

---

# CDC HEALTH DATA INNOVATION SUMMIT

Atlanta, GA | September 27-28, 2023

# Easy Genomics Democratising Access to Genomic Sequencing



**Kelsey Florek, PhD, MPH**

Senior Genomics and Data Scientist  
Wisconsin State Laboratory of Hygiene

[linkedin.com/in/kelseyflorek](https://www.linkedin.com/in/kelseyflorek)



**Evan Davey, BCS**

Partner & Global Lead  
DEPT®/HEALTH

[linkedin.com/in/evandavey](https://www.linkedin.com/in/evandavey)



# Today

- 01 Introduction
- 02 Why Genomic Sequencing?
- 03 What is Easy Genomics?
- 04 AWS HealthOmics
- 05 Demonstration

# 01

## Introduction





**Evan Davey, BCS**

Partner & Global Lead  
DEPT@/HEALTH

[linkedin.com/in/evandavey](https://www.linkedin.com/in/evandavey)

# 02

## Why Genomic Sequencing?



# What is Genomic Sequencing?

- All organisms carry a genome or blueprint composed of a sequence of nucleotides (A, C, G, and T/U) that are used to construct proteins that support a variety of functions.
- Sequencing an organism's genome is a powerful step in understanding the potential characteristics or features an organism might have.
- Genomics enhances public health outcomes by tailoring healthcare interventions, improving disease surveillance, and advancing our understanding of genetic factors affecting health and disease.



# Why is it important?

- Genomic Sequencing contributes to disease prevention, personalized healthcare, disease tracking, and the development of effective treatments.
- During the Covid-19 pandemic Genomics played a critical role in tracking variants across populations and geographies





# Revolutionary Technology

## Next Generation Sequencing

- massively parallel approach to sequencing a genome
- 360 billion ATGC's per sequencing run
- 40,000 - 150,000 words in a novel
- average word length in English is 4.79
- **one 72 hour sequencing run** would generate **791,121 novels** with **95,000 words each**

# Revolutionary Challenges

- **Cost** - Genomic sequencing can be expensive, making it inaccessible or challenging for areas of the healthcare system less resourced
- **Training** - Healthcare professionals and public health workers require additional training to effectively use genomic sequencing data in their practice.



# 03

## What is Easy Genomics?





## Open source for all.

Easy Genomics is a no-code, open-source solution for genomic data analysis. If the past few years have taught us anything, it's that genomic sequencing will play a major part in the future of (public) health. But it's an intensive, complex and expensive process to set up and run.

We've changed that.



### Simple to use.

Requires little to no technical expertise.



### Seamless experience

Powered by NextFlow, we provide a faster, hassle-free workflow.



### Secure data storage

Endless scalability as your needs grow.



### Start small, scale fast

Encrypted security to keep your patient data locked and secure.

## How it works

Run your genomic workload on a no-code front-end interface, that any non-technical user can operate in three easy steps:



### Select a Workflow

Customise to your analytic requirements.



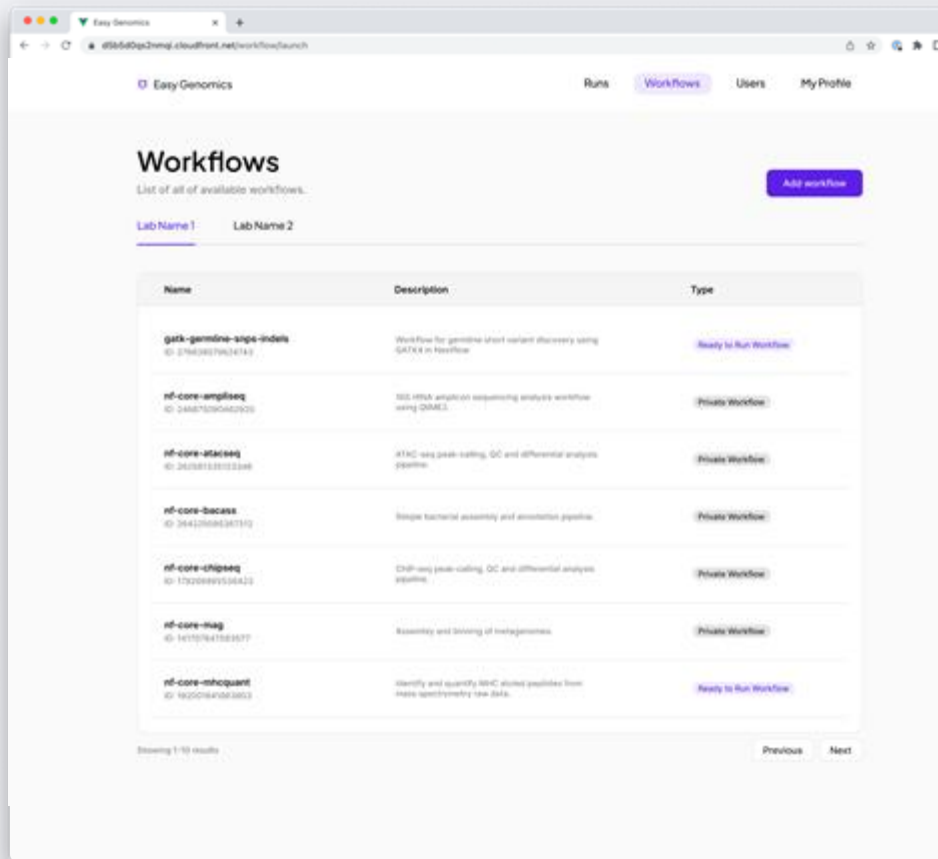
### Upload your Data

Manage access to single & multiple samples.



### Hit Submit

Monitor & receive your results via browser or email.



The screenshot shows the 'Workflows' page in the Easy Genomics web application. The page title is 'Workflows' and it includes a sub-header 'List of all of available workflows.' and an 'Add workflow' button. Below this, there are two tabs for 'Lab Name 1' and 'Lab Name 2'. A table lists several workflows with columns for Name, Description, and Type. The workflows listed are:

Name	Description	Type
gatk-gemline-snp-indels ID: 27663607962149	Workflow for gemline short variant discovery using GATK in Hadoop	<a href="#">Ready to Run Workflow</a>
nf-core-amplicon ID: 24687329042929	100-ntRNA amplicon sequencing analysis workflow using DRAGEN	<a href="#">Private Workflow</a>
nf-core-atacseq ID: 26259123102349	ATAC-seq peak-calling, QC and differential analysis pipeline	<a href="#">Private Workflow</a>
nf-core-bacseq ID: 26422666242719	Single bacterial assembly and annotation pipeline	<a href="#">Private Workflow</a>
nf-core-chipseq ID: 173209892534423	ChIP-seq peak-calling, QC and differential analysis pipeline	<a href="#">Private Workflow</a>
nf-core-mag ID: 16102764332677	Assembly and binning of metagenomes	<a href="#">Private Workflow</a>
nf-core-micquant ID: 16250264082853	Identify and quantify MIC stored profiles from mass spectrometry raw data	<a href="#">Ready to Run Workflow</a>

At the bottom of the table, it says 'Showing 1-10 results' and has 'Previous' and 'Next' navigation links.

# Easy Genomics



## DEPT®/Health

Digital health product development with care.



## Amazon Web Services

On-demand cloud computing web services.



## Datapult, an APHL Company

A secure, cloud-based, Public Health sanctioned service.



## Wisconsin State Laboratory of Hygiene

Wisconsin's Public, Environmental and Occupational Health Laboratory Since  
1903

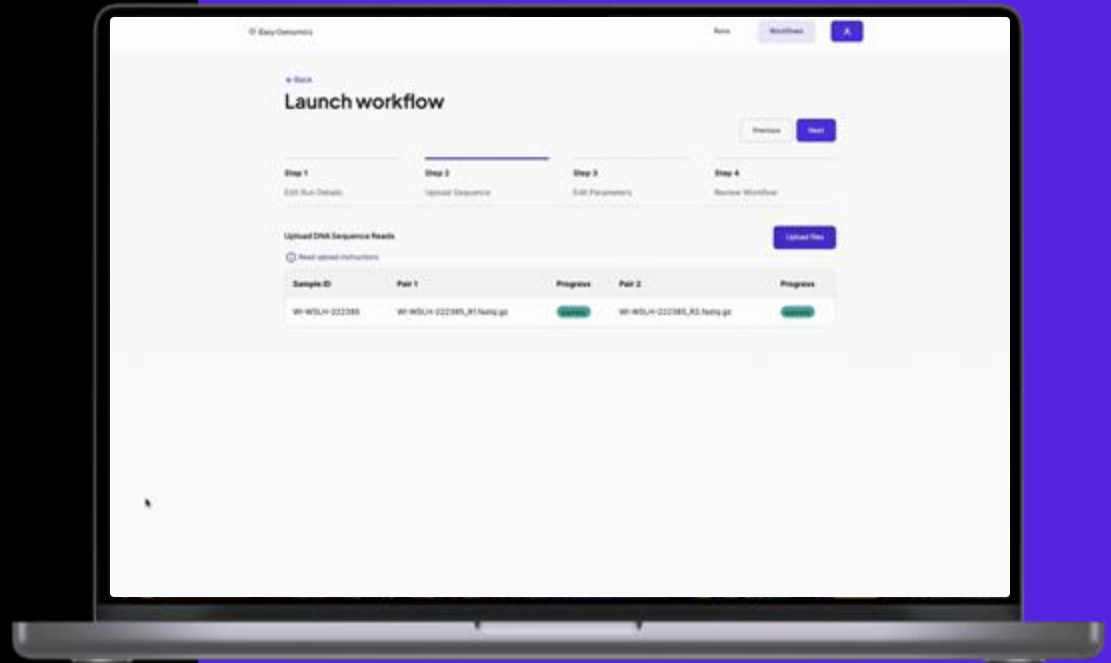
# Easy Genomics V1

- User accounts and authentication
- Simple process for uploading and/or submitting genomic sequencing data for analysis
- Allow users to select and easily configure analyses
- Access to analytical workflow status and results



# Easy Genomics V1

- 3 user types: super admin, lab admin, user
- Supports multiple "labs" with separate workflows and users
- Enables data submission through web interface into AWS S3 and automated samplesheet generation
- Runs analytical workflows through Nextflow Tower API



# 04

## AWS Health Omics



# AWS Health Omics





# AWS Health Omics

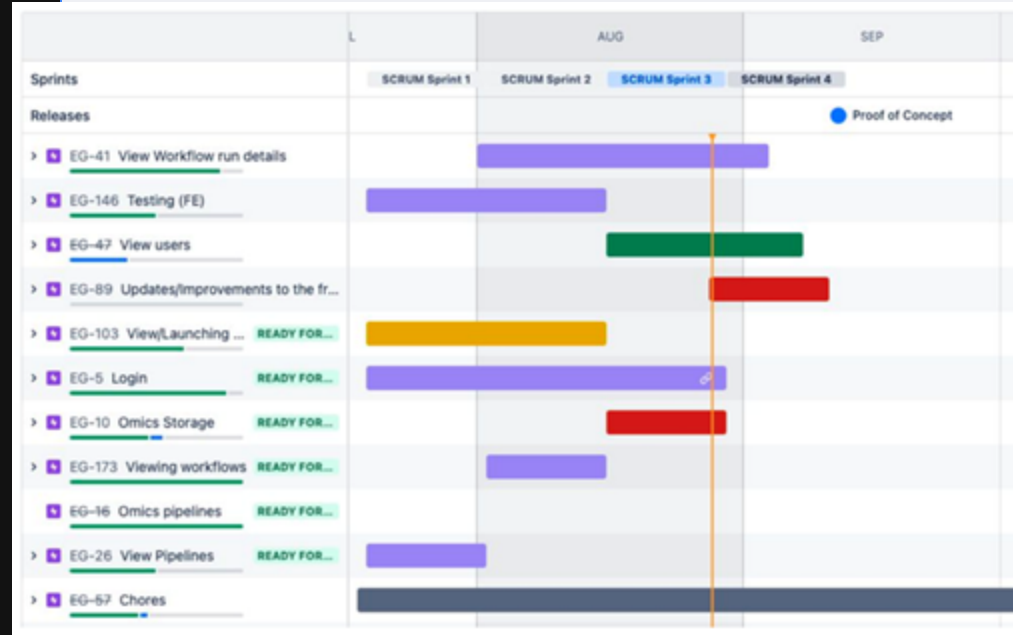
- Removes the need to configure and manage complex cloud infrastructure involving AWS Batch / EC2
- Provides tailored storage for sequence and variant data
- Supports Nextflow & WDL workflows
- Requires staging workflow containers and data
- Challenging workflow configuration
- Accessible through API



# Now we have Health Omics, let's upgrade Easy Genomics!

Easy Genomics v2

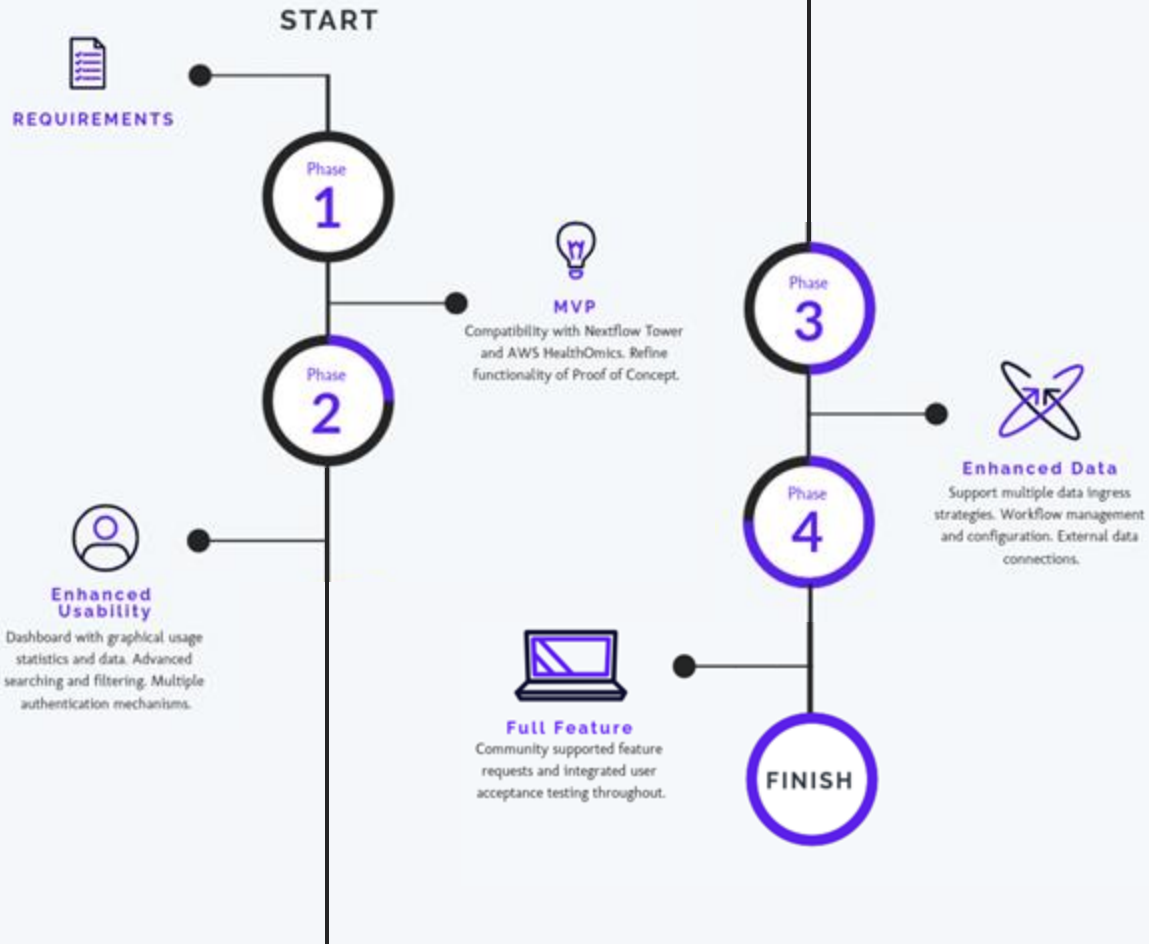
- Develop an open sourced solution powered by AWS' HealthOmics
- Provide an easy-to-use, web-based interface that will facilitate access to key genomic sequencing pipelines for all lab technicians
- With limited training and no technical skills



# Easy Genomics v2 Roadmap

→ August 2023 - POC Launched

→ Q1 2024 - Initial Release



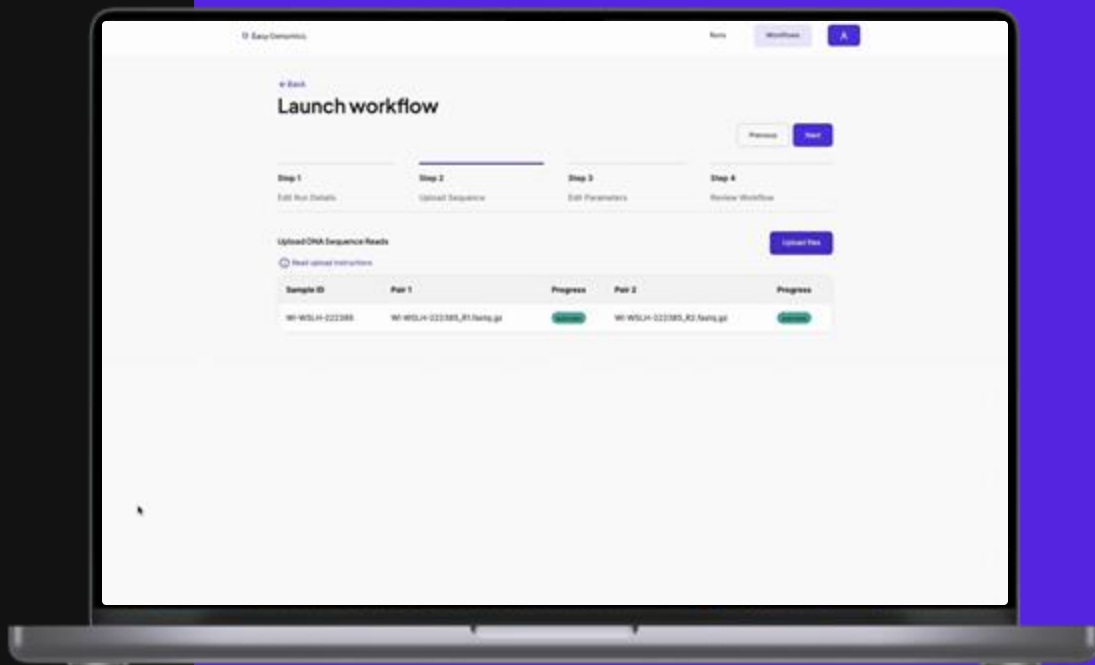
# Open source + community driven

## → Open Source

Easy Genomics is built for a field that is fundamentally built on open source projects. Genomic software and workflows are community efforts to improve our understanding of the biological world that we exist in.

## → Community Driven

Easy Genomics is a solution designed for a community of public health laboratories and is charged with meeting their unique needs.




# 05

## Demonstration



 Easy Genomics

Let's democratise access  
to Genomic Sequencing

 [evan@twobulls.com](mailto:evan@twobulls.com)

## Log in to your account

You will need an account created by your Lab Administrator.

Email address



Password



Log in

## Workflows

Select a workflow below to view a summary, or launch a pipeline and start testing.

[Launch pipeline](#)

Pipeline	User	Activity
WI-MS192-230818_4 wslh-bio/spriggan	dustin-lyfoung Submitted: 4 days ago	SUCCEEDED
WI-MS192-230818_3 wslh-bio/spriggan	dustin-lyfoung Submitted: 4 days ago	FAILED
WI-MS192-230829_4 wslh-bio/spriggan	dustin-lyfoung Submitted: 4 days ago	SUCCEEDED
WI-MS192-230829_3 wslh-bio/spriggan	dustin-lyfoung Submitted: 4 days ago	FAILED
WI-MS192-230829_2 wslh-bio/spriggan	dustin-lyfoung Submitted: 4 days ago	FAILED
WI-MS192-230818_2 wslh-bio/spriggan	dustin-lyfoung Submitted: 4 days ago	FAILED
WI-MS192-230829 wslh-bio/spriggan	dustin-lyfoung Submitted: 4 days ago	FAILED
WI-M3478-230825 wslh-bio/spriggan	dustin-lyfoung Submitted: 4 days ago	SUCCEEDED
WI-M3478-230821 wslh-bio/spriggan	dustin-lyfoung Submitted: 4 days ago	SUCCEEDED
WI-MS192-230818 wslh-bio/spriggan	dustin-lyfoung Submitted: 4 days ago	FAILED

Showing 1 to 10 results

[Next](#)

## Sequencing

Upload DNA Sequencing Reads

[Read upload instructions](#)

Sample sheet

Upload files

Sample ID	Pair 1	Progress	Pair 2	Progress
TestSample01-211229_S1_L001	TestSample01-211229_S1_L001_R1_001.fastq.gz	100%	TestSample01-211229_S1_L001_R2_001.fastq.gz	100%



## Sequencing

Upload DNA Sequencing Reads

Sample sheet

Upload Files

[Read upload instructions](#)

Sample ID	Pair 1	Progress	Pair 2	Progress
TestSample01-211229_S1_L001	TestSample01-211229_S1_L001_R1_001.fastq.gz	100%	TestSample01-211229_S1_L001_R2_001.fastq.gz	100%

sample\_sheet.csv - LibreOffice Calc

File Edit View Insert Format Styles Sheet Data Tools Window Help

Liberation 5 10 B I U A % 00 00 00

A1	sample		
1	sample	Pair 1	Pair 2
2	TestSample01-211229_S1_L001	TestSample01-211229_S1_L001_R1_001.fastq.gz	TestSample01-211229_S1_L001_R2_001.fastq.gz
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			

Sheet 1 of 1 Default English (USA) Average: Sum 0 90%

 Easy Genomics



[easygenomics.org](https://easygenomics.org)

**DEPT./HEALTH**



[deptagency.com/health](https://deptagency.com/health)



**Kelsey Florek, PhD, MPH**

Senior Genomics and Data Scientist  
Wisconsin State Laboratory of Hygiene

[linkedin.com/in/kelseyflorek](https://www.linkedin.com/in/kelseyflorek)

**DEPT®**



**Evan Davey, BCS**

Partner & Global Lead  
DEPT®/HEALTH

[linkedin.com/in/evandavey](https://www.linkedin.com/in/evandavey)

Thank you. ■