Single-cell transcriptomics reveals spatial and temporal turnover of keratinocyte differentiation regulators

Alex Finnegan

University of Illinois at Urbana-Champaign

Raymond J Cho

University of California, San Francisco

Alan Luu

University of Illinois at Urbana-Champaign <https://orcid.org/0000-0001-7363-4379>

Paymann Harirchian

University of California, San Francisco

Jerry Lee

University of California, San Francisco

Jeffrey B Cheng

University of California, San Francisco

Jun S Song (✉ songj@illinois.edu)

University of Illinois at Urbana-Champaign <https://orcid.org/0000-0002-0422-2175>

Research article

Keywords: Single-cell analysis, transcription regulation, keratinocyte, antioxidant, basal cell carcinoma, squamous cell carcinoma

Posted Date: April 15th, 2019

DOI: <https://doi.org/10.21203/rs.2.9136/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at Frontiers in Genetics on September 3rd, 2019. See the published version at <https://doi.org/10.3389/fgene.2019.00775>.

Abstract

Background: Keratinocyte differentiation requires intricately coordinated spatiotemporal expression changes that specify epidermis structure and function. **Results:** This paper utilizes single-cell RNA-seq data from 22,338 human foreskin keratinocytes to reconstruct the transcriptional regulation of skin development and homeostasis genes, organizing them by differentiation stage and also into transcription factor (TF)-associated modules. We identify groups of TFs characterized by coordinate expression changes during progression from the undifferentiated basal to the differentiated state and show that these TFs also have concordant differential predicted binding enrichment in the super-enhancers previously reported to turn over between the two states. The identified TFs form a core subset of the regulators controlling gene modules essential for basal and differentiated keratinocyte functions, supporting their nomination as master coordinators of keratinocyte differentiation. Experimental depletion of the TFs ZBED2 and ETV4, both predicted to promote the basal state, induces differentiation. Furthermore, our single-cell RNA expression analysis reveals preferential expression of antioxidant genes in the basal state, suggesting keratinocytes actively suppress reactive oxygen species to maintain the undifferentiated state. Finally, we perform in-silico lineage tracing demonstrating transcriptomic correspondence of basal cell carcinoma (BCC) to a basal keratinocyte subpopulation and squamous cell carcinoma (SCC) to a rapidly proliferating stage that is intermediate between the basal and differentiated cell states, implicating distinct candidate cells of origin for these cancers. **Conclusion:** Our work demonstrates diverse computational methods integrating single-cell and bulk RNA expression data to advance our understanding of dynamic gene regulation in normal development and in disease.

Background

Keratinocytes, the predominant cell type of mammalian epidermis, regulate their gene expression programs to fulfill specialized cellular functions within the different epidermal strata. Additionally, they must balance self-renewal against cell loss, given the epidermis' intrinsic replacement rate of ~28 days in normal human skin. How keratinocytes dynamically govern the hierarchy of self-renewal, differentiation and maturation remains poorly understood. This paper reconstructs the dynamic gene regulatory network rearrangements that occur with keratinocyte differentiation by analyzing human foreskin single-cell RNA-seq (scRNA-seq) data.

Basal keratinocytes (BKs) comprise the basal layer, the innermost layer of the epidermis. BKs divide at controlled rates that are thought to be heterogeneous across progenitor cells, ranging from rarely dividing self-renewing stem cells to rapidly cycling transit amplifying cells [1]. In addition to replicating, BKs constitute the basement membrane which is critical for adhesion of the epidermis and dermis and participate in intercellular signaling required for maintaining tissue homeostasis. Upon differentiation, differentiated keratinocytes (DKs) exit the cell cycle and travel from the basal layer through the more superficial spinous and granular layers culminating in cornification/cell death. During the differentiation process, keratinocytes synthesize components necessary for epidermal barrier function, including desmosomes (specialized adhesion structures) in the spinous layer, secretory organelles called lamellar granules that contain lipids and enzymes, and keratohyalin granules which contain proteins such as loricrin – the latter two providing vital components of the cornified lipid envelope of the epidermis' outer stratum corneum layer.

At the transcriptomic level, the stratum-specific expression patterns of many key keratinocyte genes are known, but regulators of these genes are still being identified [2]. Constructing the dynamic regulatory network of relevant transcription factors (TFs) and their target genes thus remains an active area of investigation. Previous studies have used various genomic and epigenomic data to construct regulatory networks. For example, Lopez-Pajares *et al.* [2] analyzed the time-series transcriptome of experimentally differentiated keratinocyte cultures and identified regulatory relations of genes based on temporal co-expression patterns. Joost *et al.* [3] advanced this approach to the single-cell level in murine epidermis, identifying TFs varying with differentiation pseudotime and constructing gene modules using correlation-based expression similarity. In *in-vitro* keratinocyte epigenomic studies, Cavazza *et al.* [4] and Klein *et al.* [5] mapped typical enhancers and super-enhancers (SEs) – large clusters of enhancers characterized by strong activating histone modifications, enrichment of cell-type-specific TF motifs, and regulation of cell-type-specific genes [6]. Both works identified dramatic changes in sets of

SEs between the BK and DK states and developed regulatory networks based on patterns of TF binding/motif enrichment in SEs and proximities of SEs to gene loci [4, 5]. More recently, the single-cell Perturb-ATAC method revealed changes in regulatory element chromatin accessibility during keratinocyte differentiation and targeted genetic perturbation [7]; these data permitted the grouping of TFs with correlated binding site accessibility during differentiation, the inference of interactions between TFs, and the detection of synergy in perturbations of chromatin accessibility [7].

While regulation by TFs and epigenetic modifications ultimately determine gene expression, changes in redox state and abundance of reactive oxygen species (ROS) may help guide the transition from basal to differentiated states [8]. For instance, Hamanaka *et al.* [9] demonstrated that reducing ROS through inhibition of oxidative phosphorylation impairs epidermal differentiation and increases proliferation of basal cells and that treatment of cultured keratinocytes with antioxidants impairs differentiation. Likewise, Bhaduri *et al.* [10] established MPZL3 and FDXR as proteins localizing to the mitochondria and inducing keratinocyte differentiation by increasing ROS levels. These findings demonstrate opposing roles of ROS and antioxidants in regulating differentiation; however, a genome-wide time-course examination of genes potentially modulating differentiation via their antioxidant function has not yet been described.

In this paper, we use our recently generated scRNA-seq data assaying expression in 22,338 human foreskin keratinocytes [11] to identify regulators of keratinocyte differentiation and computationally infer dynamic TF networks controlling gene expression patterns required for keratinocyte development and function. We find that expression turnover of established and predicted keratinocyte regulators coincides with previously reported change in SE sets between the BK and DK states [5]. Depletion of two predicted positive regulators of BKs – ZBED2 and ETV4 – leads to differentiation of BKs in the absence of external differentiation-inducing queues. The pattern of differential TF binding-motif enrichment between BK- and DK-specific SEs follows the pattern of TF state-specific expression, leading us to develop gene regulatory networks for TFs. These networks recapitulate known and previously predicted regulatory relationships and also identify novel regulators of differentiation-stage-specific functions. In particular, our predicted regulation of cadherins by ETV4 suggests that ETV4's established role of controlling cadherin-mediated cell sorting in branches of the neuronal lineage [12, 13] may extend to keratinocytes. Supporting the role of cellular antioxidants in suppressing ROS levels, we find that genes related to antioxidant function are preferentially expressed in BK cells and also uncover differences in subcellular localization between antioxidant genes exclusively expressed in BK state and those in DK state. Finally, we perform *in-silico* lineage tracing in the constructed transcriptomic landscape of differentiation hierarchy for the two most common skin cancers, basal cell carcinoma (BCC) and cutaneous squamous cell carcinoma (SCC), implicating the BK state and a mitotic population intermediate between the BK and DK states as respective cells of origin for these two cancer types.

Results

A subset of keratinocyte-specific transcription factors show expression and binding patterns coupled to state-specific epigenomes

To identify expression patterns of key TFs across distinct keratinocyte transcriptomic states, we examined a set of 49 established and 44 candidate keratinocyte regulators, to which we refer below as Keratinocyte TFs. Established keratinocyte regulators were obtained from a previous publication [5]; Candidate TFs were identified based on keratinocyte-specific RNA expression in the FANTOM5 (Functional ANnotation Of the Mammalian genome) cell atlas [14] (Methods; Additional file 1: Figure S1; Additional file 2: Tables S1, Table S2). Our approach of selecting candidates based on cell-type-specific expression aimed to increase the confidence that changes in TF expression across single-cell transcriptional states reflect rewiring of gene regulatory networks guiding keratinocyte differentiation and to reduce false positives in subsequent identification of TF targets from correlation analysis.

We clustered foreskin keratinocytes into 8 stages via approximate spectral clustering of imputed scRNA expression values (Figure 1A; Additional file 1: Figure S2; Methods). As observed previously [11], marker gene expression profiles indicated that these stages largely agreed with known keratinocyte states including a BK state (corresponding to stages 1-3), a mitotic state

(stage 4), and a DK state (stages 5-7) (Additional file 1: Figure S3). The mitotic state had markedly increased levels of cyclins as well as the histone H2A isoform *HIST2H2AC* known to be required for proliferation of undifferentiated mammary epithelial cells [15]. Additionally, the mitotic state had high expression of basal markers (*KRT5*, *KRT14*) and intermediate expression of early differentiation markers (*KRT1*, *KRT10*), suggesting it is a rapidly cycling subpopulation in transition from the BK to DK states (Additional file 1: Figure S3). This interpretation is supported by in-situ hybridization experiments that have identified basal and supra-basal expression of the mitotic marker gene *MKI67* [11]. Stage 8 reproduced the “channel” cluster, identified previously as a novel keratinocyte cell state not on the classic differentiation trajectory [11].

Hierarchical clustering of Keratinocyte TFs that exhibit dynamic expression across stages 1-7 clearly separated the TFs with peak expression in the BK state from those with peak expression in the DK state (Figure 1B), with a sharp transition occurring in the mitotic state (stage 4). This pattern of expression turnover coincided with the dramatic change in distribution of active SEs between the BK and DK states (previously identified from differential histone modification patterns of H3K4 monomethylation, H3K4 trimethylation and H3K27 acetylation)[4, 5]. We therefore hypothesized that the TFs with peak expression in each state may function through direct binding of state-specific SEs, thereby coupling the transcriptional and epigenetic developmental programs. To test this hypothesis, we performed a differential motif enrichment analysis for the dynamic TFs and identified 21 and 14 TFs with motifs significantly differentially enriched between BK and DK SEs, respectively (Methods, Figure 1B last row). The direction and magnitude of TF motif enrichment in BK vs. DK SEs generally agreed with the peak expression state of the corresponding TF (, one-sided Mann-Whitney U test). This observation, made possible by single-cell analysis, supported the premise that keratinocyte TF expression and chromatin conformation accessibility are coordinated during transition between keratinocyte cell states.

Next, we identified potential regulators of the switch in state-specific SEs by examining the stage-wise expression of established keratinocyte epigenetic regulators and found several of them, including *EZH2*, *DNMT1* and *UHRF1*, to have a strong expression spike in the mitotic state [16, 17] (Figure 1C). Additionally, we found that *H2A.Z* and components of the SWR1 remodeling complex, responsible for depositing this enhancer-associated histone subunit, attained peak expression in the mitotic state (Figure 1D,E). Although the sharp increase in the expression of *H2A.Z* and other histone subunits in this state may be partially explained by the abundance of rapidly dividing cells, the concurrent peak expression of SWR1 components suggested active reorganization of enhancer activities prior to differentiation. Together, these single-cell results highlighted epigenetic remodelers functioning during the mitotic state, potentially to facilitate the turnover of SEs between the BK and DK states.

Knockdown of *ETV4* and *ZBED2*, predicted promoters of the BK state, induces differentiation

To validate the regulatory function of Candidate Keratinocyte TFs, we ranked the TFs based on their predicted ability to promote the BK state. Candidates were assigned a differentiation-promoting score by first identifying highly correlated keratinocyte-specific regulatory targets and summing their log fold-changes between DK and BK states, accounting for the sign of correlation (Methods; Additional file 1: Figure S4). We filtered out TFs with low expression in undifferentiated keratinocyte cultures (less than 5 FPKM) and attempted to knock down the top six remaining TFs with greatest BK-promoting strength (strong negative differentiation-promoting score) using RNAi in the absence of external differentiation queues.

Depletion of *ETV4* and *ZBED2* transcripts resulted in a significant increase in mRNA expression of the early differentiation marker *KRT10* by 3.84 and 4.17 fold, respectively, compared to control cells transfected with non-targeting siRNA (Figure 1F; Additional file 1: Figure S5A). Depletion of *ETV4* also showed a significant increase (2.49-fold) in the mRNA expression of the late differentiation marker *FLG*, with *ZBED2* depletion also showing a similar trend (Figure 1F). These results confirmed the strong progenitor-promoting function of *ETV4* and *ZBED2*, synthetic reduction of which induced spontaneous differentiation of keratinocytes.

Depletion of *BNC1* and *HOXC11* transcripts did not significantly change the mRNA level of *KRT10* or *FLG* (Additional file 1: Figure S5B,C), suggesting that regulatory effects of these TFs do not extend to these differentiation markers or that *BNC1*

and HOXC11 protein expression was not diminished enough to have an effect. Nevertheless, previous knockout of *BNC1* in mouse significantly decreased the number of proliferating keratinocytes in the cornea of the eye [18]. Therefore, we conclude that BNC1 likely promotes the BK state in foreskin, although its regulatory targets remain to be experimentally characterized.

Attempted depletion of *SOX9* and *IRX4* resulted in low siRNA knockdown efficiency and did not significantly change the expression of differentiation marker genes (data not shown); nevertheless, previous reports supported our prediction of their role in keratinocyte differentiation. For example, overexpression of SOX9 in keratinocytes has been shown to suppress the late differentiation marker genes IVL and LOR [19]. Likewise, IRX4 was previously predicted to regulate keratinocyte proliferation and hemidesmosome assembly based on correlation with functionally annotated genes across a large set of publicly available mouse RNA-seq data [20]. Moreover, knockdown of the differentiation promoting TF GRHL3 in calcium-induced keratinocyte primary cells resulted in a gain of SEs strongly enriched for the IRX4 motif [5], suggesting antagonism between IRX4 and this established pro-differentiation TF. Overall, our prioritization of Candidate TFs revealed novel keratinocyte regulators and provided additional candidates for follow-up experiments.

Gene modules in the basal network promote tissue architecture, control of Hippo signaling and progression to the mitotic state.

We next sought to assign function to Keratinocyte TFs with motifs enriched in state-specific SEs based on their scRNA-seq expression correlation with a set of potential regulatory targets. This set was composed of the Keratinocyte TFs themselves and an additional 747 genes differentially upregulated in FANTOM5 keratinocytes compared to other cell types (Methods; Additional file 2: Table S2). Focusing first on the regulatory network governing the BK state and its progression to the mitotic state, we clustered the Keratinocyte TFs with enriched motifs in BK SEs based on their expression similarity across single cells in stages 1-4. We then clustered the regulatory targets into gene modules based on the similarity of their correlations to the TFs. Organizing the TF/target correlation matrix by TF and gene modules (Additional file 1: Figure S6A) yielded sub-matrices with strong correlation/anti-correlation delineated by module boundaries. Thresholding on the average correlation strength calculated across gene/TF pairs for each TF and gene module, we identified activating and inhibiting relationships between 13 TF and 23 target gene modules (Additional file 1: Figure S6(B-D); Additional file 2: Table S3; Additional file 3: Regulatory network construction).

Figure 2A shows regulatory relationships for four gene modules enriched in gene ontology (GO) terms (Figure 2B) (see Additional file 2: Table S4 for full GO output). Gene Module 1 was highly expressed in all BK stages and contained genes important for anchoring cells to the basement membrane and extracellular matrix via hemidesmosomes and other cell junctions, genes encoding extracellular signaling molecules, and genes participating in the key Hippo and PI3K intracellular signaling pathways. Transcription factors predicted to activate Module 1 genes recapitulated several established and independently predicted regulatory relationships. For example, TP63 and JUND are known to positively regulate *ITGB4* and *LAMA3A*, respectively [21, 22], while IRX4 and JUND are both predicted regulators of hemidesmosome assembly [20].

Notably, 4 of the 6 genes in the Hippo pathway (*AJUBA*, *WNT7A*, *WNT7B*, and *WNT3*) and 7 of the 8 genes in the PI3K pathway (*ITGA3*, *LAMB4*, *LAMB3*, *FGFR2*, *COL4A6*, *ITGB4* and *LAMA3*) were expressed as extracellular or cell membrane-associated proteins. Given that these pathways involve signaling via intracellular post-translation modification, this result suggested that the primary mechanism for pathway modulation at the transcriptional level might be via changing the expression of extracellular signaling molecules and the cell membrane proteins that transduce these signals. Examining the position of Module 1 genes in the Hippo signaling pathway [23] illustrated this mechanism and showed that Module 1 genes promoted the pro-proliferative Hippo-OFF signaling state (Additional file 1: Figure S7). Specifically, the Module 1 cell membrane-associated protein AJUBA and intracellular protein RASSF6 are known to repress MST1/2, allowing nuclear localization of YAP/TAZ, which defines the pro-proliferative Hippo-OFF state [24]. In the nucleus, TFs activated downstream of Module 1 extracellular WNT signaling proteins (*WNT7A*, *WNT7B* and *WNT3*) can interact with YAP to promote pro-proliferative genes, including the Module 1 gene *CCND2* [23].

Module 2 genes were enriched for keratins and rose sharply in expression at stage 4. Consistent with the strong mitotic signal at this stage, two of the three keratins in this module (*KRT6A* and *KRT6B*) were previously implicated in rapid keratinocyte division [25]. Moreover, *KRT6A* and *KRT6B* were also shown to suppress keratinocyte migration during wound repair [26], suggesting that the sharp rise in *KRT6A/B* expression in stage 4 and its fall beyond stage 5 could help inhibit migration of this mitotic cell population from the basal layer (Additional file 1: Figure S8). The proposed mechanism of impaired migration may explain how this mitotic population remains in or near the basal layer, despite expressing spinous layer markers (e.g. *KRT1* and *KRT10*) at higher levels than BK cells (Additional file 1: Figure S3).

The predicted function of several regulators of Module 2 genes were confirmed by previous reports. For example, TP63 knockdown was shown to increase the expression of *KRT6A* in human keratinocyte cell lines [27]. Similarly, conditional knockout of glucocorticoid receptor NR3C1 in mouse keratinocytes was shown to increase the expression of *KRT6A*, *KRT6B* and *KRT77*, another keratin in the gene module [28].

Gene Module 4 was enriched for MAPK signaling genes (*CRKL*, *FGF11*, *GADD45A*, *FLNB*, *DUSP7*, *MYC*) and rose sharply in expression at stage 2. The overall effect of Module 4 gene expression on MAPK signaling was complex, with *FGF11* and *GADD45A* activating the ERK and JNK pathways [23], *DUSP7* inhibiting ERK, JNK and p38 pathways [23, 29], and *CRKL* and *FLNB* serving structural functions. Moreover, different outcomes have been reported for activation of MAPK signaling by Module 4 genes. On the one hand, activation of JNK and P38 pathways by the DNA damage response gene *GADD45A* can promote apoptosis and cell cycle arrest [30]. On the other hand, activation of ERK signaling by growth factor *FGF11* may promote proliferation [31]. These results, together with our finding of gene Module 4 regulation by multiple TF modules, including MAPK regulatory targets *FOS*, *JUN* [29] and *FOSL1* [32], suggested complex regulation with multiple feedback mechanisms in controlling proliferation, differentiation and apoptosis.

Gene modules in the differentiated network promote keratinization, barrier formation and downregulation of basal state signaling

We next constructed regulatory relationships among gene and TF modules for the DK state using the same method described above, calculating gene correlations across cells in stages 4-7 and restricting attention to TFs with motifs enriched in DK-specific SEs (Methods). This analysis identified activating and inhibiting relationships among 21 gene and 9 TF modules (Additional file 1: Figure S9; Additional File 2: Table S3). Figure 3A shows regulatory relationships for six gene modules enriched in GO terms (Figure 3B) (see Additional file 2: Table S4 for full GO output).

Gene Module 1 decreased in expression with differentiation and was enriched for GO terms associated with intercellular signal receptors and intracellular signaling cascades. Many module genes associated with these terms were also seen to function in basal state signaling pathways. For example, module genes in the Hippo pathway included cell membrane-associated *AJUBA*, *WNT7B* and *DLG5* [23, 33, 34]. Module genes in the MAPK pathway included receptor tyrosine kinases *FGFR3* and *DDR1* [35, 36], the kinases *MAPKBP1* and *TNK1* [37, 38], the receptor *ADIPOR1* [39] and the phosphoprotein and TF *ATF5*. The decreasing expression of this signaling module thus reflected a shift in the primary cellular function upon differentiation, with basal cells balancing self-renewal and amplification via abundant signaling between and within cells, while differentiated cells began suppressing signaling proteins in favor of those needed for barrier function. Several positive regulators of this module are known to promote cell cycling, making them plausible regulators of the associated MAPK and Hippo pathways. These regulators included *KLF16* which suppresses cyclin-dependent kinase inhibitor *CDKN1A* [40] and *MYC* whose knockdown prevents keratinocyte proliferation [41].

Gene Module 4 also decreased with differentiation and was enriched for genes involved in EGF-like calcium binding and cell adhesion. Cell adhesion genes included several members of the cadherin superfamily: *CDH3*, *FAT1* and *DSG3*. Predicted activators of this module included our experimentally validated TF *ETV4* (Figure 1F), which was previously shown to positively regulate cadherins in mouse spinal cord motor neurons, promoting segregation of cells with similar function [12,

13]. Moreover, it was also demonstrated that ETV4 can positively regulate *RUNX1*, another Module 4 gene [13]. These findings thus supported that the cadherin regulatory function of ETV4 in the neuronal lineage may extend to keratinocytes.

Gene Module 3 increased its expression with differentiation and was enriched for genes related to the formation of cornified envelope and differentiated keratinocyte function. For example, the protein products of *LOR*, *SPRR1B* and *CSTA* in this module are peptides cross-linked in the cornified envelope, while the keratinocyte differentiation protein ACER1 hydrolyzes ceramides, abundant in the granular layer, producing free sphingoid bases with antimicrobial function [42]. Two other important epidermis development genes in this module were *KLK7* and *CALML5*; *KLK7* degrades cellular adhesions of the cornified layer, favoring desquamation [43], and *CALML5* is thought to regulate differentiation by mediating cytoplasmic sequestration of YAP1 and initiating the anti-proliferative Hippo-ON state [44]. This gene module did not have positive TF regulators in our network, but had two sets of negative regulators (Modules 3 and 4). Of note, TF Module 4 contained SP3, ETS1, and SMAD4 that were previously shown to interact physically and suppress hematopoiesis [45, 46]. Our analysis thus indicated that steady reduction of these TFs contributed to the de-repression of Module 3 genes during differentiation.

Gene Module 5, like Module 3, increased its expression with differentiation and was negatively regulated by the TF Modules 3 and 4. It contained genes primarily involved in barrier function, with several of these genes (*DEGS2*, *CERS3*, *ABCA12*, *TMEM79*) functioning in lipid synthesis and transport via the lamellar granule system. Other Module members were involved in cell-cell adhesion (desmosomal proteins DSC1, DSG1 and PERP), tight junctions (CLDN1 and CLDN8), and desquamation (serine-proteases *KLK8*, *KLK11*) [47]. Finally, the module also contained the enzymes TGM3 and CASP14 that promote cornification, DK-specific signaling molecules genes *KRTDAP* and *DMKN* [48, 49] and the anti-microbial gene *DEFB1* [50]. Apart from negative regulation by TF Modules 3 and 4, Gene Module 5 was positively regulated by TF Module 5. This TF module includes RORA, which is known to positively regulate *ABCA12* and other genes functioning in the granular lipid barrier [51]. Our analysis thus identified Module 3 and Module 5 genes as key components of keratinocyte terminal differentiation coordinately regulated by TFs that may preferentially localize in DK-specific SEs to either suppress or promote terminal differentiation.

Antioxidant gene expression is enriched in the basal state and coupled to the spatial organization of epidermis

Given the documented role of ROS and antioxidants in modulating keratinocyte differentiation [9, 10], we also used our scRNA-seq data to examine coordination between antioxidant gene expression and differentiation state. Clustering of annotated antioxidant genes [52] selected for dynamic expression across stages identified three distinct expression clusters (Figure 4A, Methods). The majority of antioxidant genes (20 of 32) belonged to the magenta cluster with peak expression in the basal state. The size of this cluster was significantly larger than expected by chance ($p =$, Methods), suggesting that antioxidant genes were preferentially expressed in the basal state to preserve self-renewal capacity by preventing ROS accumulation [8]. In support of this conclusion, the magenta cluster contained the gene *SOD2* whose conditional knockout in mouse keratinocytes has been shown to induce cellular senescence and elevate the expression of differentiation marker genes at wound sites [53].

The remaining two clusters (orange and green) attained peak expression in stages 4-5 and stages 5-7, respectively. Given the putative role of magenta class genes in preserving the basal state, we sought to identify distinct functions for these late peaking clusters. GO analysis revealed that magenta cluster proteins were enriched in organelle lumens; by contrast, green cluster gene products were enriched in cytoplasmic vesicles, with a similar trend holding for the group of all genes not in the magenta cluster (Figure 4B; Additional file 2: Table S5). This difference in cellular localization reflected potential differences in function, with magenta cluster proteins localized in key organelles to prevent the initiation of differentiation and green cluster proteins diffused throughout the cytoplasm to mitigate environmental oxidative stress and protect basal cells. Supporting this interpretation, the genes not in the magenta cluster were enriched for the GO term "response to oxidative stress" (Figure 4B).

***In-silico* lineage tracing assigns BCC to basal and SCC to mitotic states**

Studying the dynamic pattern of gene expression and regulation during normal development may help identify the cell of cancer origin and thus elucidate aberrant oncogenic processes leading to cancer. BCC and SCC are the two most common cancers of the epidermis. Previous studies have used experimental lineage tracing of tumor cells [54] or induction of carcinogenesis in distinct cell populations [55, 56] to identify their cell of origin. We reasoned that by comparing the differential expression profiles of individual cell stages to those of BCC and SCC, we may obtain *in-silico* lineage tracing of epidermal cancers.

We thus defined a measure of similarity between cell-stage-specific and cancer-specific differential expression profiles. The magnitude of similarity was assessed by the degree to which upregulation in tumor compared to normal tissue was skewed towards the genes with higher expression in a given stage compared to the remaining stages (Methods). This analysis showed that BCC and SCC were more similar to the basal and mitotic states than to the differentiated states, with the direction of regulation in cancer being positively correlated with the direction in these states, but mostly anti-correlated with that in differentiated stages (Figure 5). In greater detail, BCC was most similar to stage 2 ($p = 1.3 \times 10^{-4}$, one-sided Mann-Whitney U test) and also similar to stages 3 ($p = 3.8 \times 10^{-3}$) and 4 ($p = 9.7 \times 10^{-4}$), albeit to a lesser extent. SCC was found to be highly similar to stage 4 ($p = 3.5 \times 10^{-46}$), corresponding to the mitotic state, and comparatively less similar to stage 2 ($p = 2.0 \times 10^{-5}$). Given the paucity of hair follicle cells in foreskin samples, our inferred cell stages of origin likely pertain to cancers derived from the interfollicular epidermis. Additionally, BCC and SCC displayed overall differences in terms of their rewiring of gene regulation, with BCC mostly upregulating genes, whereas SCC both upregulated and downregulated a large number of genes.

Our single-cell transcriptomic data also shed light on potential oncogenic TFs. We first ranked TFs according to their correlation with genes similarly upregulated in both cancer and given cell stage (i.e., upregulated genes with high gene index in Figure 5A,B; Methods). Additional File 2: Table S6 shows the top 10 TFs upon comparing stages 2 and 4 to both BCC and SCC. Several of these TFs were previously reported to be associated with BCC and SCC or their relevant pathways. For example, high expression level of *ETS1* was implicated in local recurrence of both BCC and SCC [57]. Similarly, aberrant activation of Hedgehog signaling, a primary event in the pathogenesis of BCC, was recently shown to be associated with increased *IRF6* expression in studies of cleft palate in mice [58]. Finally, deregulation of the *PI3K* pathway, a frequent occurrence in SCC, may be affected by the upregulation of long noncoding RNA *LINC00152* by SP1 [59].

Discussion

Keratinocyte function in the basal and differentiated states depend on complex transcriptional regulation involving TFs, epigenetic modifications, and environmental queues from ROS levels and other stimuli. In this work, we have integrated bulk epigenetic profiles and single-cell expression data to better understand the coordination of these regulatory mechanisms. In particular, by considering known and predicted keratinocyte-specific TFs, we have uncovered that the turnover of this master set of TFs upon differentiation is coupled to the reported transition from BK to DK SEs. We have confirmed that synthetically suppressing the TFs ZBED2 and ETV4, identified in this work as crucial promoters of the basal state, leads to acute differentiation of BKs. We have also prioritized candidate promoters of differentiation that may be studied in subsequent experiments.

The single-cell transcriptomic data have also allowed us to identify a population of mitotic cells containing sharp expression spikes for established keratinocyte epigenetic regulators EZH2, DNMT1 and UHRF1, as well as for the enhancer-associated histone H2A.Z and the SWR1 remodeling complex that deposits the histone variant. The fact that *EZH2*, *DNMT1* and *UHRF1* peak expression coincides with the temporal stage of TF and SE turnover underscores the importance of these genes and helps localize their activity during differentiation pseudo-time. Moreover, the co-occurrence of *H2A.Z* and SWR1 complex hyper-transcription with this turnover suggests that these genes may have a previously unappreciated role in epigenetic regulation of keratinocyte transition from BK to DK states.

Network analysis has shown that TFs with differential binding in BK vs. DK SEs regulate distinct sets of gene modules enriched for important keratinocyte functions. Consistent with previous studies, our BK network analysis has highlighted the role of TP63 in basement membrane adhesion and regulation of intercellular signaling pathways including WNT [41], as well as the importance of Hippo signaling in basal keratinocytes [33]. Meanwhile, our DK analysis has identified regulators of terminal differentiation gene modules and implicated ETV4 in regulating cadherin superfamily genes, in a manner similar to its established function in motor neurons of the spinal cord [12, 13]. The role of spinal cord cadherins in segregating cells by function suggests that a subset of ETV4 targets may also mediate epidermal cell sorting to assign specific keratinocyte functions to each epidermal layer.

As a proxy for measuring the degree of ROS suppression at each keratinocyte stage, we have demonstrated preferential expression of antioxidant genes in the BK state and uncovered differences in patterns of subcellular localization between BK- and DK-specific antioxidant genes. Notably, BK-specific antioxidant proteins tend to preferentially localize in organelles, such as the mitochondria, where they may control redox levels or the transduction of redox signals, preventing the onset of differentiation. This finding complements previous results that increased expression of select proteins localizing to the mitochondria promotes differentiation by increasing ROS levels [10]. By contrast, DK-specific antioxidant proteins tend to localize in cytoplasmic vesicles where they may be more important for epidermal barrier function than for regulation of differentiation.

Identifying the cell of cancer origin is an important problem for understanding disease etiology and designing effective therapeutic strategies. To determine the origin and potential drivers of BCC and SCC, we developed a measure of similarity between differential expression patterns and compared the profiles of the epidermal cancers to those of distinct cell stages. Our method has shown that a specific basal subpopulation and mitotic states are likely the respective cells of origin for BCC and SCC, and it has also provided a transcriptome-wide perspective that informs candidate TFs involved in carcinogenesis and tumor progression.

Conclusions

Our integrative models of transcriptional regulation have shown that keratinocyte cell fate determination requires coordinating the expression level of critical TFs with the availability of their binding motifs in differentiation state-specific SEs. The inferred regulatory networks have provided insights into the transcriptional regulation of key genes essential for skin homeostasis and function. We have thus demonstrated that computational analyses of single-cell transcriptomic profiles in the context of other genomic and epigenomic data provide a powerful method for reconstructing the spatial and temporal hierarchy of cellular differentiation processes.

Methods

Source of Keratinocytes

Primary human keratinocytes were isolated from neonatal foreskin surgical tissue discards obtained with written informed consent using protocols approved by the UCSF institutional review board (#10-00944). Further details are provided in Additional file 3: Keratinocyte isolation and primary culture.

Data accession and cell selection.

Raw counts of scRNA-seq data used in this study were obtained from the European Genome-phenome Archive (EGAS00001002927). The data were generated using Chromium Single Cell 3' v2 libraries (10X genomics) from three human epidermal samples collected at each of four anatomical locations/disease conditions. Sequence demultiplexing resulted in counts of unique molecular identifiers (UMI) for genes and non-coding RNA in more than 100,000 cells (see [11] for details). Cell filtering and identification of keratinocytes followed [11], with 92,889 passing quality control metrics and 85,345 of these

identified as keratinocytes based on average marker gene expression in published cell clusters. This manuscript mainly focuses on the foreskin data from this data set.

Identification of keratinocyte-specific genes and transcription factors

Our objective of uncovering regulators and regulatory mechanisms specific to the keratinocyte lineage prompted us to focus analysis on genes and TFs with increased expression in keratinocytes compared to other types of primary cells. On the one hand, focusing on keratinocyte-specific genes and TFs had two benefits: first, it permitted discovery of gene modules particular to keratinocyte functions; and, second, it reduced false positives in our identification of keratinocyte regulators from single-cell data by adding a filter for specificity of expression across primary cells. On the other hand, recognizing that some TFs known to be important for keratinocyte regulation may also function in other cell types, we supplemented the data-driven identification of keratinocyte TFs with a set of established keratinocyte regulators from the literature.

Identification of genes and TFs with significantly increased expression in keratinocytes used the expression data from the FANTOM consortium [14]. Relative log expression normalized expression values for transcription start sites (TSSs) identified from cap analysis of gene expression (CAGE) experiments were obtained from http://fantom.gsc.riken.jp/5/datafiles/latest/extra/CAGE_peaks/hg19.cage_peak_phase1and2combined_tpm_ann.osc.txt.gz. Restricting to 495 human primary cell samples not marked for exclusion from expression analysis in Table S2 of [14], we computed gene-level expression values by associating with each gene's EntrezID the sum of CAGE peak expression values annotated with that ID. We used the Mann-Whitney U test to identify genes and TFs differentially expressed in three keratinocyte samples relative to the remaining 491 samples (due to our interest in epidermal keratinocytes, we excluded the oral keratinocyte sample from consideration). A list of annotated TFs [60] was used to distinguish TFs from other protein coding genes and non-coding RNA. Genes and TFs with Benjamini-Hochberg FDR less than 0.05 and increased average expression in keratinocytes were selected and filtered to include only those with at least 1 UMI (raw data) in at least 1% of all single-cell keratinocytes (Additional file 2: Table S1). This differential expression and filtering procedure yielded 793 genes, termed FANTOM Genes, and 49 TFs.

The set of differentially expressed TFs, prior to filtering for minimum scRNA-seq expression level, contained several members of the HES superfamily: *HES2*, *HES5*, and *HES7*. Of these, only *HES2* passed the filter. However, we observed that two other superfamily members, *HES1* and *HES4*, were robustly expressed and possessed dynamic expression patterns across our single cell data (Figure 1). For this reason and because HES genes are targets of Notch signaling that has an established function in keratinocyte differentiation [61], we elected to add *HES1* and *HES4* to the set of 49 TFs. Below, we refer to the full set of 51 TFs as FANTOM TFs. We supplemented our FANTOM TFs with additional 49 TFs previously shown to regulate keratinocyte differentiation [5]. Lowly expressed TF were filtered using the threshold in single-cell expression as described above. We refer to this set as Klein TFs.

From these FANTOM Genes, FANTOM TFs, and Klein TFs, we constructed final three sets for further analysis. The set termed Keratinocyte TFs consisted of the union of FANTOM TFs and Klein TFs and was used to study the dynamics of TF expression across single-cell stages, as well as for regulatory network analysis. The set termed Candidate Keratinocyte TFs consisted of FANTOM TFs not in the set of Klein TFs and was the focus of TF prioritization and validation. Finally, the set termed Keratinocyte Genes consisted of the union of Keratinocyte TFs and FANTOM Genes and comprised the set of candidate target genes for regulatory network analysis. Additional file 2: Table S2 lists the three gene sets.

Summary of primary computational analysis

Imputed gene expression was calculated as in [11]. Briefly, we used the ZINB-WaVE algorithm [62] to obtain a low-dimensional, bias-corrected representation of raw single-cell data, which were then used to construct a distance-based measure of cell similarity and perform imputation with the MAGIC algorithm (version 0.0) [63]. Next, we selected foreskin keratinocytes based on their membership in expression-based clusters previously characterized as keratinocytes in [11]. We identified differentiation stages within this cell population by applying Principal Components Analysis (PCA) followed by k-

means-based approximate spectral clustering [64] (Additional file 3: Identification of keratinocyte stages). To reduce false positives in downstream correlation analysis, we removed outlier cells from the eight keratinocyte stages identified by clustering, reduced MAGIC's imputation time parameter, and re-imputed (Additional file 1: Figures S10-11; Additional file 2: Table S7; Additional file 3: Calculation of gene correlations.).

To construct Figures 1 and 4, Keratinocyte TFs and antioxidant genes were filtered for dynamic expression based on stage-wise log fold-change and clustered using Pearson correlation distance among vectors of log-transformed stage-wise mean imputed counts per million (cpm) (Additional file 3: Identification of keratinocyte stages, Clustering transcription factor expression trajectories and super-enhancer differential motif enrichment, Antioxidant analysis). To prioritize Candidate Keratinocyte TFs for experimental validation, TFs were ranked by the sum of signed log-fold change of their target Keratinocyte Genes during differentiation (positive sign for activation, negative sign for repression). Targets were identified based on strength of TF-gene correlation/anticorrelation (Additional file 3: Prioritization of knockdown targets).

Regulatory analysis for the BK state used Keratinocyte TFs with motifs enriched in BK-specific SEs compared to DK-specific SEs and Keratinocyte Genes not downregulated in the BK state compared to the DK state (Methods: Differential expression). Identification of Gene and TF Modules in the BK state used hierarchical clustering on signed expression similarity scores calculated as soft-thresholded Pearson correlation [65] of log-transformed imputed expression across cells in stages 1-4. We identified regulatory relationships between Gene and TF Modules by considering the distribution of magnitude of mean similarity scores between all TF-Gene module pairs:

where, following the notation of Additional file 3: Regulatory network construction, denotes the signed similarity score of TF and target gene (Additional file 1: Figure S6B). Regulatory relationships were assigned for module pairs exceeding the threshold illustrated in Additional file 1: Figure S6(C,D). Regulatory analysis for the DK state used an analogous method (Additional file 1: Figure S9(B-D)). Further details are given in Additional file 3: Regulatory network construction.

Finally, to measure transcriptional similarity between keratinocyte stages and BCC and SCC, we used previously published bulk RNA-seq expression data and results to identify gene sets differentially expressed in each cancer subtype compared to normal cells [66, 67]. For each cancer subtype and keratinocyte stage, we indexed the genes differentially expressed both in the cancer and in the considered stage (comparing expression in the stage to that of all other stages; Methods: Differential expression). The indexing scheme, described in greater detail in Additional file 3: BCC and SCC similarity analysis, assigned to each gene an integer in the set $\{1, 2, \dots, N\}$, where N was the number of genes differentially expressed in the stage, and gave lower indices to genes that had a larger magnitude of stage-specific differential expression change in the direction matching that in cancer. P-values measuring transcriptional similarity were calculated using the one-sided Mann-Whitney U test comparing the indices of genes for each cancer-type/stage pair to the list of integers ranging from 1 to N .

Differential expression

We used differential expression analysis to identify Keratinocyte Genes specific to individual differentiation stages, as well as those specific to the BK (union of stages 1, 2, 3) and DK (union of stages 5, 6, 7) states. Log (cpm + 1) of non-imputed expression values were calculated for Keratinocyte Genes and for other genes with at least 3 UMI in 20 foreskin keratinocytes. We identified genes with stage-specific expression as those differentially expressed in a one vs. rest test at 5% FDR (Additional file 2: Table S8) using limma-trend version 3.23.9 [68]. We identified genes specific to the BK vs. DK states as those differentially expressed at 5% FDR and with magnitudes of moderated log2 fold-change between the two groups greater than 0.25 (Additional file 2: Table S9).

RNAi knockdown of predicted TFs

ON-TARGETplus siRNA pools targeting *ETV4*, *ZBED2*, *BNC1*, *HOXC11*, *SOX9*, and *IRX4* as well as the ON-TARGETplus Non-targeting Control siRNA #1 were obtained from Dharmacon (Lafayette, CO). Pooled keratinocytes from five different individuals (Additional file 3: Keratinocyte isolation and primary culture) were seeded at a density of 300,000 cells/mL in 12-

well plates. Within 30 minutes of plating, 10 nM siRNA plus 5 uL/well of Hiperfect transfection reagent (Qiagen, Germantown, MD) was added. Transfections were done in quadruplicates. At 48 hours post-transfection, siRNA media was removed and replaced with 1 mL fresh keratinocyte growth media (KGM; Medium 154CF supplemented with 0.07 mM CaCl₂ and Human Keratinocyte Growth Supplement; Life Technologies, Waltham, MA). Five days post transfection, total RNA was extracted using TRIzol reagent (Life Technologies) following the manufacturer's protocol. cDNA was synthesized using the iScript™ cDNA Synthesis Kit (Bio-Rad, Hercules, CA) following the manufacturer's protocol. qPCR was performed with POWER SYBR Green Complete Master Mix (Life Technologies) to measure the expression levels of the housekeeping gene *GUSB*, as well as *ETV4*, *ZBED2*, *BNC1*, *HOXC11*, *SOX9*, *IRX4*, *KRT10*, and *FLG*. Each sample was measured in triplicate on the Applied Biosystems StepOne System. Melting curves were manually inspected to confirm specificity. When applicable, the results are presented as mean+standard deviation. Statistical analysis was conducted using GraphPad Prism v5.0f (La Jolla, CA, USA). Student's t-test was used to compare two separate sets of independent and identically distributed samples with a p-value < 0.05 considered as significant.

Gene ontology analysis:

We used the DAVID gene ontology (GO) resource [69] to determine functional enrichment in BK and DK Gene Modules, as well as in clusters of antioxidant genes with similar dynamic gene expression patterns. For BK and DK Gene Modules, we used the R library RDAVIDWebService [70] to query DAVID with backgrounds composed of members of each gene module and a common control set of 12,516 expressed genes with at least 1 UMI in at least 1 percent of all keratinocytes. Bar plots in Figures 2B and 3B show selected GO terms with Benjamini-Hochberg adjusted p-values less than 0.05. Additional file 2: Table S4 provides the full DAVID output for all Gene Modules identified for the BK and DK states, respectively. GO analysis for clusters of dynamically expressed antioxidant genes used the set of 65 antioxidant with at least 1 UMI in at least 1 percent of all keratinocytes (Additional File 2: Table S2). Due to the small sizes of gene sets and the large number of enrichment tests performed by DAVID, we did not find any significant enrichment after adjusting for multiple hypothesis testing. We therefore reported unadjusted p-values for selected GO terms in Figure 4B; Additional file 2: Table S5 provides the full DAVID output.

Abbreviations

· TF: transcription factor, BCC: basal cell carcinoma, SCC: squamous cell carcinoma, BK: basal keratinocyte, DK: differentiated keratinocyte, SE: super-enhancer, ROS: reactive oxygen species, GO: gene ontology, cpm: counts per million, TSS: transcription start site, CAGE: cap analysis of gene expression

Declarations

Ethics approval and consent

Informed consent for utilization of surgical human skin tissue discards for research use was obtained under protocols approved by the Human Research Protection Program and Institutional Review Board of the University of California, San Francisco (#10-00944).

Consent for publication

Not applicable

Availability of data and materials

Single-cell RNA-seq data are available in the European Genome-phenome Archive repository at accession number EGAS00001002927 (<https://www.ebi.ac.uk/ega/studies/EGAS00001002927>). Bulk RNA sequence data has been deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAS00001003505. Further information about EGA can be found on <https://ega-archive.org>. Relative log expression

normalized expression values for transcription start sites (TSSs) used to identify FANTOM genes and FANTOM TFs are available at http://fantom.gsc.riken.jp/5/datafiles/latest/extra/CAGE_peaks/hg19.cage_peak_phase1and2combined_tpm_ann.osc.txt.gz. Data used to call differential expression of Smoothed inhibitor resistant BCC relative to normal are available at NCBI GEO at accession number GSE58375 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE58375>)

Source code used to generate the results is available at <https://github.com/jssong-lab/kcyteReg>.

Competing interests

The authors declare no competing interests.

Funding

This work was supported in part by funds from NIH R01CA163336 and the Grainger Engineering Breakthroughs Initiative to JSS, the L.S Edelheit Family Biological Physics Fellowship to AF, and NIH K08AR067243 to JBC.

Author Contributions

RJC, JBC, and JSS conceived and supervised the project. AF carried out most of the computational analyses. AL performed the BCC and SCC analysis and wrote the relevant sections. PH and JL performed the validation experiments. AF, RJC, JBC and JSS wrote the manuscript with contributions from other authors. All authors read and approved the final manuscript.

Acknowledgments

Not applicable.

References

1. Alcolea MP, Jones PH: **Lineage analysis of epidermal stem cells.** *Cold Spring Harb Perspect Med* 2014, **4**:a015206.
2. Lopez-Pajares V, Qu K, Zhang J, Webster DE, Barajas BC, Siprashvili Z, Zarnegar BJ, Boxer LD, Rios EJ, Tao S, et al: **A LncRNA-MAF:MAFB transcription factor network regulates epidermal differentiation.** *Dev Cell* 2015, **32**:693-706.
3. Joost S, Zeisel A, Jacob T, Sun X, La Manno G, Lonnerberg P, Linnarsson S, Kasper M: **Single-Cell Transcriptomics Reveals that Differentiation and Spatial Signatures Shape Epidermal and Hair Follicle Heterogeneity.** *Cell Syst* 2016, **3**:221-237 e229.
4. Cavazza A, Miccio A, Romano O, Petiti L, Malagoli Tagliazucchi G, Peano C, Severgnini M, Rizzi E, De Bellis G, Biciato S, Mavilio F: **Dynamic Transcriptional and Epigenetic Regulation of Human Epidermal Keratinocyte Differentiation.** *Stem Cell Reports* 2016, **6**:618-632.
5. Klein RH, Lin Z, Hopkin AS, Gordon W, Tsoi LC, Liang Y, Gudjonsson JE, Andersen B: **GRHL3 binding and enhancers rearrange as epidermal keratinocytes transition between functional states.** *PLoS Genet* 2017, **13**:e1006745.
6. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, Hoke HA, Young RA: **Super-enhancers in the control of cell identity and disease.** *Cell* 2013, **155**:934-947.
7. Rubin AJ, Parker KR, Satpathy AT, Qi Y, Wu B, Ong AJ, Mumbach MR, Ji AL, Kim DS, Cho SW, et al: **Coupled Single-Cell CRISPR Screening and Epigenomic Profiling Reveals Causal Gene Regulatory Networks.** *Cell* 2019, **176**:361-376 e317.
8. Bigarella CL, Liang R, Ghaffari S: **Stem cells and the impact of ROS signaling.** *Development* 2014, **141**:4206-4218.

9. Hamanaka RB, Glasauer A, Hoover P, Yang S, Blatt H, Mullen AR, Getsios S, Gottardi CJ, DeBerardinis RJ, Lavker RM, Chandel NS: **Mitochondrial reactive oxygen species promote epidermal differentiation and hair follicle development.** *Sci Signal* 2013, **6**:ra8.
10. Bhaduri A, Ungewickell A, Boxer LD, Lopez-Pajares V, Zarnegar BJ, Khavari PA: **Network Analysis Identifies Mitochondrial Regulation of Epidermal Differentiation by MPZL3 and FDXR.** *Dev Cell* 2015, **35**:444-457.
11. Cheng JB, Sedgewick AJ, Finnegan AI, Harirchian P, Lee J, Kwon S, Fassett MS, Golovato J, Gray M, Ghadially R, et al: **Transcriptional Programming of Normal and Inflamed Human Epidermis at Single-Cell Resolution.** *Cell Rep* 2018, **25**:871-883.
12. Livet J, Sigrist M, Stroebel S, De Paola V, Price SR, Henderson CE, Jessell TM, Arber S: **ETS gene Pea3 controls the central position and terminal arborization of specific motor neuron pools.** *Neuron* 2002, **35**:877-892.
13. Helmbacher F: **Tissue-specific activities of the Fat1 cadherin cooperate to control neuromuscular morphogenesis.** *PLoS Biol* 2018, **16**:e2004734.
14. Fantom Consortium, Forrest AR, Kawaji H, Rehli M, Baillie JK, de Hoon MJ, Haberle V, Lassmann T, Kulakovskiy IV, Lizio M, et al: **A promoter-level mammalian expression atlas.** *Nature* 2014, **507**:462-470.
15. Monteiro FL, Vitorino R, Wang J, Cardoso H, Laranjeira H, Simoes J, Caldas M, Henrique R, Amado F, Williams C, et al: **The histone H2A isoform Hist2h2ac is a novel regulator of proliferation and epithelial-mesenchymal transition in mammary epithelial and in breast cancer cells.** *Cancer Lett* 2017, **396**:42-52.
16. Ezhkova E, Pasolli HA, Parker JS, Stokes N, Su IH, Hannon G, Tarakhovsky A, Fuchs E: **Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells.** *Cell* 2009, **136**:1122-1135.
17. Sen GL, Reuter JA, Webster DE, Zhu L, Khavari PA: **DNMT1 maintains progenitor function in self-renewing somatic tissue.** *Nature* 2010, **463**:563-567.
18. Zhang X, Tseng H: **Basonuclin-null mutation impairs homeostasis and wound repair in mouse corneal epithelium.** *PLoS One* 2007, **2**:e1087.
19. Shi G, Sohn KC, Li Z, Choi DK, Park YM, Kim JH, Fan YM, Nam YH, Kim S, Im M, et al: **Expression and functional role of Sox9 in human epidermal keratinocytes.** *PLoS One* 2013, **8**:e54355.
20. Lachmann A, Torre D, Keenan AB, Jagodnik KM, Lee HJ, Wang L, Silverstein MC, Ma'ayan A: **Massive mining of publicly available RNA-seq data from human and mouse.** *Nat Commun* 2018, **9**:1366.
21. Carroll DK, Carroll JS, Leong CO, Cheng F, Brown M, Mills AA, Brugge JS, Ellisen LW: **p63 regulates an adhesion programme and cell survival in epithelial cells.** *Nat Cell Biol* 2006, **8**:551-561.
22. Virolle T, Monthouel MN, Djabari Z, Ortonne JP, Meneguzzi G, Aberdam D: **Three activator protein-1-binding sites bound by the Fra-2-JunD complex cooperate for the regulation of murine laminin alpha3A (lama3A) promoter activity by transforming growth factor-beta.** *J Biol Chem* 1998, **273**:17318-17325.
23. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K: **KEGG: new perspectives on genomes, pathways, diseases and drugs.** *Nucleic Acids Res* 2017, **45**:D353-D361.
24. Meng Z, Moroishi T, Guan KL: **Mechanisms of Hippo pathway regulation.** *Genes Dev* 2016, **30**:1-17.
25. Bolognia JL, Schaffer JV, Cerroni L: *Dermatology E-Book.* Elsevier Health Sciences; 2017.

26. Rotty JD, Coulombe PA: **A wound-induced keratin inhibits Src activity during keratinocyte migration and tissue repair.** *J Cell Biol* 2012, **197**:381-389.
27. Barbieri CE, Tang LJ, Brown KA, Pietenpol JA: **Loss of p63 leads to increased cell migration and up-regulation of genes involved in invasion and metastasis.** *Cancer Res* 2006, **66**:7589-7597.
28. Sevilla LM, Latorre V, Sanchis A, Perez P: **Epidermal inactivation of the glucocorticoid receptor triggers skin barrier defects and cutaneous inflammation.** *J Invest Dermatol* 2013, **133**:361-370.
29. Amit I, Citri A, Shay T, Lu Y, Katz M, Zhang F, Tarcic G, Siwak D, Lahad J, Jacob-Hirsch J, et al: **A module of negative feedback regulators defines growth factor signaling.** *Nat Genet* 2007, **39**:503-512.
30. Hildesheim J, Bulavin DV, Anver MR, Alvord WG, Hollander MC, Vardanian L, Fornace AJ, Jr.: **Gadd45a protects against UV irradiation-induced skin tumors, and promotes apoptosis and stress signaling via MAPK and p53.** *Cancer Res* 2002, **62**:7305-7315.
31. Kim DS, Lee HK, Park SH, Lee S, Ryoo IJ, Kim WG, Yoo ID, Na JI, Kwon SB, Park KC: **Terrein inhibits keratinocyte proliferation via ERK inactivation and G2/M cell cycle arrest.** *Exp Dermatol* 2008, **17**:312-317.
32. Gillies TE, Pargett M, Minguet M, Davies AE, Albeck JG: **Linear Integration of ERK Activity Predominates over Persistence Detection in Fra-1 Regulation.** *Cell Syst* 2017, **5**:549-563 e545.
33. Elbediwy A, Vincent-Mistiaen ZI, Spencer-Dene B, Stone RK, Boeing S, Wculek SK, Cordero J, Tan EH, Ridgway R, Brunton VG, et al: **Integrin signalling regulates YAP and TAZ to control skin homeostasis.** *Development* 2016, **143**:1674-1687.
34. Kwan J, Sczaniecka A, Heidary Arash E, Nguyen L, Chen CC, Ratkovic S, Klezovitch O, Attisano L, McNeill H, Emili A, Vasioukhin V: **DLG5 connects cell polarity and Hippo signaling protein networks by linking PAR-1 with MST1/2.** *Genes Dev* 2016, **30**:2696-2709.
35. Duperret EK, Oh SJ, McNeal A, Prouty SM, Ridky TW: **Activating FGFR3 mutations cause mild hyperplasia in human skin, but are insufficient to drive benign or malignant skin tumors.** *Cell Cycle* 2014, **13**:1551-1559.
36. Hilton HN, Stanford PM, Harris J, Oakes SR, Kaplan W, Daly RJ, Ormandy CJ: **KIBRA interacts with discoidin domain receptor 1 to modulate collagen-induced signalling.** *Biochim Biophys Acta* 2008, **1783**:383-393.
37. Lecat A, Di Valentin E, Somja J, Jourdan S, Fillet M, Kufer TA, Habraken Y, Sadzot C, Louis E, Delvenne P, et al: **The c-Jun N-terminal kinase (JNK)-binding protein (JNKBP1) acts as a negative regulator of NOD2 protein signaling by inhibiting its oligomerization process.** *J Biol Chem* 2012, **287**:29213-29226.
38. Hoare S, Hoare K, Reinhard MK, Lee YJ, Oh SP, May WS, Jr.: **Tnk1/Kos1 knockout mice develop spontaneous tumors.** *Cancer Res* 2008, **68**:8723-8732.
39. Shibata S, Tada Y, Asano Y, Hau CS, Kato T, Saeki H, Yamauchi T, Kubota N, Kadowaki T, Sato S: **Adiponectin regulates cutaneous wound healing by promoting keratinocyte proliferation and migration via the ERK signaling pathway.** *J Immunol* 2012, **189**:3231-3241.
40. Sakaguchi M, Sonogawa H, Nukui T, Sakaguchi Y, Miyazaki M, Namba M, Huh NH: **Bifurcated converging pathways for high Ca²⁺- and TGFβ-induced inhibition of growth of normal human keratinocytes.** *Proc Natl Acad Sci U S A* 2005, **102**:13921-13926.
41. Wu N, Rollin J, Masse I, Lamartine J, Gidrol X: **p63 regulates human keratinocyte proliferation via MYC-regulated gene network and differentiation commitment through cell adhesion-related gene network.** *J Biol Chem* 2012, **287**:5627-5638.

42. Houben E, Holleran WM, Yaginuma T, Mao C, Obeid LM, Rogiers V, Takagi Y, Elias PM, Uchida Y: **Differentiation-associated expression of ceramidase isoforms in cultured keratinocytes and epidermis.** *J Lipid Res* 2006, **47**:1063-1070.
43. Caubet C, Jonca N, Brattsand M, Guerrin M, Bernard D, Schmidt R, Egelrud T, Simon M, Serre G: **Degradation of corneodesmosome proteins by two serine proteases of the kallikrein family, SCTE/KLK5/hK5 and SCCE/KLK7/hK7.** *J Invest Dermatol* 2004, **122**:1235-1244.
44. Sun BK, Boxer LD, Ransohoff JD, Siprashvili Z, Qu K, Lopez-Pajares V, Hollmig ST, Khavari PA: **CALML5 is a ZNF750- and TINCR-induced protein that binds stratifin to regulate epidermal differentiation.** *Genes Dev* 2015, **29**:2225-2230.
45. Morikawa M, Koinuma D, Miyazono K, Heldin CH: **Genome-wide mechanisms of Smad binding.** *Oncogene* 2013, **32**:1609-1615.
46. Raz S, Stark M, Assaraf YG: **Binding of a Smad4/Ets-1 complex to a novel intragenic regulatory element in exon12 of FPGS underlies decreased gene expression and antifolate resistance in leukemia.** *Oncotarget* 2014, **5**:9183-9198.
47. Kishibe M, Bando Y, Terayama R, Namikawa K, Takahashi H, Hashimoto Y, Ishida-Yamamoto A, Jiang YP, Mitrovic B, Perez D, et al: **Kallikrein 8 is involved in skin desquamation in cooperation with other kallikreins.** *J Biol Chem* 2007, **282**:5834-5841.
48. Matsui T, Hayashi-Kisumi F, Kinoshita Y, Katahira S, Morita K, Miyachi Y, Ono Y, Imai T, Tanigawa Y, Komiya T, Tsukita S: **Identification of novel keratinocyte-secreted peptides dermokine-alpha/-beta and a new stratified epithelium-secreted protein gene complex on human chromosome 19q13.1.** *Genomics* 2004, **84**:384-397.
49. Tsuchida S, Bonkobara M, McMillan JR, Akiyama M, Yudate T, Aragane Y, Tezuka T, Shimizu H, Cruz PD, Jr., Ariizumi K: **Characterization of Kdap, a protein secreted by keratinocytes.** *J Invest Dermatol* 2004, **122**:1225-1234.
50. Ali RS, Falconer A, Ikram M, Bissett CE, Cerio R, Quinn AG: **Expression of the peptide antibiotics human beta defensin-1 and human beta defensin-2 in normal human skin.** *J Invest Dermatol* 2001, **117**:106-111.
51. Dai J, Brooks Y, Lefort K, Getsios S, Dotto GP: **The retinoid-related orphan receptor RORalpha promotes keratinocyte differentiation via FOXN1.** *PLoS One* 2013, **8**:e70392.
52. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, Ami GOH, Web Presence Working G: **AmiGO: online access to ontology and annotation data.** *Bioinformatics* 2009, **25**:288-289.
53. Velarde MC, Demaria M, Melov S, Campisi J: **Pleiotropic age-dependent effects of mitochondrial dysfunction on epidermal stem cells.** *Proc Natl Acad Sci U S A* 2015, **112**:10407-10412.
54. Sanchez-Danes A, Hannezo E, Larsimont JC, Liagre M, Youssef KK, Simons BD, Blanpain C: **Defining the clonal dynamics leading to mouse skin tumour initiation.** *Nature* 2016, **536**:298-303.
55. Lapouge G, Youssef KK, Vokaer B, Achouri Y, Michaux C, Sotiropoulou PA, Blanpain C: **Identifying the cellular origin of squamous skin tumors.** *Proc Natl Acad Sci U S A* 2011, **108**:7431-7436.
56. Youssef KK, Van Keymeulen A, Lapouge G, Beck B, Michaux C, Achouri Y, Sotiropoulou PA, Blanpain C: **Identification of the cell lineage at the origin of basal cell carcinoma.** *Nat Cell Biol* 2010, **12**:299-305.
57. Chrisostomidis C, Konofaos P, Karypidis D, Lazaris A, Kostakis A, Papadopoulos O: **The impact of Ets-1 oncoprotein and human endoglin (CD105) on the recurrence of non-melanoma skin cancers.** *Int J Dermatol* 2015, **54**:989-995.
58. Li J, Yuan Y, He J, Feng J, Han X, Jing J, Ho TV, Xu J, Chai Y: **Constitutive activation of hedgehog signaling adversely affects epithelial cell fate during palatal fusion.** *Dev Biol* 2018, **441**:191-203.

59. Cai Q, Wang ZQ, Wang SH, Li C, Zhu ZG, Quan ZW, Zhang WJ: **Upregulation of long non-coding RNA LINC00152 by SP1 contributes to gallbladder cancer cell growth and tumor metastasis via PI3K/AKT pathway.** *Am J Transl Res* 2016, **8**:4068-4081.
60. Zhang HM, Liu T, Liu CJ, Song S, Zhang X, Liu W, Jia H, Xue Y, Guo AY: **AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors.** *Nucleic Acids Res* 2015, **43**:D76-81.
61. Watt FM, Estrach S, Ambler CA: **Epidermal Notch signalling: differentiation, cancer and adhesion.** *Curr Opin Cell Biol* 2008, **20**:171-179.
62. Risso D, Perraudeau F, Gribkova S, Dudoit S, Vert JP: **A general and flexible method for signal extraction from single-cell RNA-seq data.** *Nat Commun* 2018, **9**:284.
63. van Dijk D, Sharma R, Nainys J, Yim K, Kathail P, Carr AJ, Burdzyak C, Moon KR, Chaffer CL, Pattabiraman D, et al: **Recovering Gene Interactions from Single-Cell Data Using Data Diffusion.** *Cell* 2018, **174**:716-729 e727.
64. Yan DH, Huang L, Jordan MI: **Fast Approximate Spectral Clustering.** *Kdd-09: 15th Acm Sigkdd Conference on Knowledge Discovery and Data Mining* 2009:907-915.
65. Zhang B, Horvath S: **A general framework for weighted gene co-expression network analysis.** *Stat Appl Genet Mol Biol* 2005, **4**:Article17.
66. Atwood SX, Sarin KY, Whitson RJ, Li JR, Kim G, Rezaee M, Ally MS, Kim J, Yao C, Chang AL, et al: **Smoothed variants explain the majority of drug resistance in basal cell carcinoma.** *Cancer Cell* 2015, **27**:342-353.
67. Cancer Genome Atlas N: **Comprehensive genomic characterization of head and neck squamous cell carcinomas.** *Nature* 2015, **517**:576-582.
68. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers differential expression analyses for RNA-sequencing and microarray studies.** *Nucleic Acids Res* 2015, **43**:e47.
69. Huang da W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.** *Nat Protoc* 2009, **4**:44-57.
70. Fresno C, Fernandez EA: **RDAVIDWebService: a versatile R interface to DAVID.** *Bioinformatics* 2013, **29**:2810-2811.

Additional Files Detail

Additional file 1: Supplementary Figures. Supplementary Figures S1-S11. (PDF)

Additional file 2: Supplementary Tables. Supplementary Tables S1-9. (individual files)

Additional file 3: Supplementary Methods and Supplementary References. Additional description of experimental and computational techniques as well as supplemental references. (PDF)

Figures

Figure 1

Turnover in keratinocyte TF expression is temporally and spatially coupled to turnover in SEs. (A) Imputed single-cell expression vectors of 22,338 foreskin keratinocytes projected onto first two principal components; stage membership was assigned by k-means-based approximate spectral clustering. (B) First seven rows show log-transformed stage-wise mean imputed expression of dynamic keratinocyte TFs normalized across stages. Bottom row shows strength of differential motif enrichment between BK and DK SEs (Methods). Gray and black cells correspond to TFs without a known binding motif and TFs not differentially enriched between SE sets, respectively. Columns are organized by hierarchical clustering on first seven rows (Methods). (C-E) Log fold-change in stage-wise mean imputed expression between stage 4 (mitotic state) and other stages for established keratinocyte epigenetic regulators (C), H2A.Z (D), and a subset of components of SWR1 remodeler complex (E). (F) Expression of KRT10 and FLG transcript following siRNA knockdown of ETV4 or ZBED2, relative to control. Asterisks indicate p < 0.05 (Student's t-test); the error bars indicate one standard deviation. See also Figures S1-5.

Figure 2

BK network analysis identifies gene and TF modules specific to basal functions. (A) Regulation of four GO-enriched Gene Modules by TF Modules, represented as a directed graph. Gene Module nodes show log-transformed stage-wise mean imputed expression normalized across stages 1-7 with shading of one standard deviation interval. TF Modules list their TF constituents. Arrows indicate regulation with width proportional to predicted strength of activation (red) or inhibition (blue). (B) Minus log of adjusted p-values for selected GO terms enriched in each gene module. See also Additional file 1: Figure S6.

Figure 3

DK network analysis identifies gene and TF modules specific to differentiated functions. (A) Regulation of six GO-enriched Gene Modules by TF Modules, represented as a directed graph. Gene Module nodes show log-transformed mean stage-wise imputed expression normalized across stages 1-7 with shading of one standard deviation interval. TF Modules list their TF constituents. Arrows indicate regulation with width proportional to predicted strength of activation (red) or inhibition (blue). (B) Minus log of adjusted p-values for selected GO terms enriched in each gene module. See also Additional file 1: Figure S9.

Figure 4

Peak expression of dynamic antioxidant genes is enriched in the BK state. (A) Log-transformed stage-wise mean imputed expression of dynamic antioxidant genes normalized across stages. Columns are organized by hierarchical clustering (Methods). (B) Minus log of unadjusted p-values (Methods) for selected GO terms enriched in selected gene sets clustered from (A). Asterisks indicates significance at 0.05 threshold.

Figure 5

Differential expression similarity analysis of cancers to single-cell stages. (A) Similarity analysis of BCC to cell stages. Genes differentially expressed in a given cell stage are ranked in order of increasing logFC of expression. Orange and blue curves correspond to smoothed and normalized histograms showing respective probabilities of upregulation and downregulation of genes in BCC with respect to gene rank. Smoothing was performed using a Gaussian kernel with radius 1000. (B) Similar to (A), but comparing SCC to cell stages.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [S2tableKeratinocyteGeneSets.xlsx](#)
- [S8tableDE.One.v.Rest.xlsx](#)
- [FinneganetalAdditionalFile1.pdf](#)

- [FinneganetalAdditionalFile3.pdf](#)
- [S7tablesummedExprFilter.pdf](#)
- [S9tableDE.DKvBK.xlsx](#)
- [S4tableGO.geneModules.xlsx](#)
- [S6tableTFsBCC.SCC.xlsx](#)
- [S3tableBK.DKmodulesmembership.xlsx](#)
- [S5tableGO.antioxidant.xlsx](#)
- [S1tableFANTOM.DE.xlsx](#)