



APERIOMICS

Female Urinary Microbiome Analysis and Natural Language Processing (NLP) Enhances the Infectious Diagnostic Yield in Precision Medicine

David Baugher, Alexander Larsen, Yuan Chen, Crystal R Icenhour, C Alexander Valencia
Aperiomics Inc. 45925 Horseshoe Dr., Suite 170, Sterling, VA 20166

Key Points

- The results showed that the NLP system reported 97% of patient samples had one pathogen more than 3 standard deviations from the abundance of that pathogen in healthy controls, and least 84% had 2 or more pathogens.
- Shotgun metagenomics allows for the identification of a comprehensive number of microbes (37,000+) compared to current testing practices with applications. Many medical fields would benefit from understanding the pathogenesis of infections and dysbiosis.

Background

- The substantial clinical load of UTIs accounts for 1% of US clinical resources equating more than \$3.7 billion annually¹.
- The complexity of the urinary microbiome makes interpretation through solely human expertise difficult. Natural Language processing can help filter through literature quickly².
- Non-infectious urinary tract issues are likely over-diagnosed due to the narrow focus and known limits of detection using culture-based, PCR, antigen, antibody, and more recently 16S methods³.
- Shotgun metagenomics permits the identification of both unculturable and culturable bacteria, viruses, and fungi as well as markers in an altered microbiome.^{4,5}



Healthy

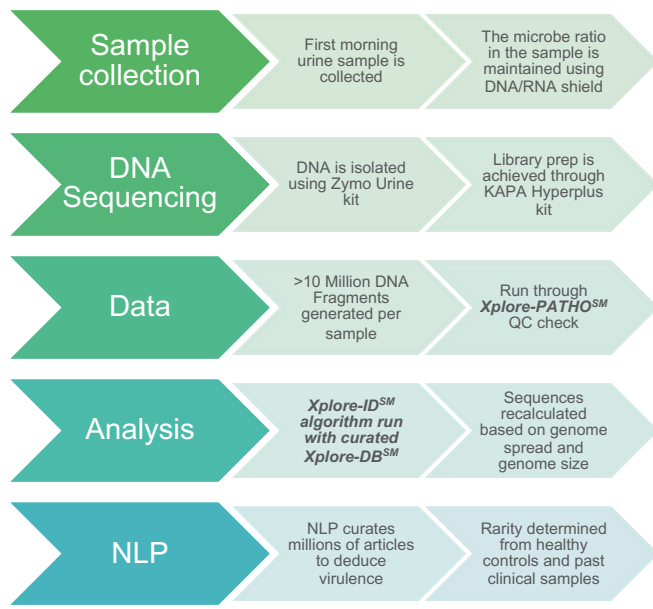
- Lactobacillus*
- Corynebacterium*
- Gardnerella*
- Prevotella*
- Bifidobacterium*
- Staphylococcus epididymitis*



Common UTI

- E. Coli*
- Klebsiella*
- Enterobacter*
- Proteus*
- Pseudomonas*
- Staphylococcus aureus*
- Candida spp.*
- Enterococcus spp.*
- Morganella*

Methodology



Results

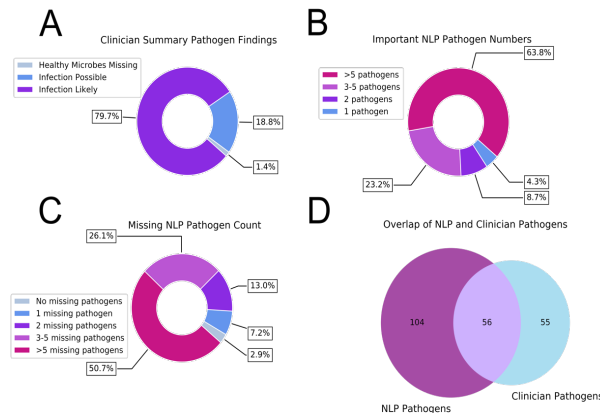


Figure 1: (A) Graph showing the three categories into which clinical summaries fell. Infection was likely in samples with well-known uropathogens, it was possible in samples with less known uropathogens, and only 1.4% were in neither of these categories but had unusual urinary microbe abundances. (B) This graph shows the percentage of the 69 patient samples that had at least one pathogenic species in them with either abundances more than 3 standard deviations from controls or no healthy control reference. (C) This graph shows the number of species found important by the NLP pipeline that were missing from the clinical reviews. It should be noted that reviewers often leave out species that are less important and/or abundant for brevity. (D) Shows the number of pathogenic species cited in each group exclusively as well as the overlap species that both groups evaluated as pathogenic.

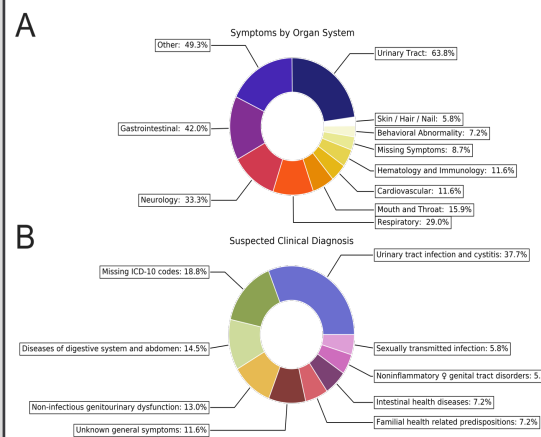


Figure 2. Clinical Symptoms and Diagnoses based on self-reporting and ICD-10 codes. A. The most common clinical symptoms grouped by organ system shown as the percentage of patients (n=67) with a cutoff of >5%. B. The most common suspected clinical diagnoses based on ICD-10 codes shown as the percentage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

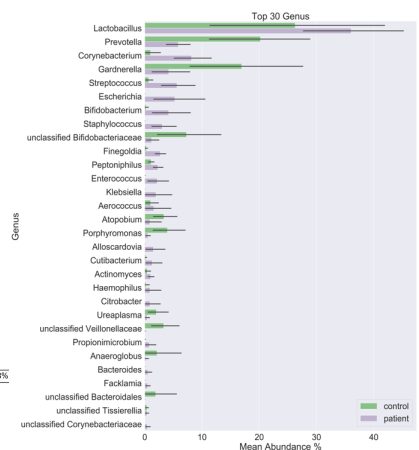


Figure 3. The genera continue to highlight various *Corynebacterium*, *proteobacteria**, *Staphylococcus* spp., and *Streptococcus* spp. as common in patients.

**proteobacteria* encompass *E. coli*, *Klebsiella*, *Citrobacter*, *Enterococcus*, etc.

Company Information

Aperiomics is a bioinformatics software company based in Sterling, VA. Using the **Xplore-PATHOSM** platform it has helped a variety of clinicians, researchers, and patients identify a variety of microbes and pathogens across sample types ranging from swab, urine, fecal and everything in-between. The company and platform were first created in 2014 with the help of leading experts across microbiology, genomics, and bioinformatics.

Citations

- Flores-Mireles, A. L., Walker, J. N., Caparon, M. & Hultgren, S. J. Urinary tract infections: Epidemiology, mechanisms of infection and treatment options. *Nat. Rev. Microbiol.* 13, 269–284 (2015).
- Cai, T. et al. Natural language processing technologies in radiology research and clinical applications. *Radiographics* 36, 176–191 (2016).
- Ranjan, R., Rani, A., Metwally, A., McGee, H. S. & Perkins, D. L. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem. Biophys. Res. Commun.* 469, 967–977 (2016).
- Caliendo, A. M. et al. Better Tests, Better Care: Improved Diagnostics for Infectious Diseases. *Clin. Infect. Dis.* 57, S139–S170 (2013).
- Breitwieser, F. P., Lu, J. & Salzberg, S. L. A review of methods and databases for metagenomic classification and assembly. *Brief. Bioinform.* (2017).