Introduction

According to the World Health Organization, approximately 180,000 deaths annually are attributed to burns, with an additional 173,000 individuals suffering from moderate to severe burns. This issue becomes a critical clinical concern, as the skin, being one of the largest organ systems, serves as the first line of defense for the body. Damage to this protective barrier can lead to severe complications, such as sepsis from bacterial infections, shock due to hypovolemia, and long-lasting mental health issues related to aesthetic concerns. To prevent these complications, it is vital to employ additional materials and substitutes that not only serve as temporary protective barriers but also act as regenerative cues. Traditionally, autologous skin grafts have been the gold standard. However, for patients with extensive and severe burns, the availability of adequate donor sites often poses a significant limitation, prompting increased interest in alternative skin substitutes such as xenografts, synthetics, and allografts. These alternatives offer temporary relief and protection, and there is a burgeoning interest in employing patient-derived iPSCs to culture cells that can be expanded and seeded onto wound healing scaffolds to accelerate recovery.

The skin is composed of three major layers: the hypodermis, dermis, and epidermis. The epidermis, the outermost layer, provides critical protection and is primarily composed of basal keratinocytes. These cells are responsible for the skin's stratified structure and the formation of junctions that give the skin its elasticity and strength. Our laboratory has recently developed a tool called OneSC, which simulates cell fate trajectories through the knockout and overexpression of key transcription factors identified from single-cell RNA sequencing data. Utilizing this tool, we have been able to pinpoint crucial transcription factors and signaling pathways necessary for differentiating cells into the basal keratinocyte lineage. With this information, we propose a novel *in vitro* protocol that can be used to culture and generate basal keratinocytes. This approach holds the potential to significantly improve clinical outcomes for burn victims by facilitating a speedier recovery.

Method/Approach

To elucidate the key driving factors and their associated signaling pathways that determine basal keratinocyte cell fate, we will analyze previously published data on developing mouse embryonic cell states. The first dataset (adEarly_CSCB24.h5ad) covers the embryonic mouse gastrula from embryonic day 8 (E8) to postnatal day 1 (P1), while the second dataset includes time points from E6 to E8 (adLate_Ectoderm.h5ad). Annotations in these datasets help identify cell types and extract identities specific to the keratinocyte lineage. We also used a list of transcriptional factors found in mice (allTFs_mm_aertslab_011924.txt).

We have developed a computational pipeline structured into five sequential stages: epiblast, definitive ectoderm, surface ectoderm, pre-epidermal keratinocytes, and basal keratinocytes. Each stage employs consistent analytical procedures, including quality control, trajectory analysis, differential gene expression analysis, and perturbation analysis. The corresponding code and results are detailed in a series of Jupyter notebooks: Epiblast_to_DefinitiveEctoderm.ipynb, DefinitiveEctoderm_to_SurfaceEctoderm.ipynb, SurfaceEctoderm_to_Pre-EpidermalKeratinocytes.ipynb, and Pre-EpidermalKeratinocytes_to_BasalKeratinocytes_with_Maturation.ipynb.

For quality control, data preprocessing, wrangling, and initial plot generation, we utilized Scanpy with Python. Our standard preprocessing, which includes removing potential doublets by filtering cells with abnormally high gene counts and excluding genes with low counts or variability, ensures the inclusion of crucial genes for stem cell development studies. We further employ UMAP and KNN clustering to ensure distinct cell clustering. Cell lineage trajectories and fates are reconstructed using trajectory inference integrated into scFate. We utilize functions such as *scf.tl.tree* and *scf.pl.graph* to infer the tree graph, set roots with *scf.tl.root*, and calculate pseudotime and trajectory with *scf.tl.pseudotime* and *scf.tl.trajectory* functions. Then we use either bifurcation analysis in scFate or differential gene expression analysis to select the top influential

transcriptional factor. We also select additional marker genes to add together with the transcriptional factors to train the classifier. Our use of OneSC for inferring the genetic regulatory network (GRN) and conducting perturbation analysis for single-cell knock-in or overexpression studies allows us to examine transcription factors' interactions—either activating or repressing with onesc.construct_cluster_graph_adata and onesc.infer_grn. OneSC predicts changes in gene expression and potential shifts in cell type using PySingleCellNet, identifying critical transcription factors that guide cell purity toward the desired state.

For our maturation analysis of basal keratinocytes, we aim to increase the expression of specific mature markers, namely *Krt5*, *Krt10*, and *Krt14*, by adjusting influential transcription factor levels. We have developed Python functions to evaluate and visualize gene expression changes using heatmaps and boxplots (*compare_gene_expression*, *create_heatmap*, *plot_gene_comparison*). We assess the mean expression of three mature marker genes in both wild-type and perturbed trajectory cells, conducting t-tests and calculating p-values to determine if the expression changes are statistically significant. This analysis is documented in the Jupyter Notebook, Pre-EpidermalKeratinocytes_to_BasalKeratinocytes_with_Maturation.ipynb.

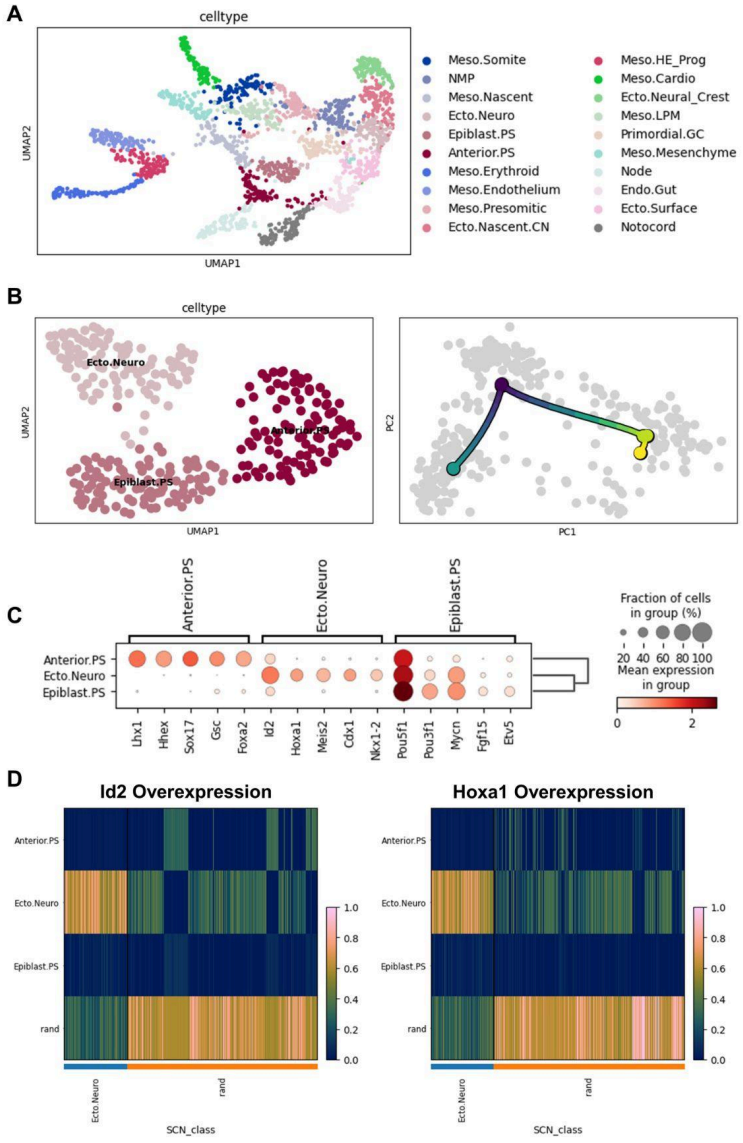


Figure 1. Identification of key TFs driving Definitive ectoderm differentiation
(A) UMAP plot of previously published early mouse gastrulation data; **(B)** Subset of cells to be used for trajectory inference (left panel), Pseudotime trajectory of the subset cells (right panel); **(C)** DotPlot visualizing cell type specific TFs; **(D)** OneSC results following overexpression of *Id2* and *Hoxa1*

Results and Discussion

Employing the detailed computational pipeline, we successfully calculated the outcomes of cell differentiation from the epiblast stage through to basal keratinocytes, including maturation processes. This approach enabled the identification of putative-enriched transcription factors (TFs) within the desired cell states. Additionally, we conducted knock-off analyses to observe the resultant changes in cell states. Crucially, we were able to identify at least two influential transcription factors for each cell transition and verify their effectiveness in driving cells toward the desired state with enhanced purity. These findings underscore the efficacy of our computational pipeline in elucidating the complex dynamics of cell differentiation. In this section, we will delve deeper into the specific roles played by the identified transcription factors, examining how they influence cell behavior and purity across different stages of differentiation. This discussion will not only validate the practical applications of our method but also explore potential avenues for further research and clinical application, particularly in regenerative medicine.

Identification of key transcription factors responsible for epiblast to definitive ectoderm cell fate transition

To commence our analysis, we leveraged previously published data on

embryonic mouse gastrulation, which cataloged a variety of cell types including the key progenitor population, the Epiblast (**Fig. 1A**). From this data, we initially focused on specific cell types of interest for differentiation: the Neuro Ectoderm, the progeny population Epiblast, and an additional population, the Anterior Primitive Streak. This selection enabled us to evaluate the specificity of the identified factors. We employed trajectory inference analysis using scFate to assess the differentiation potential of these populations, successfully confirming that scFate captured the essential biological information **(Fig. 1B)**. Next, we performed DEG analysis on a subset of our dataset specifically targeting the definitive ectoderm population. We also identified specific marker genes for identifying definitive ectoderm: Pax6, Krt8, Krt18, and Otx2. These genes are also subsetted with selected TFs and used to train the classifier and help to better identify perturbation results. This analysis revealed multiple candidate transcription factors highly enriched within this population, including Id2, Hoxa1, Meis2, Cxcl1, and Nkx1-2 (**Fig. 1C**). To investigate the role these TFs play in cell fate differentiation, we simulated overexpression of Id2 and Hoxa1 in the progenitor Epiblast population. Our results indicated that this overexpression significantly increased expression in the definitive ectoderm population, demonstrating a specific effect (**Fig. 1D**).

The outcomes from our analyses provide robust evidence supporting the critical roles of Id2 and Hoxa1 in guiding the differentiation of the Epiblast into the definitive ectoderm. As we delved deeper into the implications of these findings, we explored some biological background related to the selected genes Id2 and Hoxa1 modulation and how they can direct cell fate decisions. Previous literature shows that the Retinoic Acid Signaling (RAS) pathway is able to induce these specific TFs. The RAS pathway depends on retinoic acid which serves as the initial signaling molecule to bind to the retinoic acid receptors. The activation from the building to the receptor will interact with retinoic acid response elements on the DNA to regulate the Hoxa1 gene transcription. Therefore, we can increase the Hoxa1 expression by adding more retinoic acid to the cells. For Id2, bone morphogenetic proteins (BMPs) are important regulators that can upregulate Id2 expression by activating the Smad-dependent signaling pathway, which leads to increased transcription of the Id2 gene.

Identification of key transcription factors responsible for definitive ectoderm to surface ectoderm cell fate transition

We next decided to focus on the cell fate transition from the definitive ectoderm to the surface ectoderm. In this particular case, we took a similar approach and subset our cell types such that we had the Definitive Ectoderm (progenitor) the Surface Ectoderm (progeny) and the Neural Crest (extra population) as seen on the left panel of **Fig. 2A**. Our Pseudotime analysis was able to successfully identify the lineage relationship and predict the trajectory of the Definitive Ectoderm to the Surface Ectoderm as the final state. This included a bifurcation population

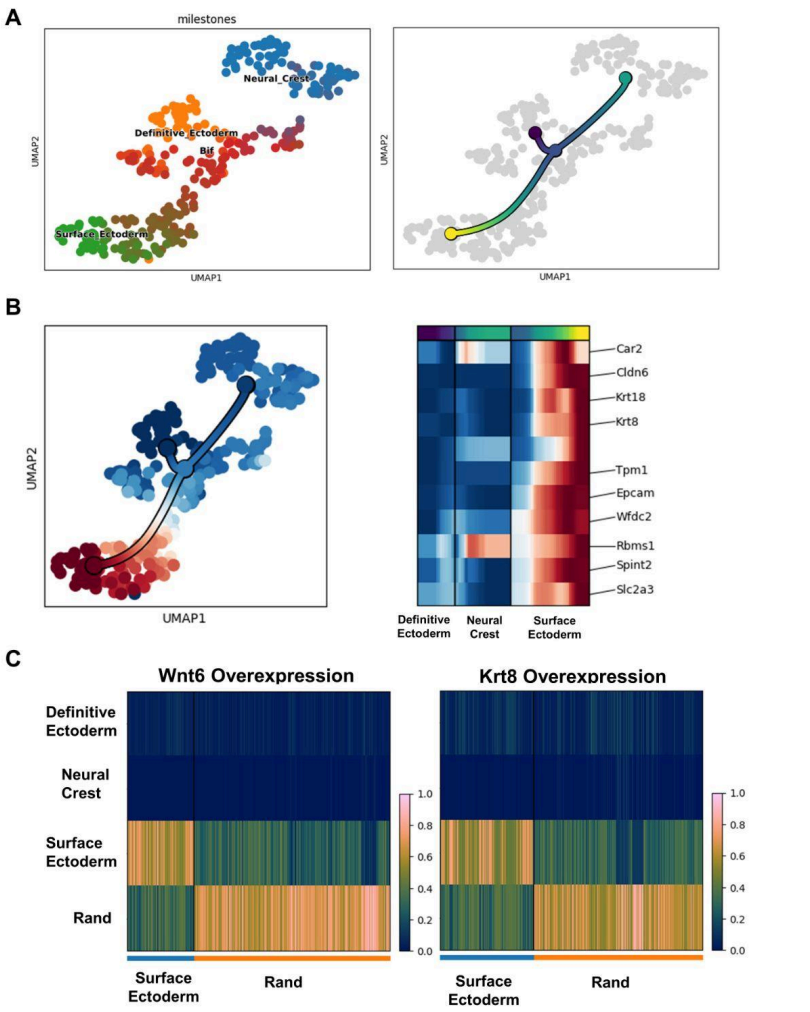


Figure 2. Identification of key TFs driving surface ectoderm differentiation
(A) UMAP subset cells (left panel), trajectory inference analysis results (right panel); **(B)** Bifurcation analysis trajectory (left panel) and identified progeny specific TFs (right panel) **(C)** OneSC results following following overexpression of *Wnt6* and *Krt8*

which can additionally be noted in the right panel of **Fig. 2A.** Using the bifurcation analysis in scFate, we identified key TFs that were found to be enriched in the progeny population and we also used surface ectoderm marker Tp63, and CRB3 to train classifer. Of interest, we identified Wnt6 and Krt8 to be enriched (**Fig. 2B**). In order to verify that these specific TFs were responsible for driving the differentiation of the surface ectoderm fate we used OneSC to overexpress these factors. From our analysis, we identified that this indeed resulted in the branching of cell fate into the surface ectoderm population, although this included a high proportion of random cells as well. Taken together, our results demonstrate that we were able to identify key factors in driving the definitive to surface ectoderm fate.

Wnt6 is one of the ligands within the Wnt family that binds to Frizzled receptors on the cell surface to activate Wnt signaling pathways. The Wnt signaling pathway plays a critical role in cell fate determination, proliferation, and epidermal patterning. Wnt6 has been implicated in various developmental processes. For instance, in the context of skin, Wnt6 has been shown to be able to promote the differentiation of certain cell types in the ectoderm. Biologically, if we want to overexpress Wnt6 through the Wnt signaling pathway, we can simply add Wnt6, which, according to our simulation, should promote differentiation into surface ectoderm. Krt8 is an intermediate filament protein of the keratin family. Though Krt8 is not typically involved in differentiation of basal keratinocytes, other keratins such as Krt5 and Krt14 are found to be crucial for the maintenance and function of basal keratinocytes. Overexpression of Krt8 can be realized through the TGF-beta pathway, which is known to regulate the expression of various keratins, including Krt8. Enhancing TGF-beta signaling through the addition of TGF-beta ligands or by activating its receptors can potentially upregulate Krt8.

Identification of key transcription factors responsible for surface ectoderm to pre-epidermal keratinocyte cell fate transition

For this differentiation step, we constructed the cell fate trajectory of surface ectoderm differentiating into pre-epidermal keratinocytes. Since the late embryonic data does not include additional cell types differentiated from surface ectoderm on the developmental tree other than the target cell type. To do this, we aimed to construct a linear trajectory of the surface ectoderm differentiating into pre-epidermal keratinocytes. We combined data from both the early (**Fig. 1A**) and late (**Fig. 3A**) embryonic data to form a new dataset (**Fig. 3B**). Pre-processing steps were then performed on the new dataset, involving normalization, filtering for highly variable genes, PCA, and clustering, and two clusters were identified (**Fig 3. B**). Tree analysis was performed on this dataset, and a root was set. Pseudotime was then assigned to each cell and cell fate trajectory was plotted (**Fig. 3B**). Differential gene analysis identified 10 significant transcription factors that were suggested to

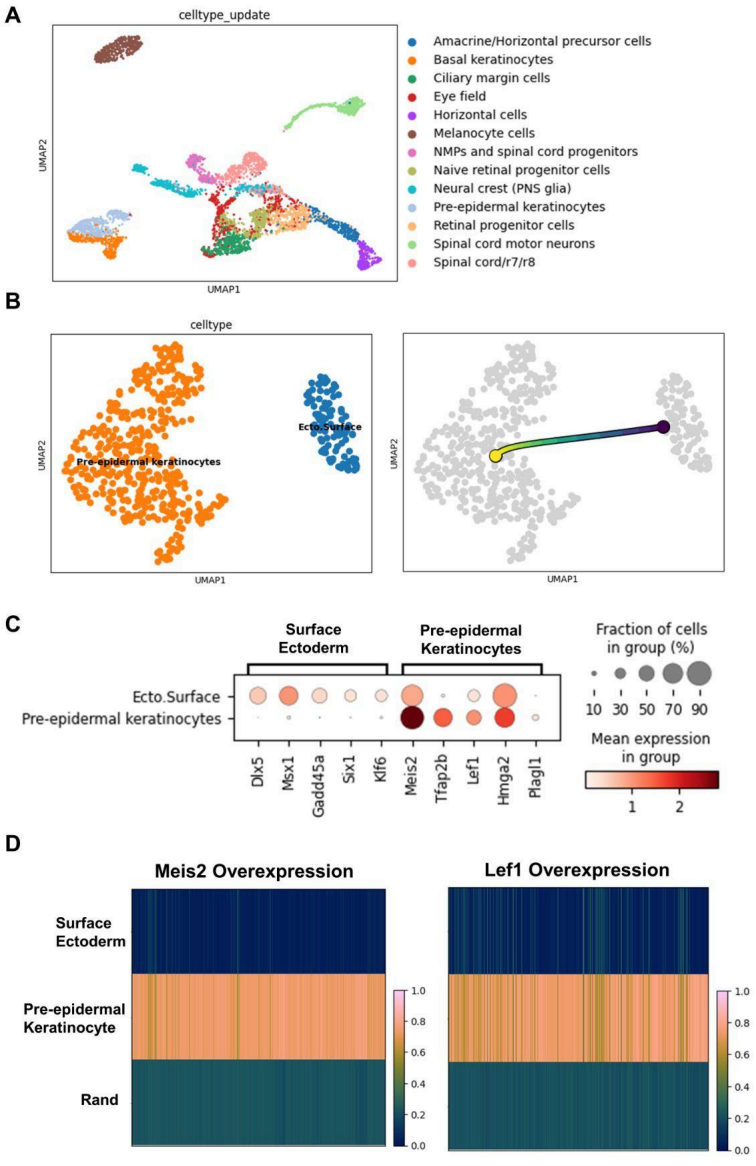


Figure 3. Identification of key TFs driving pre-epidermal keratinocyte differentiation **(A)** UMAP of late stage mouse gastrulation data set **(B)** Subset cells (left panel), trajectory inference analysis results (right panel); **(C)** DotPlot depicting DEG of TFs **(D)** OneSC results following following overexpression of *Meis2* and *Lef1*

contribute to the differentiation process (**Fig. 3C**). After extracting cells with these transcription factors from the dataset, the dataset was subsampled to include the same total count of cells for each cell type to reach a balance. We trained a classifier using pySCN to classify the target cell types with selected markers Krt1 and Krt10 for pre-epidermal keratinocytes, and inferred a GRN graph, showing how the selected transcription factors interact with each other. Finally, we simulated perturbations on this dataset, including over-expressing Meis2 and Lef1, both of which resulted in a purer population of pre-epidermal keratinocytes (**Fig 3. D**), validating the use of these two genes as markers for differentiation of surface ectoderm to pre-epidermal keratinocytes.

Meis2 is a transcriptional factor that often functions in conjunction with Hox proteins to regulate gene expression. The primary function of the Hox pathway is to provide a positional identity along the body. Though not directly related, the Hox genes can be mediated through their interactions with the Wnt pathway, which is crucial in the early stages of ectodermal patterning. Lef1 is a binding factor that is part of the Wnt signaling pathway. To upregulate Meis2 and Lef1 in a biological setting, such as in a cell culture or in an organism, we can activate the Wnt pathway to increase Lef1 expression and activate the Hox pathway to increase Meis2 expression, thus leading to differentiation of pre-epidermal keratinocytes.

Identification of key transcription factors responsible for pre-epidermal keratinocyte to basal keratinocyte cell fate transition

We next focused our analysis on the final stage of cell fate determination, specifically targeting basal keratinocytes. To achieve this, we isolated subsets of pre-epidermal and basal keratinocyte cells (**Fig. 4A**) and conducted trajectory inference using scFate (**Fig. 4B**). As anticipated, scFate successfully predicted the developmental trajectory from pre-epidermal to basal keratinocytes. To identify key TFs, we performed scFate bifurcation and DEG analysis, similar to previous stages, but limited our dataset to include only TFs detected in the single-cell data (**Fig. 4C**), such as Nfib, Bnc2, Runx1, Sox5, and Bach2. To determine the role of these TFs in driving cell fate decisions, we overexpressed selected TFs using the OneSC method. Notably, overexpression of Nfib and Bnc2 led to a nearly pure population of basal keratinocyte cells, indicating their crucial role in the commitment to pre-epidermal keratinocyte fate. Overall, our findings suggest that we have identified essential factors for the transition from pre-epidermal to basal keratinocyte cell fate.

We found that Nfib and Bnc2 guide the differentiation of pre-epidermal keratinocytes to basal keratinocytes. The TGF-β (Transforming Growth Factor-beta) signaling pathway plays a pivotal role in regulating various cellular

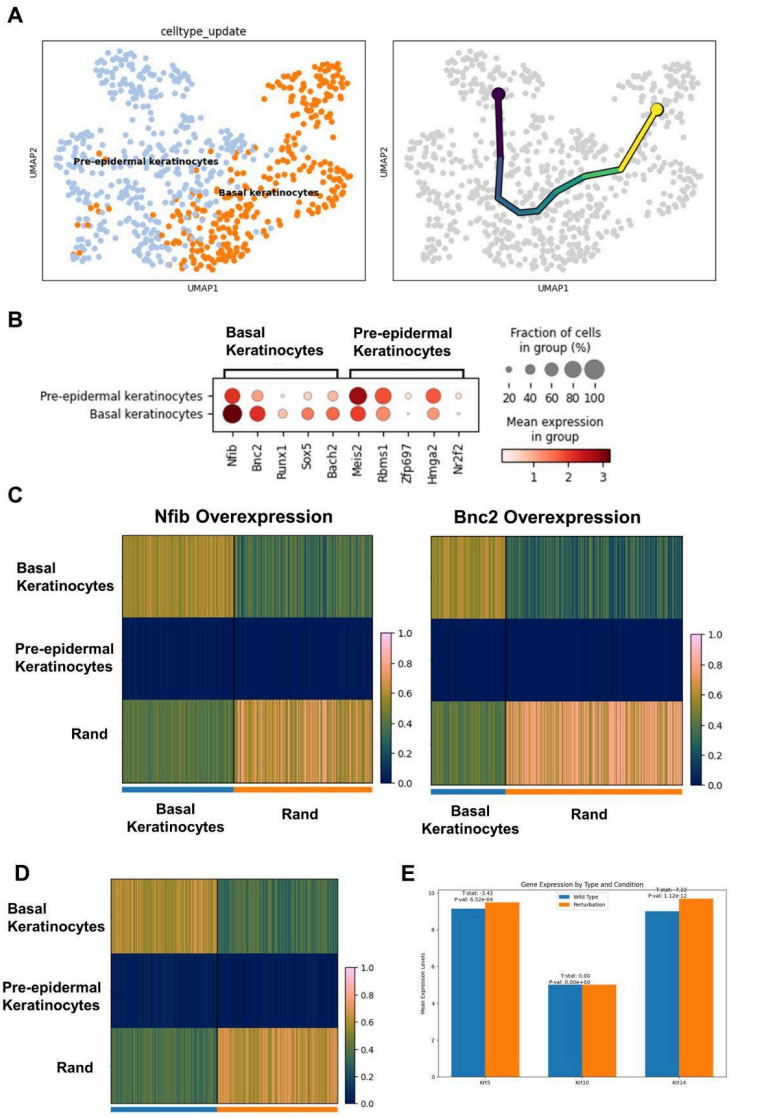


Figure 4. Identification of key TFs driving basal keratinocyte differentiation and maturation
(A) Subset cells (left panel), trajectory inference analysis results (right panel); **(B)** DotPlot depicting DEG of TFs **(C)** OneSC results following following overexpression of *Meis2* and *Lef1*; **(D)** OneSC results of *Sox5* KO; **(E)** Barplots depicting average gene expression of *Krt5*, *Krt10*, *Krt14* following *Sox5* KO

processes including cell growth, differentiation, and apoptosis. This pathway influences the expression of key genes such as Nfib and Bnc2, which are crucial in a range of developmental and physiological functions. The activation of TGF-β signaling begins when TGF-β binds to its specific receptors, leading to the phosphorylation of receptor-regulated Smads (Smad2 and Smad3). These Smads subsequently form a complex with Smad4, and this complex then translocates into the nucleus to modulate gene expression. Specifically, this pathway has been shown to boost the transcription of Nfib, particularly in contexts such as lung development and fibrosis. While the direct regulation of Bnc2 by TGF-β is less well-documented, it is likely that TGF-β could similarly influence Bnc2 expression, considering its integral role in cell differentiation. To experimentally overexpress both Nfib and Bnc1 by enhancing the TGF-β pathway, one effective method is the exogenous addition of TGF-β to cell cultures. This approach stimulates the pathway, potentially increasing the transcription of these target genes.

The factor responsible for driving basal keratinocyte maturation

Previous work has demonstrated that although one may be able to differentiate a certain progenitor into its progeny this unfortunately does not result in a completely mature cell. Suggesting there are additional factors that are necessary to drive the cell into complete cell differentiation. In order to address this problem within our protocol, we began by identifying typical mature markers found in basal keratinocytes. Upon identification of these factors, we attempted to perturb the cellular outcomes using OneSC in an effort to increase mature marker gene expression. According to the literature, Krt5, Krt10, and Krt14 are commonly considered markers associated with an indication that these cells have reached maturity. In order to assess this, using our developed function (see Methods), we found that knockout of Runx1 and Sox5 (**Fig. 4D**) resulted in statistically significant increases in gene expression of these specific mature markers (**Fig. 4E**). These results suggest that we are able to computationally predict and identify the downregulation of specific TFs that may aid in maturation in vitro.

To downregulate Runx1 and Sox5 via the Wnt/β-catenin pathway, one could manipulate this pathway's activity using inhibitors or genetic interventions. Using small molecule inhibitors that prevent β-catenin stabilization, such as those targeting the destruction of complex components (e.g., GSK-3β inhibitors), can reduce β-catenin levels in the nucleus, thereby potentially decreasing the transcriptional activation of Runx1 and Sox5 indirectly. Additionally, siRNA or shRNA targeting β-catenin mRNA can be employed to knock down β-catenin expression directly, thus diminishing its influence on these transcription factors. These strategies can modulate the downstream effects of the Wnt/β-catenin pathway on cell fate and differentiation, particularly in systems where Runx1 and Sox5 play crucial roles.

Conclusion

The effort of attempting to generate specific populations of differentiated cells from iPSCs has proven challenging. Although previous work, in the literature allows us to gain insights into the complex understanding of cell fate and differentiation, this ultimately, does not guarantee that a cell will continue down a specific trajectory. Given the increasing need for iPSC protocols, especially with goals of expanding a patient's own cells for an autologous treatment such as following a severe burn. This has made the interest in assessing the effects of knocking out or overexpressing certain TFs a dream as this would significantly reduce time and cost in developing differentiation protocols. Here we demonstrate the ability to not only identify the entire cell fate trajectory of the basal keratinocytes population but additionally, identify key TFs that aid in the progression of this lineage. According to our analysis and results using OneSC, we identified that *Id2* and *Hoxa1* were found to be critical factors guiding differentiation from *Epiblast* to *Definitive Ectoderm* cells. We found that *Wnt6* and *Krt8* were critical factors for the *Surface Ectoderm* cell fate. Subsequently, *Meis2* and *Lef1* are suggested to be the next essential TFs for *Pre-epidermal keratinocyte* differentiation. Lastly, we identified that *Nfib* and *Bnc2* were found to be necessary for final *Basal keratinocyte* differentiation. Taken together, these results suggest

multiple factors which should be targeted in order to induce differentiation of these progenitor cells down this specific cell path. This is of course with the caveat that this entire analysis has been conducted *in silico*. Because of this, it is suggested to conduct experimental work which can confirm the identified results. We suggest the necessity of performing *in vitro* cell culture with each of the given cell progenitors. Using an adenovirus vector, it would be possible to overexpress the identified TFs in order to assess their ability in driving cell fate differentiation. If the results hold, this would validate our identified results and provide additional validity to OneSC.

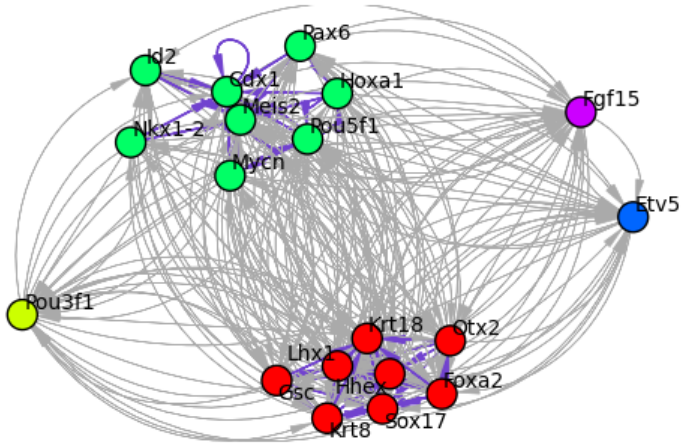
The ability to guide differentiation into a particular cell fate is a much-desired outcome, unfortunately, however, this does not guarantee complete functioning of the cell. Previous work has demonstrated that this is due to the immaturity of the *in vitro* differentiated cell, suggesting that there are additional cues or factors that drive the cell into complete maturation. In order to attempt to identify this computationally, we developed a relatively simple approach by performing perturbations on the basal keratinocytes and measuring the changes in gene expression of key maturity markers. Based on our analysis, we were able to identify two key factors which upon knockout resulted in the increased expression of keratins. Suggesting a fruitful approach to drive further maturation via *in silico* methods. In order to validate our findings, it would be essential to perform a gene knockout *in vitro*. Here we would use an adenovirus in order to knock out *Sox5* or *Runx1* or the combination of both factors. This would be followed by assessing the changes in gene expression using qRT-PCR. This would provide us with enough information to validate our findings.

In addition, another key limitation with our approach was the use of a single modality dataset. Here we used two scRNAseq datasets of the developing mouse gastrula. Although it has been shown that scRNAseq can provide great insight into cellular heterogeneity and complexity, the biggest limitation is that because mRNA is expressed it does not completely guarantee that this will result in protein translation. If a second modality dataset was available (ex. ATAC-seq, or Proteomics data) this would have allowed us to measure the expression of the key genes or assess the accessibility of some of the specific binding regions for these TFs. This would allow us to obtain an additional metric in filtering down the identified TFs within our analysis.

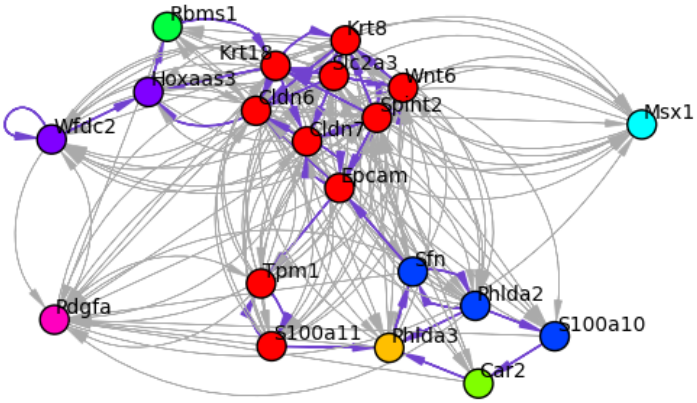
Taken together, we have developed a sound approach to the identification and guided differentiation of a pluripotent population (epiblast) to a desired cell fate (basal keratinocyte). In our analysis, we identified multiple key TFs which are shown to be critical in altering the trajectory of progenitor cells. By validating and beginning to test our identified factors. If our approach holds true, it can be applied to focus on expanding additional cell fates and saving time and money. This in turn will avoid costly trial-and-error approaches that are traditionally taken at the bench. Ultimately, allowing for increased turnaround time of translation work into the clinic and providing our patients with the cells and tissues they need.

Appendix

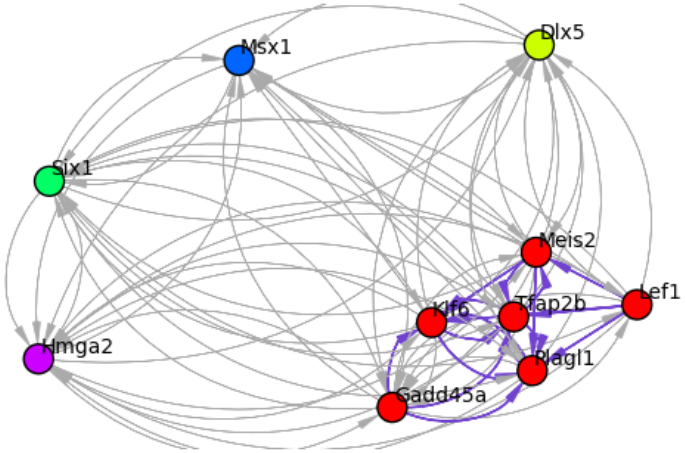
GRN for epiblast to definitive ectoderm



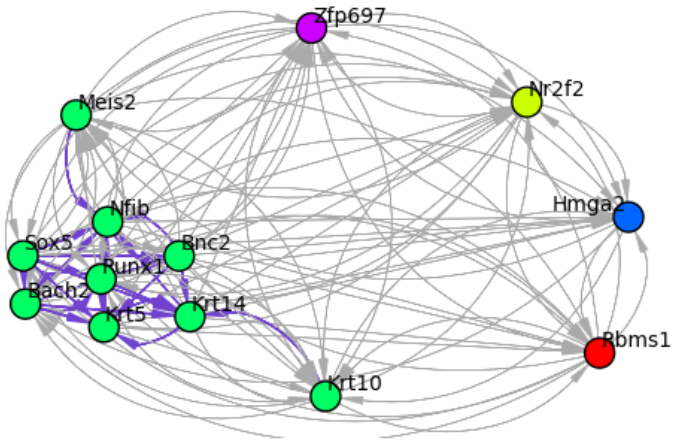
GRN for definitive ectoderm to surface ectoderm



GRN for surface ectoderm to pre-epidermal keratinocytes



GRN for pre-epidermal keratinocytes to basal keratinocytes



References

1. World Health Organization: WHO. (2023, October 13). Burns. https://www.who.int/news-room/fact-sheets/detail/burns

2. Mathes, S. H., Ruffner, H. & Graf-Hausner, U. The use of skin models in drug development. *Advanced Drug Delivery Reviews* vols 69–70 81–102 Preprint at https://doi.org/10.1016/j.addr.2013.12.006 (2014).

3. Shpichka, A. *et al.* Skin tissue regeneration for burn injury. *Stem Cell Research and Therapy* vol. 10 Preprint at https://doi.org/10.1186/s13287-019-1203-3 (2019).

4. Límová, M. Active Wound Coverings: Bioengineered Skin and Dermal Substitutes. *Surgical Clinics of North America* vol. 90 1237–1255 Preprint at https://doi.org/10.1016/j.suc.2010.08.004 (2010).

5. Strunz M, Simon LM, Ansari M, Kathiriya JJ, Angelidis I, Mayr CH, Tsidiridis G, Lange M, Mattner LF, Yee M, Ogar P, Sengupta A, Kukhtevich I, Schneider R, Zhao Z, Voss C, Stoeger T, Neumann JHL, Hilgendorff A, Behr J, O'Reilly M, Lehmann M, Burgstaller G, Königshoff M, Chapman HA, Theis FJ, Schiller HB. Alveolar regeneration through a Krt8+ transitional stem cell state that persists in human lung fibrosis. Nat Commun. 2020 Jul 16;11(1):3559. doi: 10.1038/s41467-020-17358-3. PMID: 32678092; PMCID: PMC7366678.