Amazon Omics
Transform genomic and biological data into insights

W. Lee Pang
Pr. Solutions Architect - HealthAI
Today’s Agenda

Background

Introducing Amazon Omics

Features

Use cases

Demo

Customers

Resources
Computing as a **utility**

- On Demand and fit for purpose
- Pay as you go, and only for what you use
- Focus on applications and not infrastructure
The AWS Cloud spans 99 Availability Zones within 31 geographic regions around the world, with announced plans for 12 more Availability Zones and 4 more AWS Regions in Canada, Israel, New Zealand, and Thailand.
Customer **benefits** of AWS

- Security
- Availability
- Performance
- Scalability
- Flexibility

Low cost
Customer obsessed

90% of our roadmap originates with customer requests and are designed to meet specific needs

96% of R1 Research Institutions are using AWS
How AWS can help your research

Science, not servers
Use compute when you need it to do large-scale analysis

Collaboration
Access data sets that span institutions

Share effort
Leverage work done by other scientists to save time

Reproduce research
A common platform for reproducing scientific analyses

State-of-the-art analytics
Use data science methods in your research

Security
A collection of tools to protect data and privacy
Why AWS for genomics
Challenges

Genomic and biological data has the potential to transform how we treat disease—but its scale is complicated and costly to manage.

- Tens of millions of whole genomes to be sequenced and stored in the next 5 years
- Multiple specialized tools and workflow languages required
- Special security, privacy, and compliance requirements mandated
Cost per human genome

Arrival of massively parallel sequencing technology

Cost per genome

Date

$100,000,000
$10,000,000
$1,000,000
$100,000
$10,000
$100

$95,263,072

$200


Moore’s law

>100,000X
Amazon Omics

A purpose-built service to help healthcare and life science organizations and their software partners store, query, and analyze genomics, transcriptomic, and other omics data and then generate insights from that data to improve health and advance scientific discoveries.
Benefits

- Multiomic and multimodal analysis
- Population-level scale
- Fully managed bioinformatics computation
- Built-in security, privacy, and compliance
How it works

Omics sequence data such as RNA or DNA

Amazon Omics
Store, query, and analyze genomic, transcriptomic, and other omics data

Sequence store
Support large-scale analysis and collaborative research

Bioinformatics workflows
Automatically provision and scale infrastructure to simplify running your analyses

Variant and annotation data
Optimize variant and annotation data

Multimodal and multiomic analysis
Query and analyze data to generate new insights

Clinical and medical imaging data
What is Amazon Omics?

Managed services

Compute

Storage

Workflows

Definitions

Runs

Variant Store

Reference Store

Sequence Store

Annotation Store

Analytics

Your applications

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Storage

- Discover data
- Share data
- Lower costs
1. **Oomics CreateReferenceStore**
   - Reference sequence: GCTTGGTCA
   - Reference ID: omics://...

2. **Oomics CreateSequenceStore**
   - Raw sequence:
     - Fastq (GCTTGGTCA)
     - Bam (GATAATGCA)
   - Reference ID: omics://...

3. **Reference Import Job**
   - Reference sequence: GCTTGGTCA
   - Reference ID: omics://...

4. **Read Set Import Job**
   - Raw sequence:
     - Fastq: GCTTGGTCA
     - Bam: GATAATGCA
   - Reference ID: omics://...

**Transfer Manager**
- Destination object:
  - fastq
  - bam
  - cram
- Read set:
  - fastq
  - bam
  - cram
Workflows

- Science not infrastructure
- Pay-as-you-go pricing
- Predictable cost per workflow
Analytics

Ingests variant and annotation data

Allows fine-grained access control

Scales with you
Top use cases

Scale population sequencing
Understand how genomic variation maps to phenotypes across a population

Improve clinical genomics workflows
Build reproducible and traceable clinical genomics workflows

Accelerate clinical trials
Integrate genome analysis into clinical trials to test new drug candidates’ efficacy

Enhance research and innovation
Streamline and control storage, access, and analysis of anonymized genomic data
Demo!
What we are going to build

END-TO-END GENOMIC DATA JOURNEY USING AMAZON OMICS

Create sequence and reference stores from existing Amazon S3
Create and run workflows to process sequence data
Create and query variant and annotation stores
Registry of Open Data on AWS

Direct data access from AWS native interfaces

More than 800 usage examples, tutorials, and publications

registry.opendata.aws
Recap – what we built

END-TO-END GENOMIC DATA JOURNEY USING AMAZON OMICS

Stored raw sequencing data

Ref. sequence
GCTTGGTCA
GATAATGCA

Raw sequence
GCTTGG
GAT
CATGCA

Ran analysis workflows

Workflow Definition Language

Variants
GCTTGGACA
GATCATGCA

Annotations
vcf
gvcf
gff
tsv

Stored and queried variant data

Query and interpret with Amazon Athena

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.
# Recap – What it cost

<table>
<thead>
<tr>
<th>Item</th>
<th>Size</th>
<th>Cost dimension</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>ref. sequence</td>
<td>3 Gbase</td>
<td>Free</td>
<td>$ 0.00</td>
</tr>
<tr>
<td>raw sequence</td>
<td>1.184 Gbase</td>
<td>$ Gbase-month</td>
<td>$ 0.0068</td>
</tr>
<tr>
<td>workflow run (GATK on NA12878)</td>
<td># vCPUs/task</td>
<td>$ omics instance/hr</td>
<td>$ 0.521</td>
</tr>
<tr>
<td></td>
<td># GB RAM/task</td>
<td>$ GB (Storage)-hr</td>
<td>$ 0.507</td>
</tr>
<tr>
<td></td>
<td>1.2 TB storage</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 hrs total run</td>
<td></td>
<td></td>
</tr>
<tr>
<td>variants</td>
<td>0.009 GB</td>
<td>$ GB-month</td>
<td>$ 0.0023</td>
</tr>
<tr>
<td>annotations</td>
<td>0.050 GB</td>
<td></td>
<td>$ 1.04 total</td>
</tr>
</tbody>
</table>
Multimodal analytics
Purpose-built services for healthcare and life sciences

Amazon Omics
Transform genomic, transcriptomic, and other omics data into insights

Amazon HealthLake Imaging and Analytics
Provide a complete view of individual or patient population health data

Amazon Comprehend Medical
Understand medical context using natural language processing

Amazon Transcribe Medical
Automatically convert medical speech to text
Multiomics and multimodal analyses

Clinical records
- Data (such as FHIR) normalization

Omics
- Data (such as variant) normalization

Imaging
- Metadata extraction

Analysis and query
- Amazon EMR
- Amazon Athena
- Amazon SageMaker
- Amazon QuickSight
- Third-party applications
Customers using Amazon Omics today …
We know that getting a comprehensive view of our patients is crucial to delivering the best possible care, based on the most innovative research. Combining multiple clinical modalities is foundational to achieving this. With Amazon Omics, we can expand our understanding of our patients’ health, all the way down to their DNA.

Jeff Pennington, Associate VP, Chief Research Informatics Officer
Children’s Hospital of Philadelphia
Element Biosciences is opening the world of biology to new possibilities through Element's AVITI system. Amazon Omics provides a simple solution for running workflows in the cloud, and this will enable the scientific community to easily process their sequencing data without the need to set up any infrastructure, allowing them to focus on their research.

Francisco Gracia, SVP Software and Informatics
Element Biosciences
“AWS Omics allows researchers to use tools and languages from their own domain and considerably reduces the engineering maintenance effort while taking care of cost and resource allocation considerations, which in turn reduces time to market and NRE costs of new features and algorithmic improvements.”

Ury Alon, VP Engineering
C2i Genomics
How do I start?
Reach out to us!
Amazon Omics partners

- BIOTEAM
- Cloud303.io
- DIAMOND AGE DATA SCIENCE
- lifebit
- LOKA
- ovation
- ptp
- Sentieon
Get started with Amazon Omics

Developer guide

Webpage
Additional resources to learn more

Genomics on AWS
aws.amazon.com/health/genomics

AWS for Health
aws.amazon.com/health

Open Data on AWS
aws.amazon.com/opendata

AWS Marketplace
aws.amazon.com/marketplace

AWS Partner Network
aws.amazon.com/partners/find
Thank you!

W. Lee Pang
pwyming@amazon.com

How did we do?

https://survey.immersionday.com/s_eYiF04R