

# Machine Learning Technique for Omics and Imaging Data:

An Application in TCGA-GBM

# **Motivation**

→ Histologic evaluation of tumor microenvironment shows spatial information but can have problems with variable interpretation by observers.

→ Genomic analysis has higher quantitative precision but difficulty isolating healthy from cancerous cells. Sequencing has better resolution for individual cells, but lacks clinical penetration.

## **Hypothesis and Goals**

**Our central hypothesis is that expression and methylation status of mitochondrial carrier proteins may predict the survival outcomes of glioma patients.**

**Goal 1: Incorporate imaging and omics data to improve understanding of cancer survival trajectory.**

**Goal 2: Create a novel pipeline for combining multimodal data using machine learning.**

# Data

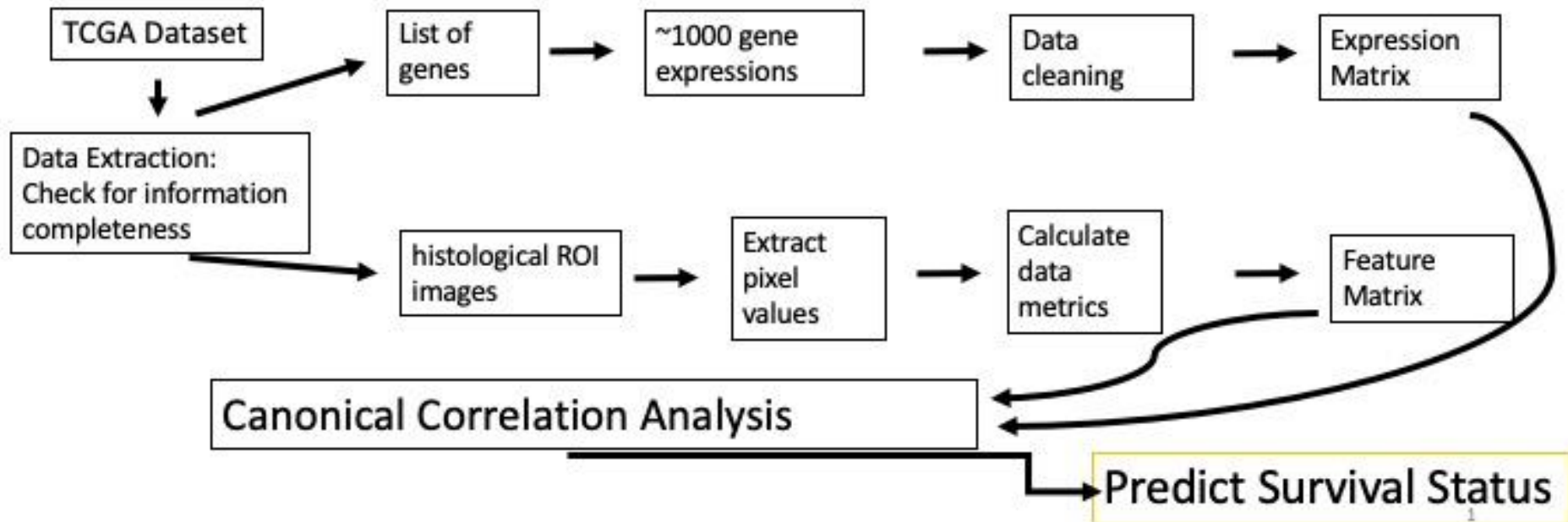


**NATIONAL CANCER INSTITUTE**  
**GDC Data Portal**

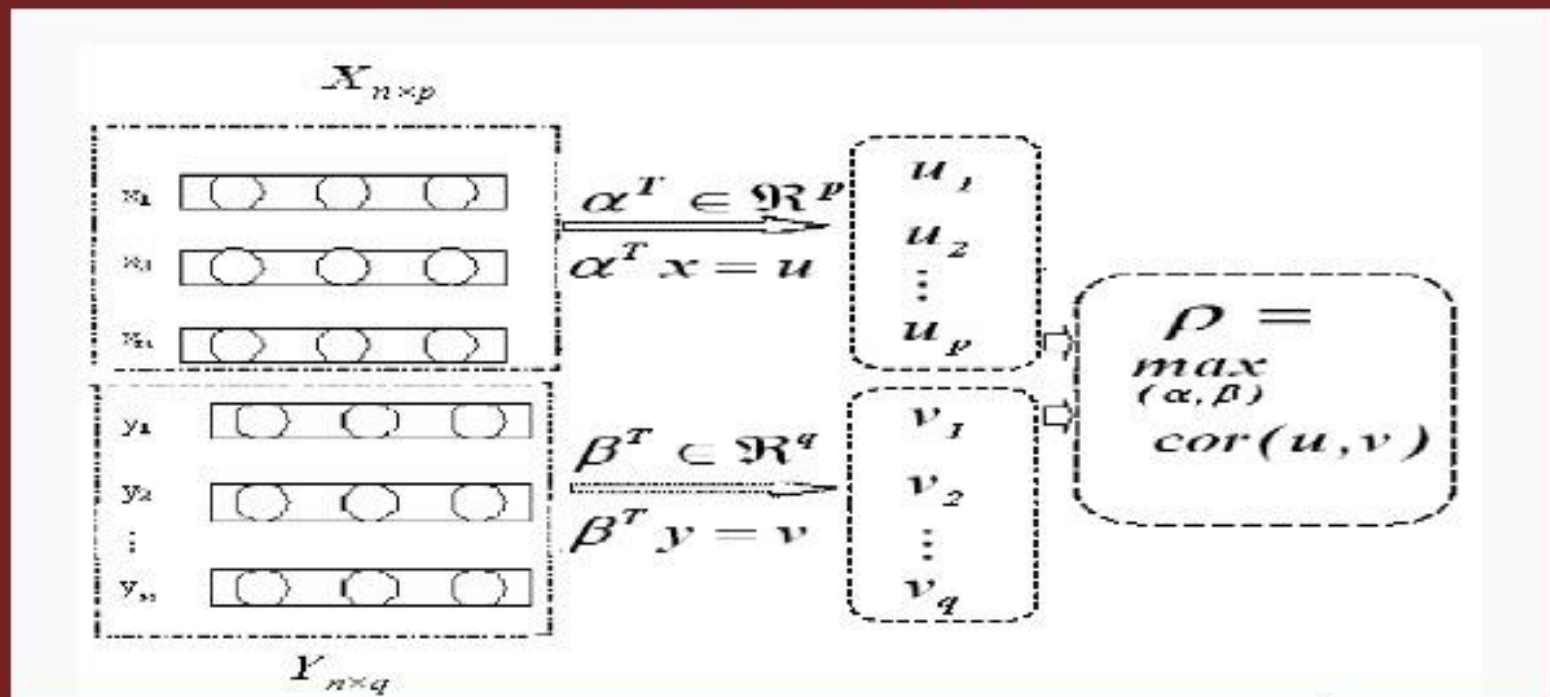
---

- Public data set (“TCGA”)
- Omics + histologic images + clinical information
- Glioma brain tumor patients
- 63 patients (52 cases, 11 control)
- 2 omics measurements: methylation and gene expression data (825 genes)
- Imaging: tissue slide (ROIs pre-selected, 1024 x 1024 2D matrix)
- Clinical outcome: Alive/Dead (number of months to death)

## Pipeline of the proposed supervised framework



## THE ARCHITECTURE OF DIMENSIONALITY REDUCTION: MULTIPLE CCA BASED APPROACH



## THE ARCHITECTURE OF CANONICAL CORRELATION ANALYSIS (CCA)



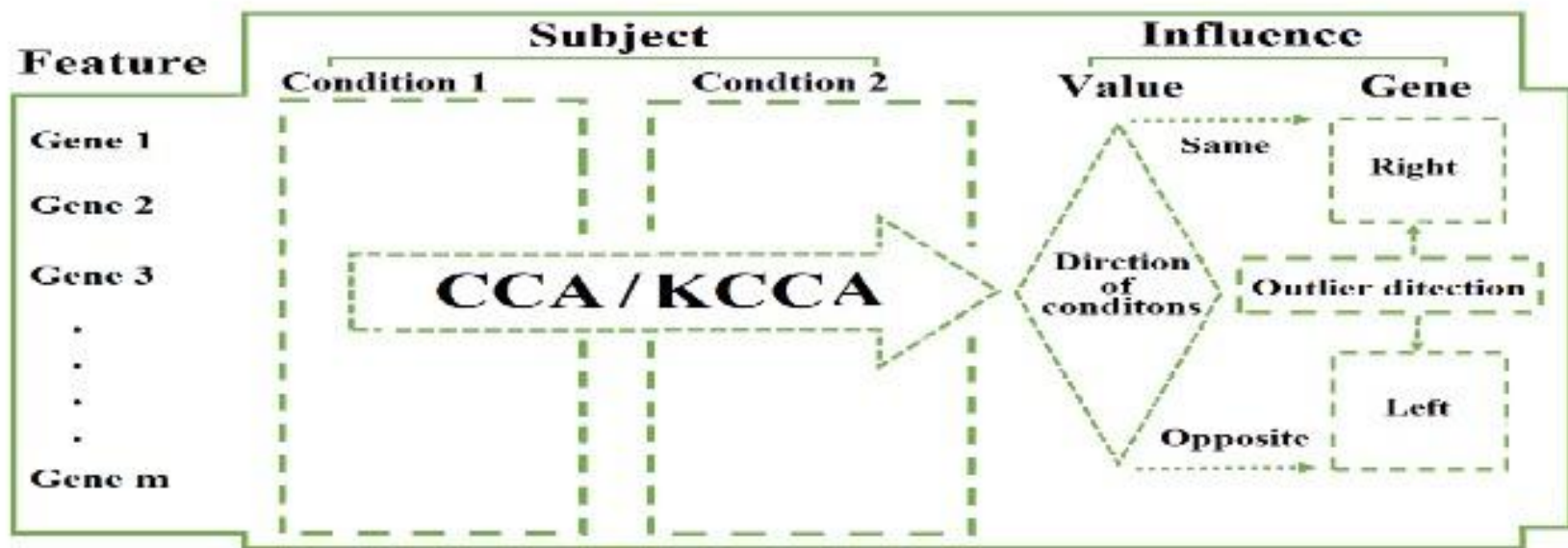
# THE ARCHITECTURE OF DIMENSIONALITY REDUCTION: MULTIPLE CCA BASED APPROACH

Multiple CCA seeks more than two sets of components for which the correlation of random variables is maximized

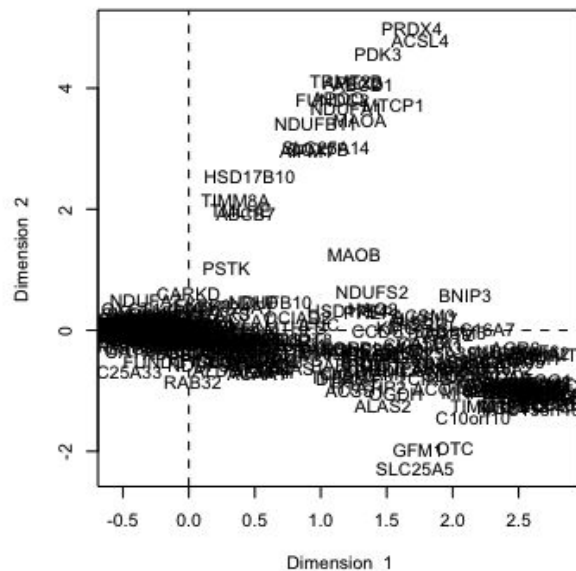
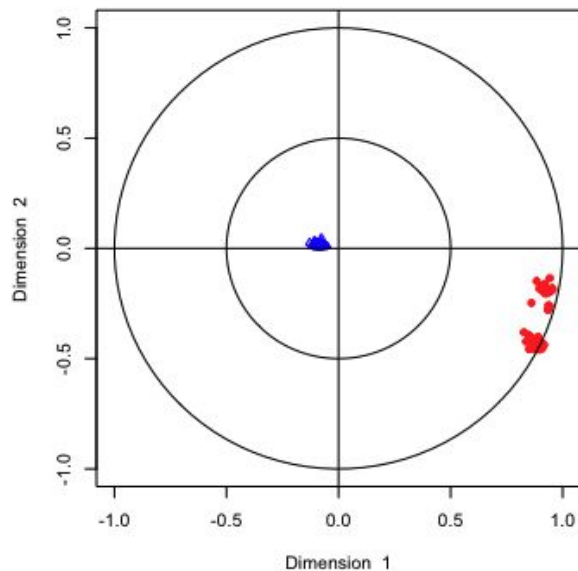
$$\max_{\substack{f_1 \in \mathcal{H}_{X_1}, \dots, f_p \in \mathcal{H}_{X_p} \\ f_1 \neq 0, \dots, f_p \neq 0}} \sum_{j=1, j' > j}^p \text{Corr}(f_j(X_j), f_{j'}(X'_{j'}))$$



# THE ARCHITECTURE OF DIMENSIONALITY REDUCTION: INFLUENCE BASED APPROACH



# Results - Regularized Canonical Correlation Analysis





# Next steps

- Incorporate the imaging data
  - Apply Fourier feature extraction
- Consider including other omics data
- Clean up code and documentation
- Generate a Manuscript

# Thanks and credits page

## **Team Members:**

Md Ashad Alam (Team Lead) ♦ Amin Cheikhi (Co-Lead & Co-Writer) ♦ Soner Koc ♦  
KuanJui "Ray" Su (Sys Admin) ♦ Becky Waugh ♦ Chengzhe Tian ♦ Udana Torian  
(Co-Writer)

## **Literature and Data References:**

Pathomic Fusion: An Integrated Framework for Fusing Histopathology and Genomic Features  
for Cancer Diagnosis and Prognosis. Richard J. Chen and Ming Y. Lu and Jingwen Wang and  
Drew F. K. Williamson and Scott J. Rodig and Neal I. Lindeman and Faisal  
Mahmood(2020)1912.08937.arXiv

National Cancer Institute GDC Data Portal (<https://portal.gdc.cancer.gov/> )