

Female Urinary Microbiome Analysis and Artificial Intelligence Enhances the Infectious Diagnostic Yield in Precision Medicine

David Baugher, Alexander W Larsen, Yuan Chen, Crystal R. Icenhour and C Alexander Valencia*

Aperiomics, Sterling, VA 20166.

*Correspondence:

C Alexander Valencia, PhD, 45925 Horseshoe Drive, Suite 170, Sterling, VA 20166, E-mail: avalenci2001@gmail.com.

Received: 20 June 2021; Accepted: 28 July 2021

Citation: Baugher D, Larsen AW, Chen Y, et al. Female Urinary Microbiome Analysis and Artificial Intelligence Enhances the Infectious Diagnostic Yield in Precision Medicine. *Microbiol Infect Dis*. 2021; 5(4): 1-8.

ABSTRACT

*We assessed the diagnostic yield of metagenomics urine sample testing in patients with urological symptoms. We conducted microbiome analysis of 86 female urine samples that included 17 healthy controls and 69 patients. Natural language processing (NLP), a subfield of artificial intelligence, was used to create a pathogen identification tool, Xplore-AI, to assess the potential pathogens in all of the samples. Meanwhile, report summaries that were written by infectious disease experts were compared to the NLP results to investigate its accuracy. The results showed that the NLP system reported 97% of patient samples had at least one pathogen over three standard deviations from values found in healthy controls. Similarly, 84% of patients had two or more classified pathogens. These diagnostic percentages were consistent with the infectious disease expert summaries. However, some pathogens like *Aerococcus urinae* were present in 13 patient samples, but only reported in one summary. In conclusion, this study demonstrated the high diagnostic yield in females with urological symptoms following metagenomic analysis and the ability of using an NLP-based system to identify pathogens to improve the accuracy of the reportable species.*

Keywords

Urinary tract infections, Urine DNA, Pathogens, Infectious diseases.

Introduction

Infectious diseases cause significant morbidity and mortality in populations worldwide [1]. The identification of the infectious agents is key to treating patients with infectious diseases by clinicians. A low detection rate has been established by conventional culture methodology, especially for microbes that require specific growth media. Thus, reaching a precision diagnosis through conventional methods has proven challenging. Although culture-independent techniques such as serologic assays and nucleic acid amplification tests permit the identification of pathogens, these methods require a priori knowledge of desired targets, which often leads to impractically time consuming and sequential testing. In addition, novel pathogens are missed by these standard methods. Estimates implicated that up to 60% of infection cases were treated without the identification of the causative infectious agent after the application of standard testing [2]. The lack of a targeted and

timely diagnosis impedes the utilization of precision antimicrobial treatment and overuse of unnecessary broad-spectrum antibiotics, causing antimicrobial resistance, and elevated healthcare costs.

In adults, urinary tract infections (UTIs) are the most common type of infection. The substantial clinical load of UTIs accounts for 1% of US clinical resources equaling more than \$3.7 billion annually [3]. Worldwide, over 150 million people will develop a UTI in a given year. Women are affected by UTIs more frequently and have a 50% chance of developing one at some point in their lives [4,5]. Although UTIs are more predominant in sexually active young women, UTI risk also rises with age predisposing menopausal and elderly women as well. UTIs are caused by a number of bacteria and fungi, but frequent uropathogens include *Escherichia coli*, *Klebsiella pneumoniae*, *Enterococcus faecalis*, *Candida albicans* and *Proteus mirabilis* [6]. UTIs are assumed to be easily treated with antibiotic therapy, however, antibiotic resistance and allergy complications frequently hinder eradicating infections [6]. This leads to prolonged cycles of infection, known as recurrent UTI (rUTI), that reduce the quality of life [4]. Additionally, UTIs can

affect the kidneys with serious consequences of pyelonephritis and life-threatening urosepsis. When refractory to antibiotic therapy, rUTI cases can last many years and potentially require a cystectomy [7]. Earlier and more comprehensive testing is required to aid doctors in circumventing these painful outcomes.

Unlike 16S/ITS sequencing, shotgun metagenomic sequencing (metagenomics) sequences all given genomic DNA from a sample, which permits unbiased result and is suitable for the elucidation of rare, novel, and atypical infectious agents. There is burgeoning interest about the application of metagenomics to infectious diseases as a diagnostic tool. Due to its sensitivity, speed, and cost-effectiveness, metagenomics may soon become a part of the standard diagnostic testing, thus replacing less comprehensive techniques such as culture and PCR-based methods [8]. Heretofore, metagenomics clinical applications have only been applied to case reports or small-scale cohort studies, with a focus pathogen detection from cerebrospinal fluid and plasma samples [9-12]. In addition to the microbe detection challenge, interpretation of metagenomics results are complex due to the large number of microbes that can be identified in a given sample based on a balance of environmental exposure and immune state, which demands further clinical contextual literature [12-16]. However, the thousands of microbes discoverable by metagenomics methods makes studying all surrounding literature manually a nearly impossible task [17].

Natural Language Processing, also known as NLP, is a branch of artificial intelligence and the intersection of linguistics and computer science. It generally started in the 1950s, and it has been quickly evolving during the past decades. Nowadays, many different classes of machine-learning algorithms have been applied to NLP tasks, such as feature learning and deep neural networks. NLP has been widely used in our daily lives, such chatbots and voice assistants, and healthcare is another promising field that NLP can help. People already started to use NLP at work such as clinical decision support, clinical trial matching and clinical documentation improvement. Several NLP-based tools that are focused on microbes have been built, such as Disbiome and AMRFinderPLUS, to link microbiome to disease and identify AMR genes [21,22]. Outside web searches, no freely available software is capable of determining microbe pathogenicity from PubMed literature.

In this study, we have assessed the diagnostic yield obtained by clinician reviews in contrast to Xplore-AI in clinical practice using consecutive unbiased urine samples utilizing metagenomics to broaden pathogen detection.

Methods

Samples and sequencing

Seventeen healthy female urine samples and 69 female patient urine samples were collected from the most recent consecutive samples. The average age of the patient group was 48 years of age with a standard deviation of 16. The age ranges had a normal distribution with a slight skew towards the older age range. Sixty-

one patients were from the United States with the second most common country being Canada at four patients. Control patients were all from the United States with a median age of 34 years of age ranging from 20-55. All subjects provided consent to participate in this study.

Urine DNA was isolated using the Quick-DNA Urine kit (Zymo Research, Irvine, CA), per the manufacturer's instructions. Libraries were prepared using the KAPA HyperPlus Prep Kit (KAPA Biosystems, Wilmington, MA). Quality control and quantification of the samples was assessed by fluorometry using the Qubit 2.0 Fluorometer (Thermo Fischer, Waltham, MA). DNA Libraries were sequenced on the NextSeq 500 (Illumina, San Diego, CA) using paired-end reads (2x75 bp). The quality of raw sequencing reads was assessed using FastQC (v 0.11.8) and filtered according to the following rules: 1) reads with average quality score below 25 were discarded; 2) 5' or 3' regions with 'N' or average quality score below 15 were trimmed; 3) trimmed reads with less than 35 bp were discarded and 4) low complexity reads were also discarded in the following analysis. All filtered sequencing data were run through Xplore-PATHO pipeline (Aperionics, Inc., Sterling, VA), the proprietary software for alignment, genome binning, relative abundance recalculation, and evaluating hit confidence scores. The core algorithm has been used by previous studies involving Pathoscope 2.0 [22-25]. A minimal of 30 reads were required to be evenly distributed across the genome, and the identified species with less than 0.1% relative abundance were removed from further analysis.

Clinical reviewers searched for well cited uropathogens within the report first and compared them to internally run historic data of healthy controls to determine outliers to report as likely pathogens. Other microorganisms that were also outside of expected values but without supporting literature including circumstantially reported uropathogens were put into the possible pathogen category by clinical reviewers.

Pathogen identification

An NLP-based system, Xplore-AI, was built with the intent of aid in the determination of the pathogen status of microbes from the scientific and medical literature. The data used for the NLP pipeline included ~20 million PubMed abstract texts, ~230 pathogen/infection terms (entities), ~300 symptom terms (entities), ~5000 disease terms (entities), ~750 environment terms (entities), ~500 antimicrobial terms (entities), ~140,000 microbe names (entities) and ~500 parse patterns (syntactic relationships between microbe and pathogen entities).

The PubMed abstracts were processed on the sentence level. Sentences were segmented by Spacy, and all sentences containing both infection and microbe terms as previously described were kept for further steps [18]. These sentences became "infection links" and were viewed as weak associations between the microbe and infection. An NLP method known as dependency parsing was then used on these links in order to find "verified infection links", which had a stronger causal nature.

Dependency parsing, included in the Spacy package, constructed a syntactic tree structure for the sentences. This method was used in conjunction with vetted parse patterns, syntactic patterns between the microbe and infection entities that were verified as implying a causal nature between the entities. Links for symptom, disease, environment, and antimicrobial entities were also generated, but not parsed.

A random forest classifier from scikit-learn was used to classify microbes as pathogenic or non-pathogenic solely from literature sources. The classifier was trained on a dataset of 143 known uropathogens and 33 known non-pathogens. The values used in feature construction were the total number of all entity links, the total number of “infection links”, and the total number of “verified infection links”. The ROC-AUC was used as a performance measure for the model and it achieved a score of 1.0, the highest score possible. In order to classify a microbe in a patient sample as being pathogenic the NLP pipeline was used in conjunction with the microbial abundance data of healthy controls. An abundance range was used in the determination and a microbe having an abundance of more than 3 standard deviations from healthy means or absent in healthy controls was considered significant. The method used to determine microbial pathogenicity in a sample can be seen in Figure 1.

Results

The average age of patients was 48 years old with a standard deviation of 16. The age ranges had a normal distribution with a slight skew towards the older age range. Sixty-one patient samples

originated from the United States with the second most common four samples from Canada. Figure 2A showed that the most common clinical symptom category by organ system was Urinary Tract symptoms at 63.8% of patients reporting those symptoms and the second most common category was Other which included fever, malaise, and chills / sweats at 49.3% of patients. Similarly, Figure 2B showed the most common ICD-10 supported clinical suspicions. The top clinical suspicion reported by their doctors, based on ICD-10 codes N30-39, were UTI and cystitis at 37.7% of patients. The second most common finding showed that many patients without any sort of ICD-10 codes included international patients for a total of 18.8%. The next most common clinical suspicion, based on ICD-10 codes R10-19, was gastrointestinal disease at 14.5% of patients.

The report clinical reviewers, who have years of infectious microbial profiling post-graduate work, created summaries that found 79.7% of the patient samples represented likely infections, while 18.8% represented possible infections (Figure 3A). The NLP pipeline found important pathogens in every sample and 92.8% of patient samples were characterized by pathogenic polymicrobial communities (Figure 3B). Specifically, 59.4% of samples had five or more microbes classified as important pathogens by the NLP pipeline (Figure 3B). Overall, the ratios indicated that 98.6% of urinary samples had at least one article supporting a clinically relevant microorganism shown (Figure 3C). Comparing the clinical review findings with the NLP findings, it was uncovered that potentially important microbes were not being mentioned in the report summaries. It was found that each process favored a

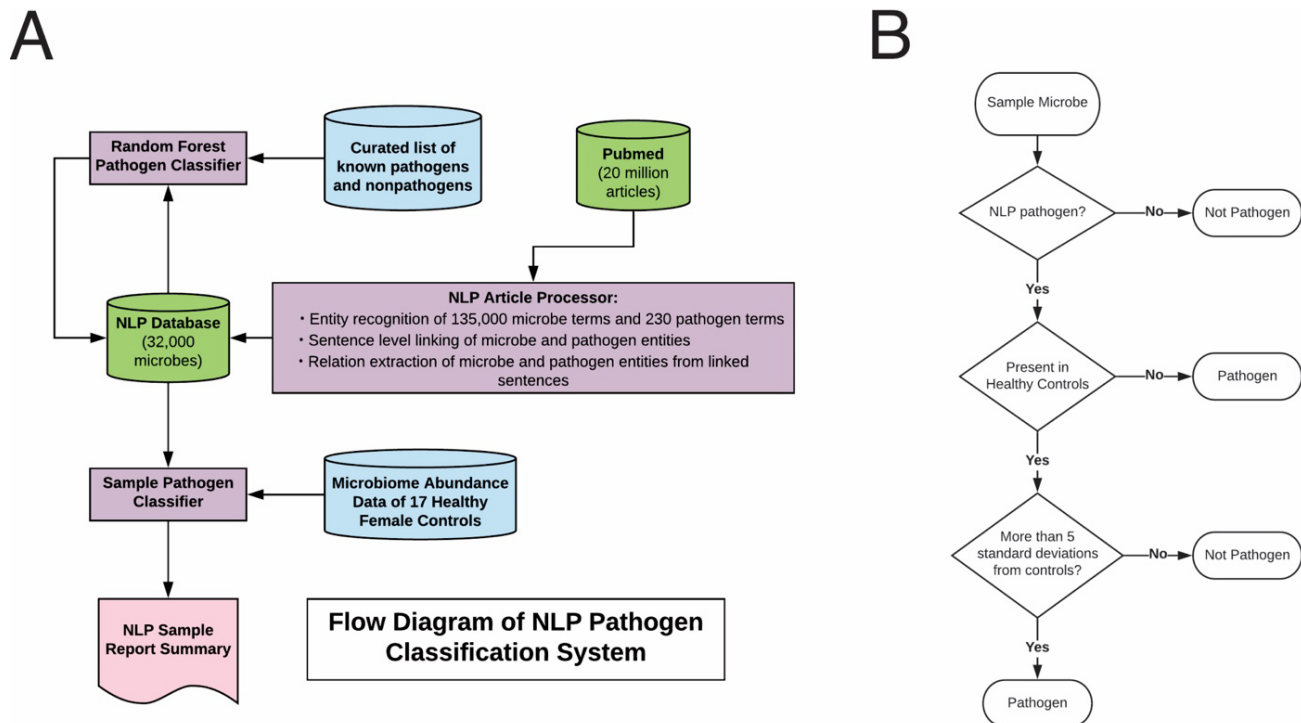


Figure 1: Flow diagram of NLP processing and curation. A. The structure of the database which started from a curated list of highly cited uropathogens compared with the abstracts of 20 million articles processed by NLP parsers for pathogenic clinical context. B. The general workflow used by NLP in binning microorganisms into a pathogen status.

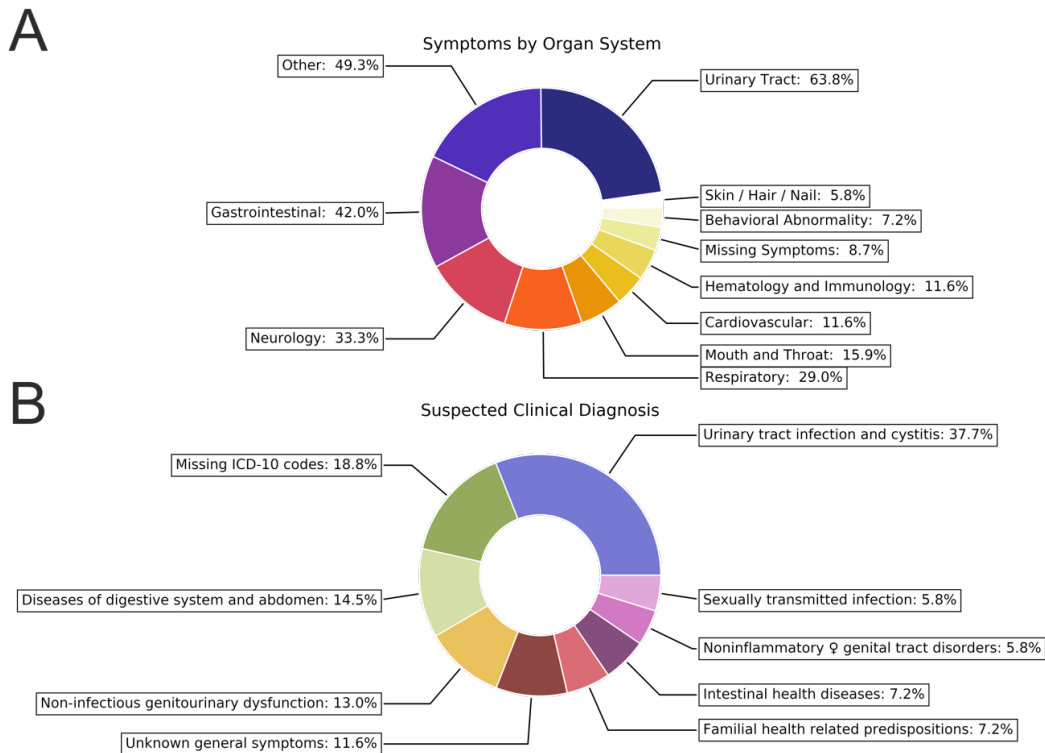


Figure 2: Clinical symptoms and diagnoses based on clinical suspicion using ICD-10 codes. A. The most common clinical symptoms grouped by organ system shown as the percentage of patients (n=67) with a cutoff of >5%. B. The most common suspected clinical diagnoses based on ICD-10 codes shown as the percentage of total patients (n=67) with a cutoff of >5%.

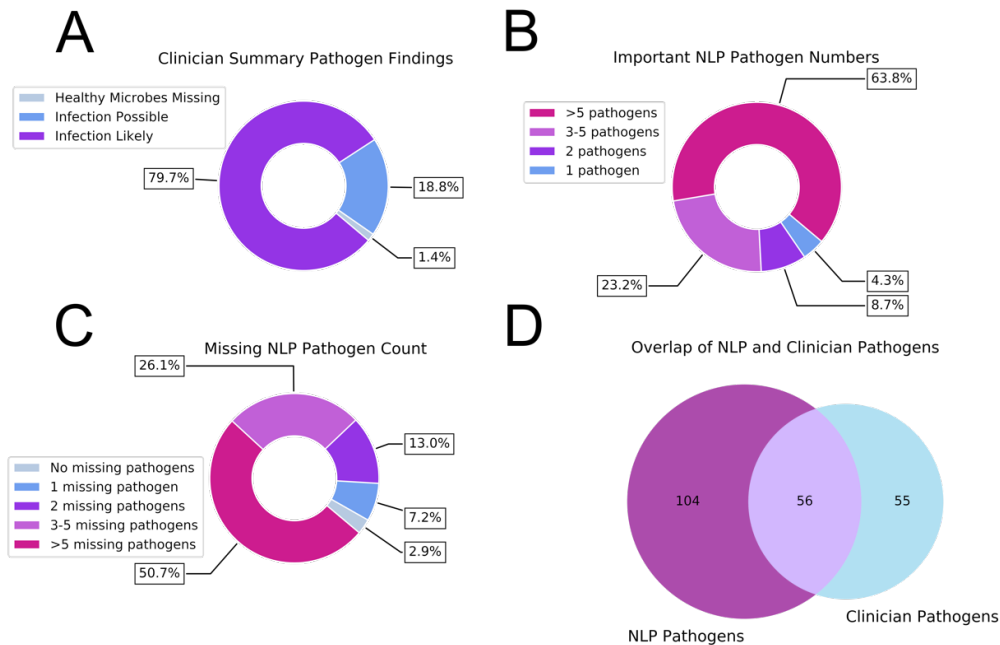


Figure 3: A). A graph showed the three categories of the clinical summaries. Infection was likely in samples with well-known uropathogens, it was possible in samples with less known uropathogens, and only 1.4% were in neither of these categories but had unusual urinary microbe abundances. B). This graph showed the percentage of the 69 patient samples that had at least n pathogenic species in them with either abundances more than 3 standard deviations from controls or no healthy control reference. C). This graph showed the number of species found important by the NLP pipeline that were missing from the clinical reviews. It should be noted that reviewers often left out species to be less important for brevity. D). Showed the number of pathogenic species cited in each group exclusively as well as the overlap species that both groups evaluated as pathogenic.

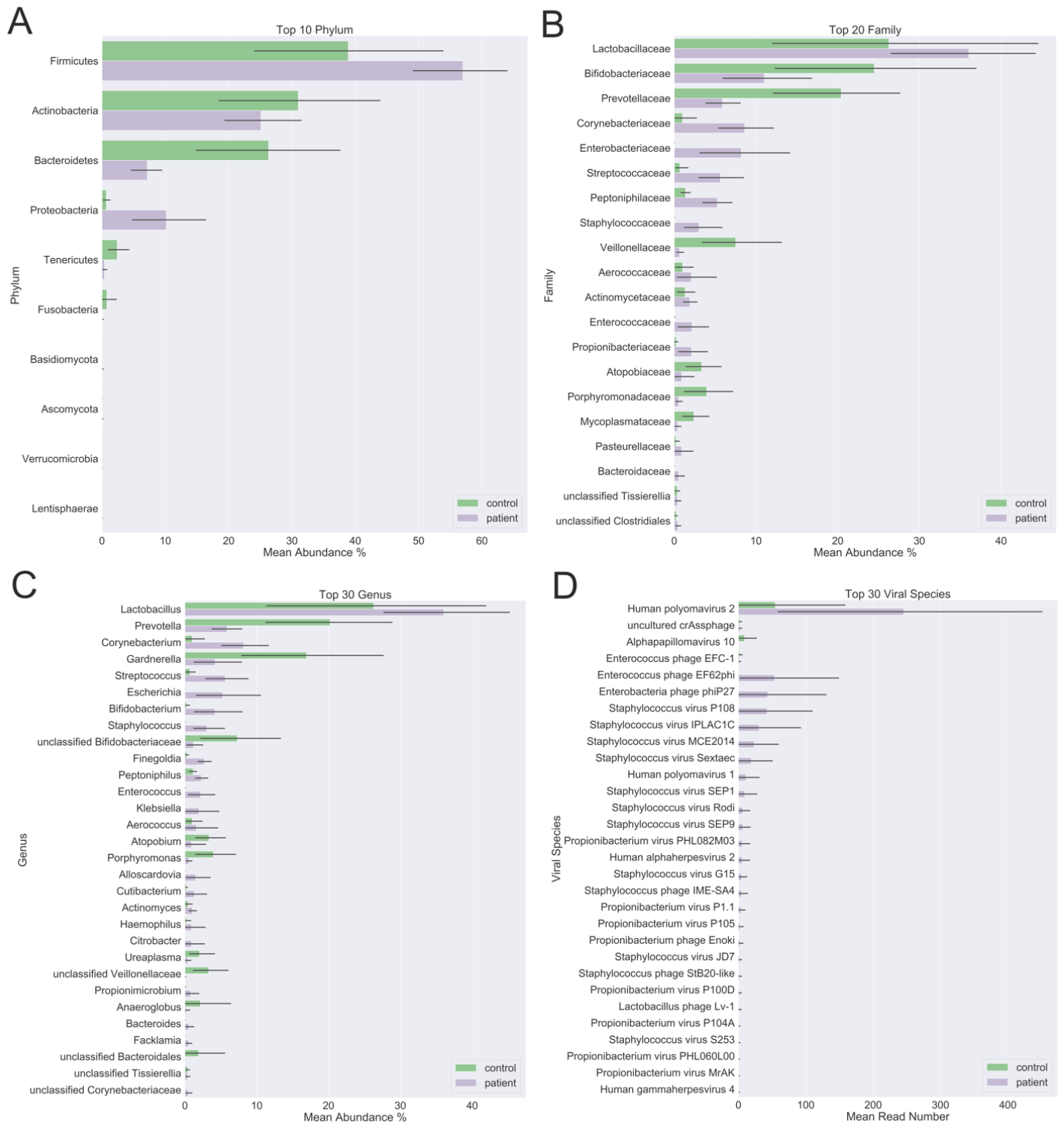


Figure 4: A). Top 10 phylum indicated that the primary difference between controls and patients was due to Bacteroidetes and Proteobacteria. B). The top families brought higher resolution towards the phyla results indicating Staphylococcaceae, Streptococcaceae, Corynebacteriaceae, and various proteobacteria related bacteria as being particularly different between groups. C). The genera continued to highlight various Corynebacteria, proteobacteria, Staphylococcus spp., and Streptococcus spp. as being common in patients. D). A variety of viruses with a human host including the JC and BK polyomaviruses, papillomaviruses, and herpesviruses were observed. Various phages were present known. Graphs A-C excluded viruses to not skew Eukaryotic and Bacterial taxa abundances. The lines indicate the range of a <95% confidence interval.

relatively distinct group of microbes with an overlap of only 45 out of 216 pathogens (Figure 3D).

After a comparison the urine microbiomes from the patient and healthy populations, we plotted the representative microbe(s) means, from phyla to viral species, and their corresponding >95% confidence intervals on Figure 4. The confidence intervals showed proteobacteria as being associated with patients, and Bacteroidetes as being associated with controls. Consistently, the same grouping was shown at specific taxa such as family and genus levels (Figures 4B and 3C) These trends resulted in several taxa, such as *Prevotella*, being associated to controls and common uropathogens such as *Streptococcus*, *Staphylococcus*, and *Escherichia* being associated to patients. Almost all viruses seemed to be associated with patients as shown on Figure 4D.

Of the clinical reviewer highlighted pathogens not found as important by NLP, 40 were not classified as pathogens by the pipeline, 13 were not found in the NLP database, due to the limitation of the commercial PUBMED section, and five were classified as pathogens but their abundance was under the threshold of three standard deviations above standard healthy control means. Various *Lactobacillus* species, nine in total, were not identified as pathogens by the NLP pipeline, but were identified as being possibly or likely pathogenic by the clinical reviewers noting their sample abundance as being significantly higher than healthy controls.

Discussion

In this study, we conducted microbiome analysis of 86 female urine samples, that 17 healthy controls and 69 patients, obtained the diagnostic yield and evaluated a manual curation process, with clinical reviewers, and compared it to NLP-based analysis for the inclusion of relevant urinary tract pathogens on clinical reports (Figure 1). The focus was on female samples as multiple microbiome phenotypes have been well characterized in peer-reviewed literature. Furthermore, urinary analyses were relatively simple as a sample type compared to fecal or oral microbiomes with a complexity on average being significantly lower. Moreover, existing literature was more easily available for the urinary tract and infection status. NLP was used for creation of a pathogen identification tool, Xplore-AI, to assess the potential pathogens in all the samples. Report summaries that were written by infectious disease experts were used to compare with the NLP result to investigate its accuracy. The results showed that the NLP system reported 97% of patient samples had at least one pathogen over three standard deviations from values found in healthy controls. Similarly, 84% of patients had two or more classified pathogens. These diagnostic percentages were consistent with the infectious disease expert summaries. However, there were a number of uropathogens that only the NLP system identified. We demonstrated a high diagnostic yield in females with urological symptoms following metagenomic analysis and that NLP improved the accuracy of the reportable species.

The demographics of the female patient population was in the middle age range reflecting a higher prevalence of UTIs

women over the age of 40 while demonstrating that UTIs can occur in women as young as 11. The cohort clinical information unsurprisingly demonstrated that a majority of the patients report some kind of urinary discomfort (Figure 2). Specifically, the ICD-10 codes for the traditional N30-39 urinary tract associated disorders were less than 50% of patients which can perhaps be explained by the fact that there are many different ICD-10 codes and many seem to describe the same organ system area. The gastrointestinal symptoms and suspected diagnoses seemed to support existing literature implicating the gut's interrelatedness to seemingly separate organ system maladies [15,19,20].

Sequencing technologies are needed to permit the promise of personalized medicine to elucidate clinically actionable findings in urology. This study demonstrated the promising combination of two technologies, metagenomic sequencing and NLP, to assist urologists in obtaining a pathogen profile in patients with suspected UTIs. Metagenomic sequencing is the most extensive microbial detection method. The number of species and several phyla within our analysis strongly supported the position of the superior breadth of metagenomics. For example, species within the bacteroidetes and proteobacteria phyla were identified in patients with UTIs (Figure 4A). Similarly, species in several families and genera were significantly found in patients with UTIs, namely, prevotellaceae and *Prevotella*, respectively. Our analysis demonstrated the identification of less commonly tested for viruses including the human polyomavirus 1 and human polyomavirus 2 and supports a breadth of proteobacteria as being strongly associated with patients experiencing urinary tract discomfort and allodynia as indicated by past investigations into interstitial cystitis (Figure 4, Larsen et al. 2021, In Press). Moreover, various viruses such as certain *polyomaviruses*, *herpesviruses*, and *papillomaviruses* were also more likely to be found in IC/BPS patients. The powerful combination of utilization patient specific bacterial, fungal, and viral microbiome data supplemented by NLP curated literature and normal ranges provided the necessary information to identify key pathogens that may be responsible for the infection. Even though we have shown the application of this technology in urology, a similar approach may be used for other disciplines of medicine.

The types of microbes in this study supported a strong association between patients experiencing urinary tract discomfort, allodynia and proteobacteria as previously reported in patients with interstitial cystitis as well as the utilization of a clinical reviewer and NLP enhanced the species that were included on the clinical reports. In addition, viruses such as certain *polyomaviruses*, *herpesviruses*, and *papillomaviruses* were also found in IC/BPS patients (Figure 4). A small fraction of the organisms identified by both the reviewers and NLP reinforce the most widely known uropathogens such as *Candida spp.*, *Pseudomonas aeruginosa*, *Klebsiella spp.*, *Enterococcus spp.*, *Proteus spp.*, and *Herpes simplex virus spp* [3]. Through metagenomics sequencing the number of species tested per sample greatly exceeded any other culture or RT-PCR-based UTI testing and had the potential to detect any one of the over 40,000, as of November 2019, curated sequenced genomes. Lesser known uropathogen species were only detected by NLP. As an

example of how the system can enhance the clinician findings, several microbes that the NLP pipeline found to be important were not mentioned in any of the clinician summaries. Included in these microbes were *Aerococcus urinae*, *Actinotignum schaalii*, and *Actinobaculum massiliense* [25,26]. All of these have garnered recent attention for causing UTIs, but *Aerococcus urinae* especially has become a notable uropathogen and was present in 13 of the 69 samples. The NLP pipeline was useful because it was difficult for clinical reviewers to both stay current with and remember all new literature associated with pathogenic microbes. The system was capable of mining the medical literature on a regular basis and updating a database that was digestible to the clinical reviewers for more accurate diagnostics, and paved the way for automated, machine learning-based diagnostics. This demonstrated how NLP can aid in the inclusion of lesser-known species more efficiently compared to reviewers. Infectious disease experts were essential in interpreting the results due to their background with the urinary microbiome. Though beneficial, there are clear limitations in an individual attempting to specialize in every species while keeping up to date on all the literature and normal ranges surrounding every microorganism. This knowledge gap is exacerbated when the species list is in the 50+ range on the clinical reports and literature is constantly changing. The comparison of the expert summaries to NLP species of interest showed the shortcomings of both systems on their own and why a synergistic relationship is necessary for enhanced accuracy as well as proper scaling for the anticipated widespread adoption of this technology.

While shotgun metagenomic technology is recent, research initiatives such as the Human Microbiome Project have already demonstrated clinical implications [21,22]. While our previous study looked at the clinical utility of using shotgun metagenomics in UTI diagnosis, the novelty of this study lies in a state-of-the-art NLP pipeline tailored towards tying together clinical symptoms, diseases, and body location to species and genera (Figure 1, Figure 4). Though NLP was initially established as a tool for clinical usage to aid in reaching a human genetic diagnosis more comprehensively. In contrast to human applications, the microbiome curated databases such as Disbiome are focused on experimental designs and are still under development, far from the mature human genetics tools such as OMIM [23,24]. Thus, there is a need to create comprehensive microbiome databases that link symptomology to specific pathogens.

The NLP pipeline was a very useful system, but similar to new technologies and methods, it had a number of limitations and ways to overcome them. These potential shortcomings fall into roughly two categories: the quality of the information being used, and the system being used to interpret it. The first exists because the NLP system took all literature as it currently exists, disregarding the possibility of incorrect information. The second exists due to the inherent complexity of the human language and grammatical idiosyncrasies of different authors. For example, the method used for segmenting sentences was not perfect and was not guaranteed to isolate all links where a microbe and an entity existed in the same sentence, especially when a sentence

was heavy with non-terminating periods. Another example was in the use of dependency parsing, which relied on the accuracy of the parser, the usefulness of the curated parse patterns being used, and the exhaustive nature of the parse patterns. Thus, the most obvious shortcoming for the NLP pipeline was the possibility of useful data not being included in the interpretation. A step towards overcoming these limitations is to urge the field to create better methods that address parsing and sentence structure interpretation.

In summary, this study demonstrated a high diagnostic yield after clinical reviewer metagenomics analyses for patients with urological symptoms. Consistent with the clinical reviewer findings, the NLP system reported 97% and 84% of patient samples had at least one and two pathogen(s), respectively compared to infectious disease expert summaries. However, in a number of cases NLP provided better insights as to the presence of rare known uropathogens. We emphasize the use of clinical reviewers in conjunction with an NLP system to accurately report pathogen species on clinical reports.

References

1. Rafael Lozano, Mohsen Naghavi, Kyle Foreman, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012; 380: 2095-2128.
2. Schlaberg R, Chiu CY, Miller S, et al. Validation of metagenomic next-generation sequencing tests for universal pathogen detection. *Arch Pathol Lab Med*. 2017; 141: 776-786.
3. Flores-Mireles AL, Walker JN, Caparon M, et al. Urinary tract infections: Epidemiology, mechanisms of infection and treatment options. *Nat Rev Microbiol*. 2015; 13: 269-284.
4. Gaitonde S, Malik RD, Zimmern PE. Financial Burden of Recurrent Urinary Tract Infections in Women: A Time-driven Activity-based Cost Analysis. *Urology*. 2019; 128: 47-54.
5. Evans MM. Urinary tract infection in older adults. *Nursing (Lond)*. 2012; 42: 72
6. Malik SA, Malik AS, Malik SA. A review on common bacteria causing urinary tract infection. *Pakistan J Med Heal Sci*. 2016; 10: 513-517.
7. Soiza RL, Donaldson AIC Myint PK. Vaccine against arteriosclerosis: an update. *Therapeutic Advances in Drug Safety*. 2018; 9: 259-261.
8. Goldberg B, Sichtig H, Geyer C. et al. Making the leap from research laboratory to clinic: Challenges and opportunities for next-generation sequencing in infectious disease diagnostics. *MBio*. 2015; 6: 1-10.
9. Erin H Graf, Keith E Simmon, Keith D Tardif, et al. Unbiased detection of respiratory viruses by use of RNA sequencing-based metagenomics: A systematic comparison to a commercial PCR panel. *J Clin Microbiol*. 2016; 54: 1000-1007.
10. Parize P, Muth E, Richaud C, et al. Untargeted next-generation sequencing-based first-line diagnosis of infection in immunocompromised adults: a multicentre, blinded, prospective study. *Clin Microbiol Infect*. 2017; 23: 574.e1-574.e6.

11. Paul Turner, Kuong Suy, Le Van Tan, et al. et al. The aetiologies of central nervous system infections in hospitalised Cambodian children. *BMC Infectious Diseases*. 2017; 17: 1-9.
12. Stefano Leo, Nadia Gaia, Etienne Ruppé, et al. Detection of Bacterial Pathogens from Broncho-Alveolar Lavage by Next-Generation Sequencing. *Int J Mol Sci*. 2017; 18: 1-13.
13. Gao Z, Kang Y, Yu J, et al. Human Pharyngeal Microbiome May Play A Protective Role in Respiratory Tract Infections. *Genomics Proteomics Bioinforma*. 2014; 12: 144-150.
14. Anna Krueger, Stéphanie Val, Marcos Pérez-Losada, et al. Relationship of the Middle Ear Effusion Microbiome to Secretory Mucin Production in Pediatric Patients with Chronic Otitis Media. *Pediatr Infect Dis J*. 2017; 36: 635-640.
15. Yige Bao, Kait F Al, Ryan M Chanyi, et al. Questions and challenges associated with studying the microbiome of the urinary tract. *Ann Transl Med*. 2017; 5: 1-8.
16. Rupinder Kaur Brara, Shaijib PS, Sahil Sehgal et al. Testing for weak D Antigen: Spectrum and its applied role in rhesus-negative transfusions in Andaman and Nicobar Islands. *Tzu Chi Med J*. 2019; 32: 167-170.
17. Varsha Dave Badal, Dustin Wright, Yannis Katsis, et al. Challenges in the construction of knowledge bases for human microbiome-disease associations. *Microbiome*. 2019; 7: 1-15.
18. Choi JD, Tetreault J, Stent A. It Depends: Dependency Parser Comparison Using A Web-based Evaluation Tool. <https://www.aclweb.org/anthology/P15-1038.pdf> 2015.
19. Bø K, Hilde G. Retest reliability of surface electromyography on the pelvic floor muscles. *HH Dederig A, Jonasson AF. Neurourol Urodyn*. 2013; 32: 215-223.
20. J Quentin Clemens, Chris Mullins, John W Kusek, et al. The MAPP research network: A novel study of urologic chronic pelvic pain syndromes. *BMC Urol*. 2014; 14: 1-6.
21. Lewis AL, Lewis WG. A New Catalog of Microbiological Tools for Women's Infectious Disease Research. *Genome Announc*. 2016; 4: 889-891.
22. Lloyd-Price, Anup Mahurkar, Gholamali Rahnavard, et al. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature*. 2017; 550: 61-66.
23. Amberger JS, Bocchini CA, Schiettecatte F, et al. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an Online catalog of human genes and genetic disorders. *Nucleic Acids Res*. 2015; 43: D789-D798.
24. Yorick Janssens, Joachim Nielandt, Antoon Bronselaer, et al. Disbiome database: Linking the microbiome to disease. *BMC Microbiol*. 2018; 18: 4-9.
25. Horton LE, Mehta SR, Aganovic L, et al. *Actinotignum schaalii* infection: A clandestine cause of sterile pyuria? *Open Forum Infect Dis*. 2018; 5: 10-12.
26. Chhibber AV, Muttaiyah S, Hill AA, et al. *Aerococcus urinae* Aortitis: A case report. *Open Forum Infect Dis*. 2019; 6: 4-7.