Assessing the Viability of Inferential Spatial Transcriptomics Data for Studying Skin Photoaging

Victoria Zhang

Emerging Diagnostic and Investigative Technologies, Department of Pathology, Dartmouth Hitchcock Medical Center

Dartmouth Health

ABSTRACT

- Frequent exposure to solar ultraviolet (UV) rays can result in serious skin damage and **photoaging**.
- Current methods to assess photoaging involve building an **age clock**. Prior clocks are mostly DNAm or scRNA-seq-based.
- Spatial transcriptomics data has the potential to **reveal unique spatial pathways behind photaging.**
- Methods to generate inferential ST data **show potential to be used in analysis**
- We use **synthetic ST data** to identify top gene markers and cell types associated with photoaging and to create **one of the first spatial aging clocks** for photoaging

INTRODUCTION

- Skin photoaging, or age acceleration caused by chronic sun exposure, disrupts various skin processes, increasing the risk for skin-related diseases such as skin cancer

Reseach in Photoaging

- Prior analyses on photoaging have used aging clocks to track age acceleration
  - Studies using DNAm and scRNA-seq data have also identified key gene markers
- Gap in knowledge in the spatial capabilities of photoaging analyses by ST data

ST Inference Models

- to reduce cost of assays needed to produce ST data, inference models are used to directly infer ST data from H&E slides

Goal: Use synthetic ST data to investigate photoaging markers and models

METHODS

Data and Preprocessing:

- 88 skin samples and 900,000+ total spots generated via an inference model in the lab
- NaN values removed, QC filtering, normalization, log transformation through Python package scanpy.
- Top 2000 highly-variable genes selected in the data, then clipped and scaled to unit variance
- Before inputting into age acceleration clock, all individual AnnData files were concatenated into one single AnnData object

Clustering and Visualization:

- Neighbors graph generated through PCA (n_components=40)
- Leiden clusters generated and visualized with UMAP plots

METHODS

Cell Type Annotations

- Top ranked genes from each Leiden cluster generated using Wilcoxon method
- Manually annotated for 16 cell types, identifiying cell type based on a score based heuristic on the gene markers present in the cluster

Differential Expression Analysis

- conducted based on solar elastosis severity, comparing across top 100 ranked genes with the Wilcoxon rank sum test
- top 20 genes from analysis visualized with expression scores

Statistical testing and correlation

- Chi-squared test to determine if cell type proportions change with age and solar elastosis severity
- Correlation test by Pearson's coefficient to determine top ten marker genes associated with age and solar elastosis

Age Acceleration Clocks

- Two models tested: 5-fold cross-validation elastic net regression and CNN
- input into regression was ST data aggregated across sample by averaging gene expression across spots
- input into CNN was gene expression concatenated with spatial coordinates

RESULTS

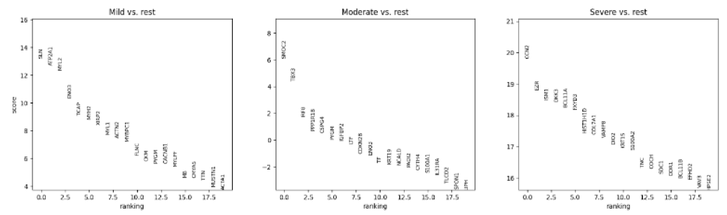

Figure 1. Top ten genes for each solar elastosis category generated from differential expression analysis

Top Gene Markers and Associated Cell Types Correlated with Age and Solar Elastosis

Correlation	Cell Type	Gene Markers
Age	Fibroblasts	SPARC, APOE, FASN, DCN, COL1A1, COL3A1
	Basal Cells	GSN, NR4A1, LMNA
	Dendritic Cells	LGALS1
Solar Elastosis	Fibroblasts	APOE, FASN, COL1A1, DCN
	Basal Cells	KRT5, KRT14, LMNA, KRT17, GSN
	T-Cells	CD74

Table 1 .Top gene markers and cell types found in correlation tests with age and solar elastosis

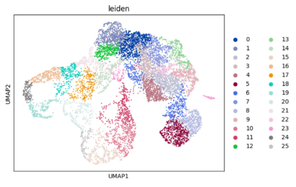

Figure 2. Leiden clustering visualization on one spot

Chi-Square Statistical Results for Cell-type Proportions

Assessed Parameter	Chi-Statistic	p-value
Age	1301.68	2.85
Solar Elastosis Severity	223.60	1.89

Table 2. Results of chi-square tests on cell type proportions

Comparison of Age Acceleration Clock Performance

Model	R^2	MSE	MAE
This paper's CNN	0.86	20.75	3.28
AltumAge	0.96	29.49	2.15
Horvath Clock	0.91	71.03	3.53
This paper's elastic-net	0.091	137.79	N/A

Table 1 .Top gene markers and cell types found in correlation tests with age and solar elastosis

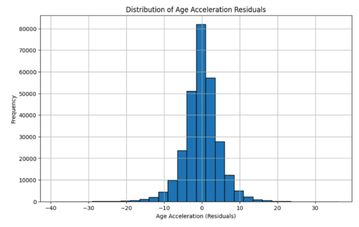

Figure 3. Distribution of calulated age acceleration over all spots by CNN

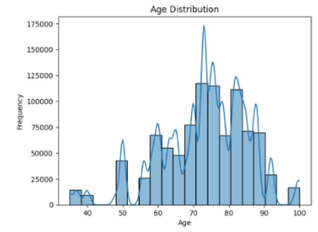

Figure 4. Age distrbution of dataset

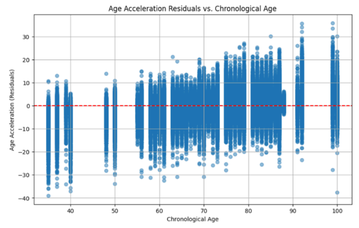

Figure 5. Age acceleration residuals across various ages

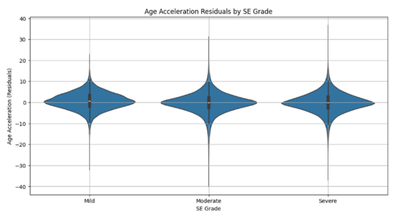

Figure 4. Age acceleration residuals by solar elastosis category

CONCLUSION

Value of ST Data and Inference:

- The CNN aging clock model performs on-standard with top deep learning and DNAm clocks in literature, showing the viability of synthetic ST data in assessing photoaging
- Gene markers identified by ST data can account for spatial relationships and pathways

Limitations:

- no valid comparision to non-synthetic ST data in this study
- cell type annotation limited to certain cells
- aggregation in elastic net may have caused loss of spatial information

Future steps may look towards conducting more thorough cell type annotations and investigating additional biological pathways revealed by ST data, as well as fine-tuning a "spatial" aging clock.

Acknowledgements: Dr. Joshua Levy, Gokul Srinivasan, Zarif Azher

edit.