RNA-seq Experiment Report

# Overview

The goal of this experiment was to compared the transcriptomics of myofibroblasts and senescent fibroblasts. To achieve this goal, RNA-seq experiment was carried out on human fetal foreskin fibroblasts 2 (HFFF2) treated with 2ng/ml TGF-beta-1 to induce myofibroblast differentiation or 10Gy gamma irradiation to induce senescence. RNA was isolated 7 days upon this treatments.

This report goes through some of the RNA-Seq analysis in R experiment starting with a read count matrix (processed using salmon). This report includes:

* Exploring count data after importing them into R and
* Normalizing RNA-seq counts
* Conducting quality assessment of counts

# Data and Metadata

We begin with our counts table and our samples metadata:  
Table

Description automatically generated

Table

Description automatically generated

# Quality control of the imported counts

## Library Sizes

The library sizes generally look good for each of our samples, but are not completely uniform:

Chart, bar chart

Description automatically generated

Thus, there are some outliers when we look at the distribution of counts:

Chart, box and whisker chart

Description automatically generated

We used the DESeq2::vst() function to compensate for different library sizes and put the data on the log2 scale. This also helps to remove dependence of variance on the mean.

## Chart Description automatically generated PCA

PCA of normalized samples colored by group (using plotPCA()):  
Chart, scatter chart

Description automatically generated

Our groups have different coloring due to their labels having different capitalization. After correction, we can see an updated plot:  
Chart, scatter chart

Description automatically generated  
  
There appears to be little to no batch effect:  
Chart, scatter chart

Description automatically generated

We also used prcomp() to try only using top genes for PCA:

Table of PCA results using prcomp() and the top 500 ranked genes only:  


# Reference:

Based on Introduction or [RNA-Seq analysis in R workshop](https://sbc.shef.ac.uk/training/rna-seq-in-r-2022-06-13/) offered by Sheffield Bioinformatics Core.

Original study: Mellone M, Hanley CJ, Thirdborough S, Mellows T, Garcia E, Woo J, Tod J, Frampton S, Jenei V, Moutasim KA, Kabir TD, Brennan PA, Venturi G, Ford K, Herranz N, Lim KP, Clarke J, Lambert DW, Prime SS, Underwood TJ, Vijayanand P, Eliceiri KW, Woelk C, King EV, Gil J, Ottensmeier CH, Thomas GJ. Induction of fibroblast senescence generates a non-fibrogenic myofibroblast phenotype that differentially impacts on cancer prognosis. Aging (Albany NY). 2016 Dec 15;9(1):114-132. doi: 10.18632/aging.101127. PMID: 27992856; PMCID: PMC5310659.