

DEList: Identifying Cancer Genes Associated With Circadian Regulator ARNTL2 Using a Novel RNA-Seq Multi-Dataset Strategy

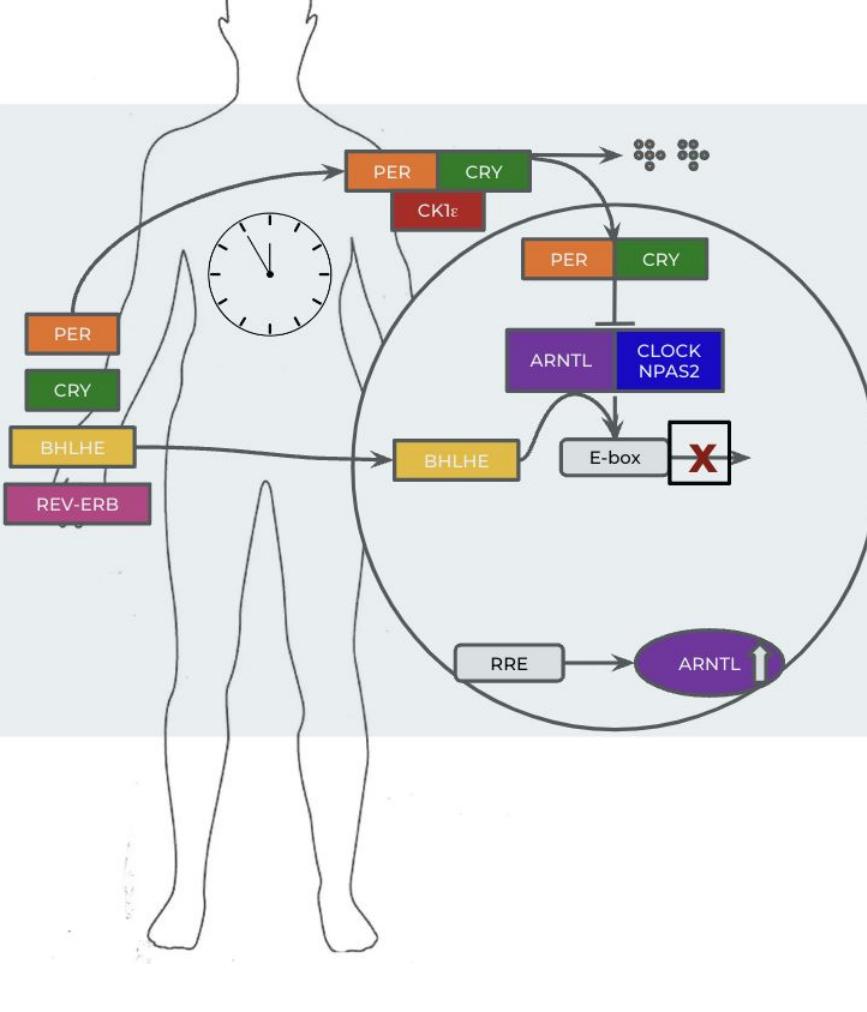
Ella Yee | BMED064

*Unless otherwise noted, all figures, images, diagrams, and illustrations were created by student

Introduction

Problem Statement

- Circadian disruptions are associated with pathologies such as cancer¹
 - In 2007, the IARC categorized "shift work involving circadian disruption" as a carcinogen¹
- Circadian disruption is a growing problem: 16% of U.S. adults worked non-daytime schedules, 70% worked indoors, 99% impacted by light pollution, growing sector aged 65+, as of 2021²
- In particular, upregulation of circadian regulator ARNTL2 (BMAL2) is associated with unfavorable prognosis in multiple cancer types³⁻⁵



Research Significance

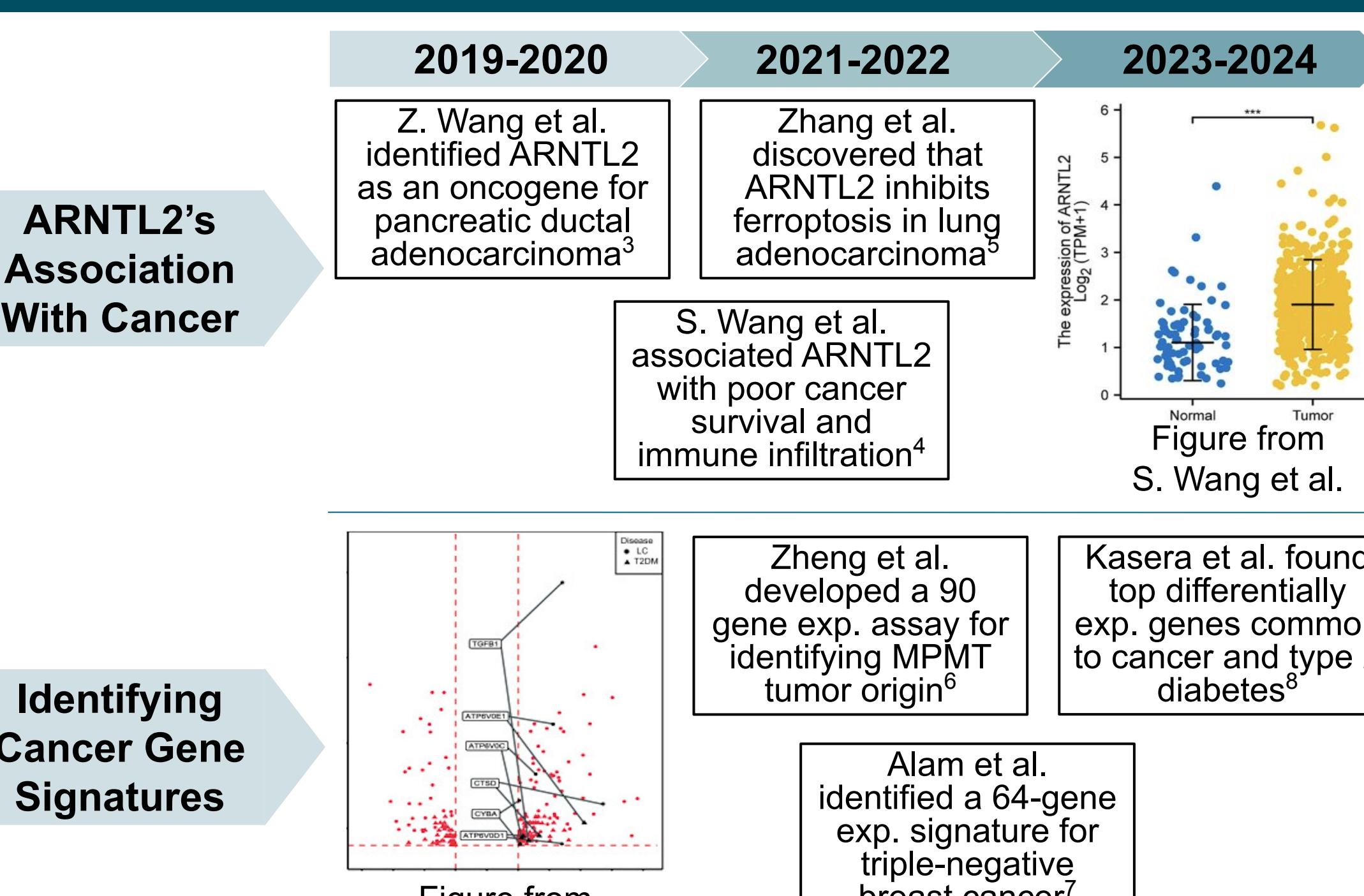
- The underlying mechanisms linking circadian rhythms and cancer remain uncertain
 - Protein links: gene expression and regulation, protein-protein interactions, secretion of endocrine factors
- This uncertainty poses significant obstacles to advancing circadian-based strategies for cancer diagnosis and treatment

Hypothesis & Research Objectives

Hypothesis: Circadian gene ARNTL2 plays a role in oncogenesis by regulating the expression of cancer-associated genes.

- Develop a strategy to investigate expression of core circadian genes, including circadian regulator ARNTL2, in tumor and normal samples
- Design a clinical/demographic subgroup study for in-depth analysis of ARNTL2 expression in tumor and normal samples
- Explore the impact of ARNTL2 blockade on gene expression in metastasis cell lines and identify ARNTL2-associated genes
- Devise data analyses to further validate ARNTL2-associated genes, including correlational, gene ontology, and RNAscope analyses

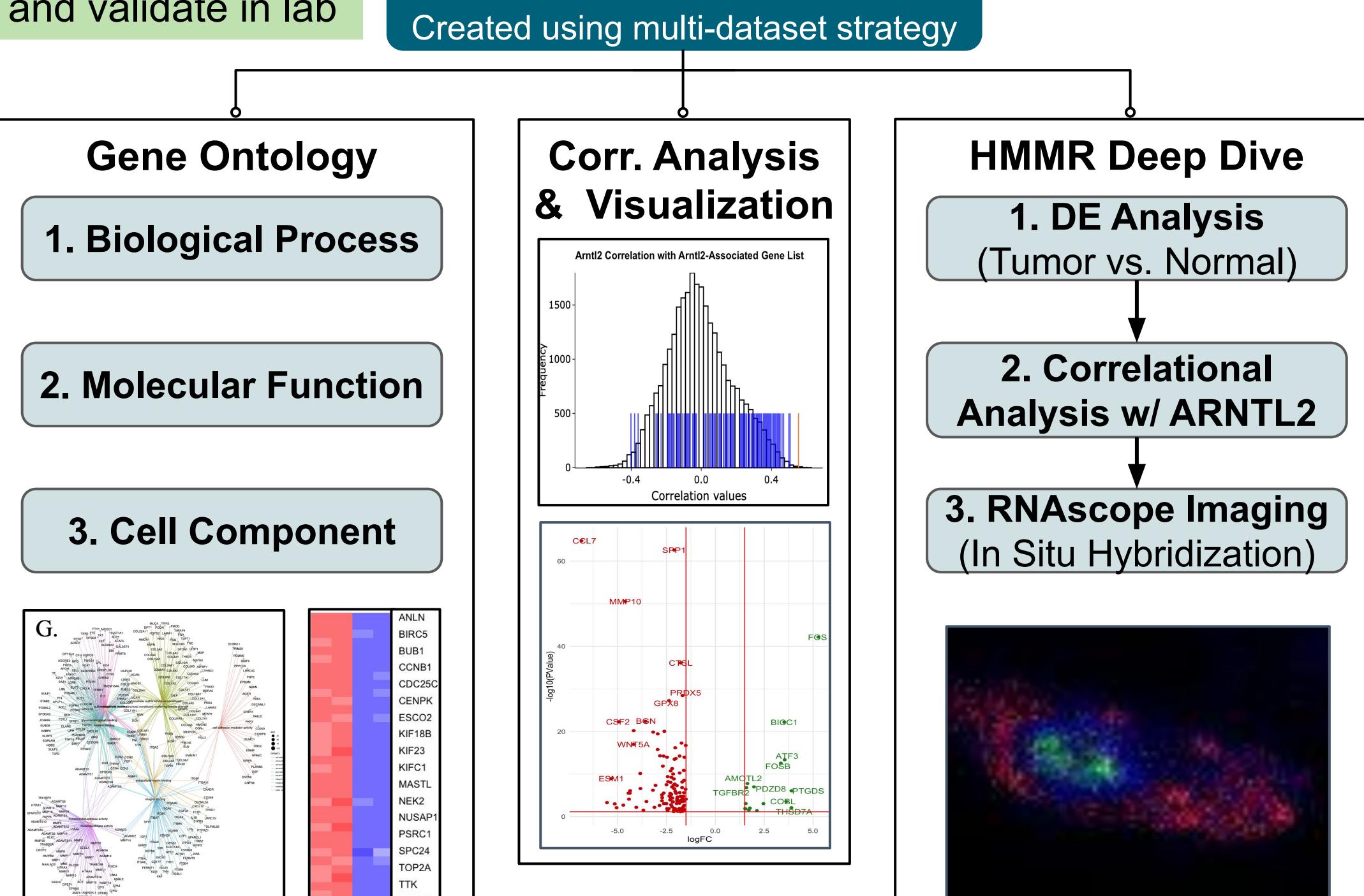
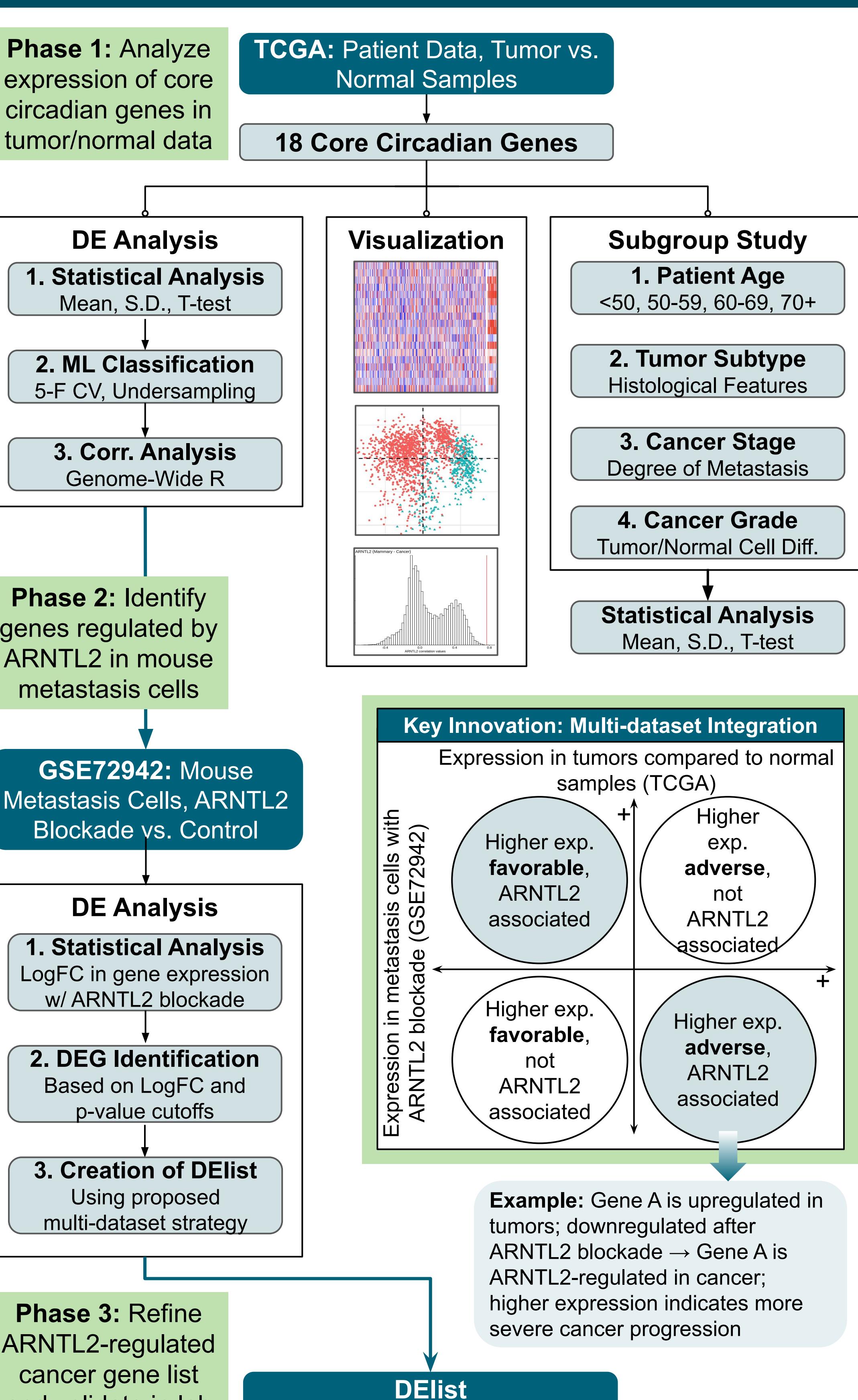
Summary of Recent Research



Research Gaps & My Research Proposal

- Previous studies have identified an association between ARNTL2 expression and cancer, but this association has not been used to identify circadian-based gene lists for predicting cancer progress
- I propose a multi-dataset integration strategy to identify ARNTL2-regulated genes with disrupted expression in cancer

Proposed Multi-Dataset Strategy



5 Key Results: Discovering 80 Genes Related to Circadian Disruption and Cancer

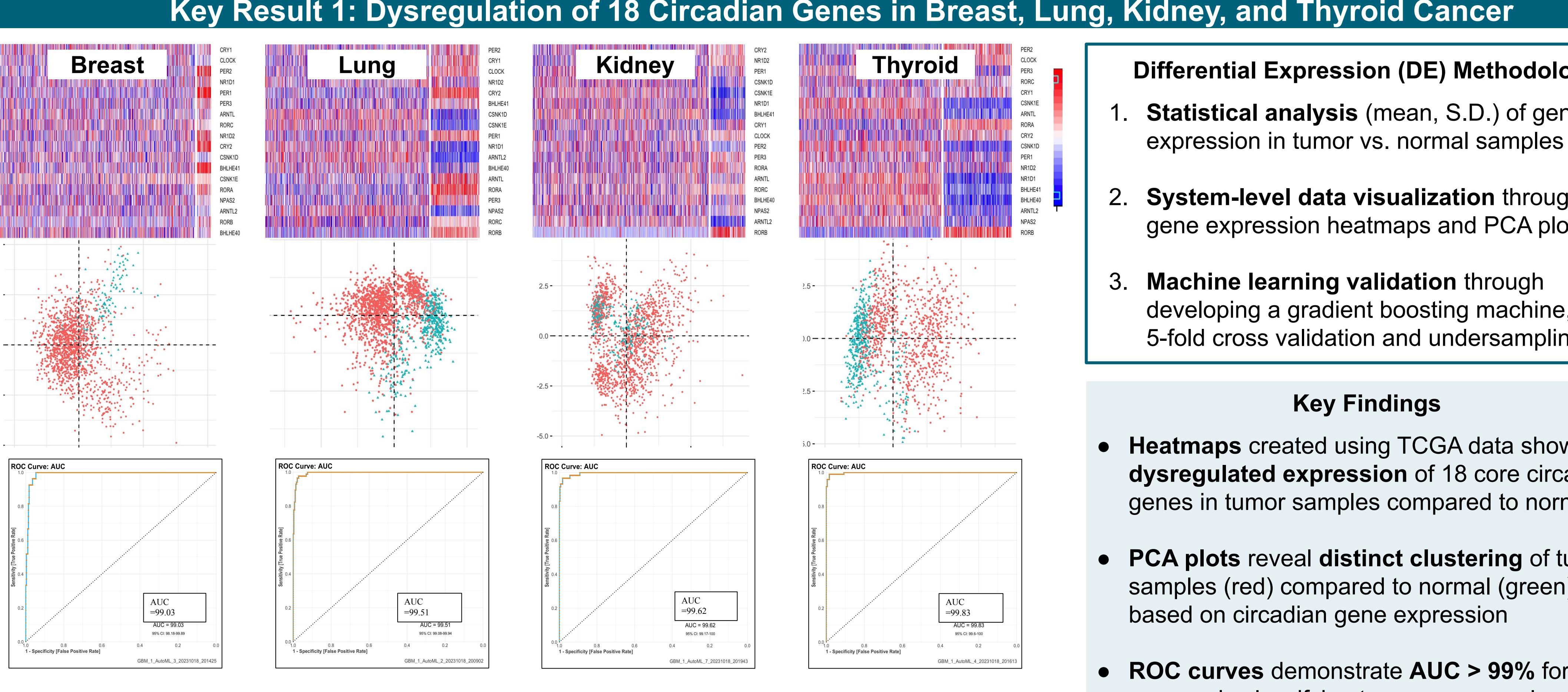


Figure 1. Heatmaps, principal component analysis plots, and ROC curves of circadian gene expression in primary tumor vs. normal samples from The Cancer Genome Atlas (TCGA).

Differential Expression (DE) Methodology

- Statistical analysis (mean, S.D.) of gene expression in tumor vs. normal samples
- System-level data visualization through gene expression heatmaps and PCA plots
- Machine learning validation through developing a gradient boosting machine, with 5-fold cross validation and undersampling

Key Findings

- Heatmaps created using TCGA data show dysregulated expression of 18 core circadian genes in tumor samples compared to normal
- PCA plots reveal distinct clustering of tumor samples (red) compared to normal (green), based on circadian gene expression
- ROC curves demonstrate AUC > 99% for all 4 cancers in classifying tumor vs. normal samples based on expression of 18 core circadian genes

Conclusions

4 Major Innovations

- Designed and implemented a systematic data mining pipeline combining DE, visualization, correlational, and ML analyses
- Explored ARNTL2 expression in clinical/demographic subgroups, which have not been thoroughly analyzed in previous reports
- Proposed novel multi-dataset integration of TCGA (multi-subject, tumor/normal) and GSE72942 (ARNTL2 blockade in metastasis cells)
- Developed RNAscope imaging procedure to analyze ARNTL2-HMMR RNA colocalization in lung squamous cell carcinoma tissue

Key Findings

- The dysregulation of 18 core circadian genes is significantly correlated with breast, lung, thyroid, and kidney cancers
- ARNTL2 was the only core circadian gene consistently upregulated in tumors compared to normal samples, regardless of cancer type, age group, cancer stage, or cancer grade
- Multidataset strategy allowed for the discovery of DEList (80 circadian-regulated cancer prognostic genes)

- RNA scope images showed colocalization of ARNTL2 and HMMR (ranked 12 on DEList) for the first time
 - HMMR is a cell motility gene, and previous reports have associated HMMR with cancer

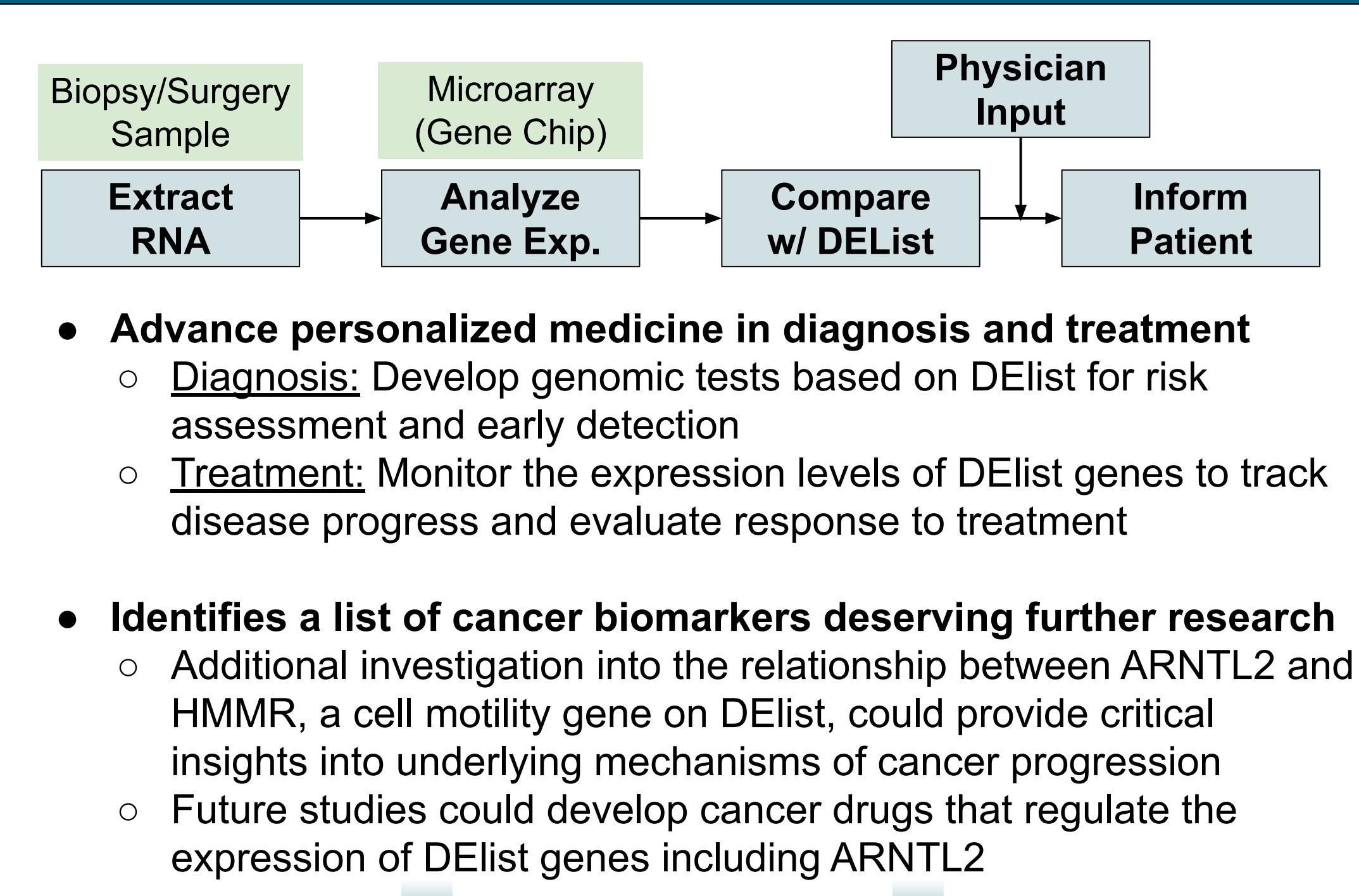
Significance of This Research

- Functionality analysis of DEList genes revealed potential mechanisms linking circadian disruption and cancer progression
 - Top-ranked DEList genes encode Ser/Thr kinases, microtubule activity proteins, and G2/M transition phase proteins
- From PDB: Ser/Thr kinase
 - From Front. Oncol.: Microtubule activity
 - From BD Biosci.: G2/M transition
- DEList opens the door to the development of circadian-based gene expression profiling for cancer patients
 - First-time identification of a circadian-regulated cancer prognosis gene list → critical applications as circadian disruptions increase due to modern lifestyle changes
 - Gene expression profiling provides increased stratification of cancer types, compared to genetic testing → especially relevant to patients w/o targetable alterations, or resistance to treatment⁶

Applications & Future Work

- 1. Developed Multi-Dataset Integration Strategy**
 - Proposed an integrated RNA-seq data analysis pipeline for the discovery of prognostic gene lists
 - Applicable to other medical conditions, such as cardiovascular or neurodegenerative disease
 - Step 1: Screen patient datasets to identify target genes with significant differential expression (ex. ARNTL2)
 - Step 2: Analyze experimental gene knockout datasets to study target gene's impact in disease model
 - Step 3: Identify genes that are both (1) regulated by the expression of the target gene and (2) differentially expressed in disease model, based on patient dataset

2. Created DEList, a Cancer Prognostic Gene List



3. Uncovered Strong Circadian-Cancer Relationship

- This research indicates a strong correlation between the circadian and cancer systems as a whole
- Underscores the need for further interdisciplinary research on circadian health and disease incidence: epidemiology, evolutionary biology, public health, and other fields

Future Work

- Computational:** Evaluate DEList through additional public cancer patient datasets
- Experimental:** Verify DEList genes at the
 - RNA level, using qPCR
 - Protein level, using ELISA & western blotting
- Experimental:** Homogenize tumor/healthy tissue sample, conduct multiplex protein screen such as Luminex

Acknowledgements & Selected References

Thank you to Mr. Chris Spenner and Dr. Qian Wang for their guidance in this project.

- Lee, V. (2021). Roles of circadian clocks in cancer pathogenesis and treatment. *Experimental & Molecular Medicine*, 53(10), 1529-1534. <https://doi.org/10.1038/s12276-021-00681-0>
- Fishbein, A. B., Knutson, K. L., & Zee, P. C. (2021, October 1). Circadian disruption and human health. *The Journal of Clinical Investigation*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8483747/>
- Wang, Z., Xie, J., Xie, Z., & Li, J. (2020). HMMR is an indicator of poor prognosis promotes epithelial-to-mesenchymal transition and inhibits ferroptosis in clear cell renal cell carcinoma. *Cancer Cell* 21, 341 (2021). <https://doi.org/10.1101/2295-0283-2020-02839-6>
- Wang, S., Ma, X., Ying, Y., et al. (2022). Gene expression profiling of multiple primary malignant tumors. *Cancer Cell* 31, 21-47 (2021). <https://doi.org/10.1166/cancer.2022.104974>
- Alam, S., Sultan, A., Wang, G., & Mollah, N. (2022). Gene expression profile analysis to discover molecular signatures for early diagnosis and therapies of triple-negative breast cancer. *Front. Mol. Biosci.* 9. <https://doi.org/10.3389/fmolb.2022.960562>
- Hijazo-Pechero, S., Alay, A., Martin, R., Vilaino, A., Santamaría, D., Nadal, E., & Solé, X. (2021). Gene Expression Profiling as a Potential Tool for Precision Oncology in Non-Small Cell Lung Cancer. *Cancers*, 13(19), 4734. <https://doi.org/10.3390/cancers13194734>