

# Building A Modern Clinical Bioinformatics Platform for Metagenomic Pathogen Detection and Reporting

Brian O'Donovan, Sirisha Thippabhotla, Nick Gimbrone, Greg Gydush, Andrew Carretta, Anita Cheung, Lakshmi Yajurvedi, Renerose Dimatulac, Fay Wang, Anna Uehara, Emily Crawford, Steve Miller

Delve Bio, 240 Newbury St, Boston, MA 02116

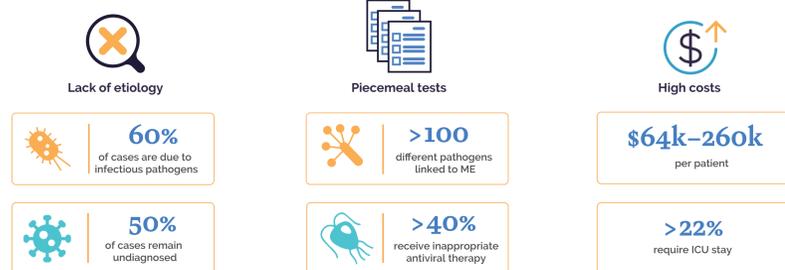
## Abstract/Summary

Metagenomic Next-Generation Sequencing (mNGS) testing for central nervous system (CNS) infections was introduced in 2014 at UCSF as part of a pioneering effort establishing the clinical utility of mNGS technology for unbiased pathogen detection. In partnership with UCSF, Delve Bio, Inc. has advanced this legacy by refactoring the UCSF clinical mNGS bioinformatics pipeline, SURPI+, into an updated and modernized platform. The refactored pipeline is designed and built with emphasis on the following facets: speed, modularity, scalability, clinical interpretation, and auditability (regulatory compliance). Here, we present the summary of the technical stack underpinning the new platform - Delve Detect and Decide. Delve Bio aims to set a new standard in mNGS pathogen detection with this flexible architecture, enabling easy expansion into other sample types and paving the way for broader applications and advancements in unbiased mNGS diagnostics.

## Background and Introduction

### Clinical mNGS for Suspected CNS Infections

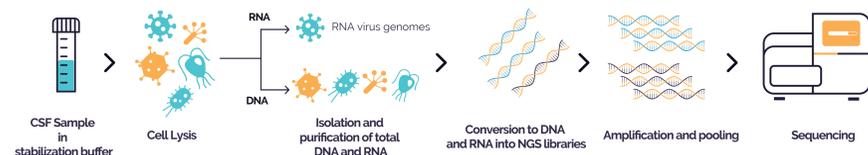
- 30,000 Suspected Meningitis and Encephalitis (ME) cases annually [1]
- Standard diagnostic paradigm is costly and inefficient [2, 3]



- **2014:** UCSF introduces mNGS testing for ME; allowing for the detection of a multitude of infectious agents simultaneously, with a single test, without a priori clinical suspicion
- **2022:** Delve Bio becomes the exclusive licensee of the UCSF assay and pipeline (SURPI+); begins the process of modernizing and scaling the process to expand access, improve clinical performance, and reduce turnaround time.

## Wet Lab - Broad Clinical Laboratory (BCL)

Automated Protocol Focused on Precision, Sterility, Speed, and Throughput



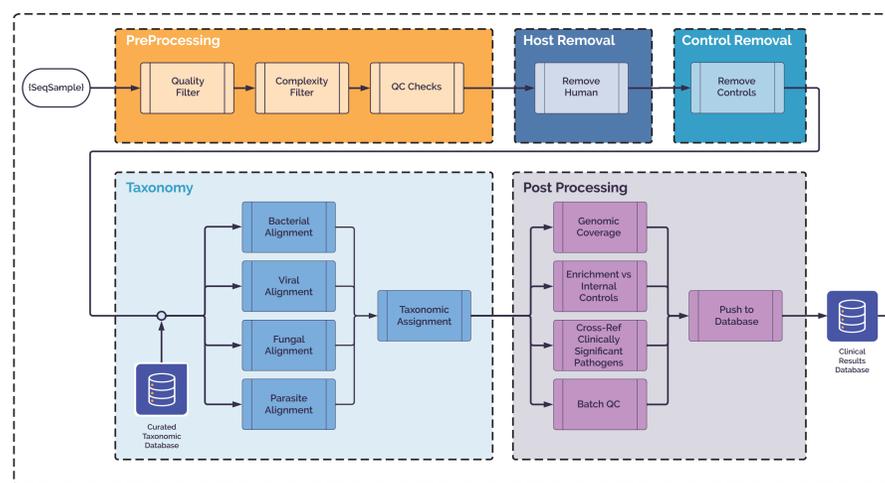
## Dry Lab - Delve Detect and Decide

Refactored for Speed, Accuracy, Scalability, and Auditability



## Methods

### Pipeline Overview/Modules

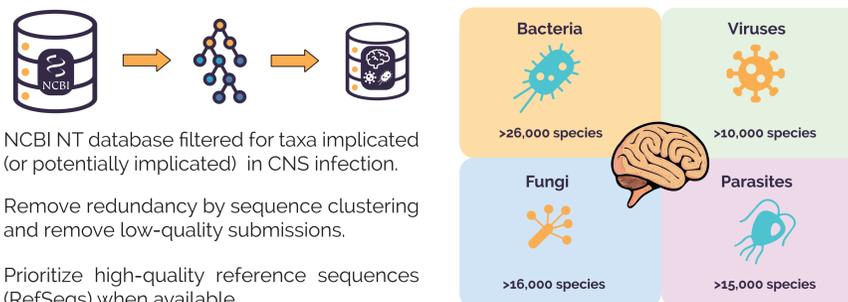


**Data** The SeqSample object is a custom class that serves as a standardized input and output of all steps/modules in the pipeline. A single FastQ file initializes a SeqSample.

**Task** Flyte docker

**Workflow** The pipeline is organized as a series of tasks managed by Flyte. Each task executes in its own containerized (docker) environment and accepts (as input) and outputs a SeqSample. Tasks are organized and grouped into Flyte workflows. Serial or parallel execution of tasks and workflows is managed by Flyte via calculation of the directed acyclic graph (DAG).

### Curated Database of Potential CNS Pathogens



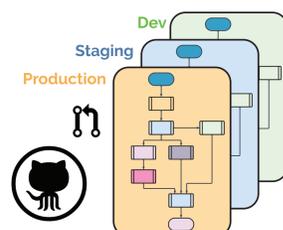
- NCBI NT database filtered for taxa implicated (or potentially implicated) in CNS infection.
- Remove redundancy by sequence clustering and remove low-quality submissions.
- Prioritize high-quality reference sequences (RefSeqs) when available.

### Pipeline Orchestration

- Execution flow and order of tasks (DAG) managed by Flyte on Elastic Kubernetes Service (EKS).
- All workflow and task executions are cached and memoized for easy iteration, retrieval, and reproducibility.
- Execution resources scale automatically; from a single sample to hundreds (in parallel) with no input or config changes required.
- Robust fault tolerance, resilience, and retry functionality - automated alerts and logging.
- Tasks and sub-workflows are built to be swappable; new workflows or additional modules can be implemented easily - either programmatically or through a drag-and-drop user-interface.

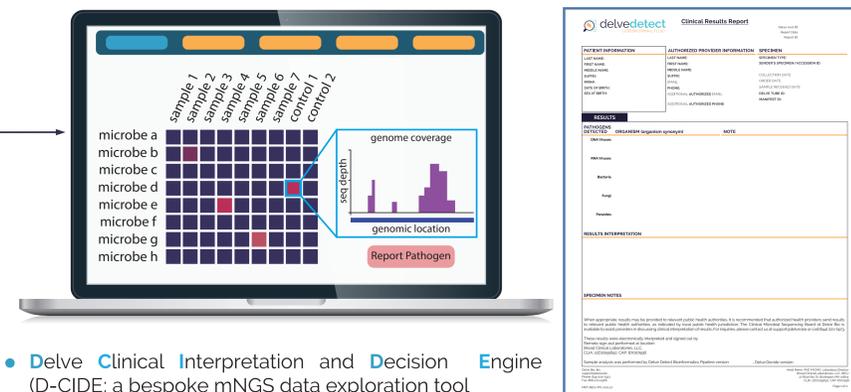
### Continuous Integration/Deployment (CI/CD)

- Pipeline changes and deployments are managed as tagged releases in GitHub.
- Independent development, staging, and production environments maintain code integrity and enable automated testing.
- Changes are rolled out continuously, robustly, and with automatic documentation and change logging (auditability).



## Methods (cont.)

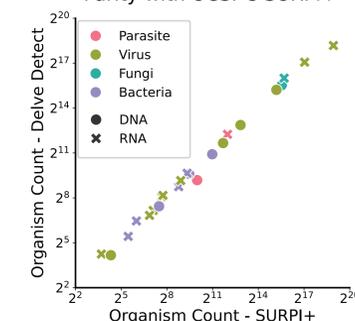
### Clinical Interpretation and Reporting



- Delve Clinical Interpretation and Decision Engine (D-CIDE; a bespoke mNGS data exploration tool)
- Results are presented in an interactive user-interface (UI) and contextualized with internal controls and statistical metrics.
- Users are provided direct interaction with primary sequencing data and methods to launch secondary analyses (BLAST, coverage maps, etc).
- Data is interpreted by a trained team of infectious disease specialists; Reports are populated and filed automatically based on user-input and sign-out.
- Access is controlled through strict user permissions and every user action is logged and auditable.

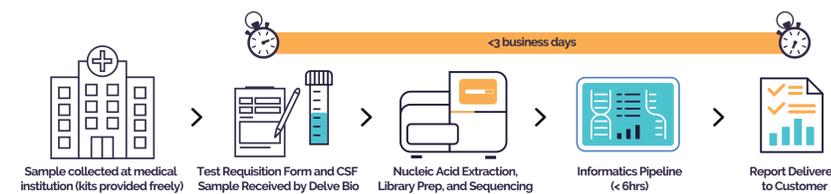
## Results

### Parity with UCSF's SURPI+



### delvedetect CEREBROSPINAL FLUID

- Parity of output for matched samples when compared to SURPI+ pipeline
- Taxonomic profiling against a database of >65,000 microbial species in under 6 hours
- Assay throughput of up to 48 samples per processing 'batch'
- Commitment to test turnaround time (TAT) of under three business days from receipt of CSF sample.



## Conclusion

Delve Detect and Decide represent a comprehensive refactoring of UCSF's clinical mNGS assay and associated SURPI+ bioinformatics pipeline. The Delve Detect pipeline retains clinical performance parity with SURPI+ while significantly improving efficiency, turnaround time, modularity, and scalability. Results are interpreted and reported through Delve Decide, a custom clinical decision support system that formalizes and records the decision-making process of our team of infectious disease experts as they interpret and contextualize the rich and often complex mNGS datasets. Built on a modern technical stack, the platform adheres to best software development practices, ensuring scalability, caching, rapid iteration, and compliance with the evolving regulatory landscape governing Laboratory Developed Tests (LDTs). Delve Bio is confident that this offering will help to further establish and advance the clinical utility and reach of mNGS.