

Unraveling DNA Damage-Induced Ageing Mechanisms through Transcriptomic Analysis

BACKGROUND

Ageing is a complex biological process driven by cumulative molecular damage, among which DNA damage and impaired repair pathways play a central role. Deficiencies in nucleotide excision repair proteins, such as ERCC1-XPF, accelerate ageing phenotypes in both humans and progeroid mouse models. However, the molecular mechanisms linking DNA repair failure to tissue-specific ageing responses remain incompletely understood. By integrating transcriptomic datasets from DNA-repair-deficient and naturally aged models, this project aims to identify key genes and pathways involved in DNA damage-induced ageing, offering new insights into genomic instability and longevity regulation.

AIMS AND OBJECTIVES

The primary aim is to investigate how DNA repair deficiency contributes to ageing at the transcriptomic level through the identification of differentially expressed genes (DEGs) and pathway alterations.

Specific Objectives:

1. Retrieve and curate public transcriptomic datasets (GSE206778 and complementary GEO datasets) involving DNA repair-deficient and naturally aged models.
2. Perform quality control, normalisation, and differential expression analysis to identify genes altered during ageing and DNA repair loss.
3. Map DEGs to biological pathways using functional enrichment tools to uncover ageing-related molecular signatures.
4. Incorporate a machine learning (ML) layer: using clustering and feature selection algorithms, to classify tissues or age groups based on gene expression and identify potential biomarkers of ageing.
5. Integrate findings into a systems-level model connecting DNA repair deficiency to ageing mechanisms.

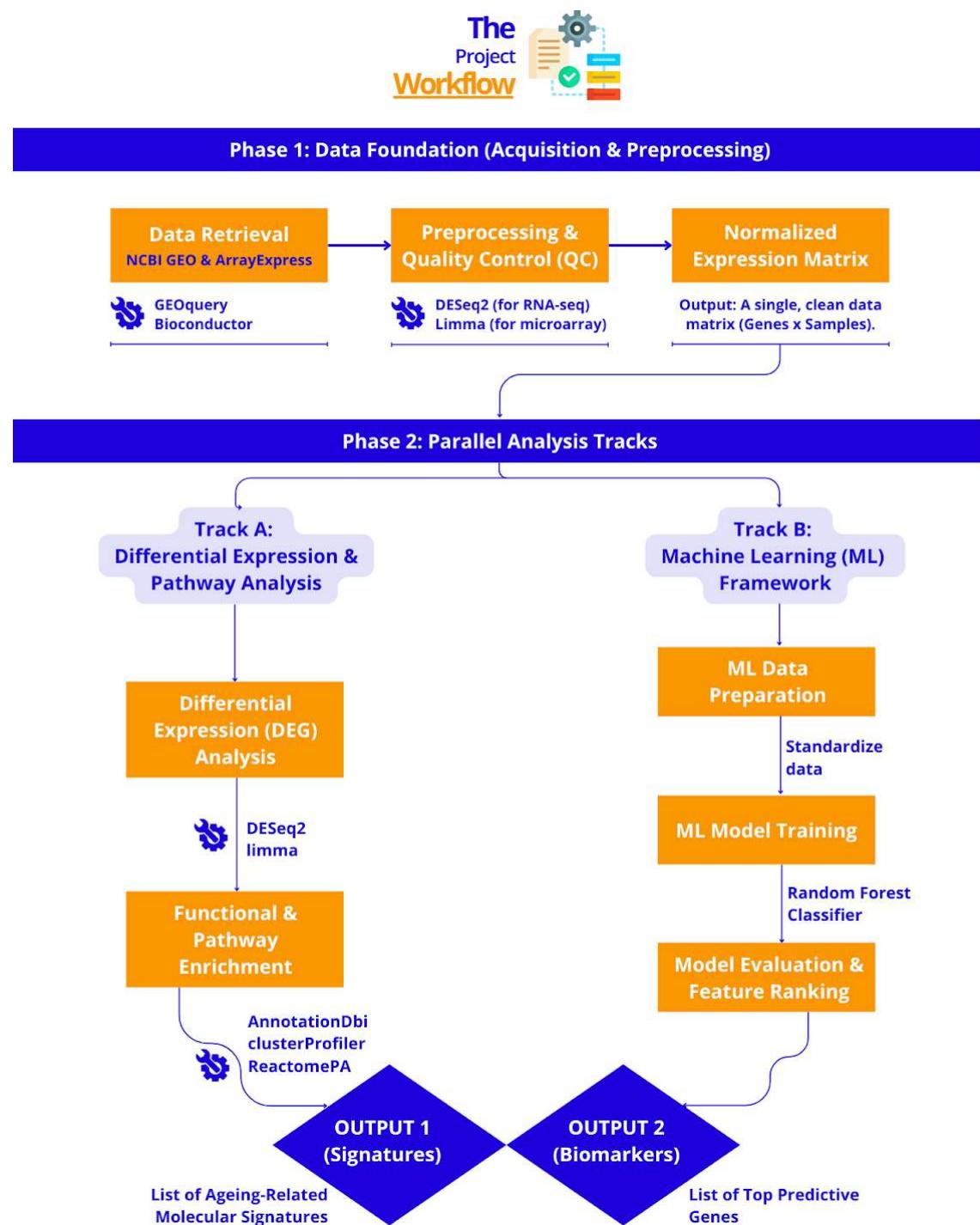
Proposed Methodology

Public RNA-seq and microarray datasets will be retrieved from the NCBI GEO and ArrayExpress databases using *GEOquery* and *Bioconductor* tools. Preprocessing (normalisation, outlier detection, and filtering) will be conducted in R using *DESeq2* or *limma*. Differentially expressed genes (DEGs) will be identified between DNA repair-deficient, aged, and control samples. Annotation and pathway enrichment will be performed using *AnnotationDbi*, *clusterProfiler*, and *ReactomePA*.

A machine learning framework will be applied to uncover discriminative expression patterns. A *Random Forest* classifier will be trained to distinguish between experimental groups, and the model's feature importance will be used to rank genes by their contribution to predictive accuracy. The top-ranked genes will be compared with DEGs and enriched pathways to identify robust molecular biomarkers of DNA damage-induced ageing.

All analyses will be implemented and version-controlled in R and GitHub. Final deliverables include. A detailed project report, a scientific poster, and an optional short video summarising the workflow and findings.

Workflow Diagram:



Group Details:

Name	Discord Username	Role
Hasmik Chilingaryan	mik0yan	Group Leader (All)
Swati	swati_29	Visualisation (Figures), Coder
James Joy	joyjames0778	Research (Literature & Data), Writer
Temitope Ayano	temiayano	Visualisation (Figures), Bioinformatics
Oluwasemilogo Akinlo	sems0396_19481	Research (Literature), Writer
Gloria	asuna4_	Research (Literature), Writer
Pavithra	pavithra0067_84434	Research (Literature), Writer
Deborah	ade_debbie	Research, (Final Project Report), Coder
Abeer Ali	abeer_ali_	Research (Literature & Data), Writer
Andrew	andronicus_11202	Research (Literature & Data), Writer
Randy	randy091754	Research (Literature), Writer