

White Paper

Accelerating Precision Biomarker Discovery:

Unlocking the Strategic Power of
Ovation Clinico-Genomic Data on the
DNAnexus Platform

Table of contents

1	Executive Summary: The Translational Stack	3
2	The Translational Chasm: Why the 'Plumbing' Delays the Science	3
3	Strategic Growth Drivers: The Joint Ovation & DNAnexus Platform Solution • Unifying Population-Scale Data and Advanced Cloud Analytics for High-Impact Precision Medicine	4
4	Real-World Case Study: Accelerating Biomarker Discovery in GLP-1 Agonist Therapeutics • Identifying Novel Machine Learning Predictors of Glycemic Response in Metabolic Disease • The GLP-1 Cohort is the Proof. The Capability is the Product.	7
5	Key Considerations & Buying Criteria • Essential Capabilities for Enterprise Clinico-Omics Platforms	9
6	Drive Precision Breakthroughs with Ovation and DNAnexus • Transform Your Biomarker Discovery Pipeline Today • About DNAnexus	10

1. Executive Summary: The Translational Stack

Biopharma R&D success is often hindered not by a lack of scientific talent, but by the "Translational Chasm" — the vast technical gap between raw clinical and omic data and actionable drug discovery insights. DNAnexus provides the "Translational Stack", a comprehensive ecosystem that bridges this gap and unlocks the value of **population-scale clinico-genomic** data with advanced cloud analytics. This integrated solution transforms the traditional months-long data engineering slog into a **streamlined analytical sprint**.

By providing a **turnkey environment** where high-depth whole genome sequencing (WGS) and longitudinal real-world data (RWD) are **already harmonized**, the platform allows researchers to move directly from defining a cohort to prioritizing evidence-backed targets. This approach **de-risks clinical development** by providing a deep, multimodal understanding of responder heterogeneity before a trial protocol is locked.

2. The Translational Chasm: Why the 'Plumbing' Delays the Science

The leading cause of late-stage clinical failure is an **incomplete understanding of which patients the drug works for**. While actual scientific analyses in a multimodal real-world evidence (RWE) program might only take weeks, the "**plumbing**"—data feasibility, harmonization, and the manual multimodal merge—frequently **consumes eight months or more**. This delay often results in findings that are real and defensible but **arrive too late** to influence the protocol.

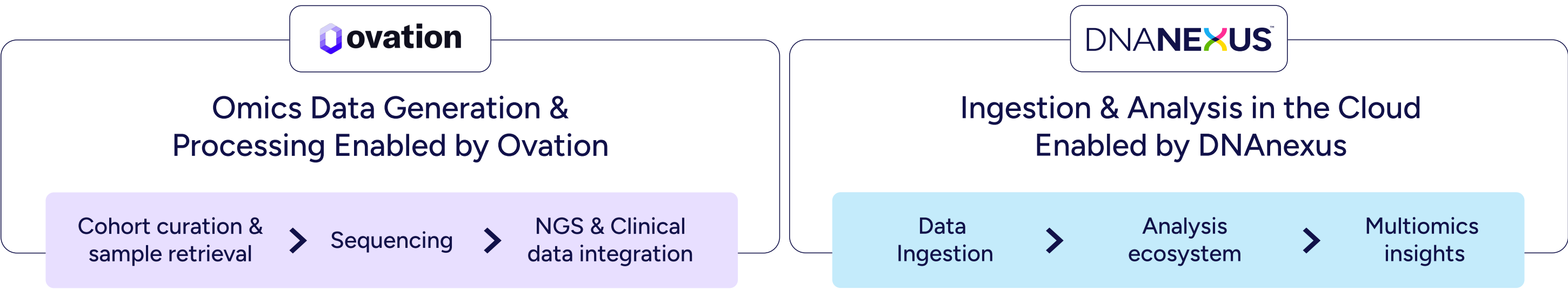


Figure 1: Overview of the Ovation and DNAnexus turnkey clinico-genomic ingestion, enrichment & analytics ecosystem

DNAnexus Trusted Research Environment for multi-omic RWD
Accelerating Translational Pharma R&D Timelines

OMICS DATA MANAGEMENT

Operationalized Pipelines & Workflows

Clinico-Genomic Data Linkage

Omics Data Catalog

Purpose-built ETL for Multi-Omic Datasets

GOVERNED DATA SHARING AND ANALYSIS ECOSYSTEM

Self - Serve Data Exploration

Cohort Browser Feasibility Sandbox

Bioinformatics, Data Science Environments

Multi-Modal Data Analysis Ecosystem

SECURITY

COMPLIANCE

QUALITY

DNAnexus Core Platform

3. Strategic Growth Drivers: The Joint Ovation & DNAnexus Platform Solution

Unifying Population-Scale Data and Advanced Cloud Analytics for High-Impact Precision Medicine

GROWTH DRIVER 1: Scale, Diversity, and Longitudinal Depth of Ovation Data

Ovation delivers targeted biological sample collection and high-throughput data production at population scale by evaluating over **250,000 patient samples monthly** across clinical lab partners. The platform links consented clinical samples directly with **tokenized, de-identified longitudinal clinical data** spanning more than eight years of medical history. This creates a robust foundation for **deep phenotypic and multiomic exploration**.

The flagship Ovation GLP-1 dataset highlights this capability, pairing **30x whole genome sequencing** with longitudinal electronic health records across **8,500 patients**.

Accelerating Precision Biomarker Discovery

4

Crucially, over 50% of the cohort represents non-Caucasian populations, ensuring researchers can validate therapeutic response signals in diverse real-world cohorts. Continuous quarterly data releases and planned **plasma proteomics upgrades** ensure biopharma teams work with expanding, cutting-edge multiomic profiles.

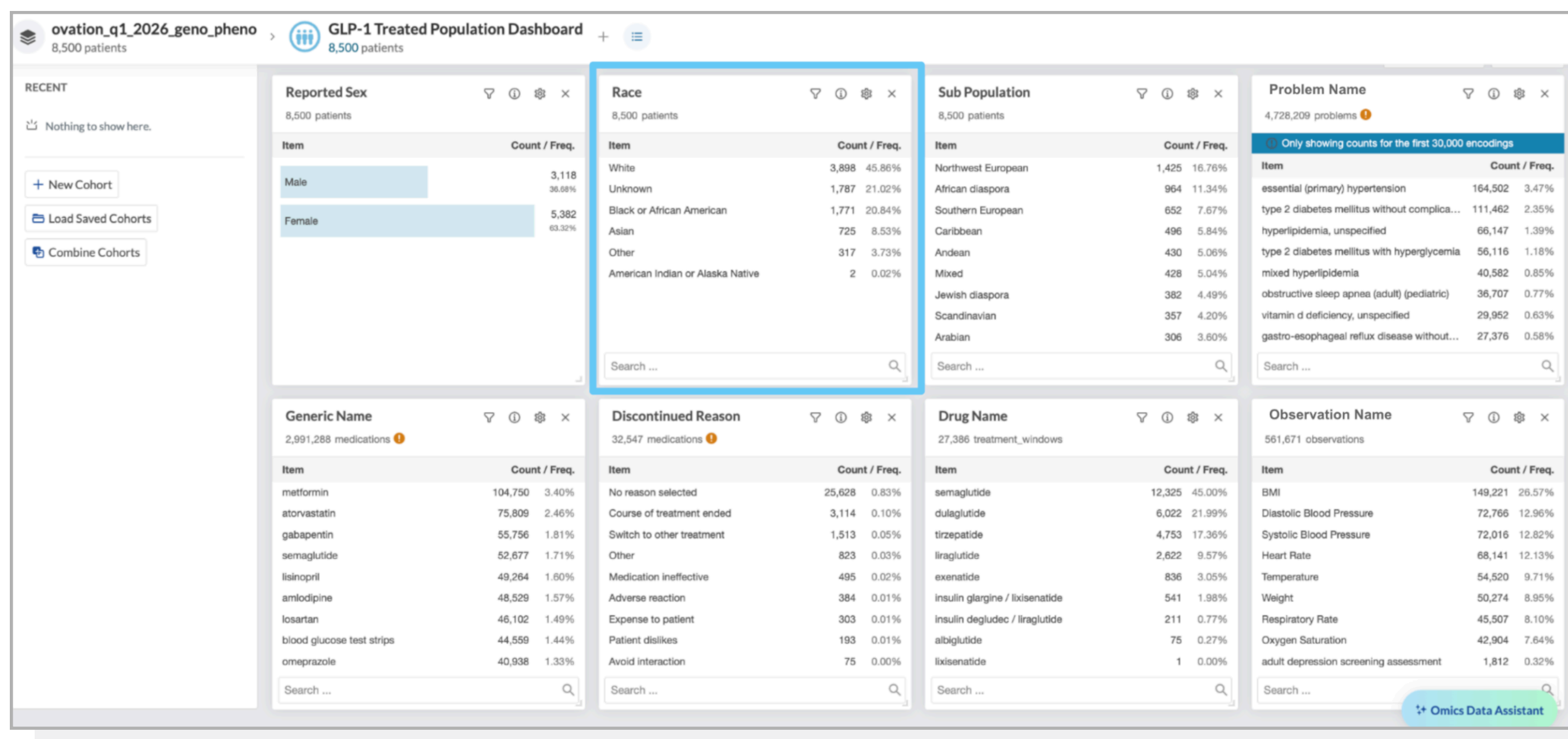


Figure 2: Cohort Browser Dashboard enables rapid query and visualization of longitudinal RWD variables, including distribution by Race, Medications Prescribed, and Comorbidities.

GROWTH DRIVER 2: Zero-Touch Ingestion and Automated Data Engineering

DNAnexus eliminates infrastructure friction through automated ingestion and specialized data orchestration engines. These tools parse raw clinical and genomic feeds into high-performance Spark datasets and Parquet files: **an automated pipeline** that transforms raw source data into **integrated phenotype and germline variant databases** harmonized under a unified semantic schema.

Bioinformaticians and translational researchers explore and collaborate on cohorts using the interactive Cohort Browser or execute natural language queries and data visualizations via the LLM-powered Omics Data Agent (ODA). This seamless abstraction reduces data preparation time from months to hours.

GROWTH DRIVER 3: Integrated Multiomic Workbench, Advanced AI/ML Analytics, Hosted Knowledge Bases

The DNAnexus platform provides a **cloud-native Trusted Research Environment (TRE)** equipped with interactive compute options and a rich bioinformatics tool library. Researchers gain native access to **JupyterLab, RStudio & AutoML Assistant** to run complex statistical models directly alongside

hosted data. This co-location of compute and data **removes the need for risky, expensive data egress**.

The platform natively integrates essential bioinformatic packages enabling scientists to fluidly transition from preliminary cohort building to **genome-wide association studies (GWAS)**, rare variant analysis, and supervised machine learning pipelines. By consolidating tools into a **single compliant workspace**, biopharma organizations **dramatically accelerate time-to-insight**.

GROWTH DRIVER 4: Enterprise-Grade Governance and Regulatory Compliance

DNAnexus stands as the global standard for omics intelligence, supporting **over 60,000 global users** and managing **125+ petabytes of multimodal data**. Trusted by 17 of the top 20 global pharmaceutical companies, the platform satisfies strict enterprise security requirements. It operates in full alignment with **HIPAA, GxP, and 21 CFR Part 11 regulations**.

Granular role-based access permissions and private workspaces ensure complete data protection and an **immutable audit trail** on every query and model run. Cross-functional teams can safely collaborate on sensitive clinico-genomic datasets without compromising patient privacy or corporate intellectual property.

DNAnexus TRE Data Governance & Ecosystem

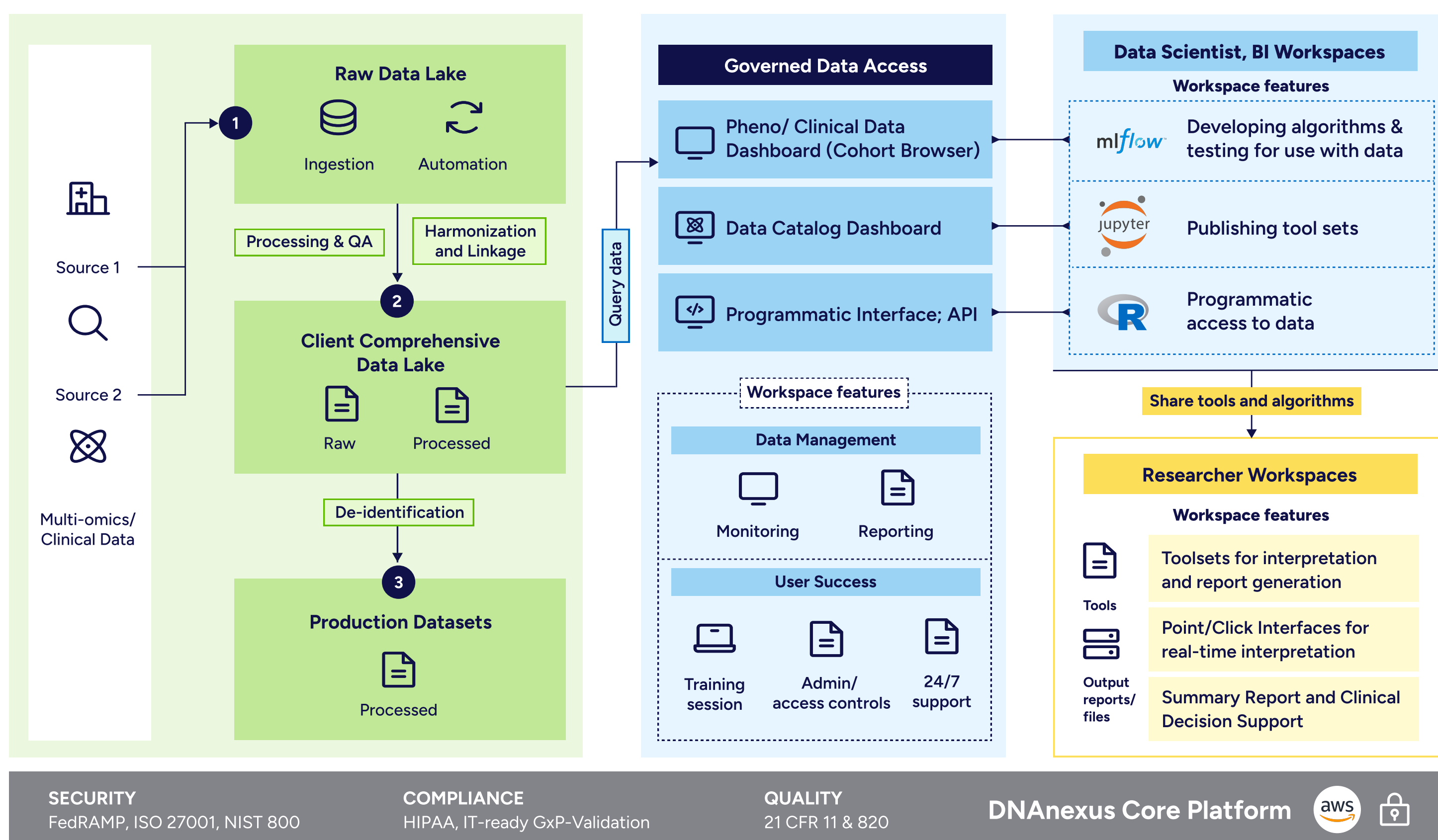


Figure 3: Enterprise data governance & ecosystem architecture within the DNAnexus Trusted Research Environment (TRE)

4. Real-World Case Study: Accelerating Biomarker Discovery in GLP-1 Agonist Therapeutics

Identifying Novel Machine Learning Predictors of Glycemic Response in Metabolic Disease

Problem

GLP-1 receptor agonists and dual GLP-1/GIP agonists have transformed diabetes and obesity management, but **real-world therapeutic response remains highly variable**. Biopharma research teams require precise predictive biomarkers to differentiate responders from non-responders. Establishing these predictors is essential for **refining clinical trial stratification** and **discovering novel drug targets**.

Solution

The DNAnexus Solution Science team evaluated a sub-cohort of **1,441 GLP-1-treated patients** from the Ovation dataset, defining response as a **1.5% reduction in HbA1c at 24 weeks**. The analysis utilized Cox proportional hazards survival modeling and logistic regression to identify strong clinical predictors, notably **baseline HbA1c and agonist class**.

To identify molecular predictors, the team extracted high-depth germline WGS samples to build a 20,158-gene feature matrix. Multiple machine learning classifiers (XGBoost, SVM, LASSO) were trained on managed GPUs, with feature importance quantified via SHAP analysis. This unconstrained genome-wide approach identified novel candidate genes such as *SIK3*, which were subsequently qualified using public evidence bases like Open Targets and GTEx.

Concrete Outcome

- **Clinical Efficacy Signals:** Dual GLP-1/GIP agonists demonstrated a 2.65-fold higher odds of early response compared to single GLP-1 agonists, with baseline HbA1c proving to be the single strongest clinical predictor (OR 2.45 per unit increase, overall clinical AUC = 0.903). Time-to-event survival curves revealed statistically significant response divergence starting at day 140.
- **Genomic Target Prioritization:** SHAP feature analysis identified a distinct panel of top candidate genes and variants where accumulated damage within a gene is suggestive of altered treatment response.
- **Mechanistic Validation:** Biological characterization via Open Targets, GTEx, and other databases available on the DNAnexus platform provided a ranked shortlist of candidates aligned to biological processes with causal genetic and multi-tissue expression support associated with metabolic disorders.

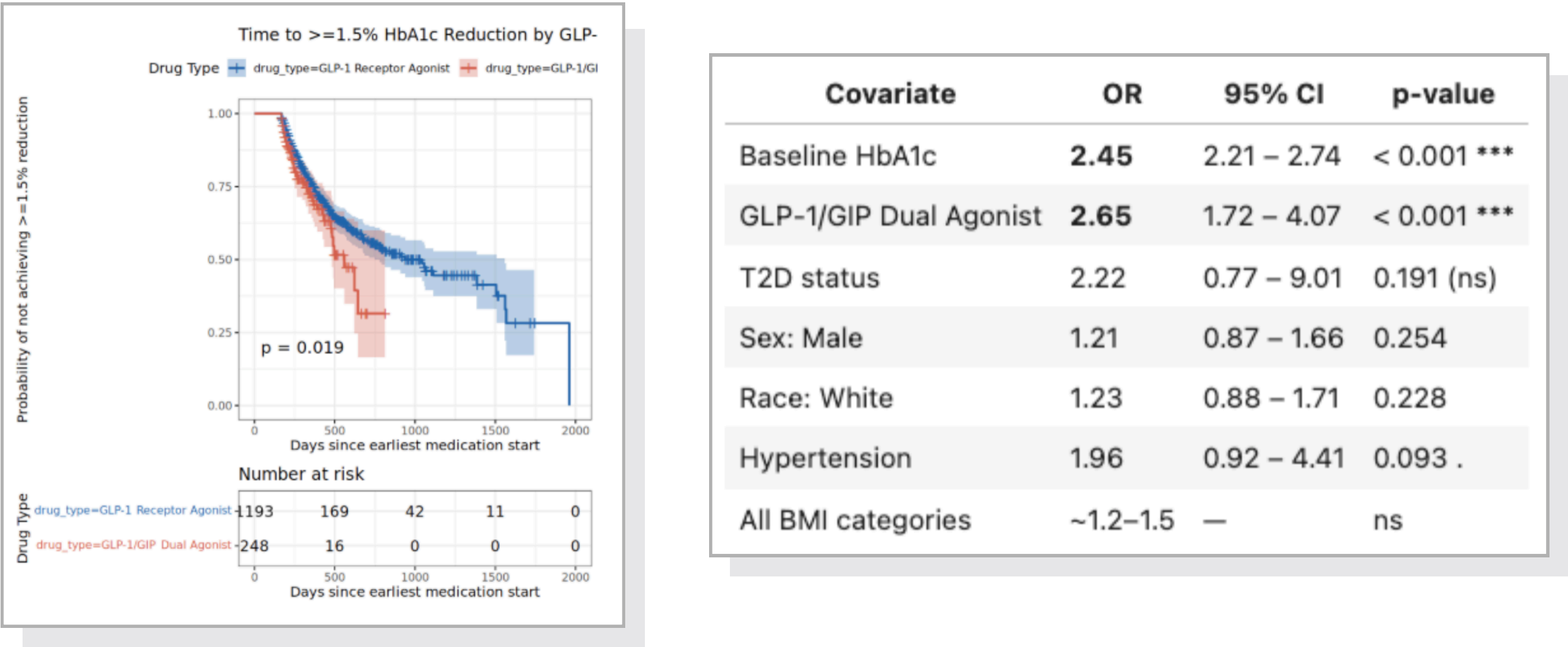
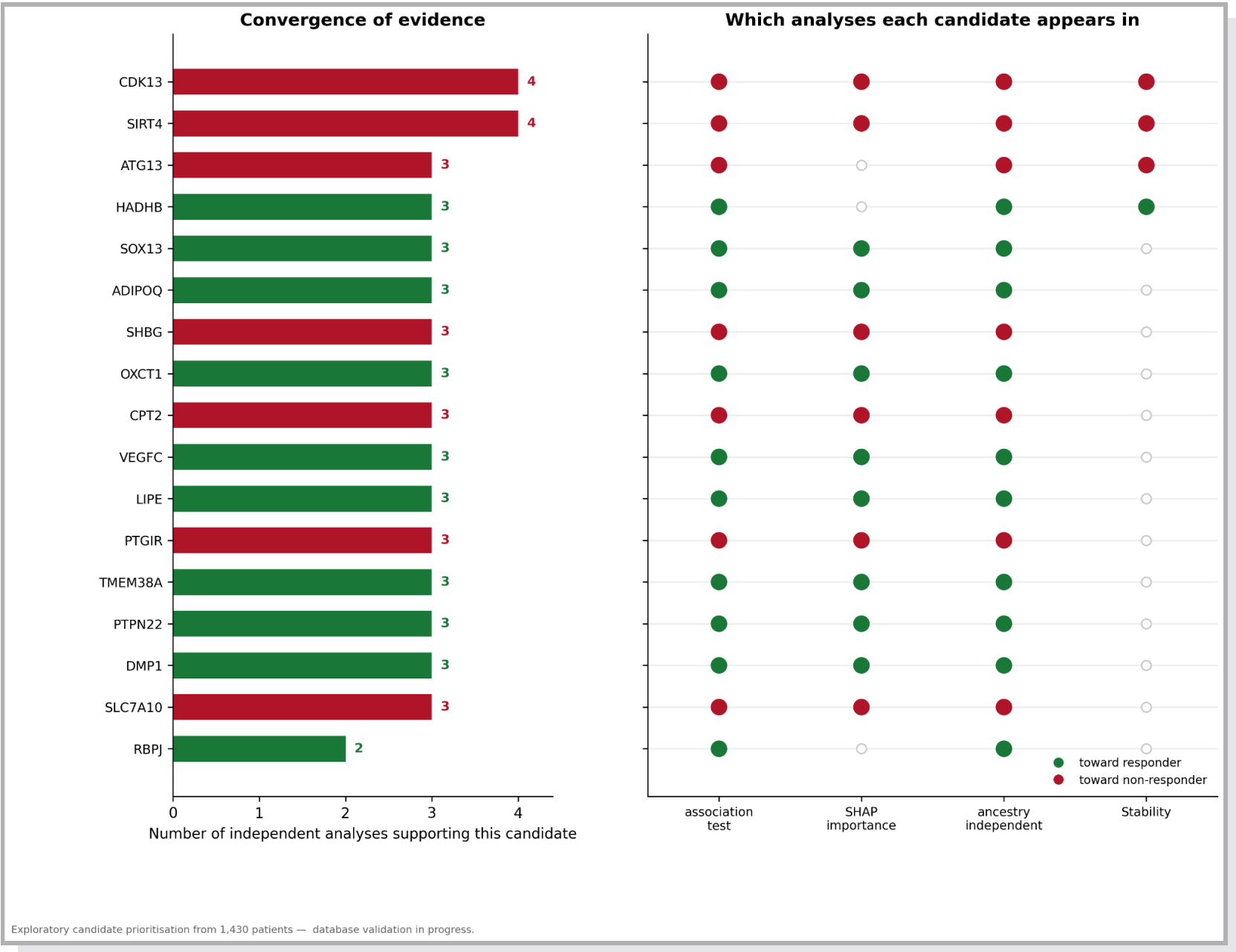
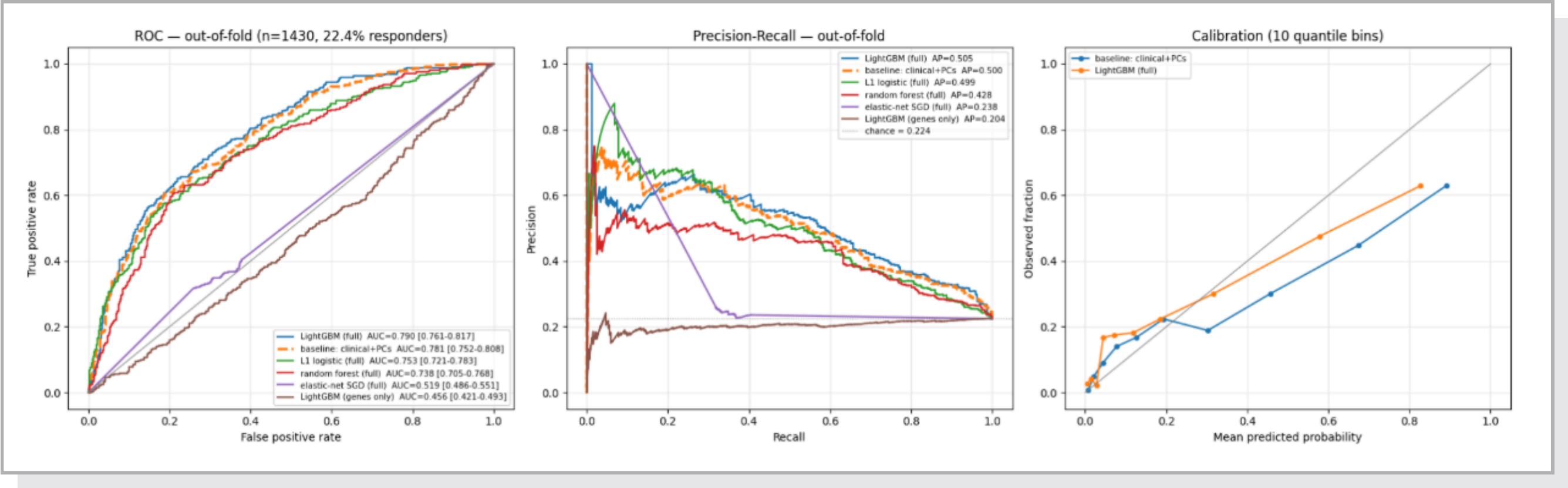


Figure 4: Kaplan-Meier time-to-event plot showing glycemic response rate divergence across GLP-1 agonist treatment groups over time



Exploratory candidate prioritisation from 1,430 patients — database validation in progress.

Figure 5: SHAP feature importance plot highlighting top genomic candidate hits from machine learning classification

The GLP-1 Cohort is the Proof. The Capability is the Product.

While the GLP-1 cohort provides the proof of execution, the underlying capability is universally applicable across therapeutic areas. The same "Translational Stack" moves that decoded GLP-1 response can be applied to complex problems in oncology, such as modeling transcriptomics, radiomics & digital pathology features together to separate responders from non-responders.

GLP-1 Case Study (The Proof)	Translational Application (The Capability)
Longitudinal EHR records joined to 8,500 whole genomes	Any longitudinal clinical dataset joined to any omics modality (WES, RNA-seq, etc.)
Responders separated by HbA1c reduction at 24 weeks	Any endpoint defined from a patient record (RECIST response, PFS, tolerability)
Genome-wide ML across ~20,000 genes	Any high-dimensional feature space (transcriptomics, radiomics, digital pathology)
Hits qualified against Open Targets & GTEx	Any public evidence base (TCGA, UK Biobank, CELLxGENE)

5. Key Considerations & Buying Criteria

Essential Capabilities for Enterprise Clinico-Omics Platforms

When evaluating enterprise solutions for clinico-genomic biomarker discovery, biopharma leaders should prioritize the following core criteria:

- **Integrated Multimodal Scope:** Native support for 30x WGS, RNA-seq, and proteomics tied to longitudinal EHR and claims records.
- **Automated Data Engineering:** Built-in ETL engines that automatically transform raw genomic VCFs and clinical tables into Spark datasets and Parquet files.
- **No-Code Cohort Exploration:** Intuitive web-based Cohort Browsers and AI-driven Omics Data Agents that streamline cohort building for non-programmer researchers.
- **Flexible Interactive Analytics:** Cloud-hosted JupyterLab and RStudio environments pre-configured with statistical genetics suites.
- **Explainable AI Frameworks:** Native support for supervised machine learning models coupled with SHAP feature attribution to accelerate target qualification.
- **Rigorous Regulatory Compliance:** Full enterprise adherence to HIPAA, GxP, and 21 CFR Part 11 within a fully audited Trusted Research Environment (TRE).

6. Drive Precision Breakthroughs with Ovation and DNAnexus

Transform Your Biomarker Discovery Pipeline Today

Unlock the full potential of real-world multiomic data. Contact our team today to explore how Ovation's population-scale datasets and the DNAnexus platform can power your biomarker discovery programs or schedule a live demonstration of our GLP-1 Responder Use Case.

About DNAnexus

DNAnexus sits at the intersection of omics data generation, management, and health innovation. As the global standard for omics intelligence, DNAnexus powers precision medicine for over **60,000 users** across global pharma, biobanks, and health systems, managing **125+ petabytes** of multimodal data in a secure, compliant environment.



DNANEXUSTM

dnanexus.com