Survival Analysis of Pediatric Neuroblastoma Patients Using Gene Expression Data: Identification of Novel Neuronal Differentiation Related Biomarkers

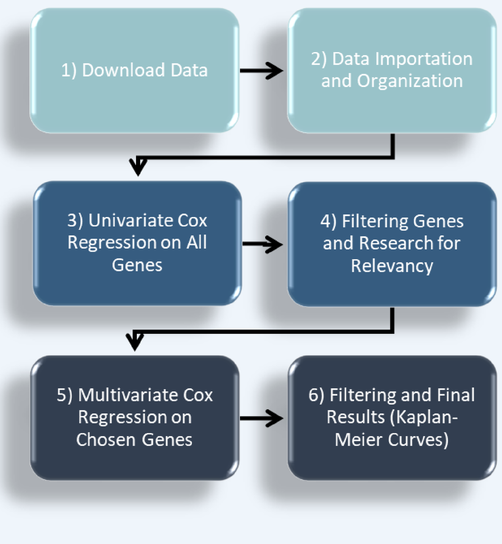
Introduction

Neuroblastoma (NBL) is the most common cancer in babies under the age of one, with a high-risk survival rate below 50%. By the time neuroblastoma is diagnosed, it has often spread beyond the tissue where the cancer starts, and common forms of neuroblastoma treatment are often ineffective or impossible if the cancer has metastasized. While gene therapy and precision medicine are far more effective treatments than the traditional chemotherapy, radiation therapy, or surgery, there is a need for biomarkers in order to develop such therapies. This project seeks to investigate such novel biomarkers, with a focus on the NEK2 gene and other genes involved in neuronal differentiation (specifically the Wnt and Notch signaling pathways) by analyzing the gene expression data of 249 neuroblastoma patients through the creation of multiple levels of Cox regression models to compare gene expression levels to patient prognosis.

Hypothesis

1. Elevated levels of **NEK2** will correlate with poorer prognosis in NBL patients as NEK2 is an important mitotic regulator and likely plays a role in tumorigenesis and metastasis of neuroblastoma
2. Genes associated with **neuronal differentiation** and the **Wnt and Notch signaling pathways** will serve as NBL prognostic biomarkers due to the nature of NBL formation and as the Wnt signaling pathway controls apical polarity of neuroblasts and activation of the Notch pathway inhibits neuronal differentiation.

Methodology




```java
public class AllGenesExpressions {

   public static final File folder = new File("C:/Users/menon/OneDrive/Documents/R/TARGET");
   private static PrintStream output;


   public static void main(String [] args) throws FileNotFoundException {
      output = new PrintStream(new File("AllGenesExpressions.txt"));
      listFilesForFolder(folder);
   }

   public static double listFilesForFolder(final File folder) throws FileNotFoundException {
      double geneExp = 0.0;
      for (final File fileEntry : folder.listFiles()) {
      Scanner files = new Scanner(fileEntry);
         if (fileEntry.isDirectory()) {
            listFilesForFolder(fileEntry);
         } else {
           String geneName = files.nextLine();
           //System.out.println(fileEntry.getName() + " ");
           while (files.hasNextLine()) {
            geneName = files.next();
            //System.out.println(geneName);
            geneExp = files.nextDouble();
            print("\n" + geneExp);
            geneName = files.nextLine();
            //System.out.println(total);
           }
         }
      }
      //System.out.println(total);
      return geneExp;
   }

   private static void print(String stat) {
      output.print(stat);
   }
}
```

Step 2 in Java

```r
library(survival)
options(expressions = 5e5)
mydataset <- read.delim("/Users/menon/OneDrive/Desktop/csrsef files/FIRST6001Genes.txt")
myFinalData6001 <- mydataset[, -1]
EFST <- myFinalData6001[ , c("EFST")]
Status <- myFinalData6001[, c("Status")]
View(myFinalData6001)
coxformula <- sapply(names(myFinalData6001)[5:ncol(myFinalData6001)], function(y) as.formula(paste('Surv(EFST, Status)
coxmodel <- lapply(coxformula, function(y) coxph(y, data = myFinalData6001))
coxreg <- lapply(coxmodel, function(y) return(cbind(HR=exp(coef(y)), exp(confint(y)), Pval=coef(summary(y))[5])))
coxreg
lapply(coxreg, write, "First6001Values.txt", append = TRUE)
```

Step 3 in R

```r
library(survival)
multivariate <- read.delim("/Users/menon/OneDrive/Desktop/csrsef files/FIRST6001GENES.txt")
eventfreesurvivaltime <- multivariate$EFST
geneValues <- multivariate$ULK4
statusvital <- multivariate$Status
agecats <- cut(multivariate$Age, breaks = c(1, 1095, 2555, 5000), include.lowest = TRUE)
gender <- multivariate$Gender
SurvObjGene<- with(multivariate, Surv(eventfreesurvivaltime, statusvital == 2))
SurvObjGene
coxallmulti <- coxph(SurvObjGene ~ geneValues + factor(agecats) + factor(gender), data =  multivariate)
summary(coxallmulti)$coefficients

quartiles <- Hmisc::cut2(geneValues, g = 4)
fitAll <- survfit(SurvObjGene ~ quartiles, data = multivariate)
ggsurvplot(fitAll, data = multivariate, pval = TRUE)
```

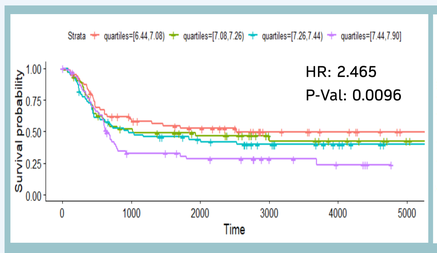
Step 5 in R

Results and Data Analysis

Step 6) Creation of Kaplan-Meier Curves and Final Results

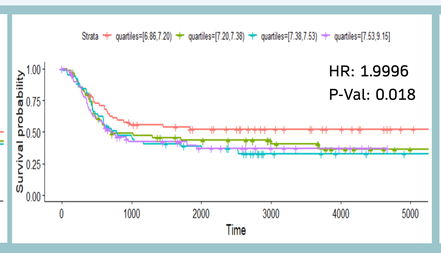
```r
library(survminer)
library(ggplot2)
quartiles <- Hmisc::cut2(geneValues, g = 4)
fitAll <- survfit(SurvObjGene ~ quartiles, data = multivariate)
ggsurvplot(fitAll, data = multivariate)
```

HES2



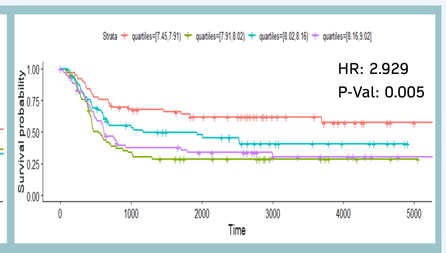
Involved in Notch/Akt pathway

HES6



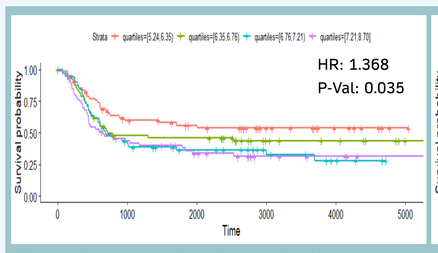
Involved in Notch/Akt pathway

TLX2



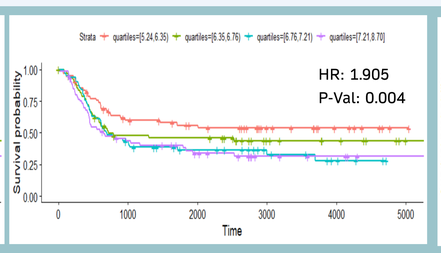
Encodes transcriptional factors for neural crest cell development

NEK2



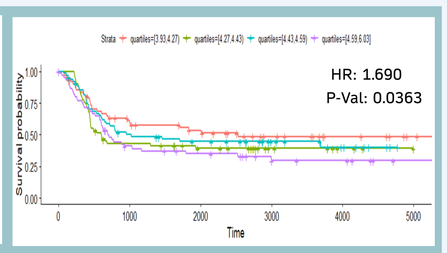
Regulates mitotic spindle formation

NOP14



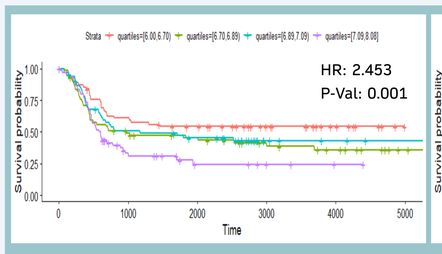
Inhibitor of Wnt pathway

ZIC1



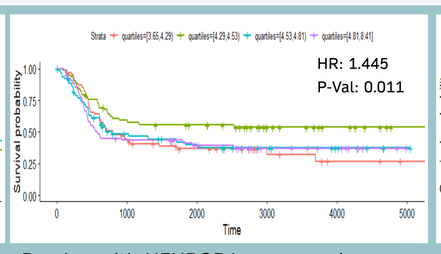
Drives neural crest cell differentiation in xenopus embryos

TEAD4



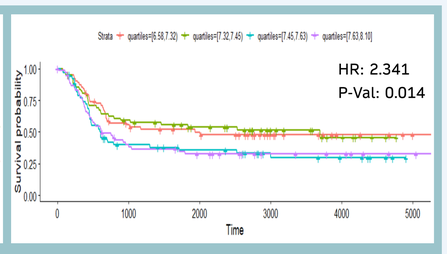
Associates with Wnt transcription factors and the Notch pathway

NEUROD4



Paralog with NEUROD1, neurogenic differentiation factor that promotes NBL

SOX12



Determining neural cell fate/differentiation (less investigated than other SoxC genes)

LHX5

Regulates neural crest differentiation, essential in mice for control of neuronal differentiation

CCDC88C

Negative regulator of WNT signaling pathway

CSNK2A2

Regulates WNT signaling by phosphorylating transcription factors

DLL3

Provides instructions for making a protein that regulates NOTCH pathway

EPN1

Involved in the Notch pathway, associated with other cancers

RSPO4

Activation of Wnt pathway, increases Wnt signaling

TRABD

Negative regulator of Wnt pathway

ULK4

Moderates Wnt Signaling

Discussion

Wnt Signaling Pathway: This pathway regulates cell differentiation. It is also highly evolutionarily conserved, and shown to control neuroblast migration in *Caenorhabditis elegans*, although this concept has yet to be thoroughly investigated in human neuroblasts. The genes identified in this project that are associated with the Wnt pathway suggest that there needs to be further investigation done with the Wnt pathway regarding how neuroblasts differentiate and become tumors.

Notch Signaling Pathway: The Notch pathway is of extreme importance in neuroblast formation/differentiation in *Drosophila melanogaster* as it helps establish apical-basal polarity, asymmetrical division, neural stem cell (NSC) self renewal, and cell differentiation, however the pathway has yet to be holistically investigated (beyond just the Notch genes) in humans.

Akt Signaling Pathway: This pathway is closely intertwined with the Notch pathway and helps establish the asymmetrical division necessary for proper neuroblast differentiation.

Conclusions

Both hypotheses that NEK2 as well as other genes associated with neuronal differentiation pathways (including those that are homologs in other animals but are yet to be investigated in humans) would serve as prognostic biomarkers, were supported with high statistical significance and significant hazard ratios (see results). In addition to NEK2, 16 other relevant and novel potential prognostic biomarkers were discovered through rigorous statistical analyses, opening new gateways that give further insight into the intricacies and pathways of NBL as well as pave the way for the creation of novel, noninvasive cancer gene therapy and rapid, preemptive cancer detection models.

Applications

Test these potential biomarkers with other existing datasets to determine whether they are consistently differentially expressed (currently pursuing)

Further research: apical cortex formation and asymmetric division of neuroblasts in humans using homologous *Drosophila melanogaster* and *Caenorhabditis elegans* pathways

Develop faster, earlier, and more accurate detections of pediatric NBL using these biomarkers, potentially before the cancer metastasizes since neural differentiation occurs early on

Develop novel targeted gene therapies using neural stem cells that deliver therapies that suppress uncontrollable neuroblast division (NSCs as delivery vehicles)