Stem Cell Reports
Article

ISSCR

OPEN ACCESS

Deciphering the heterogeneity of differentiating hPSC-derived corneal limbal stem cells through single-cell RNA sequencing

Meri Vattulainen,[1,4] Jos G.A. Smits,[2,3,4] Julian A. Arts,[2] Dulce Lima Cunha,[2] Tanja Ilmarinen,[1] Heli Skottman,[1,5,*] and Huiqing Zhou[2,3,5,6,*]

[1]Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland
[2]Department of Molecular Developmental Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University, Nijmegen, the Netherlands
[3]Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands
[4]These authors contributed equally
[5]These authors contributed equally
[6]Lead contact
*Correspondence: heli.skottman@tuni.fi (H.S.), j.zhou@science.ru.nl (H.Z.)
https://doi.org/10.1016/j.stemcr.2024.06.001

SUMMARY

A comprehensive understanding of the human pluripotent stem cell (hPSC) differentiation process stands as a prerequisite for the development of hPSC-based therapeutics. In this study, single-cell RNA sequencing (scRNA-seq) was performed to decipher the heterogeneity during differentiation of three hPSC lines toward corneal limbal stem cells (LSCs). The scRNA-seq data revealed nine clusters encompassing the entire differentiation process, among which five followed the anticipated differentiation path of LSCs. The remaining four clusters were previously undescribed cell states that were annotated as either mesodermal-like or undifferentiated subpopulations, and their prevalence was hPSC line dependent. Distinct cluster-specific marker genes identified in this study were confirmed by immunofluorescence analysis and employed to purify hPSC-derived LSCs, which effectively minimized the variation in the line-dependent differentiation efficiency. In summary, scRNA-seq offered molecular insights into the heterogeneity of hPSC-LSC differentiation, allowing a data-driven strategy for consistent and robust generation of LSCs, essential for future advancement toward clinical translation.

INTRODUCTION

Corneal limbal stem cell (LSC)-driven repair mechanisms and homeostatic renewal of corneal epithelium (CE) are essential processes in the maintenance of human ocular surface health (Deng et al., 2019). Guiding human pluripotent stem cells (hPSCs) toward specific cell types such as LSCs possesses significant promise for treating patients needing LSC transplants. Major advantages of hPSCs as the LSC source include abundance and donor-independent "off-the-shelf" nature, and, to this day, multiple CE/LSC differentiation methods for hPSCs have been introduced (Mahmood et al., 2022). Aiming to facilitate clinical translation, protocols have shifted toward simplified feeder- and serum-free versions such as our own (Hongisto et al., 2017; Mikhailova et al., 2014).

Despite constant efforts, certain universal challenges continue to impede the effective implementation of hPSC-based therapeutics. High variation between individual hPSC lines (Bock et al., 2011; Osafune et al., 2008) and clones (D'Antonio et al., 2018; Hayashi et al., 2016) are well-known problems for hPSC differentiation, which can render promising protocols infeasible for widespread use. Additionally, residual cellular heterogeneity within hPSC-derived grafts impairs safety and efficacy (Sato et al., 2019). Although hints of the efficiency variation have been provided in the previous literature (Hayashi et al., 2016; Vattulainen et al., 2021), heterogeneity in the differentiation of hPSCs into LSCs has not been systematically addressed.

Single-cell RNA sequencing (scRNA-seq) allows deciphering the heterogeneity of complex cellular systems and can serve as a powerful tool for assessing the efficiency of hPSC differentiation protocols (Cuomo et al., 2020). In recent years, the use of scRNA-seq has significantly expanded our knowledge about cell types in the cornea *in vivo* (Arts et al., 2023), but it is not yet widely applied for hPSC differentiation to LSCs. Through years of dedicated work on hPSC-LSC differentiation in our laboratory, we have obtained promising results using traditional characterization methods such as immunofluorescence (IF) staining (Hongisto et al., 2017; Vattulainen et al., 2019; 2021). Nevertheless, characterization based on a handful of markers is insufficient and provides no solutions for addressing the cell line-dependent variation. In this study, we used scRNA-seq to explore the heterogeneity of cell populations during hPSC-LSC differentiation and attempted to find feasible solutions for more consistent and robust derivation of hPSC-LSCs, using three distinct cell lines.

RESULTS

scRNA-seq recapitulates the temporal marker gene expression patterns during the corneal differentiation of hPSCs

One human embryonic stem cell (hESC) line Regea08/017 (hereafter, "hESC") and two human induced pluripotent

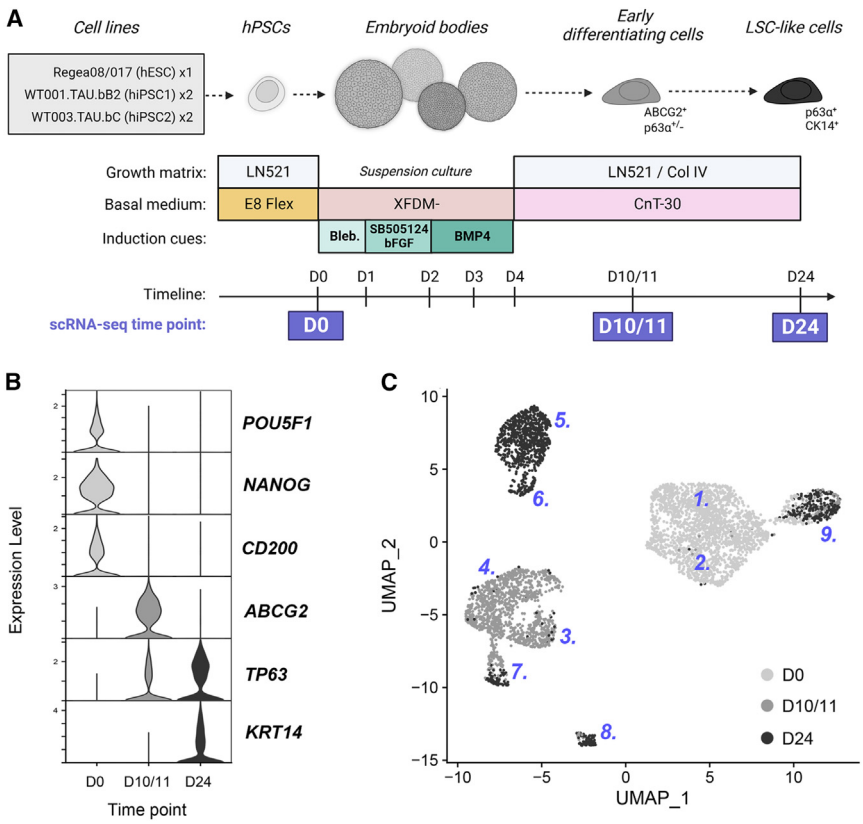

Figure 1. Single-cell RNA sequencing (scRNA-seq) of three human pluripotent stem cell (hPSC) lines differentiated toward corneal limbal stem cells (LSCs)

(A) Schematic overview of the differentiation process and scRNA-seq sample collection time points. Illustration created with Biorender.com.

(B) Expression of selected pluripotency- and LSC-associated marker genes at the studied time points.

(C) UMAP dimensionality reduction graph of the scRNA-seq data.

stem cell (hiPSC) lines WT001.TAU.bB2 (hereafter, "hiPSC1") and WT003.TAU.bC (hereafter, "hiPSC2") were used for differentiation toward LSCs (Figure 1A). The inconsistency between the cell lines was evident on morphological level: by day (D)24, the majority of hESC cultures consisted of epithelial monolayers with cuboid cell shape, whereas hiPSC1 and hiPSC2 yielded heterogeneous cultures with atypical PSC-like colonial growth (Figure S1).

To define the precise cell states within the heterogeneous populations, we carried out scRNA-seq for one hESC, two hiPSC1, and two hiPSC2 replicas. Samples were collected from undifferentiated cells at D0 and from two previously identified key time points at D10/11 and D24 (Vattulainen et al., 2019). After quality control and filtering of raw scRNA-seq data, 4,840 high-quality cells were acquired for downstream analysis. The scRNA-seq data confirmed the expected overall expression patterns for well-known marker genes over time (Figure 1B): decreased expression of pluripotency markers after D0, expression of previously detected LSC marker *ABCG2* mainly at D10/11, and strong induction of LSC/progenitor marker *TP63* and *KRT14* expression at D24. Based on the gene expression profiles, nine separate clusters were identified and visualized in high-dimensional space using uniform manifold approximations and projections (UMAPs) (Figure 1C).

scRNA-seq elaborates the heterogeneity during the hPSC-LSC differentiation

The nine clusters were characterized in detail in terms of the contributing time points, cell lines, and their associated cell-cycle phases (Figures 2A and 2B; Table S1). The main difference between clusters 1 and 2 collected at D0 arose from cell cycle and subtle cell line-specific variations, and they were merged for downstream analyses (C1–2). Cluster 3 (C3) and cluster 4 (C4) contained cells mostly collected at D10/11, with a few cells originating from D24, while cluster 5 (C5) and cluster 6 (C6) cells were collected solely at D24. Due to their association with certain collection days, we defined these clusters as "time point-specific." Among the remaining three clusters, cluster 7 (C7) and cluster 8 (C8) contained a mixture of cells from both D10/11 and D24, while cluster 9 (C9) comprised cells from all three time points. We defined these as "mixed time point" clusters. C9 contained actively cycling cells at comparable levels with C1–2 cells. (Figure 2B, and Table S1)

Importantly, the studied cell lines showed varying distributions to the identified clusters. Before the start of differentiation, C1–2 contained a similar number of cells from all three lines. In D10/11-associated clusters, hESC was slightly underrepresented in C3 and slightly overrepresented in C4. In D24-associated clusters, C5 contained a

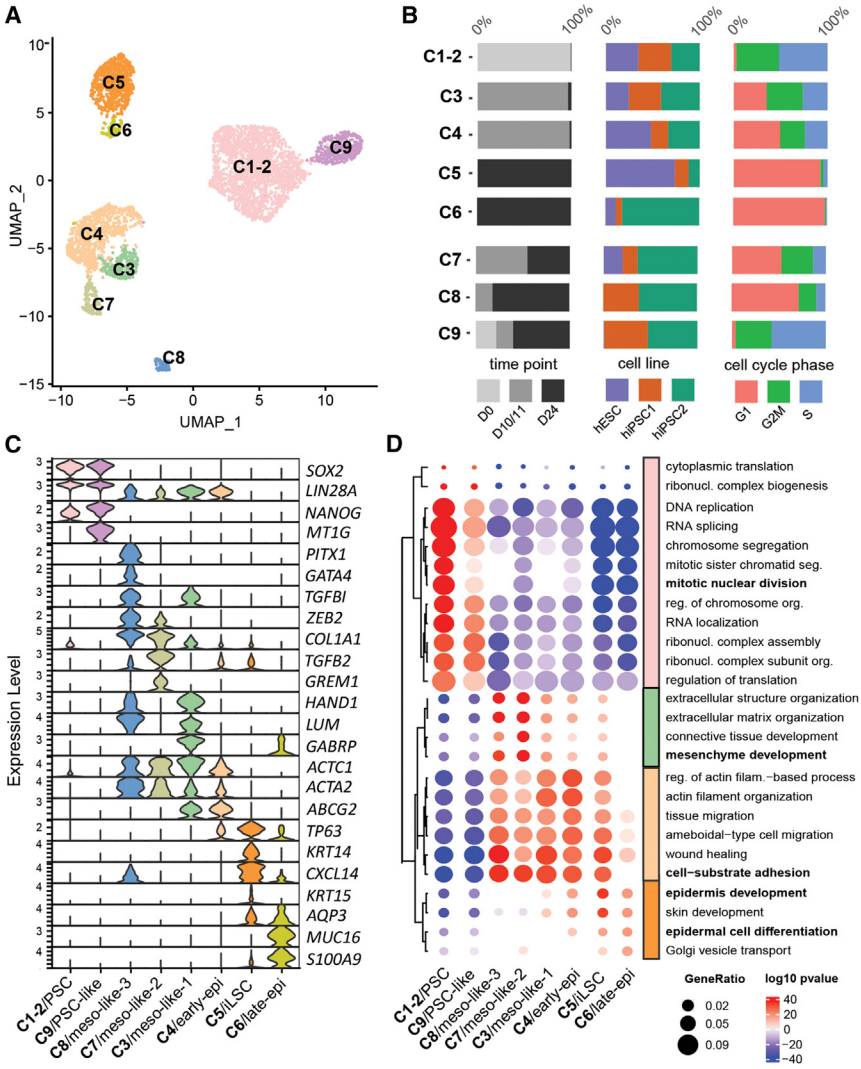

Figure 2. Single-cell RNA cluster characteristics and annotation

(A) UMAP with cluster annotations.

(B) Distribution of cells in clusters per time point, cell line, and cell cycle phase.

(C) Expression of selected marker genes across the clusters.

(D) Gene Ontology (GO) enrichment analysis based on positive cluster markers (red-pvalue) and negative cluster markers (blue-pvalue). GO terms are kmean clustered into 4 clusters.

majority of hESC cells and rather small fractions from the two hiPSC lines, while C6 was mainly derived from hiPSC2. Interestingly, C7 was the only mixed time point cluster containing a small fraction of hESC cells, whereas both C8 and C9 were solely from hiPSC1 and hiPSC2. (Figure 2B, and Table S1)

Time point-specific clusters represent cell states along the LSC differentiation trajectory

Next, we set out to identify the precise cell state in each cluster, using a combination of marker gene expression and Gene Ontology (GO) analysis (Gene Ontology Consortium, 2004). The protein-level expression of selected markers was subsequently confirmed by IF.

Many C1–2 marker genes were associated with pluripotency, including *SOX2, LIN28A*, and *NANOG* (Figures 2C, and S2) and enriched for proliferation-associated GO terms "mitotic nuclear division" and "DNA replication" (Figure 2D). IF of undifferentiated hPSCs at D0 showed uniform expression of pluripotency-associated cluster markers TRA-1-81 (recognizes an epitope expressed by PODXL), OCT3/4, and LIN28 (Figures 3A, and S2), confirming the annotation of this cluster as "PSC."

Both C3 and C4 had high expression of the stem cell marker *ABCG2* (Figures 2C, and S2). Other highly expressed marker genes for C3 included cell adhesion and migration genes such as *LUM*, *TGFBI*, and *GABRP*, and the overall marker genes were slightly enriched for "mesenchyme development" (Figure 2D), indicating mesodermal characteristics. Therefore, C3 was referred to as the "meso-like-1" cluster. C4 exhibited a slightly more epithelial-like profile, with low levels of epithelial stem cell genes such as *TP63* (Figures 2C, and S2) that gave rise to a slight enrichment for "epidermis development" in GO analysis

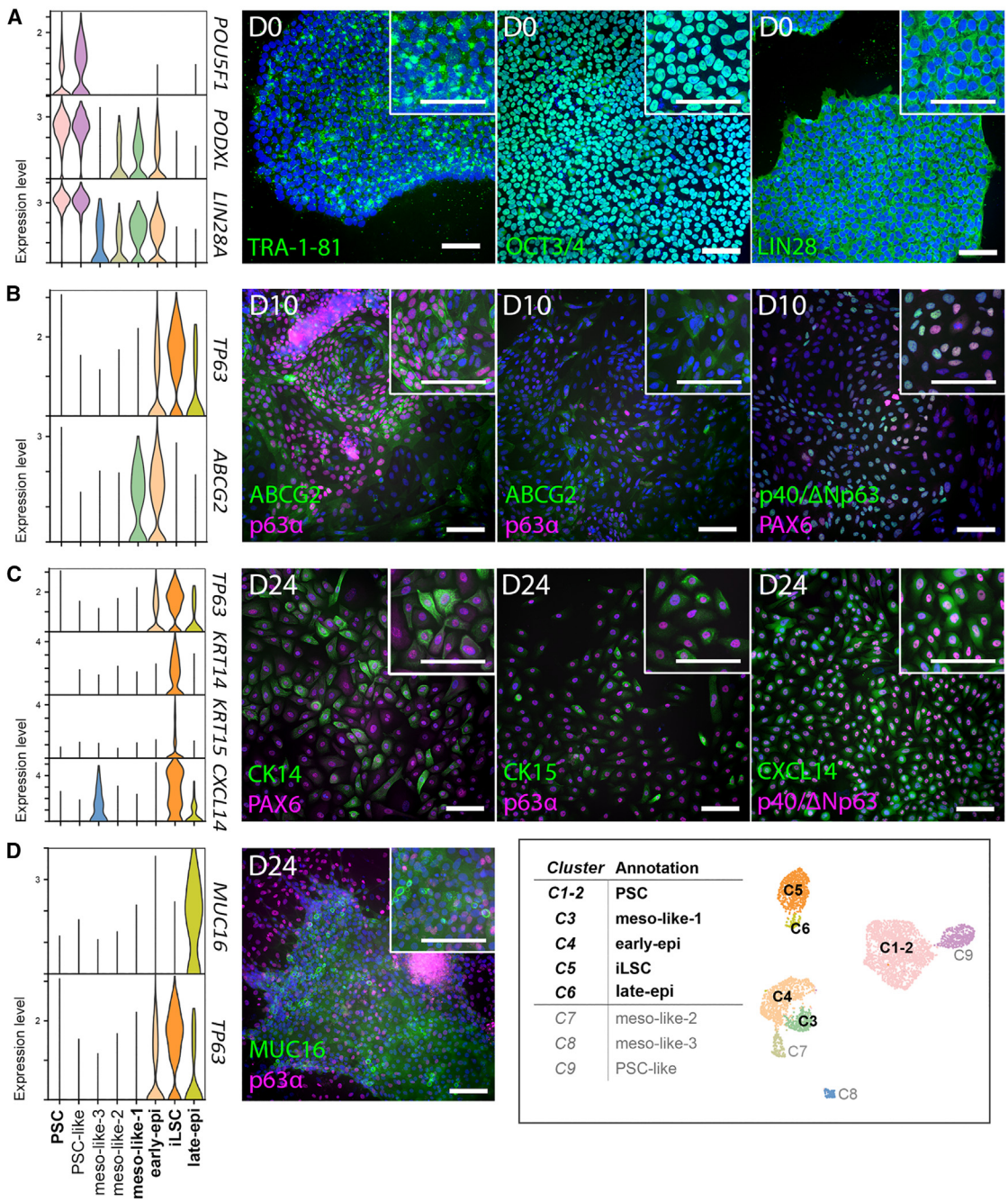

Figure 3. Immunofluorescence validation of time point-specific clusters 1–6

(A) Expression of selected C1–2/PSC markers in the undifferentiated hPSCs at D0.

(B) Expression of C3/meso-like-1 and C4/early-epi marker *ABCG2* (ABCG2), C4/early-epi marker *TP63* (p63α and p40/ΔNp63), and corneal lineage marker *PAX6* (PAX6) at D10.

(C) Expression of *PAX6* (PAX6) and C5/iLSC markers *KRT14* (CK14), *KRT15* (CK15), *TP63* (p40/ΔNp63), and *CXCL14* (CXCL14) at D24.

(D) Expression of C6/late-epi marker *MUC16* (MUC16) at D24. Representative IF images are shown for hiPSC1. Cell nuclei counterstained with DAPI or Hoechst 333420 (shown in blue). Scale bars, 100 μm.

(Figure 2D). Therefore, we named C4 as the "early-epi" cluster. Furthermore, marker genes for both clusters were significantly enriched in "cell-substrate adhesion," "tissue migration," and "wound healing" associated GO terms (Figure 2D). Consistent with the scRNA-seq data, IF of the D10 cell culture samples demonstrated wide expression of

ABCG2, with p63α co-expression in a subset of ABCG2-positive cells (Figure 3B).

Both C5 and C6 displayed clear epithelial characteristics. C5 expressed high amounts of *TP63* as well as other epithelial and LSC genes like *KRT14* and *CXCL14* (Figures 2C, and S2). C5 marker genes were highly enriched for "epidermal development/differentiation"-associated GO terms, indicating epithelial stem cell features (Figure 2D). We annotated C5 as the "induced LSC" (iLSC). Gene regulatory network analysis on the C5/iLSC using SCENIC (Aibar et al., 2017) revealed several key transcription factors of LSCs (Figures S3A and S3B). Among them, *HMGA2*, *KLF6*, *TP63*, *JUN*, and *FOSL2* had high gene expressions and SCENIC prediction scores (mean AUC, which stand for Area under the ROC Curve). Signaling pathway assessed with PROGENy (Schubert et al., 2018) identified activation of, e.g., androgen, epidermal growth factor receptor (EGFR), hypoxia, Janus kinase/signal transducers and activators of transcription (JAK-STAT), nuclear factor κB (NF-κB), and p53 pathways (Figure S3E). C6 showed high expression of diverse types of epithelial genes, such as *AQP3*, *MUC16*, and *S100A9* related to epidermis or conjunctival cells, as well as *KRT8* and *KRT18* (Figures 2C, and S2). A few cells had low expression of skin (*KRT1*)- and hair follicle *(KRT72)*-associated keratins but no typical conjunctival genes, such as *MUC2*, *MUC20*, or *KRT13*, or mature corneal epithelial markers *KRT3* or *KRT12* were observed (Figure S2). These findings suggest that C6 has a distinct epithelial cell state from C5. Therefore, we annotated C6 as "late-epi." Gene regulatory network SCENIC analysis on the C6/late-epi revealed key transcription factors such as *GATA3*, *ARID3A*, *IRF6*, *KLF5/6*, and *CREB3L2* (Figures S3C and S3D).

IF of D24 cell cultures confirmed protein-level expression of C5/iLSC markers CK14, CXCL14, p63α/p40, as well as another limbal cytokeratin CK15 (Figure 3C). MUC16 was present in morphologically distinguishable, layered cell areas accompanied with low p63α staining intensity (Figure 3D), indicating that this population possessed a distinct epithelial cell state from C5/iLSC cells, corresponding to the C6/late-epi.

Interestingly, the expression of corneal lineage marker *PAX6* was too low to reliably quantify in any of the scRNA-seq data clusters (Figure S2). However, quantitative real-time PCR (real-time qPCR) showed upregulated *PAX6* throughout the differentiation process (Figure S4A), and in IF we could see low but increasing intensity of nuclear PAX6 staining in the majority of cells on D10/11 and D24 (Figures 3B and 3C), accompanied by areas of clearly PAX6+ cells especially in the hESC and hiPSC1 cultures (Figure S4B). Coupled with epithelial markers such as p63 and CK14 (Figures 3B and 3C), these findings indicate a successful commitment to CE lineage. Taken together, cell states of the time point-specific C1–2/PSC, C3/meso-like-1, C4/early-epi, C5/iLSC, and C6/late-epi were identified by scRNA-seq and confirmed on the protein level in IF. Cells in these clusters seemed to follow the expected progression of differentiating hPSCs.

Mixed time point populations contain non-epithelial cells which deviate from the LSC differentiation pathway

Among the mixed time point clusters, C7 and C8 cells expressed mesenchymal/mesodermal genes such as *COL1A1*, *ZEB2*, *ACTC1*, and *ACTA2* (Figures 2C, and S2). Additionally, C7 cells expressed high levels of *COL8A1*, *GREM1*, and *TGFB2* and C8 marker genes included *HAND1*, *GATA4*, *PITX1*, *TGFBI*, and *LUM*. Therefore, we annotated C7 as "meso-like-2" and C8 as "meso-like-3" cluster. Interestingly, some of these markers, e.g., *ACTC1*, *ACTA2*, *HAND1*, and *LUM* were expressed in both C3/meso-like-1 and C8/meso-like-3, indicating somewhat similar cell states (Figures 2C, and S2). IF confirmed the expression of HAND1, GATA4, lumican (encoded by *LUM*), and ZEB2 on the protein level, with varying co-staining patterns at D10 and D24 (Figure 4A). However, clear expression of collagen 8A1 (encoded by *COL8A1*) was only observed at D24 (Figure 4A), indicating that it marks a distinct cell cluster.

Intriguingly, C9 had a high resemblance to C1–2/PSC. Similar pro-proliferation and pluripotency-associated genes were detected among the highly expressed markers in both clusters, but C9 cells additionally expressed distinct metallothionein (*MT*) genes, which were not expressed by C1–2/PSC (Figures 2C, andS2). Nevertheless, GO-term enrichment of C9 was similar to that of C1–2/PSC (Figure 2D), and thus it was annotated as the "PSC-like" cluster.

As the PSC-marker-expressing cells belonging to C1–2/PSC were only obtained at D0 (Figure 2B, and Table S1), we used a panel of pluripotency markers to confirm the presence of C9/PSC-like cells at D10 and D24. Indeed, strong expression of SOX2, TRA-1-81, OCT3/4, and LIN28 was confirmed by IF in colonies showing undifferentiated morphology and becoming more abundant within cells at D24 when compared to D10 (Figure 4A). Although almost no hESC-derived cells contributed to C9/PSC-like in the scRNA-seq data, formation of distinctive colonies and expression of pluripotency markers were observed also in hESC at D24 during IF validation (Figure S4C).

To investigate the potential developmental trajectory of all identified clusters, pseudotime analyses were performed with Monocle 3 (Cao et al., 2019). Using C1–2/PSC as the starting point, C5/iLSC cells were identified to be the latest point in pseudotime, both in the pooled analysis (Figure S5A) or when hPSC lines were analyzed individually (Figures S5B–S5D). This is consistent with that C5/iLSC is

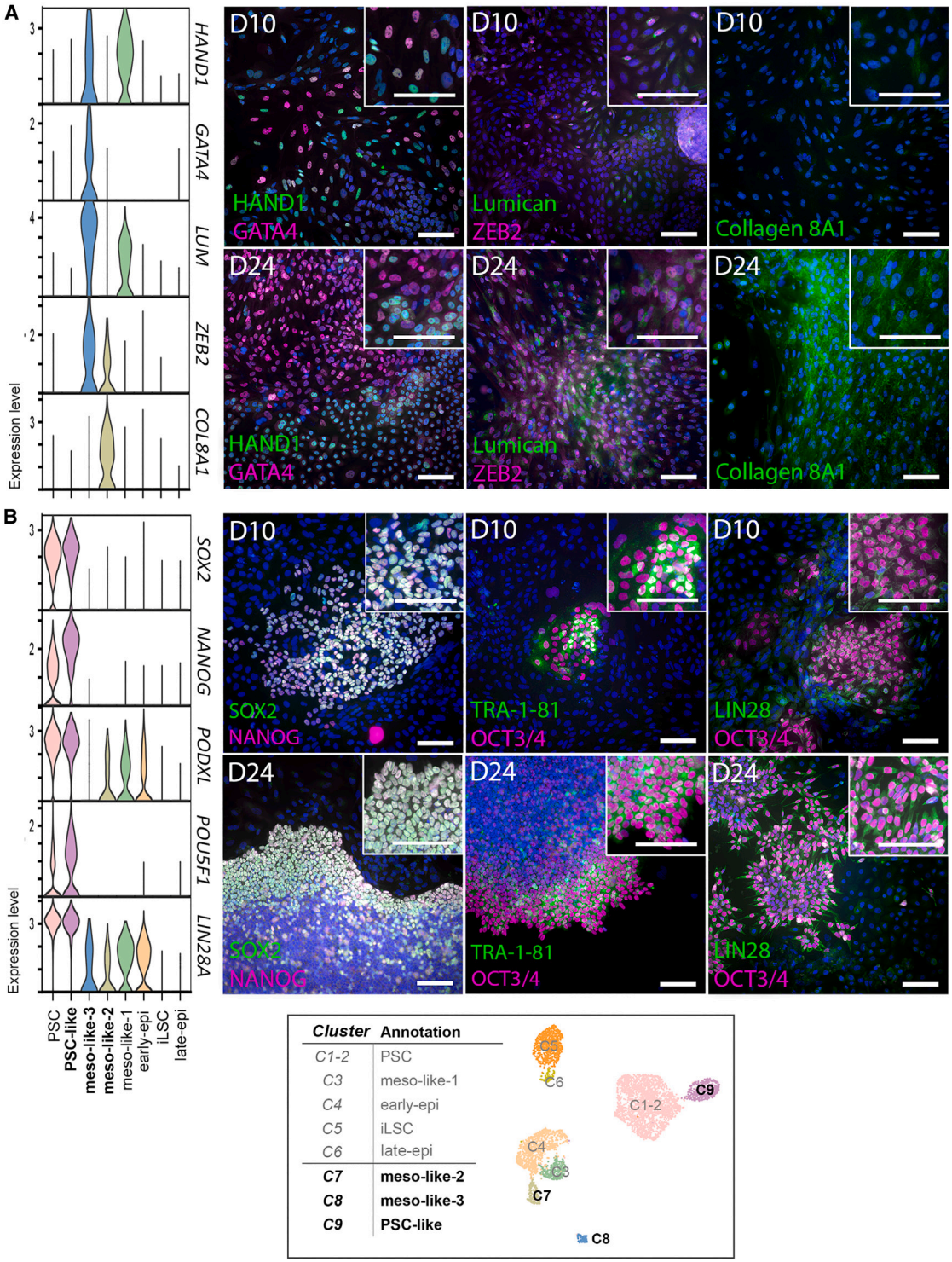

Figure 4. Immunofluorescence validation of mixed time point clusters 7–9

(A) Expression of C7/meso-like-2 and/or C8/meso-like-3 markers *HAND1* (HAND1), *LUM* (lumican), *ZEB2* (ZEB2), and *COL8A1* (collagen 8A1) at D10 and D24.

(B) Expression of pluripotency and C9/PSC-like markers *SOX2* (SOX2), *NANOG* (NANOG), *PODXL* (TRA-1-81), *POU5F1* (OCT3/4), and *LIN28A* (LIN28) among the differentiating cells at D10 and D24. Representative IF images are shown for hiPSC1. Cell nuclei counterstained with DAPI or Hoechst 333420 (shown in blue). Scale bars, 100 μm.

the endpoint of differentiation and contains the most LSC-like cells. Among other clusters, C4/early-epi was shown to be mostly consistently intermediate, between C1–2/C9 and C5, while the remaining clusters showed clear line-dependent differences in pseudotime.

Surface marker quantification confirms differences in the cell line-dependent differentiation efficiency

To assess the line-dependent heterogeneity in detail, we quantified the C5/iLSC cluster in all three lines. A clear majority (95%) of hESC-derived cells collected at D24 belonged to this cluster, instead of only 35% of hiPSC1- and 22% of hiPSC2-derived cells (Figure 5A). One possibility to deal with this line-dependent heterogeneity is to purify the C5/iLSC population. For this, we selected the top genes that showed the highest expression in C5/iLSC based on average log fold change (avg_log2FC), as compared to all other clusters (Table S7). Among them, we identified *AREG* and *ITGA6* as suitable purification marker candidates, as both proteins are known to be cell membrane associated (Figure 5B). Furthermore, as the C9/PSC-like cells constituted a particularly undesirable population, we included *PODXL* (encoding a protein which can be recognized by TRA-1-81 antibody) as the negative selection marker (Figure 5C) and proceeded to validate the markers in flow cytometry (FC).

At D24, AREG was expressed in 71.7% ± 15.3% of hESC vs. 58.4% ± 11.9% of hiPSC1 cells, and ITGA6 was expressed in 92.8% ± 4.0% of hESC vs. 82.5% ± 5.4% of hiPSC1 cells (Figure 5D). The line-dependent differences were statistically significant (paired t test, $p = 0.0319$ for AREG and $p = 0.0302$ for ITGA6). Despite major differences in the amount of *PODXL* transcript in the scRNA-seq data (Figure 5E), TRA-1-81 antigen was measured in 2.7% ± 2.4% of hESC and only in 4.7% ± 3.0% of hiPSC1 cells (Figure 5D), showing no statistically significant difference between the cell lines. Nevertheless, the quantification of C5/iLSC cells with AREG, ITGA6, and TRA-1-81 confirmed more efficient derivation of C5/iLSC from hESC, as compared to hiPSC1 (Figure 5D), consistent with scRNA-seq transcript quantification (Figure 5E).

FACS based on AREG, ITGA6, and TRA-1-81 successfully purified iLSCs

First, we simulated the purification effect by computationally selecting C5/iLSC and comparing their transcriptome to data from all D24 cells and primary adult human LSCs *in silico*. For this comparison, scRNA-seq data of C5/iLSC cells, all D24 cells, and primary adult human LSCs from Smits et al. (2023) were processed into pseudobulk data and visualized with principal-component analysis (PCA). We expected that *in silico*-purified C5/iLSC cells from all cell lines would give rise to more similar transcriptome profiles with each other as compared to all D24 cells. Indeed, the *in silico* purification demonstrated a major impact especially on hiPSC1 and hiPSC2. After purification, all three cell lines clustered closer together and moved closer to the primary LSCs on the PC1 axis, which represents the major difference in this PCA plot (58%) (Figure 6A).

Fluorescence-activated cell sorting (FACS) was used to experimentally validate the purification approach with hESC and hiPSC1 lines, using AREG and ITGA6 for positive and TRA-1-81 for negative selection. TRA-1-81$^-$/ITGA6$^+$/AREG$^+$ cells and all live cells as the unpurified control were sorted for subculture and characterized in IF after 8 days. PSC-like colonies were observed within the unpurified subcultures especially in hiPSC1, whereas TRA-1-81$^-$/ITGA6$^+$/AREG$^+$ subcultures did not display PSC-like morphologies (Figure 6B). The findings were confirmed in IF, which demonstrated expression of pluripotency markers within the unpurified cells, whereas the TRA-1-81$^-$/ITGA6$^+$/AREG$^+$ subcultures expressed LSC markers PAX6, p63α, and CK14, but not pluripotency markers such as OCT3/4 (Figure 6B). Additional IF characterization included MUC16, lumican, and CK12 (encoded by *KRT12*), to reveal the potential presence of C6/late-epi, C8/meso-like-3, or terminally differentiated CE, respectively, but these markers were not detected in any of the samples. Staining with Ki67 (Figure 6B) demonstrated the presence of smaller, proliferative cells which constituted 30.8% ± 0.6% of the hESC and 68.0% ± 1.2% of the hiPSC1 cultures when quantified through cell counting.

To further assess the molecular profile of the FACS-purified cells, we collected samples from TRA-1-81$^-$/ITGA6$^+$/AREG$^+$ hESC and hiPSC1 at 15 days post-sorting and carried out bulk RNA-seq. The expression of selected cluster marker genes in the bulk RNA-seq data of FACS-purified cells was compared to the pseudobulk datasets from "all D24 cells" and the primary LSCs from Smits et al. (2023) (Figure 6C). In line with the *in silico* and FACS-mediated purification results, expression of genes in FACS-purified hiPSC1 cells showed more similarity to hESC cells and primary LSCs. This included increased expression of LSC markers and decreased expression of mesodermal and pluripotency markers. This indicates that our cell sorting strategy, based on surface markers identified from scRNA-seq data, can be used to circumvent the line-dependent variability within the hPSCs and produce purer populations of C5/iLSC cells.

DISCUSSION

Understanding the diverse cell types involved in the *in vitro* differentiation of hPSC is crucial to gain control over the process and advance safe and effective hPSC-based therapeutics. To achieve this, streamlined data-driven tools are ultimately needed to facilitate the development of

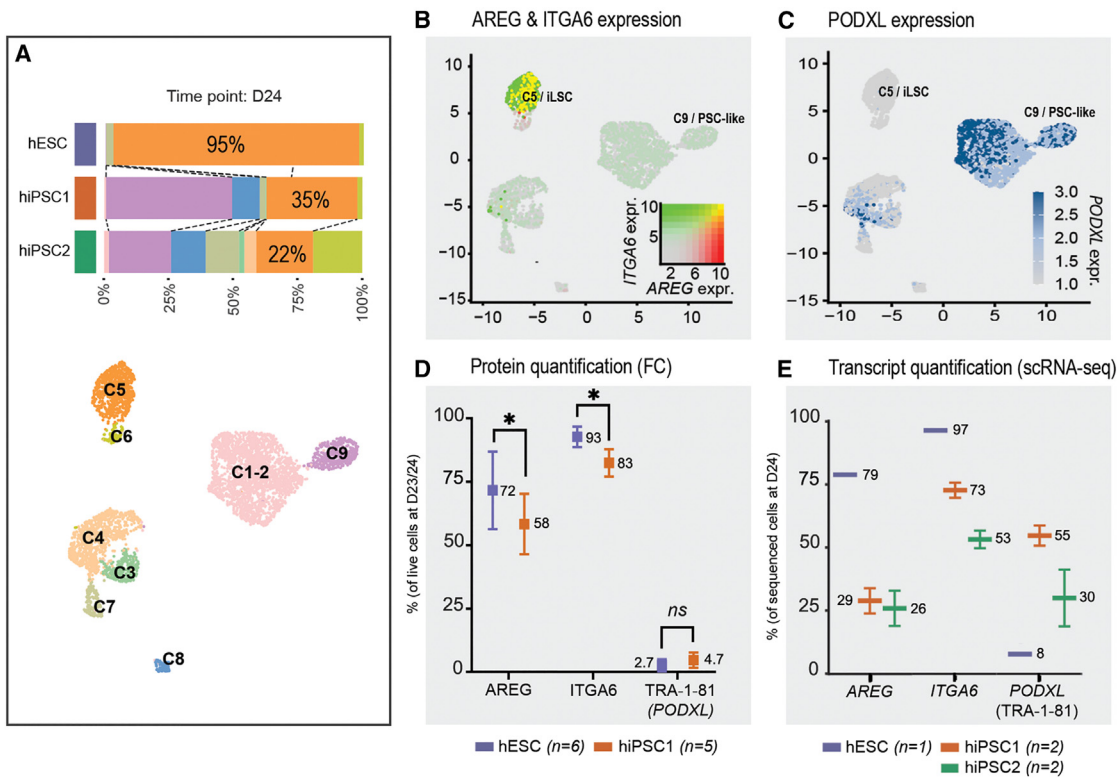

Figure 5. Measuring differentiation efficiency via quantification of surface markers with flow cytometry (FC)

(A) Distribution of cells to the identified clusters at D24, shown separately for each cell line.

(B and C) Expression (B) *AREG* and *ITGA6* and (C) *PODXL* transcripts on the UMAP graphs.

(D) FC quantification of AREG, ITGA6, and TRA-1-81 antigens at D23/24.

(E) Quantification of cells with *AREG*, *ITGA6*, and *PODXL* transcripts measured in the single-cell RNA sequencing data. Results shown as mean ± SD, statistical analysis performed with paired t test (*$p \leq 0.05$).

standardization strategies and to support the production of uniform high-quality hPSC-derived cells. In this study, we showed differentiation of three hPSCs lines toward the LSC lineage and used scRNA-seq for detailed analysis of cell types/states along the process, validating these results by IF and FC. Moreover, we demonstrated proof of principle that undesired populations such as non-epithelial or PSC-like cells and high cell-line-dependent variation can be overcome by using surface markers identified by scRNA-seq. The results support our previous works (Hongisto et al., 2017; Vattulainen et al., 2019; 2021) and, more importantly, provide new insights into the differentiation heterogeneity and strategies for improving the protocol outcome.

In this study, we successfully characterized the cell states, C1–2/PSC, C4/early-epi, and C5/iLSC, which follow the expected LSC differentiation trajectory. Additionally, we identified two previously unreported time point-specific subpopulations. C3/meso-like-1 comprising *ABCG2*$^{+}$*/TP63*$^{-}$ cells collected at D10/11 demonstrated elevated expression of genes associated with migration, invasiveness, and extracellular matrix production, indications toward epithelial-to-mesenchymal transition (EMT). This cluster of cells could therefore recapitulate unintended EMT events potentially linked to developmental neural crest-driven EMT (Weigele and Bohnsack, 2020). Whether it represents a mere side product, a transitional state, or a necessary factor influencing the outcome of the overall differentiation process is an interesting question warranting further investigation.

Although many genes in C4/early-epi were also associated with EMT, almost half of the cells also expressed epithelial *TP63*, traditional hallmark of clinically relevant LSCs (Rama et al., 2010), as well as ABCG2. We have previously demonstrated that ABCG2^{+}/p63α^{+} hPSC-LSCs have enhanced regenerative abilities compared to ABCG2^{-}/p63α^{+} cells both *in vitro* and *ex vivo* (Vattulainen et al., 2019). In this study, we demonstrated that D10/11 hPSC-LSCs co-express p63α and PAX6 proteins, which identify the corneal epithelial stem cells/progenitors both in tissue- and hiPSC-derived LSC cultures (Hayashi et al., 2016; Norrick et al., 2021). Therefore, we expect C4 to contain early LSC-like cells, and the precursors of the later-emerging

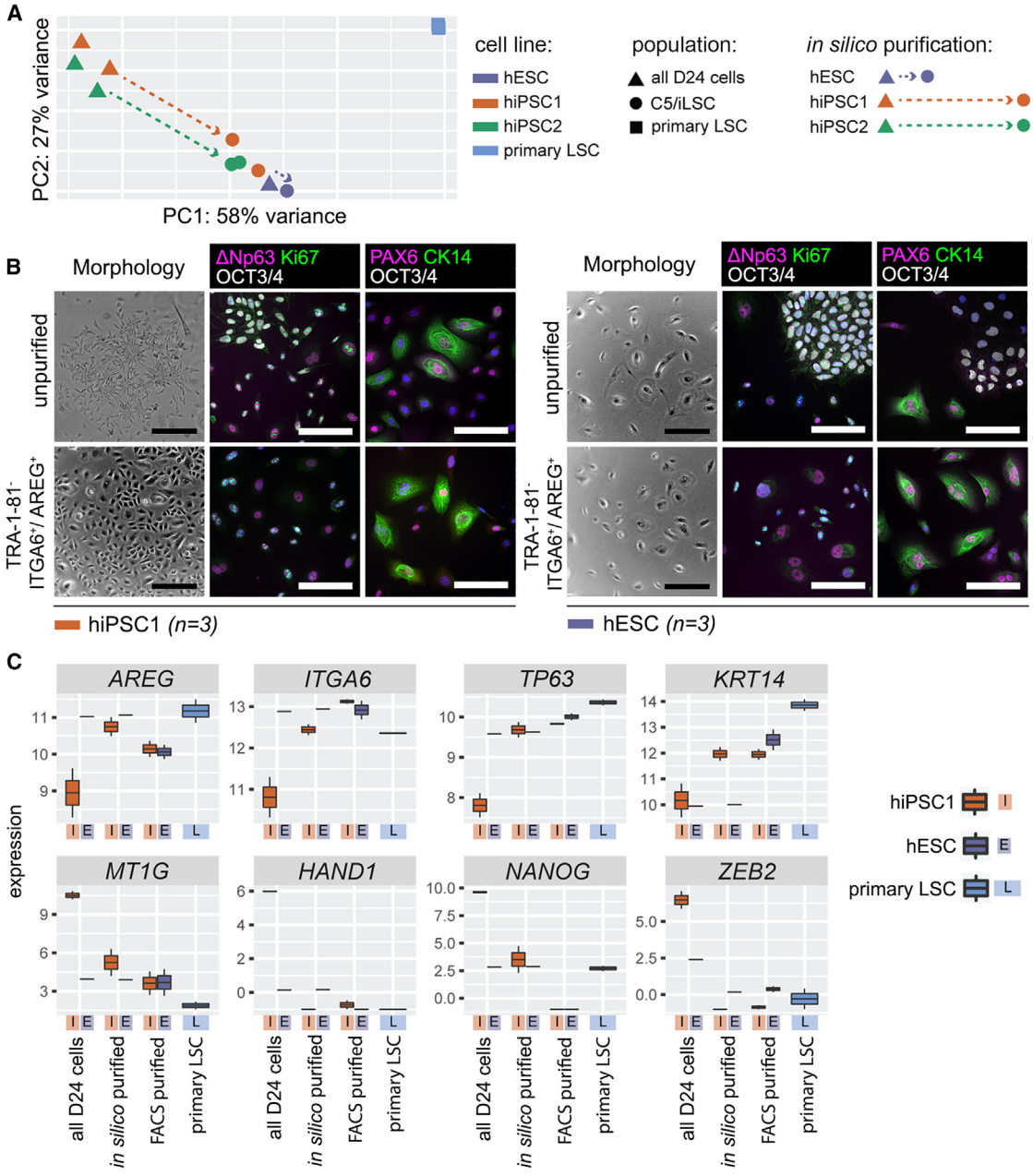

Figure 6. FACS-mediated purification of the target C5/iLSC population

(A) Principal-component analysis (PCA) of *in silico*-purified cells from the scRNA-seq data and primary cultured limbal stem cells (LSCs) control data.

(B) IF characterization of unpurified vs. TRA-1-81-/ITGA6+/AREG+ cells at day D8 post-sorting. Cell nuclei counterstained with Hoechst 333420 (shown in blue). Scale bars, 250 μm (black) and 100 μm (white).

(C) Selected marker gene expression between *in silico*- and FACS-purified hiPSC (I) and hESC (E)-derived populations, compared to primary LSCs (L). Results are shown as boxplots of the log2 values of the deseq2 normalized counts.

epithelial population represented by C5/iLSC, consistent with the pseudotime analysis.

C5/iLSC expressed several traditional limbal progenitor/LSC marker genes, e.g., *KRT14*, *KRT15*, and *TP63* (Figueira et al., 2007; Nieto-Miguel et al., 2011; Pellegrini et al., 2001). It also had a high expression of a more novel marker *CXCL14*, which was recently associated with potential LSCs in human corneas using scRNA-seq (Català et al.,

2021; Collin et al., 2021). Thus, C5/iLSC seemed to represent the most LSC-like population obtained during the differentiation process. Consistent with this, several transcription factors identified as key regulators of C5 through gene regulatory network analysis, such as *TP63*, *JUN*, and *FOSL2*, are known to be important for LSCs (Li et al., 2021; Smits et al., 2023; Wang et al., 2023). However, the distinct profile from primary LSCs, coupled with the lack of CK12 expression even after subculture, strongly indicate that C5/iLSC cells remain rather immature. Manipulation of their serum- and feeder-cell-free culture environment may be required to facilitate further lineage commitment and proper differentiation. It is worth mentioning that, in our data, PAX6 expression remained almost undetectable by scRNA-seq. However, additional analyses with real-time qPCR and IF contrasted this result by demonstrating PAX6 expression on both mRNA and protein level. This may require further attention in our upcoming studies, since correct dosage of PAX6 is known to be essential for corneal epithelial health, starting early on from the eye field development all the way to corneal morphogenesis and postnatal CE (Shaham et al., 2012). Importantly, it is one of the key transcription factors driving lineage commitment toward ocular epithelia instead of skin epidermis (Li et al., 2015; Ouyang et al., 2014; Smits et al., 2023).

C6/late-epi had a small number of cells and was identified as a distinct epithelial population appearing at D24, as cells in this cluster expressed diverse epithelial markers. MUC16 is expressed superficially throughout the adult ocular surface epithelia (Argüeso et al., 2003) and in the developing CE (Collin et al., 2021); KRT8 and KRT18 are expressed in surface ectoderm during development and in myoepithelial cells (Nguyen et al., 2018); and *KRT1* and *KRT72* keratins that were lowly expressed in a small number of cells are normally expressed in the epidermis and hair follicles, respectively (Moll et al., 2008). Several key transcription factors identified in C6/late-epi have been associated with both normal and diseased ocular surface epithelia, e.g., KLF5/6 which, among other things, regulate the development and differentiation of CE (Chiambaretta et al., 2002; Kenchegowda et al., 2011) and GATA3, which is upregulated in aniridia patients (Smits et al., 2023). It would be of interest to further define the cell state of C6/late-epi, whether it is one of the intermediate cell states during differentiation or an off-track cell state.

In addition to the time point-specific C1–5, we also detected three clusters containing cells from multiple time points (C7–9). Two of these clusters, C7/meso-like-2 and C8/meso-like-3 contained cells from both D10/11 and D24 and displayed even higher EMT genes expression than the time point-specific C3/meso-like-1 cells. Activation of unintended neural crest EMT gene expression programs could also explain the origin of these cells. However, unlike the cells in C3/meso-like-1 that either go extinct during the process or develop further into a new phenotype, the cells in C7–8 remain and overcommit to a mesodermal gene signature. Markers expressed in these clusters included genes associated with mesodermal differentiation and often linked to cardiac development, e.g., *ACTC1*, *HAND1*, and *GATA4* (Olson, 2006); this may indicate a diversion from the intended LSC fate.

Finally, C9/PSC-like represented a population deviating even further from the intended cell type. It contained PSC-like cells from all three time points, with a large percentage of D24 cells. PSC marker expression coupled with high proliferation rate is a known risk for tumorigenicity (Lee et al., 2013). The distinctive attribute of C9/PSC-like compared to C1–2/PSC was the clear upregulation of MT genes, which play multiple roles in carcinogenesis, including regulation of cell-cycle arrest, proliferation, and apoptosis (Si and Lang, 2018). Some cells in C9/PSC-like were derived from D0 time point, thus showing high initial MT levels. In a study of Lu et al. (2018), expression of MTs was shown to characterize a bizarre differentiation-arrested population, which was potentially linked to altered cellular responses to medium zinc in a rare population of hESCs. It would be of interest to further investigate the mechanisms of these genes maintaining PSC-like phenotype during differentiation and to unravel if certain hPSC line characteristics make it more prone to produce such unfavorable outcome.

In this study, the general differentiation efficiency was highest in the hESC line, with 95% of the cells belonging to the most LSC-like C5, while both hiPSC1 and hiPSC2 produced noticeably lower number of cells following the differentiation trajectory and contributed more to mixed time point C7–9. This may not be fully surprising, given that the used protocol has been initially optimized with the hESC line (Hongisto et al., 2017). The outcome is most likely also influenced by the differences in cell line origin, embryonic or reprogrammed, and even by the reprogramming method (Bock et al., 2011; Osafune et al., 2008), which also explains the challenges to reproduce the previously reported differentiation efficiencies with different cell lines, despite using a similar protocol (Sun et al., 2021). However, the observed differences represent the well-acknowledged problems of line-dependent hPSC differentiation efficiency.

To solve these issues, scRNA-seq data-driven decisions can be made to further optimize the differentiation protocol for each individual cell line. However, this approach is both time-consuming and laborious, and purification might still be needed. As implicated by our *in silico* purification results, selective enrichment efficiently mitigates the inconsistency between different hPSC lines. Indeed, by

combining positive selection of C5-specific surface markers AREG and ITGA6 with negative selection of pluripotency marker TRA-1-81 in FACS, we efficiently purified C5/iLSC cells. Ultimately, the selective sorting evened out the difference between the golden standard hESC line and the less efficient hiPSC1, by significantly enhancing the post-sort efficiency of hiPSC1 in terms of LSC-associated gene and protein level expression. When compared to the gene expression of primary adult LSCs, the purification was shown to further improve the quality of the hPSC-derived cell populations based on the elevated expression of important LSC marker genes such as *TP63* and *KRT14*, and decreased level of off-target markers like *HAND1* and *ZEB2*. In the future, transferring toward gentler in-process methods such as magnetic bead-activated cell sorting and optimization of post-sort cell culture conditions should be further explored to support scalability.

To the best of our knowledge, our study represents the first to employ scRNA-seq for a detailed investigation of differentiating hPSC-LSCs heterogeneity using several lines, to define the precise cell states, differentiation efficiency, and cell line dependency. Based on specific markers identified by scRNA-seq, we successfully improve the quality and consistency of the obtained hPSC-LSCs through cell sorting.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Further information and requests should be directed to Dr. Huiqing Zhou (j.zhou@science.ru.nl).

Materials availability

There are restrictions to the availability of human cell lines used in this study due to the Act on the Secondary Use of Health and Social Welfare Data. No other unique materials or reagents were generated or used in this study.

Data and code availability

The research data produced at Tampere University are in principle shared and open according to the findable, accessible, interoperable, resuable (FAIR) principles. All raw sequencing files generated in this study have been deposited in the Gene Expression Omnibus database (GEO) with the accession number GSE248497. Generated single cell objects are available on Zenodo at https://doi.org/10.5281/zenodo.11208468. All code used in this study is available at https://github.com/JGASmits/Heterogeneity-of-differentiating-hPSC-derived-corneal-limbal-stem-cells-through-scRNAseq. Public data of scRNA-seq data from adult LSCs were downloaded from GEO: GSM6266908 and GEO: GSM6266909.

Corneal differentiation of hPSCs

Human ESC line Regea08/017 ("hESC") was derived and characterized as described by Skottman (2010). Human iPSC line WT001.TAU.bB2 ("hiPSC1") was established and characterized in-house as described in Grönroos et al. (2021), and hiPSC line WT003.TAU.bC ("hiPSC2") as described in the supplemental information (SI) (see also Figure S6). Corneal differentiation for 23–24 days was performed as previously described in Vattulainen et al. (2019, 2021). Full details on hPSC lines and culture practices are provided in the SI. All the experiments involving cell culture work were conducted on the site of Tampere University, Faculty of Medicine and Health Technology. The faculty has the approval of the National Authority Fimea (Dnro FIMEA/2020/003758) to conduct research on human embryos, and supportive statements from Regional Ethics Committee of the Expert Responsibility area of Tampere University Hospital have been obtained by the research group, granting the permissions to derive, culture, and differentiate hESC lines (R05116), and to establish and use hiPSC lines in ophthalmic research (R16116).

scRNA-seq and data processing

Live single cells were sorted onto 384 wells using FACSAria fusion cell sorter (BD Biosciences), and libraries were constructed following the SORT-seq protocol (Hashimshony et al., 2016; Muraro et al., 2016). Data in the libraries were pre-processed using seq2science (Van Der Sande et al., 2023). Full details on scRNA-seq with SORT-seq and data processing with seq2science pipeline are provided in the SI.

Marker gene selection and GO-term enrichment

Marker genes for each cluster were identified using the Wilcoxon rank-sum test (Stuart et al., 2019). GO-term enrichment was performed using separated positive and negative marker genes for each cluster with clusterProfiler (Wu et al., 2021). A combination of marker genes, GO-term enrichment, and sample time point information was used to generate cluster annotations. The gene lists for each cluster are provided in Table S7.

Suitable markers for sorting were selected among C5/iLSC marker genes (Table S7). The top genes with the highest avg_log2FC were selected and assessed whether their GO terms were linked to the GO term "cell surface" (GO:0009986).

Pseudotime analysis

Monocle3 (1.3.1) (Cao et al., 2019) was used to perform pseudotime analysis in R (4.3.1) with Seurat (5.0.0). To prevent overplotting of time points, quasirandom geom was used to sort time points into categories from the ggbeeswarm package (0.4.0).

Single-cell pathway analysis and single-cell regulatory network inference

The Seurat object was converted to Scanpy (Wolf et al., 2018) object (h5ad file) using Seurat (4.1.1) and SeuratDisk (0.0.0.9019) in R (4.0.5). Cell-signaling pathways activity was determined using PROGENy (Badia-i-Mompel et al., 2022; Schubert et al., 2018) (decoupler-py 1.6.0). For each pathway, the top 500 genes were included. Activity for these curated pathways was inferred for each single cell by running a multivariate linear model from decoupleR.

SCENIC analysis (Aibar et al., 2017) was performed using pySCENIC (0.12.1). To identify the top transcription factors, AUC scores were summed up and sorted into individual transcription factors. Next, the top 25 factors were selected for the C5/iLSC

and C6/late-epi clusters. Finally, the most specific transcription factors were inferred by plotting the mean expression values against the mean AUC scores.

Real-time qPCR

Real-time qPCR was performed using standard methodology and sequence-specific TaqMan gene expression assays (Thermo Fisher Scientific) for GAPDH (Hs99999905_m1) and PAX6 (Hs01088112_m1). Results were analyzed with the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

Cell staining for IF and FC/FACS

Standard indirect IF protocol was performed as previously described (Vattulainen et al., 2021), using primary antibodies listed in Table S2. Olympus IX51 fluorescence microscope (Olympus Corporation) was used to capture representative IF images. ImageJ software tools (Schindelin et al., 2009) were used for determining the percentage of Ki67-stained cells against DAPI-stained nuclei. For FC/FACS, single-cell suspensions were stained with AREG-APC (17-5370-42; Invitrogen), ITGA6-PE (561894), and TRA-1-81-FITC (560194; both BD Biosciences) following standard methodology, and a minimum of 10,000 events from the primary gate were recorded for analysis with FACSAria fusion. TRA-1-81$^-$/ITGA6$^+$/AREG$^+$ cells and live cells with no selection as controls were sorted onto 24 wells and subcultured for 8 days, as described in full detail in the SI.

Statistical methods

Quantitative data from real-time qPCR, IF, and FC are presented as mean ± standard deviation (SD), except for Figure S4, where we show mean ± range. Paired t test was performed to analyze differences between the hESC and hiPSC1 lines in FC marker quantification using the GraphPad Prism 9 software (GraphPad Software Inc.). Differences were considered statistically significant when $p \leq 0.05$.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.stemcr.2024.06.001.

ACKNOWLEDGMENTS

The authors wish to thank O. Melin and H. Pekkanen for their technical assistance and contributions to cell culture. S. Harjuntausta is thanked for contributing to real-time qPCR and IF analyses. Tampere University Imaging Core and Flow Cytometry Core are thanked for providing technical assistance and equipment. Professor Aalto-Setälä group at Tampere University are acknowledged for the iPSC reprogramming plasmids. M.P.A. Baltissen, L.A. Lamers, and S. Rinzema are thanked for operating the Illumina analyzer and performing data demultiplexing.

The authors also wish to thank the following funding sources for their financial support to this study: Research Council of Finland (to H.S.), Sigrid Jusélius Foundation (to H.S.), the National Eye and Tissue Bank Foundation (to M.V.), Mary & Georg C. Ehrnrooth Foundation (to M.V.), The Ella and Georg Ehrnrooth Foundation (to M.V.), Evald and Hilda Nissi Foundation (to M.V.), Aard- en Levenswetenschappen, Nederlandse Organisatie voor Wetenschappelijk Onderzoek (to H.Z.), the European Joint Programme on Rare diseases (to H.Z.), and ZonMw Open (to H.Z.). In addition, this study is based upon the work from COST Action CA18116, "ANIRIDIA-NET," supported by COST (European Cooperation in Science and Technology).

AUTHOR CONTRIBUTIONS

Conceptualization: H.Z., M.V., J.G.A.S., and H.S.; scRNA-seq sample collection, IF, FC/FACS, and cell culture: M.V.; scRNA-seq sample preparation and data analysis: J.G.A.S. and J.A.A.; writing – original draft: M.V. and J.G.A.S.; writing – review and editing: M.V., J.G.A.S., J.A.A., T.I., D.L.C., H.Z., and H.S.

DECLARATION OF INTERESTS

T.I. and H.S. are co-inventors of a pending patent transferred from Tampere University (Tampere, Finland) to StemSight Ltd (Tampere, Finland) regarding the used cell differentiation method. T.I. and H.S. are also co-founders and shareholders in StemSight Ltd.

DECLARATION OF THE USE OF AI IN THE WRITING PROCESS

During preparation of this work, ChatGPT was used to proofread and improve the fluency of the manuscript text. All text was carefully reviewed and edited as needed by the authors, who take full responsibility of all contents of this publication.

Received: December 15, 2023
Revised: May 31, 2024
Accepted: June 3, 2024
Published: June 27, 2024

REFERENCES

Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086. https://doi.org/10.1038/nmeth.4463.

Argüeso, P., Spurr-Michaud, S., Russo, C.L., Tisdale, A., and Gipson, I.K. (2003). MUC16 Mucin Is Expressed by the Human Ocular Surface Epithelia and Carries the H185 Carbohydrate Epitope. Invest. Ophthalmol. Vis. Sci. *44*, 2487–2495. https://doi.org/10.1167/iovs.02-0862.

Arts, J.A., Laberthonnière, C., Lima Cunha, D., and Zhou, H. (2023). Single-Cell RNA Sequencing: Opportunities and Challenges for Studies on Corneal Biology in Health and Disease. Cells *12*, 1808. https://doi.org/10.3390/cells12131808.

Badia-i-Mompel, P., Vélez Santiago, J., Braunger, J., Geiss, C., Dimitrov, D., Müller-Dott, S., Taus, P., Dugourd, A., Holland, C.H., Ramirez Flores, R.O., and Saez-Rodriguez, J. (2022). decoupleR: ensemble of computational methods to infer biological activities from omics data. Bioinform. Adv. *2*, vbac016. https://doi.org/10.1093/bioadv/vbac016.

Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z.D., Ziller, M., Croft, G.F., Amoroso, M.W., Oakley, D.H., et al.

(2011). Reference Maps of Human ES and iPS Cell Variation Enable High-Throughput Characterization of Pluripotent Cell Lines. Cell *144*, 439–452. https://doi.org/10.1016/j.cell.2010.12.032.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature *566*, 496–502. https://doi.org/10.1038/s41586-019-0969-x.

Català, P., Groen, N., Dehnen, J.A., Soares, E., Van Velthoven, A.J.H., Nuijts, R.M.M.A., Dickman, M.M., and LaPointe, V.L.S. (2021). Single cell transcriptomics reveals the heterogeneity of the human cornea to identify novel markers of the limbus and stroma. Sci. Rep. *11*, 21727. https://doi.org/10.1038/s41598-021-01015-w.

Chiambaretta, F., Blanchon, L., Rabier, B., Liu, J.J., Dastugue, B., Rigal, D., and Sapin, V. (2002). Regulation of Corneal Keratin-12 Gene Expression by the Human Krüppel-like Transcription Factor 6. Invest. Ophtalmol. Vis. Sci *43*, 3422–3429.

Collin, J., Queen, R., Zerti, D., Bojic, S., Dorgau, B., Moyse, N., Molina, M.M., Yang, C., Dey, S., Reynolds, G., et al. (2021). A single cell atlas of human cornea that defines its development, limbal progenitor cells and their interactions with the immune cells. Ocul. Surf. *21*, 279–298. https://doi.org/10.1016/j.jtos.2021.03.010.

Cuomo, A.S.E., Seaton, D.D., McCarthy, D.J., Martinez, I., Bonder, M.J., Garcia-Bernardo, J., Amatya, S., Madrigal, P., Isaacson, A., Buettner, F., et al. (2020). Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic genetic effects on gene expression. Nat. Commun. *11*, 810. https://doi.org/10.1038/s41467-020-14457-z.

D'Antonio, M., Benaglio, P., Jakubosky, D., Greenwald, W.W., Matsui, H., Donovan, M.K.R., Li, H., Smith, E.N., D'Antonio-Chronowska, A., and Frazer, K.A. (2018). Insights into the Mutational Burden of Human Induced Pluripotent Stem Cells from an Integrative Multi-Omics Approach. Cell Rep. *24*, 883–894. https://doi.org/10.1016/j.celrep.2018.06.091.

Deng, S.X., Borderie, V., Chan, C.C., Dana, R., Figueiredo, F.C., Gomes, J.A.P., Pellegrini, G., Shimmura, S., and Kruse, F.E.; and The International Limbal Stem Cell Deficiency Working Group (2019). Global consensus on definition, classification, diagnosis, and staging of limbal stem cell deficiency. Cornea *38*, 364–375. https://doi.org/10.1097/ICO.0000000000001820.

Figueira, E.C., Di Girolamo, N., Coroneo, M.T., and Wakefield, D. (2007). The phenotype of limbal epithelial stem cells. Invest. Ophthalmol. Vis. Sci. *48*, 144–156. https://doi.org/10.1167/iovs.06-0346.

Gene Ontology Consortium (2004). The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. *32*, 258D–261. https://doi.org/10.1093/nar/gkh036.

Grönroos, P., Ilmarinen, T., and Skottman, H. (2021). Directed differentiation of human pluripotent stem cells towards corneal endothelial-like cells under defined conditions. Cells *10*, 331. https://doi.org/10.3390/cells10020331.

Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., De Leeuw, Y., Anavy, L., Gennert, D., Li, S., Livak, K.J., Rozenblatt-Rosen, O., et al. (2016). CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome Biol. *17*, 77. https://doi.org/10.1186/s13059-016-0938-8.

Hayashi, R., Ishikawa, Y., Sasamoto, Y., Katori, R., Nomura, N., Ichikawa, T., Araki, S., Soma, T., Kawasaki, S., Sekiguchi, K., et al. (2016). Co-ordinated ocular development from human iPS cells and recovery of corneal function. Nature *531*, 376–380. https://doi.org/10.1038/nature17000.

Hongisto, H., Ilmarinen, T., Vattulainen, M., Mikhailova, A., and Skottman, H. (2017). Xeno- and feeder-free differentiation of human pluripotent stem cells to two distinct ocular epithelial cell types using simple modifications of one method. Stem Cell Res. Ther. *8*, 291. https://doi.org/10.1186/s13287-017-0738-4.

Kenchegowda, D., Swamynathan, S., Gupta, D., Wan, H., Whitsett, J., and Swamynathan, S.K. (2011). Conditional disruption of mouse Klf5 results in defective eyelids with malformed meibomian glands, abnormal cornea and loss of conjunctival goblet cells. Dev. Biol. *356*, 5–18. https://doi.org/10.1016/j.ydbio.2011.05.005.

Lee, A.S., Tang, C., Rao, M.S., Weissman, I.L., and Wu, J.C. (2013). Tumorigenicity as a clinical hurdle for pluripotent stem cell therapies. Nat. Med. *19*, 998–1004. https://doi.org/10.1038/nm.3267.

Li, D.-Q., Kim, S., Li, J.-M., Gao, Q., Choi, J., Bian, F., Hu, J., Zhang, Y., Li, J., Lu, R., et al. (2021). Single-cell transcriptomics identifies limbal stem cell population and cell types mapping its differentiation trajectory in limbal basal epithelium of human cornea. Ocul. Surf. *20*, 20–32. https://doi.org/10.1016/j.jtos.2020.12.004.

Li, G., Xu, F., Zhu, J., Krawczyk, M., Zhang, Y., Yuan, J., Patel, S., Wang, Y., Lin, Y., Zhang, M., et al. (2015). Transcription factor PAX6 (paired box 6) controls limbal stem cell lineage in development and disease. J. Biol. Chem. *290*, 20448–20454. https://doi.org/10.1074/jbc.M115.662940.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402–408. https://doi.org/10.1006/meth.2001.1262.

Lu, J., Baccei, A., Lummertz Da Rocha, E., Guillermier, C., McManus, S., Finney, L.A., Zhang, C., Steinhauser, M.L., Li, H., and Lerou, P.H. (2018). Single-cell RNA sequencing reveals metallothionein heterogeneity during hESC differentiation to definitive endoderm. Stem Cell Res. *28*, 48–55. https://doi.org/10.1016/j.scr.2018.01.015.

Mahmood, N., Suh, T.C., Ali, K.M., Sefat, E., Jahan, U.M., Huang, Y., Gilger, B.C., and Gluck, J.M. (2022). Induced Pluripotent Stem Cell-Derived Corneal Cells: Current Status and Application. Stem Cell Rev. Rep. *18*, 2817–2832. https://doi.org/10.1007/s12015-022-10435-8.

Mikhailova, A., Ilmarinen, T., Uusitalo, H., and Skottman, H. (2014). Small-molecule induction promotes corneal epithelial cell differentiation from human induced pluripotent stem cells. Stem Cell Rep. *2*, 219–231. https://doi.org/10.1016/j.stemcr.2013.12.014.

Moll, R., Divo, M., and Langbein, L. (2008). The human keratins: Biology and pathology. Histochem. Cell Biol. *129*, 705–733. https://doi.org/10.1007/s00418-008-0435-6.

Muraro, M.J., Dharmadhikari, G., Grün, D., Groen, N., Dielen, T., Jansen, E., van Gurp, L., Engelse, M.A., Carlotti, F., de Koning,

E.J.P., and van Oudenaarden, A. (2016). A Single-Cell Transcriptome Atlas of the Human Pancreas. Cell Syst. *3*, 385–394.e3. https://doi.org/10.1016/j.cels.2016.09.002.

Nguyen, Q.H., Pervolarakis, N., Blake, K., Ma, D., Davis, R.T., James, N., Phung, A.T., Willey, E., Kumar, R., Jabart, E., et al. (2018). Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. Nat. Commun. *9*, 2028. https://doi.org/10.1038/s41467-018-04334-1.

Nieto-Miguel, T., Calonge, M., de la Mata, A., López-Paniagua, M., Galindo, S., de la Paz, M.F., and Corrales, R.M. (2011). A comparison of stem cell-related gene expression in the progenitor-rich limbal epithelium and the differentiating central corneal epithelium. Mol. Vis. *17*, 2102–2117.

Norrick, A., Esterlechner, J., Niebergall-Roth, E., Dehio, U., Sadeghi, S., Schröder, H.M., Ballikaya, S., Stemler, N., Ganss, C., Dieter, K., et al. (2021). Process development and safety evaluation of ABCB5+ limbal stem cells as advanced-therapy medicinal product to treat limbal stem cell deficiency. Stem Cell Res. Ther. *12*, 194. https://doi.org/10.1186/s13287-021-02272-2.

Olson, E.N. (2006). Gene Regulatory Networks in the Evolution and Development of the Heart. Science *313*, 1922–1927. https://doi.org/10.1126/science.1132292.

Osafune, K., Caron, L., Borowiak, M., Martinez, R.J., Fitz-Gerald, C.S., Sato, Y., Cowan, C.A., Chien, K.R., and Melton, D.A. (2008). Marked differences in differentiation propensity among human embryonic stem cell lines. Nat. Biotechnol. *26*, 313–315. https://doi.org/10.1038/nbt1383.

Ouyang, H., Xue, Y., Lin, Y., Zhang, X., Xi, L., Patel, S., Cai, H., Luo, J., Zhang, M., Zhang, M., et al. (2014). WNT7A and PAX6 define corneal epithelium homeostasis and pathogenesis. Nature *511*, 358–361. https://doi.org/10.1038/nature13465.

Pellegrini, G., Dellambra, E., Golisano, O., Martinelli, E., Fantozzi, I., Bondanza, S., Ponzin, D., McKeon, F., and De Luca, M. (2001). P63 identifies keratinocyte stem cells. Proc. Natl. Acad. Sci. USA *98*, 3156–3161. https://doi.org/10.1073/pnas.061032098.

Rama, P., Matuska, S., Paganoni, G., Spinelli, A., De Luca, M., and Pellegrini, G. (2010). Limbal stem-cell therapy and long-term corneal regeneration. N. Engl. J. Med. *363*, 147–155. https://doi.org/10.1056/nejmoa0905955.

Sato, Y., Bando, H., Di Piazza, M., Gowing, G., Herberts, C., Jackman, S., Leoni, G., Libertini, S., MacLachlan, T., McBlane, J.W., et al. (2019). Tumorigenicity assessment of cell therapy products: The need for global consensus and points to consider. Cytotherapy *21*, 1095–1111. https://doi.org/10.1016/j.jcyt.2019.10.001.

Schindelin, J., Arganda-Carrera, I., Frise, E., Verena, K., Mark, L., Tobias, P., Stephan, P., Curtis, R., Stephan, S., Benjamin, S., et al. (2009). Fiji - an open platform for biological image analysis. Nat. Methods *9*, 676–682. https://doi.org/10.1038/nmeth.2019.

Schubert, M., Klinger, B., Klünemann, M., Sieber, A., Uhlitz, F., Sauer, S., Garnett, M.J., Blüthgen, N., and Saez-Rodriguez, J. (2018). Perturbation-response genes reveal signaling footprints in cancer gene expression. Nat. Commun. *9*, 20. https://doi.org/10.1038/s41467-017-02391-6.

Shaham, O., Menuchin, Y., Farhy, C., and Ashery-padan, R. (2012). Pax6 : A multi-level regulator of ocular development. Prog. Retin. Eye Res. *31*, 351–376. https://doi.org/10.1016/j.preteyeres.2012.04.002.

Si, M., and Lang, J. (2018). The roles of metallothioneins in carcinogenesis. J. Hematol. Oncol. *11*, 107. https://doi.org/10.1186/s13045-018-0645-x.

Skottman, H. (2010). Derivation and characterization of three new human embryonic stem cell lines in Finland. In Vitro Cell. Dev. Biol. Anim. *46*, 206–209. https://doi.org/10.1007/s11626-010-9286-2.

Smits, J.G.A., Cunha, D.L., Amini, M., Bertolin, M., Laberthonnière, C., Qu, J., Owen, N., Latta, L., Seitz, B., Roux, L.N., et al. (2023). Identification of the regulatory circuit governing corneal epithelial fate determination and disease. PLoS Biol. *21*, e3002336. https://doi.org/10.1371/journal.pbio.3002336.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.

Sun, C., Wang, H., Ma, Q., Chen, C., Yue, J., Li, B., and Zhang, X. (2021). Time-course single-cell RNA sequencing reveals transcriptional dynamics and heterogeneity of limbal stem cells derived from human pluripotent stem cells. Cell Biosci. *11*, 24. https://doi.org/10.1186/s13578-021-00541-4.

Van Der Sande, M., Frölich, S., Schäfers, T., Smits, J.G.A., Snabel, R.R., Rinzema, S., and Van Heeringen, S.J. (2023). Seq2science: an end-to-end workflow for functional genomics analysis. PeerJ *11*, e16380. https://doi.org/10.7717/peerj.16380.

Vattulainen, M., Ilmarinen, T., Koivusalo, L., Viiri, K., Hongisto, H., and Skottman, H. (2019). Modulation of Wnt/BMP pathways during corneal differentiation of hPSC maintains ABCG2-positive LSC population that demonstrates increased regenerative potential. Stem Cell Res. Ther. *10*, 236 (2019). https://doi.org/10.1186/s13287-019-1354-2.

Vattulainen, M., Ilmarinen, T., Viheriälä, T., Jokinen, V., and Skottman, H. (2021). Corneal epithelial differentiation of human pluripotent stem cells generates ABCB5+ and ΔNp63α+ cells with limbal cell characteristics and high wound healing capacity. Stem Cell Res. Ther. *12*, 609 (2021). https://doi.org/10.1186/s13287-021-02673-3.

Wang, B., Guo, H., Liu, D., Wu, S., Liu, J., Lan, X., Huang, H., An, F., Zhu, J., Ji, J., et al. (2023). ETS1–HMGA2 Axis Promotes Human Limbal Epithelial Stem Cell Proliferation. Invest. Ophthalmol. Vis. Sci. *64*, 12. https://doi.org/10.1167/iovs.64.1.12.

Weigele, J., and Bohnsack, B.L. (2020). Genetics Underlying the Interactions between Neural Crest Cells and Eye Development. JDB *8*, 26. https://doi.org/10.3390/jdb8040026.

Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15. https://doi.org/10.1186/s13059-017-1382-0.

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation *2*, 100141. https://doi.org/10.1016/j.xinn.2021.100141.