Bioinformatics approach to identify genes whose tumour expression shows a dual association with patient outcome

ARNAU SOLER COSTA

MSc in Omics Data Analysis

UVIC
UNIVERSITAT DE VIC
UNIVERSITAT CENTRAL
DE CATALUNYA

IDIBELL
Institut d'Investigació Biomèdica de Bellvitge

September 21th, 2022

Contents

01

Introduction

02

Goals

03

Methods

04

Materials

05

Results

06

Discussion
& conclusions

Introduction

CANCER

Disease in which cells in the body multiply uncontrollably and spread to other parts of the body causing tumours

Genetic disease

Oncogenes
Tumour suppressor genes (TSG) } "Drivers"

Introduction

Gene expression studies

Offer information about the association of genes and the phenotype
or variable of interest

Over-expression or under-expression of a given gene the associated outcome → **Cancer outcome**

Introduction

To study of gene expression over time

Survival data → When the event ocurred

Cox Proportional-Hazards regression model (Cox model)

Analyse survival cancer data kept scientists thinking that gene expression does not change over time

This means that genes only can act as oncogenes or TSG, and this remains constant over time

Introduction

Dormant or quiescent → Active state (relapse)

Show of a dual mode of action, showing an effect in one direction that changes with time

TGF-β acts as tumour suppressor or tumour promoter depending on the cellular context

Introduction

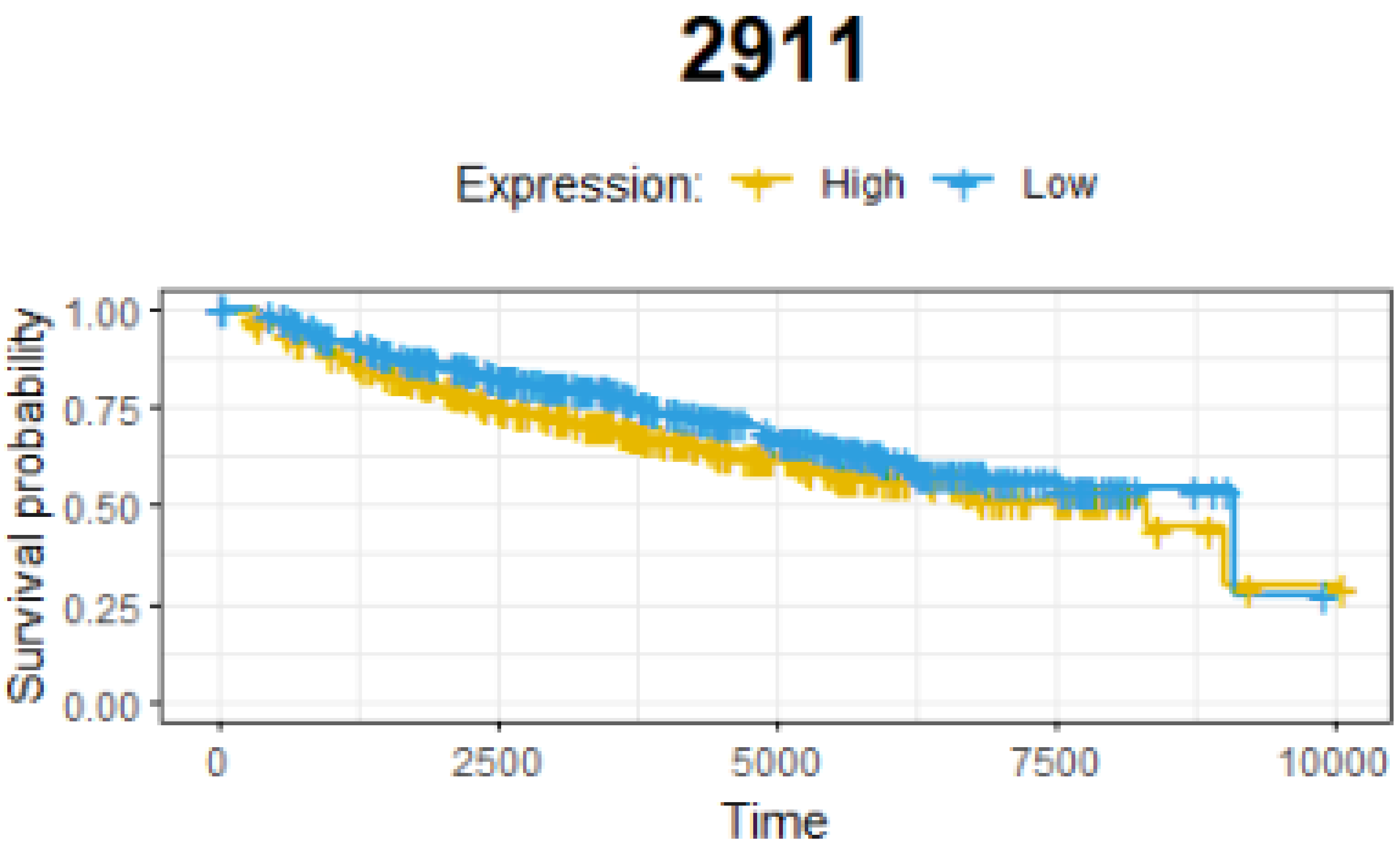
Some gene products may act as tumour suppressors or oncogenes depending on disease stage or other variables

↓

Biphasic Genes

Introduction

Kaplan-Meier plot

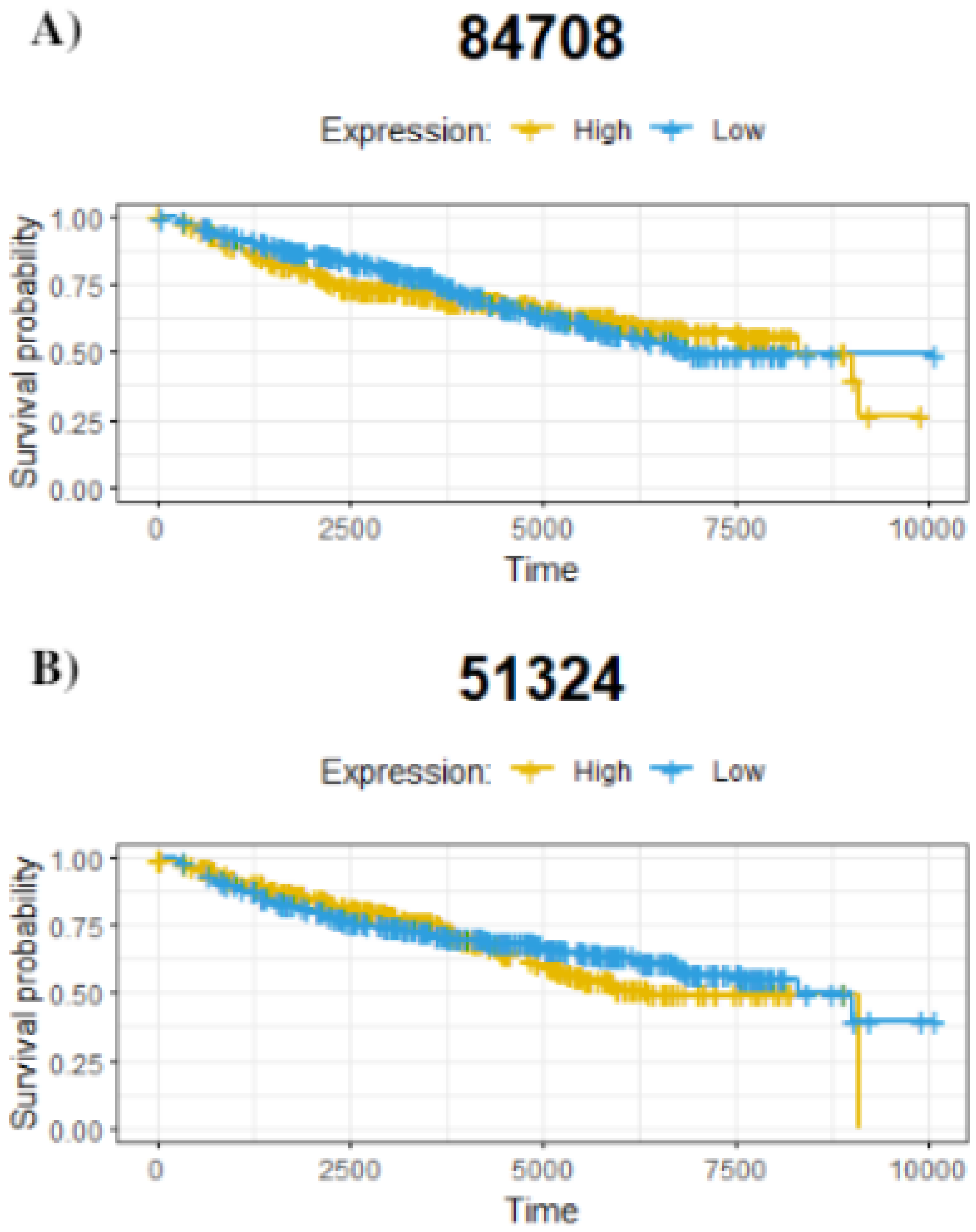


Kaplan-Meier (KM) plot of the gene 2911

Goals

Identify genes whose expression reduce or increase risk of cancer during a period of time, but subsequently change their effect direction during subsequent follow-up

Understand this novel class of cancer genes biologically and functionally



Materials

Two RNA-Seq datasets

Human breast cancer

Preprocessed and normalized

Materials

BRCA-TCGA project

Obtained from The Cancer Genome Atlas (TCGA)

Selecting the gene signature Luminal A (LumA) (subtype of breast cancer)

233 samples and 15,748 genes with PFI times up to 8,000 days

THE CANCER GENOME ATLAS

Materials

Molecular Taxonomy of Breast Cancer International Consortium (METABRIC)

Obtained from cBioportal

Selecting the gene signature LumA

679 samples and 18,492 genes with RFS times up to 10,000 days

cBioPortal
for Cancer Genomics

Materials

Only female patients remained

For each dataset there were a metadata file which contains phenotypic data and other covariates (including survival data)

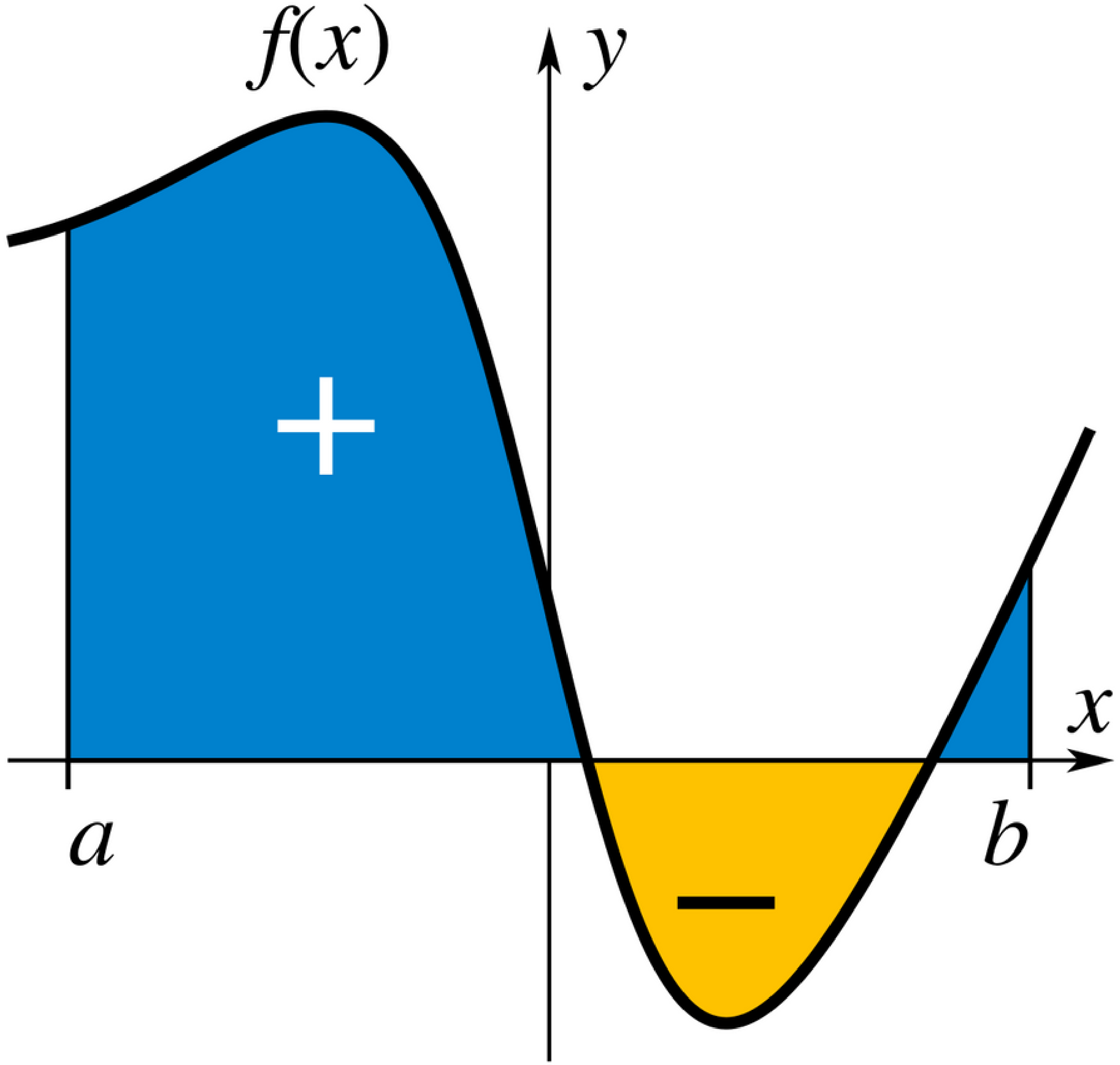
Methodology

All the methods of the study were performed in RStudio (ver. 4.1.2)

Methodology

Biphasic Genes' function

Calculating the integral between the low and high expression KM survival curves and find those tha have an intersection inside



Wikipedia (Integral - Wikipedia)

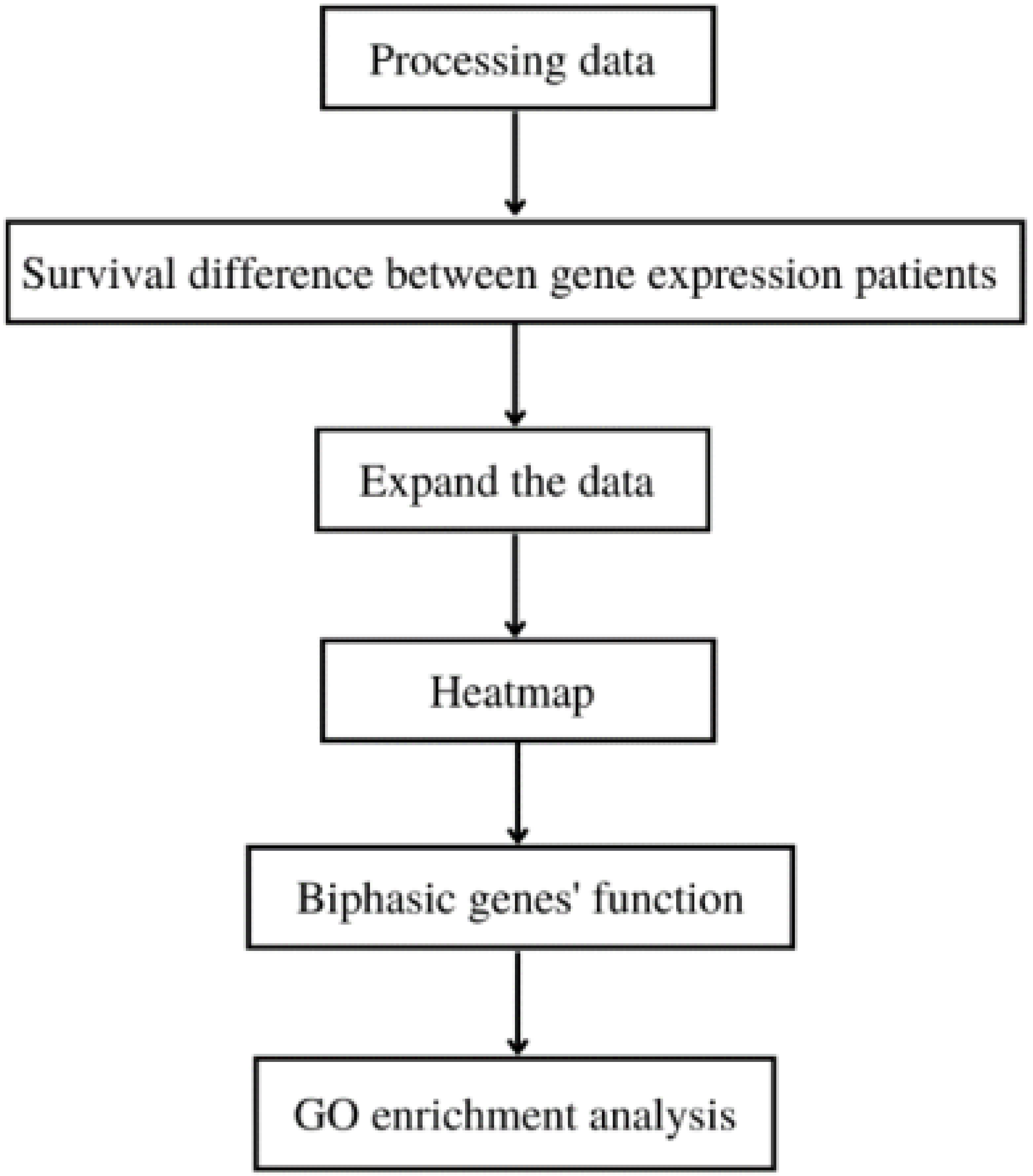
Methodology

Perform a GO enrichment analysis with *clusterProfiler* package

To understand its biology and functionality

Methodology

Bioinformatics pipeline



Processing data

Survival difference between gene expression patients

Expand the data

Heatmap

Biphasic genes' function

GO enrichment analysis

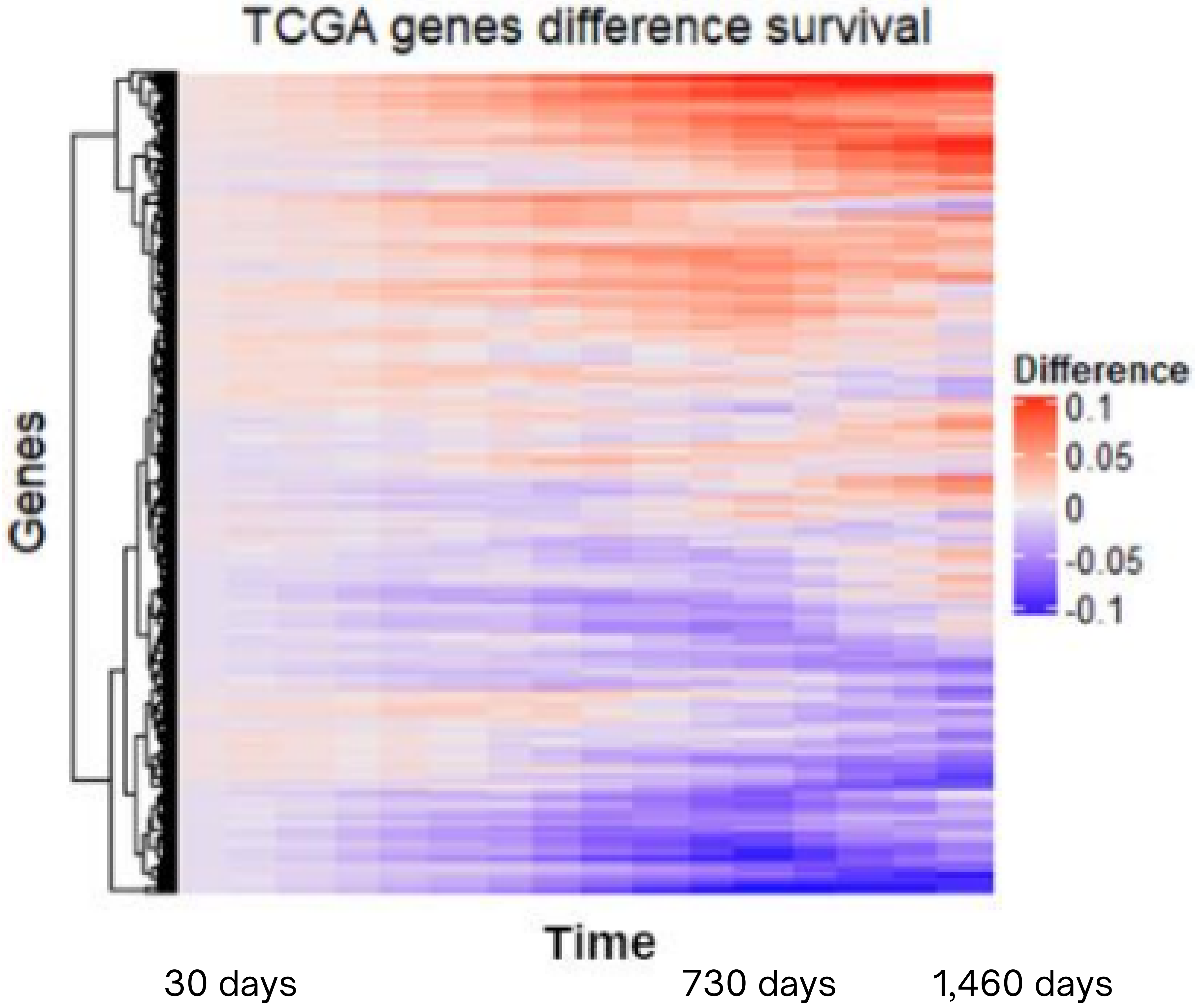
Methodology

Previous steps

DIFF_SURVIVAL

EXPAND_DATA

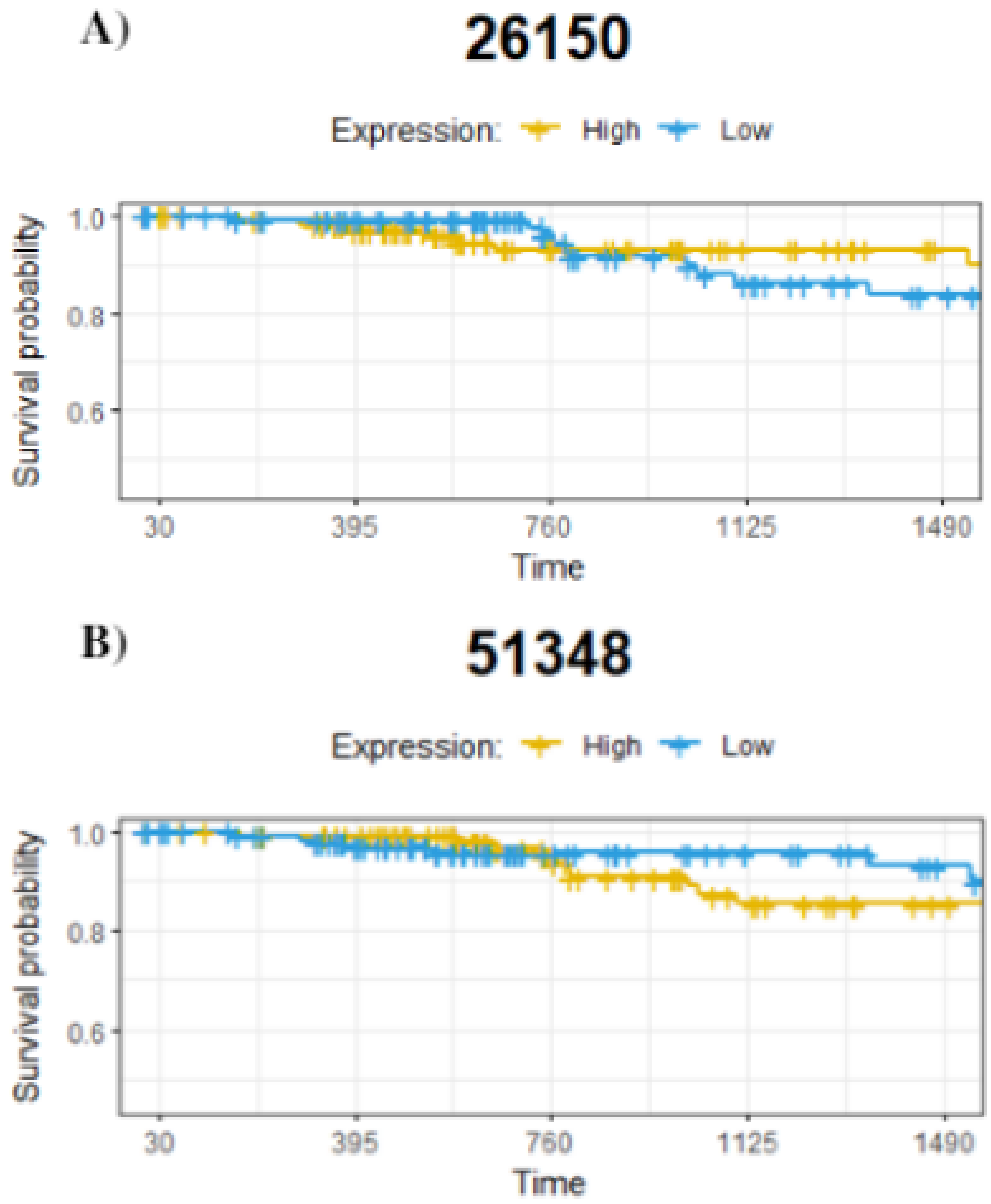
ComplexHeatmap::Heatmap



Results

BRCA-TCGA

861 genes whose expression shows a dual association with patient outcome from a total of 15,748 genes



509 genes

352 genes

Results

BRCA-TCGA

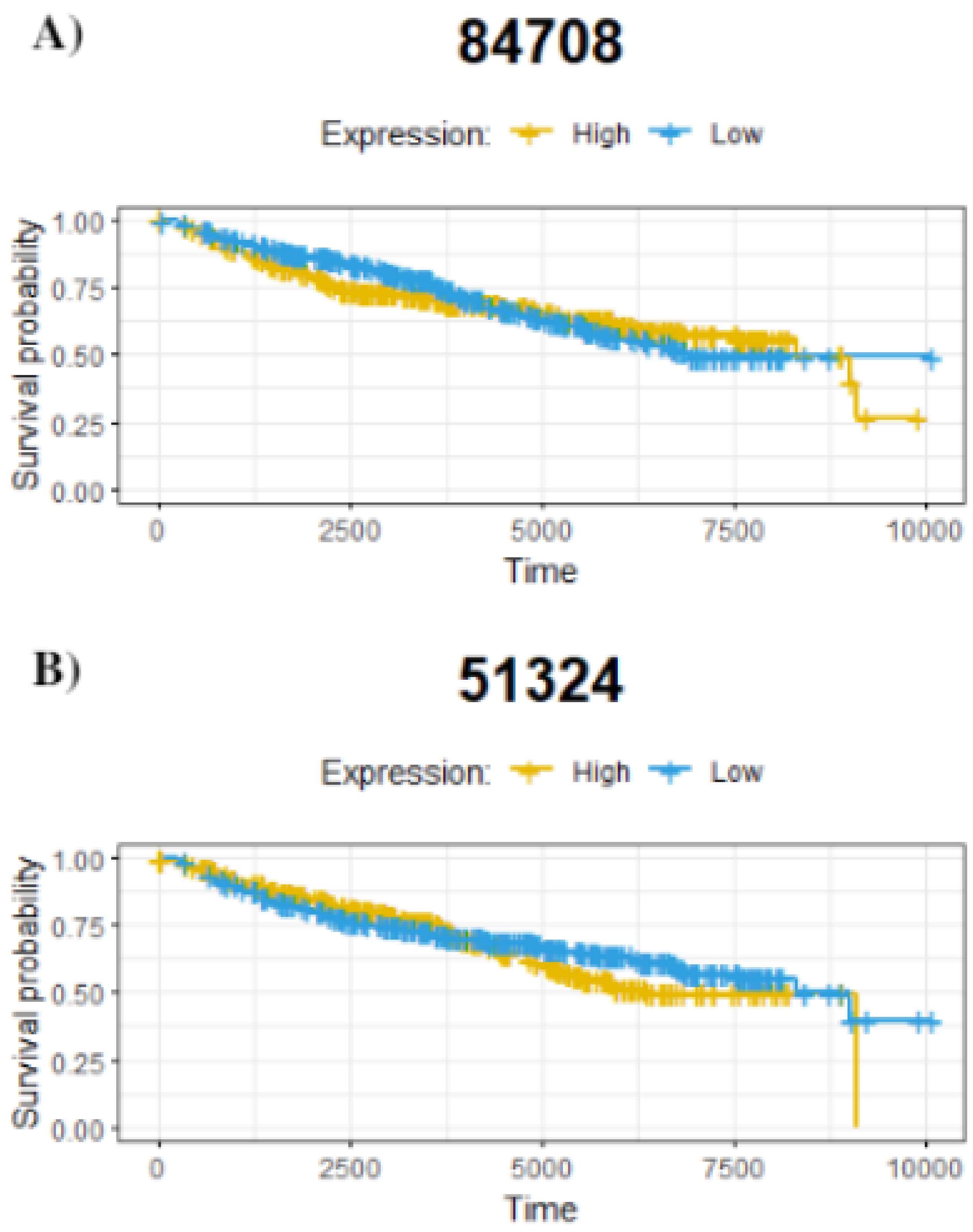
GO enrichment analysis (509 genes)

Double-strand break repair via nonhomologous end joining biological process

Results

METABRIC

249 genes whose expression shows a dual association with patient outcome from a total of 18,492 genes



120 genes

129 genes

Results

METABRIC

GO enrichment analysis (120 genes)

- Neurotransmitter transport processes

GO enrichment analysis (129 genes)

- Immune response regulating signaling pathway

Discussion & conclusions

BRCA-TCGA

509 genes
Low gene expression patients > High gene expression patients --> early stages
Low gene expression patients < High gene expression patients --> late stages

Genes that act as oncogene in early stages and TSG at late stages

Discussion & conclusions

BRCA-TCGA

352 genes
Low gene expression patients < High gene expression patients --> early stages
Low gene expression patients > High gene expression patients --> late stages

Genes that act as TSG in early stages and oncogene at late stages

Discussion & conclusions

METABRIC

120 genes
Low gene expression patients > High gene expression patients --> early stages
Low gene expression patients < High gene expression patients --> late stages

Genes that act as oncogene in early stages and TSG at late stages

Discussion & conclusions

METABRIC

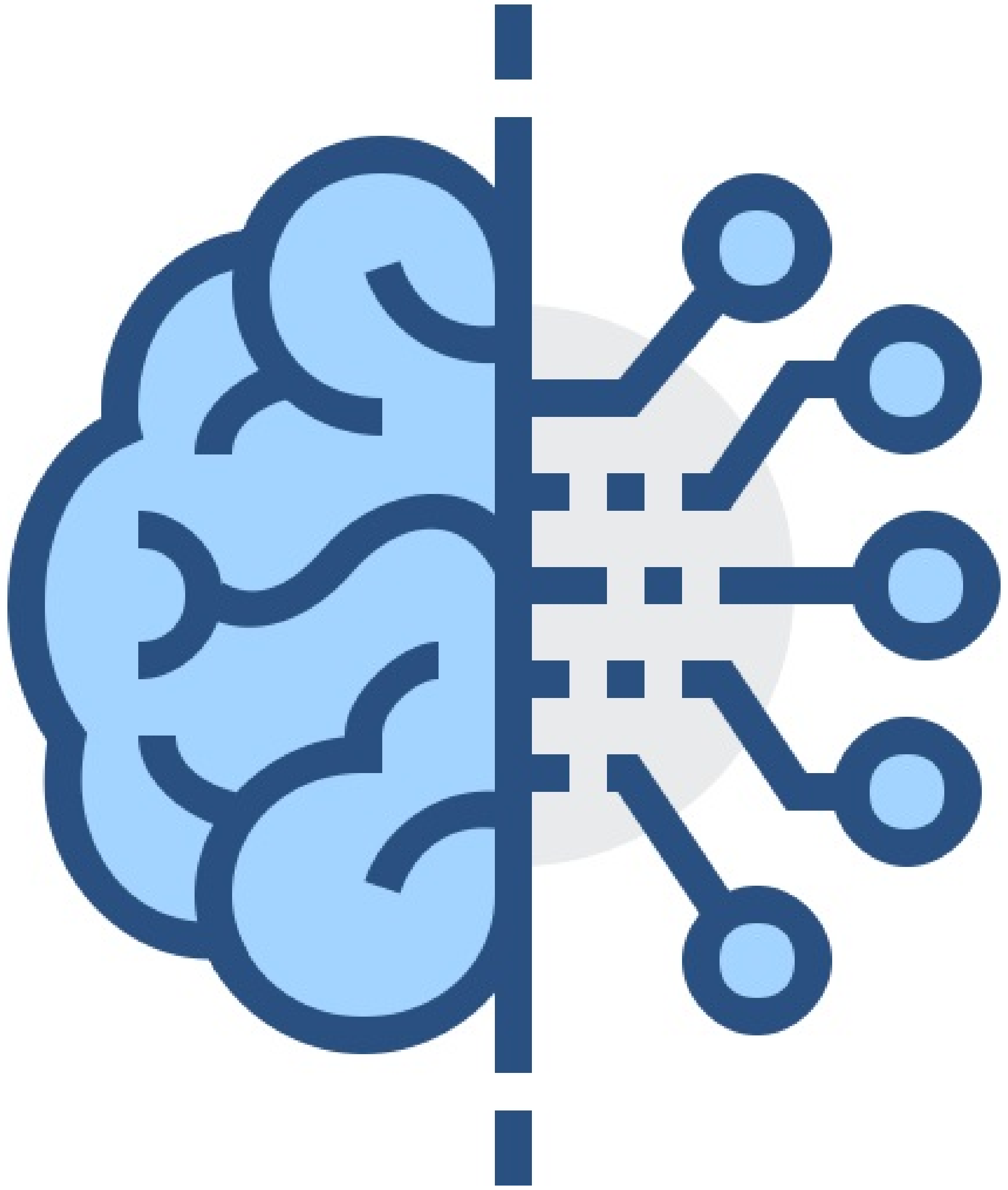
129 genes
Low gene expression patients < High gene expression patients --> early stages
Low gene expression patients > High gene expression patients --> late stages

Genes that act as TSG in early stages and oncogene at late stages

Discussion & conclusions

FUTURE DIRECTIONS

Use Machine Lerning algorithms to find a better and accurate way to identify this class of genes



UVIC
UNIVERSITAT DE VIC
UNIVERSITAT CENTRAL
DE CATALUNYA

IDIBELL
Institut d'Investigació Biomèdica de Bellvitge

Thank you very much for your attention!

You can contact me at
arnau.soler@crg.eu