Accelerating Research & Discovery on UK Biobank's Research Analysis Platform
with DNAnexus & NVIDIA

Research Analysis Platform
Powered by **DNAnexus**
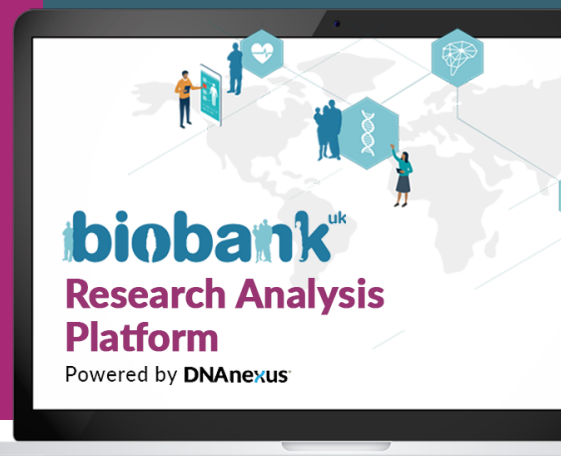
The UK Biobank

The UK Biobank is the world's most comprehensive biomedical data resource, including de-identified medical, health, and lifestyle records from more than 500,000 volunteer participants in the UK. The UK Biobank undergoes ongoing data collection, coupled with electronic health records, that has generated over 800,000 data files and more than 10 million genetic variants.

The collective aim of this wide scale data collection is to help approved researchers from around the world to better understand, prevent, and treat a wide range of diseases. It is currently being used by more than 25,000 researchers from more than 90 countries and it has been cited in more than 3,500 peer-reviewed scientific publications, and counting.

The Clara™ Parabricks Solution on DNAnexus

The UK Biobank has made NVIDIA Clara™ Parabricks on DNAnexus available to their researchers in support of the latest release of a 450,000 whole exome data set. This solution allows any UKB registered researcher to analyze the exomes at a high speed with the same custom DeepVariant model used by Regeneron Genetics center. Researchers can use the protocol for their own projects, or verify the exome variants found by Regeneron.

The Clara™ Parabricks solution and exome data set is available on the UKB Research Analysis Platform (RAP), which was built in partnership with DNAnexus to be scalable, secure, and collaborative. Enabled by DNAnexus technology and powered by Amazon Web Services (AWS), the RAP provides approved researchers with the ability to access and analyze the 11 petabytes of data in the UK Biobank database from anywhere in the world.

> *"By enabling the use of these tools within the Platform, UK Biobank researchers are able to rapidly create reproducible workflows, harmonize the data and advance scientific discovery."*
>
> *Mark Effingham, Deputy CEO, UK Biobank*

> 500k participants
> 450k Whole Exome Release
> 25k Registered Researchers

NVIDIA.

Clara™ Parabricks
GPU-Accelerated DeepVariant

Fast: Runs up to 60x faster than CPU-only solutions

Accurate: Provides identical results across a suite of variant callers

Low Cost: Reduces computing costs up to 50%

DNAnexus®
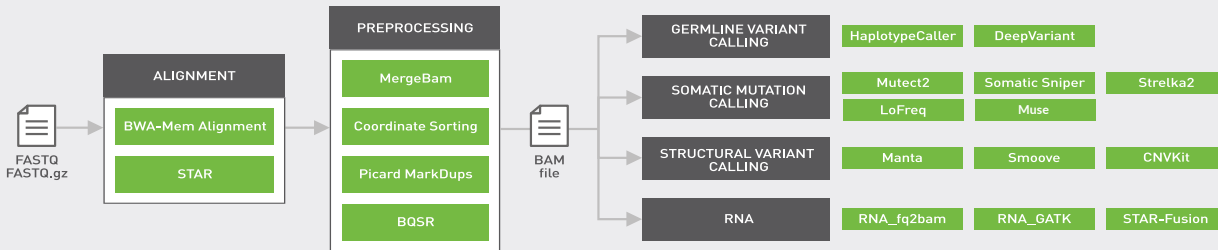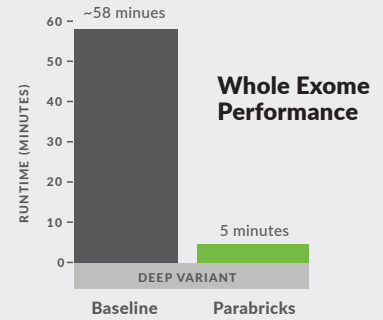
UK Biobank
Research Analysis Platform

Scalable: Balances immediate needs with long-term growth

Secure: Provides a comprehensive security, quality, and privacy framework for biomedical informatics and data management

Collaborative: Allows scientists to appropriately share data, collaborate and drive impactful research

NVIDIA Clara™ Parabricks is a software suite for secondary genomic analysis of NGS data, that delivers results quickly, with high accuracy at low costs. Running Clara™ Parabricks on DNAnexus, researchers can analyze a whole exome in under 5 minutes, or a 30x whole human genome data set in only 22 minutes. Additionally, Parabricks lowers the cost of analysis while providing equivalent results.
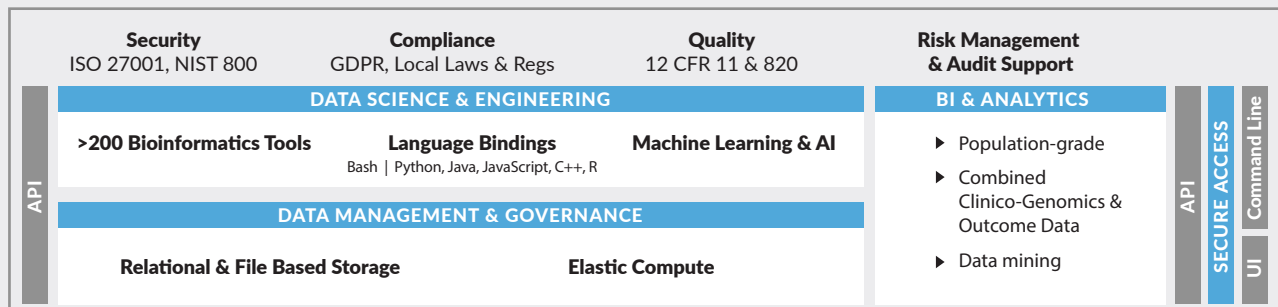
Whole Exome Performance

RUNTIME (MINUTES)

~58 minues
5 minutes

DEEP VARIANT

Baseline   Parabricks

FASTQ
FASTQ.gz

ALIGNMENT
- BWA-Mem Alignment
- STAR

PREPROCESSING
- MergeBam
- Coordinate Sorting
- Picard MarkDups
- BQSR

BAM file

GERMLINE VARIANT CALLING
- HaplotypeCaller
- DeepVariant

SOMATIC MUTATION CALLING
- Mutect2
- Somatic Sniper
- Strelka2
- LoFreq
- Muse

STRUCTURAL VARIANT CALLING
- Manta
- Smoove
- CNVKit

RNA
- RNA_fq2bam
- RNA_GATK
- STAR-Fusion

More tools available:

| JOINT GENOTYPING | ImportGVCF, SelectVariants, GenotypeGVCF |
| QUALITY CHECKING | Alignment Summary, InsertSize, SequencingArtifact, GcBias, QualityScoreDistribution, WGSMetrics, RAWWGSMetrics, BaseDistributionByCycle, MeanQualityByCycle |

| VARIANT PROCESSING | VQSR, Variant Filtration, Select Variants, CNNScore Variants |
| VARIANT QC | Variants can be annotated, filtered, and QC'd using NVIDIA generated tools, including merging multiple VCFs |

DNAnexus

DNAnexus is the leading cloud-based biomedical informatics company that facilitates secure access and utilization of biomedical data while supporting collaboration across internal and external teams. DNAnexus' Platform serves organizations across — biopharmaceutical, clinical diagnostics, healthcare, academic research, and government in 48 countries with compliant protection of data, privacy, and intellectual property. From delivering integrated diagnostic R&D and production environments, enabling national biobank informatics initiatives to providing pharmaceutical companies with a multi-omics data science platform for new target drug discovery, DNAnexus empowers the healthcare and life sciences industry to transform how they leverage biomedical data to accelerate scientific discoveries and deliver better patient care.

DNAnexus makes it easy to run Clara™ Parabricks by providing an out-of-the-box solution. Users simply supply input files and receive results delivered in their DNAnexus project. Launching Clara™ Parabricks apps is a seamless process, removing the hassle of provisioning GPU instances in the cloud manually or installing CUDA or Clara™ Parabricks apps themselves.

| **Security**
ISO 27001, NIST 800 | **Compliance**
GDPR, Local Laws & Regs | **Quality**
12 CFR 11 & 820 | **Risk Management**
& Audit Support |

DATA SCIENCE & ENGINEERING

>200 Bioinformatics Tools   **Language Bindings**   **Machine Learning & AI**
Bash | Python, Java, JavaScript, C++, R

BI & ANALYTICS
▶ Population-grade
▶ Combined Clinico-Genomics & Outcome Data
▶ Data mining

DATA MANAGEMENT & GOVERNANCE

Relational & File Based Storage   **Elastic Compute**

API

API   SECURE ACCESS   UI   Command Line

Get Started with NVIDIA Clara™ Parabricks & DNAnexus

1 Select your project

2 Select "DeepVariant Pipeline (Parabricks accelerated)" from the Tools Library

3 Click "Run"

OR in your terminal, type: dx run pbdeepvariant_ukb -h

For more information, visit: **www.dnanexus.com/partnerships/ukbiobank**