Can Untargeted RNA-Seq of Urine Samples Diagnose UTIs in Children?

Max Homm[1,2], Neelanjay Anthwal[1,2], Sojin Lee[5], Geize Tomazetto[5], Marcia Kurs-Lasky[5], Andrew C. Doxey[1,2,3,4], Nader Shaikh[5]

[1]Department of Biology, University of Waterloo, [2]Waterloo Centre for Microbial Research, University of Waterloo, [3]Cheriton School of Computer Science, University of Waterloo, [4]Faculty of Health Sciences, Department of Medicine, McMaster University, [5]Division of General Academic Pediatrics, School of Medicine, University of Pittsburgh

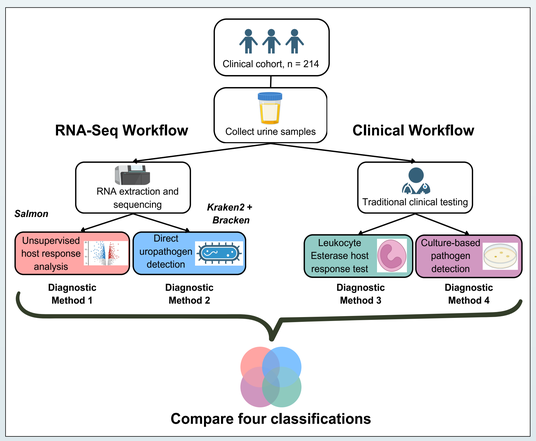
Abstract

UTI diagnosis in young children is challenging due to limitations of traditional tests. This study applied metatranscriptomic RNA-sequencing (RNA-seq) to 214 pediatric urine samples, using unsupervised learning to identify globally distinct immune response patterns and accurately detect uropathogens. RNA-seq also revealed immune-suppressing and novel uropathogens missed by standard methods. By integrating host and pathogen data, RNA-seq provides a more objective and comprehensive diagnostic approach, improving UTI detection and reducing unnecessary antibiotic use in children.

Introduction

- Urinary tract infections (UTIs) are among the most common bacterial infections in early childhood[1] but remain challenging to diagnose.
- Current diagnostic methods have limitations in detecting both pathogens (e.g., culture-based tests) and host responses (e.g., leukocyte esterase [LE] assays)[2].
- Metatranscriptomic RNA-seq (MT) offers a promising alternative by simultaneously detecting pathogens and host immune responses within environmental samples[3].
- Unlike metagenomics, MT captures real-time gene expression in both host and pathogens[3].
- This study applied MT to urine samples from 214 pediatric patients to evaluate its diagnostic potential.

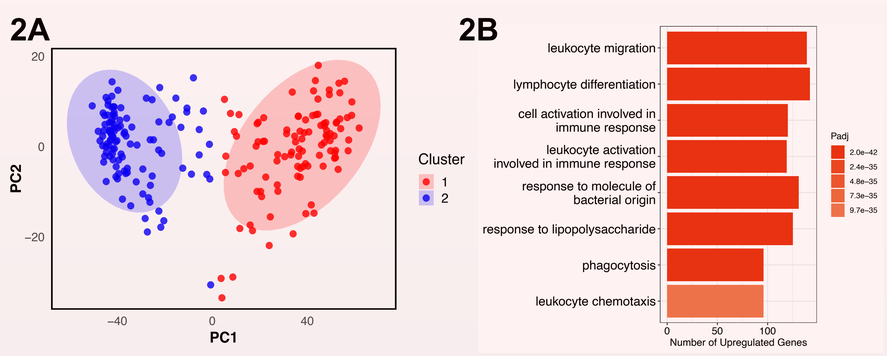
Methods



Methods workflow

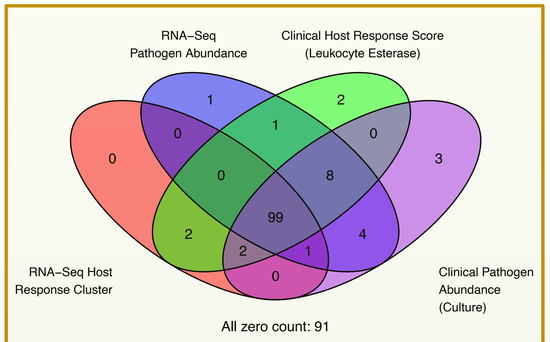
Figure 1. Urine samples from 214 children (1–36 months old) with UTI-like symptoms underwent RNA-sequencing and clinical testing, generating four independent diagnostic classifications based on host response and uropathogen profiles. Patients were assigned to positive or negative groups, and classifications were evaluated for overall agreement.

Results



Two distinct patient clusters emerge from human gene expression profiles derived by *Salmon*[4]

Figure 2. ***(A)*** PCA plot of global patient gene expression, reducing 62,198 human gene dimensions to two principal components. Points are colored by *k*-means cluster (k = 2), with ellipses indicating the 95% confidence intervals for each cluster. ***(B)*** Top 8 enriched Gene Ontology (GO) biological processes in cluster 1 relative to cluster 2, ranked by p-adjusted value. The strong immune-related process enrichment suggests that patients in cluster 1 are actively responding to an infection.

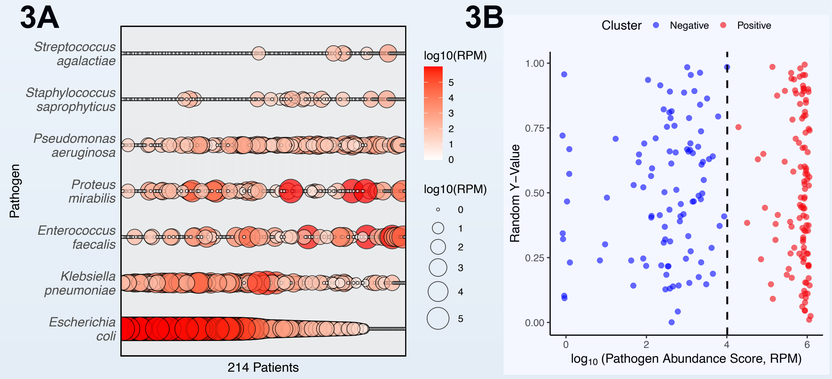


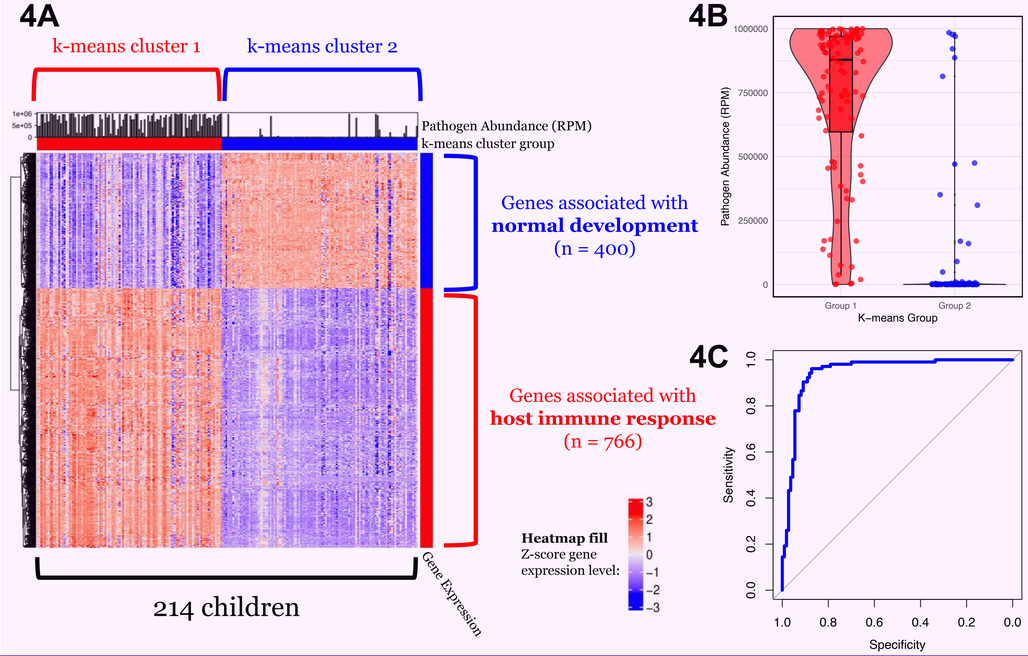
RNA-seq shows strong agreement with clinically derived diagnostics

Figure 5. Venn diagram highlighting ~89% agreement (190/214 patients) and ~11% disagreement (24/214 patients) between clinical and RNA-seq derived groups.

Patients with high and low uropathogen levels are identified via *Kraken*[5] and *Bracken*[6] taxonomic classification

Figure 3. ***(A)*** Bubble plot showing the relative abundance of seven common UTI-associated pathogens[7] (measured in reads-per-million, RPM) across 214 children, sorted in decreasing order by *E. coli* abundance. Human reads were removed before computing these RPM values. ***(B)*** One-dimensional clustering of each patient's logged pathogen abundance score (summation of RPM values from seven UTI-associated pathogens, *Fig 3A*). A random y-axis improves interpretability. Jenks Natural Breaks classification algorithm (k = 2) determined the optimal binary threshold (~4 log10RPM or ~10,000 RPM, black dotted line).





Strong agreement between host immune response and uropathogen abundance

Figure 4. ***(A)*** Heatmap of differentially expressed genes (y-axis) between two *k*-means-determined host gene clusters (x-axis), with thresholds *padj* < 1×10^{-25}, log2 fold-change > 2. An unlogged pathogen abundance score is overlaid as a bar plot above the heatmap to illustrate its strong relationship with gene expression patterns ***(B)*** Violin plots of pathogen abundance across two human gene expression clusters, with a highly significant difference (Wilcoxon test, $p = 1.92\times10^{-29}$). ***(C)*** Receiver operator curve showing that pathogen abundance scores (*Fig 3B*, x-axis) predict host response classes (*Fig 2A*) with AUC = 0.95 via logistic regression. It also accurately predicts LE activity (AUC = 0.96) and clinical culture-based tests (AUC = 0.99).

Conclusions

- RNA-seq independently identified potential UTIs and revealed distinct immune activation patterns, uncovering novel diagnostic biomarkers.
- *Kraken* and *Bracken* taxonomic classification linked uropathogen abundance to host immune response intensity, highlighting a strong association.
- Combining host and pathogen RNA-seq overcomes traditional diagnostic limitations, detecting clinically untested uropathogens and reducing human biases.
- RNA-seq enables cost-effective, less invasive diagnostics, with strong (~89%) agreement with clinical testing, potentially reducing unnecessary antibiotic use in children.

Future Directions

- Focus on human and pathogen genetic biomarkers to develop rapid, PCR-based diagnostic tests.
- Evaluate RNA-seq as a potential replacement for traditional tests, capturing host response and pathogen abundance in a single assay.
- Develop supervised models for early, non-invasive pyelonephritis detection, which is a severe kidney infection that represents a subset of UTIs.

References

1. Leung, A. K. C., Wong, A. H. C., Leung, A. A. M., & Hon, K. L. 2019. Urinary Tract Infection in Children. Recent Patents on Inflammation & Allergy Drug Discovery, 13(1), 2–18. https://doi.org/10.2174/1872213X13666181228154940
2. Bono, M. J., Reygaert, W. C., & Leslie, S. W. 2023. Urinary tract infection. Retrieved from National Library of Medicine website: https://www.ncbi.nlm.nih.gov/books/NBK470195/
3. Aguiar-Pulido, V., Huang, W., Suarez-Ulloa, V., Cickovski, T., Mathee, K., & Narasimhan, G. 2016. Metagenomics, Metatranscriptomics, and Metabolomics Approaches for Microbiome Analysis. Evolutionary Bioinformatics, 12s1, EBO.S36436. https://doi.org/10.4137/ebo.s36436
4. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nature Methods, 14(4), 417–419. https://doi.org/10.1038/nmeth.4197
5. Wood, D. E., Lu, J., & Langmead, B. 2019. Improved metagenomic analysis with Kraken 2. Genome Biology, 20(1). https://doi.org/10.1186/s13059-019-1891-0
6. Lu, J., Breitwieser, F. P., Thielen, P., & Salzberg, S. L. 2017. Bracken: estimating species abundance in metagenomics data. PeerJ Computer Science, 3, e104. https://doi.org/10.7717/peerj-cs.104
7. Mancuso, G., Midiri, A., Gerace, E., Marra, M., Zummo, S., & Biondo, C. 2023. Urinary Tract Infections: The Current Scenario and Future Prospects. Pathogens, 12(4), 623. https://doi.org/10.3390/pathogens12040623