A Single-Cell Approach To Gene Regulatory Network Analysis of The Human Cornea

Author
Julian Arts

Supervisors
Dr. Jo Huiquing Zhou
Jos Smits MSc
Dr. Jeroen de Ridder

October 15, 2021

Layman's summary

The cornea is comprised of several different tissue layers, which comprises of specific cell types. The outmost layer of the cornea, the corneal epithelium, is maintained by stem cells in the limbus called Limbal stem cells. These stem cells can also regenerate after the cornea has been damaged. However, how the transition of one cell type to the other occurs is not yet fully understood. Determining which processes drive this transition is fundamental to understand how the corneal epithelium is maintained. Changes in interconnected networks of genes are responsible for determining identity of cell types. Using single-cell corneal data is crucial to understand how the interconnected networks in all cell types exert their function, because single-cell analysis relies on full dissociation of the specific tissue. Moreover, not all corneal cell types have data available. Thus, determining how these interconnected networks are comprised in corneal cell types could shed light on these cell types transitioning from one to another.

One important question therefore is: Which important genes govern cell identity within single-cell populations in the human cornea?

By using a bio-informatics software tool called ANANSE we could try to answer this question. First, we show that combining two types of single-cell corneal datasets, regarding both information about genes and their accessibility, from the corneal atlas is feasible. Next, we demonstrate that incorporating these datasets into ANANSE reveals a subdivision of cell identity in corneal cell types. Moreover, with ANANSE we could predict changes in important genes driving interconnected networks when going from one cell population to another.

A Single-Cell Approach To Gene Regulatory Network Analysis of The Human Cornea

Julian Arts[1*]

***For correspondence:** Julian.Arts@radboudumc.nl (FMS); j.a.arts@students.uu.nl (FS)

Present address: [†]Department of Science, Utrecht University, Netherlands

[1]Department of Molecular Developmental Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University, 274, Postbus 9101, 6500HB, Nijmegen, The Netherlands

Abstract The outmost layer of the cornea, the corneal epithelium, is maintained by stem cells in the limbus called limbal stem cells. Cells in the cornea shift their cell type in a process called stratification. For cells involved in corneal stratification, it is vital to maintain the proper cell fate. Transcription factors (TFs) driving gene regulatory networks (GRNs) determine proper cell fate determination. Thus, determining which TFs drive cell fate determination in corneal cell types is necessary to get insights into corneal stratification. Here we combine data from the publicly available human corneal atlas. We show that a combination of single-cell RNA-seq and single-cell ATAC-seq can be used for determining gene regulatory networks. Motif analysis on single-cell ATAC-seq data after integration reveals a division of cell populations into stromal or epithelial cell fates. Moreover, we reveal TFs involved in driving corneal cell identity when we analyze gene regulatory networks of cell type clusters. Most importantly, KLF5, EHF, GRHL1 and ELF3 drive epithelial cell identity. We also show that limbal neural crest derived progenitor cells are separate from corneal, limbal and conjunctival cell clusters. Cell-cell interaction analysis demonstrates that this cell population could modulate the immune cells. Thus, we provide insights in corneal stratification by predicting which TFs are important for driving cell fate determination of cells in the human cornea.

Introduction

The cornea is a transparent tissue in the human eye. Its most outward layer, the corneal epithelium has a barrier function that blocks material, like dust and microbes from entering deeper into the cornea. Blindness can occur if the normal function of the cornea and the corneal epithelium is perturbed.

From basal to apical, the cornea has a distinct order in its tissue layers. First, the corneal endothelium is a small layer of cells, comprised of corneal endothelial cells that pump out fluid from

the corneal stroma to maintain corneal transparency (***Fischbarg, 2005***). The exchange of nutrient and fluids from the corneal stroma is further promoted by Descemet's membrane, which lies on top of the corneal endothelium (***Van Buskirk, 1989***). This membrane also has a protective function in the eye. The corneal stroma has been proposed to play a role in maintaining optical transparency as well as providing nutrients for the cornea (***Zhang et al., 2017***). Within the corneal stroma several cell types are present: melanocytes (***Polisetti et al., 2021***), stromal stem cells (***Hashmani et al., 2013***) and a multitude of different stromal cells.

Bowman's membrane is situated on top of the corneal stroma. This layer forms crypts in the palisades of Vogt (***MacLellan et al., 2017***). Here, limbal stem cells, for instance limbal progenitor cells and limbal neural crest derived progenitor cells reside (***Chen et al., 2019***). On one side adjoined to the limbus is a region called the conjunctiva. The conjunctiva has a role in providing moisture on top of the eyes. This tissue consists of conjunctival epithelial basal cells and conjunctival superficial epithelial cells (***Dartt, 2002***). On the other side of the limbus the corneal epithelium can be found. The corneal epithelium of several cell types, including corneal basal cells and corneal suprabasal cells (***Sosnová-Netuková et al., 2007***). Limbal stem cells give rise to the corneal epithelium during corneal stratification. The cornea is renewed and re-stratifies every 7-14 days (***Haddad, 2000***; ***Majo et al., 2008***. Additionally, when damage is inflicted upon the cornea, the cornea has the ability to regenerate and stratify up to a certain extend. Renewal of the human cornea occurs when limbal stem cell proliferate and differentiate into transit amplifying cells (***Lehrer et al., 1998***; ***Bentley et al., 2007***; ***Amitai Lange et al., 2015***). The current hypothesis is that these transit amplifying cells migrate into the corneal epithelium and give rise to cells of the corneal epithelium. The limbal stem cell niche has been found to be important for facilitating corneal stratification (***Altshuler et al., 2021***). Understanding what drives those cells' cell-fates is relevant to determine in which ways different cells contribute to healthy corneal function and corneal epithelial stratification.

Important players that drive cell-fates are transcription factors (TFs). TFs are proteins that bind to regions on the DNA, which can result in a target gene being expressed or repressed. Transcription factors have a DNA binding domain with which they can bind with a preference on a specific DNA sequence. Binding on specific genomic regions, for instance promoters or enhancers can result in target genes being expressed. Active enhancers are located in open chromatin. At these enhancers, modifications are attached to the histone tails of chromatin located at the flanks of accessible DNA: H3K27 acetylation (H3K27Ac). Upon binding of transcription factors at enhancers, DNA can form loops to downstream genes, resulting in gene expression of target genes by recruiting the transcription machinery which drives gene expression (***Mitchell and Tjian, 1989***). Where TFs bind, can be predicted by analyzing DNA motifs, where transcription factors can bind with their DNA binding domain. Motif analysis can be performed for instance in open chromatin regions. In-between different cell types, open chromatin can be differential accessible. Thus, cell identity of cell types is driven by specific transcription factors driving expression of specific target genes.

Notably, for a couple transcription factors it is known that they play an important role in the cornea. For instance, PAX6 has been shown to be important for human eye development and homeostasis of the adult cornea (***Stanescu et al., 2007***). SOX9 has been found to be important for differentiation of limbal stem cells (***Menzel-Severing et al., 2018***) in vitro. KLF5 and ELF3 promotes corneal differentiation in mice (***Gupta et al., 2011***; ***Yoshida et al., 2000***). Regulation of the corneal epithelium of mice is governed by EHF (***Stephens et al., 2013***) and FOXC1 (***Li et al., 2021***). However, these studies do not show how these transcription factors exactly can drive downstream gene expression and how these TFs exert their roles in the human cornea. Additionally, these studies do not clearly show how transcription factors work together to drive heterogeneity and cell-identity of cells in the cornea.

The way transcription factors drive gene expression can be established with several methods. To determine how transcription factors drive gene expression both gene expression needs to be measured as well as binding of transcription factors. Generally, RNA-seq is used to measure gene expression. On the other hand, determining open chromatin regions can be performed with ATAC-

seq (***Buenrostro et al., 2013***). In this technique, a Tn5 transposase inserts itself into the genome where DNA has an open conformation. Determining the position where Tn5 is inserted gives insights in where promoters, enhancers and other open chromatin regions are situated. Thus, motif analysis on open chromatin regions can be performed which gives insights in which transcription factors could bind at those regions.

Gene regulatory network (GRN) analysis can provide important insights in the importance of which transcription factors drive heterogeneity (***Zhou et al., 2021***). A gene regulatory network estimates which transcription factors are important based on predicted binding of transcription factors and how they could affect expression of genes (***Macneil and Walhout, 2011***). A software package called ANANSE (***Xu et al., 2021***) has been developed by our lab to determine such gene regulatory networks. ANANSE compares two networks based on absolute gene expression (TPM values), predicted transcription factor binding based on motif analysis from ATAC-seq and/or H3K27Ac and differential gene expression between two networks. The software works with bulk cell data of H3K27Ac and ATAC-seq data combined with bulk RNA-seq.

Determining H3K27Ac signals in single cells is technical very difficult and has not been performed yet. In contrast, techniques for determining open chromatin for single cells have been advanced. One of these techniques is called single-cell ATAC-seq (***Buenrostro et al., 2015***). This technique performs ATAC-seq in single cells. Single-cell ATAC-seq could in theory be combined with single-cell RNA-seq to determine GRNs of single-cell clusters. However, determining GRNs based on data from single cells with this tool has not been investigated. ANANSE is used commonly with available bulk data. Although for limbal progenitor cells measurements in bulk have been made, bulk data is not yet available for all cell types present in the human cornea. Therefore, a single-cell approach is needed to determine the GRNs of all available cell types in the human cornea.

A recent study provided insights in cell-fates of corneal cells at the single-cell level by establishing a single-cell corneal atlas ***Collin et al.*** (***2021***). They investigated the human cornea in regards to gene expression and gene accessibility. Although ***Collin et al.*** (***2021***) investigated gene expression and binding of transcription factors, no clear conclusions were drawn about how transcription factors drive gene expression and subsequently cell identity of single-cell populations.

We hypothesized that integrating scRNA-seq and scATAC-seq of the corneal atlas in ANANSE could answer an important question: which transcription factors drive GRNs of single-cell populations of the human corneal atlas? Additionally, several other questions could be answered: How well can scRNA-seq and scATAC-seq be integrated? Can we use scRNA-seq and scATAC-seq for GRN analysis? How to best compare GRNs of single-cell populations with ANANSE? Can we reveal important insights in GRNs during corneal differentiation?

Results

Single-cell RNA-seq clustering identifies different cell types in the human cornea

To gain insights into the transcription factors driving single-cell populations in the human cornea, we used a bioinformatics pipeline for retrieving and analyzing the single-cell cornea data. In this pipeline we started by performing quality control of scRNA-seq from corneal atlas data. After performing quality control in regard to cell quality parameters (M&M), we clustered single cells into eleven clusters (***Figure 1A***). Importantly, batch or cell cycle effects did not drive drive cell clustering (***Figure 1B*** & ***Figure 1C***).

To annotate the eleven clusters, we used the markers proposed by ***Collin et al., 2021***. The normalized counts as well as the violin plots (Supplementary information) of the markers for TYRP1, PMEL, MLANA, MITF and TYR demonstrated that cluster 10 contains melanocytes. This cluster also contains CDH19, a marker for corneal endothelial cells. We annotated this cluster as melanocytes and endothelial cells (MEC). Cell of cluster 8 highly expressed the markers ACKR1, CCL21, LYVE1 and POSTN. The expression of these markers indicated that this cluster contains both blood ves-

sels and lymph vessels. Thus, this cluster was termed as vessels (Ves). Cluster 9 solely expressed the immune cell markers CCL3 and CCL5. This cluster consists of immune cells (IC). Both cluster 7 and 4 expressed markers of several sub cell clusters: limbal fibroblasts, limbal stroma keratocytes, central stroma keratocytes. Cluster 4 showed strong expression of KERA and MMP3, terming this cluster to be central stromal stem cells (CSSCs). Further sub-clustering of cluster 7 could not be performed. Thus, cluster 7 was termed as general stromal cells (StC). Cluster 11 expressed all markers associated with fibroblastic corneal endothelial cells (FCECs): TAGLN, ACTA2, COL1A1 and COL3A1. We termed this cluster as such. Although multiple clusters expressed markers associated with limbal progenitor cells (LPCs), S100A2 was the most highly expressed in cluster 6. We annotated cluster 6 as LPCs. Likewise, multiple clusters expressed PAX6 and TP63, markers for limbal neural crest derived progenitor cells (LNPCs). In contrast, cluster 5 expressed CPVL the most highly. Accordingly, we annotated cluster 5 as LNPCs. Based on the high expression in cluster 2 of the conjunctival superficial markers: KRT13, KRT19, S100A8 and S100A9, we termed this cluster Superficial conjunctival cells (CjS). Cluster 1 expressed HES1, HES5 and more highly GJB2. Due to the fact that cluster 5 was termed as LNPCs, we termed cluster 1 as corneal basal cells (CB). The expression level of KRT24 was very high in cluster 3, thus we annotated this cluster as corneal suprabasal cells (CSB). The annotated cell clusters can be seen in (***Figure 1D***). We could annotate most of the cell types described in the original study (***Figure 1E***). This shows that clustering annotation according to the provided marker genes of the corneal atlas could be conducted fairly well with our quality control parameters used. With the clusters being correctly annotated, we predicted which cells in scATAC-seq corresponded to the annotated clusters of scRNA-seq.

Integration on single-cell ATAC-seq according to single-cell RNA-seq clusters

The cells present in scATAC-seq needed to be annotated according to scRNA-seq to perform integration. For our quality control of scATAC-seq we used snapATAC (M&M), because this method does not underrepresent single-cell clusters with a small number when determining accessible chromatin. We performed a strict quality control on the scATAC-seq datasets in snapATAC (M&M), given that the number of reads mapping to the genome is more sparce in scATAC-seq than in scRNA-seq.

Following, we imputed cell clusters of scATAC-seq based on scRNA-seq. We conducted imputation of datasets by predicting the likelihood of how well a cell in scATAC-seq resembled the annotated cells in scRNA-seq. This imputation was based on how well the differential expressed genes for each cell cluster were correspondingly differential accessible. All annotated single-cell populations from scRNA-seq were visible in scATAC-seq (***Figure 2A***). Still, not all populations had a high number (>100) of cells in scATAC-seq. In short, for eight out of the eleven populations imputation seems to work well, which had a high number of cells. The success of imputation is further demonstrated by the fact that the accessibility around expected marker genes was high in the associated clusters as expected. For instance, S100A9, a marker for CjS was found to be the most accessible in this population (***Supplementary Figure 1A***). Next, MMP3, a marker of stromal cells had the highest accessibility in CSSCs (***Supplementary Figure 1B***). S100A2, a marker for LPCs, showed the highest accessibility in the expected population (***Supplementary Figure 1C***). The marker CDH19 showed highest accessibility in MEC (***Supplementary Figure 1D***). These results demonstrate that imputation of single-cell clusters of scATAC-seq according to scRNA-seq works well for the corneal atlas data.

The importance of peak calling on binned regions is visible from the UMAP generated in the study of ***Collin et al.*** (***2021***) (***Figure 2B***). For instance, LNPCs seem very closely related to red blood cells in terms of chromatin accessibility, which is very unexpected. Moreover, in our UMAP of scATAC-seq cells seem less scattered over other single-cell populations (***Figure 2A***) than in the original study (***Figure 2B***).

TF motif enrichment separates clusters into epithelial and stromal cell fates

With our annotated clusters in scATAC-seq, we needed to determine which clusters had a high enough quality to perform motif enrichment on. We selected only the clusters that had a minimum number of cells (M&M) to obtain enough signal to predict TF binding given the sparsity of scATAC-seq data.

To determine which TFs bind onto accessible DNA in each single-cell population, we performed motif analysis on our scATAC-seq clusters with GimmeMaelstrom (M&M). This tool determines the likelihood that DNA motifs across the genome are accessible by chance, also termed motif enrichment. It also links which TFs are estimated to bind these DNA sequence motifs. The heatmaps of the top 10 motifs for each single-cell populations that shows both the relative gene expression and motif enrichment revealed clustering of single-cell populations into three larger populations (***Figure 2D***). The individual cells were split into limbal, corneal and one conjunctival cluster (LPCs, LNPCs, CSB, CB and CjS) against stromal cell clusters (StC and CSSCs). Melanocytes and endothelial cells (MEC) was found to be the outgroup. Likewise, the split was also visible in the clustering dendrogram of scATAC-seq (***Figure 2C***). These results suggest that distinct transcription factor binding drive clustering of epithelial and stromal cell fates in our single-cell populations.

From comparing motif enrichment and gene expression, we could estimate how well integration of single-cell datasets worked by determining which TFs are found. A clear correlation between the relative expression level and motif enrichment can be observed (***Figure 2D***). This is particularly true for MEIS1, GRHL2, PAX6, TP63 and TGIF1. The success of our data integration is also shown by finding these TFs, because PAX6 and TP63 are both associated with the limbus and corneal epithelial development (***Table 2***). These TFs seem to bind at accessible DNA as well as being expressed relatively high in the limbal, conjunctival and corneal cell populations. In contrast, the TFs HLTF, NR3C1 and MYEF2 showed relative high gene expression and motif enrichment in the stromal cell populations (CSSCs and StC). Based on finding these TFs, we demonstrate that our integration worked well.

Comparing single-cell corneal atlas populations against pluripotent cells

Only performing motif enrichment is not sufficient to determine which transcription factors drive GRNs of single-cell corneal atlas populations, because our motif enrichment only predicted binding of TFs in cell populations based on scATAC-seq. To perform gene regulatory network analysis, predicted TF binding also needed to be linked to the upregulation of target genes from TFs. Therefore, expression from scRNA-seq, including the expression of target genes and TFs, needed to be combined with chromatin accessibility from scATAC-seq. To predict which transcription factors drive gene regulatory networks of the single-cell populations we used a gene regulatory network analysis software tool: ANANSE.

In short, ANANSE determines an influence score of a transcription factors by comparing two networks in regards of predicted TF binding based on ATAC signal, absolute gene expression (TPM value for each cluster) of TFs and target genes and differential gene expression. However in order to calculate the influence of TFs, we needed to compare each cell cluster network with a comparison network. Raising the question: “against what type of cells network can the single-cell populations best be compared?” Generally, pluripotent cells are very different from differentiated cells in regards of gene expression and chromatin accessibility. Thus, we hypothesized that a comparison against pluripotent cells would reveal important transcription factors in our differentiated single-cell populations. A very distinct type of pluripotent cells are embryonic stem cells. We estimated that comparing single-cell corneal atlas populations against ESCs would reveal important transcription factors involved in the cornea when compared to pluripotent ESCs. Additionally, comparing each cell population against the same ESC dataset could perhaps also reveal subtle differences in TF influence between the single-cell clusters.

We estimated how well GRNs of single-cell populations could be determined by revealing in which populations overlapping transcription factors could be found. We expected that some tran-

scription factors would be found more than once over multiple cell populations, because these would be the transcription factors driving differences in developmental fates between the eye and the embryo. Unexpectedly, factors that were present in the top 20 transcription factors of all the single-cell populations revealed a large number of overlapping transcription factors across all cell populations (***Figure 3A*** & ***Figure 4A***). This demonstrates that GRNs of single-cell populations cannot be determined sufficiently with this method.

By determining which literature annotated transcription factors were found we further see how well GRNs of single-cell populations were predicted. We found transcription factors related to the eye and eye development: FOSB, KLF2, HSF4, ATF3 and HES4 (***Table 1***). Additionally two epithelial related factors, MAFB and JUNB (***Table 3***) were also found to be important in driving differences between the cells from the eye and the embryo. Additionally, these shared factors are not expressed relatively the same in each single-cell population, despite having a similar influence score (***Figure 4A***). Additionally, KLF4 has been found to play a specific role in corneal epithelium and not in limbal stem cells, stromal stem cells, stromal cells and endothelial cells (***Table 2***). These data demonstrate that comparing single-cell corneal atlas clusters against ESCs does reveal important transcription factors in eye tissue. However, this approach did not seem to predict transcription factors driving GRNs of single cells sufficiently.

This conclusion was supported by analyzing distinct transcription factors for each single-cell population. Only one distinct transcription factor was found in the GRN of CB (***Figure 3A***). For LNPCs and LPCs only two unique factors were found in their GRNs (***Figure 3A***). Importantly, none of those distinct transcription factors had an high influence score (***Figure 4A***). StC, MEC, CjS and CSSCs were the only four clusters which showed distinct transcription factors that had a high influence score (***Figure 4A***). For StC this included no transcription factors involved in the epidermis or the cornea. SIX6 is not specifically found in the cornea and PAX3 was expected in neural crest cells (***Table 2***). These data show that comparing single-cell corneal atlas clusters against ESCs cannot distinguish transcription factors between the single-cell populations.

Of note, a very small number of transcription factors were present in their expected cell populations. CjS showed two factors involved in corneal development: ZBTB33 and GRHL2 (***Table 1***). Interestingly LMX1B, associated with the corneal endothelium and corneal stromal cells (***Table 2***) was found to be important solely for CSSCs (***Figure 4A***). Besides the shared transcription factors and very few relevant transcription factors, no clear structure was visible in the data. Thus, a different approach was needed which could determine important transcription factors driving GRNs of single-cell populations in the human cornea.

Gene regulatory networks of stromal and corneal epithelial single-cell clusters are governed by distinct TFs

Since a comparison of our clusters with pluripotent cells did not provide relevant insights in important TFs driving GRNs of our single-cell populations, we tried an alternative approach. We compared the gene regulatory network of each single-cell population against the gene regulatory network of the median from all other cell populations (M&M). Importantly, we scaled populations with a high amount of cells to the median of all cell populations to make sure large populations were not over-representing the median network. We termed this new method the median approach.

We hypothesized that our median approach would reveal distinct differences between our single-cell corneal atlas populations and their cell fates. Our hypothesis was supported by a number of different findings. The relative expression pattern and the transcription factor influence score pattern correlated very well (***Figure 4B***). TFs that had a high influence score correspondingly demonstrated a high relative gene expression. A lack of influence score corresponded with a relative down-regulation of the transcription factor (***Figure 4B***). Next, single-cell populations clustered similarly like scATAC-seq (***Figure 2C***) and our motif analysis (***Figure 2D***). Additionally, we expected that the number of shared factors across all cell populations would be diminished, due to the fact that no TFs should be shared anymore across all cell populations. Indeed, no shared factors across

all cell populations were present (***Figure 3B*** & ***Figure 4B***).

To estimate how much better our median approach worked in determining distinct differences in cell fates of corneal atlas populations, we analyzed the number of shared factors in populations with a stromal cell fate against populations with a epithelial fate. Stromal cells (StC) and stromal stem cells (CSSCs) shared a large number of TFs that were not present in all other single-cell populations (***Figure 4B***). Likewise, we observed a number of shared factors in-between all clusters of LPCs, LNPCs, CB, CSB and CjS (***Figure 4B***). The number of shared TFs across these single-cell populations increased from three, when compared to ESCs (***Figure 3A***) to seven with the median approach (***Figure 3B***). These results indicate that distinct differences between single-cell populations and their corresponding cell fates can be found with using our median approach.

Examples of TFs involved in driving stromal cell identity

To determine which important TFs drive the gene regulatory networks of single-cell clusters with stromal cell fates, we took the top 25 TFs of each single-cell population (***Supplementary Figure 2***). Of the important TFs, only two TFs have been annotated in literature: PITX2 and KLF2 (***Table 1***), suggesting that our median approach could not be sufficient to predict important TFs driving GRNs of stromal cell clusters. We looked at which distinct literature annotated TFs we could find to see if we could determine which TFs could potentially drive the individual populations of stromal cells. As expected, we found a couple of distinct TFs. The literature annotated factor SALL4 (***Table 1***) was distinctly present in CSSCs. Due to the fact that the corneal endothelium is closely related to the stroma (***Figure 2C***), we were not surprised to find ZEB1 in StC (***Figure 4B***). Additionally, LMX1B, a factor associated with the corneal endothelium and the corneal stroma was found in the CSSCs (***Figure 4B***). Unexpectedly, SIX2, a transcription factor associated with the epidermis (***Table 3***) was also found only in this single-cell cluster (***Figure 4B***). These results imply that a small number of distinct TFs influence the GRNs of single-cell populations distinctly with a stromal cell fate.

Examples of TFs involved in driving corneal cell identity

To reveal which important TFs were driving the gene regulatory networks of single-cell corneal clusters with an epithelial fate, we took the top 25 TFs predicted to drive the GRN of each single-cell population. We found three separate clusters showing different trends of influence score and their transcription factors (***Figure 4C***). The first cluster consisted of TFAP2C, USF1 and DBP. These factors had the highest influence scores in LNPCs and lower influence scores in the other clusters. The second cluster showed a trend of an increasing influence score from CB to CSB and a lower influence score in CjS. The third cluster generally contained transcription factors with a low influence score in LPCs and a high influence in CjS. We show with this data that distinct transcription factors drive GRNs of more than one single-cell population with a corneal fate.

The three clusters contained transcription factors with distinct roles annotated in literature. For instance, the TFs TFAP2C and DBP in the first cluster are associated with neural crest development (***Table 2***). The second cluster contained genes involved in corneal differentiation: KLF5, EHF, ELF3 (***Table 2***). Additionally genes reported for early eye development: HSF4, HES4 and KLF6 (***Table 1***) and for epidermal development: GRHL1 (***Table 3***) was also present in this cluster. The third cluster contained a mix of transcription factors related to the epidermis: JUNB and MAFB (***Table 3***) and transcription factors with an currently unprecedented role in the cornea. These data show that transcription factors: KLF5, EHF, ELF3 and GRHL1 are important in driving GRNs of cells with an epithelial cell fate. These results could also suggest that a combinational effect of transcription factors with different roles might contribute to cell identity from single-cell clusters in the epithelial arm.

To determine if distinct TFs could drive GRNs of single-cell clusters with an epithelial fate, we looked at which literature annotated TFs we could find that have been shown to play important roles in the epidermis and/or cornea. We estimated that only a small number of TFs would be found for each single-cell cluster due to the similarity of the clusters. Distinct transcription factors could

be observed in several single-cell clusters (***Figure 4B***). Of the three transcription factors unique to LNPCs, nothing is known in literature. However, for CSB, we found two literature annotated factors: SMAD2 and MTF1 (***Table 3*** & ***Table 1***). To our surprise, LPCs did not include any distinct TFs (***Figure 3B*** & ***Figure 3C***), despite that they closely resemble CjS on the level of scATAC-seq (fig 2C). This could suggest that although this cluster is distinct from LNPCs on the chromatin accessibility level, they are transcriptionally very similar to other single cell clusters in the transcription factors they express and how they potentially regulate downstream target genes. This data demonstrates that only a small number of distinct transcription factors are expected to be important in driving GRNs of single-cell populations. Additionally, this contributes to our proposed hypothesis that a combinatorial effect of transcription factors with distinct roles could be important in driving GRNs of single-cell populations.

Pseudo time analysis of single-cell clusters with an epithelial cell fate

A large number of transcription factors predicted to be important for GRNs overlapped between the epithelial arm clusters: LPCs, LNPCs, CB, CSB and CjS (***Figure 4B*** & ***Figure 4C***). Due to the fact that limbal stem cells give rise to the corneal epithelium during stratification, we wanted to determine how our single-cell clusters with an epithelial fate were ordered in pseudo time. Of course, we estimated that the limbal progenitor cells should most likely ordered first in pseudo time and that the corneal, more stratified, clusters should be ordered last. Likewise, we predicted that CSB were ordered in pseudo time after CB, because CSB cells are situated above CB cells. Still, it was not clear where CjS would fit in pseudo time and which of the two limbal progenitor cell populations: LPCs and LNPCs would be ordered first in pseudo time. We tried to answer which of the two limbal progenitor cell populations could give rise to the other and the additional cell clusters with Slingshot pseudo time (M&M). Slingshot pseudo time analysis showed that LPCs is predicted to be the first population in pseudo time (***Figure 4D***). Following, LNPCs and CB can be seen on the pseudo time scale. Lastly, CSB and CjS are predicted to be at the end of pseudo time. Still, CjS cells had of a large spread (***Figure 4D***), perhaps suggesting that they do not belong in the pseudo time or CjS does not originate directly from one of the limbal progenitor cell clusters. Interestingly, LNPCs are not predicted directly at the beginning of pseudo time, which implied that LNPCs more likely originate from LPCs. We concluded that LPCs are probably the initial population that could give rise to LNPCs, corneal clusters and possibly the conjunctiva.

TF influence of single-cell cluster with an epithelial fate over predicted pseudo time

With the cell populations now ordered according to pseudo time, we could try to answer: how do the important transcription factors influencing GRNs of the single-cell populations according to pseudo-time time scale change? To determine influence changes of important transcription factors according to our pseudo-time scale, we looked at transcription factors that would explain our predicted pseudo time and corneal differentiation the best. We based our criteria on increasing influence scores over predicted pseudo time and the amount of literature annotated factors present in the cluster. Therefore, we chose cluster 2 (***Figure 4C***) to further zoom in on.

Cluster 2 consisted of 10 transcription factors: HSF4, MAFF, KLF6, KLF5, EHF, FOXA1, GRHL1, ELF3, HES4 and MAFK (***Figure 5A***). In LPCs some transcription factors seem to exert an influence, except GRHL1, ELF3, HSF4, HES4 and MAFF (fig 5A). None of the transcription factors had an influence score in LNPCs (***Figure 5A***). In CB, all transcription factors except KLF6, FOXA1, MAFK and MAFF demonstrated an influence score above 0.5 (***Figure 5A***). Likewise, all transcription factors increase in their influence score in CSB, except HES4. KLF6, FOCA1, MAFK and MAFF showed a high influence in CSB (***Figure 5A***). In CjS many transcription factors go slightly down again in their expression score, except HES4, MAFK and MAFF. These results demonstrate that transcription factors annotated in literature to be important for corneal differentiation increase over our predicted pseudo time. Furthermore, these results suggest that the CjS population probably does not belong in our predicted pseudo time despite sharing many important transcription factors with corneal

epithelial cell populations.

Although the influence score trend generally corresponded well with gene expression of these transcription factors over our predicted pseudo time (***Figure 5B***), not for all transcription factor this trend could be seen. Based on gene expression only, HSF4, FOXA1 and MAFK would not be found as important transcription factors due to their low expression. Additionally, KLF6 is highly expressed in CB, but does not have an influence score in this single-cell cluster. This implies that although the transcription factor is expressed, the predicted target genes could not be expressed. These data also show that for some transcription factors implicated to be important in driving GRNs, trends can be seen that are not observed when only looking at gene expression. We noticed that LNPCs had no influence score for any of the transcription factors. Interestingly, EHF and KLF5, transcription factors involved in corneal differentiation are not found in LNPCs (***Table 2***). These TFs might not upregulate their target genes, which could suggest that LNPCs are perhaps not derived from LPCs. If LNPCs are not derived from LPCs, what could be the role of LNPCs in the human cornea be?

Both LNPCs and LPCs are predicted to interact with immune cells

One way to predict a function of LNPCs in the cornea is to see which cell-cell interactions could occur from LNPCs with the other limbal and corneal clusters. ***Collin et al.*** (***2021***) showed that LPCs modulate the immune cell axis by cell-cell interactions. We wanted to see if we could replicate their findings and we predicted with which cells LNPCs might interact. Analysis with the cell-cell interaction prediction tool (M&M) SingleCellSignalR (***Cabello-Aguilar et al., 2020***) replicated the findings of ***Collin et al.*** (***2021***): LPCs are predicted to interact with immune cells. (***Figure 5D***). Additionally, we found that LPCs modulate stromal cells and stromal stem cells (StC and CSSCs) to a high extent. In contrast, we predicted LNPCs to interact with CB, CSB and LPCs besides demonstrating a modulating immune cells (***Figure 5C***). It is unlikely that LNPCs modulate CB, CSB and LPCs with cell-cell interactions, because only one ligand-receptor interaction is shown with each of these cell populations. Given the fact that we find both LPCs and LNPCs to have a multitude of interactions with immune cells, we predict that LNPCs and LPCs could modulate the immune cell axis together.

Discussion

The cornea consists of several tissue layers comprised with many different cell types. Understanding regulatory mechanisms driving cell fates of these cell types provides insights into how processes like corneal stratification is governed. Gene regulatory networks drive cell fates of these cell types. In this study we show that scRNA-seq and scATAC-seq can be integrated very well based on scRNA-seq annotated clusters. Furthermore, our motif analysis demonstrates that distinct transcription factor binding drives clustering of epithelial and non-epithelial fates in single-cell corneal atlas populations. Additionally, we uncover that gene regulatory networks of stromal and corneal epithelial single-cell clusters are governed by distinct transcription factors. For single-cell clusters with an epithelial fate, we predict that factors like KLF5, EHF, ELF3 and GRHL1 are important in driving their GRNs. Lastly, we predict a role for LNPCs in the cornea. Together with LPCs, LNPCs are predicted to interact immune cells.

Prediction of TF cooperation driving cell fate of corneal cell populations

We have found a multitude of important transcription factors by both looking at the transcription factors with the highest influence score (***Figure 4C***) and by determining distinct transcription factors (***Figure 4A*** & ***Figure 4B***). Rather than the absence or presence of distinct TFs, we estimate that it is more likely that a combination of multiple TFs, their relative expression level and their interactions with each other together on downstream target genes are the most important in driving GRNs in closely related cell types. A combination principle holds true for example with cellular reprogramming. Four TFs have together been shown to alter cell fate, driving adult cells to go into pluripotent cells (***Okita et al., 2008***).

It has been shown that a combination of transcription factors drives cell identity, for example in cancer (***Chen et al., 2020***). Interestingly, this study established that KLF5, ELF3 and EHF bind cooperatively at super enhancers, which drives genes important in pathways that contribute to cancer disease phenotype. It is very likely that KLF5, ELF3 and EHF also work cooperatively during corneal differentiation, because the three transcription factors are individually found in literature to be important for corneal differentiation (***Loughner C et al., 2017***; ***Yoshida et al., 2000***; ***Stephens et al., 2013***) and we find them to be very important in our cell clusters ordered according to our predicted pseudo time (***Figure 5A***). Additionally, they also have high expression in corneal single-cell populations (***Figure 5B***). More specifically, analysis of core transcription regulatory circuitry and the regulation of super enhancers in limbal stem cells revealed that super enhancers drive important epithelial regulators (***Li et al., 2021***). Among these important corneal epithelial regulators, KLF5 and EHF, were found.

Of note, transcription factors involved in these types of regulatory mechanisms are not found in important transcription factors driving stromal cell identity of single-cell clusters. Unfortunately, there is a lack of studies establishing roles of transcription factors involved in the corneal stroma. This is demonstrated by the fact that from all shared transcription factors between CSSCs and StC, we do not find any transcription factors associated with processes in stromal cells. Furthermore, the fact that we only find one distinct factor in CSSCs, LMX1B, that is found to be involved in functioning of stromal keratocytes (***Liu and Johnson R, 2010***). What role LMX1B exactly plays in CSSCs is still an unanswered question. Likewise, it is has to be determined if PITX2 and GZF1 both actually play a role in stromal cells and what role they play exactly.

Differentiation of limbal stem cells to conjunctival cells

With our pseudo time analysis we estimate that limbal cell clusters are ordered earlier in pseudotime than corneal epithelial cell clusters (***Figure 4D***), which is consistent with the current knowledge about corneal stratification. However, our pseudo time analysis also included the CjS single-cell cluster, because the conjunctival cell population shows an corneal epithelial fate in regards to open chromatin (***Figure 2C***) and the influence scores of important predicted TFs for corneal stratification (***Figure 5A***). If the limbus actually contributes to renewal of conjunctival cells still remains unanswered. Although there are limbal stem cells present in the conjunctiva, it seems that conjunctival cells have their own associated stem cells (***Bertolin et al., 2019***). These are more likely than limbal stem cells to play a direct role in the renewal of the conjunctiva, because culturing conjunctival goblet cells from these stem cell ex vivo is possible (***Bertolin et al., 2019***). Still, we cannot exclude the possibility that LPCs have a separate developmental trajectory towards CjS. In this trajectory, LPCs could give rise to conjunctival stem cells which then give rise to conjunctiva epithelial cells. Thus, interpretations about CjS within our pseudo time analysis of predicted corneal stratification should be made with caution.

Cell-cell interactions in the limbal niche

We predict that LPCs and LNCPs both modulate immune cells with ligand-receptor interactions. ***Collin et al.*** (***2021***) also demonstrated a modulation of immune cells by LPCs. Cell-cell interactions seem to be very important in the limbal stem cell niche. For instance, MSCs in the limbal niche stimulate limbal epithelial stem cell functions (***Yamada et al., 2015***). As expected, immune cells are also present in the limbal niche (***Polisetti et al., 2021***). Cell signaling molecules from the limbal niche seem to enhance interaction between limbal niche cells and immune cells. The recruitment of immune cells is also stimulated by cell signaling molecules from the limbal niche. Moreover, a recent study showed that depletion of immune cells severely inhibited corneal regeneration by limbal stem cells in mice (***Altshuler et al., 2021***). Additionally, other cells also populate the limbal niche: melanocytes (***Higa et al., 2005***) and stromal cells (***Katikireddy et al., 2016***). Studies need to be performed to see if these cells also might contribute to corneal regeneration.

A multitude of studies propose that adult stem cells interplay with immune cells during regeneration (***Lechner et al., 2017***; ***Goh et al., 2013***; ***Simkin et al., 2017***). If immune cells are playing an important role during corneal regeneration still needs to be determined. Interestingly, stromal cells have been suggested to be important for maintaining stemness of limbal stem cells (***Huang et al., 2015***). From our analysis we predict that LPCs interact with stromal cells and stromal stem cells (***Figure 5D***), which could suggest that limbal progenitor cells and stromal cells interact reciprocally.

Improving the annotation of single-cell clusters

We showed a lower number of clusters than present in the corneal atlas study (***Collin et al., 2021***). In the cell clustering approach they describe, a maximum amount of clusters that could be determined was taken and in retrospect clustered cells together as one single-cell cluster. In contrast, we found that using our established optimal QC parameters in regards to clustering resolution (M&M), that 11 clusters could be seen. If this resolution is not taken into account, then over clustering of cell populations can occur. This can result in finding clusters that are actually not present in your dataset.

As we pointed out earlier, with both our ESC comparison and our median approach not all single-cell clusters contained distinct transcription factors that had a high influence score (***Figure 4A*** & ***Figure 4B***). Thus, it is possible that the in-between approach we use still does not have the proper resolution to determine transcription factors for cell populations where we could not find distinct transcription factors for. If very specific factors need to be determined, it is possible to compare a single-cell population of interest with the closest related single-cell population. However, all information in regard to overlapping factors will then be lost. Instead of our median approach (M&M), a direct comparison can be made if the single-cell populations consist of an equal number of cells in scATAC-seq. Even if single-cell populations are annotated very well in scRNA-seq, a minimum number of cells in scATAC-seq is necessary for downstream analysis. Thus, the main limiting factor in imputation is the number of remaining cells after quality control in scATAC-seq. To overcome this limiting factor both scRNA-seq and scATAC-seq could be retrieved from the same cell (***Stuart et al., 2019***). Moreover, by analyzing both datasets from the same cell will give insights in which predictions from Seurat label transfer are correct. Alternatively, more scATAC-seq on similar tissues can be performed to increase the number of cells. The maximum prediction threshold, which calculates the likelihood that a scATAC-seq cell corresponds to a scRNA-seq cell could then also be increased to be even more certain that your cells are imputed correctly.

Technical points of improvement for pseudotime analysis

We used Slingshot pseudotime analysis to determine pseudotime of our single-cell clusters. Slingshot pseudo time analysis first to determine lineages in single-cell cluster data by using minimum spanning trees with constraints (***Street et al., 2018***). Next, Slingshot uses simultaneous principal curves to fit branching curves to lineages. This is performed to estimate cell-level pseudo time variables for each linage stably. To infer direction of lineages in a more direct way, RNA velocity could be used instead (***La Manno et al., 2018***). This technique relies on a difference between spliced RNA in the cytoplasm and unspliced RNA in the nucleus. They propose that unspliced RNA is ordered later on the time scale than spliced RNA, because it is being processed. Cell populations that are ordered later in time have unspliced RNA in the nucleus that resembles spliced RNA from the cell population that is earlier in time. They state that on the time scale of several hours the tool can predict direction of single-cell populations very well. Performing RNA velocity should also be possible on our single-cell populations.

Technical points of improvement in the current version of ANANSE

ANANSE determines influence score of transcription factors by seeing how binding on DNA affects expression of downstream genes. Although ANANSE can provide important insights in gene regu-

latory networks, technical issues still exist. This is apparent when we analyze influence scores of transcription factors in our LNPCs cluster. We demonstrate that LNPCs do not have an influence score for KLF5 and EHF, which are found to be important during corneal differentiation (***Figure 5A***). Unfortunately, we cannot determine for certain if this is due to actual biology or due to technical problems with determining GRNs with our median approach. One possible technical explanation is that because these transcription factors are slightly higher expressed in all pseudo time related clusters except LNPCs (***Figure 5B***), that during differential expression calculation the genes are not found to be differential expressed. This would result in a lack of influence score from these transcription factors. Supporting the notion that technical problems might account for the lack of TFs is the result that gene expression of for instance EHF and KLF5 is still high, but lower than all other epithelial branch populations (***Figure 5B***). Moreover, LNPCs seem to be closely related to CB and CSB at the level of scATAC-seq (***Figure 2C***), which means that binding at target genes of TFs in LNPCs should largely be similar to CB and CSB.

An important issue is that for some transcription factors a binding motif is shared between multiple transcription factors. This could result in a binding model that is taken for the specific transcription factor to be a general transcription factor binding model determined by ANANSE. Alternatively, a different binding model can be chosen if it is predicted to share in target genes according to a Jaccard index. An instance where the alternative model does not resemble actual binding, is the case for PAX6. Binding of PAX6 was calculated with the PDX1 binding model that receives a negative score when ATAC signal increases. With the newest version of ANANSE a 'jaccard-cutoff' can be implemented to link binding models of transcription factors together only when a certain threshold is passed. This way the likelihood that a wrong binding model is chosen is reduced. However, not all alternative models that are chosen are wrong. Binding models of closely related transcription factors could still provide better insights in actual binding than when a general model is used, even if this is on accident. Thus, an overview of how well the 'jaccard-cutoff' is dictating the binding model and how this relates to actual binding should be supplied in the output of ANANSE.

Even if the binding models are right for all transcription factors, not all important transcription factors will be shown with ANANSE. ANANSE assumes that transcription factors will only be important if they drive expression of target genes. Transcription factors do not always play important roles when activating transcription of downstream genes. They can also bind and repress their target genes, which is the case for example with PAX6 (***Xie et al., 2014***). ***Xie et al.*** (***2014***) showed that different mutations in PAX6 lead to an upregulation of normally downregulated target genes. To get a full overview of all transcription factors, transcriptional repression should be taken into account. However, implementing transcription factors that downregulate genes will be difficult.

ANANSE oversimplifies target gene expression driven by TFs and distal elements. The influence score is both based on promoter binding and binding at regions that are predicted to drive expression of genes in close proximity to that region. Nuances in if TF binding at promoters drives gene expression and binding at distal regions represses gene expression cannot be determined currently. Moreover, predicting distal regulatory elements driving gene expression from only open chromatin data is not fully correct. Open chromatin regions are not always flanked by active enhancer histone marks: H3K27Ac. This implies that not all open chromatin regions function as an active enhancer.

One of these two principles could account for not finding SMAD3 being present in our GRN of LPCs. SMAD3 interacts with PAX6 and RUNX1 , where the three factors act as a core transcriptional machinery in limbal stem cells (***Li et al., 2021***). In the same study, knocking down SMAD3 resulted in a downregulation of PAX6, implicating that SMAD3 is an important TF in limbal stem cells. It is possible that SMAD3 causes gene expression when it is bound at promoters and is associated with repressing gene expression when it is bound at regulatory elements. Alternatively, SMAD3 might bind in open chromatin regions which are not enhancers thereby confounding the prediction of how important SMAD3 is in driving GRNs of LPCs.

There is another, small issue with the current version of ANANSE. When determining GRNs of

single cells, on a few occasions one of two errors (ValueError or RecursionError) is raised during the procedure when calculating the influence scores of TFs. Currently, for these transcription factors where this error is raised, no influence score is calculated. This problem needs to be solved, because it is still possible that these transcription factors could be important in driving GRNs in the single-cell corneal atlas populations.

Validating transcription factors driving the GRNs of single-cell populations in the adult corneal atlas

We show that with our in-between approach we can determine subtle differences between very similar single-cell clusters of the human corneal atlas. For instance, we find KLF5, ELF3 and EHF as important literature annotated factors during our estimated pseudo time of corneal differentiation. We can also determine which genes they are likely to target. To validate if these transcription factors indeed target the predicted downstream genes, a recently developed new technique, called Spear-ATAC (***Pierce et al., 2021***) could be used. This method makes use of perturbing expression of one or more transcription factors, by posing repressive epigenetic marks with the CRISPR/Cas complex at promoters of transcription factors. At the same time, scATAC-seq is performed. Thus, if this technique is used on cornea tissue, for example in mice, it can be determined if knocking down KLF5, ELF3 and EHF separately or simultaneously really affects only the corneal cell populations.

Additionally, during differentiation experiments from limbal stem cells to corneal epithelium it can be established if and how these factors affect corneal differentiation and/or regeneration. Moreover, we can establish if KLF5, ELF3 and EHF are not important in LNPCs, which is our prediction based on the approach we used. For instance, it could be determined when these genes are perturbed in the cornea, if we can find back LNPCs. If they are not found back or significantly reduced, the factors could be of importance for driving GRNs in LNPCs. Combining scRNA-seq and Spear-ATAC from in single cells could directly reveal the influence transcription factors have on the expression of downstream target genes.

Developments of predicting cell-cell interactions

A limitation of cell-cell signaling inference is that only scRNA-seq is used where all real physical cell-cell contacts are not present anymore due to tissue dissociation. Developments in single-cell analysis techniques could also provide more insights in cell-cell interactions. Instead of a full dissociation of tissue to single cells, one could perhaps dissociate the tissue where one cell still interacts with a limited amount of other cells. Sub-selecting these cells, for instance with fluorescent associated cell sorting (***Penter et al., 2018***), and then performing dissociation again, the actual physical cell-cell interactions could be determined. Alternatively, different developments, like combining scRNA-seq and spatial transcriptomics (***Andersson et al., 2020***) or the development of an even more promising technique: spatial single-cell RNA-seq could also provide relevant insights in cell-cell interactions. Currently, predicted paracrine chemical signaling could be experimentally validated with perturbation assays to determine if the interactions indeed occur and how they change after perturbation.

Future applications of single-cell corneal atlas GRNs

Knowing which transcription factors are important in gene regulatory networks is necessary to determine which factors are needed to culture cells. Additionally, cell culture conditions of specific cell populations for the human eye can be improved that currently have a low efficiency. Our data shows that rare cell populations are present in the human cornea which have an unknown function, like LNPCs. The ability to selectively culture these cells and for instance performing experiments in bulk on these cells which are not yet available for single cells and performing perturbation assays could reveal insight in the function of these cells. Understanding GRNs of single-cell populations from the human corneal atlas not only contributes to the knowledge of improving cell culture conditions and understanding function of these cells. Surgically transplanting grafts of limbal stem

cells in eyes of patients to restore corneal damage is already performed (***Rama et al., 2010***). The efficiency of this procedure is very low and is dictated by the type of damage the eye received. Thus, knowing how to improve cell culture conditions could also improve the efficiency by keeping a high number of cells, which can be transplanted. Alternatively, selective culturing of necessary cell populations during corneal regeneration and grafting a combination of cell populations back in the eye can also improve the efficiency and long term outcomes of surgical transplantations.

Several eye diseases are driven by multiple deregulated genes at once. An illustrative example is glaucoma. Among a multitude of GWAS studies, mutations in multiple genes have been associated with glaucoma (***Chen et al., 2014***; ***Gharahkhani et al., 2014***). The principle that multiple deregulated genes are associated with eye diseases also holds true for several other ocular diseases, regardless of age onset. These diseases include for example Cataract, Myopia, Marfan syndrome and Stargardt's disease (***Singh et al., 2018***). Knowing which genes are affected in these syndromes and knowing in which single-cell population or across which single-cell populations the affected gene seems to exert an influence, can provide insights in disease mechanisms of these syndromes.

In conclusion, we show that single-cell RNA-seq and single-cell ATAC-seq of the single-cell corneal atlas can be integrated to determine gene regulatory networks of single-cell populations. Our gene regulatory network analysis reveals transcription factors involved in driving corneal cell identity of single-cell corneal atlas populations.

Materials and Methods

Data availability

Datasets are available under the Gene Expression Omnibus (GEO) identifier GSE155683. These datasets were also used in the study of ***Collin et al.*** (***2021***) .The full processing workflow and documentation of code is available in the github repository: **scananse**. GRCh38 was downloaded with genomepy and was used for all downstream analysis.

Pre-processing single-cell RNA-seq and single-cell ATAC-seq data

Datasets were downloaded from GEO with **seq2science**. The 10X sra files were split into fastq files with seq2science. Cellranger count was run with Cellranger 6.0.1 to retrieve the matrix, barcodes and features files necessary for Seurat analysis in R. Similarly, Cellranger-atac count was run with Cellranger-atac 2.0.0 to generate both the position sorted bam file and the singlecell.csv file, containing barcode information, necessary for snapATAC analysis in bash and in R. The position sorted bam and singlecell.csv file were used to generate snap files, as described in the **SnapATAC** package.

Quality control of scRNA-seq and scATAC-seq in R

scRNA-seq datasets were analysed in R, as described in the Github repository. scRNA-seq Cells were selected with a minimum count of 2000, a feature number higher than 1000 and a mitochondrial percentage lower than 30 percent. scATAC-seq datasets were also processed in R with **SnapATAC**. Due to the higher sparsity of the data compared to scRNA-seq, cells had a stricter quality control. Only cells with a mitochondrial ratio lower than 20 percent were selected. Additionally, the fragment number for each cell was between 5000 and 500000. The number of unique counts was between 3.75 and 5. To minimize the number of doublets which could skew the data, a maximum of 500000 for the fragment number and 5 for unique counts were used. A Log10 coverage above 3.6 and a promoter ratio between 2 and 8 were used. To minimize clustering based on quality control parameters, cells with a duplicate ratio higher than 0.8 and a low mapq (low mapping quality) value higher than 5000 were also excluded. Lastly, a bin coverage of 5000 was used. For scRNA-seq, a number of 10 PCs and a clustering resolution of 0.2 were used based on the Elbow plot, principle component analysis and the Clustree respectively. After annotating the cell clusters

in scRNA-seq as described, scATAC-seq cell cluster were imputed with Seurat label transfer with a maximum prediction score of 0.2.

Bam file generation and motif analysis

After scATAC-seq quality control of cells in R, cell barcodes were retrieved. The barcodes were generated for each single-cell population. For downstream statistical analysis, splitting of the barcodes additionally occurred on three of the four datasets. We performed this for the imputed cell populations consisting of more than 100 cells in scATAC-seq. For motif analysis peaks were generated from the .bam files with peak calling software MACS2. Both the .bam and .Narrowpeak files were used to generate accessible summit files and peak count files with motif analysis software Gimmemotifs. These two types of files were used in q-quantile normalization of peak count files and were used for downsteam ANANSE analysis. Comparison of accessible peaks together with Z-scores of gene expression was conducted in R with Complex Heatmap.

Gene regulatory network analysis of single-cell populations

ANANSE was performed in Jupyter notebook with the Python3 kernel to determine GRNs of single-cell populations consisting of more than 100 cells present in scATAC-seq. The accessible summit files were used to generate a pre-computed pfmscore files with Gimmemotifs. The hg38 genome, the bam files from scATAC-seq for each single-cell population and the accessible summit file of all populations were also used for performing ANANSE binding. For ANANSE network the output network file from ANANSE binding was implemented, next to the TPM count file and hg38 as the reference genome. For ANANSE influence we performed two different network comparisons. We used a network from ESCs (GSM466732) for comparison against our networks from our single-cell populations. All our single-cell cluster networks were compared to this network in ANANSE influence. 500.000 edges were selected when calculating the influence score. For the median approach, we first determined the median of the number of cells across all single-cell populations. To reduce the large influence of large cell populations on the comparison networks, cell populations with a high number of cells were scaled to the median, in the number of cell barcodes for .bam file generation, the number of reads mapping to genes and the signal of peak accessibility. Cell populations with a lower number of cells than the median were not scaled to the median. All other parameters used for ANANSE binding, network and influence have been described for the ESC network comparison. Analysis of the results was performed in R.

Pseudotime analysis

For Slingshot pseudotime analysis the Seurat object from scRNA-seq was loaded in R. A subselection on the single-cell clusters LNPCs, LPCs, CB, CSB and CjS was made. LPCs were selected as the starting population due to the fact that selecting LNPCs as the starting population resulted in a severe exclusion of LPCs cells from pseudotime.

Network analysis

For **SingleCellSignalR** the Seurat object from scRNA-seq was loaded in R. The significant paracrine interactions were calculated for all scRNA-populations present in the annotated scRNA-seq Seurat object. From all significant interactions specific one single-cell of interest, LPCs or LNPCs, and all available receptor cells were chosen.

References

Alldredge A, Fuhrmann S. Loss of Axin2 Causes Ocular Defects During Mouse Eye Development. Investigative ophthalmology visual science. 2016 10; 57:5253–5262. doi: 10.1167/IOVS.15-18599.

Altshuler A, Amitai-Lange A, Tarazi N, Dey S, Strinkovsky L, Hadad-Porat S, Bhattacharya S, Nasser W, Imeri J, Ben-David G, Abboud-Jarrous G, Tiosano B, Berkowitz E, Karin N, Savir Y, Shalom-Feuerstein R. Discrete limbal epithelial stem cell populations mediate corneal homeostasis and wound healing. Cell Stem Cell. 2021 jul; 28(7):1248–1261.e8. doi: 10.1016/J.STEM.2021.04.003.

Amitai Lange A, Altshuler A, Bubley J, Dbayat N, Tiosan B, Shalom-Feuerstein R. Lineage tracing of stem and progenitor cells of the murine corneal epithelium. Stem cells (Dayton, Ohio). 2015 jan; 33(1):230–239. doi: 10.1002/STEM.1840.

Andersson A, Bergenstråhle J, Asp M, Bergenstråhle L, Jurek A, Fernández Navarro J, Lundeberg J. Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. Communications Biology 2020 3:1. 2020 oct; 3(1):1–8. doi: 10.1038/S42003-020-01247-,Y.

Barbollat-Boutrand L, Joly-Tonetti N, Dos Santos M, Metral E, Boher A, Masse I, Berthier-Vergnes O, Bertolino P, Damour O, Lamartine J. MicroRNA-23b-3p regulates human keratinocyte differentiation through repression of TGIF1 and activation of the TGF-ß-SMAD2 signalling pathway. Experimental dermatology. 2017 1; 26:51–57. doi: 10.1111/EXD.13119.

Bentley AJ, Nakamura T, Hammiche A, Pollock HM, Martin FL, Kinoshita S, Fullwood NJ. Characterization of human corneal stem cells by synchrotron infrared micro-spectroscopy. Molecular Vision. 2007 feb; 13:237.

Bertolin M, Breda C, Ferrari S, Van Acker S, Zakaria N, Di Iorio E, Migliorati A, Ponzin D, Ferrari B, Lužnik Z, Barbaro V. Optimized Protocol for Regeneration of the Conjunctival Epithelium Using the Cell Suspension Technique. Cornea. 2019 apr; 38(4):469–479. doi: 10.1097/ICO.0000000000001670.

Boglev Y, Wilanowski T, Caddy J, Parekh V, Auden A, Darido C, Hislop N R, Cangkrama M, Ting S B, Jane S M. The unique and cooperative roles of the Grainy head-like transcription factors in epidermal development reflect unexpected target gene specificity. Developmental biology. 2011 1; 349:512–522. doi: 10.1016/J.YDBIO.2010.11.011.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for multimodal regulatory analysis and personal epigenomics. Nature methods. 2013 dec; 10(12):1213. doi: 10.1038/NMETH.2688.

Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, Chang HY, Greenleaf WJ. Single-cell chromatin accessibility reveals principles of regulatory variation. Nature. 2015 jul; 523(7561):486–490. doi: 10.1038/nature14590.

Cabello-Aguilar S, Alame M, Kon-Sun-Tack F, Fau C, Lacroix M, Colinge J. SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. Nucleic Acids Research. 2020 jun; 48(10):e55–e55. doi: 10.1093/NAR/GKAA183.

Chen L, Huang M, Plummer J, Pan J, Jiang YY, Yang Q, Silva TC, Gull N, Chen S, Ding LW, An O, Yang H, Cheng Y, Said JW, Doan N, Dinjens WNM, Waters KM, Tuli R, Gayther SA, Klempner SJ, et al. Master transcription factors form interconnected circuitry and orchestrate transcriptional networks in esophageal adenocarcinoma. Gut. 2020 apr; 69(4):630. doi: 10.1136/GUTJNL-2019-318325.

Chen SY, Cheng AMS, Zhang Y, Zhu YT, He H, Mahabole M, Tseng SCG. Pax 6 Controls Neural Crest Potential of Limbal Niche Cells to Support Self-Renewal of Limbal Epithelial Stem Cells. Scientific Reports. 2019 12; 9. doi: 10.1038/S41598-019-45100-7.

Chen Y, Lin Y, Vithana E, Jia L, Zuo X, Wong T, Chen L, Zhu X, Tam P, Gong B, Qian S, Li Z, Liu X, Mani B, Luo Q, Guzman C, Leung C, Li X, Cao W, Yang Q, et al. Common variants near ABCA1 and in PMM2 are associated with primary open-angle glaucoma. Nature genetics. 2014 sep; 46(10):1115–1119. doi: 10.1038/NG.3078.

Chen S Y, Hayashida Y, Chen M Y, Xie H T, Tseng S C. A new isolation method of human limbal progenitor cells by maintaining close association with their niche cells. Tissue engineering Part C, Methods. 2011 5; 17:537–548. doi: 10.1089/TEN.TEC.2010.0609.

Chung D D, Zhang W, Jatavallabhula K, Barrington A, Jung J, Aldave A J. Alterations in GRHL2-OVOL2-ZEB1 axis and aberrant activation of Wnt signaling lead to altered gene transcription in posterior polymorphous corneal dystrophy. Experimental eye research. 2019 11; 188. doi: 10.1016/J.EXER.2019.107696.

Collin J, Queen R, Zerti D, Bojic S, Dorgau B, Moyse N, Molina MM, Yang C, Dey S, Reynolds G, Hussain R, Coxhead JM, Lisgo S, Henderson D, Joseph A, Rooney P, Ghosh S, Clarke L, Connon C, Haniffa M, et al. A single cell atlas of human cornea that defines its development, limbal progenitor cells and their interactions with the immune cells. The Ocular Surface. 2021 7; 21:279. doi: 10.1016/J.JTOS.2021.03.010.

Coomer C E, Wilson S G, Titialii-Torres K F, Bills J D, Krueger L A, Petersen R A, Turnbaugh E M, Janesch E L, Morris A C. Her9/Hes4 is required for retinal photoreceptor development, maintenance, and survival. Scientific reports. 2020 12; 10. doi: 10.1038/S41598-020-68172-2.

Dartt DA. Regulation of mucin and fluid secretion by conjunctival epithelial cells. Progress in Retinal and Eye Research. 2002 nov; 21(6):555–576. doi: 10.1016/S1350-9462(02)00038-1.

Deal K K, Chandrashekar A S, Beaman M M, Branch M C, Buehler D P, Conway S J, Southard-Smith E M. Altered sacral neural crest development in Pax3 spina bifida mutants underlies deficits of bladder innervation and function. Developmental biology. 2021 8; 476:173–188. doi: 10.1016/J.YDBIO.2021.03.024.

Fischbarg J. The Corneal Endothelium. Advances in Organ Biology. 2005 jan; 10:113–125. doi: 10.1016/S1569-2590(05)10004-4.

Fujimoto M, Izu H, Seki K, Fukuda K, Nishida T, Yamada S, Kato K, Yonemura S, Inouye S, Nakai A. HSF4 is required for normal cell growth and differentiation during mouse lens development. The EMBO journal. 2004 10; 23:4297–4306. doi: 10.1038/SJ.EMBOJ.7600435.

Gao Y, Zhou J, Xie Z, Wang J, Ho C K, Zhang Y, Li Q. Mechanical strain promotes skin fibrosis through LRG-1 induction mediated by ELK1 and ERK signalling. Communications biology. 2019 12; 2. doi: 10.1038/S42003-019-0600-6.

García-Frigola C, Herrera E. Zic2 regulates the expression of Sert to modulate eye-specific refinement at the visual targets. The EMBO journal. 2010 9; 29:3170–3183. doi: 10.1038/EMBOJ.2010.172.

Gharahkhani P, Burdon K, Fogarty R, Sharma S, Hewitt A, Martin S, Law M, Cremin K, Bailey J, Loomis S, Pasquale L, Haines J, Hauser M, Viswanathan A, McGuffin P, Topouzis S, Foster P, Graham S, Casson R, Chehade M, et al. Common variants near ABCA1, AFAP1 and GMDS confer risk of primary open-angle glaucoma. Nature genetics. 2014 sep; 46(10):1120–1125. doi: 10.1038/NG.3079.

Goh Y, Henderson N, Heredia J, Red Eagle A, Odegaard J, Lehwald N, Nguyen K, Sheppard D, Mukundan L, Locksley R, Chawla A. Eosinophils secrete IL-4 to facilitate liver regeneration. Proceedings of the National Academy of Sciences of the United States of America. 2013 jun; 110(24):9914–9919. doi: 10.1073/PNAS.1304046110.

Guha S, Chaurasia S, Ramachandran C, Roy S. SLC4A11 depletion impairs NRF2 mediated antioxidant signaling and increases reactive oxygen species in human corneal endothelial cells during oxidative stress. Scientific Reports 2017 7:1. 2017 6; 7:1–11. doi: 10.1038/s41598-017-03654-4.

Gupta D, Harvey SAK, Kaminski N, Swamynathan SK. Mouse Conjunctival Forniceal Gene Expression during Postnatal Development and Its Regulation by Krüppel-like Factor 4. Investigative Ophthalmology Visual Science. 2011 7; 52:4951–4962. doi: 10.1167/IOVS.10-7068.

Haddad A. Renewal of the rabbit corneal epithelium as investigated by autoradiography after intravitreal injection of 3H-thymidine. Cornea. 2000 may; 19(3):378–383. doi: 10.1097/00003226-200005000-00024.

Hashmani K, Branch M, Sidney L, Dhillon P, Verma M, McIntosh O, Hopkinson A, Dua H. Characterization of corneal stromal stem cells with the potential for epithelial transdifferentiation. Stem cell research therapy. 2013 jun; 4(3). doi: 10.1186/SCRT226.

Hendee K E, Sorokina E A, Muheisen S S, Reis L M, Tyler R C, Markovic V, Cuturilo G, Link B A, Semina E V. PITX2 deficiency and associated human disease: insights from the zebrafish model. Human molecular genetics. 2018 5; 27:1675–1695. doi: 10.1093/HMG/DDY074.

Higa K, Shimmura S, Miyashita H, Shimazaki J, Tsubota K. Melanocytes in the corneal limbus interact with K19-positive basal epithelial cells. Experimental eye research. 2005 aug; 81(2):218–223. doi: 10.1016/J.EXER.2005.01.023.

Hosokawa R, Urata M M, Ito Y, Bringas P, Chai Y. Functional significance of Smad2 in regulating basal keratinocyte migration during wound healing. The Journal of investigative dermatology. 2005 12; 125:1302–1309. doi: 10.1111/J.0022-202X.2005.23963.,X.

Huang M, Wang B, Wan P, Liang X, Wang X, Liu Y, Zhou Q, Wang Z. Roles of limbal microvascular net and limbal stroma in regulating maintenance of limbal epithelial stem cells. Cell and tissue research. 2015 feb; 359(2):547–563. doi: 10.1007/S00441-014-2032-4.

Katikireddy KR, Schmedt T, Price MO, Price FW, Jurkunas UV. Existence of Neural Crest–Derived Progenitor Cells in Normal and Fuchs Endothelial Dystrophy Corneal Endothelium. The American Journal of Pathology. 2016 10; 186:2736. doi: 10.1016/J.AJPATH.2016.06.011.

Knight RD, Nair S, Nelson SS, Afshar A, Javidan Y, Geisler R, Rauch GJ, Schilling TF. lockjaw encodes a zebrafish tfap2a required for early neural crest development. Development. 2003 12; 130:5755–5768. doi: 10.1242/DEV.00575.

Kole C, Brommer B, Nakaya N, Sengupta M, Bonet-Ponce L, Zhao T, Wang C, Li W, He Z, Tomarev S. Activating Transcription Factor 3 (ATF3) Protects Retinal Ganglion Cells and Promotes Functional Preservation After Optic Nerve Crush. Investigative ophthalmology visual science. 2020 2; 61. doi: 10.1167/IOVS.61.2.31.

La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastriti ME, Lönnerberg P, Furlan A, Fan J, Borm LE, Liu Z, van Bruggen D, Guo J, He X, Barker R, Sundström E, Castelo-Branco G, Cramer P, et al. RNA velocity of single cells. Nature 2018 560:7719. 2018 aug; 560(7719):494–498. doi: 10.1038/S41586-018-0414-6.

Lechner A, Driver I, Lee J, Conroy C, Nagle A, Locksley R, Rock J. Recruited Monocytes and Type 2 Immunity Promote Lung Regeneration following Pneumonectomy. Cell stem cell. 2017 jul; 21(1):120–134.e7. doi: 10.1016/J.STEM.2017.03.024.

Lehrer MS, Sun TT, Lavker RM. Strategies of epithelial repair: modulation of stem cell and transit amplifying cell proliferation. Journal of Cell Science. 1998 oct; 111(19):2867–2875. doi: 10.1242/JCS.111.19.2867.

Li M, Zhu L, Liu J, Huang H, Guo H, Wang L, Li L, Gu S, Tan J, Zhong J, Wang B, Mao Z, Fan Y, Liu C, Yuan J, Ouyang H. Loss of FOXC1 contributes to the corneal epithelial fate switch and pathogenesis. Signal Transduction and Targeted Therapy 2020 6:1. 2021 1; 6:1–11. doi: 10.1038/s41392-020-00378-2.

Li W, Cornell R A. Redundant activities of Tfap2a and Tfap2c are required for neural crest induction and development of other non-neural ectoderm derivatives in zebrafish embryos. Developmental biology. 2007 4; 304:338–354. doi: 10.1016/J.YDBIO.2006.12.042.

Liu P, Johnson R L. Lmx1b is required for murine trabecular meshwork formation and for maintenance of corneal transparency. Developmental dynamics : an official publication of the American Association of Anatomists. 2010 8; 239:2161–2171. doi: 10.1002/DVDY.22347.

Loughner C L, Tiwari A, Kenchegowda D, Swamynathan S, Swamynathan S K. Spatiotemporally Controlled Ablation of Klf5 Results in Dysregulated Epithelial Homeostasis in Adult Mouse Corneas. Investigative ophthalmology visual science. 2017 9; 58:4683–4693. doi: 10.1167/IOVS.17-22498.

Lumb R, Buckberry S, Secker G, Lawrence D, Schwarz Q. Transcriptome profiling reveals expression signatures of cranial neural crest cells arising from different axial levels. BMC Developmental Biology. 2017 4; 17. doi: 10.1186/S12861-017-0147-,Z.

MacLellan B, Tan B, Hileeto D, Mason E, Bizheva K, Sorbara L, Hosseinaee Z. In-vivo imaging of the palisades of Vogt and the limbal crypts with sub-micrometer axial resolution optical coherence tomography. Biomedical Optics Express, Vol 8, Issue 9, pp 4141-4151. 2017 sep; 8(9):4141–4151. doi: 10.1364/BOE.8.004141.

Macneil L, Walhout A. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. Genome research. 2011; 21(5):645–657. doi: 10.1101/GR.097378.109.

Majo F, Rochat A, Nicolas M, Jaoudé A, Barrandon Y. Oligopotent stem cells are distributed throughout the mammalian ocular surface. Nature. 2008 nov; 456(7219):250–254. doi: 10.1038/NATURE07406.

Menzel-Severing J, Zenkel M, Polisetti N, Sock E, Wegner M, Kruse FE, Schlötzer-Schrehardt U. Transcription factor profiling identifies Sox9 as regulator of proliferation and differentiation in corneal epithelial stem/progenitor cells. Scientific Reports 2018 8:1. 2018 7; 8:1–18. doi: 10.1038/s41598-018-28596-3.

Mitchell P, Tjian R. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science (New York, NY). 1989; 245(4916):371–378. doi: 10.1126/SCIENCE.2667136.

Okita K, Nakagawa M, Hyenjong H, Ichisaka T, Yamanaka S. Generation of mouse induced pluripotent stem cells without viral vectors. Science (New York, NY). 2008 nov; 322(5903):949–953. doi: 10.1126/SCIENCE.1164270.

O'Shields B, McArthur A G, Holowiecki A, Kamper M, Tapley J, Jenny M J. Inhibition of endogenous MTF-1 signaling in zebrafish embryos identifies novel roles for MTF-1 in development. Biochimica et biophysica acta. 2014; 1843:1818–1833. doi: 10.1016/J.BBAMCR.2014.04.015.

Penter L, Dietze K, Bullinger L, Westermann J, Rahn HP, Hansmann L. FACS single cell index sorting is highly reliable and determines immune phenotypes of clonally expanded T cells. European Journal of Immunology. 2018 jul; 48(7):1248–1250. doi: 10.1002/EJI.201847507.

Pierce SE, Granja JM, Greenleaf WJ. High-throughput single-cell chromatin accessibility CRISPR screens enable unbiased identification of regulatory networks in cancer. Nature Communications 2021 12:1. 2021 may; 12(1):1–8. doi: 10.1038/s41467-021-23213-w.

Polisetti N, Giessl A, Zenkel M, Heger L, Dudziak D, Naschberger E, Stich L, Steinkasserer A, Kruse FE, Schlötzer-Schrehardt U. Melanocytes as emerging key players in niche regulation of limbal epithelial stem cells. The Ocular Surface. 2021 oct; 22:172–189. doi: 10.1016/J.JTOS.2021.08.006.

Rama P, Matuska S, Paganoni G, Spinelli A, De Luca M, Pellegrini G. Limbal Stem-Cell Therapy and Long-Term Corneal Regeneration. http://dxdoiorgproxylibraryuunl/101056/NEJMoa0905955. 2010 jul; 363(2):147–155. doi: 10.1056/NEJMOA0905955.

Sadl V, Jin F, Yu J, Cui S, Holmyard D, Quaggin S, Barsh G, Cordes S. The mouse Kreisler (Krml1/MafB) segmentation gene is required for differentiation of glomerular visceral epithelial cells. Developmental biology. 2002; 249:16–29. doi: 10.1006/DBIO.2002.0751.

Simkin J, Gawriluk T, Gensel J, Seifert A. Macrophages are necessary for epimorphic regeneration in African spiny mice. eLife. 2017 may; 6. doi: 10.7554/ELIFE.24623.

Singh K, Camera E, Krug L, Basu A, Pandey R K, Munir S, Wlaschek M, Kochanek S, Schorpp-Kistner M, Picardo M, Angel P, Niemann C, Maity P, Scharffetter-Kochanek K. JunB defines functional and structural integrity of the epidermo-pilosebaceous unit in the skin. Nature communications. 2018 12; 9. doi: 10.1038/S41467-018-05726-,Z.

Skowronska-Krawczyk D, Zhao L, Zhu J, Weinreb R N, Cao G, Luo J, Flagg K, Patel S, Wen C, Krupa M, Luo H, Ouyang H, Lin D, Wang W, Li G, Xu Y, Li O, Chung C, Yeh E, Jafari M, et al. P16INK4a Upregulation Mediated by SIX6 Defines Retinal Ganglion Cell Pathogenesis in Glaucoma. Molecular cell. 2015 9; 59:931–940. doi: 10.1016/J.MOLCEL.2015.07.027.

Sosnová-Netuková M, Kuchynka P, Forrester V John. The suprabasal layer of corneal epithelial cells represents the major barrier site to the passive movement of small molecules and trafficking leukocytes. British Journal of Ophthalmology. 2007 mar; 91(3):372–378. doi: 10.1136/BJO.2006.097188.

Stanescu D, Iseli H P, Schwerdtfeger K, Ittner L M, Remé C E, Hafezi F. Continuous expression of the homeobox gene Pax6 in the ageing human retina. Eye (London, England). 2007; 21:90–93. doi: 10.1038/SJ.EYE.6702166.

Stephens DN, Klein RH, Salmans ML, Gordon W, Ho H, Andersen B. The Ets Transcription Factor EHF as a Regulator of Cornea Epithelial Cell Identity. The Journal of Biological Chemistry. 2013 11; 288:34304. doi: 10.1074/JBC.M113.504399.

Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, Purdom E, Dudoit S. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics 2018 19:1. 2018 jun; 19(1):1–16. doi: 10.1186/S12864-018-4772-0.

Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck W, Hao Y, Stoeckius M, Smibert P, Satija R. Comprehensive Integration of Single-Cell Data. Cell. 2019 jun; 177(7):1888–1902.e21. doi: 10.1016/J.CELL.2019.05.031.

Tian F, Zhao J, Bu S, Teng H, Yang J, Zhang X, Li X, Dong L. KLF6 Induces Apoptosis in Human Lens Epithelial Cells Through the ATF4-ATF3-CHOP Axis. Drug design, development and therapy. 2020; 14:1041–1055. doi: 10.2147/DDDT.S218467.

Ullah E, Wu D, Madireddy L, Lao R, Ling-Fung Tang P, Wan E, Bardakjian T, Kopinsky S, Kwok P Y, Schneider A, Baranzini S, Ansar M, Slavotinek A. Two missense mutations in SALL4 in a patient with microphthalmia, coloboma, and optic nerve hypoplasia. Ophthalmic genetics. 2017 7; 38:371–375. doi: 10.1080/13816810.2016.1217550.

Van Buskirk EM. The anatomy of the limbus. Eye 1989 3:2. 1989; 3(2):101–108. doi: 10.1038/EYE.1989.16.

Webb CF, Bryant J, Popowski M, Allred L, Kim D, Harriss J, Schmidt C, Miner CA, Rose K, Cheng HL, Griffin C, Tucker PW. The ARID Family Transcription Factor Bright Is Required for both Hematopoietic Stem Cell and B Lineage Development. Molecular and Cellular Biology. 2011 mar; 31(5):1041–1053. doi: 10.1128/MCB.01448-10.

Xie Q, Ung D, Khafizov K, Fiser A, Cvekl A. Gene regulation by PAX6: Structural-functional correlations of missense mutants and transcriptional control of Trpm3/miR-204. Molecular Vision. 2014 mar; 20:270.

Xu J, Liu H, Lan Y, Jiang R. Cis-Repression of Foxq1 Expression Affects Foxf2-Mediated Gene Expression in Palate Development. Frontiers in cell and developmental biology. 2021 4; 9. doi: 10.3389/FCELL.2021.665109.

Yamada K, Young RD, Lewis PN, Shinomiya K, Meek KM, Kinoshita S, Caterson B, Quantock AJ. Mesenchymal–epithelial cell interactions and proteoglycan matrix composition in the presumptive stem cell niche of the rabbit corneal limbus. Molecular Vision. 2015 dec; 21:1328.

Yoshida N, Yoshida S, Araie M, Handa H, Nabeshima Y. Ets family transcription factor ESE-1 is expressed in corneal epithelial cells and is involved in their differentiation. Mechanisms of development. 2000 10; 97:27–34. doi: 10.1016/S0925-4773(00)00419-6.

Zhang L, Anderson MC, Liu CY. The role of corneal stroma: A potential nutritional source for the cornea. Journal of nature and science. 2017 aug; 3(8).

Zhenhua X, Takeshi Y, Lijuan W, Debasish M, Liudmila C, Elia J D. Transcription Factor MEF2C Suppresses Endothelial Cell Inflammation via Regulation of NF-,B and KLF2. Journal of cellular physiology. 2015 6; 230:1310. doi: 10.1002/JCP.24870.

Zhou S, Huang Ye, Liu H, Zhou X, Yuan M, Hou F, Wang L, Jiang W. Single-cell RNA-seq dissects the intratumoral heterogeneity of triple-negative breast cancer based on gene regulatory networks. Molecular Therapy - Nucleic Acids. 2021 mar; 23:682–690. doi: 10.1016/J.OMTN.2020.12.018.

Zhu Y T, Han B, Li F, Chen S Y, Tighe S, Zhang S, Tseng S C. Knockdown of both p120 catenin and Kaiso promotes expansion of human corneal endothelial monolayers via RhoA-ROCK-noncanonical BMP-NFB pathway. Investigative ophthalmology visual science. 2014; 55:1509–1518. doi: 10.1167/IOVS.13-13633.

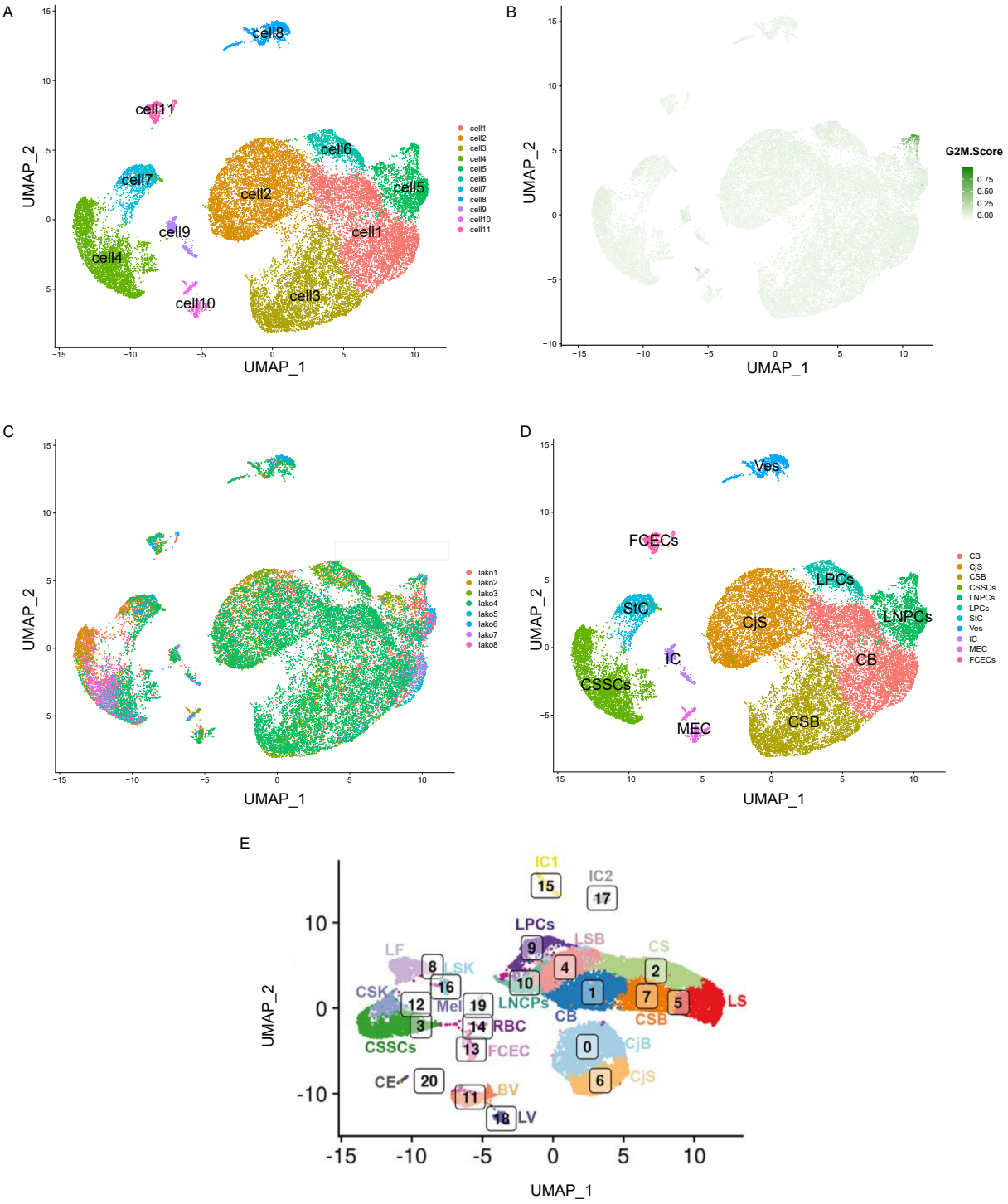


Figure 1. scRNA-seq of the single-cell corneal atlas populations. (A) UMAP of unannotated single-cell clusters from the corneal atlas. (B) Cell cycling score (G2M score) of cells on the UMAP. (C) UMAP with the different scRNA-seq experiment replicates plotted. (D) Annotated UMAP of single-cell clusters from the corneal atlas. (E) Annotated UMAP of single-cell clusters from the corneal atlas used in the study of ***Collin et al.*** (***2021***).

Abbreviations for panel 1D: limbal progenitor cells (LPCs), limbal neural crest derived progenitor cells (LNPCs), corneal basal epithelium (CB), corneal suprabasal epithelium (CSB), melanocytes and endothelial cells (MEC), conjunctival superficial epithelium (CjS), immune cells (IC), blood and lymph vessels (Ves), fibroblastic corneal endothelial cells (FCECs), central stromal stem cells (CSSCs), stromal cells (StC).

Abbreviations for panel 1E not present in 1D: corneal stroma keratocytes (CSK), blood vessels (BV), corneal superficial epithelium (CS), corneal endothelium (CE), limbal fibroblasts (LF), limbal suprabasal epithelial cells (LSB), limbal stroma keratocytes (LSK), limbal superficial epithelium (LS), lymphatic vessel (LV), melanocytes (Mel) and red blood cells (RBC).

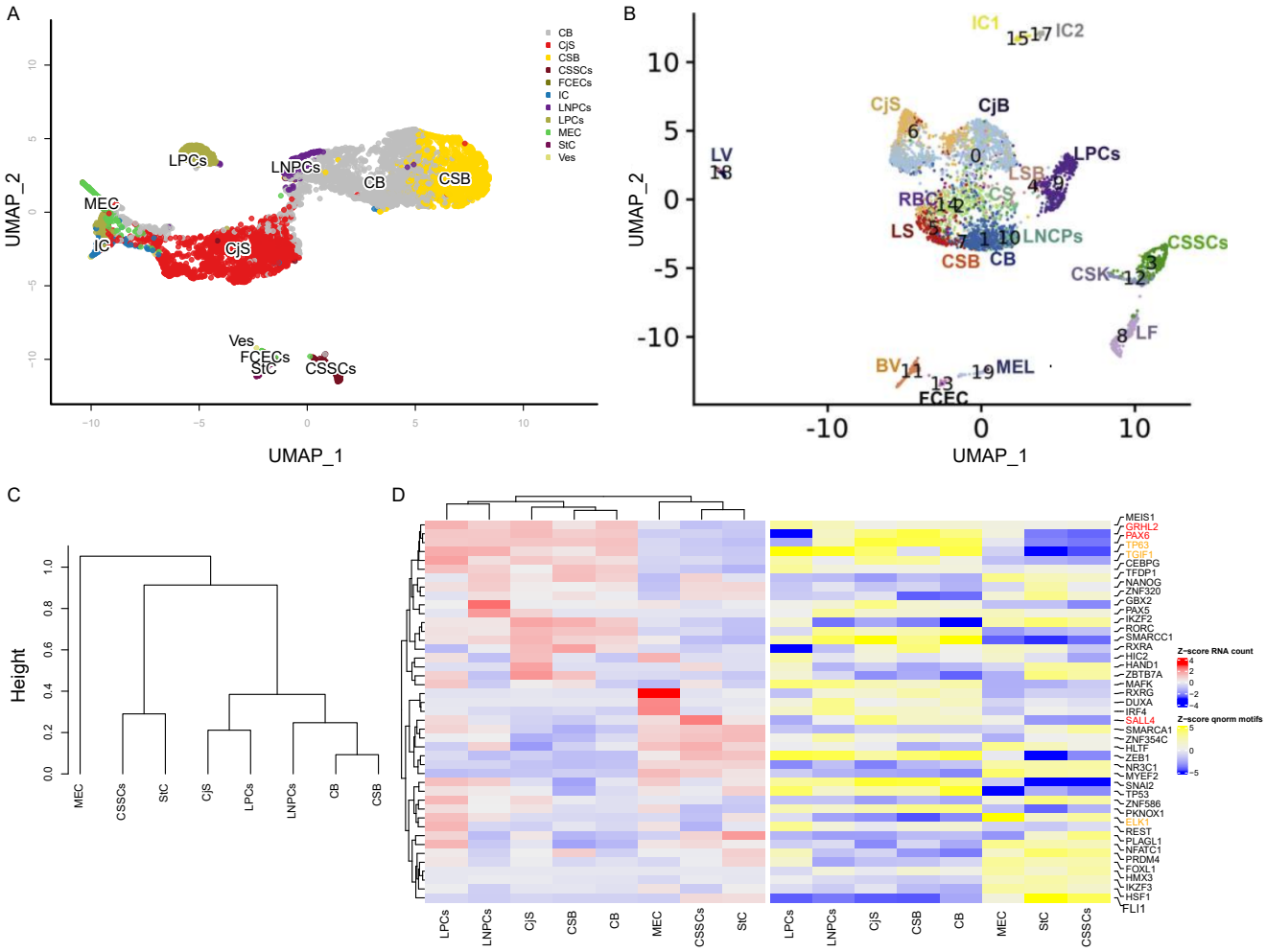


Figure 2. scATAC-seq of the single-cell corneal atlas populations and motif analysis. (A) UMAP of scATAC-seq based on accessibility of variable genes in scRNA-seq. (B) UMAP of scATAC-seq based on accessibility of variable genes in scRNA-seq from ***Collin et al.*** (***2021***). (C) Clustering dendrogram of scATAC-seq showing single-cell corneal atlas populations. (D) Heatmaps of relative gene expression (Z-score RNA count) and relative motif accessibility of the top 10 transcription factors and motifs associated with these transcription factors for each single-cell population. Red colors indicate transcription factors that have been found in literature to be associated with the human eye and/or cornea. Yellow colors indicate an association with the epidermal development.

Abbreviations for panel 2A: limbal progenitor cells (LPCs), limbal neural crest derived progenitor cells (LNPCs), corneal basal epithelium (CB), corneal suprabasal epithelium (CSB), melanocytes and endothelial cells (MEC), conjunctival superficial epithelium (CjS), immune cells (IC), blood and lymph vessels (Ves), fibroblastic corneal endothelial cells (FCECs), central stromal stem cells (CSSCs), stromal cells (StC).

Abbreviations for panel 2B not present in 2A: corneal stroma keratocytes (CSK), blood vessels (BV), corneal superficial epithelium (CS), corneal endothelium (CE), limbal fibroblasts (LF), limbal suprabasal epithelial cells (LSB), limbal stroma keratocytes (LSK), limbal superficial epithelium (LS), lymphatic vessel (LV), melanocytes (Mel) and red blood cells (RBC).

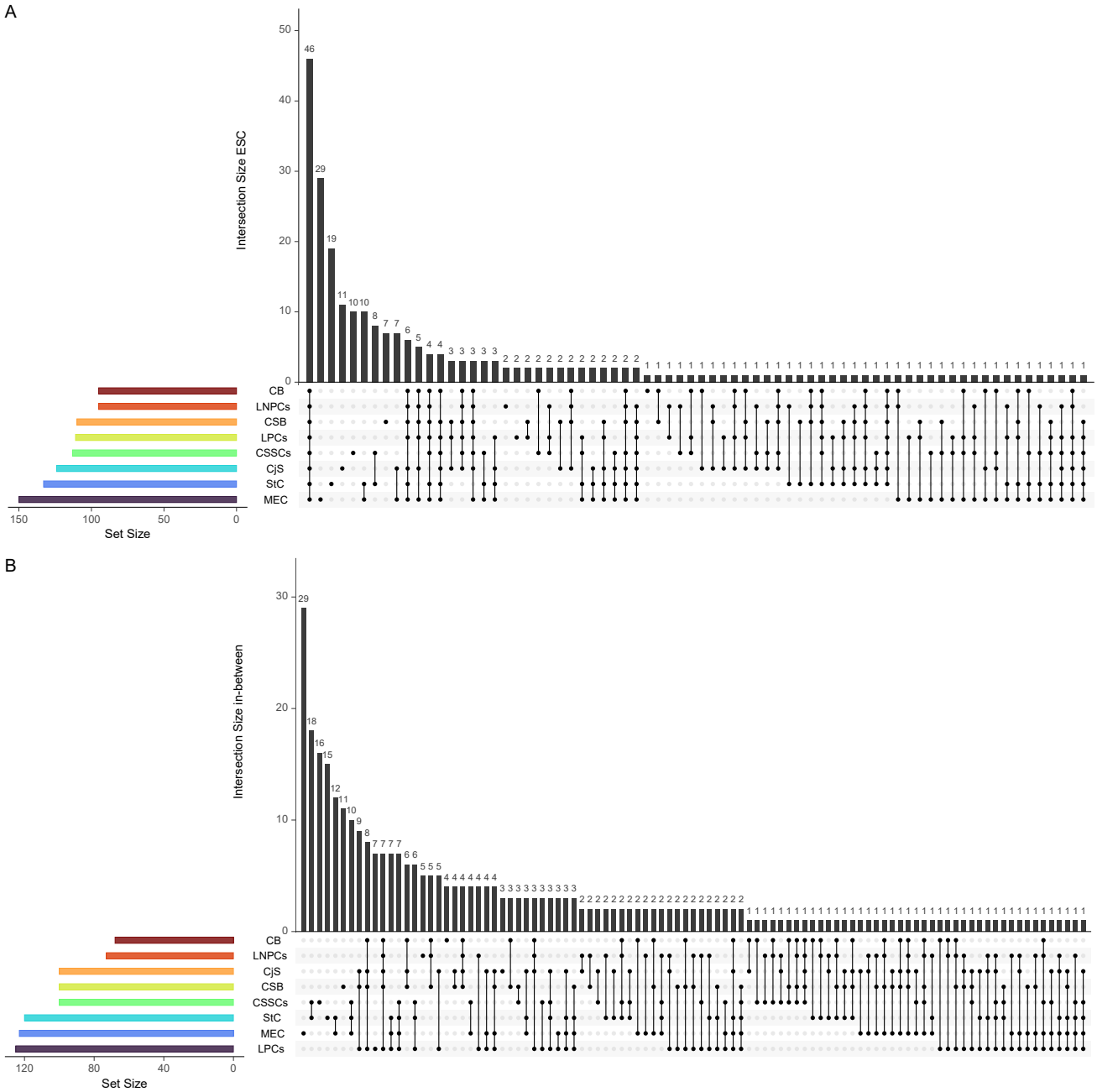


Figure 3. Shared transcription factors in gene regulatory networks in-between single-cell corneal atlas populations. (A) Interaction plot showing the number of overlapping and unique factors with intersection size in single-cell corneal atlas populations with the comparison to embryonic stem cells. Set Size indicates the total number of transcription factors found influencing the GRN for each single-cell population. (B) Interaction plot showing overlapping and unique factors in single-cell corneal atlas populations with the median approach of ANANSE.

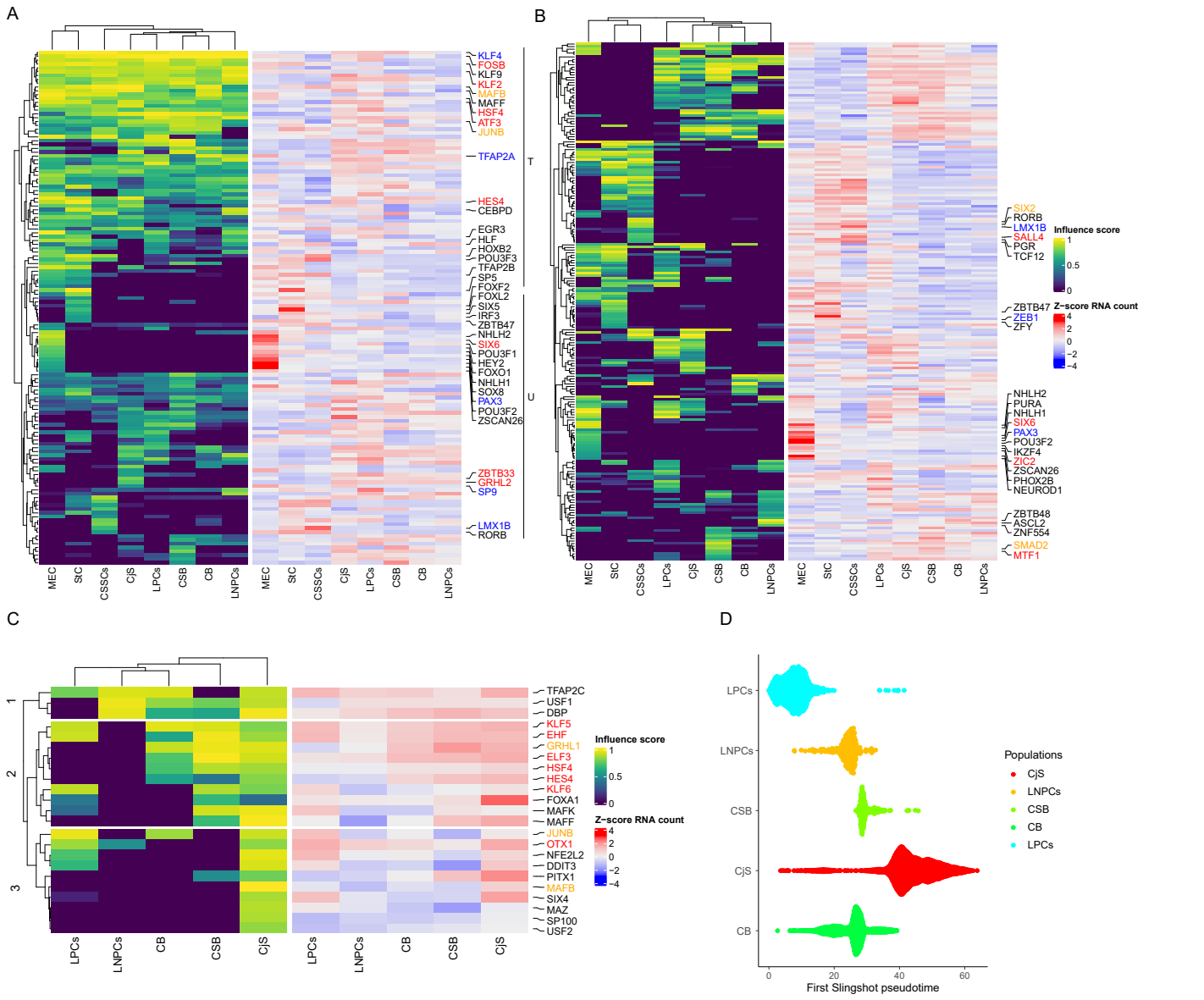


Figure 4. Unique transcription factors in gene regulatory networks of single-cell populations and predicted pseudo time of corneal differentiation. (A) Heatmaps of the influence score (>0.5) and relative gene expression (Z-score RNA count) from transcription factors influencing GRNs of the single-cell corneal atlas populations when compared to embryonic stem cells. (B) Heatmaps of the influence score (>0.5) and relative gene expression (Z-score RNA count) from transcription factors influencing GRNs of the single-cell corneal atlas populations when using the in-between ANANSE comparison. (A-B) Transcription factors found to be present in only one single-cell population are marked at the side. Marked transcription factors with a "U" indicate that transcription factors were found in one single population. Marked transcription factors with a "T" indicate that they were found in the top 20 of single-cell populations. (C) Sub-selection of 4B based on shared factors between the stromal cell clusters against the limbal, corneal and conjunctival cell clusters. This shows transcription factors driving the difference between CSSCs and StC against LPCs, LNPCs, CB, CSB and CjS. (A-C) Red colors indicate transcription factors that have been found in literature to be associated with the human eye. Yellow colors indicate an association with the epidermis. Blue colors indicate that a transcription factor is involved in regulating one or a multitude of the single-cell clusters. (D) Single-cell populations: LPCs, LNPCs, CB, CSB and CjS ordered according to Slingshot pseudo time. Each dot is representing one cell from scRNA-seq.

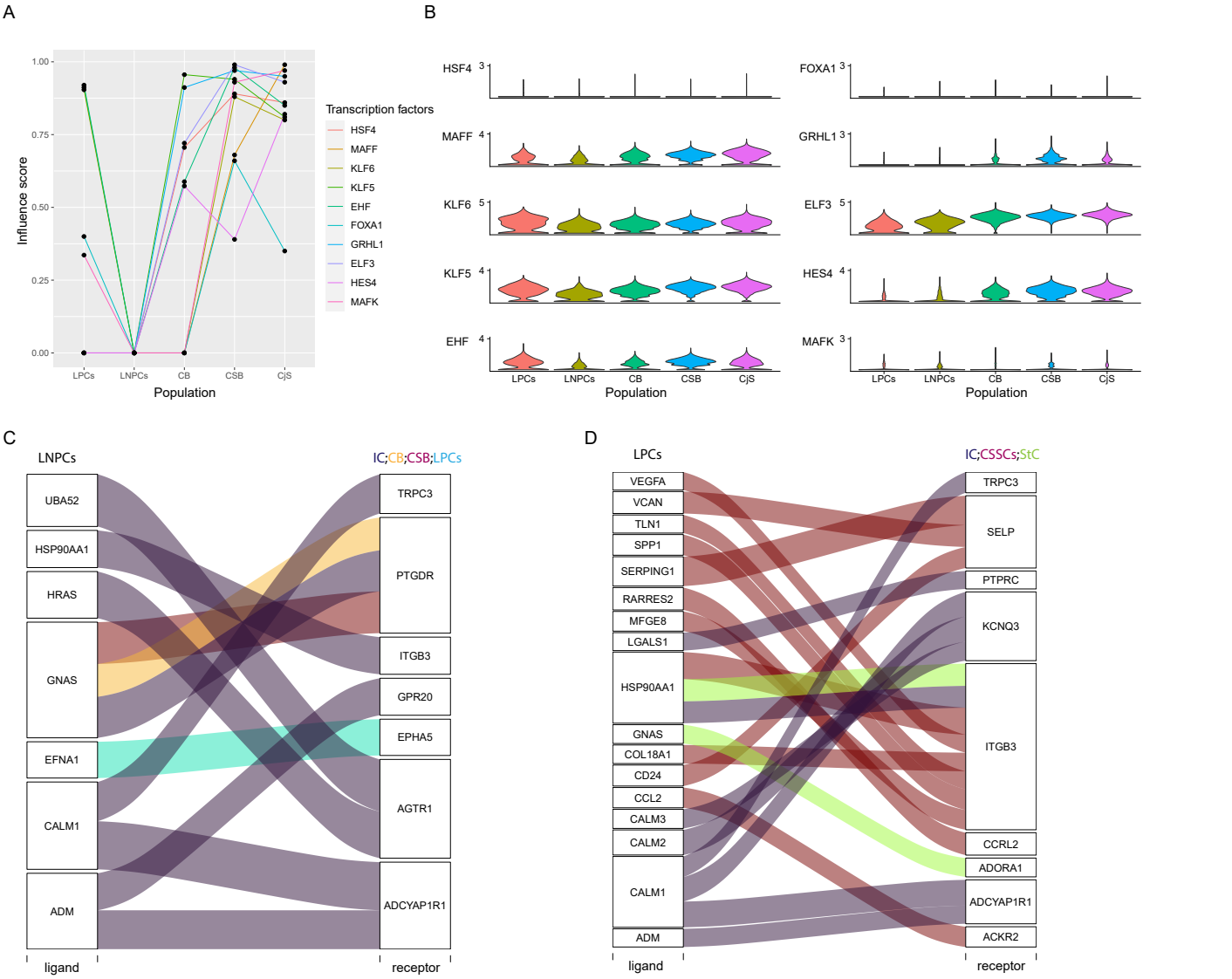
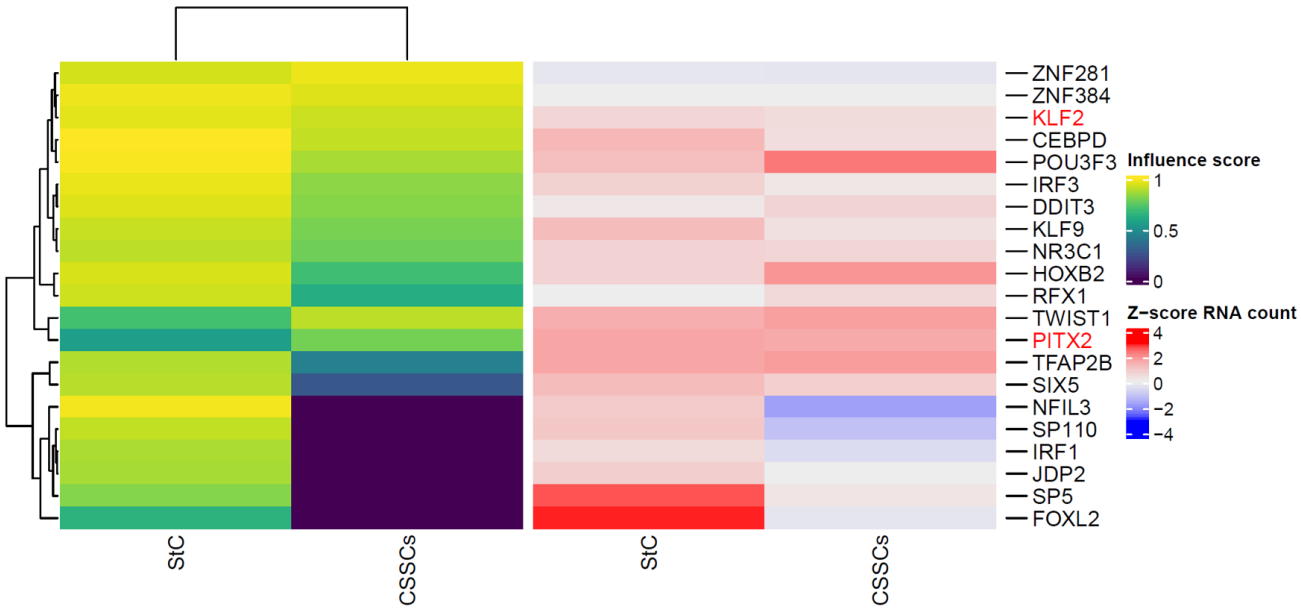


Figure 5. Differences of important transcription factors over pseudo time and paracrine interactions of limbal progenitor cell populations. (A) Line plots with influence score of transcription factors with the highest influence over predicted pseudotime. (B) Violin plots of normalized gene expression of genes from single-cell populations LPCs, LNPCs, CB, CSB and CjS. (C) Paracrine interactions of LNPCs with IC, CB, CSB and LPCs based on scRNA-seq. Purple indicates receptors upregulated in IC. Yellow indicates receptors upregulated in CB. Brown indicates receptor upregulated in CSB. Azure shows the upregulation of one receptor in LPCs. (D) All paracrine interactions of LPCs with IC, CSSCs and StC based on scRNA-seq. Purple indicates receptors upregulated in IC. Brown indicates receptors upregulated in CSSCs. Green shows the upregulation of receptors in StC.

Supplementary Figure 1. UCSC genome browser screenshots of gene accessibility of marker genes in integrated scATAC-seq populations. (A-D) Bigwig files from bam files are indicated on the left for each track. Peaks correspond to overall accessibility (A) Bigwig tracks of S100A9, a marker gene for CjS. (B) Bigwig tracks of MMP3, a marker gene for CSSCs. (C) Bigwig tracks of S100A2, a marker gene for LPCs. (D) Bigwig tracks of CDH19, a marker gene for endothelial cells.



Supplementary Figure 2. Distinct transcription factors found in cell populations with a stromal cell fate. Heatmaps of the influence score (>0.5) and relative gene expression (Z-score RNA count) from the top 25 transcription factors influencing GRNs of the single-cell corneal atlas populations with a stromal cell identity.

Table 1. Literature annotated transcription factors associated with the eye

TF	Tissue	Function	Associated disease	Organism	Reference
FOSB	Eye	Inflammation	None	Rattus norvegicus	***Fujimoto et al.*** (***2004***)
KLF2	Eye	Leukocyte adhesion	None	Mus musculus	***Zhenhua et al.*** (***2015***)
HSF4	Eye	Inflammation	None	Rattus norvegicus	***Fujimoto et al.*** (***2004***)
ATF3	Eye	Retinal protection	None	Mus musculus	***Kole et al.*** (***2020***)
HES4	Eye	Retinal devlopment	None	D. Rerio	***Coomer C et al.*** (***2020***)
SIX6	Eye	RGC senescence	Glaucoma	H. Sapiens	***Skowronska-Krawczyk et al.*** (***2015***)
ZBTB33	Eye	Corneal endothelium proliferation and EMT	None	H. Sapiens	***Zhu Y et al.*** (***2014***)
GRHL2	Eye	Maintenance corneal endothelial cell state	Posterior polymorphous corneal dystrophy	H. Sapiens	***Chung D et al.*** (***2019***)
SALL4	Eye	Early eye development	Coloboma	H. Sapiens	***Ullah et al.*** (***2017***)
ZIC2	Eye	Axonal refinement RGCs	None	Mus musculus	***García-Frigola and Herrera*** (***2010***)
MTF1	Eye	Early eye development	None	D. Rerio	***O'Shields et al.*** (***2014***)
KLF6	Eye	Apoptosis lens cells	None	H. Sapiens	***Tian et al.*** (***2020***)
OTX1	Eye	Unknown	None	Mus musculus	***Alldredge and Fuhrmann*** (***2016***)
PITX2	Eye	Eye development	Axenfeld-Rieger syndrome	D. Rerio	***Hendee K et al.*** (***2018***)

Table 2. Literature annotated transcription factors specific to corneal atlas cell populations

TF	Tissue	Function	Associated disease	Organism	Reference
KLF4	Cornea	EMT repression corneal epithelium	None	H. Sapiens	***Fujimoto et al.*** (***2004***)
TFAP2A	Neural crest cells	Early neural crest development	None	D. Rerio	***Knight et al.*** (***2003***)
PAX3	Neural crest cells	Sacral neural crest development	Spina bifida	Mus musculus	***Deal K et al.*** (***2021***)
SP9	Neural crest cells	Unknown	None	D. Rerio	***Lumb et al.*** (***2017***)
LMX1B	Stromal cells	Sacral neural crest development	None	D. Rerio	***Liu and Johnson R*** (***2010***)
ZEB1	Corneal endothelium	Regulation of EMT	Posterior polymorphous corneal dystrophy	H. Sapiens	***Guha et al.*** (***2017***)
KLF5	Cornea	Promoting cornea differentiation	None	Mus Musculus	***Gupta et al.*** (***2011***)
EHF	Cornea	Regulator cornea epithelium	None	Mus Musculus	***Stephens et al.*** (***2013***)
PAX6	Cornea	Eye development and corneal homeostatsis	Aniridia	H. Sapiens	***Stanescu et al.*** (***2007***)
TP63	Limbal progenitor cells	Unknown	None	H. Sapiens	***Chen S et al.*** (***2011***)
TFAP2C	Neural crest cells	Early neural crest induction	None	D. Rerio	***Li and Cornell R*** (***2007***)
DBP	Neural crest cells	Early neural crest development	None	Mus Musculus	***Webb et al.*** (***2011***)

Table 3. Literature annotated transcription factors associated with the epidermis

TF	Tissue	Function	Associated disease	Organism	Reference
TGIF1	Epidermis	Negative regulator of epidermal differentiation	None	H. Sapiens	***Barbollat-Boutrand et al.*** (***2017***)
ELK1	Epidermis	Angiogenesis	Skin fibrosis	H. Sapiens	***Gao et al.*** (***2019***)
MAFB	Epidermis	Epidermal differentiation	None	Mus musculus	***Sadl et al.*** (***2002***)
JUNB	Epidermis	Skin homeostasis	None	Mus musculus	***Singh et al.*** (***2018***)
SIX2	Epidermis	Palatogenesis	Cleft palate	Mus musculus	***Boglev et al.*** (***2011***)
SMAD2	Epidermis	Basal keratinocyte migration	None	H. Sapiens	***Hosokawa et al.*** (***2005***)
GRHL1	Epidermis	Epidermal development	None	Mus musculus	***Boglev et al.*** (***2011***)