PURDUE UNIVERSITY | The Data Mine

TRANSLATIONAL BIOMEDICAL DATA SCIENCE FOR NEURODEGENERATIVE DISEASES

Translational HDPM

Wefae Ali, Keira Bosland, Katie Connolly, Deeya Prabhu, Bella Schaetzle

INTRODUCTION

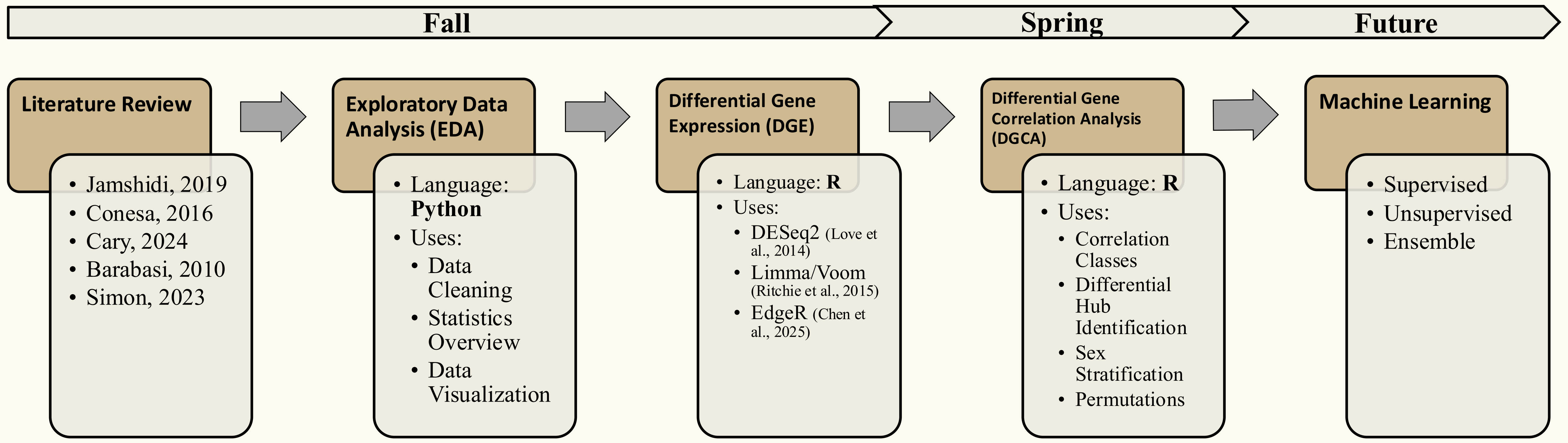
BACKGROUND

- Translational HDPM is an early-stage biotechnology company focused on novel drug development, through in-depth understanding of topology of gene modules, known as network medicine.
- Pharmacodynamic biomarkers can contain cell-free RNA; which serve as a communication system of the body.
  - Analysis of such provides a noninvasive approach to investigate the progression of biological process disruptions, in addition to identifying both potential patients and medicines.

MISSION

- The purpose of this biomedical data science project is to develop transparent 'clinical grade' high dimensional cell free RNA (liquid biopsy) co-expression networks.
  - These networks would provide insight into the rewired network of underlying pathological mechanisms of neurodegenerative diseases (NDD), specifically Alzheimer's Disease (AD).

METHODOS AND TIMELINE



Fall | **Spring** | **Future**

Literature Review
- Jamshidi, 2019
- Conesa, 2016
- Cary, 2024
- Barabasi, 2010
- Simon, 2023

Exploratory Data Analysis (EDA)
- Language: **Python**
- Uses:
  - Data Cleaning
  - Statistics Overview
  - Data Visualization

Differential Gene Expression (DGE)
- Language: **R**
- Uses:
  - DESeq2 (Love et al., 2014)
  - Limma/Voom (Ritchie et al., 2015)
  - EdgeR (Chen et al., 2025)

Differential Gene Correlation Analysis (DGCA)
- Language: **R**
- Uses:
  - Correlation Classes
  - Differential Hub Identification
  - Sex Stratification
  - Permutations

Machine Learning
- Supervised
- Unsupervised
- Ensemble

Figure 1. Schematic of Fall and Spring Workflow

Our team used a multi-stage analytical framework to process and interpret our high-dimensional biological data. The process began with a literary review to establish foundational knowledge, followed by Exploratory Data Analysis (EDA) to ensure data quality and visualize key trends. We then performed Differential Gene Expression (DGE) and Differential Gene Correlation Analysis (DGCA) using R-based statistical tools, including DESeq2, Limma/Voom, and EdgeR, to identify significant changes in gene frequency and hub gene interactions, as well as how that differs between men and women.

Extended Methods:

RESULTS

DIFFERENTIAL GENE EXPRESSION (DGE)

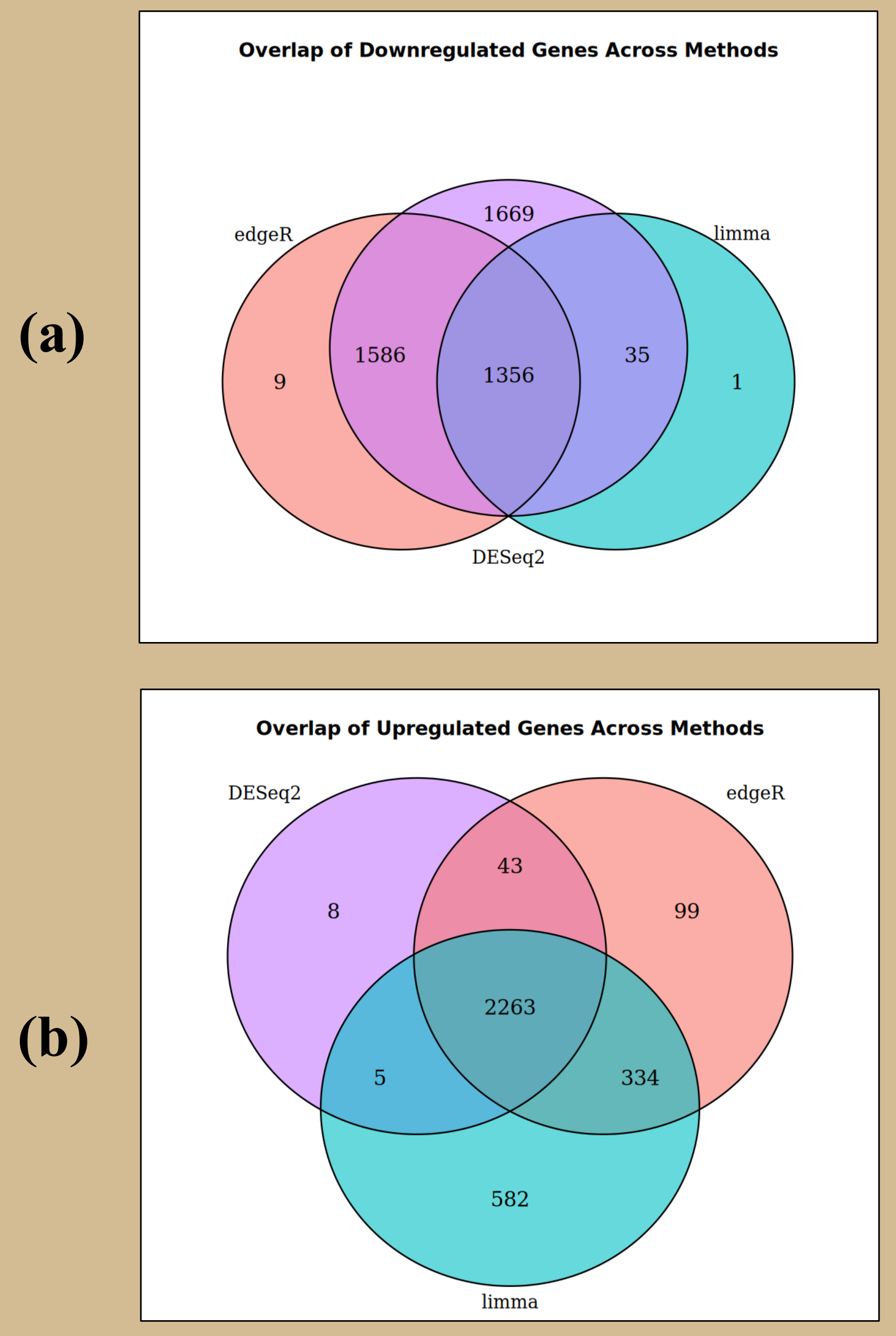
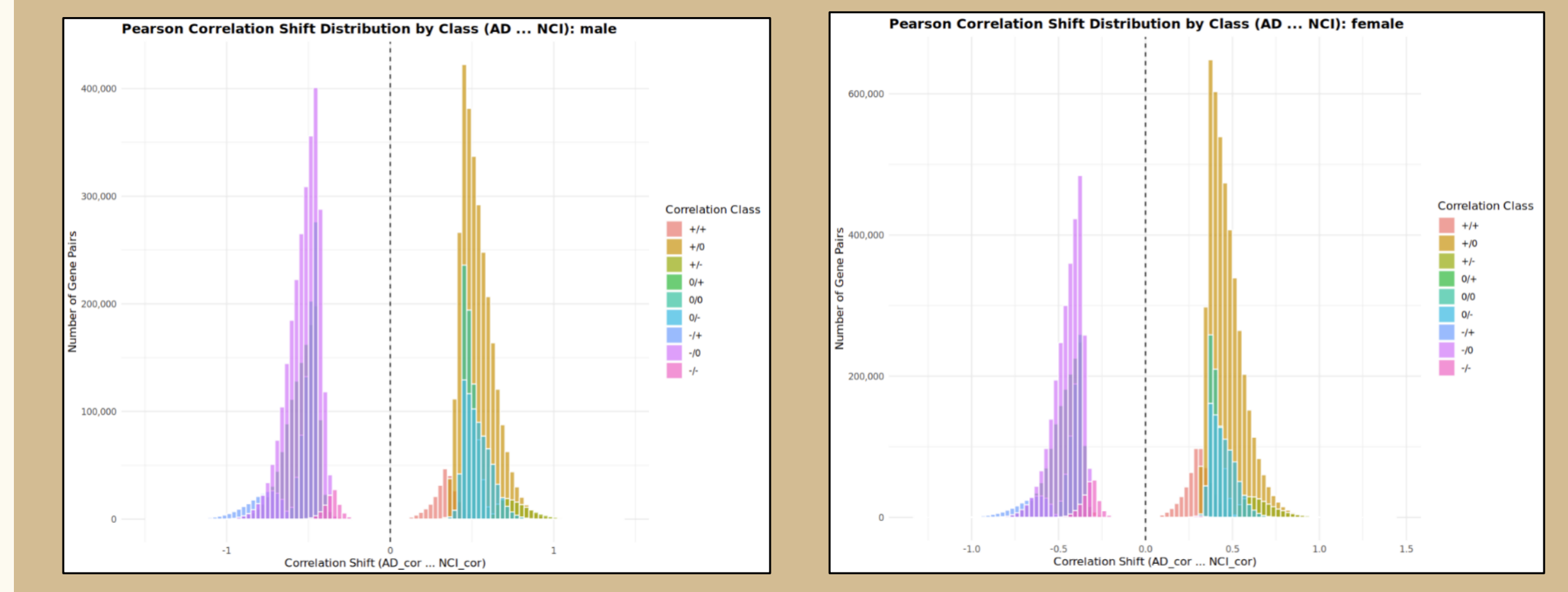


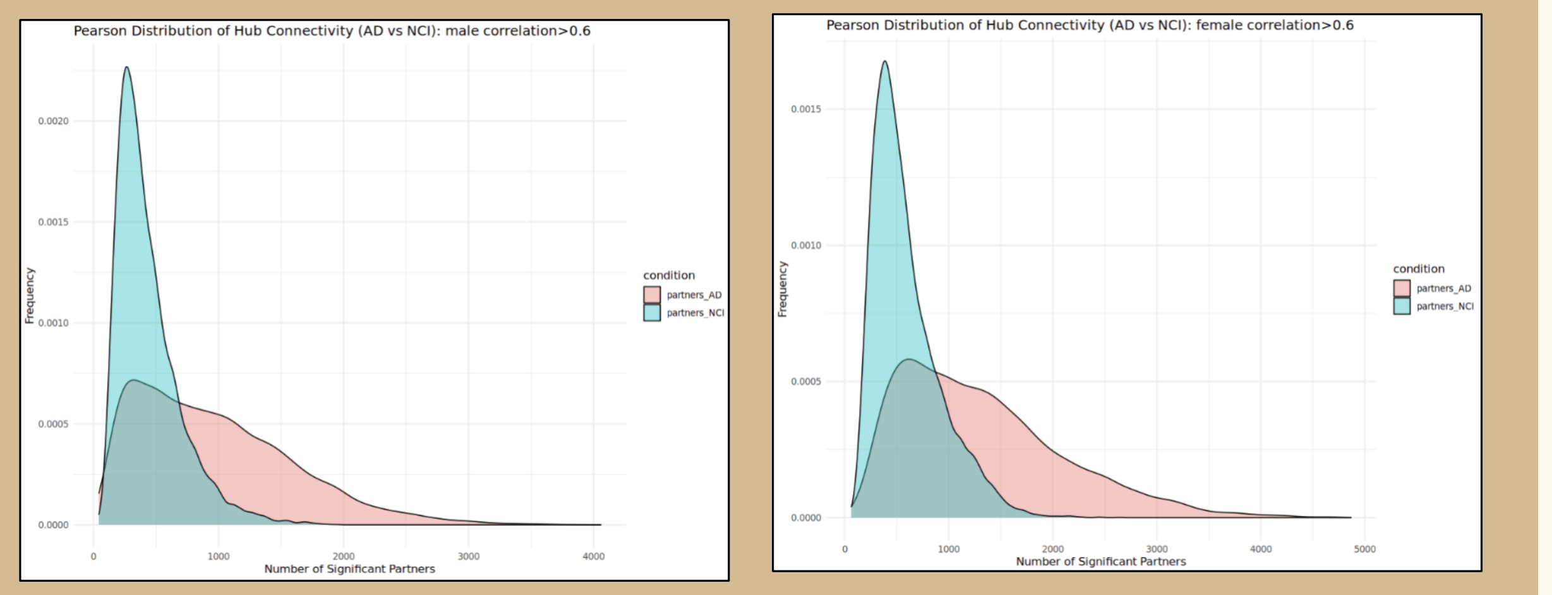
Figure 2. Venn diagram depicting count of overlapping genes between methods
(a) downregulated genes (b) upregulated genes

DIFFERENTIAL GENE CORRELATION ANALYSIS (DGCA)



(a) (b)

Figure 3. Histograms depicting bimodal correlation of gene classes stratified by sex.
(a) Males (b) Females



(a) (b)

Figure 4. Density plots of difference in number of significant gene partners between AD and NCI
(a) Males (b) Females

Table 1. Class-stratified gene overlap between male and female AD co-expression networks after filtering with permutation data.

Class	Male Genes	Female Genes	Shared	Male Only	Female Only	% Shared of Male	% Shared of Female
+/0	3601	3599	1943	1658	1656	54	54
-/0	2964	1793	804	2160	989	27.1	44.8
+/+	709	2128	585	124	1543	82.5	27.5
0/+	630	435	93	537	342	14.8	21.4
+/-	329	188	93	236	95	28.3	49.5
0/-	223	144	10	213	134	4.5	6.9
-/+	457	30	5	452	25	1.1	16.7
-/-	17	229	4	13	225	23.5	1.7

Class represented as AD/NCI, (+) coupling of genes, (-) decoupling of genes, (0) no significant co-expression

DISCUSSION

DIFFERENTIAL GENE EXPRESSION

- Preliminary data analysis, concluded a negative logarithmic trend between gene expression and variability highlighting genes low in expression having higher variability

DIFFERENTIAL GENE CORRELATION ANALYSIS

- Permutation analysis increases the robustness of gene-gene correlation data for biological interpretation
- AD expression varies significantly between sexes
  - In NCI, the sexes share fewer co-expression genes than the AD state (14% vs 43%, respectively)
  - Whilst the sexes share 2975 genes in AD, 32% are male-specific and 25% female-specific. Non-sex specific genes showing greater connectivity.
  - Between sexes, 54% rewiring differently in AD. The hallmark of each sex in AD is males undergo decoupling and females see reinforcement of co-expression with sparce decoupling.
  - The hub genes are autosomal – the top male AD hub,TWF2, has no female AD connectivity and the top AD female hub has 219 connections vs only 21 in males

CHALLENGES

- Historically NDD such as AD have been limited with a reproducibility of experiments and results.
  - Due in part to the high heterogeneity of AD causing larger feature overlap along with high fail rate and cost of research.

FUTURE DIRECTIONS

Translational HDPM will move forward with:
- Developing AI and Machine Learning models to bridge the gap between high dimensional biomedical data and the underlying biological knowledge of disease
- To develop pharmacotherapies and fill unmet clinical needs for other NDDs.

REFERENCES AND ACKNOWLEDGEMENTS

We would like to thank Dr. John Sninsky and Dr. Samantha Khoury from TranslationalHDPM for their leadership, as well as Traven Cargill, Nicholas Lauw, Ahad Al-Fawwaz, Logan Portscheller, Vikram Kavalipati, Rachel Miskelly, Bhargavi Yallamraju, Venkat Mamidi, and Venkata Chelliboyina for their work on this project. We also thank the Data Mine for providing facilities and resources.

References: