Identification of cellular senescence signatures for classifying senescence status based on machine learning approaches

Dong Jun Shin[1], Seon Jeong Woo[1], and Sohyun Hwang[1,2]

[1]Department of Biomedical Science, CHA University

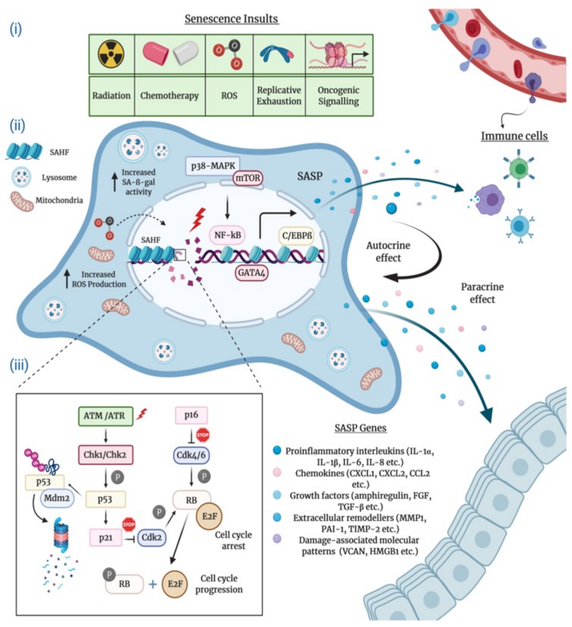
[2]Department of Pathology, CHA Bundang Medical Center, CHA University School of Medicine

Abstract

Cellular senescence is a permanent cell-cycle arrest which prevents damaged cells from unusual proliferation. By the secretion of senescence-associated secretory phenotype factors, senescent cells affect various age-related diseases including cancers. Selective elimination of senescent cells, named as 'senolysis', has emerged as a potential therapeutic method for various diseases. However, since senescent cells possess heterogeneous features according to cell types or senescence inducers, it was difficult to identify the senolysis markers. Therefore, we collected raw 192 samples of RNA-seq data in various human senescence cells by various inducers such as replicative senescence, oncogene induced senescence, and therapy induced senescence from different experiments to find out consensus features of cellular senescence regardless of inducer types or cell types. Meta-analysis was conducted to raise statistical power and machine learning approaches such as lasso and support vector machine were used for selecting representative features of cellular senescence and for the construction of senescence classification model with selected features. Our model can discriminate senescent samples from non-senescent samples regardless of their inducer types with great performance. Moreover, our model was also able to discriminate various senescence cell types, compared to other senescence gene sets from various studies. As a result, we offered these 17 signature genes as core characters of cellular senescence.

Introduction

Cellular senescence is irreversible cell cycle arrest triggered by stressful insults or processes, characterized by senescence-associated secretory phenotype (SASP), macromolecular damage, and altered metabolism. It arises in response to diverse triggers including telomere attrition, macromolecular damage and signaling from activated oncogenes, sharing different traits according to its inducers. Recent findings revealed that senescent cell is a contributor to aging and age-related diseases. Senolysis, selective elimination of senescent cells, has been emerged as an effective therapy for the treatment of age-related diseases including cancers when combined with pre-existing therapies. However, pre-existing senescence markers differ by cell types or inducer types of senescent cells, making it difficult for researchers to define consensus markers for cellular senescence. Therefore, it is necessary to find characteristic markers for cellular senescence, regardless of their heterogeneous types.



Demirci, Dilara, et al. *Cells* 10.2 (2021): 208.

Materials & Methods

▻ Salmon pipeline and workflow of the analysis

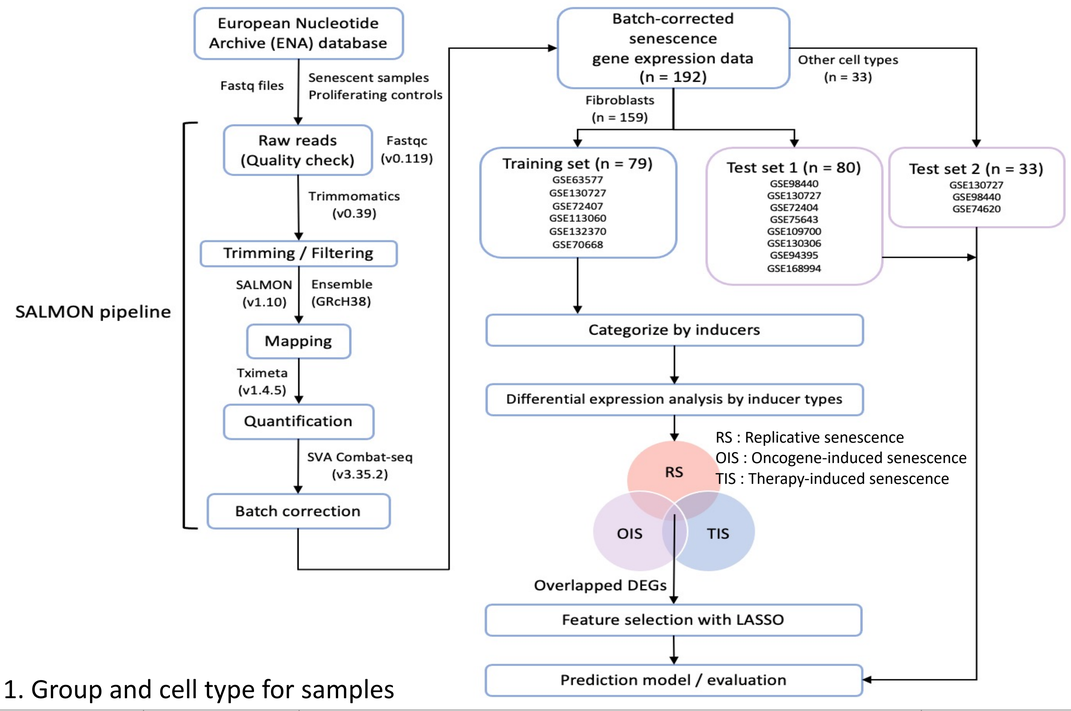


Table 1. Group and cell type for samples

Group	Cell type	Controls	Replicative	OIS	TIS	Total
Training	Fibroblasts	37	13	15	14	79
Test 1	Fibroblasts	32	13	24	9	80
Test 2	Other cell types	15	7	-	11	33
Total	-	73	28	41	23	**192**

▻ Total 192 of cellular senescence expression data as Fastq files of RNA-seq were collected from ENA-EBI database. All samples were processed with in-built salmon pipeline, which consists of quality control, trimming, mapping, quantifying, and batch-correction steps. Data were divided into training set and test sets. Test set 1 consists of 80 fibroblasts and test set 2 of 33 samples with various cell types such as Mesenchymal stem cells, endothelial cells, and liposarcoma. Samples were categorized by their senescence inducer types.

▻ Differential expression (DE) analysis between controls samples and various inducers in training sets were performed using *edgeR* 3.10. DE genes were filtered by adjusted p-value (<0.05) and log2 fold change (log_2FC >1 or log_2FC < -1). Among inducer specific DE genes, common 465 genes across all types of senescence inducers were selected for further analysis.

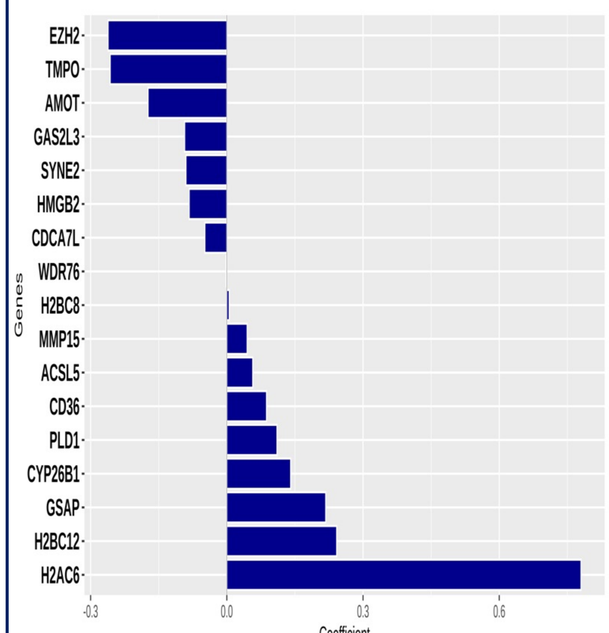
▻ TMM normalized and log transformed expression values of filtered 465 genes were used for lasso regression. Final 17 genes were selected by lasso (*glmnet*, v4.1-1), and gene ontology analysis on 17 genes was performed with *gprofiler2* v.0.2.1.

▻ Principal component analysis (PCA) with 14,037 genes after filtering lowly expressed genes and 17 lasso selected genes on test set were plotted respectively by *stats* v3.6.2 and *factoextra* v1.0.7.

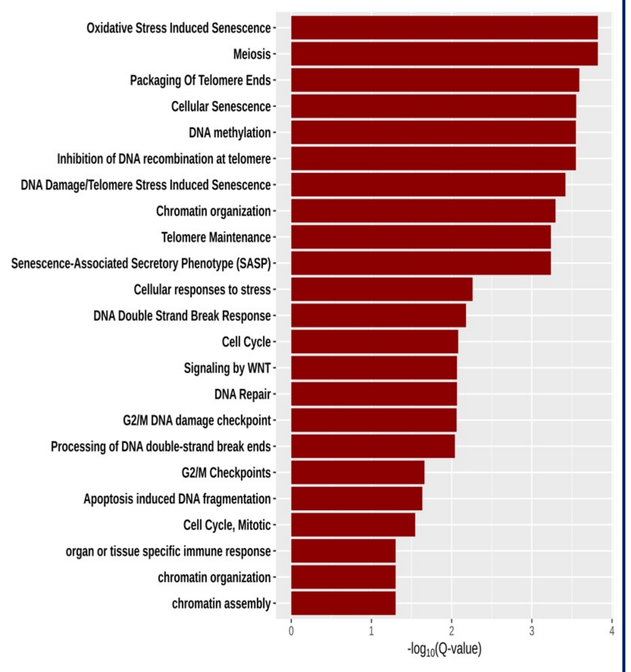
▻ SVM model (*e1071*, v1.7-6) for cellular senescence prediction was built with 17 genes, as well as various pre-defined four senescence gene sets from different studies. Model performance on test sets was compared by ROC curves (*ROCR*, v1.0-11).

Results

▻ **Non-zero coefficients of lasso selected 17 genes**

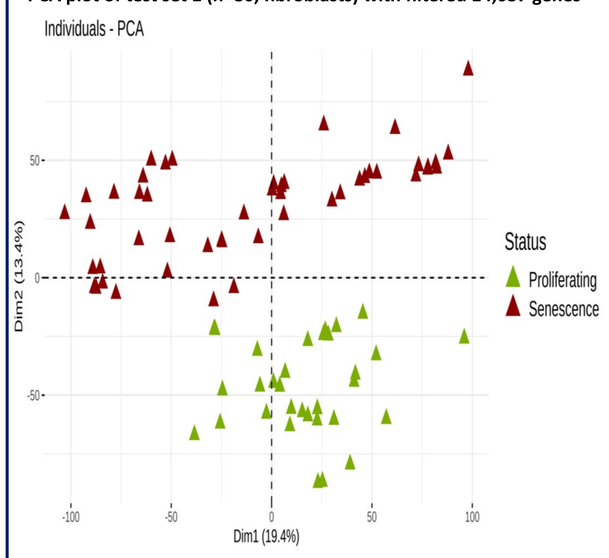


▻ **Enriched GO terms in selected 17 genes**

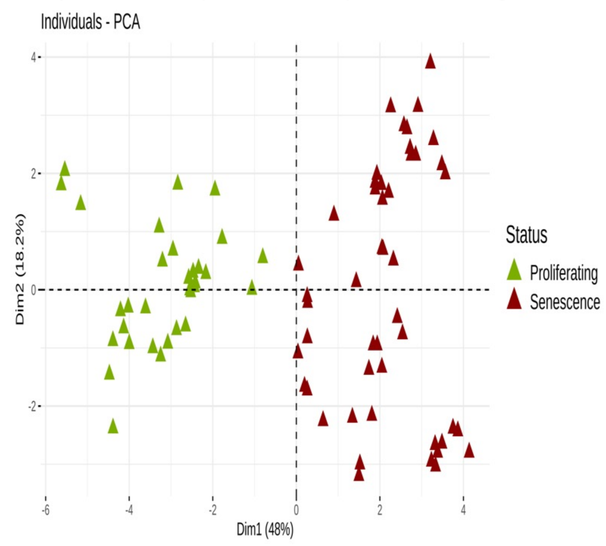


▻ **Principal component analysis on Test set**

PCA plot of test set 1 (n=80, fibroblasts) with filtered 14,037 genes

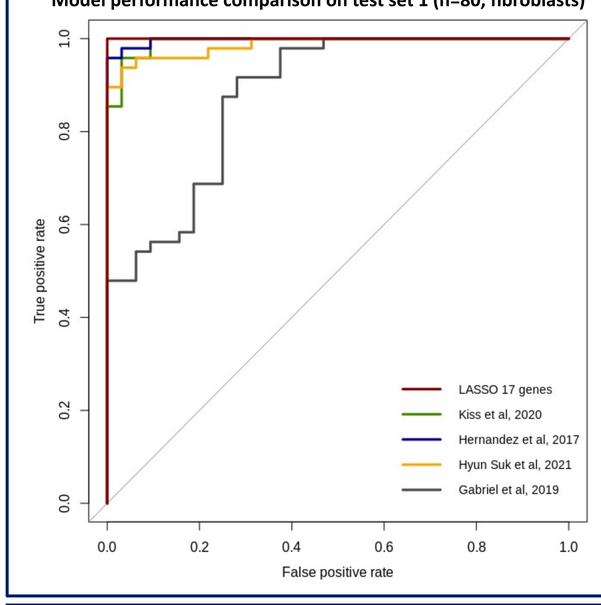


PCA plot of test set 1 (n=80, fibroblasts) with selected 17 genes

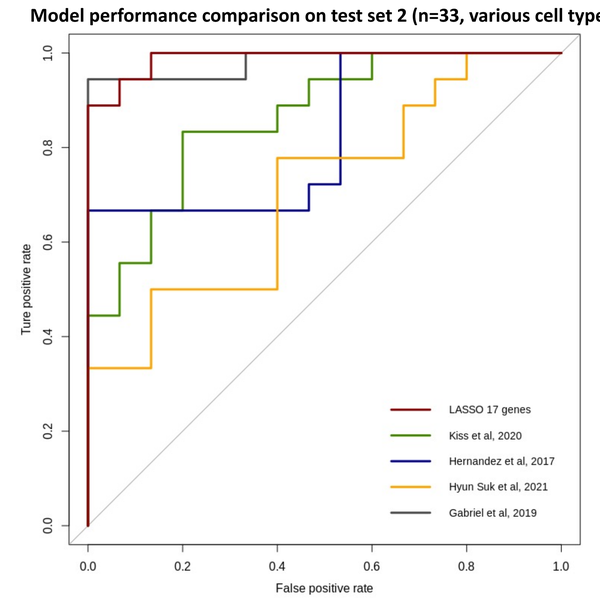


▻ **Prediction performance for senescent cell in several models**

Model performance comparison on test set 1 (n=80, fibroblasts)



Model performance comparison on test set 2 (n=33, various cell types)



Conclusions

▻ We identified 17 cellular senescence signature genes based on meta-analysis and ML-approaches.

▻ Those signature genes are related to senescence pathways such as telomere, cell cycle, immune response, chromatin assembly, DNA damage, and SASP.

▻ PCA results on test set prove that our signature genes can discriminate cellular senescence with proliferating control samples regardless of its inducer type.

▻ SVM prediction model of our 17 signatures outperformed other four models from various studies, which showed the almost perfect performance regardless of cell types in test set 1: AUC = 1.0 and test set 2: AUC = 0.98.

▻ We offered these 17 signature genes as core characters of cellular senescence.

References

- Hernandez-Segura, Alejandra, et al. Unmasking transcriptional heterogeneity in senescent cells. *Current Biology*, 2017, 27.17: 2652-2660. e4.
- Casella, Gabriel, et al. Transcriptome signature of cellular senescence. *Nucleic acids research*, 2019, 47.14: 7294-7305.
- Park, Hyun Suk, et al. Endothelial cell senescence: A machine learning-based meta-analysis of transcriptomic studies. *Aging research reviews* 65, 2021, 65.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. NRF-2018R1C1B5032617 and No. NRF-2019R1A6A1A03032888).