BulkRNA-Agent

An AI-Powered Agent for Automated Bulk RNA-Seq Data Analysis

A comprehensive, interactive web interface for transcriptomics analysis—from Quality Control to Enrichment—powered by a Dual LLM system.

Intelligent Automation for Transcriptomics

Workflow Automation

Handles the full pipeline from raw count matrices to biological insights. Automates the transition from data ingestion to actionable results.

Interactive Interface

Built on Gradio, offering a user-friendly GUI rather than just CLI scripts. Accessible to researchers without deep coding expertise.

Hybrid Analysis

Combines standard statistical rigor (`DESeq2/PyDESeq2`) with AI-driven interpretation. Uses a **Dual LLM Advantage** for reasoning versus knowledge.

Key Outputs: Quality Control reports ✤ Differential Expression tables ✤ Enrichment Analysis (GO/KEGG/Reactome) ✤ Interactive visualizations

The Architecture: A Dual LLM System

The agent automatically detects intent and routes queries to the most capable model.

User Query / Data Context

Router

Reasoning Engine

GPT-OSS (`gpt-oss:20b`)

- Tool selection
- Analysis planning
- Statistical decision-making
- Code execution logic

Biomedical Engine

Biomistral
(`cniongolo/biomistral`)

- Gene function
- Pathway contextualisation
- Biological interpretation

Integrated Response

Synthesized Biological Insights

The Analysis Workflow

From raw CSV data to actionable biological insights in five streamlined steps.

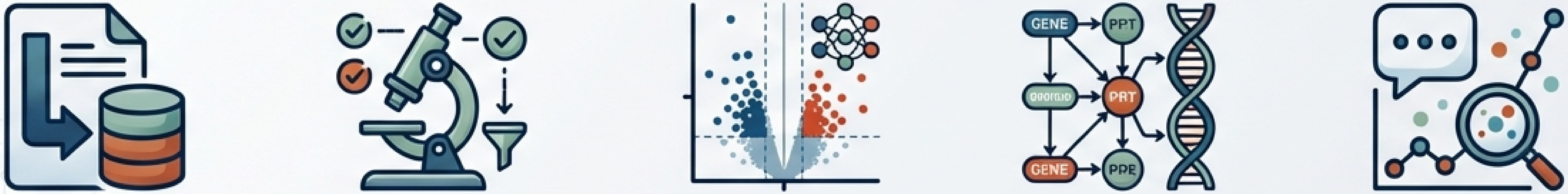
1. Input
Upload Counts & Metadata

2. Quality Control
Filter & Visualise

3. Differential Expression
AI Design & Execution

4. Enrichment
Pathway Analysis

5. Exploration
Chat & Interactive Plots



AI-driven analysis pipeline. Supports raw count data from CSV. Generates comprehensive reports & interactive tools.

Data Requirements and Input

Web Interface Upload

Drag & Drop files here
or click to browse

Supported Formats: CSV, TSV

File 1: Count Matrix

Format: CSV/TSV | **Structure:** Genes (rows) × Samples (columns)

```
gene_id, sample1, sample2...
ENSG001, 200, 150...
ENSG002, 0, 12...
```

File 2: Sample Metadata

Format: CSV/TSV | **Structure:** Samples (rows) × Conditions (columns)

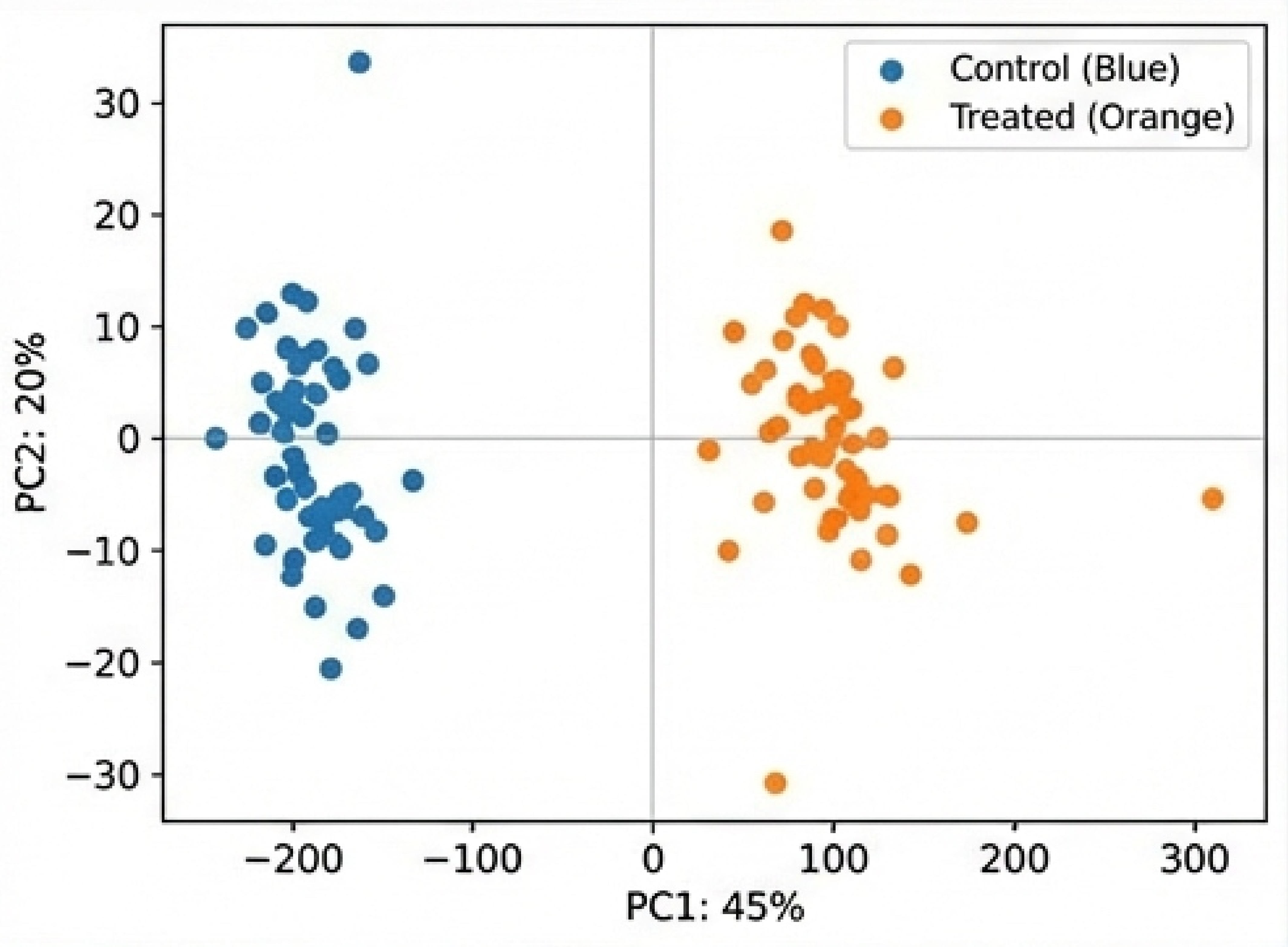
```
sample_id, condition, batch
sample1, control, A
sample2, treated, A...
```

Supports custom metadata columns (e.g., Batch, Treatment) essential for complex experimental designs.

Automated Quality Control

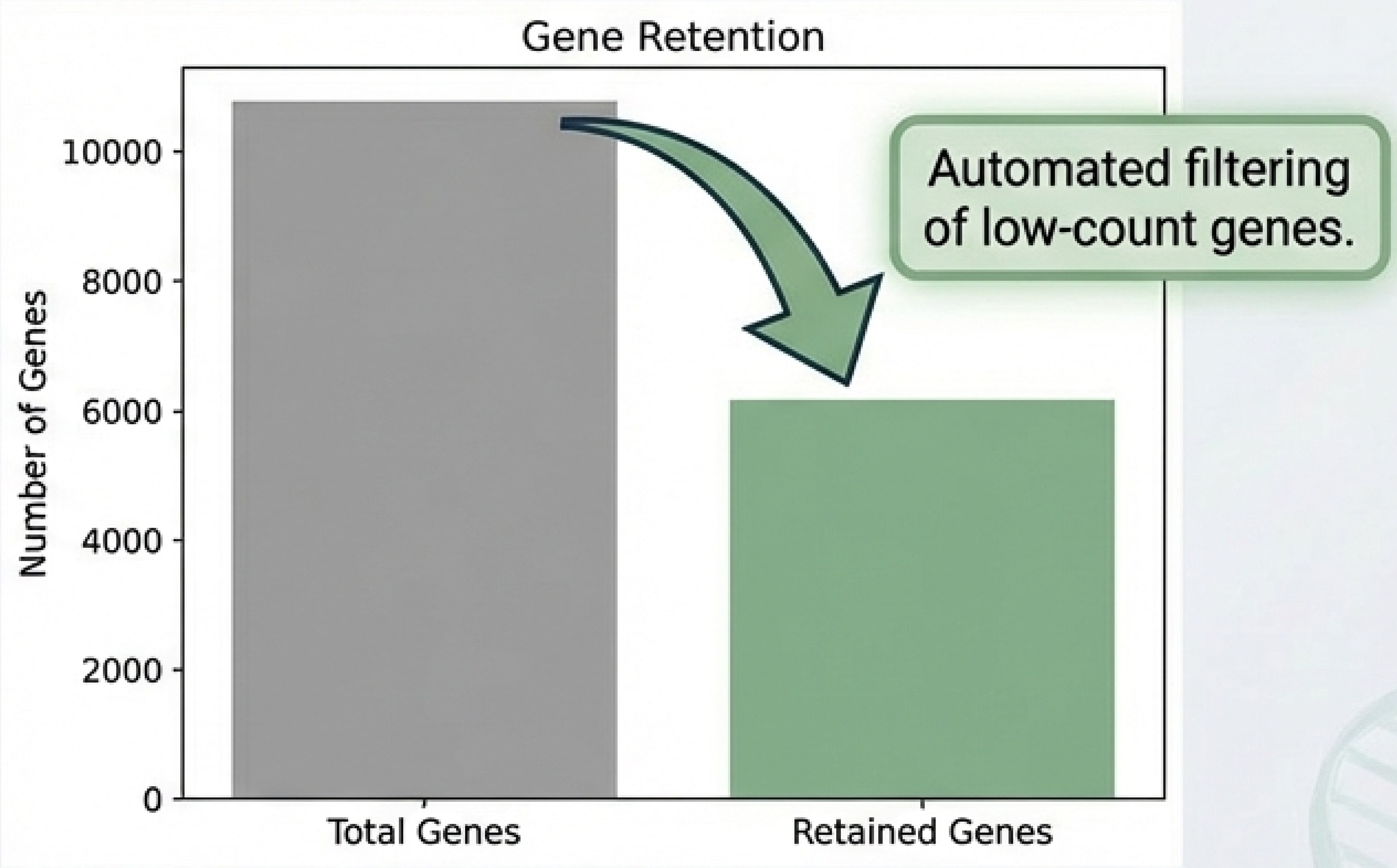
Ensures statistical analysis is performed only on high-quality, relevant data.

PCA Plot



Principal Component Analysis visualises sample clustering and detects outliers.

Filter Chart



Automated filtering of low-count genes.

Integration: Leverages **RNAseqQC** for standardized metrics.

AI-Assisted Differential Expression

The Engine

Core Engine: **DESeq2 / PyDESeq2**

Industry standards for differential expression analysis.

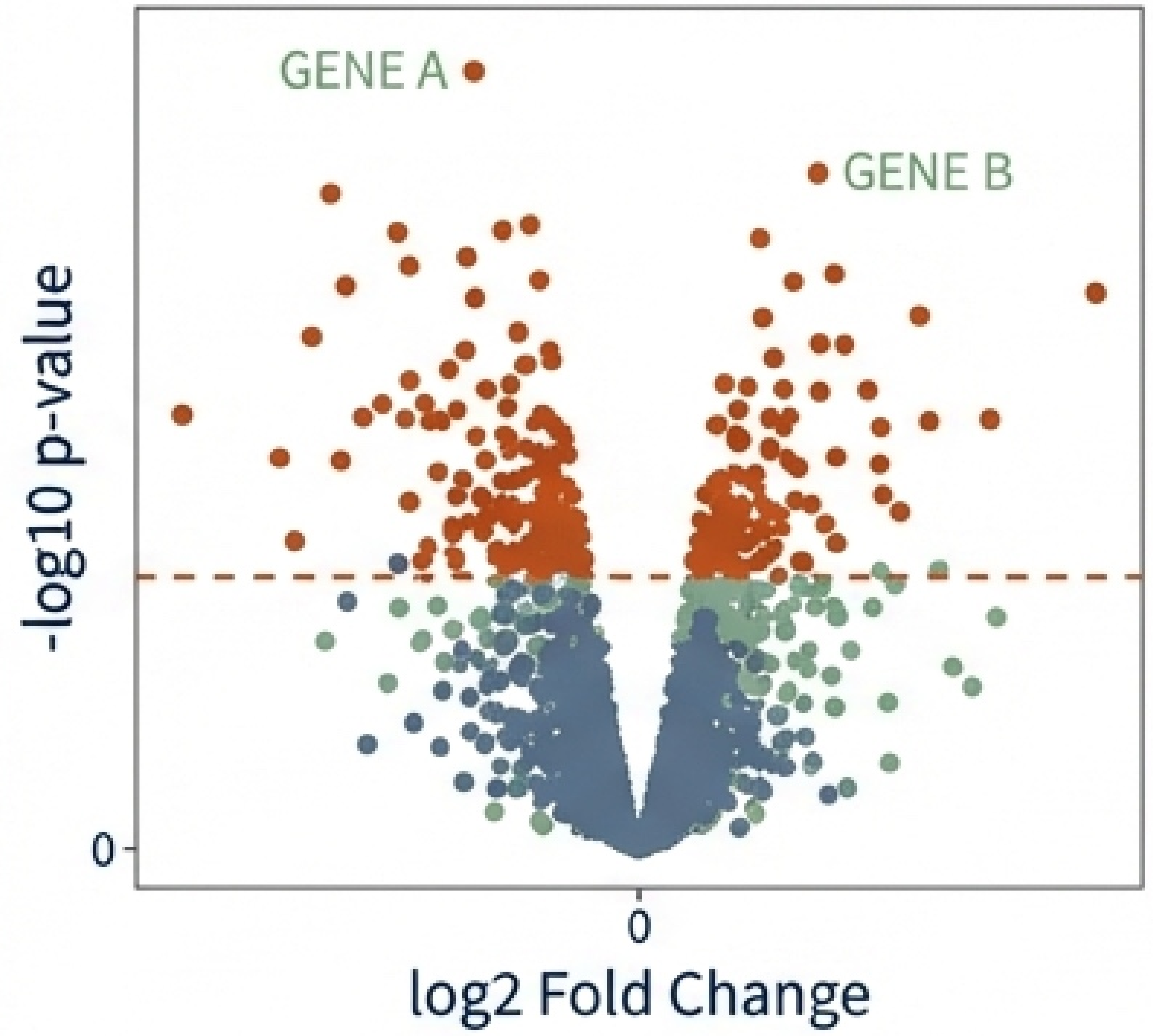
Fira Code

```
design_formula = '~ batch + condition'
```

Inter **Roboto**

Based on your metadata, I suggest using '~ batch + condition' to account for batch effects.

The Output

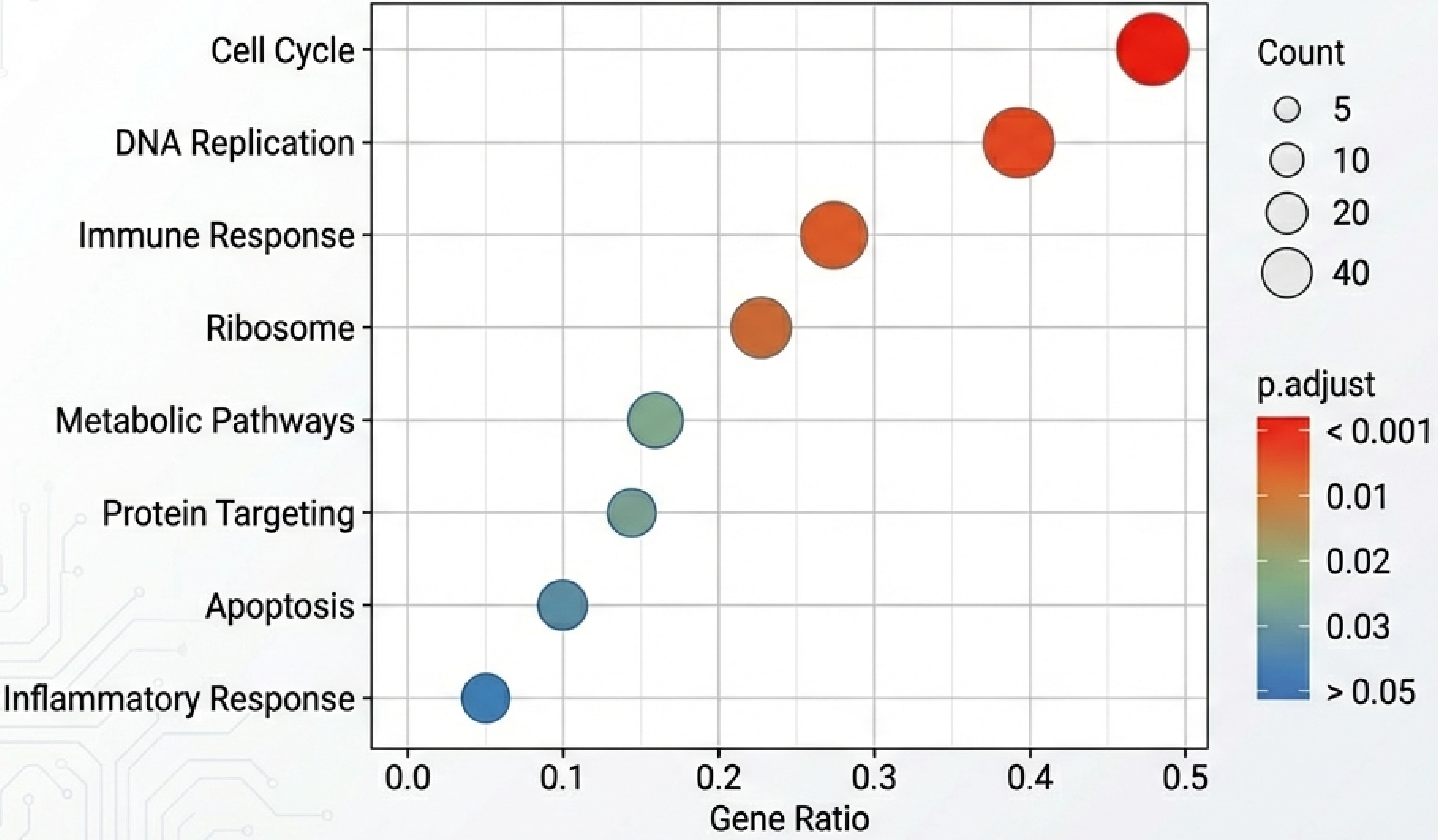


Interactive Volcano Plots showing significance vs. fold change.

Flexibility: Supports both Python-native implementation and optional R integration.

Comprehensive Enrichment Analysis

Automatically maps significantly differentially expressed genes to biological databases.



Supported Databases

GO (Gene Ontology)

Molecular Function, Biological Process, Cellular Component

KEGG

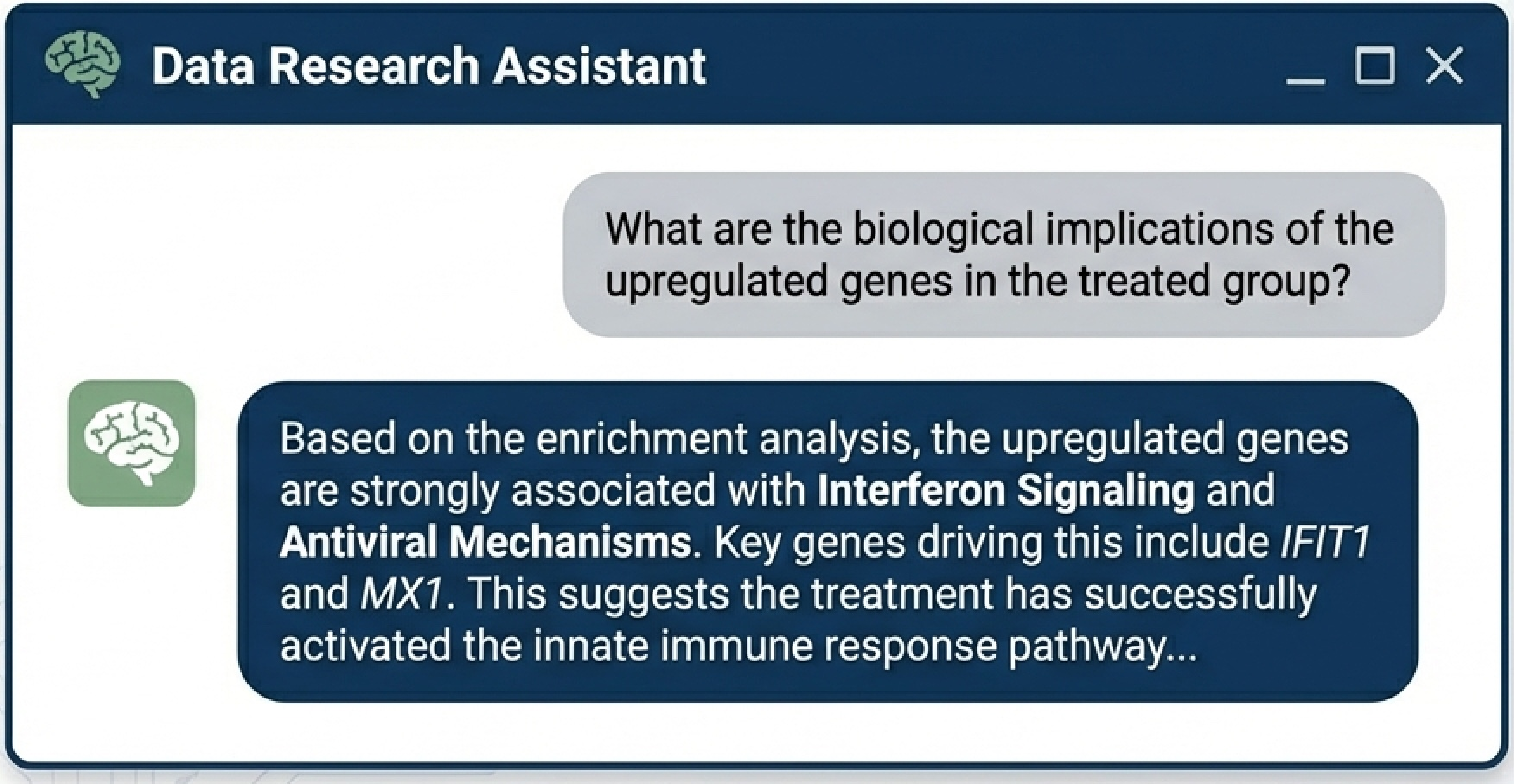
Pathway mapping.

Reactome

Detailed reaction pathways.

Conversational Data Exploration

An intelligent research assistant that knows your dataset.



What are the biological implications of the upregulated genes in the treated group?

Based on the enrichment analysis, the upregulated genes are strongly associated with **Interferon Signaling** and **Antiviral Mechanisms**. Key genes driving this include *IFIT1* and *MX1*. This suggests the treatment has successfully activated the innate immune response pathway...

Capabilities

- Ask questions about your data.
- Request specific comparisons.
- Ask for biological context.
- Powered by **Biomistral** (Biology) + **GPT-OSS** (Structure).

Interactive Visualization Suite

Built on **Gradio** for responsive, browser-based interaction.

Gradio - Volcano Plot

P-value Cutoff 0.05

Fold Change Cutoff 2.0

Gene Labeling: Gene Labeling

Color Scheme: Color Scheme

Grsdie - MA Plet

Log Ratio Cutoff 1

Filter by Gene Set: --

Gradie - PCA Plet

PC1 vs PC2: PC1 vs PC2

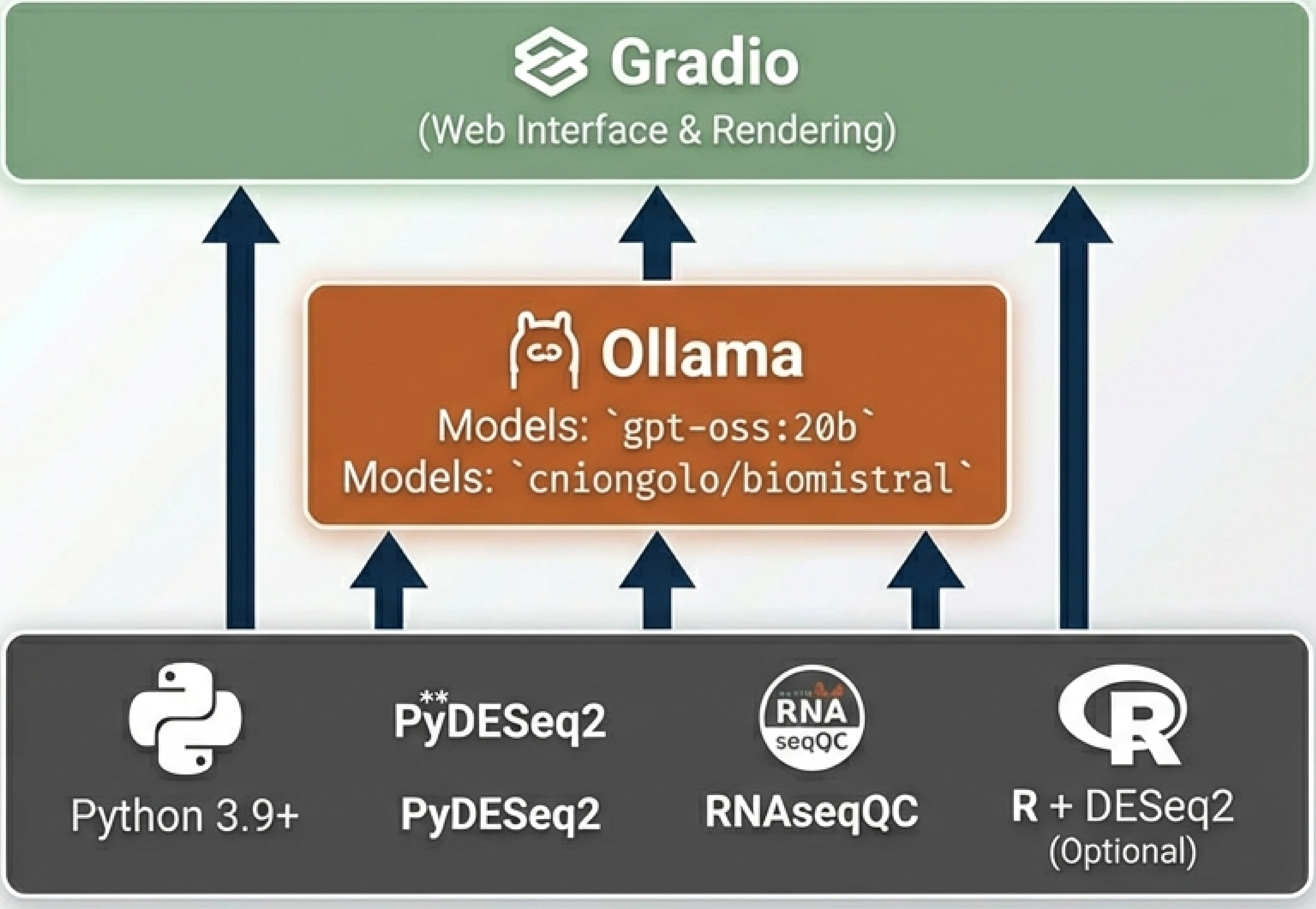
Sample Size 2.0

Zoom • Toggle • Adjust Parameters

Zoom • Toggle • Adjust Parameters

Technology Stack & Dependencies

A high-fidelity vecto “layer cake” diagram.



Gradio

(Web Interface & Rendering)

Ollama

Models: `gpt-oss:20b`

Models: `cniongolo/biomistral`

Python 3.9+

PyDESeq2

RNAseqQC

R + DESeq2

(Optional)

Prerequisites

System Environment

- ☑ **Python**: Version 3.9 or higher
- ☑ **Conda**: Recommended for environment management

Ollama Setup (Local Inference)

Must be installed locally. Required model pulls:

```
ollama pull gpt-oss:20b
ollama pull cniongolo/biomistral
```

Optional Extensions

- ☑ R and Bioconductor packages (DESeq2, RNAseqQC) if R-based workflow is preferred.

Installation & Launch

Installation

```
$ git clone https://github.com/shivaprasad-patil/BulkRNA-Agent.git
$ cd BulkRNA-Agent
$ ./install.sh
```

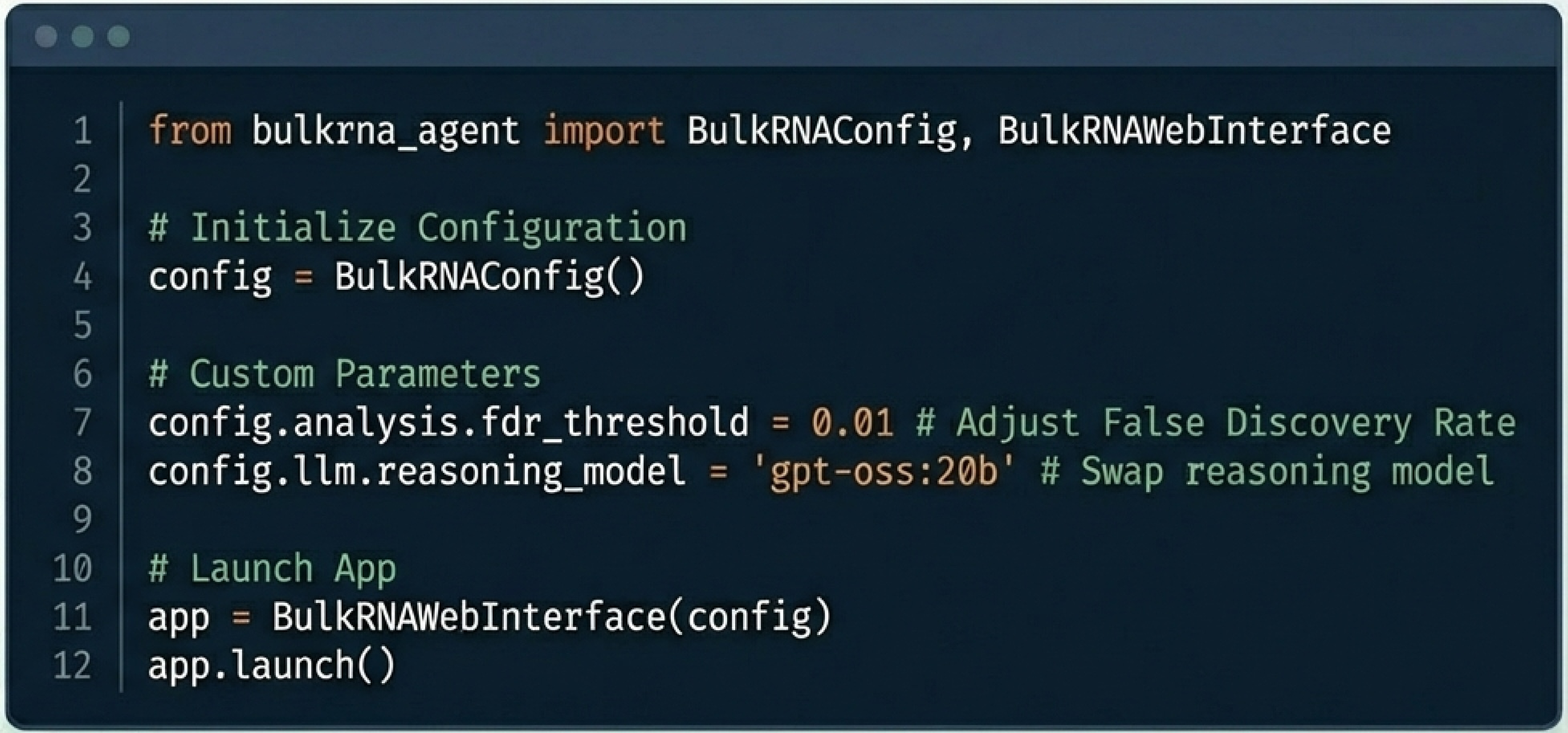
Or manual setup via conda/pip

Launch

```
$ ./start.sh
> Starting Ollama service...
> Launching Gradio interface...
> Running on local URL:
http://localhost:7860
```

Advanced Configuration

Customize parameters via the Python API.



```python
from bulkrna_agent import BulkRNAConfig, BulkRNAWebInterface

# Initialize Configuration
config = BulkRNAConfig()

# Custom Parameters
config.analysis.fdr_threshold = 0.01 # Adjust False Discovery Rate
config.llm.reasoning_model = 'gpt-oss:20b' # Swap reasoning model

# Launch App
app = BulkRNAWebInterface(config)
app.launch()
```

R Integration

Install Bioconductor packages via `Rscript` for native DESeq2 backend support.

Resources & Citation

1. Citation

Patil, Shivaprasad (2026).
BulkRNA-Agent: AI-Powered
Bulk RNA-seq Analysis Tool.

2. Documentation

- `GETTING_STARTED.md` - Detailed guides
- `CONTRIBUTING.md` - Development guidelines
- **License**: **MIT License (Open Source)**

3. Repository

github.com/shivaprasad-patil/BulkRNA-Agent

Status: **Active development**. **Contributions welcome**.