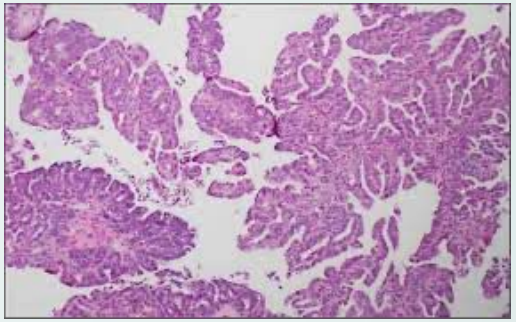
UNSUPERVISED MACHINE LEARNING FOR SEROUS OVARIAN CANCER SUBTYPING

Rebecca Aviles,
Jeanne Michelle Revilla,
Andrea Velazquez

Ovarian Cancer

- Ovarian Cancer
  - A cancer that affects the reproductive system.
  - 5th overall deadliest type of cancer
    - Late prognosis and lack of symptoms

      Age at Diagnosis: 55-64 years
  - 5 year survival rate : 49.7% (2011-2018)



Lisio, et. al, 2019

Serous Ovarian Cancer

Serous Ovarian Cancer

- Epithelial cancer cell
- 75% of overall cases

Low Grade Serous Ovarian Cancer (LGSOC)

- Slower abnormal cell growth
- Cancer cells are more similar

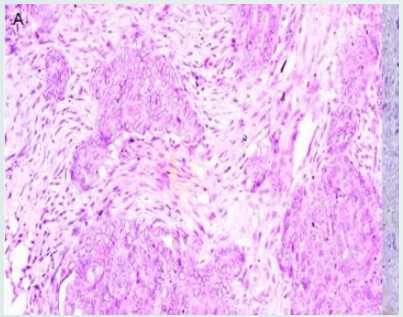
High Grade Serous Ovarian Cancer (HGSOC)

- More common within Serous subtype
- Faster abnormal cell growth
- High metastasis rate

HGSOC Subtypes

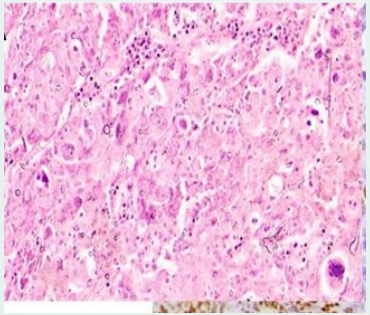
Mesenchymal

- Cellular stroma



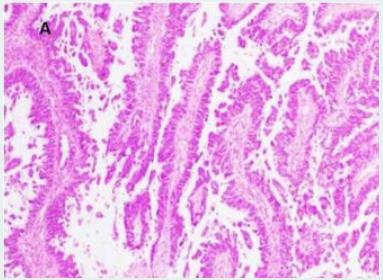
Proliferative

- Mitotic figures
- Nuclear aggregates



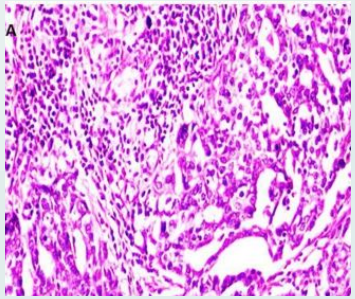
Differentiated

- Papillary structures
- Intratumoral lymphocytes



Immunoreactive

- Lymphocytes-infiltrating tumors



Khashaba et. al, 2022

Motivations

- Identify genetic markers for subtype characterization
- No existing test for Ovarian Cancer detection
    - Current diagnosis is invasive
- Optimize treatment and predict prognosis

Research Question

Using transcriptomic TCGA data of **High-Grade Serous Ovarian Cancer**, can an **unsupervised machine learning** model cluster the data into distinct **molecular subtypes**?

Methods

- "TCGA-OV" dataset **(n=429)**
  - RNAseq
  - R and Python
- Principal Feature Analysis: **246 genes selected**
  - 95% variance
- Clustering
  - 5 clusters– 1 outlier cluster (n=5)
- Analysis- **4 clusters**
  - Clinical factors: age and survival
  - Pathway analysis
  - Specific gene expression

Figure 1 : Clusters

- Cluster 0 (dark blue), n = 13
- Cluster 1 (teal), n = 166
- Cluster 2 (green), n = 36;
- Cluster 3 (yellow), n = 209.
- Uncertainty (dark purple), n =5 .

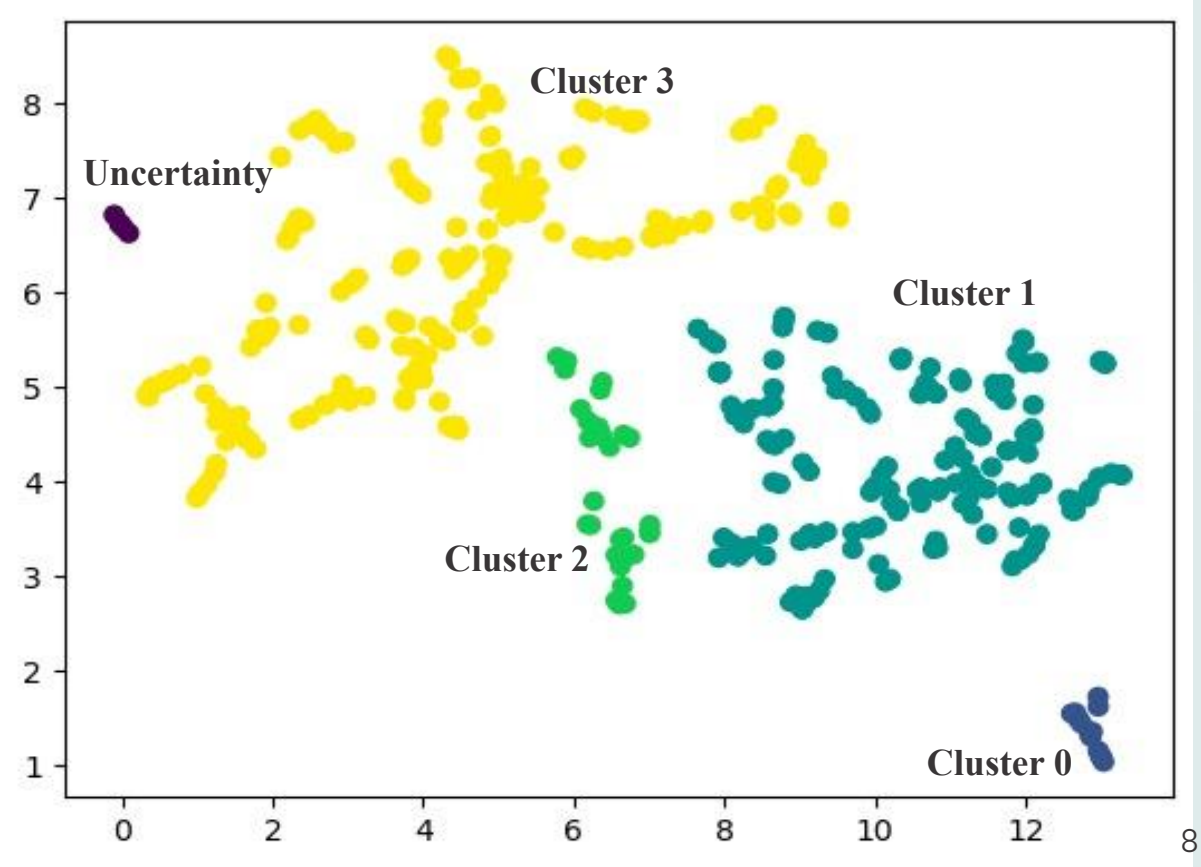
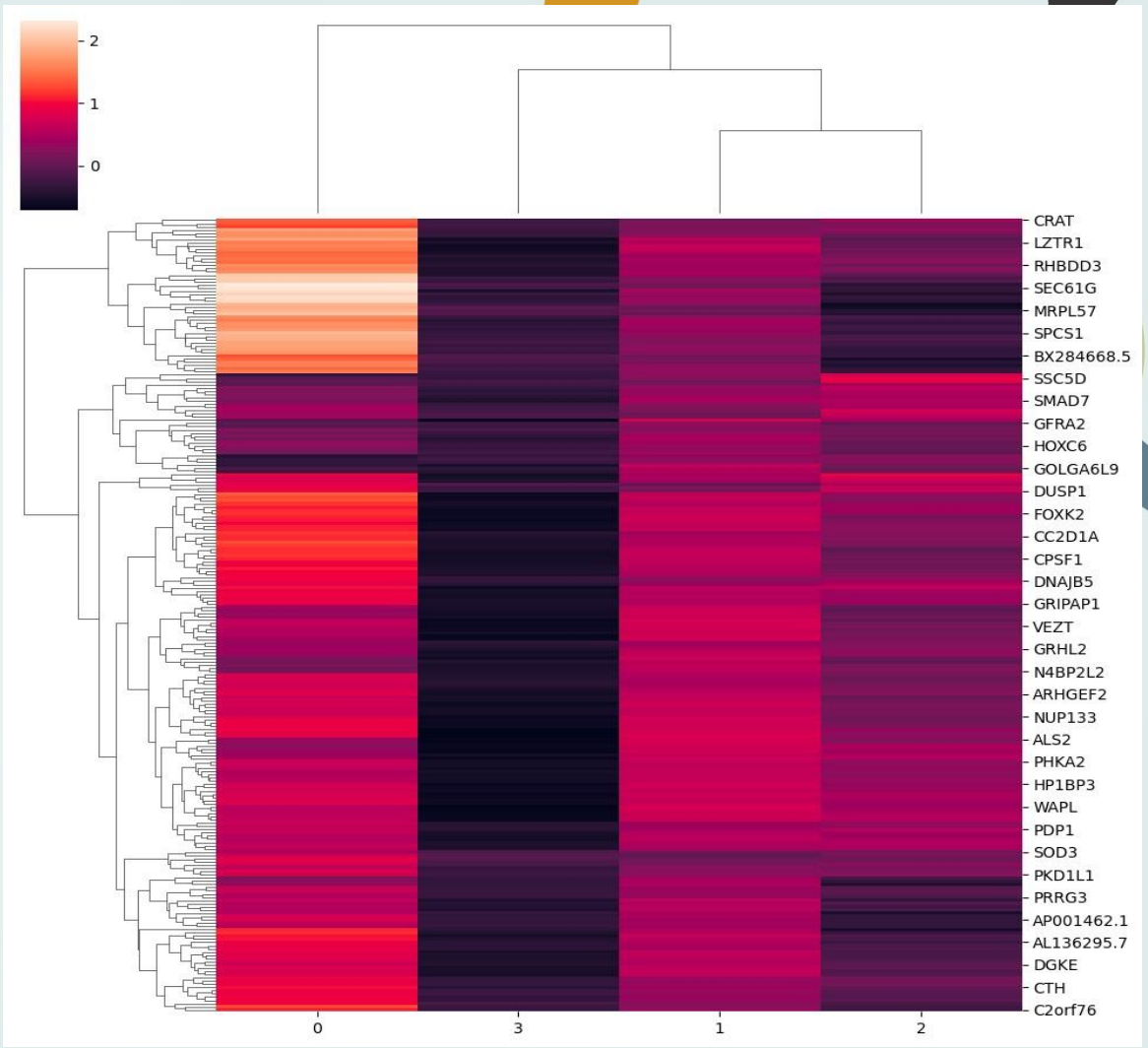


Figure 2: Gene Expression by Cluster

Heat-map for gene expression by clusters (labeled at the bottom)



Common Identifiers

mesenchymal	immunoreactive	differentiated	proliferative

● ⇧stromal components
● ⇧epithelial-mesenchymal transition (EMT)
● **Angiogenesis**
● **More likely to be metastatic**
	● Immune-related
	● ⇧T-cell markers
	● ⇧cell death protein PD1
	  ○ PDCD1
	● **⇧programmed death-ligand (PDL1)**
	  ○ **CD274**
		● **⇧ ovarian tumor markers**
		  ○ **MUC1**
		  ○ **MUC16**
		● **high expression of transcription factors/proliferative markers**
			● **⇧MKI67**
			● **⇧PCNA**
			● **high expression of transcription factors and proliferative markers**

Kasshun et. al, 2021, Khashaba et. al, 2022

Mesenchymal Analysis

- *No TCGA samples are defined as metastatic*

PANTHER Analysis – pathway

UNIQUE genes from Top 50 most expressed in **Cluster 3**

2 / 21 genes related to angiogenesis pathway

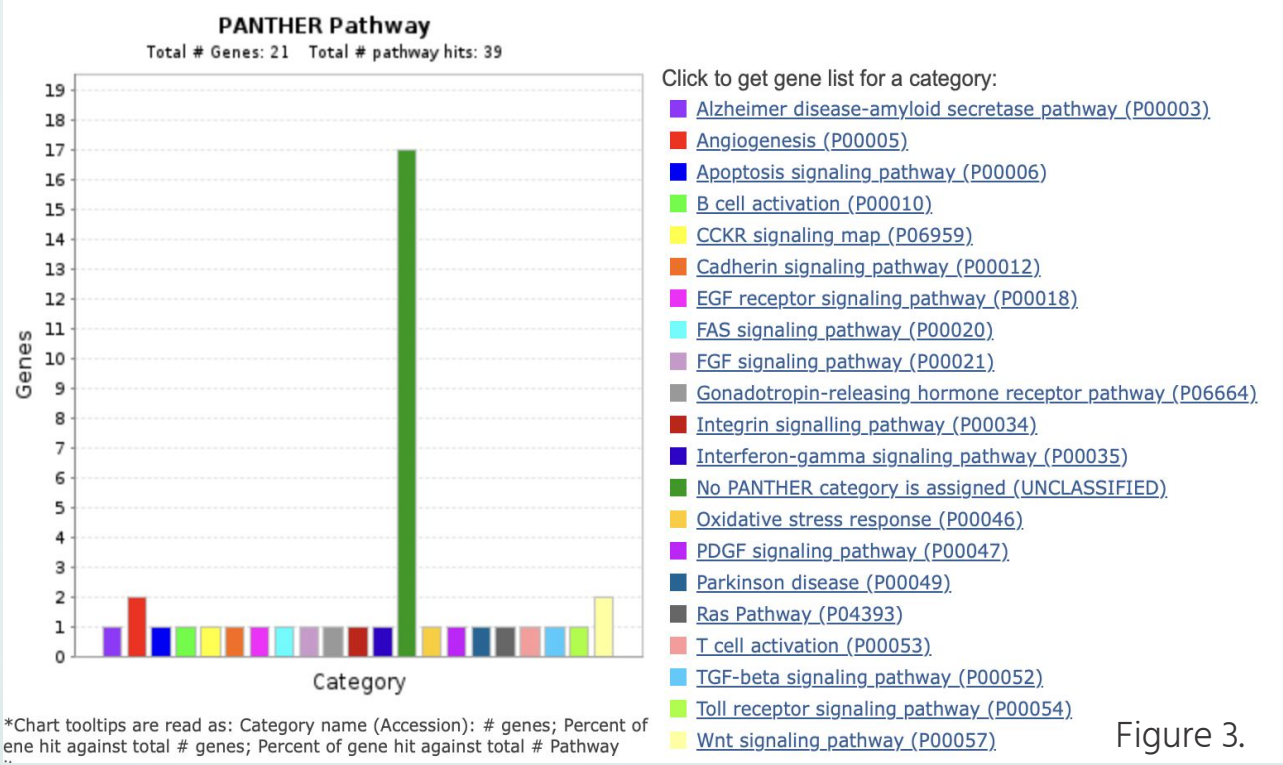


Figure 3.

Immunoreactive Analysis

Analysis of PDL1 Protein (gene CD274):

- ★ Relatively similar expression
- ★ High outliers in **Cluster 1** have possible significance

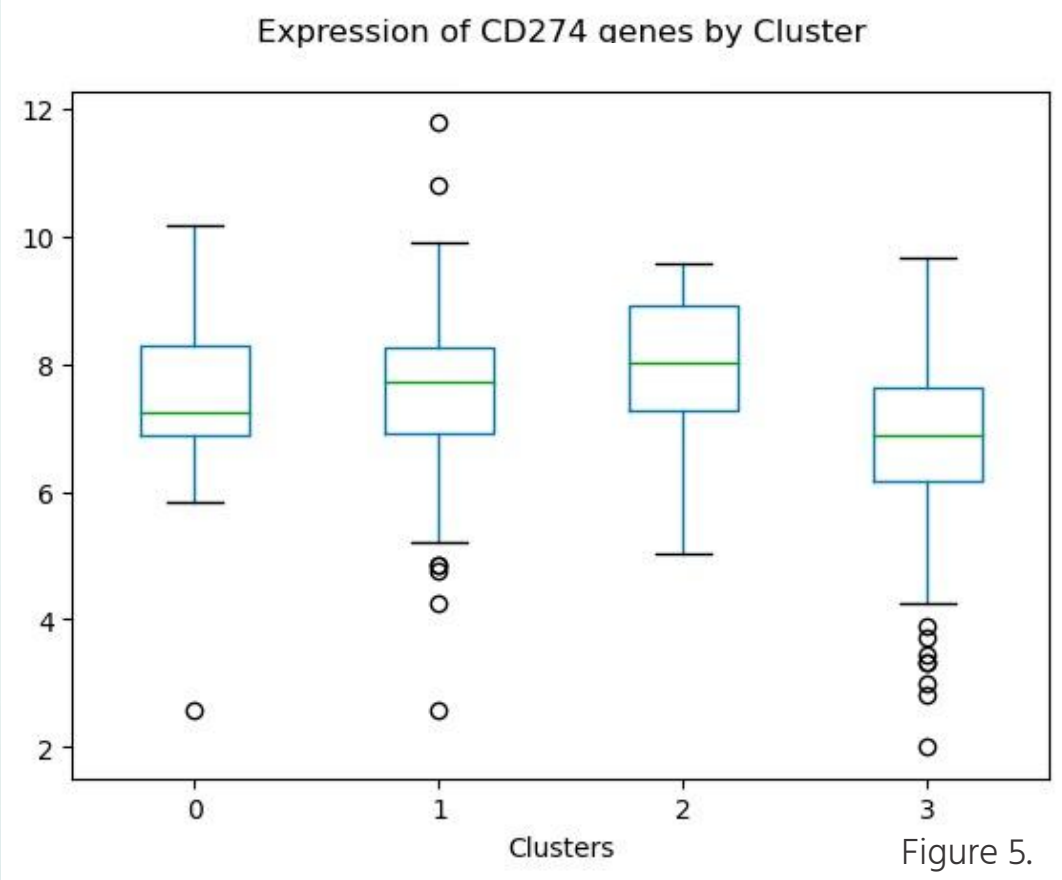


Figure 5.

Differentiated Analysis

Analysis of MUC1 and MUC16 levels:

- ★ **Cluster 0** has higher levels of MUC1
- ★ Range of **Cluster 0** is higher & smaller
- ★ High spread in **Cluster 3**
- ★ Similar levels of MUC16

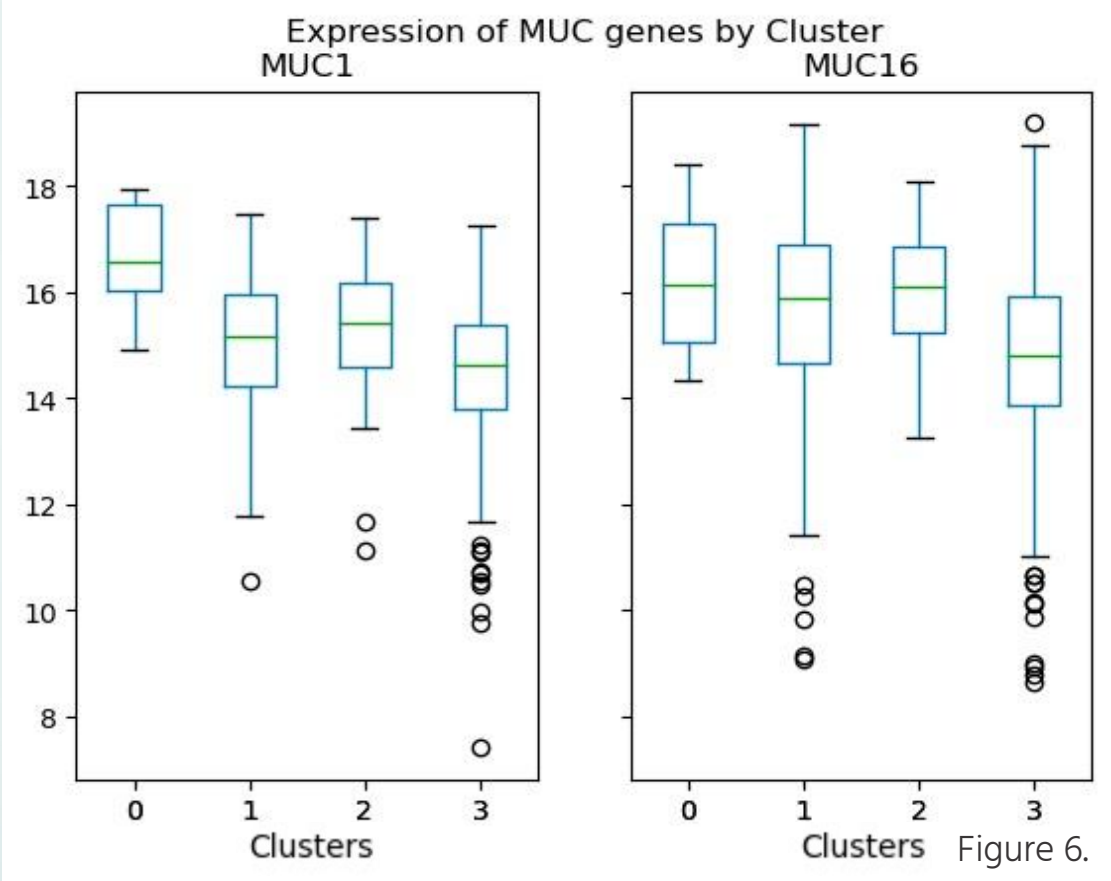


Figure 6.

Proliferative Analysis

Analysis of MKI67 and PCNA levels:

- ★ **Cluster 0** has higher levels of PCNA
- ★ Range of **Cluster 0** is smaller
- ★ High spread in **Cluster 3**

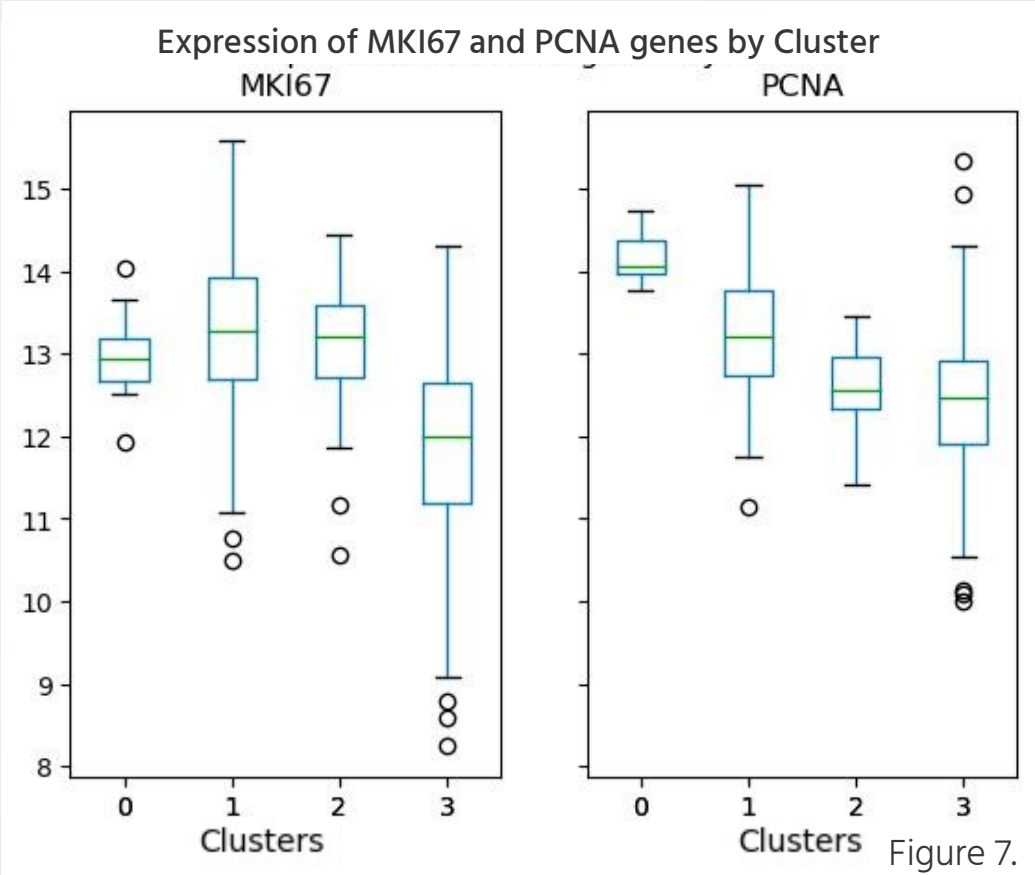


Figure 7.

Transcription Factors

PANTHER Analysis – protein class

- ★ UNIQUE genes from Top 50 most expressed in **Cluster 1**
- ★ 11/35 related to transcription
- ★ 1-2 in other clusters

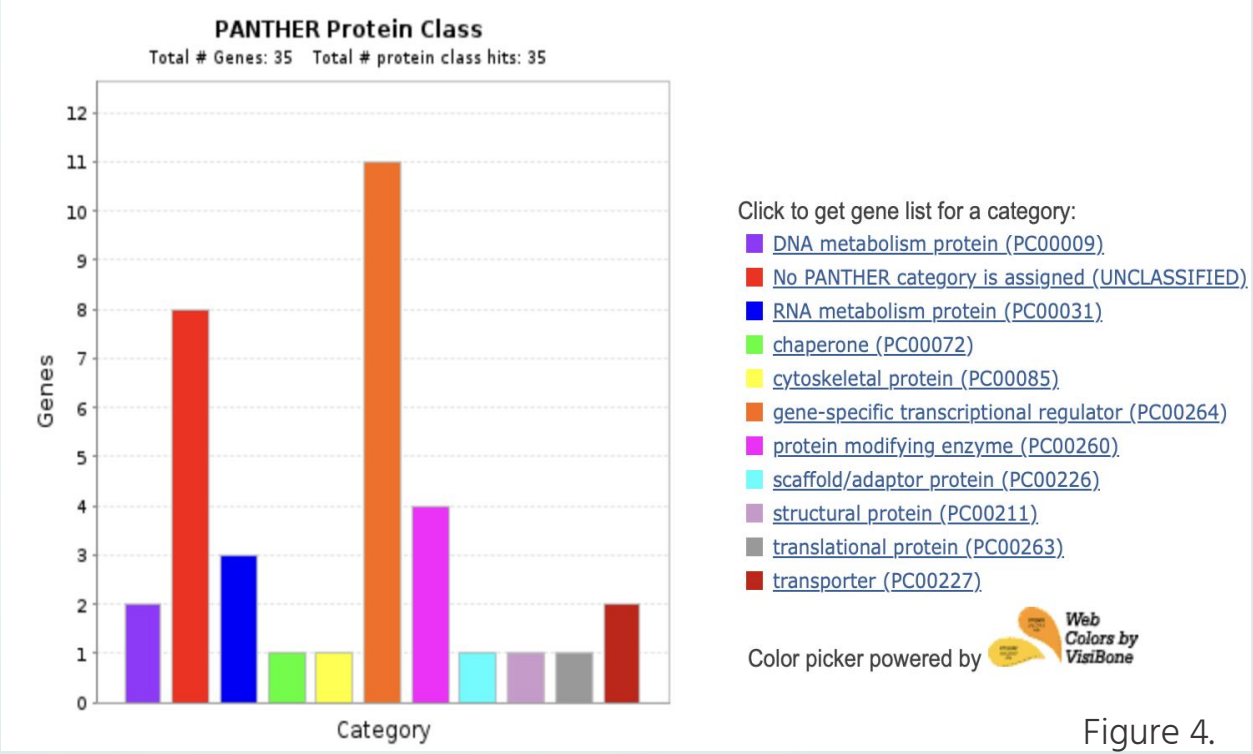


Figure 4.

Clinical Variables: Age, KM plots

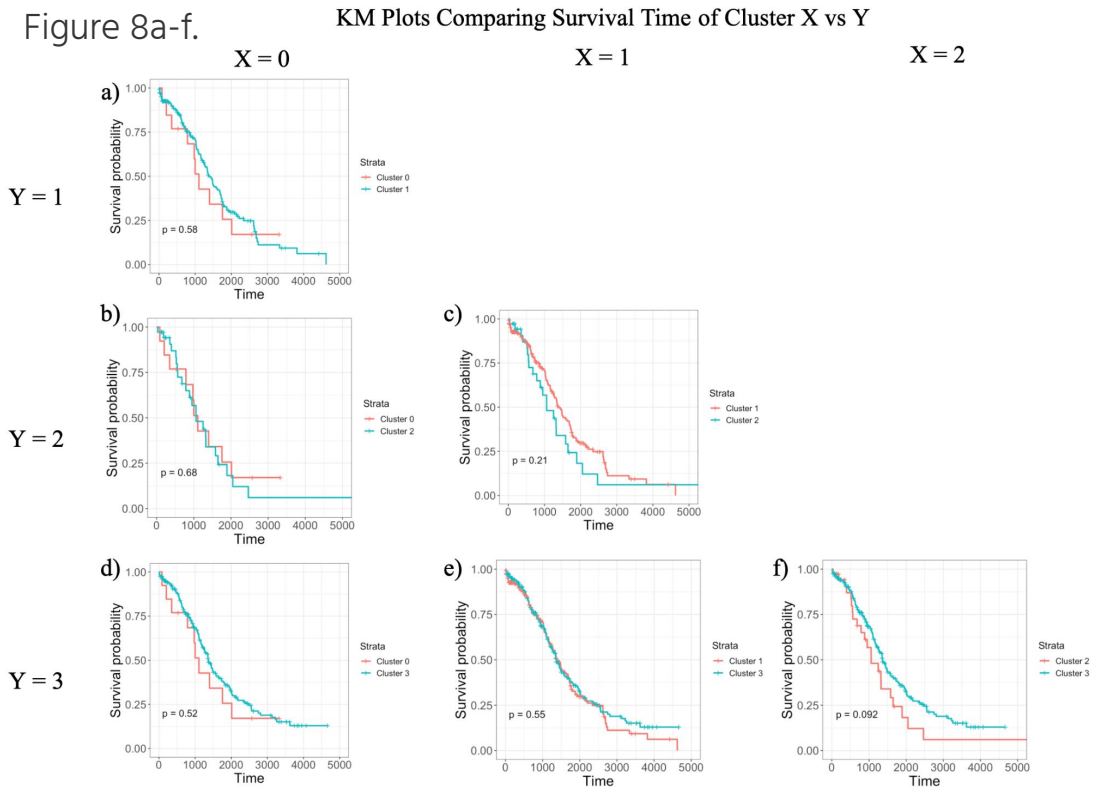
Cluster	0	1	2	3	Overall
Mean Age (yrs)	55.46	60.36	59.33	60.35	60.12

Table 1.

★ similar average age -> clustering not based on age

Clinical Variables: Age, KM plots

Figure 8a-f.



Pairwise Analysis of Survival Time for every combination of clusters

P-values between 0.5-0.6

Clinical Variables: Age, KM plots

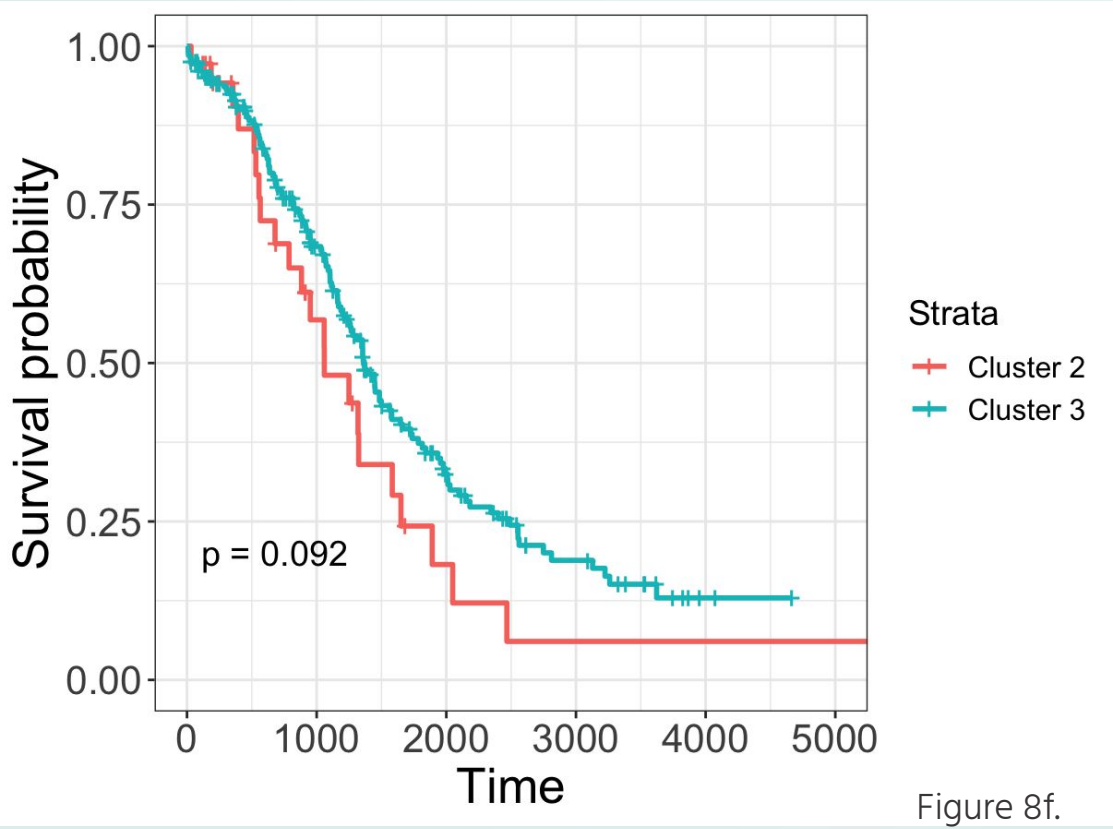


Figure 8f.

Most significant finding.

p-value = 0.092

Discussion

Cluster	0	1	2	3
Possible Subtype	• Differential • Proliferative	• Immunoreactive • Proliferative/ Differentiated	• None	• Mesenchymal

★ Clusters show traits of multiple subtypes

★ Cluster 2 showed no traits

★ Unclear, no consensus

★ Cluster accuracy?

Future Research Opportunities

- Connection to Existing Literature
- Machine Learning
  - Supervised ML could be useful for a starting point
- Subtype characterization
  - Combine histopathological analysis and omic data currently
- Explore
  - Treatment outcomes in clusters

WORKS CITED

Budden, Timothy, et al. “Development and Validation of the Gene Expression Predictor of High-Grade Serous Ovarian Carcinoma Molecular SubTYPE (PrOTYPE).” *Clinical Cancer Research*, vol. 26, no. 20, 2020, pp. 5411–23, https://doi.org/10.1158/1078-0432.CCR-20-0103.

Canadian Cancer Society / Société canadienne du cancer. “Grading Ovarian Cancer.” *Canadian Cancer Society*, cancer.ca/en/cancer-information/cancer-types/ovarian/grading#:~:text=Low%20grade%20usually%20means%20that,are%20poorly%20differentiated%20or%20undifferentiated.

Chien, J., et al. “Integrated Genomic Analyses of Ovarian Carcinoma.” *Nature (London)*, vol. 474, no. 7353, 2011, pp. 609–15, https://doi.org/10.1038/nature10166.

Kassuhn, Wanja Nikolai, et al. “Classification of Molecular Subtypes of High-Grade Serous Ovarian Cancer by MALDI-Imaging.” *Journal of Clinical Oncology*, vol. 39, no. 15_suppl, 2021, pp. e17544–e17544, https://doi.org/10.1200/JCO.2021.39.15_suppl.e17544.

Khashaba, Marwa, et al. “Subtyping of High Grade Serous Ovarian Carcinoma: Histopathological and Immunohistochemical Approach.” *Journal of Egyptian National Cancer Institute*, vol. 34, no. 1, 2022, pp. 6–6, https://doi.org/10.1186/s43046-022-00104-9.

Lisio, Michael-Antony, et al. “High-Grade Serous Ovarian Cancer: Basic Sciences, Clinical and Therapeutic Standpoints.” *International Journal of Molecular Sciences*, vol. 20, no. 4, 2019, pp. 952-, https://doi.org/10.3390/ijms20040952.

Leong, Huei San, et al. “Efficient Molecular Subtype Classification of High-Grade Serous Ovarian Cancer.” *The Journal of Pathology*, vol. 236, no. 3, 2015, pp. 272–77, https://doi.org/10.1002/path.4536.

Momenimovahed, Zohre et al. “Ovarian cancer in the world: epidemiology and risk factors.” *International journal of women's health* vol. 11 287-299. 30 Apr. 2019, doi:10.2147/IJWH.S197604.

“Ovarian Cancer by the Numbers.” *OCRA*, ocrahope.org/get-the-facts/statistics/.

Rojas, Veronica, et al. “Molecular Characterization of Epithelial Ovarian Cancer: Implications for Diagnosis and Treatment.” *International Journal of Molecular Sciences*, vol. 17, no. 12, 2016, pp. 2113–2113, https://doi.org/10.3390/ijms17122113.

Wang, Yuanshuo Alice, et al. “Multi‐omics‐based Analysis of High Grade Serous Ovarian Cancer Subtypes Reveals Distinct Molecular Processes Linked to Patient Prognosis.” *FEBS Open Bio*, vol. 13, no. 4, 2023, pp. 617–37, https://doi.org/10.1002/2211-5463.13553.