

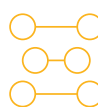
# Unlock Rapid Clinico- Omics Insights with *the GenAI-powered* Omics Data Assistant

## KEY TAKEAWAYS:



- **Accelerate Discovery:** Rapidly translate complex clinico-omics data into actionable insights through intuitive, natural language queries.
- **Empower Non-Technical Users:** Enable researchers, clinicians, and therapeutic area leaders to independently create cohorts and explore datasets without reliance on bioinformatics or data engineering expertise.
- **Collaborative Efficiency:** Leverage the DNAAnexus Trusted Research Environment (TRE) to streamline internal and external collaborations, facilitating real-time, context-rich conversations around datasets.
- **Proven Scalability:** ODA works with thousands of fields and millions of rows, as demonstrated by applying it on existing large-scale datasets hosted by DNAAnexus.

## The Problem:

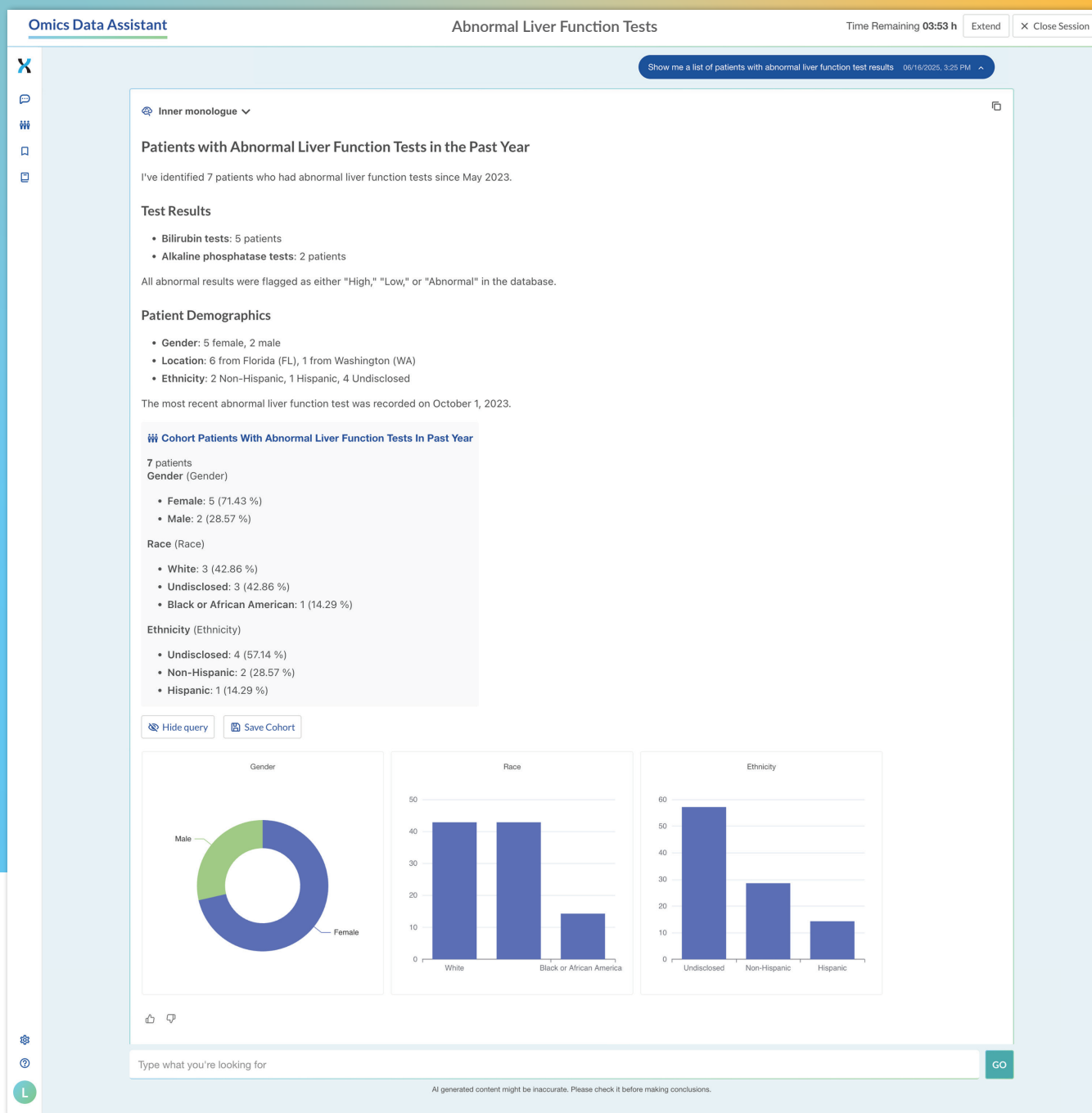


Clinico-omics research generates vast amounts of complex, fragmented data across genomic, proteomic, and other omics modalities. Therapeutic area leaders and innovation teams face significant hurdles when attempting to navigate these complex datasets without deep technical expertise. Reliance on data specialists creates bottlenecks, delaying insights, slowing down decision-making, and impeding collaboration.

## The Need:



The biopharma industry urgently requires solutions that democratize access to intricate clinico-omics data. Therapeutic area heads and innovation leaders need tools that provide rapid insights, enable real-time collaborative exploration, and significantly reduce dependence on specialized technical teams.



## Our Solution:



The DNAnexus Omics Data Assistant (ODA) is a GenAI-powered conversational interface that integrates seamlessly into the DNAnexus Trusted Research Environment. ODA empowers researchers and clinicians to effortlessly query complex datasets in natural language. By leveraging Large Language Models, ODA efficiently handles cohort creation, data summarization, and sophisticated longitudinal and cross-modal queries.

→

## Key Functionalities Include:

- **Natural Language Queries:** Users can ask dataset-specific questions conversationally, drastically simplifying data exploration.
- **Cohort Definition and Refinement:** Intuitively define, adjust, and save cohorts without technical hurdles.
- **Insight Generation:** Automated summaries and actionable insights derived directly from data queries.
- **Longitudinal Data Exploration:** Easily perform complex temporal queries to track and analyze patient outcomes over time.
- **Cross-Modal Analysis:** Seamlessly integrate and interrogate data across multiple omics modalities to reveal comprehensive insights.

Omics Data Assistant

3 prime UTR variant in IL6

Time remaining 03:58 h

Extend

Close Session

Find all of the patients with 3 prime UTR variant in IL6 06/16/2025, 4:45 PM

Inner monologue

### Finding Patients with 3 Prime UTR Variants in IL6 Gene

To find patients with 3 prime UTR variants in the IL6 gene, I've constructed the following query.

This query:

1. Identifies IL6 gene variants (located on chromosome 7, positions 22725884-22732002)
2. Filters specifically for 3 prime UTR variants using the effect field
3. Links these variants through the annotation and genotype tables to patient IDs

The query joins:

- Patient data from participant\_0001 table
- Sample ID mapping from pheno\_geno\_sample\_ids table
- Genotype data from genotype\_alt\_read\_optimized table
- Variant annotation from annotation\_read\_optimized table

Cohort Patients with 3 Prime UTR Variant In IL6

10 patients

Age at recruitment (p21022)

- Average: 60.2
- Minimum: 43.0
- Maximum: 69.0

Sex (p31)

- Male: 6 (60%)
- Female: 4 (40%)

Ethnic background (p21000\_I0)

- British: 9 (90%)
- Irish: 1 (10%)

Hide query

Save Cohort

RAW SQL

```
SELECT DISTINCT p.eid
FROM apollo.ukbrap_synth_pheno_100k_db.participant_0001 p
JOIN apollo.ukb_pheno_100k_db.pheno_geno_sample_ids pg ON p.eid = pg.p_sample
JOIN apollo.ukb_synth_geno_100k_db.genotype_alt_read_optimized g ON pg.g_sample = g
JOIN apollo.ukb_synth_geno_100k_db.annotation_read_optimized a ON g.a_id = a.a_id
WHERE a.gene_name = 'IL6'
AND a.chr = '7'
AND a.bin = 9
AND array_contains(a.effect, '3_prime_UTR_variant')
LIMIT 5;
```

USE CASE

EXAMPLE

Data element discovery

Does the dataset include ethnicity information?

Clinico-genomic cohort building

Build a cohort of those diagnosed with type 2 diabetes with frameshift GCK variants.

Longitudinal (time axis) querying

Find those who were admitted for myocardial infarction and then within a month were diagnosed with GERD.

SQL-based analytics

For patients that took vedolizumab, which lab results changed the most? Only include lab tests recorded within 3 months of taking vedolizumab.

Type what you're looking for

AI generated data might be inaccurate. Check it before making conclusions.

DNAexus © 2025



## Your Benefits and Key Outcomes:

- **Reduced Technical Barriers:** Non-technical therapeutic leaders easily access, explore, and analyze data, promoting self-sufficiency and efficiency.
- **Enhanced Collaboration:** Foster context-rich collaboration across internal teams and external partners within the secure DNAnexus TRE.
- **Accelerated Time-to-Insight:** Swiftly move from data query to meaningful insight, expediting research cycles and enhancing competitive advantage.
- **Optimized Resource Allocation:** Shift bioinformatics resources away from routine data queries towards high-value tasks, maximizing the strategic use of specialized talent.
- **AI/ML Integration:** Easily port ODA-generated cohorts and interrogate them using DNAnexus AI/ML tools in a secure, closed-box system, ensuring compliance and protecting sensitive data.
- **Secure and Compliant Exploration:** Maintain robust data security and compliance standards while maximizing data utility and insights in a trusted environment.



**Join leading biopharma organizations** in transforming your clinico-omics research process. Contact DNAnexus today to learn more about Omics Data Assistant and how the DNAnexus platform can help you start unlocking your data's full potential.

**[dnanexus.com](https://dnanexus.com)**