IDENTIFYING GENES OF INTEREST THROUGH CLUSTERING AND OUTLIER ANALYSIS OF GENOMIC TIME SERIES DATA

Robin L. Belton
The Ohio State University Comprehensive Cancer Center - Arthur G. James Cancer Hospital and Richard J. Solove Research Institute

Introduction

Oncogenic phenotypes are commonly thought of to arise from a few “driver genes” that often behave differently than the majority of genes. Their detection can be performed through what is known as outlier gene analysis. Simultaneously, time series gene expression experiments are a powerful tool for aligning phenotypic with genomic data and revealing subtle gene expression shifts. However, standard clustering algorithms are often unable to distinguish between real and random patterns.

To overcome these limitations, we hypothesize that cancer driver genes may be detected from multidimensional gene expression time series data via outlier analysis based on the Mahalanobis distance when applied to a Gaussian mixture clustering model.

We apply this tool to the development of peritoneal metastasis in ovarian cancer. Specifically, peritoneal metastasis is local spread of disease within the abdomen. Unfortunately, untreated patients with peritoneal metastasis will live for approximately one year. Furthermore, this phenomenon is seen in multiple cancers. A mechanism that has been proposed for peritoneal metastasis is the breaking of small portions of tumor which then form spheroids and reattach elsewhere. Little information is known about the molecular drivers of spheroid formation. To this end, our laboratory has an in vitro model of this spheroid formation process and captured time series data for which we have performed outlier analysis using the novel methodology described.

Aims

- Explore whether Gaussian mixture cluster modeling / Mahalanobis distance can be used to detect outliers in temporal genomic data.
- Validate these targets in an ovarian cancer model of peritoneal metastasis.

Methods

To determine the genes of interest, we analyzed Illumina HT-12 genomic profiles of the HEYA8 cell line (ovarian cancer) as the cell line was induced to form spheroids. Expression values from 35,000 genes were recorded every six hours starting at time 0 and then every 6 hours to 72 hours. We then performed a clustering and outlier analysis to determine genes of interest using R. The clustering method used was a Gaussian mixture, and we determined outliers by computing the Mahalanobis distance between each relative gene expression and its corresponding mean relative expression. We chose the 100 genes with the largest Mahalanobis Distance to be the outlier genes.

Results

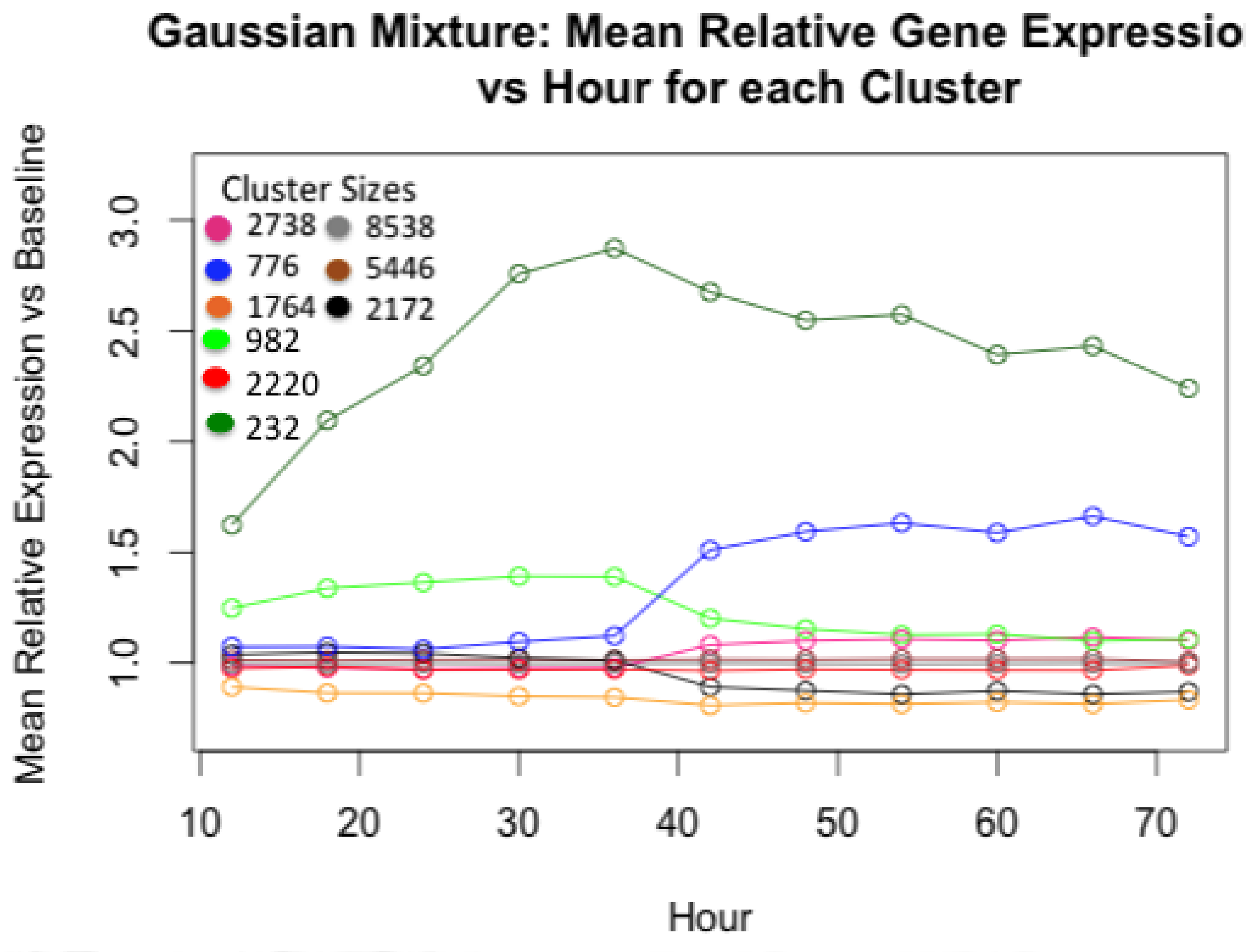
We established 9 clusters using the Gaussian mixture. The mean-relative expressions for these clusters are given in Figure 1, where each color represents a different cluster. The amount of genes assigned to each cluster is given in the upper left.

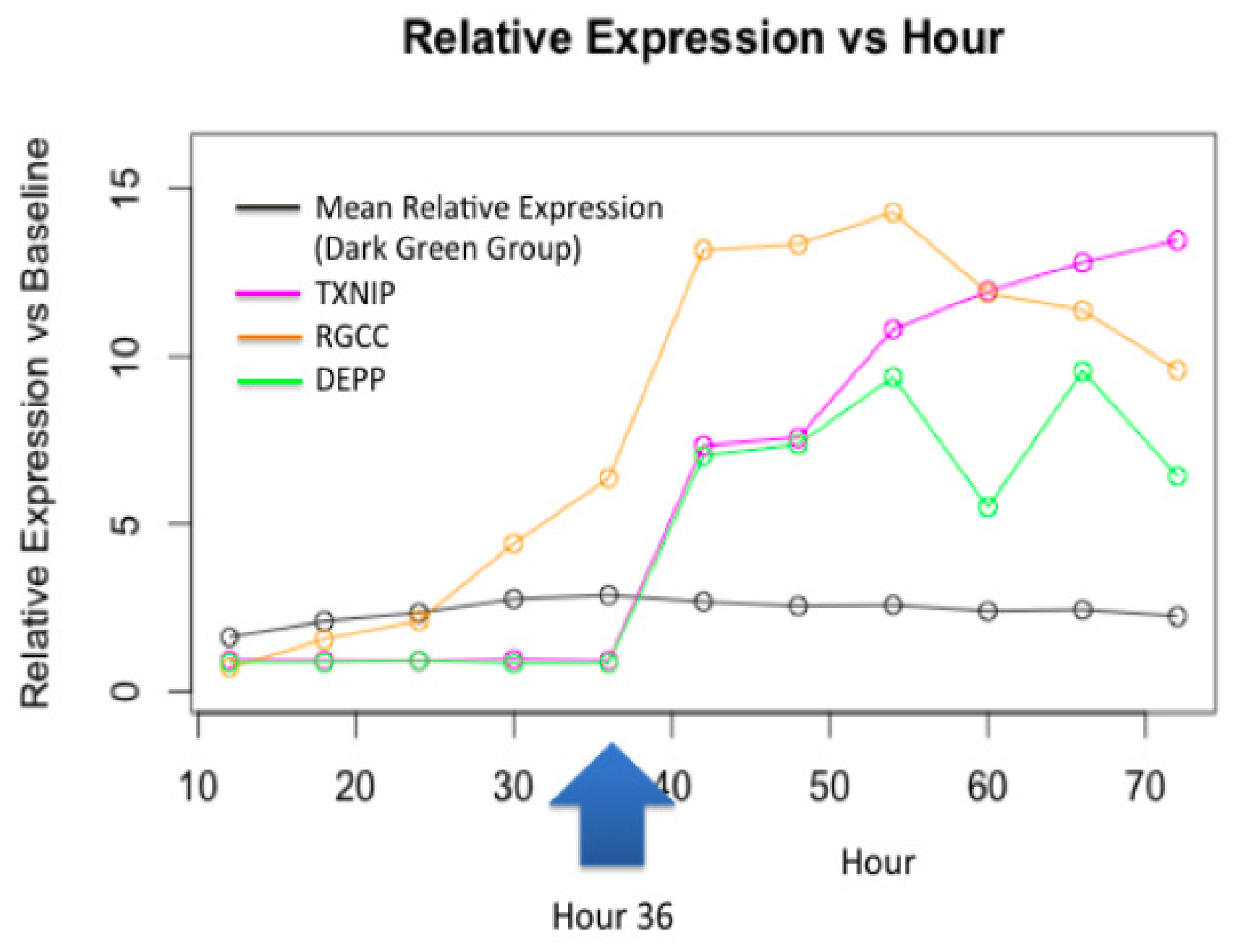
After computing the Mahalanobis distances between each gene and its assigned cluster’s center, we found that the outlier genes distances ranged from 5.3-26.

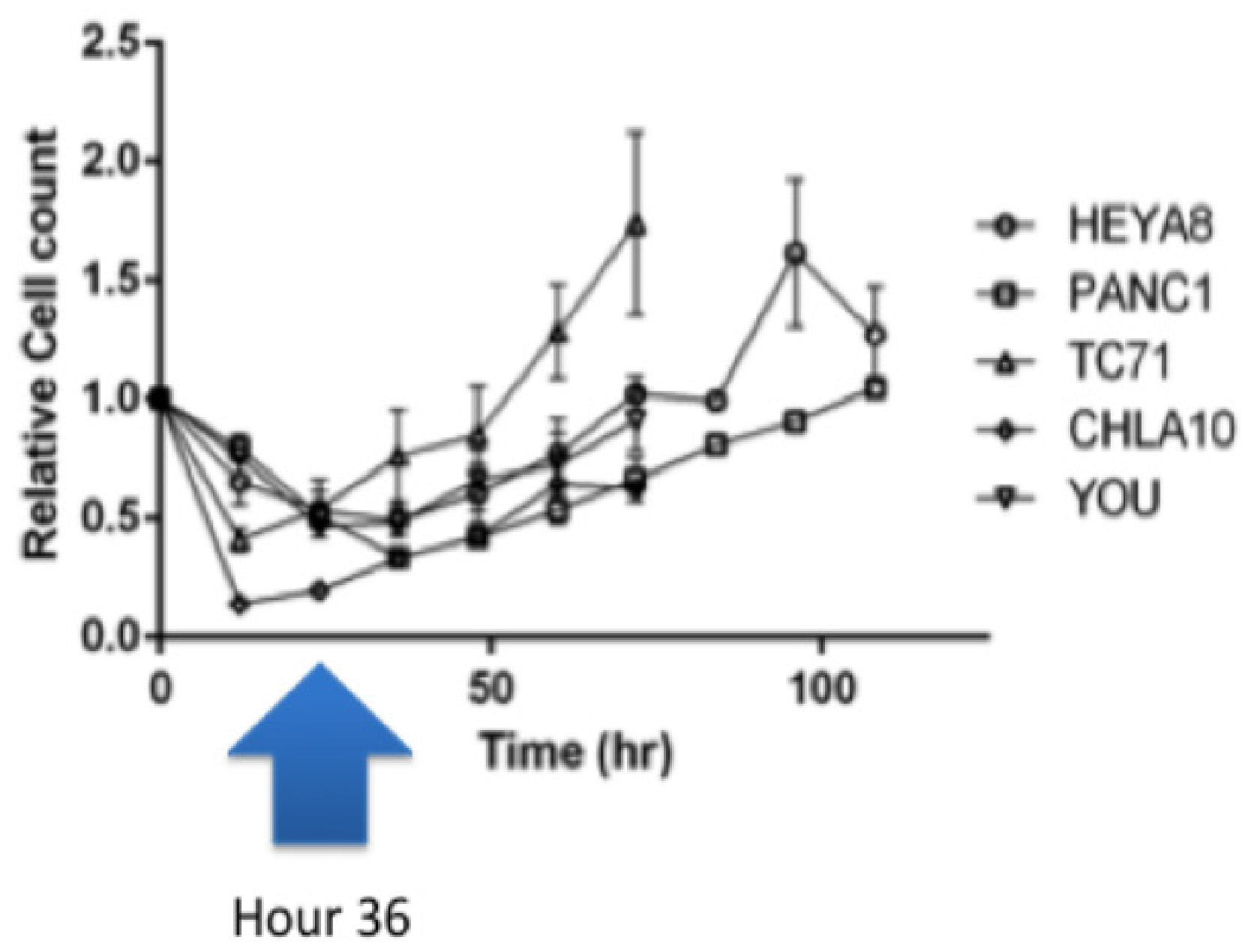
The three genes with the largest Mahalanobis distance were TXNIP, RGCC, and DEPP.

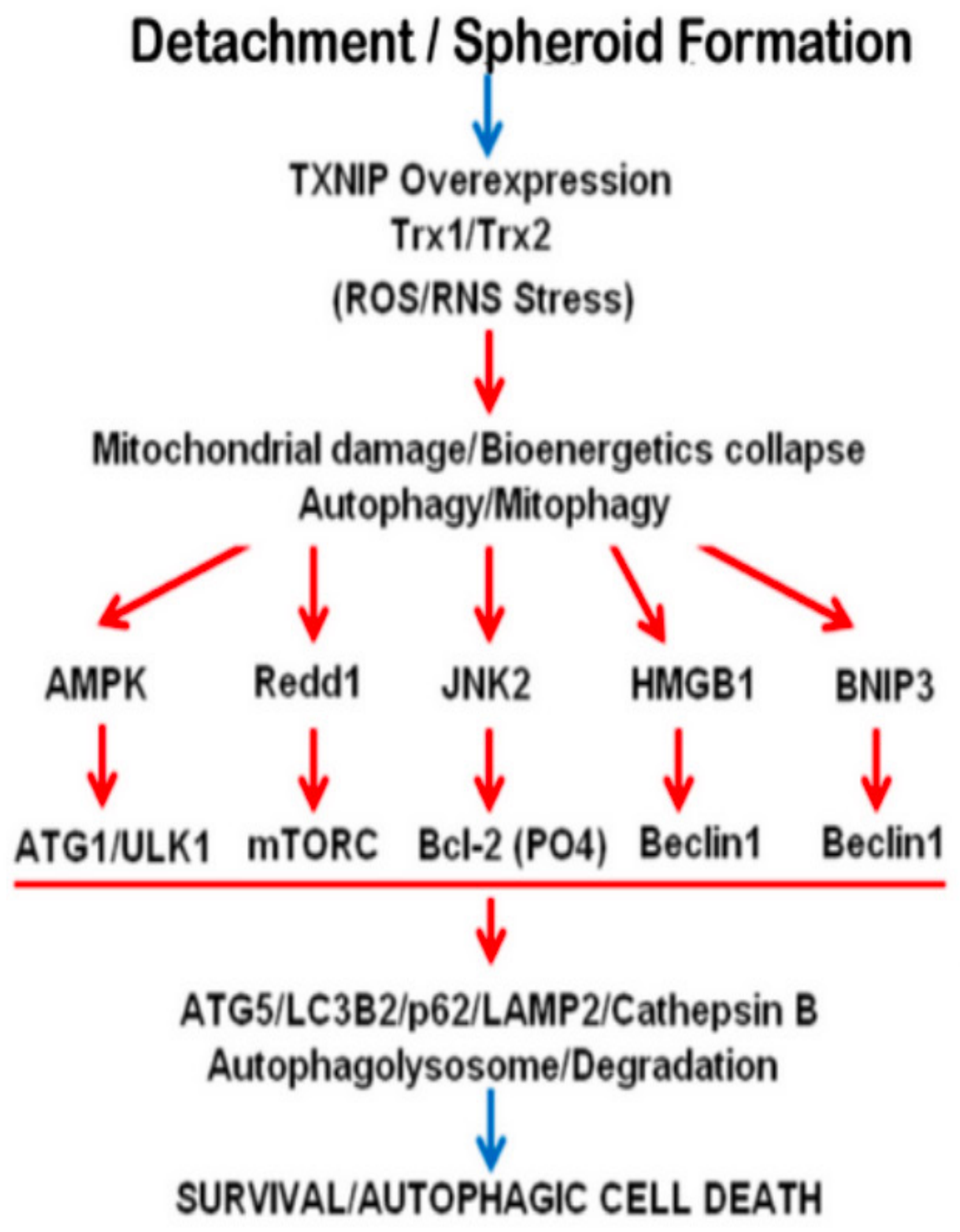
From the Figure 2, we observe that our outlier genes have low relative expression until about hour 36, in which relative gene expression increases greatly. We have also observed that around this time point, relative cell count increases greatly for several cell lines including HEYA8.

Furthermore, we have observed that formation of spheroids may lead to an over expression in TXNIP and release of stress that ultimately turns on autophagy. This autophagy may be linked directly to the spread of the cancer. Recent studies performed in the lab have shown that targeting autophagy using chloroquine prevents the formation of spheroids (see Figure 3). TXNIP protein expression was confirmed with WB validating our RNA findings.








Detachment / Spheroid Formation

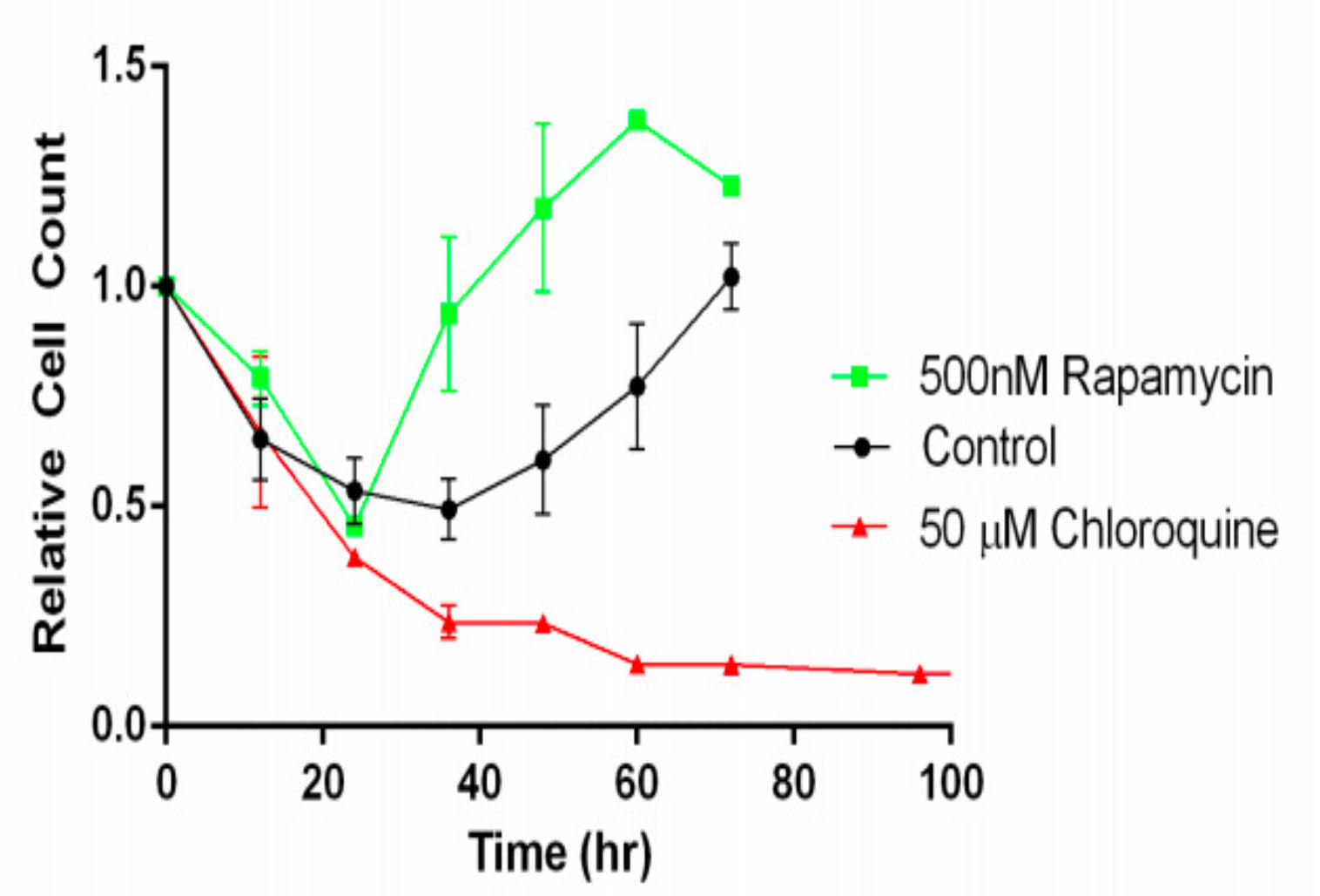
TXNIP Overexpression
Trx1/Trx2
(ROS/RNS Stress)

Mitochondrial damage/Bioenergetics collapse
Autophagy/Mitophagy

AMPK → ATG1/ULK1
Redd1 → mTORC
JNK2 → Bcl-2 (PO4)
HMGB1 → Beclin1
BNIP3 → Beclin1

ATG5/LC3B2/p62/LAMP2/Cathepsin B
Autophagolysosome/Degradation

SURVIVAL/AUTOPHAGIC CELL DEATH



Conclusion

- Outlier analysis of time series data using Gaussian Mixture and Mahalanobis Distance is feasible
- Identified genes are indicative of poor prognosis in ovarian cancer, although this will need to be verified
- Public Access: The software we developed to modify data, and perform clustering and outlier analysis is available at the BMI server.

Acknowledgements

I would like to thank: The Ohio State University Comprehensive Cancer Center, Pelotonia, Dr. Jim Chen, Dr. John Hays, Mark Burch, Kenyon College.

The Ohio State University

TDA@Ohio State