

**Exploratory Analysis of Gene Expression in Neuroblastoma: A Comparison of Sex-Specific
Mortality Cases Using Microarray Probe Intensities**

Justin T. Harris

The University of Birmingham

DL BIO - 37441 Health Data Analytics

26 June 2025

Abstract

This study investigates the gene expression profiles of 498 primary neuroblastomas obtained from RNA-Seq and Microarray data obtained from the study Zhang et al. (2015) [1]. After cleaning and preprocessing the data, the top expressed genes were identified in each sex and trends were observed based on age and genetic subgroups. The analysis revealed similar expression intensities among the top-expressed genes in both sexes, despite the subgroup differences (ST4S in males vs MNA in females). Due to the limited number of death cases (n=6) above the age of 25 months, it is recommended to have exploratory findings and future studies done with larger cohorts and unsupervised methods such as PCA to strengthen biological interpretations.

Introduction

Neuroblastoma is a pediatric cancer that arises from neural crest tissues and is biologically heterogeneous. Genetic subfeatures including MYCN amplification and clinical genetic subgrouping play a crucial role in prognosis. Previous studies (Matthay et al., 2016; Schleirmacher et al., 2012 [2,3]) have shown the relations between clinical outcomes and gene expression patterns; however limited research has examined sex-specific differences in expression among deceased patients.

Understanding whether differential expression has an effect on sex-specific survival outcomes can contribute to establishing treatment or prevention protocols, as well as progressing the knowledge of tumor biology. This study performs exploratory data analysis on clinical metadata linked with microarray probe intensity data to assess expression difference in male vs female patients that died from neuroblastoma.

Aim and Hypothesis

Aim: To identify and compare the most highly expressed genes in male and female neuroblastoma patients who died using microarray probe intensity data.

Hypothesis: There are sex-specific differences in the expression profiles of the top expressed genes in neuroblastoma death cases.

Method

Dataset Summary

Clinical and expression data were obtained from a published study (Zhang et al., 2015) which RNA-seq and microarray probe data compiled from neuroblastoma patients. In this study,

the microarray probe data was the primary focus. The probe file has log2-transformed intensity values with probe and gene IDs. Clinical data includes MYCN status, gender, INSS stage, and mortality or death outcomes.

Preprocessing

NaNs within the probe ID and gene symbol columns were removed due to their mapping role. Samples with 0 or unknown age (months) were filtered out; ages above 240 months or 20 years were considered biologically implausible and were excluded. Only deceased patients over 25 months of age were selected in the study, which resulted in 6 patients (3 male and 3 female).

Software and Tools

Raw data in csv format was initially visualized using excel. All analysis was performed using Python with Pandas, Matplotlib, and Seaborn libraries. Descriptive statistics and grouped comparisons were computed using `.mean()`, `.describe()`, `groupby()`, and `crosstab()` functions.

Statistical Methods

Mean expression per gene was calculated separately for female and male mortality groups. The top 10 expressed genes per group were visualized and identified using bar plots.

Results

Age and Mortality

Females tended to die earlier (approximately 30 to 100 months), while males died later at 100-150 months. MYCN was amplified (value=1) in all 6 death cases

Top Expressed Genes

The top 10 expressed genes per group had similar gene expression values, all within a high intensity range (≥ 11), which suggests strong transcriptional activity. Despite subgroup differences (ST4S in males vs MNA in females), no drastic differences in top gene expression were observed.

Visualisations

Bar plots showed similar expression ranges in the top genes between genders. No outliers or strong divergent gene signals were identified

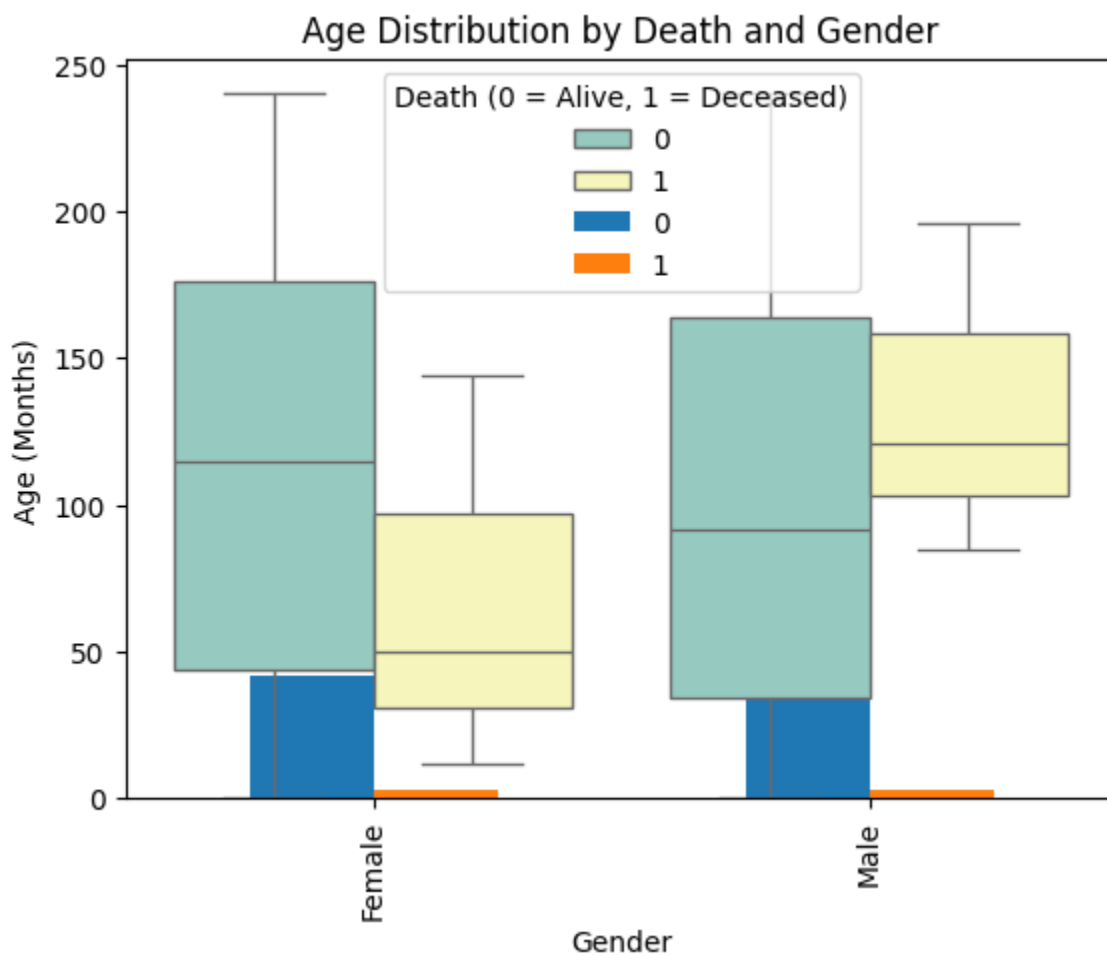


Figure 1 Age Distribution by Death and Gender: *The plot above shows the age and mortality distribution between male and female neuroblastoma patients. Green and blue bars indicate that the patients are alive while beige yellow and orange indicate that the patient has died. The orange bars show that there are patient deaths >25 months which were then excluded from the analysis.*

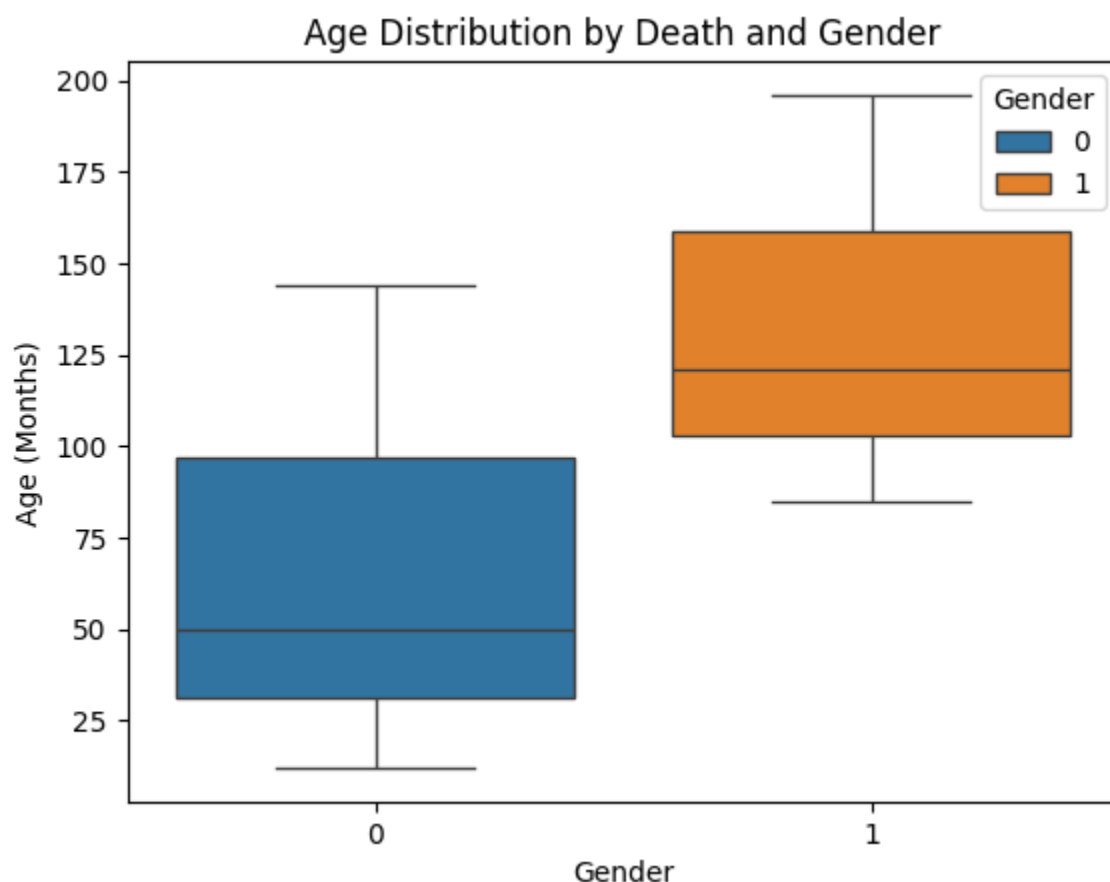


Figure 2 Age Distribution by Age and Gender >25 Months: *The graph above shows the age vs gender deaths with neuroblastoma patients age >25 months. The males are labeled orange or 1 while the females are blue or 0. There is a separation in the range of deaths for each as females died around 30 to 100 months while males died 100 to 150 months.*

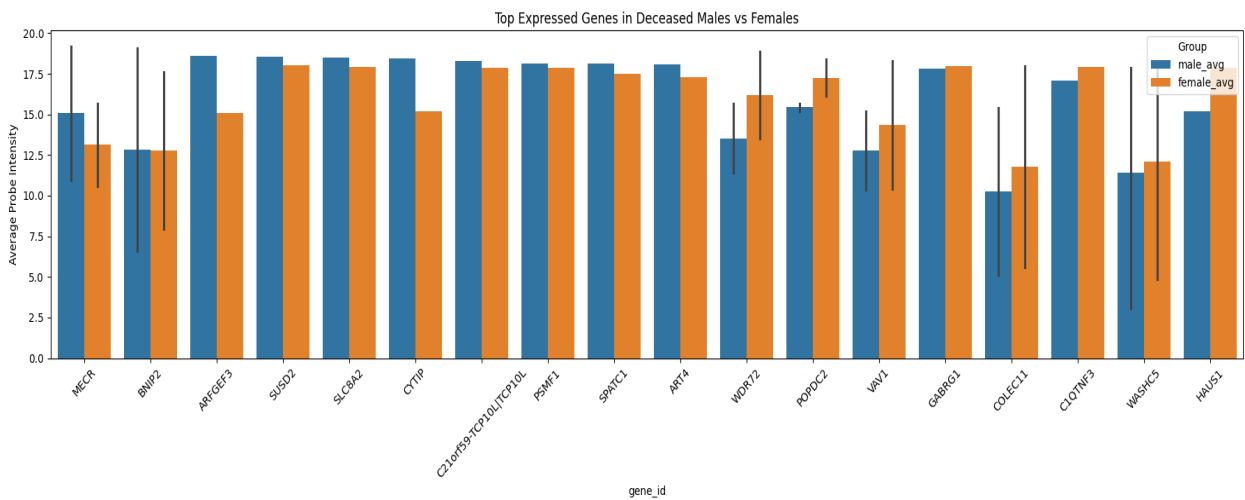


Figure 3 Top Expressed Genes in Deceased Male vs Females: *The bar plot shown above contains the top 10 expressed genes using probe intensities for males and females. Blue bars indicate male and orange bars indicate female gene expression data. Trends were very similar between genders with an overall log2 probe intensity of ≥ 11 .*

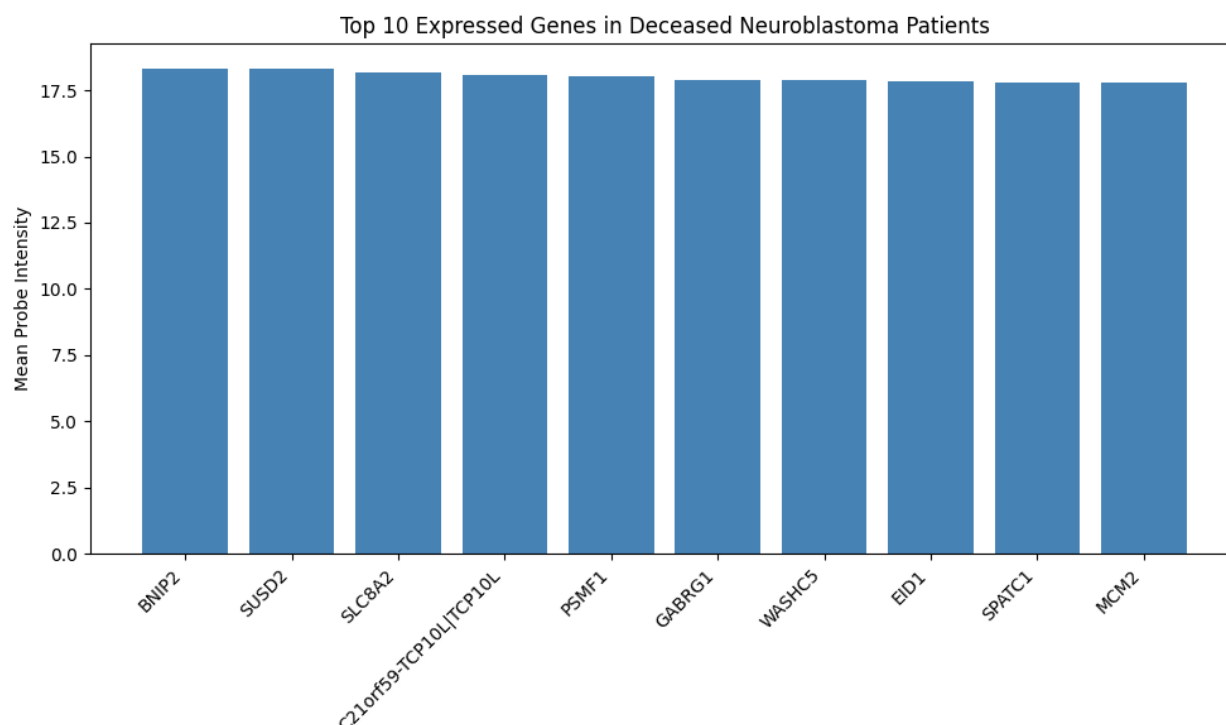


Figure 4 Top 10 Expressed Genes in Deceased Neuroblastoma Patients: *The bar plot displayed shows the top 10 expressed genes in both males and female neuroblastoma patients that have died. The mean intensities were calculated for each and the top means were ranked and plotted in the bar chart. All 10 genes have a high mean probe intensity which indicates high expression and significance.*

Discussion & Conclusion

The results suggest while clinical subgroups differed between female and male mortality cases, their top gene expression profiles were relatively similar due to their probe intensity. This may imply a shared set of genes active in advanced neuroblastoma stage(s) regardless of sex, or may reflect the limited statistical power due to a small sample size. MYCN uniform amplification in both sexes help support the notion of aggressive tumor profiles. Females were exclusively MNA and males were ST4 and ST4S, a pattern that requires more biological

exploration with larger datasets. Some of the limitations in the study include the small number of deceased patients and possible batch effects not accounted for. Also, the analysis did not incorporate unsupervised learning (e.g, PCA) or supervised classification due to the low number of samples. Future study or research should incorporate PCA or clustering to the full dataset linking probe expression or survival duration. This analysis provides a starting point for the integration of expression profiles linked to clinical outcomes and highlights the necessity of sex-specific stratification in neuroblastoma studies.

Top 10 Genes in Deceased Neuroblastoma Patients	
Gene ID	Mean Intensity (\log_2)
BNIP2	18.33
SUSD2	18.3
SLC8A2	18.19
C21orf59-TCP10L TCP10L	18.07
PSMF1	18.01
GABRG1	17.9
WASHC5	17.89
EID1	17.82
SPATC1	17.8
MCM2	17.79

Figure 6 Top 10 Genes in Deceased Neuroblastoma Patients: *The table above displays the top 10 expressed genes by mean probe intensities in deceased patients*

References

1. Zhang, W., Yu, Y., Hertwig, F. *et al.* Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. *Genome Biol* 16, 133 (2015).
<https://doi.org/10.1186/s13059-015-0694-1>
2. Matthay, K. K. et al. (2016). Neuroblastoma. *Nature Reviews Disease Primers*, 2, 16078.
3. Schleiermacher, G. et al. (2012). Segmental chromosomal alterations have prognostic impact in neuroblastoma. *British Journal of Cancer*, 107(8), 1418–1422.