Genomics in the Cloud

Accelerate genomic discoveries on AWS

THE GROWTH OF GENOMICS DATA:

Time and cost of genome sequencing has dropped by a **factor of 1M in less than 10 years.**[1]

It is estimated that between **100M and 2B** human genomes will be sequenced by 2025.

Projections show genomic data acquisition will hit **1 zetta-bases per year in 2025.**

Estimates show **2–40 exabytes** of storage capacity will be needed just for human genomes by 2025.[2]

DATA TRANSFER & STORAGE

CHALLENGE

As genomics sequencing gets less expensive, the volume and velocity of data becomes harder to manage and store while still offering rapid and secure access.

SOLUTION

AWS services offer high-throughput data ingestion, cost-effective storage, secure access and efficient searching.

Macrogen manages 20+PB of data, and using AWS it cut backup costs by 35% compared with on-premises.

Genuity Science uses AWS Direct Connect with 10G pipe for data transfer and manages >6PB genomics data in the cloud.[3]

Illumina reduced genomics data storage costs by $90K per month by leveraging AWS tiered data storage options.[4]

SECONDARY ANALYSIS & WORKFLOW AUTOMATION

CHALLENGE

Companies struggle with tracking the origins of data and enabling researchers to run reproducible and scalable workflows while minimizing IT overhead.

SOLUTION

Cromwell, Nextflow or AWS native services offer scalable, cost-effective data analysis and simplified orchestration for running parallelizable workflows.

Fred Hutch can perform 7 years of compute time in 7 days on AWS, translating gigabytes of genomic data into insights.[5]

Automation and orchestration on AWS cut genomics research time by 50% for the University of Tubingen.[6]

Mission Bio processes millions of genomes and billions of data points on AWS from their single-cell DNA analyses.[7]

DATA AGGREGATION & GOVERNANCE

CHALLENGE

Successful genomic research and interpretation often depend on multiple, diverse datasets representing large populations, relying on data and methods to be findable, accessible, interoperable and reusable (FAIR).

SOLUTION

AWS enables organizations to harmonize multi-omic datasets and govern robust data access controls and permissions across a global infrastructure. Simplify the ability to store, query and analyze genomics data, and to link with clinical information.

Lifebit's federated technology platform provides access to 20+ PB of Genomics England's data for research analysis- without ever needing to copy or move data.[8]

Mount Sinai School of Medicine uses AWS to help scientists analyze more than 100TB of data generated by The Cancer Genome Atlas Consortium.[9]

Biogen is analyzing 500K UK Biobank whole exomes in the cloud, and it is using the knowledge to prioritize existing drug targets and identify new ones.[10]

INTERPRETATION & DEEP LEARNING

CHALLENGE

Broader adoption of sequencing is unlocking the opportunity to expand the discovery and translational potential of genomics in precision medicine. This requires incorporation of available datasets and knowledge bases, along with intensive computational power.

SOLUTION

Turn big genomic data into actionable insights by leveraging machine learning and high-performance computing. Advances in cloud computing enable greater efficiencies of scale, reproducible data processing and access to public data for clinical annotation, all within a compliance-ready environment.

DNAnexus Apollo on AWS can explore millions of phenotypic variants and billions of genotypes from the UK Biobank dataset in seconds.[11]

Fabric Genomics software on AWS can interpret an entire genome's variant set within minutes.[12]

Benchling reduced their CRISPR off-target search times by 90% and scaled to hundreds of genomes.[13]

REFERENCES

1. https://www.semanticscholar.org/paper/Big-Data-in-Genomics%3A-Challenges-and-Solutions-Is-a-Costa/0e387bd00952c8450deefcdbcdebf5c946c20f54?p2df
2. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4494865/
3. https://www.youtube.com/watch?v=ZKj8QxjOqog
4. https://aws.amazon.com/solutions/case-studies/illumina
5. https://aws.amazon.com/solutions/case-studies/fredhutch-case-study/
6. https://aws.amazon.com/solutions/case-studies/quantitative-biology-center/
7. https://aws.amazon.com/solutions/case-studies/mission-bio/
8. https://awslifesciences.ufcontent.com/web-day-2020-lifesciences-technical-tracks/powering-genomics-englands-research-environment-lifebit
9. https://aws.amazon.com/solutions/case-studies/mt-sinai/
10. https://awslifesciences.ufcontent.com/web-day-2020-lifesciences-business-tracks/biogen-ukbiobank-sequencing
11. https://www.dnanexus.com/product-overview/apollo/apollo-for-ukb
12. https://aws.amazon.com/solutions/case-studies/fabricgenomics/?trk=hcls_case-studies_card
13. https://aws.amazon.com/solutions/case-studies/benchling/

Learn how AWS can help your organization by visiting:

https://aws.amazon.com/health/genomics/