

SenPred

Bethany Hughes

31 March 2024

Machine learning model built from single cell RNA sequencing of replicative senescent and early proliferative human dermal fibroblasts

Funding sources - BBSRC iCASE Unilever PhD studentship (BB/V509668/1)

Table of Contents

- [Lead contacts](#)
- [Project background](#)
- [Input data](#)
- [Installation](#)
- [File overview](#)
- [Analysis pipeline](#)
- [Output](#)

Lead contacts

Primary author: Beth Hughes, Queen Mary University of London, b.k.hughes@qmul.ac.uk

Principal Investigator: Professor Cleo Bishop, Queen Mary University of London, c.l.bishop@qmul.ac.uk

Project background

This analysis was run based on the Seurat 4.3.0.1 and scPred 1.9.2 vignettes. This analysis has been version controlled on GitHub, and can be accessed via the following URL https://github.com/bethk-h/SenPred_HDF.

The input data was generated by Dr Deborah Milligan, and sequenced using the v2 Chromium single cell 3' Kit (10X Genomics), and Illumina NextSeq 500 High Output Run Sequencing platform using paired end sequencing.

Input data

The input data for this project can be found on GEO, accession number ??????. The contents of the Input-files folder will not be tracked via this repository, and will need to be downloaded before running. These downloaded files need to be placed into a folder called Inputfiles within the R project.

Installation

R version 4.3.1 (2023-06-16)

Package versions are stored in [renv.lock](#) file. To use this folder, one must remove the line “renv.lock” from the [.gitignore](#) file. Relevant package versions can then be installed using `renv::restore()`.

File overview

R Scripts All scripts can be found in the [R_scripts](#) folder. The number prefix indicates the order the files should be run (from 1 to 18). To run the complete analysis, first run 95_Make_Clean.rmd to create or clean the results folder. Then, run 99_Run_All.rmd to run the scripts in order.

Analysis pipeline

- [1_HDF_data.Rmd](#) - loading data and converting into Seurat object.
- [2_HDF_filtering_normalisation.Rmd](#) - Filtering and normalisation of the Seurat object.
- [3_HDF_2DCells.Rmd](#) - filtering Seurat object to select the 2D cells, and investigating clustering and differential expression of 2D fibroblasts.
- [4_Scpred_2D.Rmd](#) - Building machine learning models of proliferative and senescent HDFs in 2D, using ScPred package (see <https://github.com/powellgenomicslab/scPred>).
- [5_ScPred2d_Chan.Rmd](#) - Testing ML models on external replicative senescence fibroblast single cell RNA seq data set from Chan et al. (Chan, M., Yuan, H., Soifer, I., Maile, T.M., Wang, R.Y., Ireland, A., O'Brien, J.J., Goudeau, J., Chan, L.J., Vijay, T. and Freund, A., 2022. Novel insights from a multiomics dissection of the Hayflick limit. *Elife*, 11, p.e70283.).
- [6_ScPred2D_Chan_EPESDS.Rmd](#) - Building PDL50 cells from Chan et al. into the model, as ES or 'Early Senescent' cells.
- [7_Teo.Rmd](#) - Applying the EP/ES/DS SenPred model to an alternative dataset, with OIS and paracrine triggers. (Teo, Y.V., Rattanaivirotkul, N., Olova, N., Salzano, A., Quintanilla, A., Tarats, N., Kiourtis, C., Müller, M., Green, A.R., Adams, P.D. and Acosta, J.C., 2019. Notch signaling mediates secondary senescence. *Cell reports*, 27(4), pp.997-1007.)
- [8_Tabib.Rmd](#) - Testing the 2D model on external whole skin in vivo single cell RNA seq dataset from Tabib et al. (Tabib, T., Morse, C., Wang, T., Chen, W. and Lafyatis, R., 2018. SFRP2/DPP4 and FMO1/LSP1 define major fibroblast populations in human skin. *Journal of Investigative Dermatology*, 138(4), pp.802-810.)
- [9_soleboldo.Rmd](#) - Testing the 2D model on external whole skin in vivo single cell RNA seq dataset from Solé-Boldo et al. (Solé-Boldo, L., Raddatz, G., Schütz, S., Mallm, J.P., Rippe, K., Lonsdorf, A.S., Rodríguez-Paredes, M. and Lyko, F., 2020. Single-cell transcriptomes of the human skin reveal age-related loss of fibroblast priming. *Communications biology*, 3(1), p.188.)
- [10_soleboldo_tabib_combined2D.Rmd](#) - Integrating Tabib and Sole Boldo datasets to combine donor predictions from 2D EP/ES/DS SenPred model.
- [11_ScPred_3D.Rmd](#) - Building an ML model to predict senescence of 3D fibroblasts.
- [12_3Dpred_Tabib.Rmd](#) - Testing the 3D model on external whole skin in vivo single cell RNA seq dataset from Tabib et al. (Tabib, T., Morse, C., Wang, T., Chen, W. and Lafyatis, R., 2018. SFRP2/DPP4 and FMO1/LSP1 define major fibroblast populations in human skin. *Journal of Investigative Dermatology*, 138(4), pp.802-810.)
- [13_3Dpred_SoleBoldo.Rmd](#) - Testing the 3D model on external whole skin in vivo single cell RNA seq dataset from Solé-Boldo et al. (Solé-Boldo, L., Raddatz, G., Schütz, S., Mallm, J.P., Rippe, K., Lonsdorf, A.S., Rodríguez-Paredes, M. and Lyko, F., 2020. Single-cell transcriptomes of the human skin reveal age-related loss of fibroblast priming. *Communications biology*, 3(1), p.188.)
- [14_soleboldo_tabib_combined3D.Rmd](#) - Combining the 3D SenPred predictions for Tabib and Sole Boldo datasets. Generating heatmaps of common senescence marker correlations in cells predicted to be EP or DS.
- [15_matrisome.Rmd](#) - Importing NABA_MATRISOME genes from the matrisome project <https://sites.google.com/uic.edu/matrisome/home?authuser=0>, to compare ECM expression in fibroblasts grown in 2D, 3D, and in vivo.
- [16_Ganier.Rmd](#) - Preprocessing, for testing SenPred on an additional whole skin dataset by Ganier et al, and combining these donor predictions with the Tabib and Sole Boldo datasets to investigate age related trends in senescence predictions. Paper: Ganier, C., Mazin, P., Herrera-Oropeza, G., Du-Harpur, X., Blakeley, M., Gabriel, J., Predeus, A.V., Cakir, B., Prete, M., Harun, N. and Darrigrand,

- J.F., 2024. Multiscale spatial mapping of cell populations across anatomical sites in healthy human skin and basal cell carcinoma. *Proceedings of the National Academy of Sciences*, 121(2), p.e2313326120.
- [17_Ganier2.Rmd](#) - Identifying and filtering fibroblasts for testing SenPred on an additional whole skin dataset by Ganier et al, and combining these donor predictions with the Tabib and Sole Boldo datasets to investigate age related trends in senescence predictions. Paper: Ganier, C., Mazin, P., Herrera-Oropeza, G., Du-Harpur, X., Blakeley, M., Gabriel, J., Predeus, A.V., Cakir, B., Prete, M., Harun, N. and Darrigrand, J.F., 2024. Multiscale spatial mapping of cell populations across anatomical sites in healthy human skin and basal cell carcinoma. *Proceedings of the National Academy of Sciences*, 121(2), p.e2313326120.
 - [18_Ganier3.Rmd](#) - Testing SenPred on an additional whole skin dataset by Ganier et al, and combining these donor predictions with the Tabib and Sole Boldo datasets to investigate age related trends in senescence predictions. Paper: Ganier, C., Mazin, P., Herrera-Oropeza, G., Du-Harpur, X., Blakeley, M., Gabriel, J., Predeus, A.V., Cakir, B., Prete, M., Harun, N. and Darrigrand, J.F., 2024. Multiscale spatial mapping of cell populations across anatomical sites in healthy human skin and basal cell carcinoma. *Proceedings of the National Academy of Sciences*, 121(2), p.e2313326120.
 - [19_alldermal_celltypes.Rmd](#) - Testing 3D SenPred on all skin cells from Tabib and Sole Boldo datasets.
 - [20_allskin_ganier.Rmd](#) - Testing 3D SenPred on all skin cells from Ganier dataset, and combining with Tabib and Sole Boldo dataset to investigate age related trends in senescence predictions.
 - [21_lungs.Rmd](#) - Processing data to test 3D SenPred on fibroblasts from lung tissue. The data was obtained from the human cell atlas using the following paper: Sikkema, L., Ramírez-Suástegui, C., Strobl, D.C., Gillett, T.E., Zappia, L., Madissoon, E., Markov, N.S., Zaragosi, L.E., Ji, Y., Ansari, M. and Arguel, M.J., 2023. An integrated cell atlas of the lung in health and disease. *Nature Medicine*, 29(6), pp.1563-1577. Data was downloaded from cellxgene at the following link: <https://cellxgene.cziscience.com/collections/6f6d381a-7701-4781-935c-db10d30de293>
 - [22_lungs.Rmd](#) - Testing 3D SenPred on fibroblasts from lung tissue. The data was obtained from the human cell atlas using the following paper: Sikkema, L., Ramírez-Suástegui, C., Strobl, D.C., Gillett, T.E., Zappia, L., Madissoon, E., Markov, N.S., Zaragosi, L.E., Ji, Y., Ansari, M. and Arguel, M.J., 2023. An integrated cell atlas of the lung in health and disease. *Nature Medicine*, 29(6), pp.1563-1577. Data was downloaded from cellxgene at the following link: <https://cellxgene.cziscience.com/collections/6f6d381a-7701-4781-935c-db10d30de293>

Output

The resulting output files are saved to the [Output](#) folder in this repository. This will not be tracked in GitHub.

Please note, each time the script is run, the partitioning of training and testing dataset is random and therefore slightly different. If you want to run the scripts multiple times with reproducible outputs, you must copy the training and testing dataset files from [Output/Seurat_Objects](#), and save them to Input files. You must then remove the # in front of “training <- ReadRDS...” and “testing <- ReadRDS” in the scripts [4_Scpred_2D.Rmd](#), [6_ScPred2D_Chan_EPESDS.Rmd](#), and [9_ScPred_3D.Rmd](#).

```
sessionInfo()
#> R version 4.3.1 (2023-06-16)
#> Platform: aarch64-apple-darwin20 (64-bit)
#> Running under: macOS Monterey 12.0.1
#>
#> Matrix products: default
#> BLAS: /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRblas.0.dylib
#> LAPACK: /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRlapack.dylib; LAPACK v
#>
```

```
#> locale:
#> [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
#>
#> time zone: Europe/London
#> tzcode source: internal
#>
#> attached base packages:
#> [1] stats      graphics  grDevices datasets  utils      methods    base
#>
#> loaded via a namespace (and not attached):
#> [1] compiler_4.3.1      BiocManager_1.30.21.1 fastmap_1.1.1
#> [4] cli_3.6.1           htmltools_0.5.5      tools_4.3.1
#> [7] rstudioapi_0.15.0   yaml_2.3.7           rmarkdown_2.23
#> [10] knitr_1.43          xfun_0.39            digest_0.6.33
#> [13] rlang_1.1.1         renv_1.0.0           evaluate_0.21
```

Session information This document was processed on: 2024-03-31.