



Actionable Clinical Metagenomics with One Codex

Fast and scalable bioinformatics for building the next generation of infectious disease diagnostics.

PLATFORM FEATURES

One Codex Database

Consists of 148k microbial whole genomes including XXX pathogens.

- Bacteria
- Viruses
- Fungi
- Protists
- Archaea

Supported Sequencing Types

- Whole genome sequencing (WGS)
- Targeted enrichment (e.g., hybrid capture, amplicon)
- Long read sequencing (ONT, PacBio)

Supported NGS Platforms

- Illumina MiniSeq / MiSeq / NextSeq / NovaSeq
- PacBio RS II / Sequel
- Ion Torrent PGM / S5 / Proton
- Oxford Nanopore MinION / GridION / PromethION

POWERFUL TOOLS FOR CLINICAL METAGENOMICS

- Intuitive web-based interface
- REST API for simplified integration with LIMS and EHR systems
- Secure data sharing & management
- Rapid and accurate analysis
- Support for first party and custom workflows
- Version control for auditing, documenting
- Automated data formatting, import/export for public health reporting and sharing information in public databases

Motivation

Clinical metagenomics promises to revolutionize infectious disease diagnostics.^{1,2} Despite advances in lab automation and increased availability of sequencers in clinical laboratories, there are few clinical metagenomics solutions available that meet the diverse needs of modern clinical and diagnostic laboratories.³

One Codex is a fast, easy to use, and purpose-built analysis platform for clinical metagenomics supporting diagnostic testing in individual clinical laboratories.⁴ We've served as the backbone for clinical trials, clinical microbiome analysis, diagnostics, and public health surveillance.⁵⁻⁸ We offer the world's leading curated reference database of microbial genomes.⁹ Combined with our experience working in highly regulated industries and our expertise in microbial genomics, One Codex is poised to power the next-generation of clinical metagenomics workflows.

Clinical metagenomics testing has challenges that make it difficult to operationalize in diagnostic laboratories.¹⁰ We've carefully considered these and have enhanced the One Codex Platform to provide the tools and data infrastructure necessary for laboratories to kickstart their clinical NGS programs. We are excited to deliver accessible, clinically-actionable microbial genomics capabilities to meet the growing needs of laboratories performing clinical metagenomics sequencing.

Table 1. Key features of the One Codex Clinical Metagenomics solution

Turn-around time	Raw sequence to actionable report in <20 minutes
Clinically actionable results	Automated reporting of informative results based on lab-director's criteria
LDT support	Validation, documentation and monitoring all streamlined in the One Codex web application
Security	Secure, HIPAA compliant platform
One Codex Database	The world's leading curated microbial reference database. Benchmarked to be contamination and misclassification-free.
Breadth of pathogen detection	Specificity of pathogen detection based on lab-selected pathogen lists
Transparent quality control	Accessible interface to track and evaluate quality of data underlying results when needed
Alert capabilities	Automated alerting for trends in specific organisms or sample types.

Challenges & One Codex Solution

Challenge: *Different labs have different sequencing technologies, sample types, nucleic acid extraction and library prep workflows*

Solution: **Sample-type and sequencing technology agnostic support**

We ingest raw sequencing data files that can be uploaded automatically to the One Codex platform via our API, Illumina Basespace integration or straight from the web browser.^{11,12} Sequencing technology agnostic support means that raw sequence files can be generated by the sequencing platform (Illumina, ONT, PacBio, Ion Torrent, etc.) and any genomic data type being generated (shotgun metagenomics, targeted amplicon, hybrid capture) by any lab. This flexibility allows us to support the wet lab and sequencing workflow that works best for each individual lab. We account for differences in sequencing quality and data input types automatically in downstream analyses.

Challenge: *Data analysis & reporting methods lack transparency or can be difficult to interpret*

Solution: **Transparent, version controlled pipelines that incorporate laboratory quality control and assurance parameters and are customizable to meet the needs of individual labs**

The sensitive and highly curated One Codex Database is the backbone of our clinical mNGS workflow.⁹ Clinical metagenomics requires a level of specificity beyond microbial community classification¹³. We achieve this by combining a sensitive *k*-mer based microbial classification with read mapping and alignment to curated pathogen lists, which can be customized by lab, patient population or sample type. This two-step approach enables the hypothesis-free testing inherent with metagenomics with the data metrics required for interpretability. For example, sequencing quality, reference genome used, depth, coverage and ambiguous sequences mapping to multiple organisms can together provide confidence measures for detected organisms.^{3,14} Transparency in our analysis and ability to dig in to informative genomic information when samples are difficult to interpret enables better lab director and clinical review and decision making for what a given mNGS result means for a given patient and sample type.

Challenge: *Developing and maintaining clinical grade bioinformatics pipelines for mNGS requires software engineering and support as well as advanced compute infrastructure*

Solution: **One Codex is a scalable, cloud-based data platform capable of rapidly and reproducibly analyzing thousands of microbial genomes**

Clinical bioinformatics pipelines for microbial genomics analysis requires a level of quality, reproducibility, accessibility and speed beyond standard analysis and interpretation practices found in research settings.² One Codex specializes in creating user-friendly software that can rapidly analyze thousands of genomes in a secure cloud environment.⁴ The platform strictly versions all software and data, including sequencing data and metadata, to enable reproducibility of analyses and results.¹⁵ Our system allows historical versions of a workflow to be rerun, enabling auditing, and ensuring reproducibility of results at any point in the future. All data storage and sharing systems can be accessed from the command line, web app GUI, API, or via integrations with laboratory information systems. Clinical mNGS analysis begins upon upload of raw data with results available in approximately 20 minutes.

Challenge: *Hypothesis-free testing makes it impractical to validate all possible targets*

Solution: **Infrastructure for supplementing traditional methods-based validations with in silico data**

mNGS can theoretically detect an infinite number of microbes making traditional target validation impractical.^{2,16-18} While there is no substitute for methods-based validation, the One Codex platform can pick up where wet lab and person-power falls short: infrastructure is available to support in silico analytical validation of thousands of microbes across relevant sample type background matrices. Example features include analysis and determination of accuracy, precision, LOD, and reportable range. Positive control monitoring, negative control subtraction and sample quality metrics including microbial and host proportions, contamination detection are also available. Ongoing quality assurance measures are readily accessible for review of results, controls monitoring and quality improvement initiatives.

Challenge: *Metagenomics diagnostics should provide clinically meaningful results that are relevant for specific labs and patient populations*

Solution: **The curated One Codex Database enables sensitive microbial sequence detection, clinically relevant reporting names, automated reporting and accessible review and auditing**

The curated One Codex database supports broad microbial classification, however reporting every microbe in a sample is not necessarily clinically informative.¹⁰ We provide the information to allow laboratories to leverage their expertise in clinically meaningful results by providing accessible views into the genomic data, thereby supporting labs to decide at what thresholds, sample types and for which patient populations specific organisms should be reported. Automate reporting of

straightforward cases, while easily reviewing those that are not: the continuum of raw data to curated report from every sample is accessible via the platform. Performance of internal controls and negative control subtraction are always available for director review and conferencing with clinicians via PDF reports, custom dashboards and our intuitive GUI.

Challenge: Aggregate data monitoring for tracking clinical testing, and informing antimicrobial stewardship and infection prevention should be readily available

Solution: Data infrastructure that enables seamless sharing for individual and population-based data

While the primary goal of clinical mNGS is to diagnose individual patients, genomic data is inherently informative for population-level insights, for example antimicrobial resistance and epidemiologic trends. The One Codex mNGS workflow seamlessly integrates with other analyses that can inform these initiatives, for example variant calling, strain designation, genomic epidemiology investigations and export to public databases (NCBI, GISAID, SRA etc). Institutional profiles and custom projects allow users to collaborate (e.g. different data types can be added to a sample by different users) and access different views of the data. To illustrate, for several laboratory customers, raw sequence data and informatics are accessible to bioinformaticians, while summary results in PDF format are available for clinicians. Automated alerts enable streamlined reporting and sharing of data for priority healthcare pathogens to infection prevention teams and genomic detection of resistance mechanisms for antimicrobial stewardship.

Challenge: Clinical metagenomics data and patient metadata requires a high-level of security, documentation and compliance throughout testing, analysis, reporting and data storage

Solution: Secure platform for analyzing and storing microbial genomics data at scale

One Codex is hosted in the Amazon Web Services cloud, and implements a comprehensive range of technical and procedural controls to ensure the security, privacy, and integrity of all data hosted on the platform. The One Codex platform is HIPAA compliant and undergoes annual third-party penetration testing. All data is encrypted in transit to or from the One Codex application using Transport Layer Security, and at rest using 256-bit Advanced Encryption Standard (AES-256). Access

to data in One Codex is secured via unique user identifiers and passwords, with role-based access controls following the principle of least privilege for all sensitive data access. One Codex maintains a comprehensive Information Security Management System, with detailed data handling procedures and regular training for all staff with access to sensitive data. One Codex implements a secure Software Development Lifecycle (SDLC), with a combination of automated and manual testing and code review performed on all code deployed to a production environment. Scientific pipelines are packaged in Docker containers and all versions are preserved. Our storage system links raw and processed sequence data with metadata making it easy to export compliance documentation for LDTs.

References

1. Gu W, Miller S, Chiu CY. Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection. *Annu Rev Pathol Mech Dis.* 2019;14(1):319–38.
2. Chiu CY, Miller SA. Clinical metagenomics. *Nat Rev Genet.* 2019 Jun;20(6):341–55.
3. Dulanto Chiang A, Dekker JP. From the Pipeline to the Bedside: Advances and Challenges in Clinical Metagenomics. *J Infect Dis.* 2020 Mar 28;221(Supplement_3):S331–40.
4. One Codex | A fast, easy-to-use platform for microbiome sequencing and analysis [Internet]. [cited 2023 Oct 12]. Available from: <https://www.onecodex.com/>
5. VE303, a Defined Bacterial Consortium, for Prevention of Recurrent Clostridioides difficile Infection: A Randomized Clinical Trial | Gastroenterology | JAMA | JAMA Network [Internet]. [cited 2023 Oct 12]. Available from: <https://jamanetwork.com/journals/jama/fullarticle/2803996>
6. Precision identification of diverse bloodstream pathogens in the gut microbiome | Nature Medicine [Internet]. [cited 2023 Oct 12]. Available from: <https://www.nature.com/articles/s41591-018-0202-8>
7. Qi C, Hountras P, Pickens CO, Walter JM, Kruser JM, Singer BD, et al. Detection of respiratory pathogens in clinical samples using metagenomic shotgun sequencing. *J Med Microbiol.* 2019 Jul;68(7):996–1002.
8. Comprehensive Viral Research Panel - Twist Bioscience [Internet]. [cited 2023 Feb 2]. Available from: <https://www.twistbioscience.com/products/ngs/fixd-panels/comprehensive-viral-research-panel>
9. One Codex Database [Internet]. [cited 2023 Feb 2]. Available from: <http://docs.onecodex.com/en/articles/3761205-one-codex-database>
10. Greninger AL. The challenge of diagnostic metagenomics. *Expert Rev Mol Diagn.* 2018 Jul 3;18(7):605–15.
11. Importing Data | One Codex Docs [Internet]. [cited 2023 Oct 12]. Available from: <http://docs.onecodex.com/en/articles/3764397-importing-data>
12. Importing Sequence Read Archive (SRA) Data Into One Codex [Internet]. 2023 [cited 2023 Apr 6]. Available from: <https://onecodex.com/blog/2023/01/05/importing-sequence-read-archive-sra-data-into-one-codex/>
13. Batool M, Galloway-Peña J. Clinical metagenomics—challenges and future prospects. *Front Microbiol* [Internet]. 2023 [cited 2023 Oct 12];14. Available from: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1186424>
14. Govender KN, Eyre DW. Benchmarking taxonomic classifiers with Illumina and Nanopore sequence data for clinical metagenomic diagnostic applications. *Microb Genomics.* 2022 Oct 21;8(10):mgen000886.
15. Database Versions | One Codex Docs [Internet]. [cited 2023 Oct 12]. Available from: <https://docs.onecodex.com/en/articles/3764811-database-versions>
16. Duncavage EJ, Abel HJ, Merker JD, Bodner JB, Zhao Q, Voelkerding KV, et al. A Model Study of In Silico Proficiency Testing for Clinical Next-Generation Sequencing. *Arch Pathol Lab Med.* 2016 Oct;140(10):1085–91.
17. Ebinger A, Fischer S, Höper D. A theoretical and generalized approach for the assessment of the sample-specific limit of detection for clinical metagenomics. *Comput Struct Biotechnol J.* 2021 Jan 1;19:732–42.
18. Schlager R, Chiu CY, Miller S, Procop GW, Weinstock G, the Professional Practice Committee and Committee on Laboratory Practices of the American Society for Microbiology, et al. Validation of Metagenomic Next-Generation Sequencing Tests for Universal Pathogen Detection. *Arch Pathol Lab Med.* 2017 Feb 7;141(6):776–86.



Discover the power of One Codex.

Register for a complimentary account at www.onecodex.com

One Codex is a data platform for applied microbial genomics, enabling new and valuable applications in clinical diagnostics, food safety, and biosecurity.

hello@onecodex.com
(226) 406-8524