

AWS HealthOmics ?? ? ?? ??

_____, ____ ____

AWS HealthOmics 是 一个 托管 的 基因组 数据 分析 平台 , 它 允许 用户 在 云上 存储 基因组 数据 , 并 使用 各种 分析 工具 对 数据 进行 分析 . 它 支持 各种 基因组 数据 格式 , 如 FASTQ , BAM , CRAM 等 . 用户 可以 使用 各种 分析 工具 , 如 AWS HealthOmics 提供 的 各种 分析 工具 , 或 使用 自己 的 分析 工具 .

HealthOmics 是 一个 托管 的 基因组 数据 分析 平台 . HealthOmics Storage 是 一个 托管 的 基因组 数据 存储 平台 . HealthOmics Analytics 是 一个 托管 的 基因组 数据 分析 平台 . HealthOmics Workflows 是 一个 托管 的 基因组 数据 分析 平台 .

AWS HealthOmics? ???????

?? ??

HealthOmics 是 一个 托管 的 基因组 数据 分析 平台 , 它 允许 用户 在 云上 存储 基因组 数据 , 并 使用 各种 分析 工具 对 数据 进行 分析 . 它 支持 各种 基因组 数据 格式 , 如 FASTQ , BAM , CRAM 等 . 用户 可以 使用 各种 分析 工具 , 如 AWS HealthOmics 提供 的 各种 分析 工具 , 或 使用 自己 的 分析 工具 .

HealthOmics 是 一个 托管 的 基因组 数据 分析 平台 , 它 允许 用户 在 云上 存储 基因组 数据 , 并 使用 各种 分析 工具 对 数据 进行 分析 . 它 支持 各种 基因组 数据 格式 , 如 FASTQ , BAM , CRAM 等 . 用户 可以 使用 各种 分析 工具 , 如 AWS HealthOmics 提供 的 各种 分析 工具 , 或 使用 自己 的 分析 工具 .

- [Storage](#) - ?? (FASTQ, BAM, CRAM? ?? ????? ?? ????? ?? ???)
- [Analytics](#) - ?????? ??? (? , VCF? ??? ?? ???)
- [Workflows](#) - ????? (Nextflow, WDL, CWL ? ????? ?????? ????? ??? ? ??? ?? ????? ?? ??? ???)

??

- 它 允许 用户 在 云上 存储 基因组 数据 , 并 使用 各种 分析 工具 对 数据 进行 分析 . HealthOmics 是 一个 托管 的 基因组 数据 分析 平台 . AWS 是 一个 托管 的 基因组 数据 分析 平台 . Amazon Athena 是 一个 托管 的 基因组 数据 分析 平台 . AWS 是 一个 托管 的 基因组 数据 分析 平台 .

- -


Getting started

- Storage**
 - Reference store
 - Sequence stores
- Analytics**
 - Variant stores
 - Annotation stores
- Workflows**
 - [Ready2Run workflows](#)
 - Private workflows
- Runs
- Run groups


Data shares

All data shares


Documentation



Step 1
Select a Ready2Run workflow - new
 Ready2Run workflows are pre-built pipelines from third-party software companies and open-source communities.



Step 2
Add run parameters
 Create a run by setting pipeline run parameter values, run priority, and output location.



Step 3
Start a run
 Choose Start run to perform a Ready2Run workflow your run in the Runs list page.

GATK-BP Germline fq2vcf for 30x genome Info

DetailsParametersRun tasks

Details

Workflow ID

9500764

Name

GATK-BP Germline fq2vcf for 30x genome

Workflow description

This workflow follows the GATK Best Practices workflow recommendations. It uses GATK4 tools to generate a gVCF from FASTQ files for up to 30x WGS data.

Workflow version

1.0

Workflow language

WDL

Run storage capacity (TB)

1.2

Workflow accelerators

None required

Workflow software

GATK 4.2.6.1, Samtools 1.16, Picard 2.27.4, BWA 0.7.17

Workflow files

[View test and example parameter files](#)

Workflow help

[Support information](#)

Max total input size

64GiB

List price/run

\$10.00

Publisher

Broad Institute

Subscription

No

Estimated run time (hh:mm)

12:30

Amazon Resource Name (ARN)

[arn:aws:omics:us-west-2:workflow/9500764](#)

Published

May 15, 2023, 09:00 (UTC -8)

Private workflow

Private R2R Run Task

AWS HealthOmics

Getting started

Storage

Analytics

Workflows

Data shares

Documentation

my run 20240410 Info

Last refresh: Today, a few seconds ago

DeleteClone runRe-

Console workflow run summary views and logs are here.

Run details include improved run and task metrics inside the HealthOmics console. The run summary includes enhanced run status and descriptions, as well as direct links to run and engine logs for faster troubleshooting. You can also see run, task and engine logs directly in the HealthOmics console.

Run summary Info

Metric	Description
Status	Completed
Run ARN	arn:aws:omics:us-east-1:run/473352
Total run time	0 days, 2 hours, 53 mins, 44 seconds
Output URI	View run output
S3 log location	View logs in S3

Run and workflow details Info

Run tasksParametersTags

Run tasks (107) Info

Choose View logs to view a tasks logs in CloudWatch.

Find task

< 1 2 >

ID	Name	Status	CloudWatch logs	Run time (dd:hh:mm:ss)	vCPUs	Memory(GiB)	GPU(s)	Start (UTC -5)	Str
3261604	MergeGVCFs	Completed	View logs	00:00:03:04	2	8	0	April 11, 2024, 02:02	Ap
1667819	HaplotypeCaller-13-1321	Completed	View logs	00:00:09:21	2	8	0	April 11, 2024, 01:44	Ap

Analytics

Healthomics Athena

HealthOmics API

- VCF files can be converted to TSV/CSV or GFF3 files
- HealthOmics can convert VCF files to TSV/CSV or GFF3 files
- Amazon Athena can query VCF files (Athena can query 3 TB files)

Annotation files: VCF (Variant Effect Predictor (VEP) annotations ?? ??.)

Annotation files: TSV, VCF, or GFF

?? ??

AWS CLI — HealthOmics can convert VCF files to TSV/CSV or GFF3 files.

AWS CLI (AWS CLI) — AWS CLI can convert VCF files to TSV/CSV or GFF3 files. AWS CLI can be installed on Windows, macOS, and Linux. AWS CLI can be installed on Windows, macOS, and Linux. AWS CLI can be installed on Windows, macOS, and Linux.

AWS SDK — AWS SDK can convert VCF files to TSV/CSV or GFF3 files. AWS SDK can be installed on Windows, macOS, and Linux. AWS SDK can be installed on Windows, macOS, and Linux. AWS SDK can be installed on Windows, macOS, and Linux.

AWS API — AWS API can convert VCF files to TSV/CSV or GFF3 files. AWS API can be installed on Windows, macOS, and Linux. AWS API can be installed on Windows, macOS, and Linux. AWS API can be installed on Windows, macOS, and Linux.

?? ??? ? ? Quota

<https://docs.aws.amazon.com/general/latest/gr/healthomics-quotas.html>

?? ??

HealthOmics Private workflow can be installed on Windows, macOS, and Linux.

- WDL workflows
 - [GATK Best Practice workflows](#)
 - [Analysis ready germline BAM to VCF](#)
 - [CRAM to BAM](#)
 - [FASTQs to analysis ready BAM](#)

- [Germline FASTQs to VCF](#)
- [Somatic SNPs and InDELS](#)
- ☐ [Protein folding workflows](#)
 - [AlphaFold](#)
 - [ESMFold](#)
- ☐ [Other WDL workflows](#)
 - [HISAT-Genotype HLA Caller](#)
- Nextflow workflows
 - ☐ [NF-Core workflows](#)
 - [FASTQC](#)
 - [RNAseq](#)
 - [scRNAseq-cellranger](#)
 - [scRNAseq](#)
 - [TaxProfiler](#)
 - ☐ [Other Nextflow workflows](#)
 - [VEP](#)

<https://github.com/aws-samples/amazon-omics-tutorials/tree/main/example-workflows>

Learn more about HealthOmics from these workshops and tutorials:

- HealthOmics workshop – [HealthOmics end to end workshop](#)
 - [Migrating Nf-core Workflows Into AWS HealthOmics](#)
- AWS genomics resources – [Public Amazon ECR repositories](#) related to genomics
- Python tutorials – [Jupyter notebook tutorials](#) on GitHub, covering HealthOmics storage, analytics, and workflows

Become familiar with additional HealthOmics tools that AWS provides:

- WDL linter – [HealthOmics linter for WDL](#)
- Nextflow linter – [HealthOmics linter for Nextflow](#)
- HealthOmics Amazon ECR helper tool – [Amazon ECR helper tool for HealthOmics](#)
- HealthOmics tools on GitHub – [Tools for working with HealthOmics](#) (Transfer manager, URI parser, Omics rerun, Run analyzer).

?? ???

- Secure Your Genomic Workflows and Data with AWS HealthOmics
- Category: AWS HealthOmics / AWS for Industries

Revision #9

Created 11 April 2024 01:44:55 by Hyunmin Kim

Updated 19 February 2025 06:19:35 by Hyunmin Kim