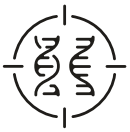
illumina®

Track emerging infectious diseases with wastewater surveillance

Broad detection of pathogens with the MiSeq™ i100 Series

Comprehensive workflow includes library preparation, target enrichment, sequencing, and data analysis



Accurate detection and characterization of a wide range of RNA and DNA viral pathogens

Fast, flexible sequencing that delivers same-day results for effective and efficient wastewater surveillance

M-GL-02915 v1.0

Introduction

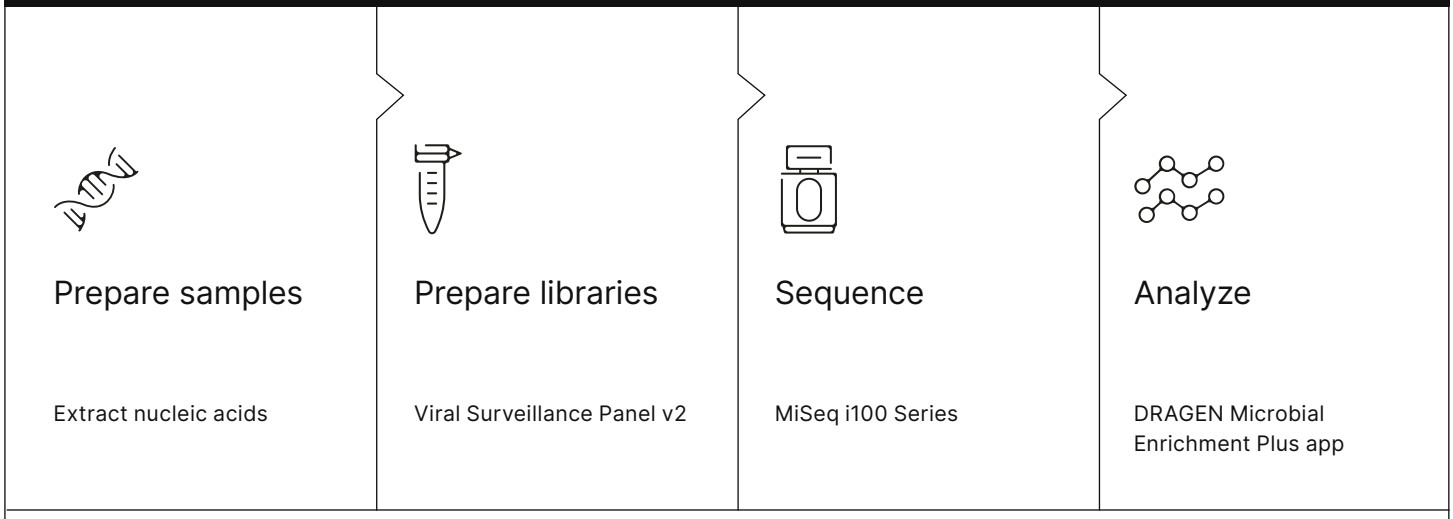
The modern world is witnessing an alarming increase in the threat posed by infectious diseases. Globalization and international travel facilitate the rapid spread of infectious agents across borders.[1] Additionally, expanding urbanization and growing population density create ideal conditions for the transmission of infectious agents. Moreover, the emergence of drug-resistant strains of bacteria and viruses poses a significant challenge in the treatment and control of infections.[2]

Wastewater surveillance is a method for detecting, identifying, tracking, and characterizing potential pathogens to assess the health of a population.[3] This method provides data to help monitor outbreaks and serves as an early warning of infectious threats at the community level, as demonstrated through genomic surveillance of SARS-CoV-2.[4] By knowing where these threats are, communities can better allocate resources during a public health response. PCR based–methods can provide rapid and relatively low-cost snapshots of the presence or absence of expected pathogens, but their performance for monitoring and detection depends on preexisting knowledge of microorganism sequence variability and of variant determining mutations. Next-generation sequencing (NGS) workflows are relatively more expensive in time and cost, but are more tolerant to mutation, enable variant discovery, and can provide detailed genomic information beyond qualitative detection of pathogens.[5–7]

Accurate and comprehensive detection of viral pathogens in wastewater depends on both the upfront enrichment of viral genomes by filtration or concentration techniques and NGS library prep enrichment methods to overcome the analytical challenges inherent to studying the relatively small genomes of viruses, the most abundant of which make up only a small proportion of the total genomic material in wastewater.[8–10] By increasing the relative abundance of genomic content of interest in the library, enrichment also unlocks the potential for sequencing on benchtop instruments.

This application note demonstrates detection and characterization of viral pathogens in real-world wastewater samples using an NGS workflow that integrates the Illumina Viral Surveillance Panel v2, the MiSeq i100 Series, and onboard DRAGEN™ secondary analysis (Figure 1). The MiSeq i100 Plus System delivers same-day results for efficient wastewater surveillance to enable a rapid public health response.

Figure 1: Comprehensive NGS workflow for wastewater surveillance



Combine Illumina library preparation with fast, simple sequencing on the MiSeq i100 Series and DRAGEN secondary analysis for accurate and efficient wastewater surveillance.

Methods

Samples

Raw wastewater samples were collected from wastewater treatment plants by Wisconsin State Laboratory of Hygiene (WSLH) (n = 12) and from student dormitories by Colorado State University (CSU) (n = 12); both sites are located in the United States. Samples were collected from each site over multiple time points from November 4, 2022 to December 16, 2022. 10–50 ml of sample wastewater was prepared by WSLH by capture and concentration of viruses with Nanotrap Microbiome A Particles (Ceres Nanosciences, Inc., Catalog no. 44202). Nucleic acids were extracted using the Wizard Enviro Total Nucleic Acid Kit (Promega Corporation, Catalog no. A2991). Samples prepared by CSU involved removal of solids via centrifugation at ~2000 × g, followed by capture and concentration of viruses with the CP Select Concentrating Pipette (InnovaPrep, Inc.). Nucleic acids were extracted using the QIAamp Viral RNA Mini Kit (QIAGEN, Catalog no. 52904).

Library preparation

Sequencing-ready libraries were prepared from a maximum volume of 8.5 µl (≥ 100 ng of extracted total nucleic acid (TNA)) input using the Viral Surveillance Panel v2 Kit, Set A (96 samples) (Illumina, Catalog no. 20108081). Libraries were combined into four pools of six libraries each (6-plex) with a loading concentration of 80 pM for sequencing.

Sequencing

Prepared libraries were sequenced on the MiSeq i100 Plus System using a 25M flow cell with a run configuration of 2 × 150 bp. For larger studies, sequencing runs can be scaled up to the NextSeq™ 1000, NextSeq 2000, NovaSeq™ 6000, and NovaSeq X Systems.

Data analysis

After sequencing was complete, data were downsampled to 1M and 4M clusters/fragments per sample using the FASTQ Toolkit App. Downsampled and nondownsampled data was analyzed using the DRAGEN Microbial Enrichment Plus (DME+) app onboard the MiSeq i100 Plus System. The app can also be accessed in the cloud in BaseSpace™ Sequence Hub.

Results

Sequencing metrics

The four wastewater pools were sequenced across four runs on the MiSeq i100 Plus System. All four runs resulted in an average % Q30 greater than 90% and percent reads passing filter (PF) of ~80%, indicating both high-quality reads and consistent instrument loading concentrations. The total number of paired-end (PE) reads obtained exceeds the 50M specification of the flow cell. For all four runs, the combined instrument run time and onboard analysis times were under eight hours (Table 1). This demonstrates that the MiSeq i100 Plus System offers speed and efficiency necessary to provide timely results important for public health monitoring and response.

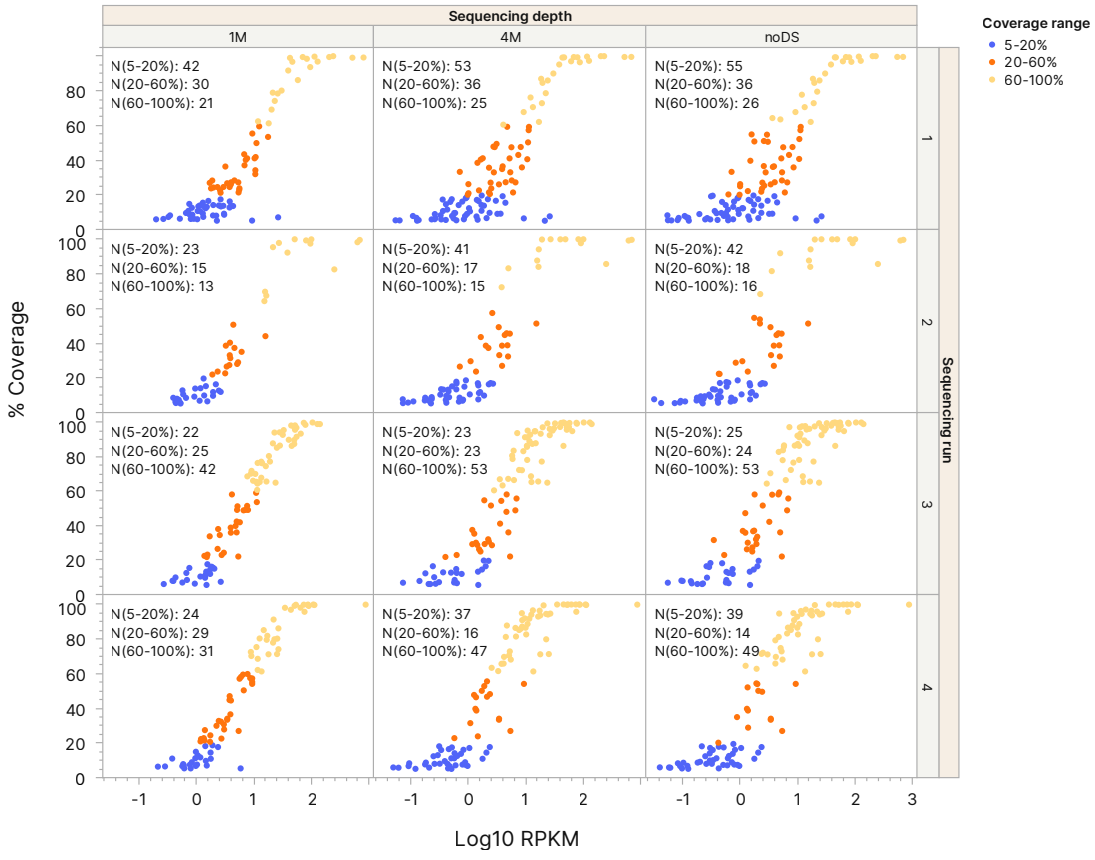
Table 1: Sequencing metrics for the MiSeq i100 Series

MiSeq i100 Plus run	Average % Q30	% PF	Total no. of paired-end reads	Total no. of paired-end reads PF	Run time	Onboard analysis time
Run 1 (CSU)	93.65%	78.49%	79,073,280	62,064,116	7 hr 11 min	18 min
Run 2 (CSU)	93.35%	78.06%	79,073,280	61,723,176	7 hr 12 min	17 min
Run 3 (WSLH)	94.57%	81.62%	79,073,280	64,539,830	7 hr 12 min	19 min
Run 4 (WSLH)	94.38%	81.43%	79,073,280	64,389,944	7 hr 12 min	19 min

Detection of viral genomes

To help establish and compare the amount of sequencing depth needed for effective pathogen detection, analysis of the reads was simulated at various depths. In addition, data normalization was needed for accurate quantification due to the large variation in length of the detected viral genomes. Therefore, the percent of the genomes covered was plotted against the log10 RPKM (Reads Per Kilobase per Million) reads across the four wastewater sample library pools and simulated at three sequencing depths: downsampled to 1M fragments, downsampled to 4M fragments, and full depth (not downsampled or ~5M fragments per sample). RPKM is a common means to normalize NGS data that combines the depth of coverage with the length of the target region. This permits more accurate comparisons across different pathogens, as well as for a single pathogen over time. Coverage of the detected viruses can be segregated into three general ranges, detection (5–20%), surveillance (20–60%), and full genome coverage (60–100%) (Figure 2). Viral pathogens in the lowest range were confidently detected, but lacked sufficient coverage to track viral evolution or accurate variant information. Viral genomes in the mid-range had sufficient coverage for effective molecular characterization and changes in viral concentrations over time. Genomes detected in the highest range were able to be characterized with variant determination. At greater read depths, the total number of viral genomes was higher, but plateaued at around 4M clusters/fragments (8M PE reads) (Figure 2 and Table 2).

Figure 2: Varying genome coverage for viral pathogens detected in wastewater samples



Percent genome coverage was plotted against log10 RPKM over four MiSeq i100 Plus sequencing runs. Data were downsampled to 4M and 1M reads or not downsampled. Genome coverage is segregated into three ranges: detection range (5–20%), surveillance range (20–60%), and full genome coverage range (60–100%). N represents the number of viral genomes detected in each range. Even at a simulated depth of 1M reads per library, coverage of ≥ 60% was achieved for ≥ 22% of detected viral targets in each of the four runs.

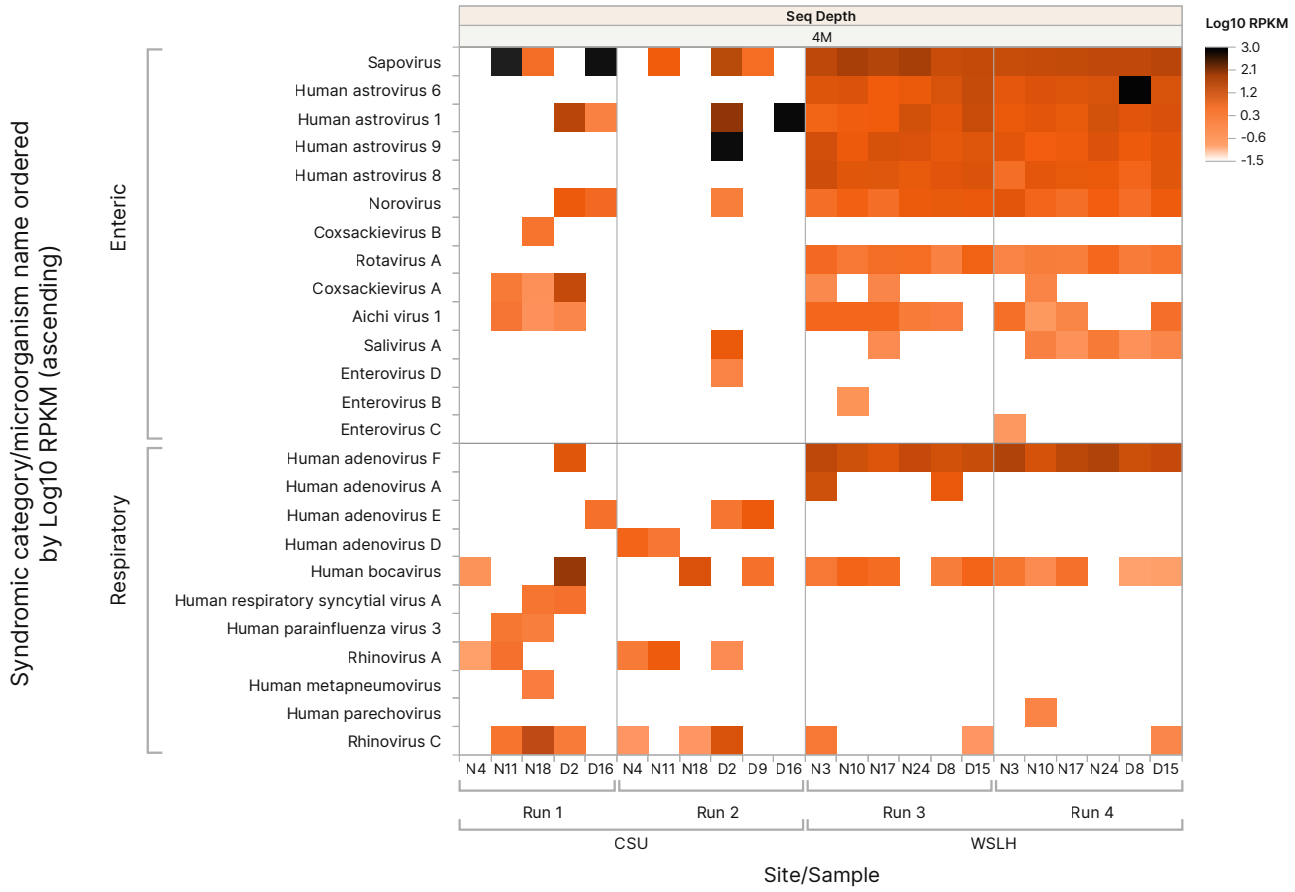
Table 2: Total number of viral genomes detected by site

Site	Genome coverage	Sequencing depth		
		1M reads	4M reads	Full depth
CSU	5–20%	65	94	97
	20–60%	45	53	54
	60–100%	34	40	92
WSLH	5–20%	46	60	64
	20–60%	54	39	38
	60–100%	73	100	102

Characterization of viral genomes

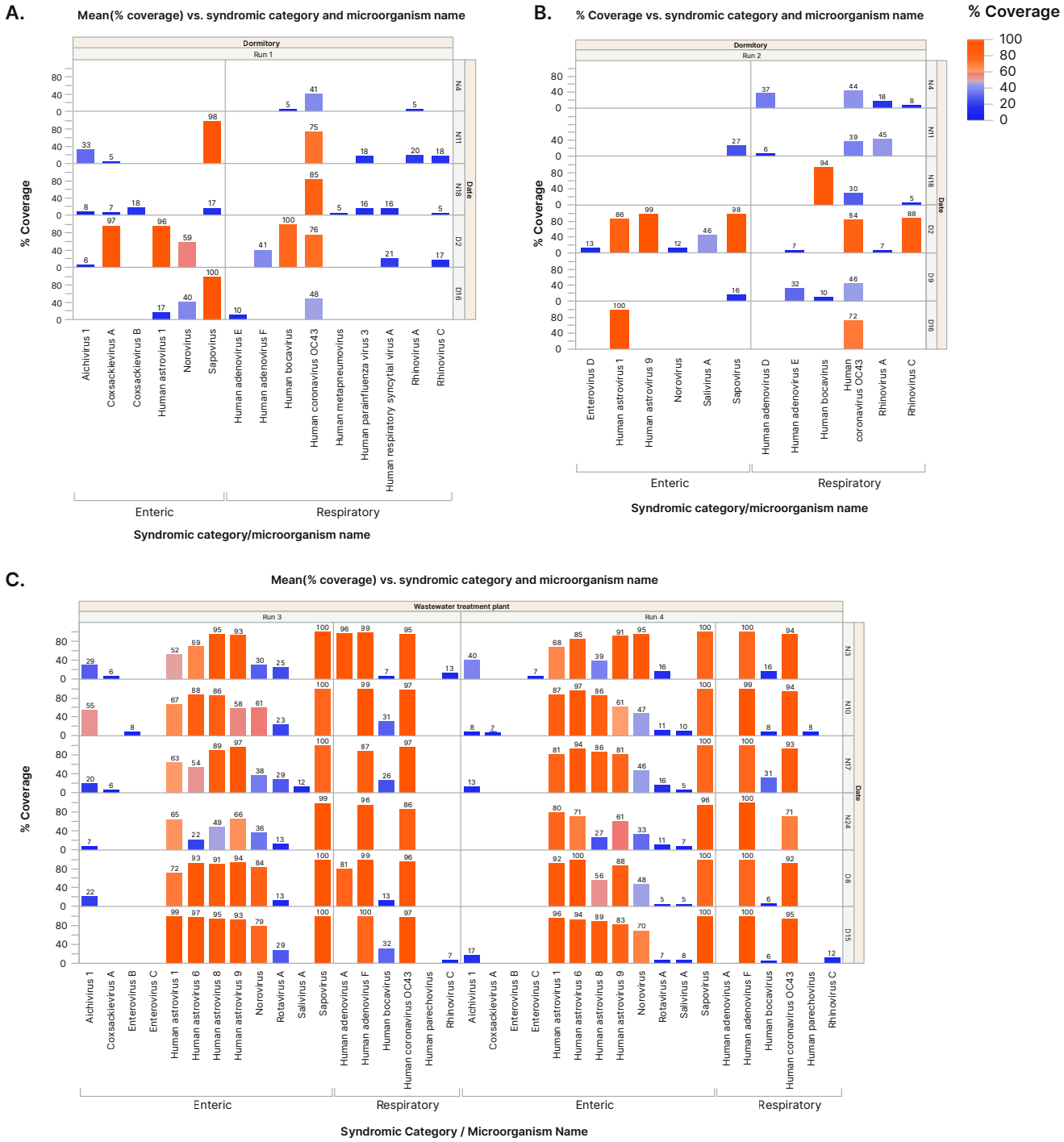
As part of the longitudinal study, detected viral genomes were plotted and arranged with ascending log10 RPKM values in a heat map that showed various levels of detection for enteric and respiratory viruses across collection sites and over time, as measured by relative abundance (Figure 3) and by genome coverage (Figure 4). Plotting genome coverage over time showed increased levels of enteric viruses, including Coxsackie virus A/B, Norovirus, Mamastrovirus (human astrovirus), and Salivirus in December time points (Figure 4), consistent with previous reports detecting high abundance of these and other pathogenic viruses shed in human stool.[11]

Figure 3: Heat map of viral abundance in wastewater samples



Viral genomes detected from CSU and WSLH samples are arranged with ascending Log10 RPKM values. Viruses corresponding to enteric and respiratory categories are shown in this heat map. WSLH samples (N = 145) showed higher viral genome detection compared to CSU samples (N = 63). Due to the relatively lower number of people contributing to the dorm wastewater samples, fewer pathogen detections are observed compared to samples from the treatment plants. Collection dates in November (N) and December (D) are indicated above each collection site.

Figure 4: Changes in viral pathogens detected over time



Monitoring of viral pathogen levels as part of a longitudinal study showed changes in enteric and respiratory viruses in samples collected from (A) both sites, (B) CSU, and (C) WSLH. Genome coverage is plotted against viruses from enteric and syndromic categories over time (dates indicated on the right side of plot).

Similarly, the respiratory viruses Bocavirus and Rhinovirus A/C showed elevated levels in samples collected in December (Figure 4). These data are consistent with historically observed seasonal increases in transmission of respiratory pathogens in early winter, perhaps accelerated by travel over the Thanksgiving holiday.[12]

Summary

The MiSeq i100 Series is part of a fast, comprehensive NGS workflow that enables broad detection of viral pathogens for effective wastewater surveillance as part of public health efforts.

Learn more →

Viral Surveillance Panel v2

MiSeq i100 Series

DRAGEN secondary analysis

References

1. CDC. Disease Patterns in Travelers. cdc.gov/travel/yellowbook/2024/introduction/disease-patterns-in-travelers. Updated May 1, 2023. Accessed February 5, 2025.
2. Reyes R, Ahn R, Thurber K, Burke TF. Urbanization and Infectious Diseases: General Principles, Historical Perspectives, and Contemporary Challenges. *Challenges in Infectious Diseases*. 2012;123-146. Published 2012 May 19. doi:10.1007/978-1-4614-4496-1_4.
3. CDC. Wastewater Monitoring—How Does It Work? cdc.gov/nwss/pdf/Wastewater-COVID-infographic-h.pdf. Accessed January 24, 2025.
4. Gupta P, Liao S, Ezekiel M, et al. Wastewater Genomic Surveillance Captures Early Detection of Omicron in Utah. *Microbiol Spectr*. 2023;11(3):e0039123. doi:10.1128/spectrum.00391-23.
5. US GAO. Science & Tech Spotlight: Wastewater Surveillance. gao.gov/products/gao-22-105841. Published April 11, 2022. Accessed January 24, 2025.
6. Diamond MB, Keshaviah A, Bento AI, et al. Wastewater surveillance of pathogens can inform public health responses. *Nat Med*. 2022; 28(10):1992-1995. doi:10.1038/s41591-022-01940-x
7. Crits-Christoph A, Kantor RS, Olm MR, et al. Genome Sequencing of Sewage Detects Regionally Prevalent SARS-CoV-2 Variants. *mBio*. 2021;12(1):e02703-20. Published 2021 Jan 19. doi:10.1128/mBio.02703-20.
8. Dueholm MKD, Nierychlo M, Andersen KS, et al. MiDAS 4: A global catalogue of full-length 16S rRNA gene sequences and taxonomy for studies of bacterial communities in wastewater treatment plants. *Nat Commun*. 2022;13(1):1908. Published 2022 Apr 7. doi:10.1038/s41467-022-29438-7.
9. Kantor RS, Jiang M. Considerations and Opportunities for Probe Capture Enrichment Sequencing of Emerging Viruses from Wastewater. *Environ Sci Technol*. 2024;58(19):8161-8168. doi:10.1021/acs.est.4c02638.
10. Cancela F, Lizasoain A, Panzera Y, et al. Targeted Enrichment Sequencing Utilizing a Respiratory Pathogen Panel for Genomic Wastewater-Based Viral Epidemiology in Uruguay. *Food Environ Virol*. 2025;17(1):14. Published 2025 Jan 9. doi:10.1007/s12560-024-09629-9.
11. Tisza M, Javornik Cregeen S, Avadhanula V, et al. Wastewater sequencing reveals community and variant dynamics of the collective human virome. *Nat Commun*. 2023;14(1):6878. Published 2023 Oct 28. doi:10.1038/s41467-023-42064-1.
12. CDC. Respiratory Illnesses Data Channel. cdc.gov/respiratory-viruses/data/index.html#cdc_data_surveillance_section_2-weekly-national-summary. Updated February 21, 2025. Accessed February 25, 2025.

illumina®

1.800.809.4566 toll-free (US) | +1.858.202.4566 tel
techsupport@illumina.com | www.illumina.com

M-GL-02915 v1.0