

# Cancer survival prediction and integration of multi-omics integration with Supervised Autoencoders, Stacked Autoencoders and Concrete Supervised Autoencoders for multiple correlated driver genes

Team Papaki



Alexis Action

University of Alabama, Birmingham



Vasileios Alevizos

Karolinska Institutet



Yishu Qu

Northwestern University



Vanessa Xiao

Massachusetts Institute of  
Technology



Zongliang Yue

University of Alabama, Birmingham

[github.com/u-brite/team\\_papaki](https://github.com/u-brite/team_papaki)



# Project Overview

Breast cancer is the second most diagnosed cancer among women in the United States.

- Heterogeneous
- Several different molecular subtypes
- Clinical, pathological, and molecular characteristic differences contribute to breast cancer progression
  - Patient prognosis, survival and therapeutic significance

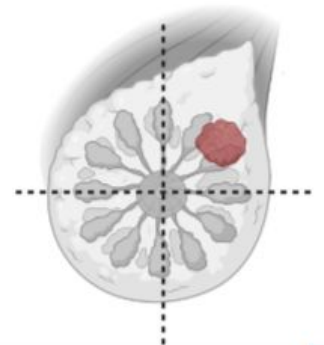
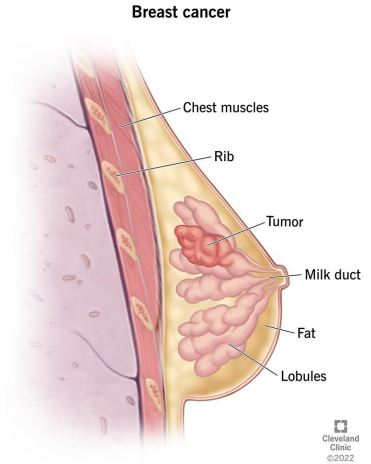
## Purpose:

- Deep learning methods are utilized in cancer prognosis prediction using genomic information
  - Single layer omics data (mRNA)
- Autoencoders are tools that allow for the integration of multi-omics data for cancer prognosis prediction

## Goal:

To compare these existing algorithms to advance and optimize the methods for better utilities

- Categorizing molecular subtypes
- Risk-level for cancer staging



# Remarks on selected frameworks

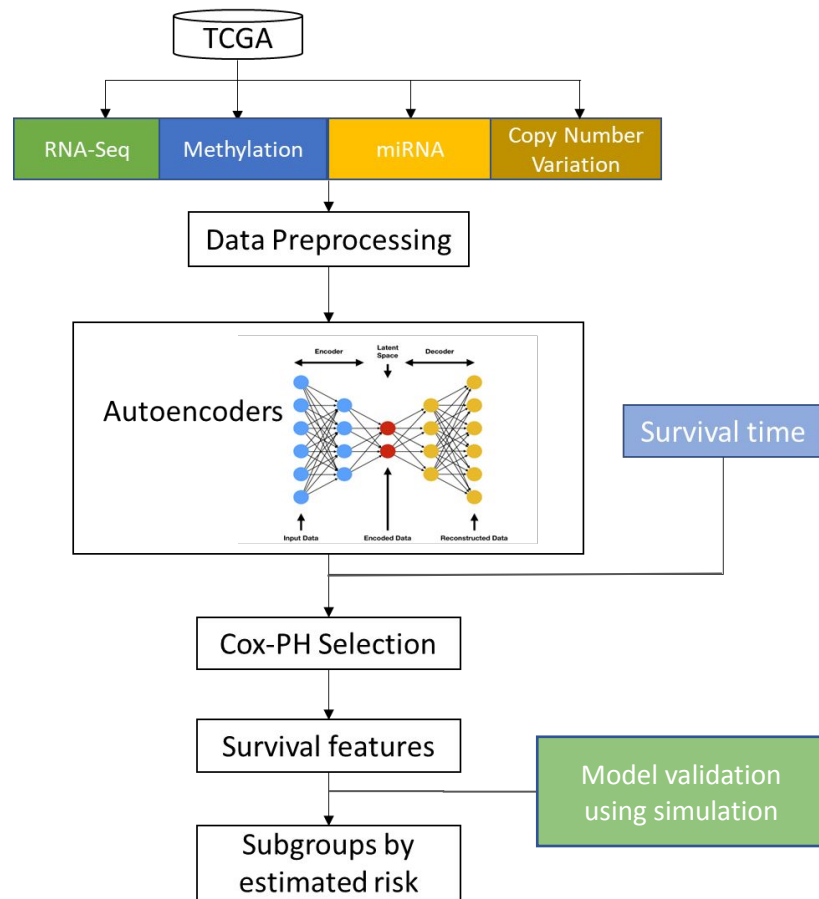
*Denoising Autoencoder for accurate cancer prognosis prediction* (DCAP) is a framework that allows for the integration of multi-omics data by denoising that is utilized to accurately estimate cancer risks through the Cox model.

The DCAP in this study improved the C-index values by 6.5% compared to that of previous methods.

*eXtreme Gradient Boosting* (XGboost) was used for selection of a small number of genomic features that are considered to be correlated to tumorigenesis and risk factor level of breast cancer

The XGboost models were shown to achieve an average C-index values of 0.627.

Moanna (Multi-Omics Autoencoder-based Neural Networks Algorithm) DL based model facilitating classification with semi-supervised autoencoders



# Future work and next steps

## Experiments

1. Integration of GAN based autoencoders (Ahmed et al, Bioinformatics 2022 & Yang et al, Bioinformatics 2021)
2. Cancer survival analysis utilizing subtype detection techniques
3. Prioritize omics profiles/modifications

## Room for improvements

1. Assessment through k-fold Cross-validation
2. Scale-up training pipeline with Apache Spark
3. Integrating other types of omics data, e.g., metabolomics, proteomics (Alakwaa et al, J Proteome Res 2018 & Lewis et al, Nat Commun 2021)

# Thank you for your kind attention!

Questions and Feedback?



Thanks again to all the team members and organizers for hosting this amazing hackathon!