HS-007

Discovering Molecular Endotypes of Bladder Cancer and Urinary Pathology Using Urine Proteomic Network Analysis

Shri Chada
Cypress Woods High School, Cypress, Texas

Abstract

Background: Distinguishing bladder cancer (BC) from non-malignant causes of hematuria remains challenging due to overlapping biological processes such as inflammation and tissue remodeling. High-dimensional urinary proteomics may enable identification of underlying molecular states beyond traditional diagnostic categories.

Methods: Urinary proteomic data (7596 proteins) from BC, urinary controls (UC), and healthy controls (HC) were analyzed. After preprocessing, a protein–protein correlation network (r > 0.95) was constructed and protein modules were identified using the Walktrap algorithm. Modules (>20 proteins) were retained, and module activity scores were computed per patient. Patients were clustered using hierarchical clustering, and module-derived scores were evaluated for classification performance between BC and UC.

Results: Eight protein modules representing coordinated biological processes were identified. Clustering revealed a BC-enriched group (G4) with 80% positive predictive value in this cohort. Modules associated with immune activation (Module 6), adaptive remodeling (Module 3), and proliferative signaling (Module 0) were differentially active across patient groups. A module-based score (Module 0 – [Module 6 + Module 3]) improved in-sample discrimination between BC and UC (AUC = 0.84; sensitivity = 93.3%; specificity = 70%). Overlap between BC and UC reflects shared biological processes rather than distinct protein markers.

Conclusions: Urinary proteomic analysis identifies biologically distinct states defined by coordinated pathway activity and supports preliminary classification of BC and UC. These findings highlight the potential of module-based approaches to improve diagnostic specificity, pending validation in larger cohorts.

Study Objective

To determine whether high-dimensional urinary proteomics can identify biologically distinct patient states and support classification of bladder cancer versus non-malignant conditions.

Select References

1. Chaussabel, D., Quinn, C., Shen, J., Patel, P., Glaser, C., Baldwin, N., Stichweh, D., Blankenship, D., Li, L., Munagala, I., Bennett, L., Allantaz, F., Mejias, A., Ardura, M., Kaizer, E., Monnet, L., Allman, W., Randall, H., Johnson, D., Lanier, A., ... Pascual, V. (2008). A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. Immunity, 29(1), 150–164. https://doi.org/10.1016/j.immuni.2008.05.012

2. Schiess, R., Wollscheid, B., & Aebersold, R. (2009). Targeted proteomic strategy for clinical biomarker discovery. Molecular oncology, 3(1), 33–44. https://doi.org/10.1016/j.molonc.2008.12.001

3. George, B. J., Reif, D. M., Gallagher, J. E., Williams-DeVane, C. R., Heidenfelder, B. L., Hudgens, E. E., Jones, W., Neas, L., Hubal, E. A., & Edwards, S. W. (2015). Data-driven asthma endotypes defined from blood biomarker and gene expression data. PloS one, 10(2), e0117445. https://doi.org/10.1371/journal.pone.0117445

4. Saelens, W., Cannoodt, R., & Saeys, Y. (2018). A comprehensive evaluation of module detection methods for gene expression data. Nature communications, 9(1), 1090. https://doi.org/10.1038/s41467-018-03424-4

5. Suzuki, M., Cole, J. J., Konno, S., Makita, H., Kimura, H., Nishimura, M., & Maciewicz, R. A. (2021). Large-scale plasma proteomics can reveal distinct endotypes in chronic obstructive pulmonary disease and severe asthma. Clinical and translational allergy, 11(10), e12091. https://doi.org/10.1002/clt2.12091

6. Vasunilashorn, S. M., Dillon, S. T., Chan, N. Y., Fong, T. G., Joseph, M., Tripp, B., Xie, Z., Ngo, L. H., Lee, C. G., Elias, J. A., Otu, H. H., Inouye, S. K., Marcantonio, E. R., & Libermann, T. A. (2022). Proteome-Wide Analysis Using SOMAscan Identifies and Validates Chitinase-3-Like Protein 1 as a Risk and Disease Marker of Delirium Among Older Adults Undergoing Major Elective Surgery. The journals of gerontology. Series A, Biological sciences and medical sciences, 77(3), 484–493. https://doi.org/10.1093/gerona/glaa326

Procedure and Methodology

Methods and Data Overview

Datasets

- Urinary proteomic measurements were obtained from patients with bladder cancer (BC), urology clinic controls (UC; presenting with hematuria), and healthy controls (HC), comprising 38 patients and 7,596 proteins measured per sample.

Data Preprocessing

- Protein measurements log-transformed to reduce skew and enable fair comparison across patients.
- Four extreme outlier samples were excluded during quality control.
- Low-variance and highly redundant proteins filtered prior to network construction to focus the analysis on informative molecular signals.

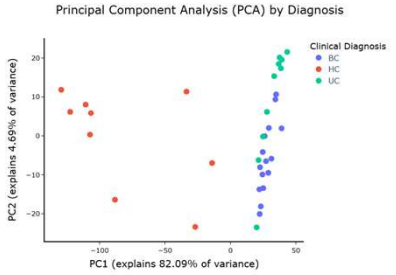
Module and Patient Analysis

- Proteins were organized into co-expression modules using a network-based approach (Walktrap algorithm).
- **Module activity scores** were computed for each patient across all protein modules.
- Patients were then clustered into **molecular groups** based on module activity profiles, **without using clinical disease labels.**

Biological Interpretation

- Modules showing differential activity across patient clusters were interpreted using pathway and gene ontology enrichment analysis (Reactome, Gene Ontology, KEGG).

Figure 1. Exploratory dimensionality reduction of patient proteomic profiles



Each point represents a patient and is colored by clinical diagnosis. Principal component analysis shows that nearly all variation in the dataset is explained by a single axis separating healthy from disease samples. Within the disease group, a smaller secondary axis reveals variation between bladder cancer and urinary controls.

Results

Figure 2: Protein Modules Represent Coordinated Biological Programs

Module 0: Proliferation & Metabolic Reprogramming

- Nucleotide metabolism
- Pyrimidine salvage
- Metabolic regulation

Interpretation: Active biosynthesis supporting cell growth and division

Module 6: Immune Activation & Cellular Stress

- IRF5-dependent signaling
- Viral response pathways
- DNA repair

Interpretation: Inflammatory and stress response to tissue damage

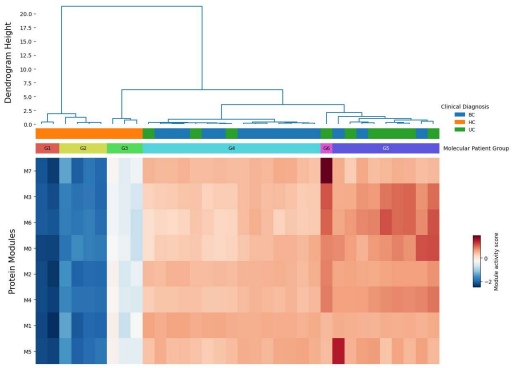
Module 3: Adaptive Remodeling & Repair

- Fatty acid metabolism
- mRNA splicing
- Axon guidance

Interpretation: Cellular adaptation and tissue reorganization

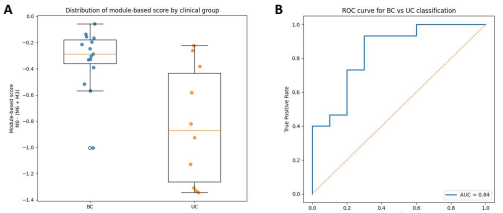
Functional enrichment analysis was performed using Reactome, Gene Ontology (GO), and KEGG pathway databases. Enriched pathways were summarized into core biological themes for each module. These modules represent fundamental biological response programs - growth, stress, and repair - that are shared across both malignant and non-malignant conditions.

Figure 3. Hierarchical clustering of patients based on protein module activity



Hierarchical clustering of patients based on protein module activity reveals distinct molecular subgroups. Annotation bars indicate clinical diagnosis and molecular cluster assignment, while the heatmap shows the module activity scores that drive patient similarity.

Figure 4: Module-Based Score Enables Classification of Bladder Cancer vs Urinary Controls



A module-based score (M0 – [M6 + M3]) was used to classify BC vs UC, reflecting proliferative activity adjusted for inflammatory and remodeling signals.
(A) The score shows separation between BC and UC.
(B) ROC analysis demonstrates in-sample discrimination (AUC = 0.84). Sensitivity = 93.3%, specificity = 70%.

Acknowledgments

I am grateful to Dr. Chandra Mohan, Department of Biomedical Engineering, University of Houston, for providing access to the urine proteomic dataset used in this research.

Discussion

- This study demonstrates that network-based analysis of high-dimensional urinary proteomic data can reveal biologically meaningful patient groupings. The clear separation of healthy controls from disease-associated clusters suggests that this approach captures robust signals related to pathological processes in the urinary system.
- The overlap between bladder cancer and urinary control groups highlights a key clinical challenge: distinguishing malignancy from non-malignant causes of hematuria. The observed differences in module activity suggest that patient groups reflect distinct biological states, potentially representing stable disease-associated processes versus more acute or high-intensity responses such as inflammation or tissue injury.
- The presence of urinary control patients within the bladder cancer-enriched cluster is particularly notable. These individuals may share underlying molecular features with cancer patients, likely reflecting overlapping biological pathways. Alternatively, they may represent early or undiagnosed disease; however, this interpretation remains speculative and requires further validation.

Conclusion

- Network-based proteomic analysis identified structured patterns of protein co-expression that distinguish healthy individuals from those with urinary pathology. While bladder cancer patients were enriched within a distinct cluster, overlap with urinary controls underscores the complexity of disease-specific biomarker discovery. These findings support the use of module-based approaches to reduce dimensionality and uncover biologically meaningful patterns, with potential to support classification pending further validation.

Limitations

- Small sample size: The limited number of patients reduces statistical power and increases the risk of overfitting.
- Lack of external validation: Findings have not been validated in independent cohorts.
- Cross-sectional design: The absence of longitudinal data limits the ability to assess disease progression or clinical outcomes.

Future Work

- Validation in larger cohorts: Replicate findings in larger, independent datasets to assess robustness and generalizability.
- Longitudinal follow-up: Track urinary control patients clustering with bladder cancer to determine whether they exhibit increased clinical risk or disease progression.
- Refinement of scoring models: Further develop and optimize module-based scoring approaches for improved classification performance.