Rapamycin treatment 48 h after continuous Doxycycline treatment. Double staining of Krt10 and pS6 with Krt14 is shown. **i** Kaplan-Meir plots representing percent survival rates of control ($n = 12$), DTG ($n = 12$), and DTG treated with Rapamycin ($n = 11$) mice after continuous Doxycycline treatment. $P < 0.0001$. Values in the graphs represent means ± SD. Student's *t*-test was used for calculating P values in **f**, **g**. Representative images are shown. Epidermis and dermis are demarcated with broken line. Scale bars, 25 μm (**c**, **h**).

that it will now be feasible to physically sort Paget cells from human EMPD skin for downstream genomic and epigenetic interrogation.

Progress in understanding the pathogenesis of PD is hampered by the lack of mouse models of the disease. Here we demonstrated that a transgenic mouse model with epidermis-specific Msi1 overexpression partially recapitulates human PD phenotype, mimicking multiple histopathological, single-cell and molecular aspects of this human disease. On histopathology, Paget-like cells begin to appear in the DTG mouse epidermis as early as 36 h after Msi1 overexpression. Furthermore, ectopic expression of multiple established EMPD and MPD markers, including PAS, Mucin1, Cam5.2, Krt7, Krt8, Gata3, Ceacam1, Podoplanin, Her2 and Ar,[1,6–8,11,12,52–54] becomes prominently activated in DTG epidermis upon Msi1 overexpression. At single-cell level, a cluster of Paget-like cells was identified in DTG epidermis, resembling human EMPD Paget cells. Moreover, many of the abnormally activated signaling pathways, likely mediators of the disease, were shared in both the transgenic mouse epidermis and human EMPD skin samples. Thus, we conclude that Msi1-overexpressing transgenic mice could serve as a mouse model for human PD. In human, MPD specifically occurs at nipple site of mammary gland, and EMPD is specifically located at the vulvar, perianal, scrotal, penile and axillary regions. In line with this, we found that the ventral skin of DTG mouse, which is more similar to human skin in the specific regions, gave rise to more severe Paget-like phenotypes compared to back skin. We also noticed that Msi1 overexpression results in additional mouse phenotypes, including epidermal-dermal detachment and subsequent animal death, which are not characteristics of human PDs. A likely explanation for these observed phenotypes in DTG mice could be that sustained overexpression of Msi1 leads to downregulation of cell adhesion proteins in basal keratinocytes, thereby affecting skin barrier and accounting for skin barrier-associated death. Indeed, the Msi1 orthologue — MSI2 has been shown to downregulate tight-junction-associated Claudins to increase tumor metastasis in non-small cell lung cancers.[55] The regulation of Msi1 on cell adhesion and skin barrier merits further investigation. Despite this limitation, this mouse model can provide a definitive and quantifiable phenotype in studies designed to test the effectiveness of anti-PD candidate drugs.

While the histopathological features of PD are well characterized,[1] the molecular mechanism of the disease remains largely unknown. Our findings show that Msi1 activates the mTOR pathway in human PD and that Rapamycin rescues the Msi1-induced Paget-like phenotype in DTG mice and effectively alleviates disease symptoms in EMPD patients, indicating the importance of mTOR signaling in the pathogenesis of PD. In agreement with these findings, one study reported that the PI3K-AKT-mTOR pathway is activated in EMPD tissues[56] and that the expression levels of mTOR and other pathways' members are correlated with the invasiveness of EMPD.[56,57] Our previous study showing that Msi1 overexpression

induces intestinal transformation by activating mTOR signaling[23] suggests that the Msi1-Pten-mTOR signaling cascade is important for tumor initiation in cancers originating from both the ectoderm and the endoderm. Here, we also found that Msi1 can activate the mTOR pathway via promoting Her2 activity. Indeed, Her2 oncoprotein, known for its ability to activate PI3K-mTOR signaling,[43] is frequently overexpressed in EMPD and MPD tissues.[8,41] We conclude that the abnormal activation of mTOR signaling in PD can be achieved through the Msi1-Pten-mTOR and/or Msi1-Her2-mTOR pathways (Supplementary information, Fig. S27). The PI3K-AKT-mTOR pathway is one of the most frequently dysregulated pathways in cancer.[58] Temsirolimus and Everolimus, two mTOR inhibitors derived from Rapamycin, have been approved by the FDA for the treatment of numerous cancer types, including advanced-stage renal cell carcinoma[59] and pancreatic neuroendocrine tumors.[60] Specific to skin, topical Rapamycin has been extensively used to treat benign facial angiofibromas associated with tuberous sclerosis complex.[46] In this work we show, for the first time, that topical Rapamycin treatment effectively alleviates EMPD symptoms in a cohort of Chinese human patient volunteers. This finding is significant, as no drugs were previously known to be effective for treating PD. Our pilot clinical data warrants further, large-scale, clinical trials to validate the efficacy of mTOR inhibitors in PD. In addition, it merits further investigation on the upstream regulation of MSI1 activation. Identifying driver mutations using exome sequencing would be therefore critical for understanding the pathogenesis of PD.

Our findings that Msi1 overexpression in the basal epidermal cells of DTG mice induces Paget-like cells, combined with subsequent pseudotime, RNA velocity, and lineage tracing studies, reveal a basal keratinocyte-Paget cell conversion in PD, supporting the in situ transformation theory. To support this idea, clinical observations have shown that most EMPD lesions and some MPD lesions are initially limited to the epidermis and do not appear to be associated with distant adenocarcinomas.[61,62] Moreover, genetic alterations in Paget cells, including the loss of heterozygosity and mitochondrial DNA displacement loop sequences, are distinct from those observed in the underlying breast cancer cells.[63] In contrast, the Epidermotropic theory, postulating that Paget cells originate from ductal carcinoma cells that have migrated into the overlying epidermis, is primarily supported by the common histopathological features and HER2 reactivity between PD and the underlying adenocarcinoma,[16] as well as the clinical observation that most cases of MPD have an underlying HER2+ breast carcinoma.[15] It is worthwhile to note that MSI1 is also upregulated in certain breast cancers[64] and in colorectal cancers[23,65]; therefore, the morphological similarity between Paget cells and bona fide adenocarcinoma cells could be the result of the shared Msi1-activated signaling in both cell types. Therefore, although we cannot formally rule out the Epidermotropic model, our findings provide compelling evidence and lend support to the in situ transformation model.
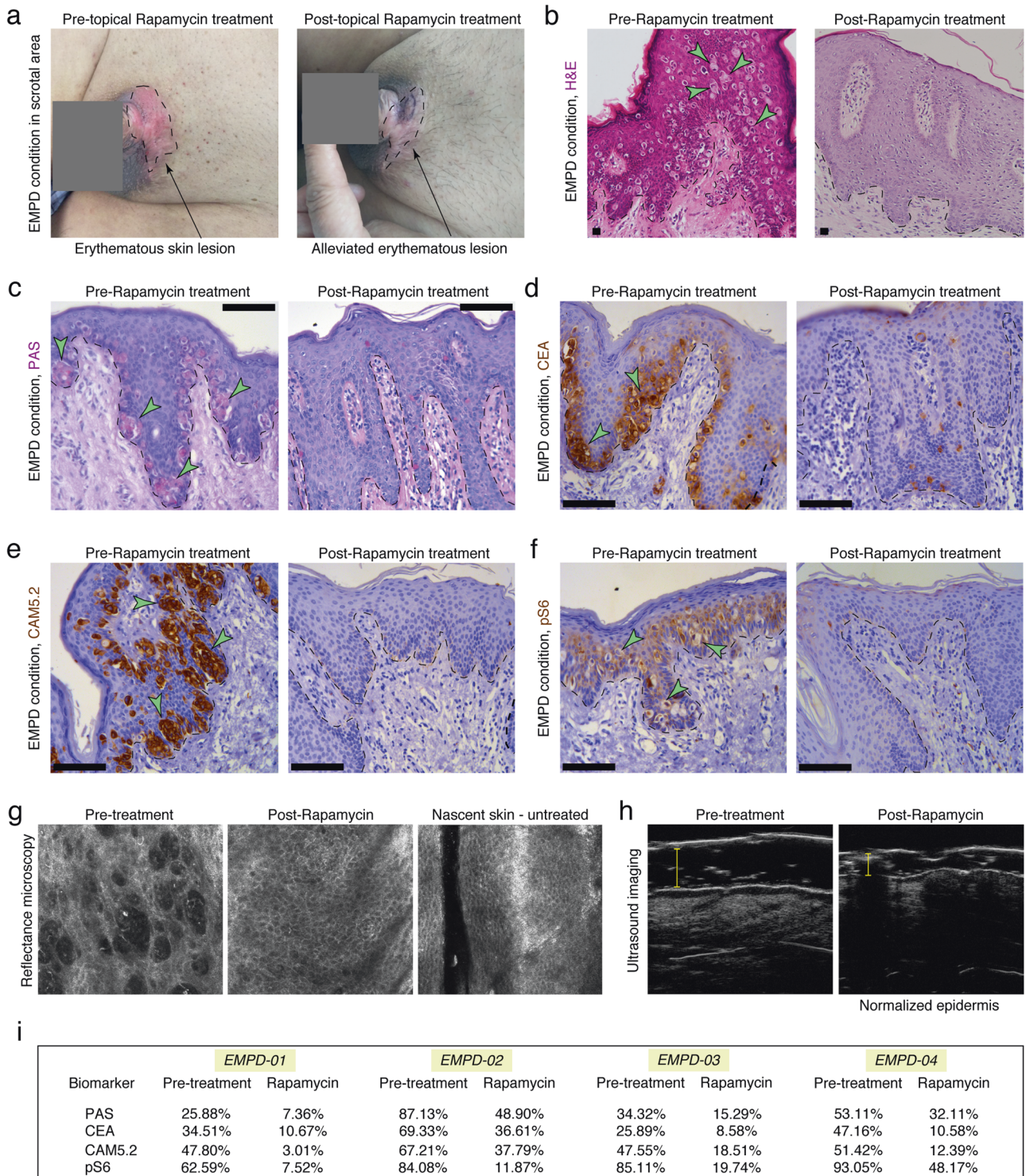
Fig. 8   Topical Rapamycin treatment effectively alleviates PD-associated pathological symptoms in a cohort of human EMPD patients.
a Gross image of the erythematous lesion of skin around the genital region of a 68-year-old patient (Patient ID: EMPD01) with scrotal EMPD first diagnosed in 2015 before and after two weeks of topical Rapamycin treatment. Black arrows point to the eczematous-like lesion. **b** Histopathology of EMPD skin from patient EMPD01 before and after two weeks of topical Rapamycin treatment. **c**–**f** PAS and immunohistochemical staining against CEA, CAM5.2 and pS6 in EMPD skin from patient EMPD01 before and after two weeks of topical Rapamycin treatment. **g** Reflectance confocal microscopy of lesioned skin from patient EMPD01 before and after two weeks of topical Rapamycin treatment. **h** Ultrasound images of lesioned skin from patient EMPD01 before and after two weeks of topical Rapamycin treatment. **i** Quantification of the percentage of CEA-, CAM5.2-, PAS- and pS6-positive cells in the skin lesions of a cohort of EMPD patients before and after two weeks of topical Rapamycin treatment. Arrows point to Paget cells. Representative images are shown. Epidermis and dermis are demarcated with broken line. Scale bars, 100 μm (**c**–**f**).

MATERIALS AND METHODS

Ethics statements

Human clinical studies were approved by the Ethics Committee of Shanghai Skin Disease Hospital (Shanghai, China). We certify that all applicable institutional regulations concerning the ethical use of information and samples from human volunteers were strictly followed in this work. Each subject provided written informed consent. This study is registered at the Chinese Clinical Trial Registry http://www.chictr.org.cn (ID: ChiCTR1800014979). All animal procedures were authorized by the Beijing Laboratory Animal Management committee. All animal studies were performed in strict adherence to the Institutional Animal Care and Use Committee (IACUC) guidelines of China Agricultural University (Approval number: SKLAB-2011-04-03).

Human samples

The use of human samples was approved by China Agricultural University (Beijing, China), Shanghai Skin Disease Hospital (Shanghai, China), and Tianjin Medical University General Hospital (Tianjin, China). EMPD and MPD lesions were clinically diagnosed by histopathological analysis of lesion at Shanghai Skin Disease Hospital by board certified Pathologist/Dermatologist. All EMPD and MPD samples and adjacent normal skin tissues were obtained with written informed consent from patient volunteers (See Ethics Statement). Human sample information is summarized in Supplementary information, Tables S1–S3.

Fluorescence-activated cell sorting

Human. Single cell suspensions of human normal and EMPD epidermal cells were obtained and stained with FVD450 (eBioscience,1:1000) for 20 min. Viable cells were physically sorted with a BD LSRFortessa fluorescence-activated cell sorter (BD Biosciences). Viability rate for normal and EMPD epidermal cells was 74.29% and 87.21%, respectively. **Mouse**. Single cell suspensions of mouse control and DTG epidermal cells were obtained and stained with FVD450 (eBioscience,1:1000) for 20 min. Viable cells were physically sorted with a BD LSRFortessa fluorescence-activated cell sorter (BD Biosciences). Viability rate for control and DTG epidermal cells was 93.0% and 92.3%, respectively.

Isolation of primary cells for scRNA-seq

Human. Extra-Mammary Paget's Disease (EMPD) patient was clinically diagnosed by histopathological analysis of lesion at Shanghai Skin Disease Hospital. Human EMPD and normal skin (site adjacent to EMPD lesion) samples were surgically excised and placed in tissue storage solution (MACS). Samples were processed for single cell isolation within 10 h post-surgery. Excess subcutaneous adipose and dermal tissues were excised. Samples were cut into small pieces with a maximum diameter of 4 mm. Skin specimens were then washed twice with cold 1× PBS. Single cells were isolated using the human epidermis isolation kit (MACS) as per the manufacturer's instructions with minor modifications. **Mouse**. Skin epidermal cells were collected from DTG ($n = 5$) and control ($n = 4$) newborn pups (post-natal day 0.5) from the same litter 5 days after in utero Doxycycline treatment. Mice were euthanized, and consecutively immersed in 70% ethanol and cold 1× PBS twice for 3 min. Fore and hind limbs and tail were cut off and skin was peeled along the ventral midline axis with blunt tweezers. Subcutaneous adipose tissue was micro-dissected. 12 mm skin biopsies were laid on EpiLife CF basal media (STEM CELL) droplets (300 μL) on a 12-well plate with skin dermis side down. 2 mL (2.5 U/mL) Dispase solution (STEM CELL) was added to skin and incubated at 4 °C for 12 h. Epidermis was separated from dermis with two sharp tweezers and placed on Accutase (ThermoFisher) droplets (500 μL) for 30 min (dermis side down) at RT to create single cell suspensions. Dead single cells were stained with FVD450 (EBioscience). Live cells were physically

sorted with a BD LSRFortessa fluorescence-activated cell sorter (BD Biosciences).

3′-end scRNA-seq

Single cell capture, library preparation and sequencing were performed by Beijing CapitalBio Technology (Beijing, China). Single cell capture was performed using the Chromium system (10× Genomics) with single Cell 3′ Chemistry V3. cDNA libraries were sequenced as paired-end with Illumina HiSep 2500 (ST-E00205).

Processing of 3′-end scRNA-seq data

The web summary of Cell Ranger statistics on sequenced cells, sequencing and mapping metrics pre-quality control assessment for downstream bioinformatic analyses were as follows: **Human**. Human transcripts were mapped to the human reference transcriptome (GRCh38) using Cell Ranger Version 3.0.2. *Normal skin*: Sequencing metrics: ~322,137,849 total number of reads, ~97.3% valid barcodes; Mapping metrics: ~66.9% reads mapped confidently to transcriptome; Cell metrics: 9031 estimated number of cells, ~79.0% fraction reads in cells, ~35,670 mean reads per cell, ~2601 median genes per cell, ~22,081 total genes detected, ~9789 median UMI counts per cell. *EMPD skin*: Sequencing metrics: 332.368.998 total number of reads, ~97.6% valid barcodes; Mapping metrics: ~63.0% reads mapped confidently to transcriptome; Cell metrics: 14,480 estimated number of cells, ~81.7% fraction reads in cells, ~22,953 mean reads per cell, ~976 median genes per cell, ~24,277 total genes detected, ~2490 median UMI counts per cell. **Mouse**: Murine transcripts were mapped to the mouse reference transcriptome (mm10) using Cell Ranger Version 2.0.1. *Control mouse*: Sequencing metrics: ~456,436,577 total number of reads, ~97.7% valid barcodes; Mapping metrics: ~74.7% reads mapped confidently to transcriptome; Cell metrics: 11,145 estimated number of cells, ~62.1% fraction reads in cells, ~40,954 mean reads per cell, ~2747 median genes per cell, ~19,151 total genes detected, ~8,886 median UMI counts per cell. *DTG mouse*: Sequencing metrics: ~473,602,702 total number of reads, ~97.7% valid barcodes; Mapping metrics: ~74.4% reads mapped confidently to transcriptome; Cell metrics: 8,589 estimated number of cells, ~71.3% fraction reads in cells, ~55,140 mean reads per cell, ~3248 median genes per cell, ~19,299 total genes detected, ~11,445 median UMI counts per cell. scRNA-seq data has been submitted to GEO repository through the access code: GSE140956.

Quality control analyses

Quality control filtering was performed to discard low quality cells.[26] For species-specific integrated analyses, the same quality control metrics were performed to ensure the same cells were used for downstream analyses. Cells were kept only if they met the following collective quality control criteria: > 350 genes/cell, < 5000 genes/cell; and < 10% mitochondrial gene expression. Outliers were removed by implementing a quadratic model ($P = 1.03e^{-3}$) as in the PAGODA2 pipeline.[66] Low quality cells and outliers were discarded, and only "valid" cells were used for downstream analyses.

Downstream analysis and data integration using Anchoring

We performed data integration of mouse control/DTG and human normal/EMPD data sets using the Seurat package (Version 3.0.0, R Studio Version 3.5.1).[27] In brief, individual digital gene expression matrices were column-normalized using the *NormalizeData* function. Variable features were identified using the *FindVariableFeatures* function with the *selection.method = "mean.var.plot"* *mean.cutoff = c(0.0125, 3), dispersion.cutoff = c(0.5, Inf))* functions enabled. Integration anchors between data sets were identified using the *FindIntegrationAnchors* function with dimensions set to 50. Data sets were integrated using the *IntegrateData* function

with dimensions set to 50. Integrated data was scaled using the *ScaleData* function. Principal component analysis (PCA) was first performed on the list of highly variable features identified using the *FindVariableFeatures* and significant PCs used for clustering and finding neighbors were identified using a combination of statistical (PC heatmaps and Jackstraw method — 100 permutations) and heuristic methods (elbow plot). Clusters and neighbors were identified using the *FindNeighbors* and *FindClusters* functions with dimensions specified by user and visualized using Uniform Manifold Approximation and Projection (UMAP). The following PCs were used for clustering of data sets; human epithelial cells — 1:8, 15:25, 30:35, 40:45, 50:55; human immune cells — 1:9, 12:17, 20, 25, 30:32, 35:37, 40; mouse epithelial cells — 1:4, 5:8, 9:12, 13, 15, 16, 19, 22, 23, 25, 26, 28, 33, 34. For gene marker identification (cluster biomarkers), a minimum required average Log fold change in expression (2×) and a minimum percent of cells that must express a gene marker in either one cluster (0.25 or 25%) were set. Gene markers were identified using the *FindAllMarkers* function with the Wilcoxon Rank Sum Test (Wilcox).

Cell cycle discrimination analyses
We used cell cycle-related genes, previously defined as a core set of 43 G1/S and 54 G2/M genes,[30] to identify cycling cells in our data set using the *Cell Cycle Scoring* function in Seurat. We projected scored cells onto UMAP space and colored cells based on the Seurat-defined cell cycle clusters. Cells present in each cell cycle state were also quantified.

Agglomerative hierarchical clustering
Agglomerative hierarchical clustering based on Euclidian distance with Ward's linkage was performed using the MATLAB function *plot_Cladogram* with gene-cell counts and cluster-specific labels as input.

High-density plots
High-density plots (i.e., tight subplots) were generated using the MATLAB function *tight_subplot* with gene-cell counts and cluster-specific labels as input and visualized as before.[67]

Gene module scoring
Aggregate biomarker module scores were assigned to Paget cells; inflamed epithelial cells; and cytotoxic, exhaustive, and naïve T cells using the *AddModuleScore* function in Seurat with the following conditions enabled: $ctrl = 5$. Paget cells were identified by implementation of a 'Paget cell aggregate biomarker score' defined by a core set of known PD-associated biomarkers. These PD biomarkers included: *KRT7, KRT8, CAM5.2, CEA* — *CEACAM5, Mucin1* — *MUC1, Her2* — *ERBB2*, cyclin D1 — *CCDN1*, p21 — *CDKN1A, AR*, EMA — *MUC1*, GCDFP-15 — *PIP, GATA3*, and *P16* — *CDKN2A*. Inflammation scores were assigned to epithelial cells by implementation of scores defined by a previously-identified core set of inflammation-associated biomarkers.[68] These biomarkers included the following: Inflammation — *CXCL1, CXCL8, CCL20, RARRES2, S100S, IL36G, IL1B, TNF, IL17A, IL18, IL6, CXCL10, IL23A, CCL2, CCL5, CCL13, IL22, S100A7A15, RAGE*; Scalp — *S100A8, S100A9, S100A7, CSTB, SPRR1B, CALML3, HSPA1A, IFI27, S100A6, S100A2, S100A1*; Psoriasis — *SERPINB4, PI3, SPRR2D, SPRR2A, SPRR2E, SERPINB3, SPRR1B, IFI6, FABP5, CRABP2, CYCS, LY6D, LYZ*. Cytotoxic, exhaustive, and naïve T cells were scored by implementation of aggregate scores defined by previously identified core set of cytotoxic-, exhaustive- and naïve-associated genes in melanoma tissues.[34]

Mice
TRE-Msi1 mice have been previously described.[23] *Krt14-rtTA* (JAX stock: 008099), *Krt14-CreERT2* (JAX stock: 007678), and *mTmG* (JAX stock: 007676) mice were purchased from Jackson Laboratory. *Krt14-rtTA;TRE-Msi1* double transgenic (DTG) and *Krt14-rtTA;TRE-Msi1;Krt14-CreER;mTmG* quadrupled transgenic (QTG) mice were

genotyped by PCR. Briefly, genomic DNA was isolated from tail tissue and PCR amplified using the following primers: *TRE-Msi1-*Col/frtB: CCCTCCATGTGTGACCAAGG, *TRE-Msi1-*Col/frtA1: GCACAG CATTGCGGACATGC; *TRE-Msi1-*Col/frtC1: GCAGAAGCGCGGCCGT CTGG; *K14-rtTA-F*: CACGATACACCTGACTAGCTGGGTG; *K14-rtTA-R*: CATCAC CCACAGGCTAGCGCCAACT; *Krt14-CreERT1*: GCGGTCTGGC AGTAAAAACTATC; *Krt14-CreERT2*: GTGAAACAGCATTGCTGTCACTT; *Krt14-CreERT3*: CTAGGCCACAGAATTGAAAGATCT; *Krt14-CreERT4*: GTAGGTGGAAATTCTAGCATCATCC; *mTmG1*: CTCTGCTGCCTCCT GGCTTCT; *mTmG2*: CGAGGCGGATCACAAGCAATA; *mTmG3*: TCAAT GGGCGGGGGTCGTT.

Pseudotime analyses
Pseudo-ordering of individual cells was performed using Monocle2 (Version 2.10.1).[69–71] Briefly, interfollicular epidermal cells (*Msi1*[+] basal and Paget-like cells) were subclustered in Seurat using the *SubsetData* function. A *cellDataSet* object was created in Monocle2 with the function *newCellDataSet* and the following arguments enabled: *lowerDetectionLimit = 1.0, expressionFamily = negbinomial.size()*. Subclustered cells were ordered based on variable features identified using the *VariableFeatures* function. Dimensionality reduction was performed using the *reduceDimensions* function with the *reduction method = 'DDRTree'* argument enabled. Cells were then ordered using the *orderCells* function. For correct ordering of pseudotime-dependent genes, biased ordering of cells was performed by enabling the *root_state* function in Monocle2. To identify pseudotime-dependent gene expression changes, we applied the single-cell Energy path (scEpath; Version 1; MATLAB Version 9.5)[72] algorithm on Monocle2-ordered cells. To identify statistically significant pseudotime-dependent gene changes, we compared the standard deviation of the observed smoothed expressions with a set of similarly permuted expressions by randomly permuting the cell order (*nboot = 100* permutations). We considered all genes with a standard deviation greater than 0.01 and a Bonferroni-corrected *P*-value below a significance level $\alpha = 0.05$ to be pseudotime-dependent. To analyze pseudotime-dependent mouse transcription factors, we used genes annotated in the Animal Transcription Factor Database (AnimalTFDB 2.0)[73] by enabling the *TF_Ifo.mouse.Symbol* function for mouse-specific transcription factors in scEpath. Pseudotime-dependent genes were represented and visualized using a rolling wave plot with user-defined optimal K clusters.

RNA velocity
RNA velocity vectors were calculated on the basis of spliced and unspliced transcript reads as previously reported.[74] Briefly, annotation of spliced and unspliced reads and generation of Loom files was performed using the Python script *velocyto.py* on the Cell Ranger outs folder for each individual condition. Anaconda (Version 3.6–5.0.1) and samtools (Version 1.9) were implemented in *velocyto.py*. Mouse (mm10) *rmsk.gtf* files were downloaded from UCSC Genome Browser and used in *velocyto.py*. The R version of velocyto (velocyto.R) was used for velocity estimations and vector projections (Version 0.6). Data and genome annotations were loaded, and count matrices were read from the Loom files. Loom files were combined using *loompy*. Only cells part of the pseudotime trajectories were used for calculation of RNA velocities. PCA, kNN, and n sight parameters were user defined. PCA analyses was performed using PAGODA2 (Version 0.0.0.9002) based on the top 40 PCs.[66] RNA velocity was estimated using a gene-relative model with kNN cell pooling and used $n = 2000$ neighbors in mouse projections. RNA velocity vector fields were projected onto pseudotime trajectory calculated by Monocle2.

Lineage tracing
Krt14-rtTA;TRE-Msi1;Krt14-CreER;mTmG quadrupled transgenic (QTG) mice were used for lineage tracing studies. *Krt14-rtTA;Krt14-CreER; mTmG* triple transgenic (TTG) mice were used as control. Briefly,

QTG and TTG mice were induced with Tamoxifen for 5 consecutive days at a concentration of 20 mg/mL to label all Krt14$^+$ basal keratinocytes. Paget phenotype was then induced with continuous Doxycycline administration for 48 h. Mice were then sacrificed, and lineage of Paget-like cells was evaluated in dorsal and ventral skin tissues. Sections were counterstained with DAPI mounting media (Invitrogen). Sections were imaged with a Leica DM5500 B fluorescence microscope (Leica, Germany).

Microarray analysis
Total RNA was isolated from whole dorsal skin of littermate control ($n = 4$) and DTG ($n = 4$) mice using TRIzol reagent (LifeTechnologies) as per the manufacturer's instructions. Mice had been induced with Doxycycline for 24 h. Total RNA was isolated from normal human skin ($n = 3$) and EMPD ($n = 4$) skin samples. Genomic DNA was digested using a RNAse-free DNAse kit (Qiagen). Purified RNA was labeled and hybridized to Agilent Mouse/Human Gene arrays (CapitalBio Technology). Microarray data was summarized, normalized and quality control-checked using the GeneSpring software V13 (Agilent). The q value, a measure of the false discovery rate (FDR), was computed using the Significance Analysis of Microarrays function from Cluster 3.0 (Berkeley Lab) for each probe set and significance was attributed based on unpaired t-test. Microarray data was Log$_2$-transformed, and the median was centered by genes, clustered using hierarchical clustering with average linkage using the *Adjust Data* function of Cluster 3.0 software. Tree visualization was performed using Java TreeView (Stanford University School of Medicine, Palo Alto, CA, USA) as an intensity heatmap. Microarray data has been submitted to GEO repository and is available at GSE117284 (mouse) and GSE117285 (human).

GSEA
GSEA was performed using the Java application from the Broad Institute.[75] GSEA was used to identify significant patterns in the gene sets of interest.

Gene class and GO analysis
Manual curation of Gene Classes was performed using the Protein Analysis Through Evolutionary Relationships (PANTHER) on-line database.[32] Manual curation of GOs was performed using KEGG or Enrichr and visualized in the form of heatmap using $-\text{Log}_2(P \text{ value})$.[33,76]

Comparative analysis across conditions
Data integration was performed as before (see Downstream analysis and data integration using Anchoring). After integration, comparative analysis was performed between specific clusters, including: a) control and Msi1-overexpressing keratinocytes; and b) normal and EMPD keratinocytes. To identify DEGs in basal keratinocytes across conditions, we plotted the average expression of both the control and Msi1-overexpression/EMPD conditions and looked for outlier genes in scatterplot. DEGs were used as input in KEGG or Enrichr to identify significant pathways.[33,76] To identify unique gene markers in Paget-like cells in DTG conditions, we visualized gene expression changes in Paget-like clusters using the *split.by* function applied to both feature and violin plots. To identify gene biomarkers in Paget/Paget-like cells in EMPD and DTG conditions, we visualized gene expression markers in the Paget cluster using the feature and violin plots.

Dual luciferase activity assay
10 ng of WT and mutant *Pten* and *Cmtm5* reporter constructs were transfected into HEK 293 T cells and cultured in 96-well plates in a water-jacketed incubator at 37 °C with 5% CO$_2$ output. Firefly and Renilla luciferase activity were measured using the Dual-Glo Luciferase Assay system (Promega, Madison, WI).

Capture and Ligation Probe (CLIP)-qPCR
CLIP-qPCR assays were performed as previously described with minor modifications.[77] Briefly, skin epidermal cells from *K14-rtTA; TRE-Msi1* DTG mice treated with Doxycycline for 48 h were isolated as previously described.[78] For crosslinking, single cell suspensions were irradiated twice at 400 mJ/cm^2 using the 2400 Stratalinker system (Stratagene). Epidermal cells were then lysed using PXL buffer (PBS, 0.1% SDS, 0.5% deoxycholate, 0.5% NP-40, including protease and RNase inhibitors). Cells were then centrifuged at 1000 rpm for 5 min and the supernatant was added to anti-Msi1-(MBL) or goat anti-rat IgG (Jackson ImmunoResearch, West Grove, PA)-conjugated protein A Dynabeads (Dynal, 100.02, Thermo-Fisher, Fremont, CA) and incubated for 4 h at 4 °C. The beads were washed and digested with Proteinase K (Roche). Total RNA was extracted from the beads and then quantified by qRT-PCR.

Rapamycin administration
Rapamycin (53123–88–9; LC laboratory) was reconstituted in 100% ethanol to a final concentration of 10 mg/mL and diluted in 5% Tween-80 (Beijing Lablead Biotechnology Co., Ltd.) and 5% PEG-400 (Tianjin Zhiyuan Chemical Reagent Co., Ltd.) to a final concentration of 1 mg/mL. Rapamycin was administered daily via intraperitoneal injection (4 mg per kg of body weight) for 5 consecutive days. The final volume of all injections was 200 µL. Doxycycline was administered to DTG mice for 48 h after the third dose of Rapamycin.

Topical Rapamycin treatment
EMPD skin lesions were treated topically with FDA-approved 0.4% Rapamycin ointment twice daily for two weeks. Skin lesions were then evaluated macroscopically, histologically, and by reflectance confocal microscopy and ultrasound imaging.

Statistical analysis
The detailed information of statistical analysis has been stated in each figure legend. Wilcox test was used for calculating P value in Fig. 1j; Log-rank test was used for calculating P value in Figs. 2e and 7i. Unless noted, unpaired Student's t-test were used for calculating P value using Graphpad (Version 5.0). $P < 0.05$ was considered statistically significant. All data are reported as mean ± SD. The means and standard deviations from at least three independent experiments are represented in all graphs.

Partial materials and methods are included in Supplementary information, Data S1.

AUTHOR CONTRIBUTIONS
Z. Yu designed research; Y.S., C.F.G.-J., Z.C., Y. Tang, X.M., C.L., X.B., M.D., L.B., Y. Tian, R.L., R.Z., J.X., X.S., S.D., Y.L., Y. Zhu and S.S. performed research; C.F.G.-J. performed

scRNA-seq and bioinformatic analyses and curated scRNA-seq data; Q.N. and Z. Yu supervised scRNA-seq and bioinformatic analyses; Y.S., C.F.G.-J., H.C., G.Z., J.S., F.R., L.X., Z. Ying, Y. Zhao, X.D., C.J.L., B.A., M.V.P, Q.N., and Z. Yu analyzed data; C.F.G-J. and Y.S. produced figures; C.F.G.-J., Y.S., Q.N., M.V.P. and Z. Yu wrote the manuscript.

ADDITIONAL INFORMATION

Supplementary information accompanies this paper at https://doi.org/10.1038/s41422-020-0334-5.

Competing interests: The authors declare no competing interests.

REFERENCES

1. Kanitakis, J. Mammary and extramammary Paget's disease. *J. Eur. Acad. Dermatol. Venereol.* **21**, 581–590 (2007).
2. Lopes Filho, L. L. et al. Mammary and extramammary Paget's disease. *An. Bras. Dermatol.* **90**, 225–231 (2015).
3. Sandoval-Leon, A. C., Drews-Elger, K., Gomez-Fernandez, C. R., Yepes, M. M. & Lippman, M. E. Paget's disease of the nipple. *Breast Cancer Res. Treat.* **141**, 1–12 (2013).
4. Lam, C. & Funaro, D. Extramammary Paget's disease: summary of current knowledge. *Dermatol. Clin.* **28**, 807–826 (2010).
5. Smith, K. J., Tuur, S., Corvette, D., Lupton, G. P. & Skelton, H. G. Cytokeratin 7 staining in mammary and extramammary Paget's disease. *Mod. Pathol.* **10**, 1069–1074 (1997).
6. Liegl, B. et al. Mammary and extramammary Paget's disease: an immunohistochemical study of 83 cases. *Histopathology* **50**, 439–447 (2007).
7. Hikita, T., Ohtsuki, Y., Maeda, T. & Furihata, M. Immunohistochemical and fluorescence in situ hybridization studies on noninvasive and invasive extramammary Paget's disease. *Int. J. Surg. Pathol.* **20**, 441–448 (2012).
8. Keatings, L. et al. c-erbB-2 oncoprotein expression in mammary and extramammary Paget's disease: an immunohistochemical study. *Histopathology* **17**, 243–247 (1990).
9. Fu, W., Lobocki, C. A., Silberberg, B. K., Chelladurai, M. & Young, S. C. Molecular markers in Paget disease of the breast. *J. Surg. Oncol.* **77**, 171–178 (2001).
10. Mori, O., Hachisuka, H., Nakano, S., Sasai, Y. & Shiku, H. Expression of ras p21 in mammary and extramammary Paget's disease. *Arch. Pathol. Lab. Med.* **114**, 858–861 (1990).
11. Liegl, B., Horn, L. C. & Moinfar, F. Androgen receptors are frequently expressed in mammary and extramammary Paget's disease. *Mod. Pathol.* **18**, 1283–1288 (2005).
12. Morbeck, D. et al. GATA3 expression in primary vulvar Paget disease: a potential pitfall leading to misdiagnosis of pagetoid urothelial intraepithelial neoplasia. *Histopathology* **70**, 435–441 (2017).
13. Zhang, G., Zhao, Y., Abdul-Karim, F. W. & Yang, B. P16 expression in primary vulvar extramammary Paget Disease. *Int. J. Gynecol. Pathol.* https://doi.org/10.1097/PGP.0000000000000602 (2019).
14. Ito, T., Kaku-Ito, Y. & Furue, M. The diagnosis and management of extramammary Paget's disease. *Exp. Rev. Anticancer Ther.* **18**, 543–553 (2018).
15. Sek, P., Zawrocki, A., Biernat, W. & Piekarski, J. H. HER2 molecular subtype is a dominant subtype of mammary Paget's cells. An immunohistochemical study. *Histopathology* **57**, 564–571 (2010).
16. Caliskan, M. et al. Paget's disease of the breast: the experience of the European Institute of Oncology and review of the literature. *Breast Cancer Res. Treat.* **112**, 513–521 (2008).
17. Marucci, G. et al. Toker cells are probably precursors of Paget cell carcinoma: a morphological and ultrastructural description. *Virchows Arch.* **441**, 117–123 (2002).
18. Mehta, N. J., Torno, R. & Sorra, T. Extramammary Paget's disease. *South Med. J.* **93**, 713–715 (2000).
19. Nakamura, M., Okano, H., Blendy, J. A. & Montell, C. Musashi, a neural RNA-binding protein required for *Drosophila* adult external sensory organ development. *Neuron* **13**, 67–81 (1994).
20. Sakakibara, S. et al. Mouse-Musashi-1, a neural RNA-binding protein highly enriched in the mammalian CNS stem cell. *Dev. Biol.* **176**, 230–242 (1996).
21. Sakakibara, S., Nakamura, Y., Satoh, H. & Okano, H. Rna-binding protein Musashi2: developmentally regulated expression in neural precursor cells and subpopulations of neurons in mammalian CNS. *J. Neurosci.* **21**, 8091–8107 (2001).
22. Kayahara, T. et al. Candidate markers for stem and early progenitor cells, Musashi-1 and Hes1, are expressed in crypt base columnar cells of mouse small intestine. *FEBS Lett.* **535**, 131–135 (2003).
23. Li, N. et al. The Msi family of RNA-binding proteins function redundantly as intestinal oncoproteins. *Cell Rep.* **13**, 2440–2455 (2015).
24. Yousefi, M. et al. Msi RNA-binding proteins control reserve intestinal stem cell quiescence. *J. Cell Biol.* **215**, 401–413 (2016).
25. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
26. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
27. Stuart, T. et al. Comprehensive Integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
28. Hu, H. B., Yang, X. P., Zhou, P. X., Yang, X. A. & Yin, B. High expression of keratin 6C is associated with poor prognosis and accelerates cancer proliferation and migration by modulating epithelial-mesenchymal transition in lung adenocarcinoma. *Genes Genom.* **42**, 179–188 (2019).
29. Wang, S. et al. Single cell transcriptomics of human epidermis reveals basal stem cell transition states. *bioRxiv* https://doi.org/10.1101/784579 (2019).
30. Kowalczyk, M. S. et al. Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res.* **25**, 1860–1872 (2015).
31. Mori, O., Karashima, T., Matsuo, K. & Hashimoto, T. Epidermal cell cultures from involved skin of patients with mammary and extramammary Paget's disease. *Kurume Med. J.* **44**, 165–169 (1997).
32. Thomas, P. D. et al. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* **13**, 2129–2141 (2003).
33. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
34. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
35. Iga, N. et al. Accumulation of exhausted CD8+ T cells in extramammary Paget's disease. *PLoS One* **14**, e0211135 (2019).
36. Sugiyama-Nakagiri, Y., Akiyama, M., Shibata, S., Okano, H. & Shimizu, H. Expression of RNA-binding protein Musashi in hair follicle development and hair cycle progression. *Am. J. Pathol.* **168**, 80–92 (2006).
37. Benoit, S. et al. Elevated serum levels of calcium-binding S100 proteins A8 and A9 reflect disease activity and abnormal differentiation of keratinocytes in psoriasis. *Br. J. Dermatol.* **155**, 62–66 (2006).
38. Marionnet, C. et al. Modulation of gene expression induced in human epidermis by environmental stress in vivo. *J. Invest. Dermatol.* **121**, 1447–1458 (2003).
39. Leyva-Castillo, J. M., Hener, P., Jiang, H. & Li, M. TSLP produced by keratinocytes promotes allergen sensitization through skin and thereby triggers atopic march in mice. *J. Invest. Dermatol.* **133**, 154–163 (2013).
40. Hildebrand, J. D. Shroom regulates epithelial cell shape via the apical positioning of an actomyosin network. *J. Cell Sci.* **118**, 5191–5203 (2005).
41. Wolber, R. A., Dupuis, B. A. & Wick, M. R. Expression of c-erbB-2 oncoprotein in mammary and extramammary Paget's disease. *Am. J. Clin. Pathol.* **96**, 243–247 (1991).
42. Xu, T. et al. [CMTM5 inhibits the tumor cell behavior of prostate cancer by downregulation of HER2]. *Beijing Da Xue Xue Bao Yi Xue Ban* **42**, 386–390 (2010).
43. Fu, Y. F., Gui, R. & Liu, J. HER-2-induced PI3K signaling pathway was involved in the pathogenesis of gastric cancer. *Cancer Gene. Ther.* **22**, 145–153 (2015).
44. Zhong, H. et al. Modulation of hypoxia-inducible factor 1alpha expression by the epidermal growth factor/phosphatidylinositol 3-kinase/PTEN/AKT/FRAP pathway in human prostate cancer cells: implications for tumor angiogenesis and therapeutics. *Cancer Res.* **60**, 1541–1545 (2000).
45. Hudson, C. C. et al. Regulation of hypoxia-inducible factor 1alpha expression and function by the mammalian target of rapamycin. *Mol. Cell Biol.* **22**, 7004–7014 (2002).
46. Norrenberg, S. et al. Retrospective study: Rapamycin or rapalog 0.1% cream for facial angiofibromas in tuberous sclerosis complex: evaluation of treatment effectiveness and cost. *Br. J. Dermatol.* **179**, 208–209 (2018).
47. Lin, N. et al. Expression of the p38 MAPK, NF-kappaB and cyclin D1 in extramammary Paget's disease. *J. Dermatol. Sci.* **45**, 187–192 (2007).
48. Chen, S. Y. et al. Concordant overexpression of phosphorylated ATF2 and STAT3 in extramammary Paget's disease. *J. Cutan. Pathol.* **36**, 402–408 (2009).
49. Cosgarea, I., Zaremba, A. & Hillen, U. Extramammary Paget's disease. *Hautarzt* **70**, 670–676 (2019).
50. Ghazizadeh, Z. et al. Prospective isolation of ISL1(+) cardiac progenitors from human ESCs for myocardial infarction therapy. *Stem Cell Rep.* **10**, 848–859 (2018).
51. Smith, N. R. et al. Cell adhesion molecule CD166/ALCAM functions within the crypt to orchestrate murine intestinal stem cell homeostasis. *Cell. Mol. Gastroenterol. Hepatol.* **3**, 389–409 (2017).
52. Du, X. et al. Extramammary Paget's disease mimicking acantholytic squamous cell carcinoma in situ: a case report. *J. Cutan. Pathol.* **37**, 683–686 (2010).
53. Tanskanen, M., Jahkola, T., Asko-Seljavaara, S., Jalkanen, J. & Isola, J. HER2 oncogene amplification in extramammary Paget's disease. *Histopathology* **42**, 575–579 (2003).

54. Cho, Z. et al. Podoplanin expression in peritumoral keratinocytes predicts aggressive behavior in extramammary Paget's disease. *J. Dermatol. Sci.* **87**, 29–35 (2017).

55. Kudinov, A. E. et al. Musashi-2 (MSI2) supports TGF-beta signaling and inhibits claudins to promote non-small cell lung cancer (NSCLC) metastasis. *Proc. Natl. Acad. Sci. USA* **113**, 6955–6960 (2016).

56. Chen, S. et al. Immunohistochemical analysis of the mammalian target of rapamycin signalling pathway in extramammary Paget's disease. *Br. J. Dermatol.* **161**, 357–363 (2009).

57. Hata, H. et al. mTOR expression correlates with invasiveness and progression of extramammary Paget's disease. *J. Eur. Acad. Dermatol. Venereol.* **30**, 1238–1239 (2016).

58. Janku, F., Yap, T. A. & Meric-Bernstam, F. Targeting the PI3K pathway in cancer: are we making headway? *Nat. Rev. Clin. Oncol.* **15**, 273–291 (2018).

59. Kato, R. et al. Efficacy of everolimus in patients with advanced renal cell carcinoma refractory or intolerant to VEGFR-TKIs and safety compared with prior VEGFR-TKI treatment. *Jpn. J. Clin. Oncol.* **44**, 479–485 (2014).

60. Yao, J. C. et al. Everolimus for advanced pancreatic neuroendocrine tumors. *N. Engl. J. Med.* **364**, 514–523 (2011).

61. Sasaki, M. et al. Anorectal mucinous adenocarcinoma associated with latent perianal Paget's disease. *Am. J. Gastroenterol.* **85**, 199–202 (1990).

62. Yim, J. H., Wick, M. R., Philpott, G. W., Norton, J. A. & Doherty, G. M. Underlying pathology in mammary Paget's disease. *Ann. Surg. Oncol.* **4**, 287–292 (1997).

63. Morandi, L. et al. Intraepidermal cells of Paget's carcinoma of the breast can be genetically different from those of the underlying carcinoma. *Hum. Pathol.* **34**, 1321–1330 (2003).

64. Katz, Y. et al. Musashi proteins are post-transcriptional regulators of the epithelial-luminal cell state. *Elife* **3**, e03915 (2014).

65. Espersen, M. L., Olsen, J., Linnemann, D., Hogdall, E. & Troelsen, J. T. Clinical implications of intestinal stem cell markers in colorectal cancer. *Clin. Colorectal Cancer* **14**, 63–71 (2015).

66. Fan, J. et al. Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. *Nat. Methods* **13**, 241–244 (2016).

67. Guerrero-Juarez, C. F. et al. Single-cell analysis reveals fibroblast heterogeneity and myeloid-derived adipocyte progenitors in murine skin wounds. *Nat. Commun.* **10**, 650 (2019).

68. Cheng, J. B. et al. Transcriptional programming of normal and inflamed human epidermis at single-cell resolution. *Cell Rep.* **25**, 871–883 (2018).

69. Qiu, X. et al. Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **14**, 309–315 (2017).

70. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

71. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).

72. Jin, S., MacLean, A. L., Peng, T. & Nie, Q. scEpath: energy landscape-based inference of transition probabilities and cellular trajectories from single-cell transcriptomic data. *Bioinformatics* **34**, 2077–2086 (2018).

73. Zhang, H. M. et al. AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors. *Nucleic Acids Res.* **43**, D76–D81 (2015).

74. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).

75. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).

76. Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform.* **14**, 128 (2013).

77. Wang, S. et al. Transformation of the intestinal epithelium by the MSI2 RNA-binding protein. *Nat. Commun.* **6**, 6517 (2015).

78. Ma, X. et al. Msi2 maintains quiescent state of hair follicle stem cells by directly repressing the Hh signaling pathway. *J. Invest. Dermatol.* **137**, 1015–1024 (2017).