Microsoft Genomics

Microsoft's perspective for pharmaceutical and life sciences

Better experiences. Better insights. Better care.

Industry scenarios

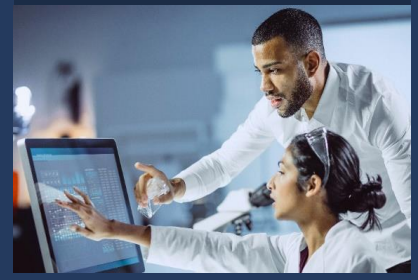
Empower health consumers

Accelerate scientific innovation

Accelerate scientific innovation

Enhance workforce experience

Build operational agility



Customer outcomes

Engage directly and securely with health citizens to drive market efficiency and reduce costs

BD

Modernize discovery, R&D, and improve clinical trial management

NOVARTIS

Protect sensitive health data to support privacy and effective security end-to-end

ZUELLIG PHARMA
making healthcare more accessible

Equip your workforce with skills and tools needed to streamline processes and remove barriers to increased collaboration

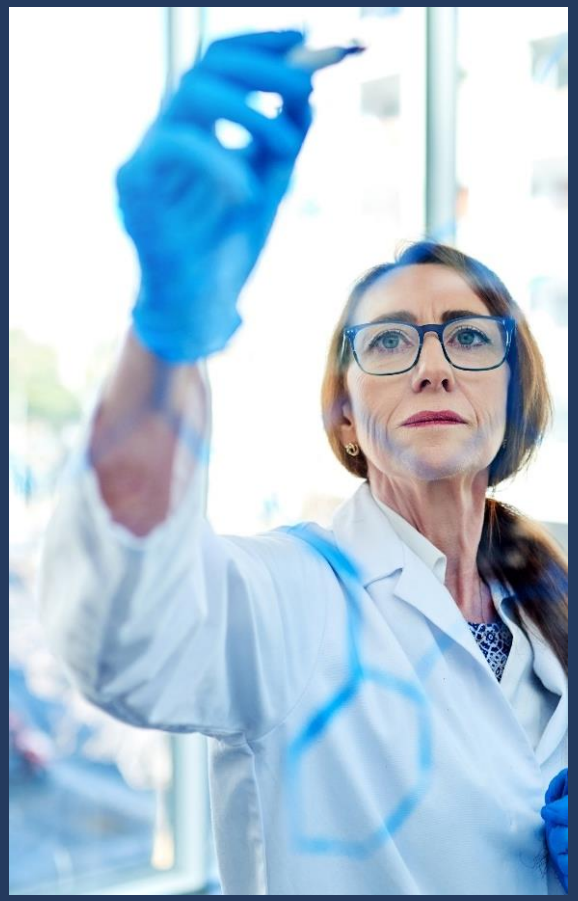
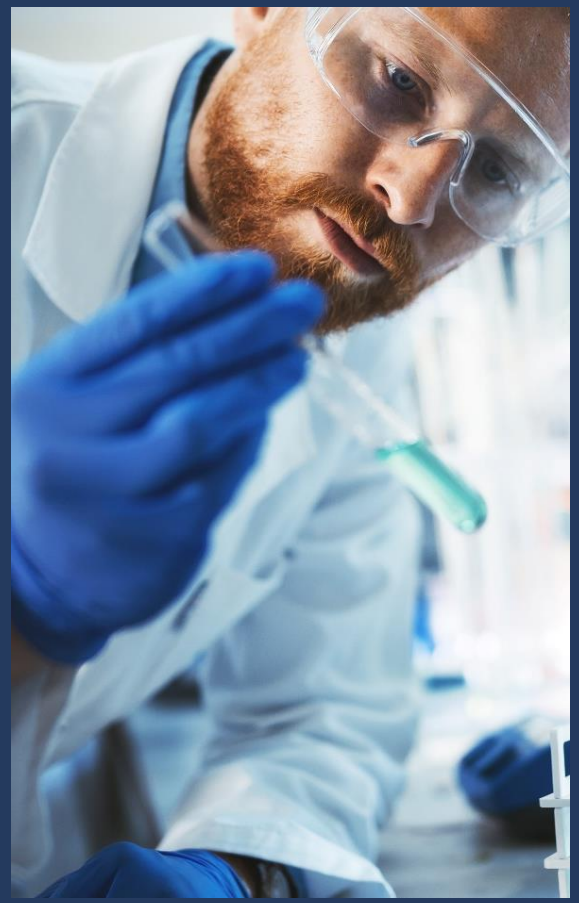
Allergan

Use secure actionable insights to transform operations, increase productivity and reduce time-to-market

stryker

Accelerate scientific innovation

Modernize discovery, R&D, and improve clinical trial management


The future of healthcare is powered by genomics

Precision medicine

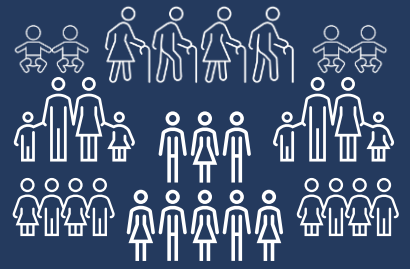
- Early detection
- Disease prediction
- Personalized treatment
- Targeted drugs
- Gene therapy

Research institutions	Academic Medical Centers	Pharma & Biotech	Provider Systems
Fundamental research	Clinical research	Drug development	Clinical decision support

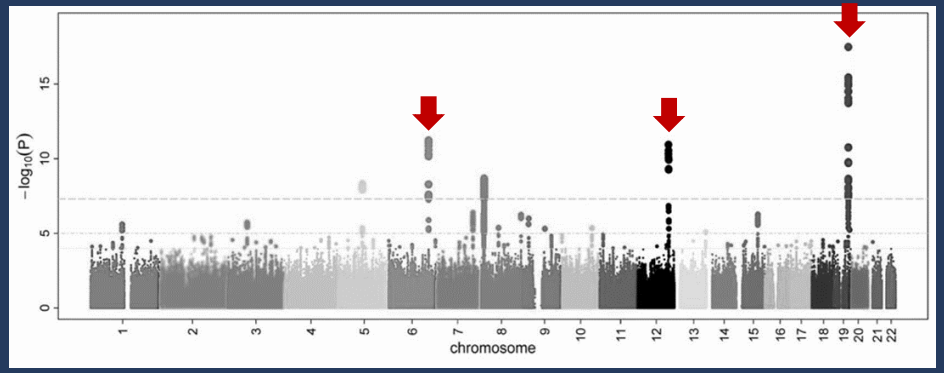
Gene—Disease association requires population scale sequencing

Sequence individual genome

~ 4 Million variations in persons genome



Sequence genomes for large cohorts



Uncover insights
Curate knowledge base
Build disease models

Sequence individual genome

Disease models and knowledge base

Genomics at scale is about big data & hyperscale compute

						New cancer cases in US every year	Childbirth in US every year
Number of people	1	1,000	10,000	100,000	1,000,000	1,800,000	3,800,000
Whole genome storage size	100 GB	100 TB	1 PB	10 PB	100 PB	180 PB	380 PB
Compute power for data analysis	2000 core hrs.	2 M core hrs.	20 M core hrs.	200 M core hrs.	2 B core hrs.	3.6 B core hrs.	7.6 B core hrs.

On-prem → Cloud

Technology is driving the democratization of genomics

Next-generation sequencing…

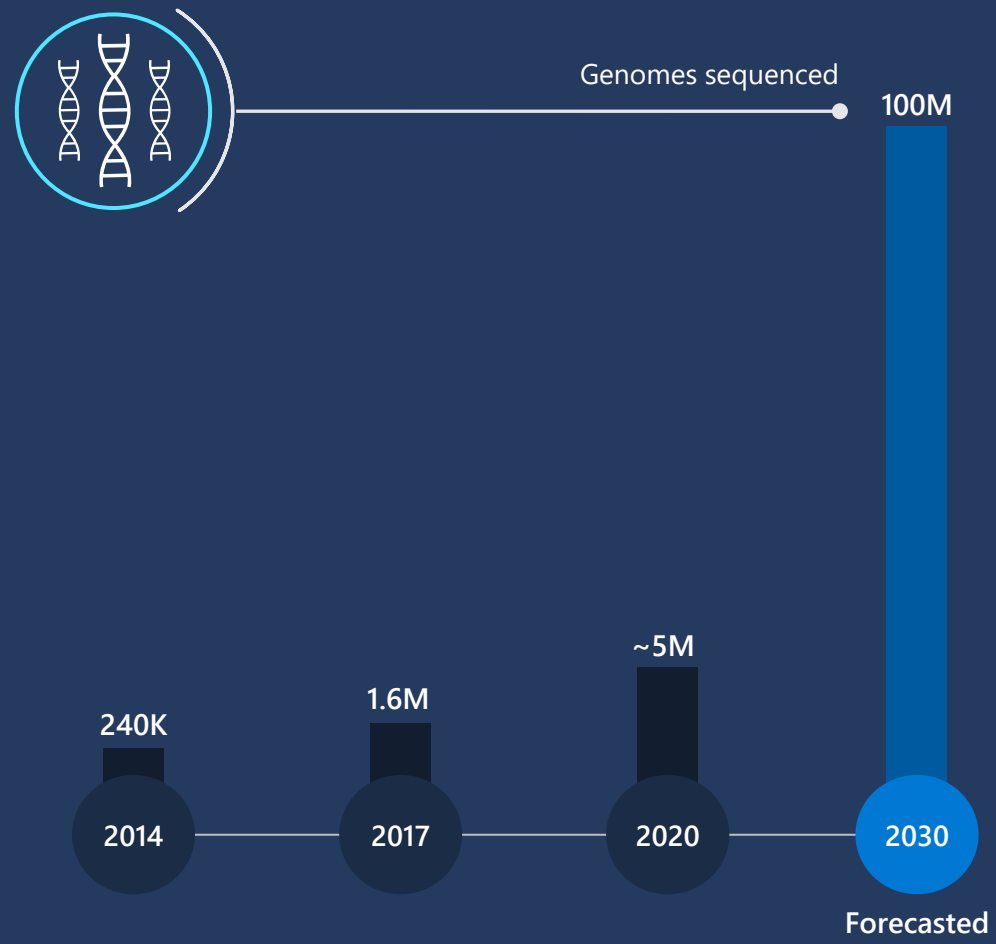
Higher throughput

Better precision

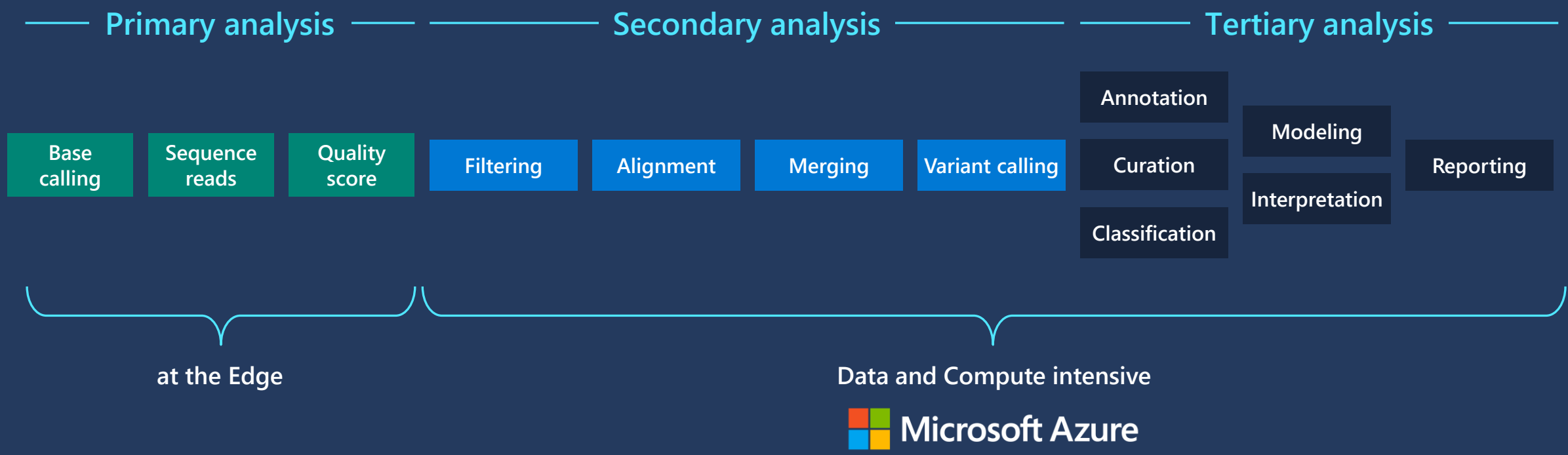
Lower cost

Drug development…

Clinical application…



Genomic analysis workflow—30,000-foot view



Microsoft Azure

Microsoft enables genomics work at all levels

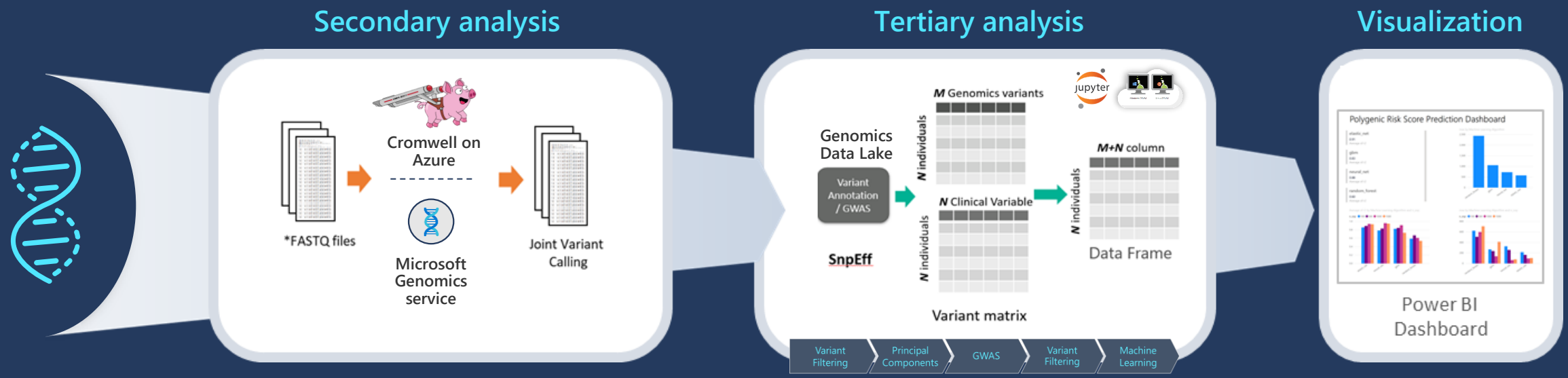
Focus areas	Solution approach
Research & discovery	Collaboration platform with easy access to open-source and community developed bioinformatics and data science tools to support genomics R&D on Azure
Automation & scale	Workflow management platform to enable rapid prototyping, scaling and automation of pipelines for large scale genomics analysis on Azure
Clinical genomics	Turnkey genomics variant calling service on Azure that is secure, compliant (ISO certified, HIPPA compliant), and covered under Microsoft BAA

Enable innovation | Support open ecosystem | Promote interoperability

Microsoft enables genomics work at all levels

Focus areas	Microsoft genomics solutions
Research & discovery	**Genomics Notebooks for Azure** preconfigured Jupyter notebooks **Bioconductor tools on Azure** container with bioinformatics & data science tools **Genomics Data Science VM** preconfigured Azure VM templates **Genomics Data Lake** collection of public genomics datasets
Automation & scale	**Cromwell on Azure** workflow & task execution engine
Clinical genomics	**Microsoft Genomics service on Azure** turnkey secondary analysis pipeline

End-2-end genomics analysis workflow on Azure



AZURE VM

AZURE BATCH

AZURE CONTAINER SERVICE

docker

AZURE BLOB STORAGE

DATA LAKE STORE

DATA FACTORY

SQL DB

SQL DW

DATA SCIENCE VM

AZURE ML

POWER BI

jupyter

Biotia

Biotia leverages Cromwell on Azure to rapidly deploy a novel COVID-19 detection genomics workflow

Customer Biotia (biotia.io) works with hospitals and researchers to leverage high-complexity tests and artificial intelligence for rapid precision infectious disease diagnostics to guide patient treatment and improve health outcomes

Challenge Biotia's novel genomics workflows are data rich and computationally intensive, needing to deliver results timely while ensuring proper version control for compliance

Solution Biotia has deployed their novel genomics workflow with Cromwell on Azure orchestrating the massive parallelization to satisfy HPC demands.

Benefits

- Efficient parallelization to address growth demand
- Version control

Biotia

> *"At Biotia, we have achieved substantial parallelization, thorough version control, and novel COVID-19 detection results by using Cromwell on Azure to back our compute-intensive genomics workflows. We are pleased to include Cromwell on Azure in our bioinformatics software stack."*

— **Joe Barrows**
Head of Software Engineering
Biotia

St. Jude Children's Hospital

Microsoft, St. Jude, DNAnexus join forces to fuel pediatric cancer research

Customer St. Jude Children's Research Hospital is consistently rated as one of the leading pediatric cancer hospital by U.S. News & World Report.

Challenge Secure platform for sharing and collaborative analysis of petabyte size genomics datasets, essential in the quest to discover cures for pediatric cancer.

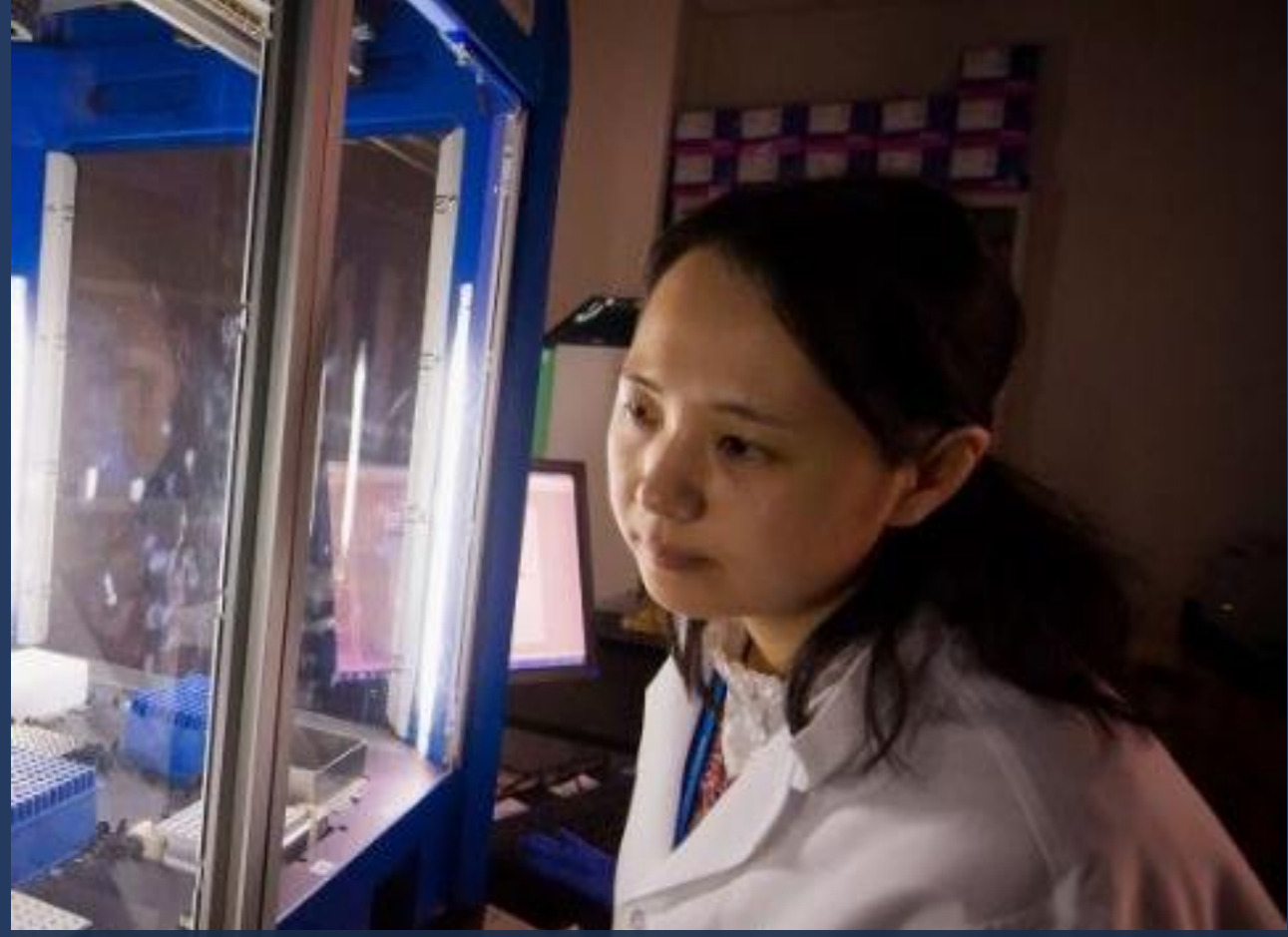
Solution St. Jude Cloud, world's largest repository of pediatric cancer genome data (~2 petabytes) developed by scientists at St. Jude Children's Research Hospital using Microsoft Genomics service and DNAnexus and hosted on Azure.

Benefits

- Consistent, high quality data streamed to the cloud
- Proven genomic processing pipeline for clinical genomics application providing speed, scalability, and ease of use
- Researchers can easily access the data and powerful tools to enable novel discovery.

St. Jude Children's Research Hospital

DNAnexus



"Access to high-quality clinical genomic data, generated leveraging Microsoft Genomics service and streamed to St. Jude Cloud, will help further research in precision medicine for childhood cancer and other diseases."

— **Dr. Jinghui Zhang**
Chair of the St. Jude Department of Computational Biology
St. Jude Children's Research Hospital

Canadian Department of Fisheries & Oceans

Microsoft enables accelerated genomics research with profound effects for fish conservation and overcoming challenges presented by human behavior and climate change

Customer Department of Fisheries and Oceans Canada (DFO) is the federal lead for safeguarding & managing Canada's fisheries, oceans and freshwater resources.

Challenge The sheer size of genomic data combined with limited computational resources led to significant barriers to leveraging whole genome sequencing data at population scale.

Solution DFO deployed Cromwell on Azure and migrated their genomics workflow for variant identification to the cloud.

Benefits

- Consistent, high quality data
- Significantly accelerated analysis time
- Easy to use cloud-based reproducible data analysis and sharing solution

> *Leveraging Cromwell on Azure for running our genomics pipelines give us the ability to scale our analysis to 1000's of fish genomes in an automated fashion. We can essentially eliminate 3 months' time of manual work generating and connecting the genomic variant data with other data sources we have.*

— **Dr. Tony Kess,**
Visiting Scientist,
Belford Institute of Oceanography

Precision medicine

Clinical insights at population scale
Take advantage of cloud-based compute, storage, and data analysis to gain actionable research insights

Customization of healthcare
Tailor medical decisions, practices, or products to individual patients

Prevent future health problems
by using analytics to identify patient risks before symptoms are manifested

DNAnexus

PRECISIONFDA
Tap into a community platform for NGS assay evaluation and regulatory science exploration.

BC

BC PLATFORM
Enable clinicians and researchers to combine and use information seamlessly.

SOPHiA™

SOPHIA FOR GENOMICS
Clinical-grade solutions to detect genomics associated with cancers & hereditary disorders.

2bPrecise

PRECISION HEALTH PLATFORM
Data is delivered in a meaningful vocabulary and actionable format.

<epam>

EPAM CLOUD PIPELINE
Scripts & workflows that support genomics analysis, modeling & simulation, and ML.

PACBIO®

SMRT SEQUENCING PLATFORM
Sequencing technology providing the most comprehensive view of genomes, transcriptomes, and epigenomes.

Get more from your data with open integration