BRILLIANCE. TOGETHER.

HOW ARTIFICIAL INTELLIGENCE (AI) IS TRANSFORMING GENOMICS

UKHSA Conference 2023

Vibin Vijay - Data and AI Solution Specialist

OCF Limited

ABOUT ME

VIBIN VIJAY

- ✓ **Data and AI Solution Specialist** at OCF Limited
- ✓ **Associate Lecturer** at Sheffield Hallam University
- ✓ **14 years of experience** in Data, Business Intelligence,
- ✓ Machine Learning, Internet of Things (IoT), Artificial Intelligence (AI) and Big Data
- ✓ **Industries** – Healthcare & Life Sciences, Higher Education & Research, Manufacturing, Telecom and Banking

HOW AI IS TRANSFORMING GENOMICS

AGENDA

- About OCF
- Solutions
- Healthcare and life sciences
- Whole Genome Sequencing (WGS)
- Convergence of AI and WGS
- Accelerated genomics research – use cases

BRILLIANCE. TOGETHER.

ABOUT OCF

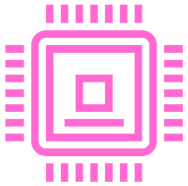
- ✓ HPC, Storage, Cloud & AI Integrator
- ✓ Headquartered in Sheffield, South Yorkshire - UK-wide coverage
- ✓ 50 Strong team
- ✓ Significant in-house expertise
- ✓ 2 decades experience in the market
- ✓ Un-matched partner ecosystem
- ✓ Worked with UKHSA (previously PHE & HPA) for over 10 years – supporting HPC and scientific research

OCF'S

SOLUTIONS & SERVICES

For over 20 years, OCF has delivered brilliance in high-performance computing.

Taking a consultative role between client and technology vendor, we combine technical expertise with trusted partnerships to develop cutting-edge solutions that help forward-thinking organisations break boundaries.



HPC

Bespoke high-performance computing (HPC) solutions

STORAGE

Tailored storage solutions to meet any performance or capacity requirement

AI & ML

AI Infrastructure, software & professional services

CLOUD

Public cloud, private cloud, cloud native, cloud burst or containers

PROFESSIONAL SERVICES

Consultancy, Support & Managed Services

OCF – HEALTH CARE &

LIFE SCIENCES

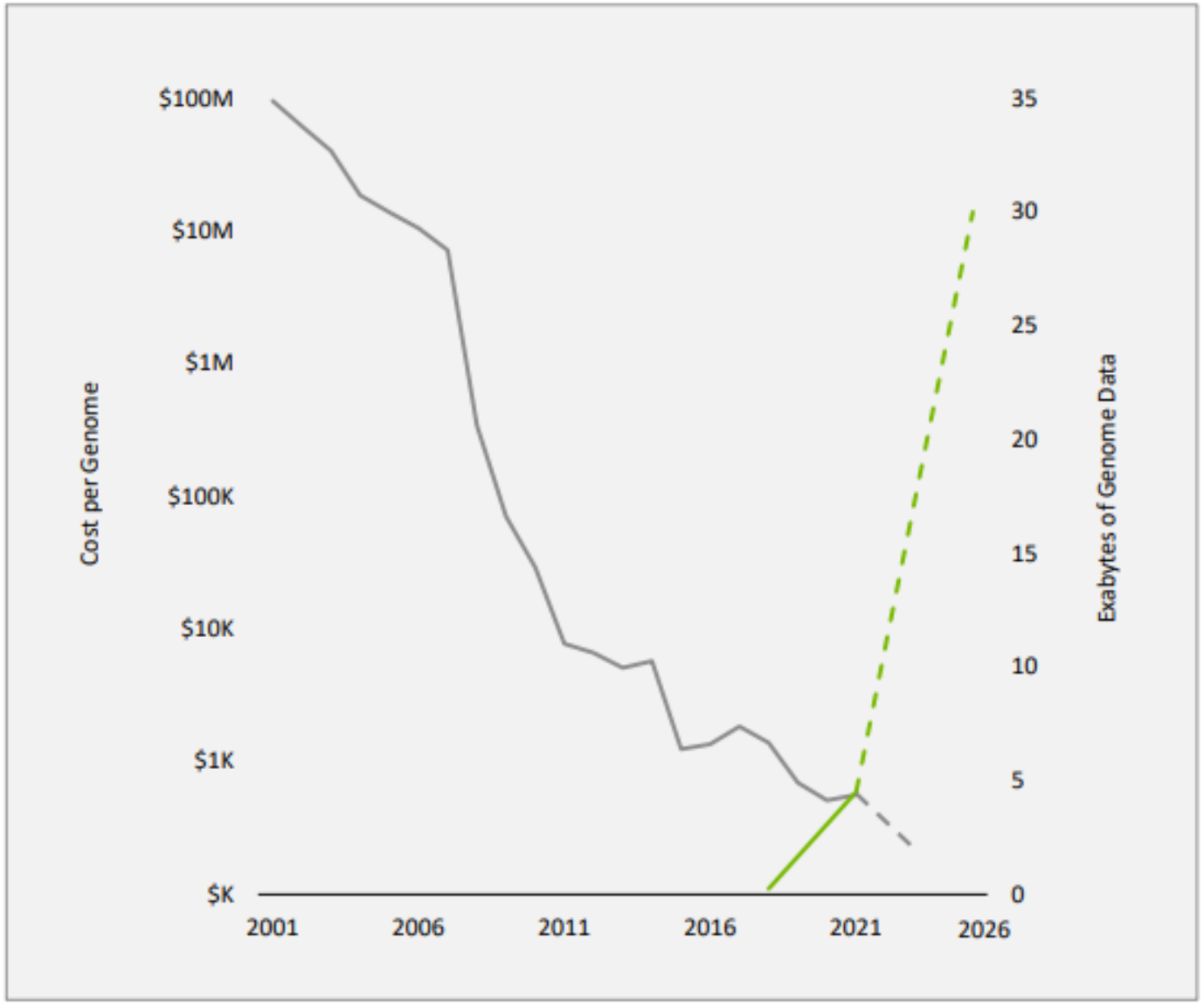
Customers include:

- UKHSA
  - Supplied, supported & assist in managing:
    - High Performance Computing (HPC) infrastructure
    - High Performance, large capacity storage (LUSTRE)
    - High Performance networks
    - OpenStack & OpenShift Scientific Private Cloud & Containers
    - GPU capability for AI workloads
  - Supported UKHSA team in increase capacity from 2,000 samples to 25,000 samples per week.
  - Contact - UKHSA Technology/HPC team
- Other customers include UK Biobank, global pharmaceuticals, Cancer Research organisations, research & teaching universities and radiology & pathology services

WHOLE GENOME SEQUENCING

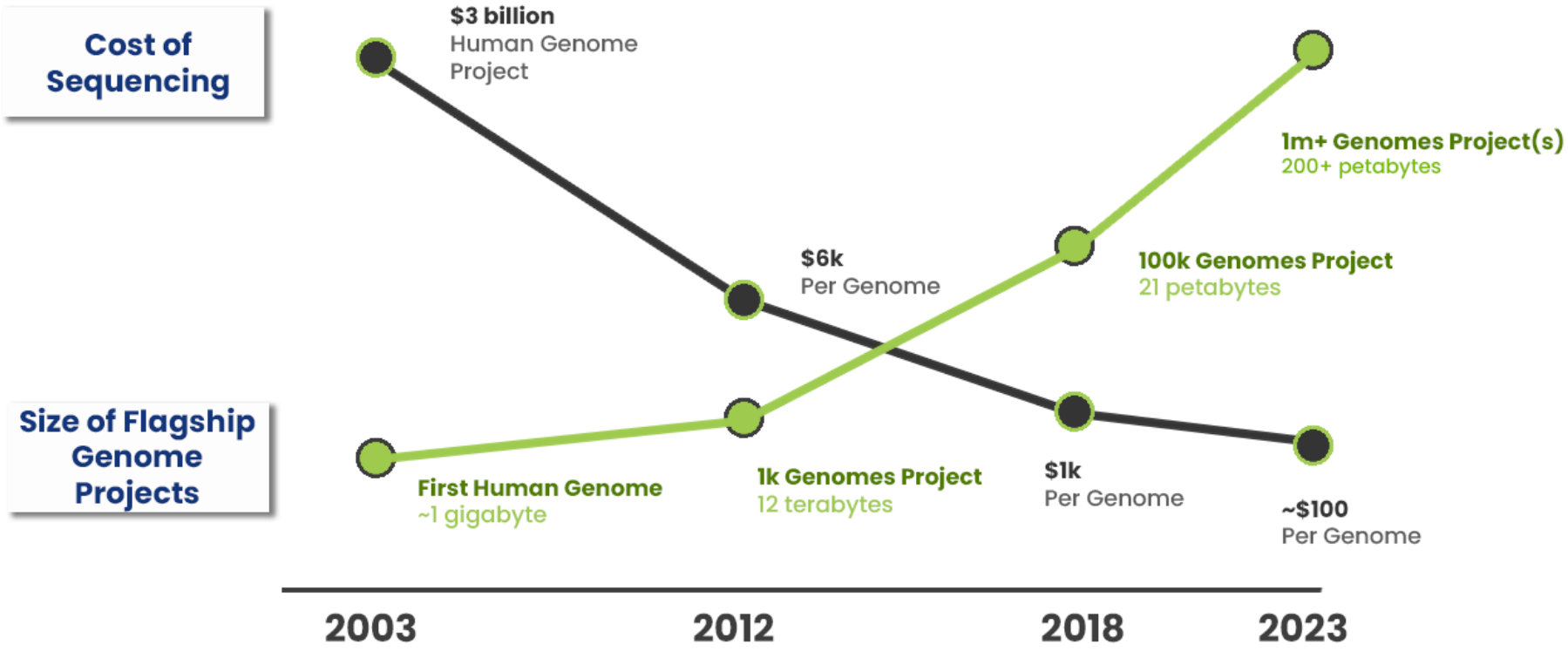
VOLUME OF GENETICS DATA

- Human whole genome generates roughly 100 gigabytes of raw data.
- 40 exabytes is estimated to store whole human genome data by 2026.
- Cost of sequencing decreasing volumes of sequencing data are exponentially increasing.



FLAGSHIP PROJECTS

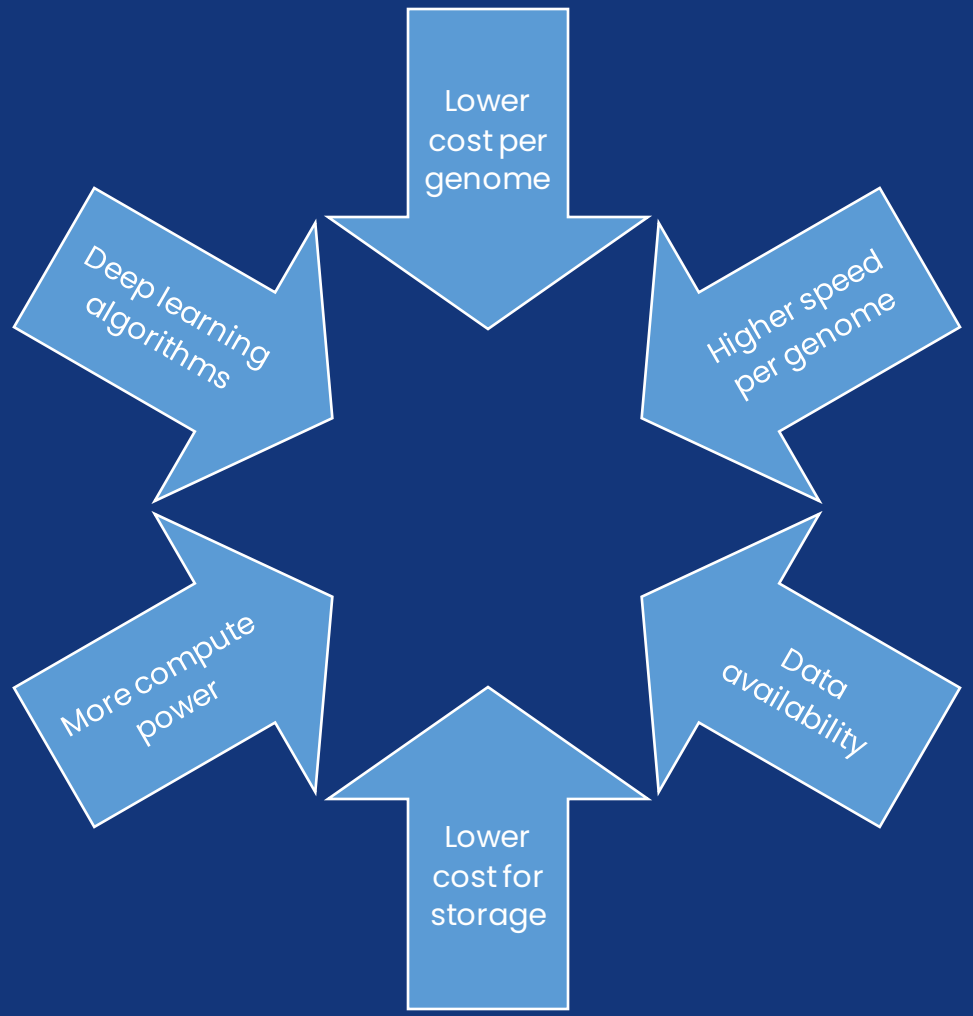
OVER THE YEAR



ROLE OF

AI & TECHNOLOGY

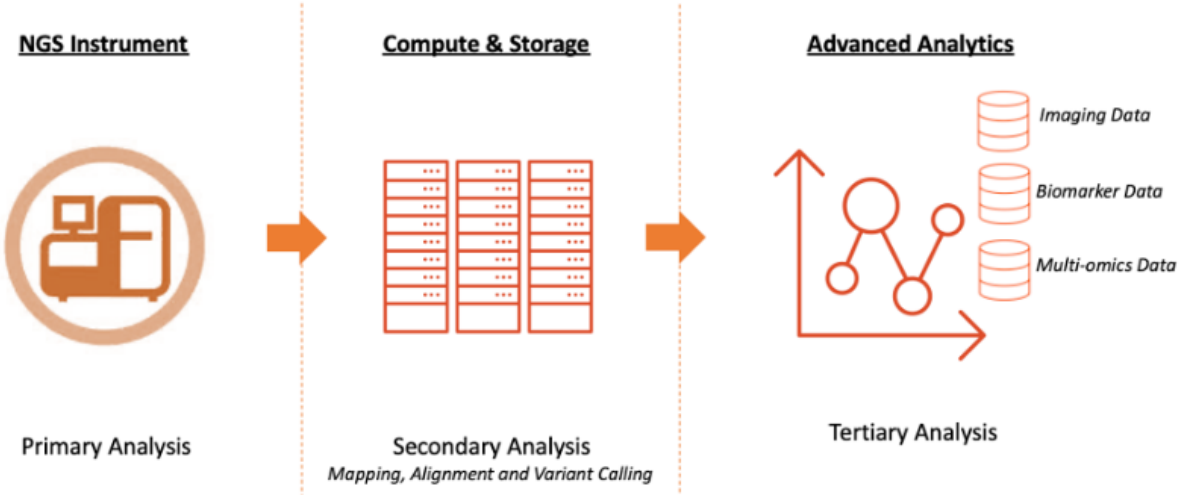
- ✓ Actionable insights on large data (Identify and act)
- ✓ Availability of parallel computing (GPUs)
- ✓ Vast training dataset
- ✓ Programming Languages(Python)
- ✓ Deep Learning algorithms (Neural networks)
  - Convolutional Neural Network (CNN)
  - Recurrent Neural Network (RNN)
  - Denoising Auto Encoder (DAE)
  - Deep Belief Networks (DBNs)
  - Long Short-Term Memory (LSTM)

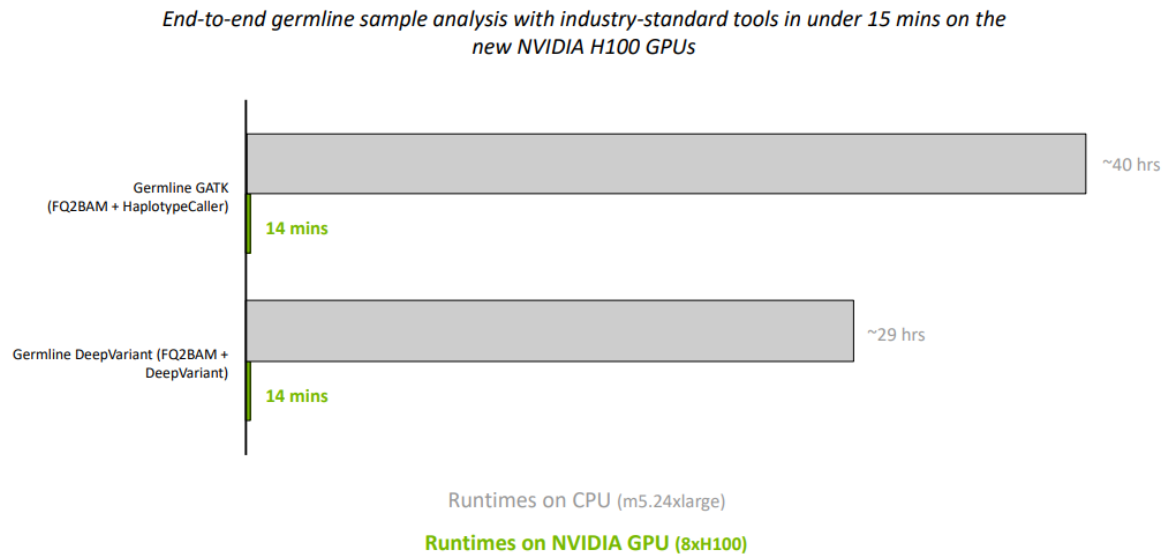


CONVERGENCE OF

AI & WGS

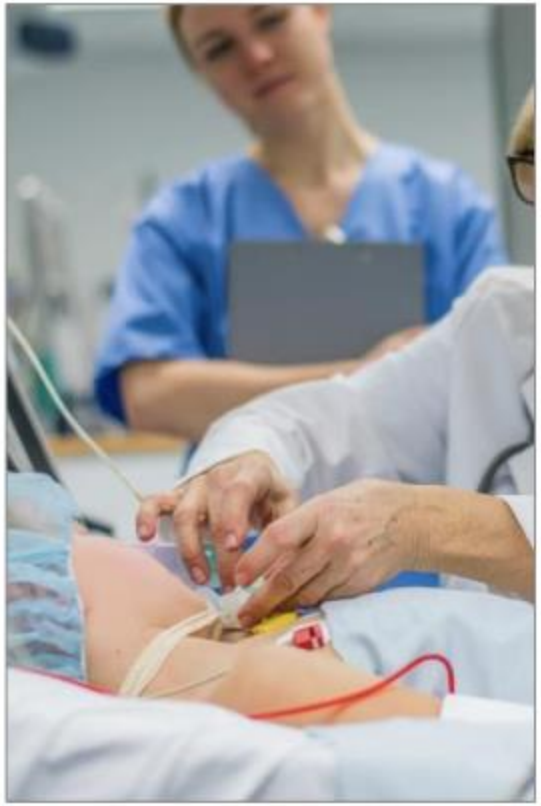
- Base calling using Recurrent Neural Network (RNN) and convolutional neural network (CNN).
- Broad Institute's GATK — a genome analysis toolkit for germline variant calling
- Google's DeepVariant increase accuracy of variant calls.
- NVIDIA GPU-accelerated Parabricks.





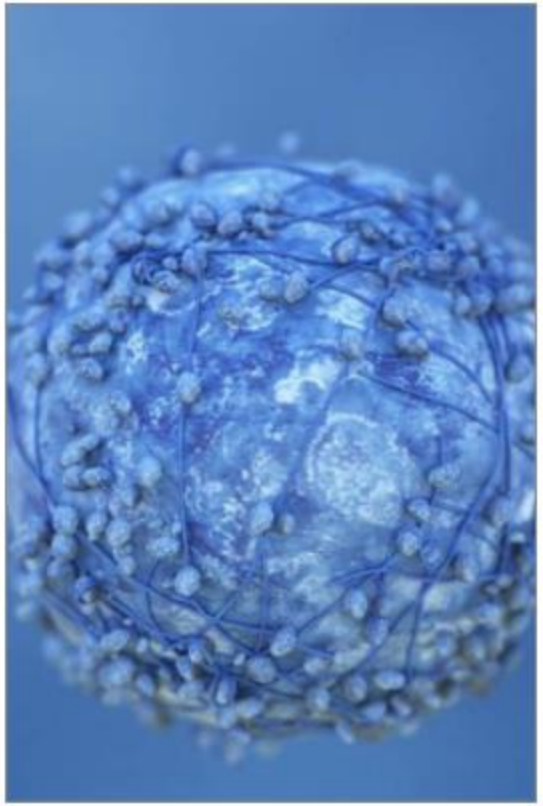
ACCELERATED GENOMIC WORKFLOWS

SAVING LIVES, COSTS & TIME



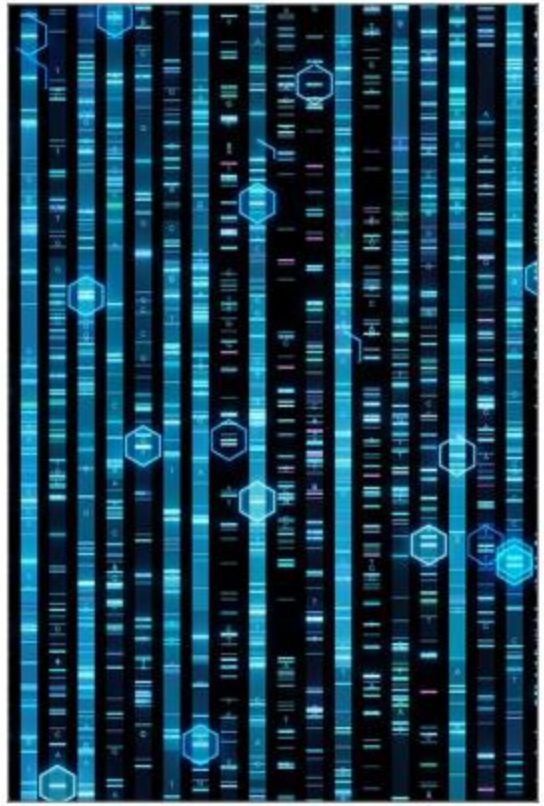
Critical Care Patients

Diagnosing rare diseases in newborns and children



Cancer Genomics

Identifying genetic causes of cancer and matching patients with treatments based on variants



Large Population Studies

Sequencing thousands of genomes and exomes to identify common genes associated with diseases

Research

Fast turn around times for critical research studies and publication deadlines

USE CASES
NVIDIA CLARA PARABRICKS

" As demonstrated by Regeneron, GPU acceleration with NVIDIA Parabricks achieves the throughputs, speed and reproducibility needed when processing genomic datasets at scale. There are a number of research groups in the U.K. who were pushing for these accelerated tools to be available in our platform for use with our extensive dataset."

— Mark Effingham, deputy CEO of UK Biobank.

REGENERON ANALYZES EXOMES OF OVER 500,000 PATIENTS FROM UK BIOBANK, SAVING TIME & COST

Challenge

With more than 100 trillion genotypes generated since 2015, Regeneron Genetics Center (RGC) is a large-scale, fully integrated genomics program that builds on Regeneron's longstanding strength in genetics and related technologies. The program supports applications from target discovery pipelines through public-industry collaborations, including the UK Biobank and others. More than 1.2 million exomes have been processed with this analytical pipeline.

Goal is to glean insights about germline variants from exomes- the protein coding part of the genome- and work with UK Biobank to create a genome database that will be utilized by research institutions, pharmaceutical companies, and university hospitals.

Solution

RGC used NVIDIA Parabricks, a GPU-accelerated software suite for secondary genomic analysis of NGS data. Specifically, they implemented DeepVariant within Parabricks as the foundation for generating scalable, high-quality data that can be reproduced across the genomics community.

Parabricks' GPU-accelerated genomic analysis software contains accelerated industry standard tools such as GATK and BWA-MEM plus deep learning tools such as DeepVariant.

Their previous 32-vCPU-based genomic analysis tool was taking over 1 hour to sequence 1 exome. With Parabricks, Regeneron is now able to analyze an exome in under 5 minutes, while also reducing cost by approximately 40 percent.

NVIDIA Parabricks
Accelerated genomic analysis supporting long and short read DNA

<5 mins For analysis of 1 exome with Parabricks vs 60 minutes with CPU tools

40% Reduction of cost of genomic analysis with Parabricks vs. Previous CPU tools

NVIDIA

***UKHSA has infrastructure to support similar projects**

" ...We're talking about being able to get an answer before the end of a hospital ward round. That's a dramatic jump."

— Euan Ashely, MD DPhil, Professor of Medicine & Biomedical Data Science, Stanford

WORLD RECORD DNA SEQUENCING TECHNIQUE HELPS FIND ANSWERS FOR CRITICAL CARE PATIENTS IN JUST 5 HOURS

Challenge

Whole genome sequencing is not part of standard clinical care due to its cost, traditional long turnaround times, and need for expertise across sequencing, bioinformatics and variant interpretation.

Critical care patients can benefit from whole genome testing early to potentially uncover the genetic cause of a disease and match the patient with the right treatment early. Cutting down the turnaround for a genetic diagnosis from a couple weeks to just a few hours can provide clinical teams with answers needed to treat patients.

Panels are the standard of care and take 2 weeks to deliver results and often do not always contain the right variants.

Solution

A team lead by Dr. Euan Ashely from Stanford with collaborators from Oxford Nanopore Technologies, Baylor College of Medicine, Google, UC Santa Cruz and NVIDIA, created an ultra-fast DNA sequencing technique for whole genome analysis.

The team achieved a record speed by optimizing every stage of the sequencing workflow. They used high throughput nanopore sequencing on Oxford Nanopore's PromethION flow cells to generate more than 100 gigabases (100 Billion nucleotides) of data per hour. NVIDIA technology accelerated base calling inside the PromethION as well as variant calling with NVIDIA Parabricks AI and accelerated software suite for secondary analysis.

This technique helped uncover a variant causing epilepsy in a 3 month-old baby as well as a boy with heart failure. Results published in NEJM and Nature Biotechnology 2022.

NVIDIA Parabricks
Accelerated genomic analysis supporting long and short read DNA

NVIDIA A100 GPUs
Accelerated base calling of Dorado AI model in Oxford Nanopore Technologies genomic sequencer PromethION

~5 Hours Ultra-Fast Whole Genome Sequencing Technique

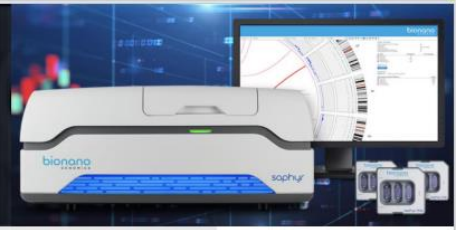
Download ultra-fast pipeline

NVIDIA

" Bionano is focused on continuous upscaling of our end-to-end genome mapping solutions, and we see this collaboration with NVIDIA as an important part of that effort...The solution we're developing with NVIDIA will accelerate OGM analysis and enable future expansion into areas including CAR T-cell therapy research and bioprocessing."

— Erik Holmlin, PhD, president and chief executive officer of Bionano Genomics

bionano



CONCLUSIONS

BRILLIANCE. TOGETHER.

01 AI is not all about ChatGPT

02 AI is proven with real benefits for life sciences

03 Further innovation and research

04 It's a complicated mix – Experienced Partners, collaboration and consultancy is key for success

05 Cost per genome gets cheaper

06 Compute and storage affordability

OCF

BRILLIANCE. TOGETHER.

Q&A

THANK YOU!

0114 257 2200 ocf.co.uk info@ocf.co.uk

OCF Limited, Unite 5 Rotunda Business Centre,
Thorncliffe Rd, Chapeltown, Sheffield, S35 2PG