



## Data Article

# Single cell RNA-sequencing data generated from human pluripotent stem cell-derived lens epithelial cells



Rachel Shparberg<sup>a</sup>, Chitra Umala Dewi<sup>a</sup>,  
Vikkitharan Gnanasambandapillai<sup>b</sup>, Liwan Liyanage<sup>c</sup>,  
Michael D. O'Connor<sup>a,\*</sup>

<sup>a</sup> School of Medicine, Western Sydney University, Campbelltown, NSW 2560, Australia

<sup>b</sup> Garvan-Weizmann Centre for Cellular Genomics, Garvan Institute, UNSW Cellular Genomics Futures Institute, University of New South Wales, Sydney 2010, Australia

<sup>c</sup> School of Computer, Data and Mathematical Sciences, Western Sydney University, Campbelltown, NSW 2560, Australia

## ARTICLE INFO

## Article history:

Received 29 October 2020

Revised 9 December 2020

Accepted 10 December 2020

Available online 8 January 2021

## Keywords:

Human lens epithelial cell

ROR1

Pluripotent stem cell

Micro-lens

Cataract

Single cell RNA-sequencing

Bioinformatics

## ABSTRACT

Detailed transcriptomic analyses of differentiated cell populations derived from human pluripotent stem cells is routinely used to assess the identity and utility of the differentiated cells. Here we provide single cell RNA-sequencing data obtained from ROR1-expressing lens epithelial cells (ROR1e LECs), obtained via directed differentiation of CA1 human embryonic stem cells. Analysis of the data using principal component analysis, heat maps and gene ontology assessments revealed phenotypes associated with lens epithelial cells. These data provide a resource for future characterisation of both normal and cataractous human lens biology. Corresponding morphological and functional data obtained from ROR1e LECs are reported in the associated research article "A simplified method for producing human lens epithelial cells and light-focusing micro-lenses from pluripotent stem cells" (Dewi et al., 2020).

DOI of original article: [10.1016/j.exer.2020.108317](https://doi.org/10.1016/j.exer.2020.108317)

\* Corresponding author.

E-mail address: [m.oconnor@westernsydney.edu.au](mailto:m.oconnor@westernsydney.edu.au) (M.D. O'Connor).

<https://doi.org/10.1016/j.dib.2020.106657>

2352-3409/© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

## Specifications Table

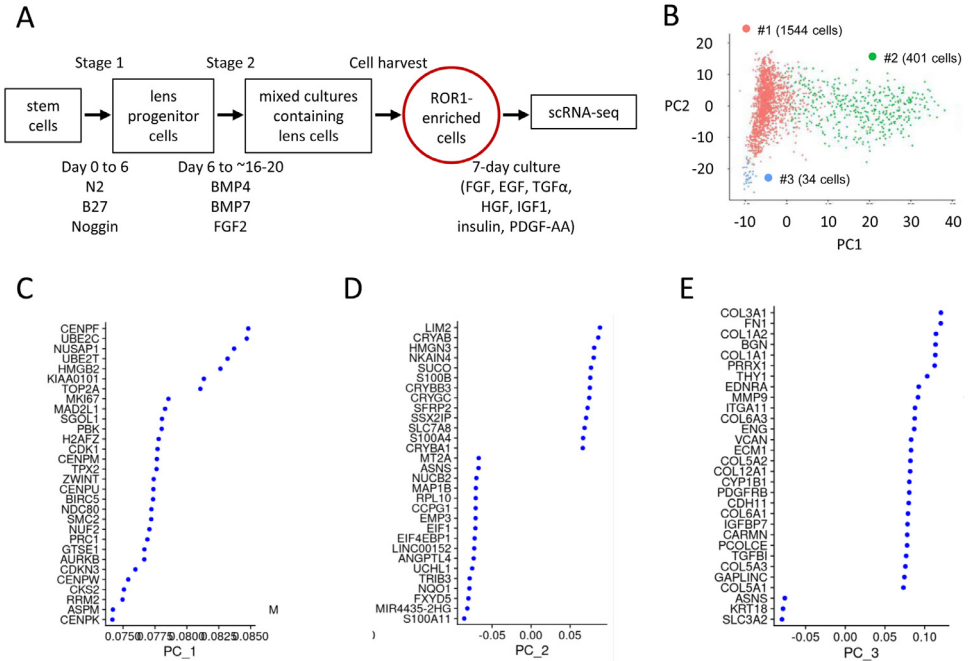
Subject	Biology
Specific subject area	Molecular Biology, Cell Biology, Human Pluripotent Stem Cells, Lens Epithelial Cells, Single Cell Transcriptomics
Type of data	Table Figure
How data were acquired	Stem cell culture and differentiation. RNA-Seq (10X Genomics single cell 3' mRNA-prep kit and Illumina NextSeq 500). Transcriptomic data analyses (Seurat guided clustering suite version 3; Bioconductor R package; DAVID gene ontology).
Data format	Raw fastq files
Parameters for data collection	Human pluripotent stem cells were differentiated to lens epithelial cells using established methods (Murphy et. al, 2018). ROR1e LECs were harvested on day 16 and expanded for a further 7 days in optimised medium (Murphy et. al, 2018). Cells were collected 7 days after harvest and immediately processed for scRNA-seq.
Description of data collection	Total RNA was isolated from ROR1e LECs. cDNA library was constructed and scRNA-seq performed.
Data source location	Bioinformatics performed using Seurat guided clustering suite version 3. School of Medicine, Western Sydney University, Campbelltown, Australia
Data accessibility	ArrayExpress accession number E-MTAB-9178 <a href="https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-9178/">https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-9178/</a>
Related research article	Umala Dewi et al, 2020, A simplified method for production of human lens epithelial cells and light-focusing micro-lenses from pluripotent stem cells, Exp. Eye Res. Experimental Eye Research. <a href="https://doi.org/10.1016/j.exer.2020.108317">doi.org/10.1016/j.exer.2020.108317</a> .

## Value of the Data

- These single cell RNA-sequencing profiles, obtained from lens epithelial cells derived from human embryonic stem cells, provide insights into the transcriptional heterogeneity of human lens cell cultures.
- Exploration of these data will provide molecular insights into regulation of gene expression in human lens epithelial cells that will be of use to researchers investigating lens and cataract development.
- These data can be further analysed to better understand lens transcriptional regulators, and to provide molecular hypotheses to guide functional investigations of normal lens biology and cataract initiation.

## 1. Data Description

Here we report initial scRNA-seq data analyses for human, pluripotent stem cell-derived, ROR1e LECs. These LECs were generated from CA1 human embryonic stem cells as outlined in Fig. 1. The culture protocol included 16 days guided differentiation of the stem cells to lens cells, followed by harvesting of ROR1e LECs and then an additional 7-days culture in an optimised LEC medium [1]. Phenotypic characterisation of similar ROR1e LEC populations demonstrated: they are morphologically-similar to antibody-purified ROR1<sup>+</sup> LECs; they express expected LEC



**Fig. 1.** Single cell RNA sequencing analysis of ROR1e LECs. (A) Schematic diagram outlining generation of ROR1e LECs from CA1 human embryonic stem cells. (B) PCA analysis revealed three cell clusters within the ROR1e LEC scRNA-seq data. (C-E) Genes that show high cell-cell variability across the data set in the top three principal components.

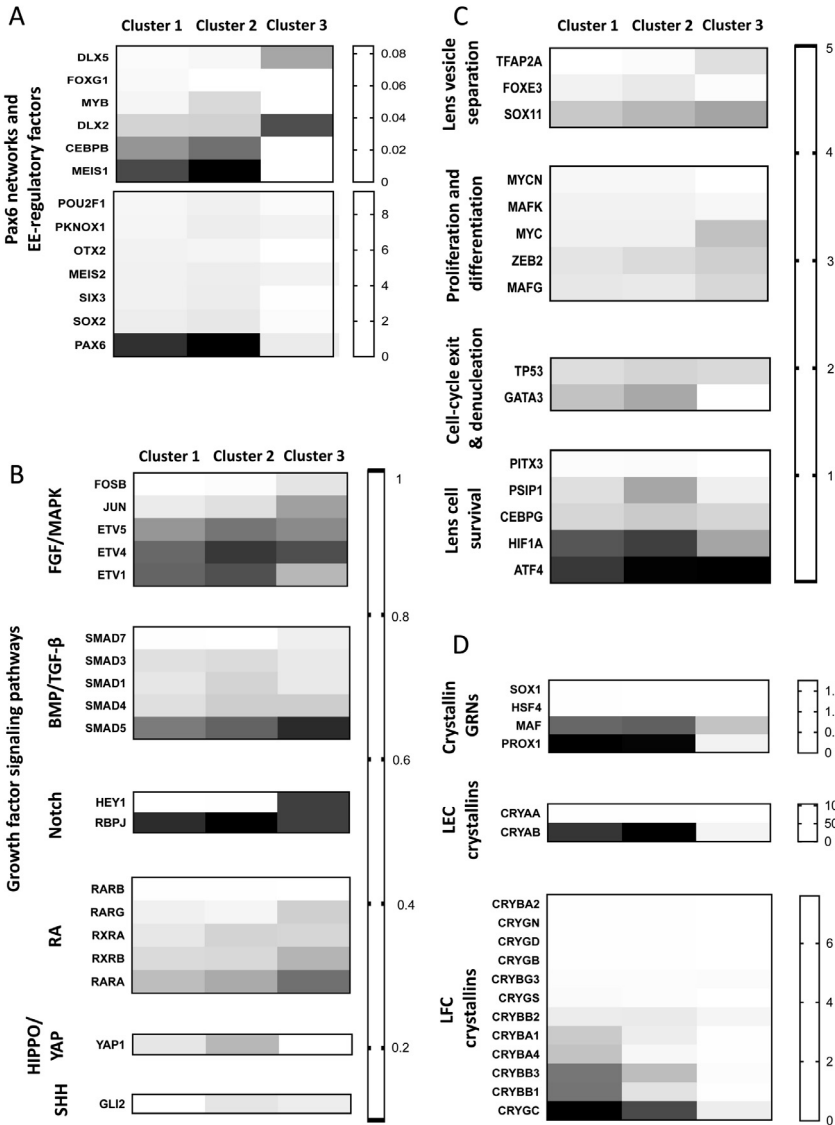
crystallin proteins; and they can be aggregated to generate light-focusing micro-lenses that have similar morphology and protein expression to primary human lenses [2].

### 1.1. Principal component analyses

The ROR1e LECs were dissociated immediately prior to processing for scRNA-seq. RNA was extracted from the cultured LECs and processed using a 10X Genomics single cell 3' mRNA-prep kit. Sequencing was performed using an Illumina NextSeq 500 sequencer. The scRNA-seq data can be accessed via Array Express accession E-MTAB-9178. Analysis using the Seurat guided clustering suite yielded 1799 cells, with a total of 17,944 genes expressed. Assessment via principal component analysis (PCA) revealed three distinct cell clusters (Fig. 1B). Together, clusters 1 and 2 accounted for 98.3% of the total cell population (78% and 20.3%, respectively). Cluster 3 accounted for 1.7% of the total cell population. The main genes responsible for the observed variation between the cells are shown (Fig. 1C-E).

### 1.2. Heat map analysis of critical lens genes

Heat map analyses were used to investigate expression of critical lens genes in each of the three cell clusters. The initial analysis examined expression of 65 genes known to be required for lens development. This included growth factor signaling genes, crystallin genes and their regulators, proliferation and cell survival genes, and various transcription factors [3] (Fig. 2). Clusters 1 and 2 expressed similar levels of all 65 genes – including high expression of critical lens genes

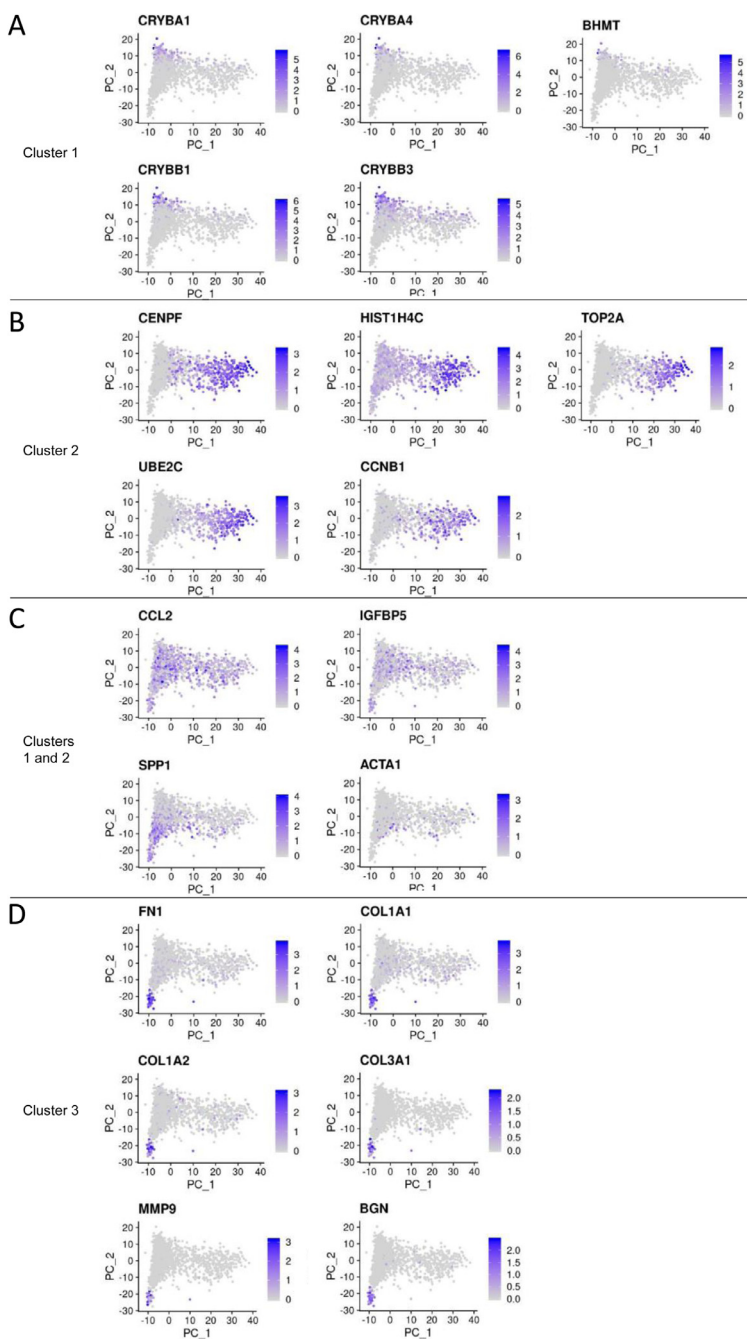


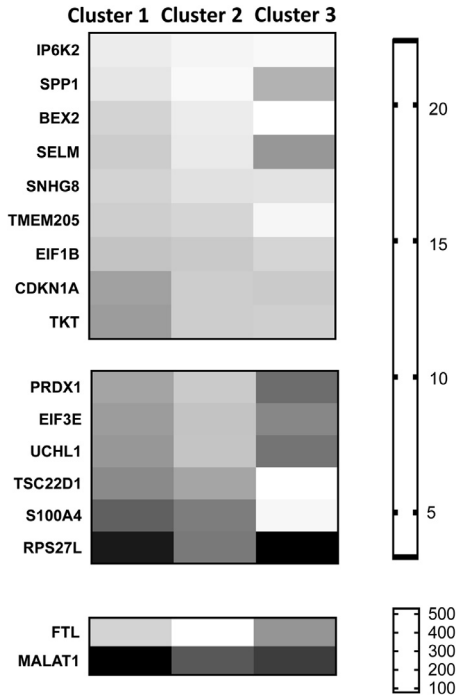
**Fig. 2.** Expression of key lens genes. Heat maps showing average expression per cell in each cluster for important lens-related genes, with gene groupings based on gene function.

such as PAX6 and CRYAB – while some genes had reduced expression in cluster 3. Overall, these data suggest the ROR1e cells have a lens phenotype.

### 1.3. Analysis of the top 20 cluster marker genes

Assessment of the top 20 genes most differentially expressed across the 3 clusters (Fig. 3) showed 70% of these genes (14/20) were more highly expressed in Clusters 1 and/or 2, and 30% (6/20) more highly in Cluster 3. In particular: Cluster 1 had higher expression of crystallin genes





**Fig. 4.** Heat maps of cluster 1 marker genes. The 17 marker genes that were significantly and positively associated with cluster 1 were grouped into three heatmaps based on average gene expression level per cell, then sorted based on lowest to highest expression in cluster 1.

associated with lens fibre cells (Fig. 3A); Cluster 2 had higher expression of some genes associated with cell proliferation such as CENPF, TOP2A, UBE2C and CCNB1 (Fig. 3B); and Cluster 3 had higher expression of some genes associated with epithelial-to-mesenchymal transition (Fig. 3D).

#### 1.4. Analysis of cluster 1 marker genes

Heat maps and GO analyses were then used to assess the 17 genes identified as marker genes for cluster 1. The heat map analysis showed cluster 1 marker genes (Fig. 4) were more similarly expressed in cluster 2 than in cluster 3 – suggesting cluster 1 is similar to cluster 2. No significant GO terms were associated with the cluster 1 marker genes. Comparison with published lens transcriptomic datasets revealed all the cluster 1 marker genes are expressed in adult human lenses [4] and/or mouse lenses [5] – including protein-coding and non-coding genes required for normal lenses (e.g., BEX2, TKT, SNHG8, MALAT1) [6].

#### 1.5. Analysis of cluster 2 marker genes

For the 314 genes identified as markers of cluster 2 (Fig. 5), most were expressed much higher in cluster 2 compared to both clusters 1 and 3 – particularly for lowly-expressed cluster 2 marker genes. GO analysis of the cluster 2 marker genes identified terms relating to cell cycle and mitosis (Table 1). These data indicate cluster 2 contains cells that had higher expression of some proliferation-related genes when the cell population was captured.



**Table 1**

Top 20 significant GO terms associated with Cluster 2 markers.

		GO Term p-value (Benjamini )
GO:1903047	mitotic cell cycle process	1.68E-79
GO:0007067	mitotic nuclear division	3.68E-63
GO:0000280	nuclear division	2.04E-61
GO:0000819	sister chromatid segregation	3.78E-47
GO:0098813	nuclear chromosome segregation	2.49E-43
GO:0044770	cell cycle phase transition	2.41E-41
GO:0044772	mitotic cell cycle phase transition	4.51E-40
GO:0000070	mitotic sister chromatid segregation	4.74E-34
GO:0006259	DNA metabolic process	2.71E-33
GO:0006260	DNA replication	4.01E-29
GO:0007062	sister chromatid cohesion	4.77E-29
GO:0007346	regulation of mitotic cell cycle	2.61E-28
GO:0010564	regulation of cell cycle process	1.36E-27
GO:0071103	DNA conformation change	2.66E-25
GO:0007088	regulation of mitotic nuclear division	2.35E-21
GO:0051783	regulation of nuclear division	1.71E-20
GO:0000226	microtubule cytoskeleton organization	4.20E-20
GO:0006281	DNA repair	9.75E-20
GO:0000075	cell cycle checkpoint	2.74E-17
GO:0065004	protein-DNA complex assembly	5.60E-17

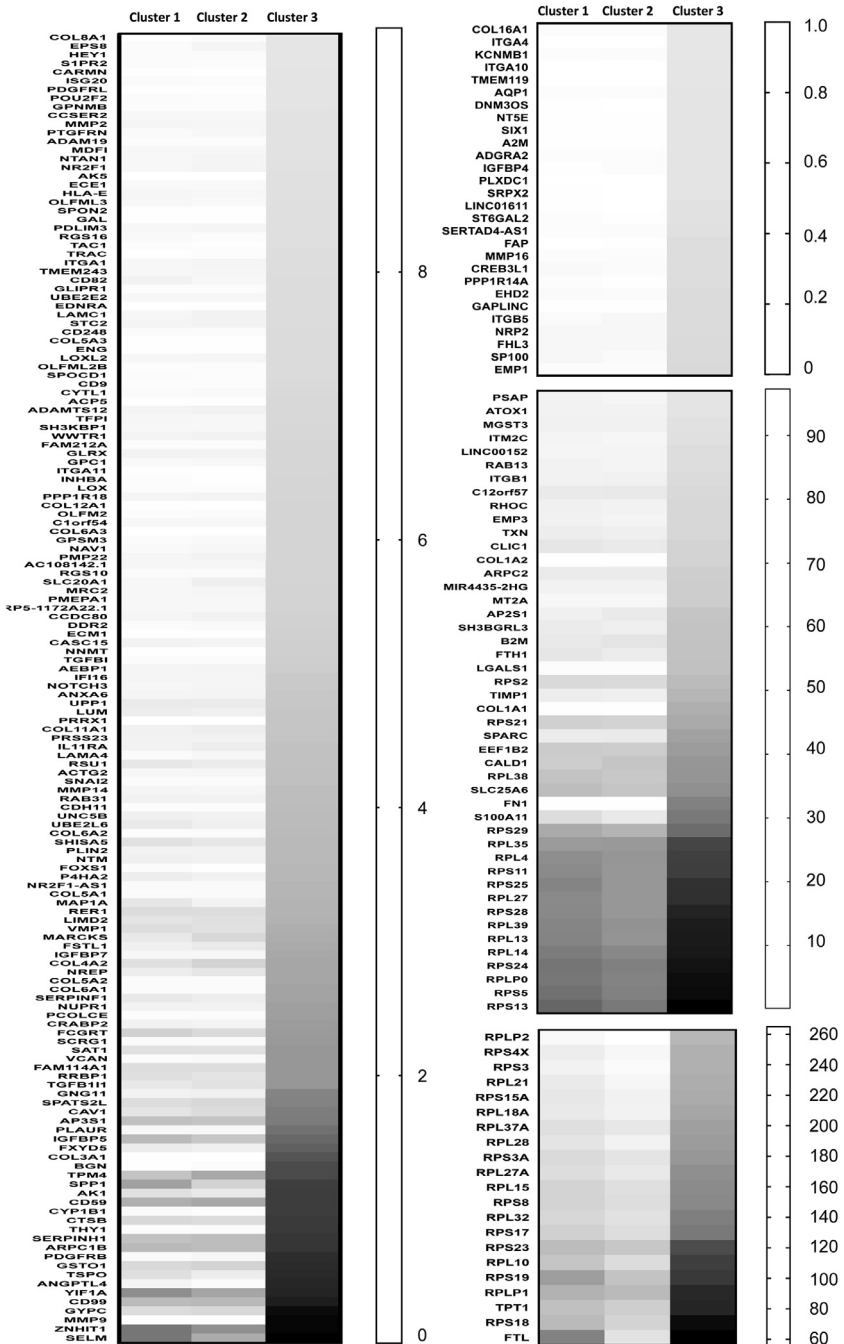
### 1.6. Analysis of cluster 3 marker genes

For the 245 cluster 3 marker genes (Fig. 6), most were detected in clusters 1 and 2 but at lower levels. GO analysis of the cluster 3 marker genes identified terms relating to collagen metabolism and extracellular matrix degradation (Table 2). As the cluster 3 marker gene list includes fibronectin, MMP9 and collagen 1/3 genes (Fig. 3D), it is possible cluster 3 represents cells primed for epithelial-to-mesenchymal transition (though this cluster represents only 1.7% of the total population of ROR1e LECs).

### 1.7. Comparison of GO terms obtained from all expressed genes in each cluster

To gain a more detailed insight into the relative difference in transcriptomes between the three ROR1e LEC clusters, additional GO analyses were performed using the expressed genes for each cluster. When the expressed genes from all three cluster were combined, the GO terms identified were consistent with known lens biology. This includes GO terms relating to establishment of a polarised epithelium (a key feature of primary LECs), and terms related to growth factor signalling pathways required by LECs (e.g., FGF and WNT) [7]. When the expressed genes from each cluster were analysed separately, 15 of the top 20 GO terms were shared across all the clusters [2]. The first 12 of these 15 shared GO terms were the most statistically-significant GO terms for each of the three clusters (excluding generic terms). Strikingly, these shared GO terms related to establishment of a polarised epithelium, morphogenesis of an epithelium, FGF signalling and WNT signalling. A GO term relating to epithelial to mesenchymal transition was identified in the very small cluster 3. Overall these GO analyses are consistent with stem cell-derived ROR1e LECs being a highly purified population of human LEC-like cells. These findings are also consistent with morphological, lens function, proteomic and electron microscopy data obtained using ROR1e LECs [2].





**Fig. 6.** Heat maps of cluster 3 marker genes. The 245 marker genes that were significantly and positively associated with cluster 3 were grouped into four heatmaps based on average gene expression level per cell, then sorted based on lowest to highest expression in cluster 3.

**Table 2**

Top 20 significant GO terms associated with Cluster 3 markers.

		GO Term p-value (Benjamini)
GO:0032963	collagen metabolic process	4.62E-20
GO:0030574	collagen catabolic process	2.52E-16
GO:0001944	vasculature development	5.55E-15
GO:0001568	blood vessel development	6.02E-14
GO:0035987	endodermal cell differentiation	4.33E-13
GO:0072359	circulatory system development	1.00E-12
GO:0072358	cardiovascular system development	1.00E-12
GO:0030199	collagen fibril organization	2.61E-12
GO:0001706	endoderm formation	4.03E-12
GO:0001525	angiogenesis	1.59E-11
GO:0016477	cell migration	1.16E-09
GO:0001501	skeletal system development	4.53E-09
GO:0001704	formation of primary germ layer	7.09E-09
GO:0048513	animal organ development	4.75E-08
GO:0022617	extracellular matrix disassembly	4.57E-07
GO:0007369	gastrulation	8.63E-07
GO:0001649	osteoblast differentiation	4.03E-06
GO:0007167	enzyme linked receptor protein signaling pathway	1.81E-05
GO:0007166	cell surface receptor signaling pathway	1.94E-05
GO:0030334	regulation of cell migration	4.98E-05

## 2. Experimental Design, Materials and Methods

### 2.1. Generation of ROR1e LECs

CA1 human embryonic stem cells were obtained from A. Nagy [8]. Approval for their use was obtained from the Western Sydney University Human Research Ethics Committee (Australia). CA1 cells were maintained on Matrigel-coated plates (Corning, Sydney, Australia) in mTeSR1 (StemCell Technologies, Sydney, Australia) as previously described [9,10]. Generation of heterogeneous differentiation cultures containing lens cells occurred in: DMEM:F12 (Thermo Fisher Scientific, Sydney Australia) containing 500 ng/mL noggin and 10 nM SB431542 (6 days); followed by DMEM:F12 containing 20 ng/mL BMP4, 20 ng/mL BMP7 and 100 ng/mL FGF2 [2]. Growth factors were sourced from Miltenyi Biotec and Peprotech (Sydney, Australia). ROR1e LECs were harvested using TrypLE (Thermo Fisher Scientific), filtered through a 40-micron cell strainer (Interpath Services, Sydney, Australia), centrifuged (300 g, 5 min) and counted, then resuspended in medium containing 10 ng/mL FGF2, 5 ng/mL EGF and TGF $\alpha$ , 10 ng/mL HGF, 10 ng/mL IGF1, 10  $\mu$ g/mL insulin and 10 ng/mL PDGF-AA [1]. After 7 days, ROR1e LECs were harvested using TrypLE and immediately processed for scRNA-seq.

### 2.2. Single cell RNA sequencing

The ROR1e LECs were processed using a 10X Genomics single cell 3' mRNA-prep kit as per the manufacturer instructions (part numbers PN-120267, PN-1000009 and PN-120262). Sequencing was performed by the Ramaciotti Centre for Genomics (Sydney, Australia) using an Illumina NextSeq 500 (NextSeq Control Software v 2.2.0.4 / Real Time Analysis 2.4.11), with 150 cycles as follows: 26bp (Read 1), 98bp (Read 2) and 8bp (Index). Read alignment, filtering, barcode

counting, and UMI counting were performed using CellRanger (v3.1.0) and FASTQ files. The reads were aligned to Genome Reference Consortium Human Build 38.

### 2.3. Bioinformatic analyses

Processed scRNA-seq data were analysed using R and the Seurat guided clustering suite version 3 [11]. The data was filtered to include cells with unique feature counts between 200 and 6000 and with <10% mitochondrial counts. Global-scaling normalization (LogNormalize) was used to normalize the feature expression measurements for each cell. A subset of features exhibiting high cell-to-cell variation was calculated using the module FindVariableFeatures. A standard, pre-processing linear transformation scaling was applied (ScaleData function). The data then underwent linear dimension reduction via PCA using the previously determined variable genes. Marker genes were identified for each cell cluster using ROC analysis and taking the default value of 0.25 log(feature count) threshold-within FindAllMarkers. Identified clusters of genes were visualised using PCA plots where individual genes were overlaid on the plot. GO analyses were performed using the Functional Annotation tool of the DAVID Bioinformatics Resources, version 6.8 [12–14]. For each cluster, GO analysis was performed using both all expressed genes, and genes with average expression count of  $\geq 1$  per cell. The top 20 GO terms are reported for each cluster based on Benjamini p-values, excluding generic GO terms related to common cellular processes (e.g., cell cycle, macromolecule processing, respiration, protein processing, regulation of apoptosis, etc.).

### Ethics Statement

Approval for this study was provided by the Western Sydney University Human Research Ethics Committee (Australia).

### CRediT Author Statement

**Rachel Shparberg:** Visualisation, Writing – Original draft preparation, Writing – Reviewing and Editing **Chitra Umala Dewi:** Methodology, Investigation, **Vikkitharan Gnanasambandapillai:** Methodology, **Liwan Liyanage:** Data curation, Methodology, Software, Formal analysis, Validation **Michael O'Connor:** Conceptualization, Methodology, Resources, Supervision, Writing – Original draft preparation, Writing – Reviewing and Editing, Project Administration, Funding acquisition

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

### Acknowledgments

The authors thank the Ramaciotti Centre for Genomics (Sydney, Australia) for performing the single cell RNA sequencing, and J. E. Powell and the Garvan-Weizmann Centre for Cellular Genomics (Sydney, Australia) for assistance with mapping the single cell RNA sequencing data. The authors wish to also acknowledge the following funding organisations that supported the study: Western Sydney University; The Medical Advances Without Animals Trust (MAWA); the Rebecca L. Cooper Medical Research Foundation; and Stem Cells Australia.

## References

- [1] P. Murphy, M.H. Kabir, T. Srivastava, M.E. Mason, C.U. Dewi, S. Lim, A. Yang, D. Djordjevic, M.C. Killingsworth, J.W.K. Ho, D.G. Harman, M.D. O'Connor, Light-focusing human micro-lenses generated from pluripotent stem cells model lens development and drug-induced cataract in vitro, *Development* 145 (2018) pii: dev155838.
- [2] C. Umala Dewi, M. Mason, T. Cohen-Hyams, M.C. Killingsworth, D.G. Harman, V. Gnanasambandapillai, L. Liyanage, M.D. O'Connor, A simplified method for producing human lens epithelial cells and light-focusing micro-lenses from pluripotent stem cells, *Exp. Eye Res.* doi.org/10.1016/j.exer.2020.108317.
- [3] A. Cvekl, X. Zhang, Signaling and gene regulatory networks in mammalian lens development, *Trends Genet.* 33 (2017) 677–702.
- [4] J.R. Hawse, C. DeAmicis-Tress, T.L. Cowell, M. Kantorow, Identification of global gene expression differences between human lens epithelial and cortical fiber cells reveals specific genes and their associated pathways important for specialized lens cell functions, *Mol. Vis.* 11 (2005) 274–283.
- [5] A. Kakrana, A. Yang, D. Anand, D. Djordjevic, D. Ramachandruni, A. Singh, H. Huang, J.W.K. Ho, S.A. Lachke, iSyTE 2.0: a database for expression-based gene discovery in the eye, *Nucleic Acids Res.* 46 (2018) D875–D885.
- [6] C.J. Bult, J.A. Blake, C.L. Smith, J.A. Kadin, J.E. Richardson, M.G.D. Group, Mouse genome database (MGD) 2019, *Nucleic Acids Res.* 47 (2019) D801–D806.
- [7] F.J. Lovicu, J.W. McAvoy, R. de Jongh U., Understanding the role of growth factors in embryonic development: insights from the lens, *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 366 (2011) 1204–1218.
- [8] O. Adewumi, B. Aflatoonian, L. Ahrlund-Richter, M. Amit, P.W. Andrews, G. Beighton, P.A. Bello, N. Benvenisty, L.S. Berry, S. Bevan, B. Blum, J. Brooking, K.G. Chen, A.B.H. Choo, G.A. Churchill, M. Corbel, I. Damjanov, J.S. Draper, P. Dvorak, K. Emanuelsson, R.A. Fleck, A. Ford, K. Gertow, M. Gertsenstein, P.J. Gokhale, R.S. Hamilton, A. Hampl, L.E. Healy, O. Hovatta, J. Hyllner, M.P. Imreh, J. Itskovitz-Eldor, J. Jackson, J.L. Johnson, M. Jones, K. Kee, B.L. King, B.B. Knowles, M. Lako, F. Lebrin, B.S. Mallon, D. Manning, Y. Maysyar, R.D.G. McKay, A.E. Michalska, M. Mikkola, M. Mileikovsky, S.L. Minger, H.D. Moore, C.L. Mummery, A. Nagy, N. Nakatsuji, C.M. O'Brien, S.K.W. Oh, C. Olsson, T. Otonkoski, K.-Y. Park, R. Passier, H. Patel, M. Patel, R. Pedersen, M.F. Pera, M.S. Piekarczyk, R.A.R. Pera, B.E. Reubinnoff, A.J. Robins, J. Rossant, P. Rugg-Gunn, T.C. Schulz, H. Semb, E.S. Sherrer, H. Siemen, G.N. Stacey, M. Stojkovic, H. Suemori, J. Szatkiewicz, T. Turetsky, T. Tuuri, S. van den Brink, K. Vintersten, S. Vuorio, D. Ward, T.A. Weaver, L.A. Young, W. Zhang, Characterization of human embryonic stem cell lines by the International Stem Cell Initiative, *Nat. Biotechnol.* 25 (2007) 803–816.
- [9] M. O'Connor D., M.D. Kardel, C.J. Eaves, Functional assays for human embryonic stem cell pluripotency, in: *Embryonic Stem Cell Ther. Osteo-Degenerative Dis*, Humana Press, Totowa, NJ, 2010, pp. 67–80.
- [10] M.D. O'Connor, M.D. Kardel, I. Isofina, D. Youssef, M. Lu, M.M. Li, A. Nagy, C.J. Eaves, S. Vercauteren, Alkaline phosphatase-positive colony formation is a sensitive, specific, and quantitative indicator of undifferentiated human embryonic stem cells, *Stem Cells* 26 (5) (2008) 1109–1116.
- [11] T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W.M. Mauck, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive integration of single-cell data, *Cell* 177 (2019) 1888–1902.
- [12] W. Huang da, B.T. Sherman, R.A. Lempicki, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists, *Nucleic Acids Res.* 37 (2009) 1–13.
- [13] W. Huang da, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, *Nat. Protoc.* 4 (2009) 44–57.
- [14] M.H. Kabir, R. Patrick, J.W.K. Ho, M.D. O'Connor, Identification of active signaling pathways by integrating gene expression and protein interaction data, *BMC Syst. Biol.* 12 (2018) 120.