Unveiling Potential Biomarkers for Urinary Tract Infection: An Integrated Bioinformatics Approach

Reza Maddah[1], Fahimeh Ghanbari[2], Maziyar Veisi[3], Eman Koosehlar[4], Marzieh Shadpirouz[5], Zarrin Basharat[6], Alireza Hejrati[7], Bahareh Shateri Amiri[7], Lina Hejrati[8]

[1]Department of Bioprocess Engineering, Institute of Industrial and Environmental Biotechnology, National Institute of Genetic Engineering and Biotechnology, Tehran, Iran, [2]Applied Physiology Research Center, Isfahan University of Medical, Isfahan, Iran, [3]Department of Veterinary Medicine Shahrekord, Shahrekord University, Chaharmahal and Bakhtiari Province, Shahrekord, Iran, [4]Department of Biotechnology, Faculty of Biological Science and Technology, University of Isfahan, Isfahan, Iran, [5]Department of Applied Mathematics, Faculty of Mathematical Sciences, Shahrood University of Technology, Semnan, Iran, [6]Alpha Genomics Private Limited, Islamabad 45710, Pakistan, [7]Internal Medicine Department, School of Medicine, Hazrat-e Rasool General Hospital, Iran University of Medical Sciences, Tehran, Iran, [8]Internal Medicine Department, School of Medicine, Iran University of Medical Sciences, Tehran, Iran

Abstract

Background: Urinary tract infections (UTIs) are a widespread health concern with high recurrence rates and substantial economic impact, and they can increase the prevalence of antibiotic resistance. This study employed an integrated bioinformatics approach to identify key genes associated with UTI development, offering potential targets for interventions.

Materials and Methods: For this study, the microarray dataset GSE124917 from the Gene Expression Omnibus (GEO) database was selected and reanalyzed. The differentially expressed genes (DEGs) between UTIs and healthy samples were identified using the LIMMA package in R software. In this section, Enrichr database was utilized to perform functional enrichment analysis of DEGs. Subsequently, the protein-protein interaction (PPI) network of the DEGs was constructed and visualized through Cytoscape, utilizing the STRING online database. The identification of hub genes was performed using Cytoscape's cytoHubba plug-in employing various methods. Receiver operating characteristic (ROC) analysis was performed to assess the diagnostic accuracy of hub genes.

Results: Among the outcomes of the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis, the tumor necrosis factor (TNF) signaling pathway was identified as one of the notable pathways. The PPI network of the DEGs was successfully established and visualized in Cytoscape with the aid of the STRING online database. Using cytoHubba with different methods, we identified seven hub genes (STAT1, IL6, IFIT1, IFIT3, IFIH1, MX1, and IRF7). Based on the ROC analysis, all hub genes showed high diagnostic value.

Conclusion: These findings provide a valuable baseline for future research aimed at unraveling the intricate molecular mechanisms behind UTI.

Keywords: Bioinformatics, computational, molecular biologies, urinary tract infections

INTRODUCTION

Urinary tract infections (UTIs) are recognized as one of the most common bacterial infections, affecting approximately 150 million individuals worldwide annually.[1,2] According to data estimates, sixty percent of females report having a history of at least one UTI, and over 50% of women experience recurrent infections within 6 months.[3] The most common pathogen is urinary tract pathogenic *Escherichia coli* (UPEC). Approximately seventy-five percent of uncomplicated UTIs and sixty-five percent of complicated UTIs are caused by

Access this article online

Quick Response Code:

Website:
www.advbiores.net

DOI:
10.4103/abr.abr_355_23

How to cite this article: Maddah R, Ghanbari F, Veisi M, Koosehlar E, Shadpirouz M, Basharat Z, *et al*. Unveiling potential biomarkers for urinary tract infection: An integrated bioinformatics approach. Adv Biomed Res 2024;13:44.

UPEC.[4,5] The hemolysin, fimbriae, and iron acquisition system are the primary virulence factors of UPEC responsible for causing UTIs. This infection can result in a significant economic burden on both the public and society. While antibiotics are the current standard treatment for UTIs, the rising rates of drug resistance have diminished the therapeutic efficacy of these drugs for treating UTIs.[6,7] Advancements in technology such as microarray and bioinformatics have revolutionized the exploration of molecular mechanisms of diseases and improved disease diagnosis and treatment. A comprehensive analysis of microarray data combined with bioinformatics approaches is essential. By these techniques, researchers gain a comprehensive understanding of the genetic and molecular factors driving UTIs. This knowledge can aid in the development of more accurate diagnostic tools for early detection of UTIs, as well as the identification of potential therapeutic targets. Furthermore, it can provide insights into personalized treatment strategies based on an individual's genetic profile, allowing for more precise and effective interventions.[8,9]

In this study, we employed bioinformatics methodologies to identify pivotal genes implicated in the progression of UTI, which can be offered as an intervention target for patients with UTIs.

MATERIAL AND METHODS

Source of data and analysis

The microarray profile dataset was retrieved from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) databases.[10] The dataset GSE124917, obtained from the Agilent-072363 SurePrint G3 Human GEN v3 8x60K Microarray (GPL21185), consisted of a total of six samples. This dataset included three normal samples and three samples infected with UPEC. The LIMMA package[11] of R software was employed to identify differentially expressed genes (DEGs) in the UPEC group compared with the corresponding control. DEGs were filtered based on an adjusted *P* value of less than 0.05 and an absolute log-fold change greater than 1.0 and were considered for downstream analysis. To visualize DEGs that were upregulated or downregulated, a volcano plot was generated. This plot provides a graphical representation of the fold change and statistical significance of the DEGs. A comprehensive list of DEGs can be found in Supplementary File 1. This file contains detailed information regarding the identified DEGs, including their corresponding fold change values, statistical significance, and additional relevant annotations ethic commete approval ID (IR.IUMS.REC.1402.179).

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional enrichment analysis

The enrichment analysis of DEGs was conducted using the Enrichr online tool.[12] This analysis aimed to explore the molecular and functional enrichments associated with the DEGs. Within Enrichr, both GO and KEGG enrichment analyses were performed on the DEGs.

The GO enrichment analysis provides insights into the functional categorization of DEGs across three domains: biological process (BP), molecular function (MF), and cellular component (CC). However, the KEGG enrichment analysis identifies the key pathways in which the DEGs are likely to be involved. To determine statistical significance, a *P* value threshold of less than 0.05 was applied.

Protein-protein interaction (PPI) network and screening hub genes

STRING, an online biological database dedicated to predicting PPIs, was employed for the construction of the PPI network.[13] To visualize the PPI network, Cytoscape software was utilized, which is specifically designed for visualizing complex biological networks.[14] Furthermore, the cytohubba Cytoscape plug-in[15] was employed for the identification of hub genes among the DEGs. The PPI network construction involved utilizing the STRING database, with a focus on interactions with median confidence scores exceeding 0.7, which were deemed statistically significant. Subsequently, Cytoscape was employed to visually analyze the constructed PPI network. To identify hub genes, the cytoHubba plug-in was employed, utilizing the degree, closeness, and betweenness centrality methods, which enabled the sorting of hub genes. The intersection of hub genes among these methods was designated as hub genes in this study.

Evaluation of the diagnostic value

To assess the diagnostic effectiveness of these biomarkers, receiver operating characteristic (ROC) curves were generated using GraphPad Prime 9.[16] The area under the ROC curve (AUC) was then calculated, with an AUC value greater than 0.9 indicating that the biomarkers possessed favorable diagnostic value.

RESULTS

Data analysis

In the present investigation, a microarray dataset consisting of individuals with UTI and healthy individuals was subjected to analysis using the LIMMA package in the R software. The analysis revealed a total of 801 DEGs, with 579 genes showing upregulation and 222 genes showing downregulation. The DEGs are visually represented in a volcano plot, as depicted in Figure 1. The list of DEGs is shown in Supplementary File 1.

GO and KEGG functional enrichment analysis

The Enrichr database was employed to conduct GO enrichment and KEGG pathway analyses on a set of 1599 DEGs. The GO analysis demonstrated significant enrichments of the DEGs in BP terms [Figure 2], including the cytokine-mediated signaling pathway (GO: 0019221), cellular response to interferon-gamma (GO: 0071346), and cellular response to type I interferon (GO: 0071357). Regarding MF terms [Figure 2], the DEGs exhibited notable enrichments in ribonucleic acid (RNA) polymerase II transcription regulatory region sequence-specific deoxyribonucleic

acid (DNA) binding (GO: 0000977), cis-regulatory region sequence-specific DNA binding (GO: 0000987), and RNA polymerase II cis-regulatory region sequence-specific DNA binding (GO: 0000978). Furthermore, in terms of CC terms [Figure 2], the DEGs were predominantly enriched in the nucleus (GO: 0005634), intracellular membrane-bounded organelle (GO: 0043231), and integral component of plasma membrane (GO: 0005887).

Additionally, the KEGG pathway analysis highlighted several significant pathways [Figure 3], including the tumor necrosis factor (TNF) signaling pathway, viral protein interaction with cytokine and cytokine receptor, cytokine-cytokine receptor interaction, influenza A, and nucleotide oligomerization domain (NOD)-like receptor signaling pathway. These pathways were identified as the most important in relation to the DEGs.

PPI network and screening hub genes
A PPI network [Figure 4], comprising 369 nodes and 1808 edges, was obtained from the STRING database and subsequently visualized using Cytoscape. To identify hub genes within the PPI network, cytoHubba analysis was performed. Notably, the analysis identified STAT1, IL6, IFIT1, IFIT3, IFIH1, MX1, and IRF7 as hub genes shared among three different methods: degree, closeness, and betweenness.

### Evaluation of the diagnostic value
The accuracy of the prognostic potential of the hub genes was evaluated using the ROC curve. The AUC was calculated to assess and compare the diagnostic capabilities of the hub genes. The findings revealed significant discriminative power, as indicated by estimated AUC values ranging from 0.9 to 1 [Figure 5]. Notably, based on the ROC analysis, all hub genes exhibited high diagnostic values with AUC values exceeding 0.9.

DISCUSSION

UTIs are highly prevalent, affecting millions of individuals worldwide each year. Recurrence rates among women are
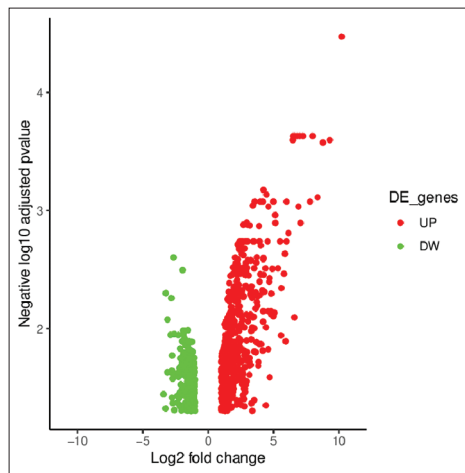

Figure 1: A volcano plot illustrating the DEGs. In this plot, downregulated genes are shown by green points, while upregulated genes are depicted by red points
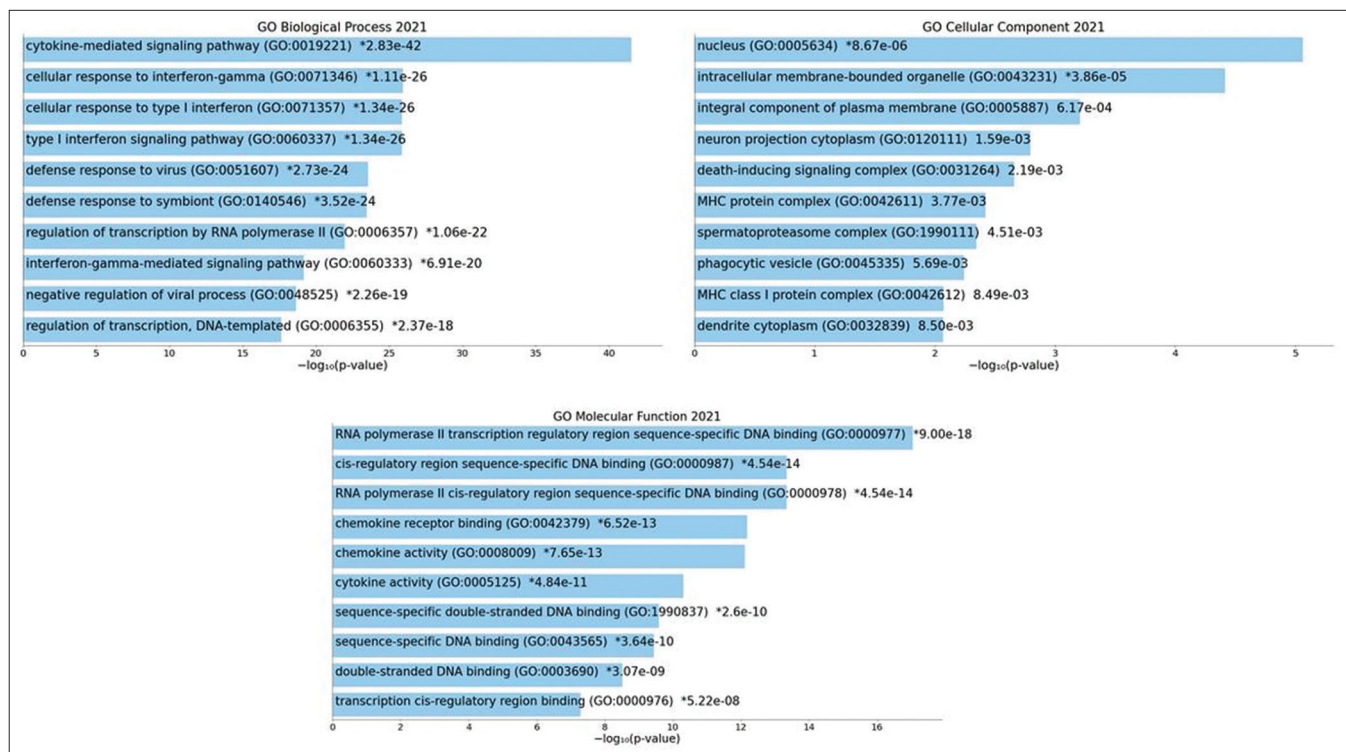
Figure 2: Results of the Gene Ontology analysis performed on the DEGs. This analysis provides insights into the functional categorization associated with the identified DEGs
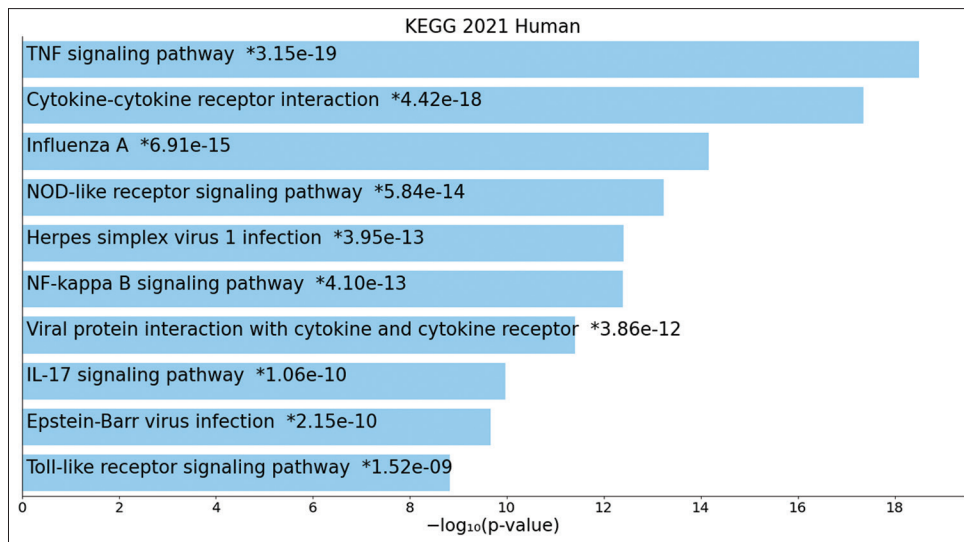
Figure 3: Outcomes of the KEGG pathway enrichment analysis conducted on the DEGs. This analysis provides valuable insights into the specific biological pathways that are significantly enriched with the identified DEGs
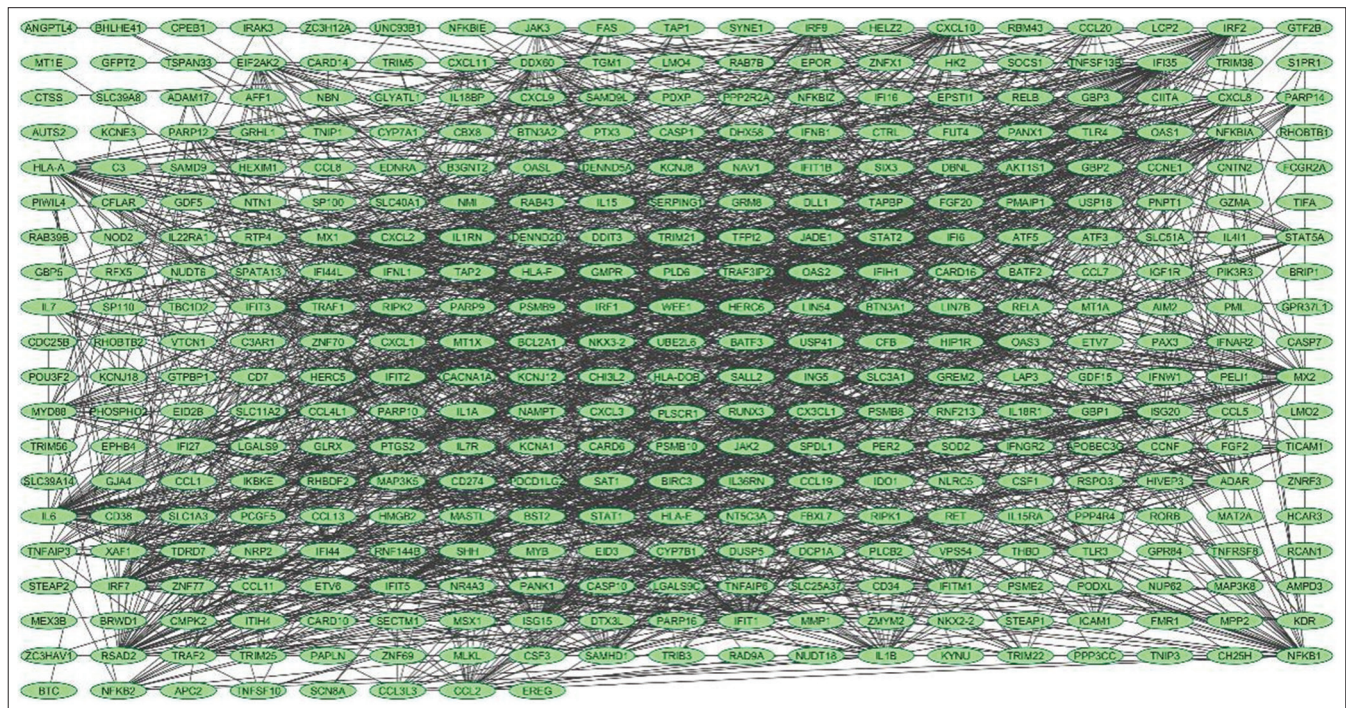

Figure 4: PPI network of genes, wherein the blue ellipses represent individual genes, and the gray lines depict the interactions between them. Notably, the PPI network consisted of 369 nodes (genes) and 1808 edges (interactions)

high, with a considerable proportion experiencing multiple infections within a brief period.[17,18] This study utilized a comprehensive approach integrating various bioinformatics methodologies to discover key genes associated with the development of UTIs, highlighting their potential as targets for interventions.

In this study, KEGG pathway enrichment analysis revealed that the TNF signaling pathway exhibits notable significance, suggesting its potential importance in the initiation and progression of UTIs. The TNF signaling pathway plays an important role in the immune response to infections. Upon bacterial attack, macrophages and neutrophils, the protective cells of the urinary tract, are recruited to the site of infection. These cells produce TNF, which binds to its receptors on various cell types, including immune cells, epithelial cells, and endothelial cells. Activation of TNF receptors leads to a signaling cascade, and activating transcription factors such as nuclear factor kappa B (NF-Kb) and activator protein 1 (AP-1), result in increased expression of genes related to pro-inflammatory cytokines and chemokines. This recruitment
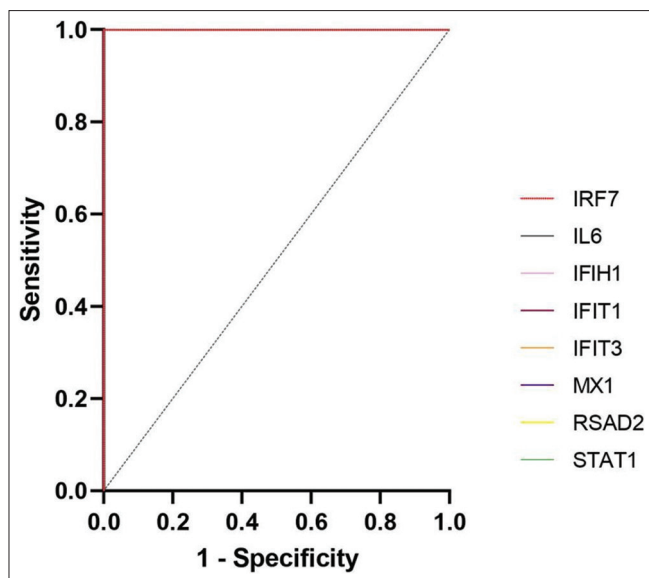
Figure 5: Validation of hub genes through receiver operating characteristic (ROC) curves

and activation of additional immune cells aid in bacterial clearance. However, excessive production of TNF can also contribute to tissue damage and inflammation. Modulation of the TNF signaling pathway may serve as a therapeutic target to reduce inflammation and tissue injury. Overall, the TNF signaling pathway plays a crucial role in the immune response to UTIs, but dysregulation of this pathway can also contribute to tissue damage and inflammation.[19]

The present study identified seven hub genes (STAT1, IL6, IFIT1, IFIT3, IFIH1, MX 1, and IRF7) that are significantly associated with the progression of UTIs. These findings underscore the crucial role of these genes in the development and advancement of UTIs.

Based on reports, not only has STAT1 been recognized as an essential player in the biological response to various Toll-like receptors (TLRs), including TLR-4, but it also plays a vital role in the modulatory mechanism that contributes to the development of inflammatory diseases or infections.[20,21]

In this regard, C. H. Ho *et al*. showed that STAT-1 is essential for controlling UPEC invasion and infection in uroepithelial cells. They also showed that sugar could increase UPEC infection in uroepithelial cells by overexpression of the TLR-4, STAT-1, and pro-inflammatory IL-6.[22,23]

In 2016, Cheng *et al*. found that UPEC infection in the epididymis could increase IL-6 and TNF-α through NF-kB signaling with TLR-4 or TLR-5.[24] In 2018, Ching *et al*. demonstrated that IL-6 deficiency was linked to a rise in the development of intracellular bacterial populations (Ching *et al*., 2018).[24] According to the study of Engelsöy *et al*. (2019), UPEC infection causes an increase in TNF-α and IL-6 gene levels, reducing biofilm formation and hemolytic activity.[25] Furthermore, Vegaj-Hernández *et al*. in 2021 revealed that the UPEC proteins FimH and FliC cause the secretion of IL-6.[26]

The signaling pathway associated with IFIT1 and IFIT3 is a key component of the host immune response to infection. Bacteria such as *Escherichia coli (E. coli)* can activate TLRs on the surface of epithelial cells in the urinary tract, leading to the production of IFN-alpha and IFN-beta interferons. These cytokines activate the JAK-STAT signaling pathway, which regulates the expression of other interferon-stimulated genes, including IFIT-3 and IFIT1. IFIT3 inhibits the replication of *E. coli* and other uropathogenic bacteria and modulates the host immune response. In addition to its direct antimicrobial effects, IFIT3 signaling also plays a role in the activation of adaptive immune cells. IFIT3 promotes the differentiation of cluster of differentiation (CD) 8+ cells into cytotoxic T lymphocytes (CTLs), which can eliminate infected cells. Overall, the signaling pathway associated with IFIT3 is an important part of the host immune response to infection.[27-29]

Despite the valuable insights provided by this study, there are some limitations that should be acknowledged:

Reliance on a single dataset: The analysis was conducted using a specific microarray dataset (GSE124917), which may limit the generalizability of the findings. Including multiple independent datasets or replicating the study with different datasets would strengthen the robustness of the results. Bioinformatics approach limitations: While the bioinformatics approach employed in this study is valuable for identifying potential key genes and pathways, it relies on computational predictions. The findings should be interpreted with caution and further validated through experimental studies. Lack of clinical data integration: This study primarily focused on molecular and genetic aspects of UTIs. The integration of clinical data, such as patient characteristics, treatment regimens, and outcomes, would provide a more comprehensive understanding of the disease and strengthen the relevance of the findings in a clinical context. Addressing these limitations would contribute to a more comprehensive understanding of the molecular mechanisms underlying UTI development and facilitate the translation of these findings into clinical applications. To advance our understanding further and address gaps in current knowledge, future studies in this field should focus on several key aspects. Firstly, exploring the functional roles of the identified hub genes in the pathogenesis of UTIs is essential. Furthermore, the validation of the diagnostic potential of the identified hub genes in diverse patient populations and clinical settings is crucial. Lastly, as the field of bioinformatics continues to evolve, integrating multi-omics data and employing advanced computational approaches could provide a more holistic view of UTI pathogenesis.

Conclusion

In conclusion, this study brings us one step closer to understanding the complex puzzle of UTIs. By unraveling the key regulatory pathways involved, we now have a better grasp of how UTIs develop and progress. However, there is still much more to learn, and further research is crucial to validate

these findings. By unraveling the mysteries of this condition, we can develop innovative strategies and interventions that go beyond simply treating the symptoms. Our collective efforts will pave the way for a future where UTIs are more effectively understood, prevented, and managed, leading to happier and healthier lives for all.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Waller TA, Pantin SA, Yenior AL, Pujalte GG. Urinary tract infection antibiotic resistance in the United States. Prim Care 2018;45:455-66.
2. Zalewska-Piątek B, Piątek R. Phage therapy as a novel strategy in the treatment of urinary tract infections caused by *E. coli*. Antibiotics (Basel) 2020;9:304.
3. Medina M, Castillo-Pino E. An introduction to the epidemiology and burden of urinary tract infections. Ther Adv Urol 2019;11:1756287219832172.
4. Rosen DA, Hooton TM, Stamm WE, Humphrey PA, Hultgren SJ. Detection of intracellular bacterial communities in human urinary tract infection. PLoS Med 2007;4:e329.
5. Foxman B. The epidemiology of urinary tract infection. Nat Rev Urol 2010;7:653-60.
6. Jacobsen SM, Stickler DJ, Mobley HL, Shirtliff ME. Complicated catheter-associated urinary tract infections due to *Escherichia coli* and Proteus mirabilis. Clin Microbiol Rev 2008;21:26-59.
7. Mancuso G, Midiri A, Gerace E, Marra M, Zummo S, Biondo C. Urinary tract infections: The current scenario and future prospects. Pathogens 2023;12:623.
8. Khan J, Bittner ML, Chen Y, Meltzer PS, Trent JM. DNA microarray technology: The anticipated impact on the study of human disease. Biochim Biophys Acta 1999;1423:M17-28.
9. Kang JU. Medical implementation of microarray technology. Korean J Clin Lab Sci 2020;52:310-6.
10. Clough E, Barrett T. The gene expression omnibus database. Methods Mol Biol 2016;1418:93-110.
11. Smyth GK. Limma: Linear models for microarray data. Bioinformatics and Computational Biology Solutions Using R and Bioconductor. New York, NY: Springer New York, 2005. p. 397-420.
12. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, *et al*. Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 2016;44:W90-7.
13. Mering CV, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: A database of predicted functional associations between proteins. Nucleic Acids Res 2003;31:258-61.
14. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, *et al*. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498-504.
15. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: Identifying hub objects and sub-networks from complex interactome. BMC Syst Biol 2014;8:1-7.
16. Swift ML. GraphPad prism, data analysis, and scientific graphing. J Chem Inf Comput Sci 1997;37:411-2.
17. Zalewska-Piątek BM, Piątek RJ. Alternative treatment approaches of urinary tract infections caused by uropathogenic *Escherichia coli* strains-Review. Acta Biochim Pol 2019;66:129-38.
18. Foxman B. Urinary tract infection syndromes: Occurrence, recurrence, bacteriology, risk factors, and disease burden. Infect Dis Clin North Am 2014;28:1-3.
19. Yamada HI, Odonnell MA, Matsumoto TE, Luo YI. Interferon-γ up-regulates toll-like receptor 4 and cooperates with lipopolysaccharide to produce macrophage-derived chemokine and interferon-γ inducible protein-10 in human bladder cancer cell line RT4. J Urol 2005;174:1119-23.
20. Liu L, Yue Y, Xiong S. NK-derived IFN-γ/IL-4 triggers the sexually disparate polarization of macrophages in CVB3-induced myocarditis. J Mol Cell Cardiol 2014;76:15-2
21. Chaves de Souza JA, Nogueira AV, Chaves de Souza PP, Kim YJ, Silva Lobo C, Pimentel Lopes de Oliveira GJ, *et al*. SOCS3 expression correlates with severity of inflammation, expression of proinflammatory cytokines, and activation of STAT3 and p38 MAPK in LPS-induced inflammation *in vivo*. Mediators Inflamm 2013;2013:650812.
22. Ho CH, Liu SP, Fan CK, Tzou KY, Wu CC, Cheng PC. Insulin downregulated the infection of uropathogenic *escherichia coli* (UPEC) in bladder cells in a high-glucose environment through JAK/STAT signaling pathway. Microorganisms 2021;9:2421.
23. Ho CH, Fan CK, Wu CC, Yu HJ, Liu HT, Chen KC, *et al*. Enhanced uropathogenic escherichia coli-induced infection in uroepithelial cells by sugar through TLR-4 and JAK/STAT1 signaling pathways. J Microbiol Immunol Infect 2021;54:193-205.
24. Cheng L, Chen Q, Zhu W, Wu H, Wang Q, Shi L, *et al*. Toll-like receptors 4 and 5 cooperatively initiate the innate immune responses to uropathogenic *escherichia coli* infection in mouse epididymal epithelial cells. Biol Reprod 2016;94:58-1.
25. Ching CB, Gupta S, Li B. Interleukin-6/Stat3 signaling has an essential role in the host antimicrobial response to urinary tract infection. Kidney international. 2018; 93:1320-9.
26. Vega-Hernández R, Ochoa SA, Valle-Rios R, Jaimes-Ortega GA, Arellano-Galindo J, Aparicio-Ozores G, *et al*. Flagella, Type I fimbriae and curli of uropathogenic *Escherichia coli* promote the release of proinflammatory cytokines in a coculture system. Microorganisms 2021;9:2233.
27. Moskophidis D, Kioussis D. Contribution of virus-specific CD8+ cytotoxic T cells to virus clearance or pathologic manifestations of influenza virus infection in a T cell receptor transgenic mouse model. J Exp Med 1998;188:223-32.
28. Naranjo NM, Salem I, Harris MA, Languino LR. IFIT3 (interferon induced protein with tetratricopeptide repeats 3) modulates STAT1 expression in small extracellular vesicles. Biochem J 2021;478:3905-21.
29. Zhao H, Boije H, Granberg F, Pettersson U, Svensson C. Activation of the interferon-induced STAT pathway during an adenovirus type 12 infection. Virology 2009;392:186-95.