QBIO 490 Literature Presentation on Ovarian Cancer

Matthew, Alice, Lucia

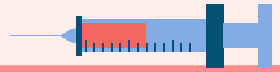
Introduction to the Papers

Review Paper: The Role of Omics Approaches to Characterize Molecular Mechanisms of Rare Ovarian Cancers: Recent Advances and Future Perspectives

- Covers genomics, transcriptomics, proteomics, and metabolomics

Research Paper: Integrated Multi-omics Analysis of Ovarian Cancer Using Variational Autoencoders

- This study discusses **integrated** multi-omics data
  - Covers genomics, transcriptomics, and epigenomics



Review Paper

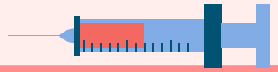
Goals:

- Look at the state of individual and multi -omics approaches and their use in studying ovarian cancer

- Gain a better overall understanding of cancer research; look into therapies, drug resistance, cancer recurrence, prognosis, etc. that have been discovered through -omic approaches

Methods:

- Survey of 104 journal-published papers about ovarian cancer -omic data



Review Paper

Results:

- Genomics
  - identify new subtypes of rare ovarian cancers such as clear cell carcinoma, mucinous carcinoma, LGSOC
- Transcriptomics
  - insights on biomarkers, temporal gene upregulations, gene fusions, and potential druggable sites
  - scRNA-sc is emerging as a way to look at the transcriptome of single cells
- Proteomics
  - still a young field of study, but has been used to find post-translational protein modifications
- Metabolomics
  - found potential biomarkers in ovarian cancers but not many studies have been published

Review Paper

More Results:

- Since cancer is multi-faceted and affects many layers of the central dogma, multi-omic approaches using machine learning and AI are starting to be built and deployed
  - Can better find linear and non-linear processes
  - Better identify biomarkers, therapeutic targets, and causes of cancer
- Cancer recurrence and drug resistance pathways have been studied and some causes of both have been found

Takeaways:

- Individual omics and multi-omics approaches have greatly increased knowledge of both rare and common ovarian cancers
- Research of rare ovarian cancers has especially blossomed with greater availability of biological data and we have found much more information on them than previously known
- Novel biomarkers, therapies, and personalized care has been achieved through omics data

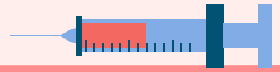
Research Paper

Hypothesis:

- DL based integrated multi-omics analysis difficult, as data is imbalanced with too many molecular features and relatively few patient samples
- DL-based dimensionality reduction technique, including variational autoencoder (VAE), is a potential solution to balance high dimensional multi-omics data.

Method:

- Monomics, diomics, tri-omics datasets from TCGA (i.e., mRNA, CNV/CNA, DNA methylation, RNAseq, and miRNA)
- **Preprocessed**: intersected the mono-omics datasets to find the common and same size samples; identified and removed the missing/zero/NA values; normalise datasets using the min-max technique so that they have **equal importance** in analysis; concatenated to form the di- and tri-omics datasets



Research Paper

Method:

- Construct Standard VAE and MMD-VAE (an improved version of VAE, namely Maximum Mean Discrepancy) respectively
- Implement them using the same architecture
- **Clustering and classification in cancer:** 4 ovarian cancer transcriptional subtypes; VAE or MMD-VAE generated latent and compressed features (z or LFs) can be used to cluster and classify cancer samples, subtypes, including existing transcriptional or molecular subtypes of ovarian cancer
- **Survival analysis:** identification of robust survival subgroups; existing subtypes may not be useful, so we need to find the prognostic biomarkers to predict new subgroups

Research Paper

Main Findings:

- VAE and MMD-VAE outperform existing dimensionality reduction techniques
- Integrated di- and tri-omics based LFs can perform better or similar to their mono-omics counterparts based LFs
- Molecular subtypes clustering and classification results show that MMD-VAE is outperforming VAE in most datasets

Limitations:

- Multi-omics based LFs and subgroups predicted based on them do not improve the survival prediction performances than their mono-omics counterparts
- Straightforward integration (concatenation based) strategy, treating omics measurements from different platforms equally and performing integration in a **parallel fashion** may not always be useful

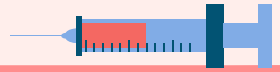
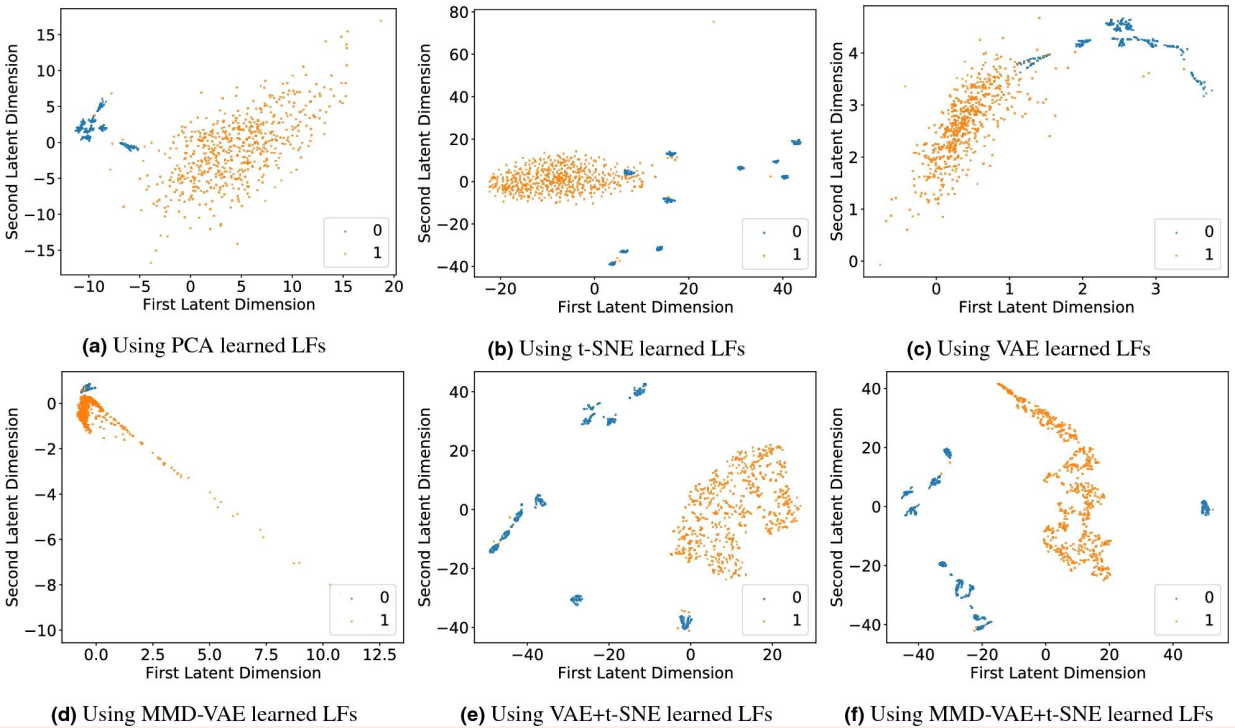


Figure 1



(a) Using PCA learned LFs

(b) Using t-SNE learned LFs

(c) Using VAE learned LFs

(d) Using MMD-VAE learned LFs

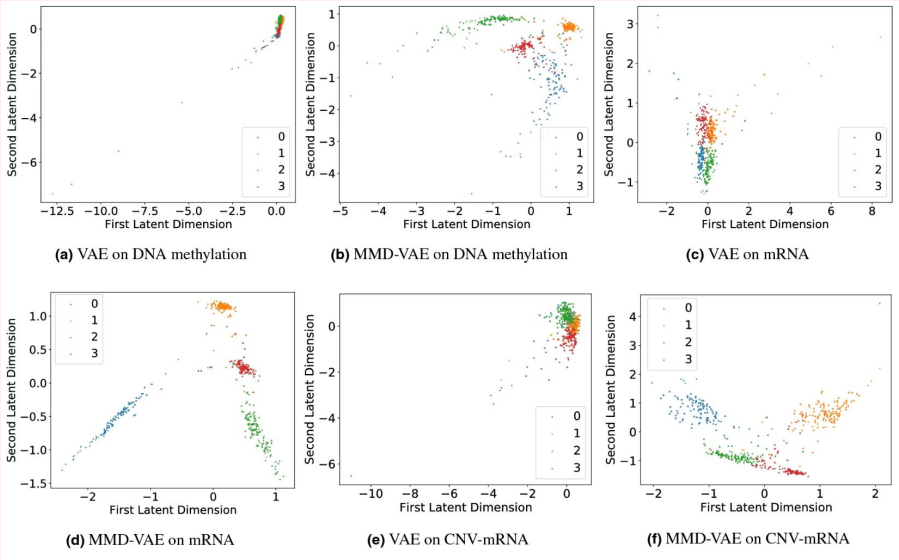
(e) Using VAE+t-SNE learned LFs

(f) Using MMD-VAE+t-SNE learned LFs

- Used the GDC TCGA Ovarian Cancer cohort (DNA methylation)
- Demonstrates clustering between normal and cancer data when using unsupervised VAE, unsupervised MMD-VAE, PCA, and t-SNE to learn the labeling functions
- Chose clustering as visual way to demonstrate separation between cancer/normal data

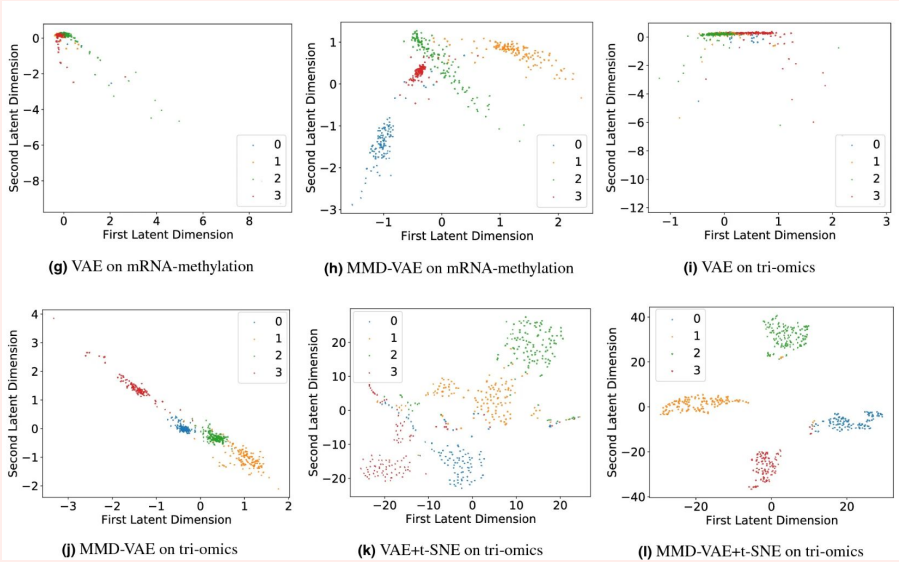
Figure 2

0—Immunoreactive, 1—Differentiated, 2—Proliferative and 3—Mesenchymal.



(a) VAE on DNA methylation (b) MMD-VAE on DNA methylation (c) VAE on mRNA

(d) MMD-VAE on mRNA (e) VAE on CNV-mRNA (f) MMD-VAE on CNV-mRNA



(g) VAE on mRNA-methylation (h) MMD-VAE on mRNA-methylation (i) VAE on tri-omics

(j) MMD-VAE on tri-omics (k) VAE+t-SNE on tri-omics (l) MMD-VAE+t-SNE on tri-omics

- Used TCGA Ovarian Cancer cohort mRNA, CNV/CNA + RNAseq, DNA methylation, and combinations of the above (di-omics and tri-omics)
- Demonstrates clustering between four molecular subtypes with supervised VAE, supervised MMD-VAE, t-SNE, and combinations of the above
- Chose clustering as visual way to demonstrate separation between molecular subtypes

Questions/Future Investigations

- How can we better improve AI so that it is more accurate in predicting cancers
- Look into making transcriptomic, proteomic, and metabolomic more accessible (easier/faster) so that we can get better data
- Set international standards on data collections so that we can get a more standard approach
- How can we weigh different types of omics data efficiently