

NIH STRIDES

Accelerating biomedical research

Workflow for Bulk RNA-seq and AI-enabled Visualizations

August 11, 2026 and August 13, 2026



NIH STRIDES

Accelerating biomedical research

Part 2: AI-enabled Visualizations

Aug 13, 2026



Part 2 Agenda

Part 2: AI-enabled Visualizations

- Welcome & Context
- Azure Environment Setup for AI
- What is a Prompt ?
- The Four Prompt Layers
- Detailed Prompt Walkthrough
- Demonstration: AI-Generated RNA-Seq Figures



Welcome & Context

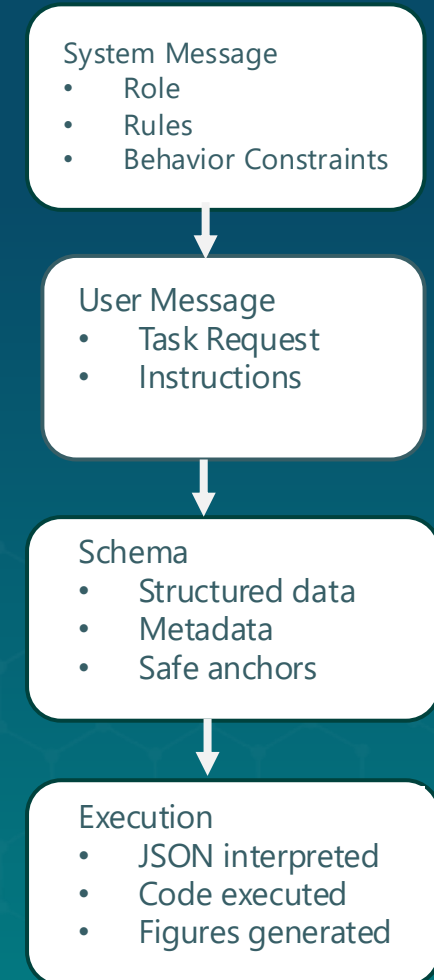
- Recap of Part 1 (Bulk RNA-seq workflow) using FASTQ → QC → Trimming → Alignment → DESeq2 → Result tables
- AI-enabled visualizations using outputs from Part 1:
 - res_full (DESeq2 output)
 - res_shrunken (DESeq2 shrunken output)
 - top50 up (Top 50 upregulated genes)
 - top50 down (Top 50 downregulated genes)
 - gene_samples_df (gene x sample matrix)
- AI-enabled visualizations using outputs from Part 2:
 - pca_df (pca coordinates)
 - dist_corr_df (correlation matrix for samples)

Azure Environment Setup for AI

- Launching and using Azure OpenAI (GPT and embeddings)
- Quick overview of model endpoint, API keys, and workspace environment

What is a Prompt?

- A prompt is the set of instructions given to an AI model to determine what it should do.
- Prompts usually contain multiple layers, each shaping the model's behavior and output.
- This layered design helps create predictable, safe, reproducible AI operations.
- In practical AI workflows, The prompt is composed of:
 1. System Message
 2. User Message
 3. Schema / Context
 4. Execution / Output



The Four Prompt Layers

1

System Message

- Defines role
- RNA-seq domain rules
- Output format rules

2

User Message

- Task:
"Create heatmap.."
- Plot type, description
- What Data frame to use

3

Schema

- `df.columns`, `df.shape`
- `head()` preview
- Safe anchor points

4

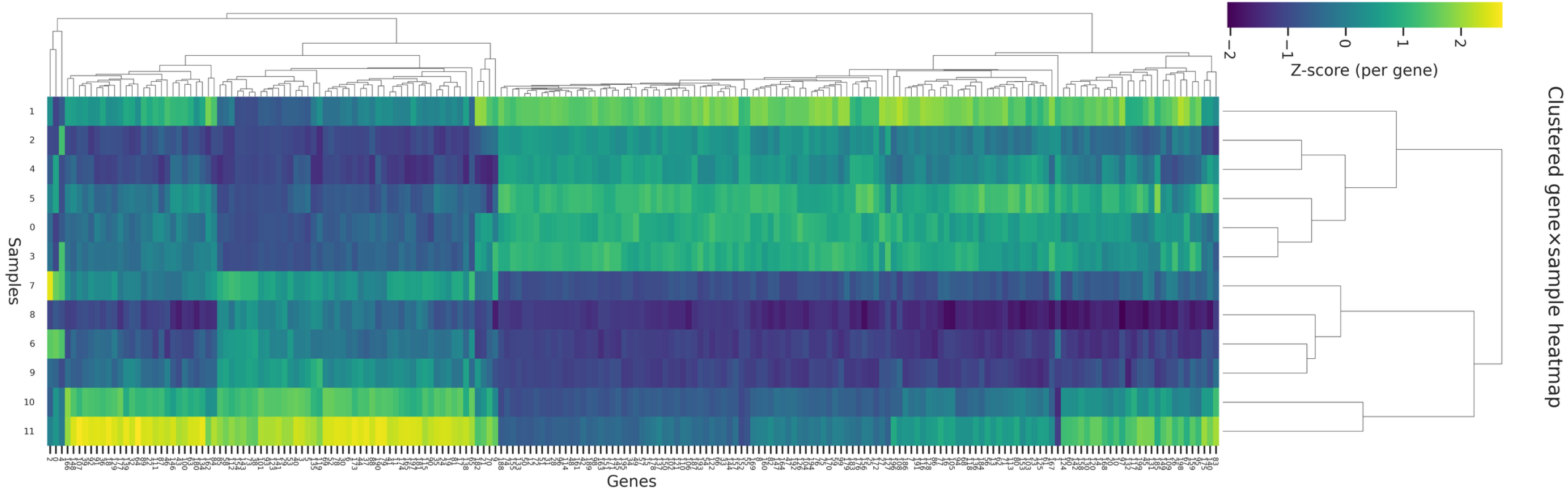
Execution

- Run generated code
- Apply cleaning rules
- Save PNG/SVG
- Error fallback

Detailed Prompt Walkthrough

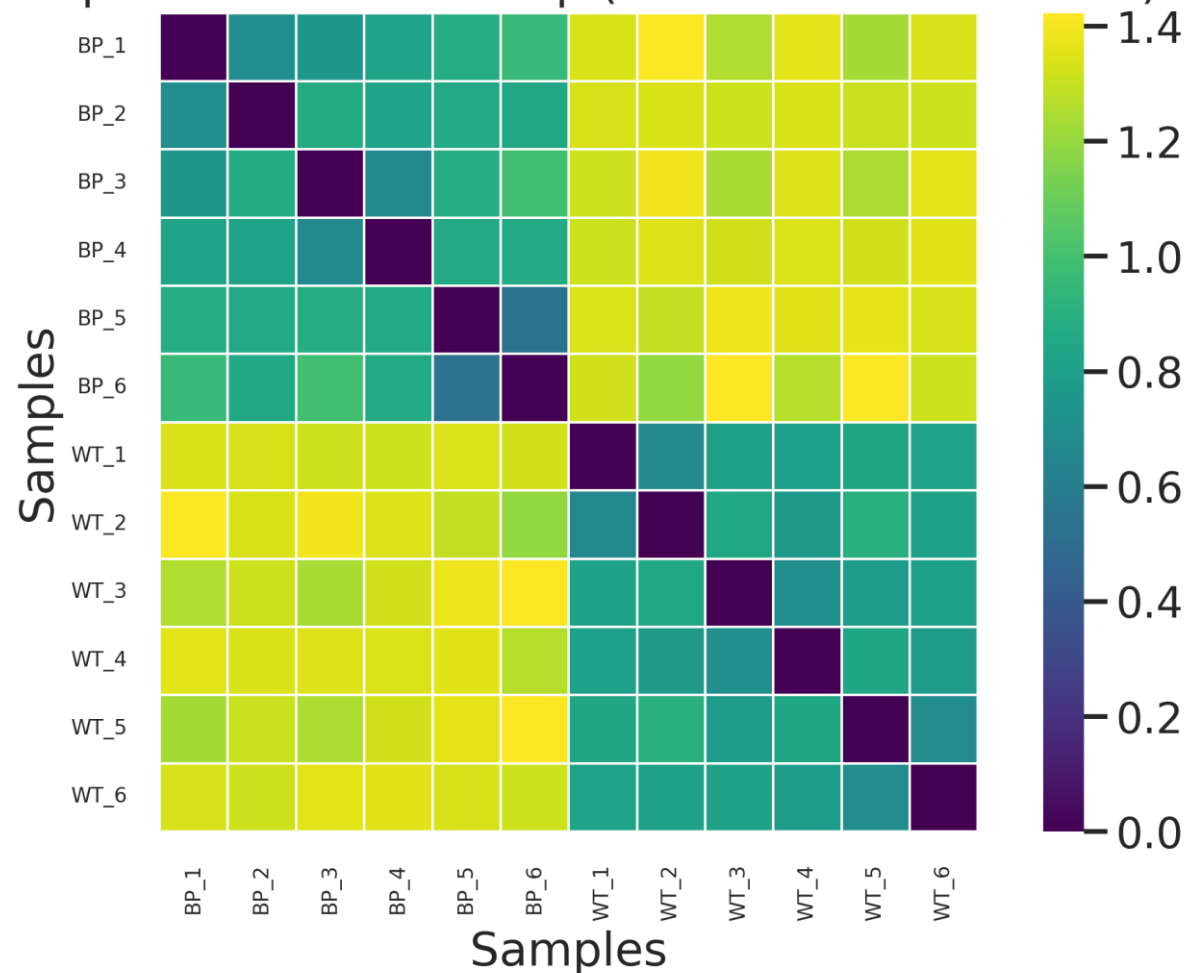
- How you compose the system message
(allowed libraries, palette, strict code rules)
- How the user message defines the figure
(description, plot type, dependencies)
- How the schema anchors the model to correct data
(sample IDs, gene names, numeric columns)
- How output is validated and executed inside the notebook

AI-Generated RNA-Seq Figures: Gene x Sample heatmap

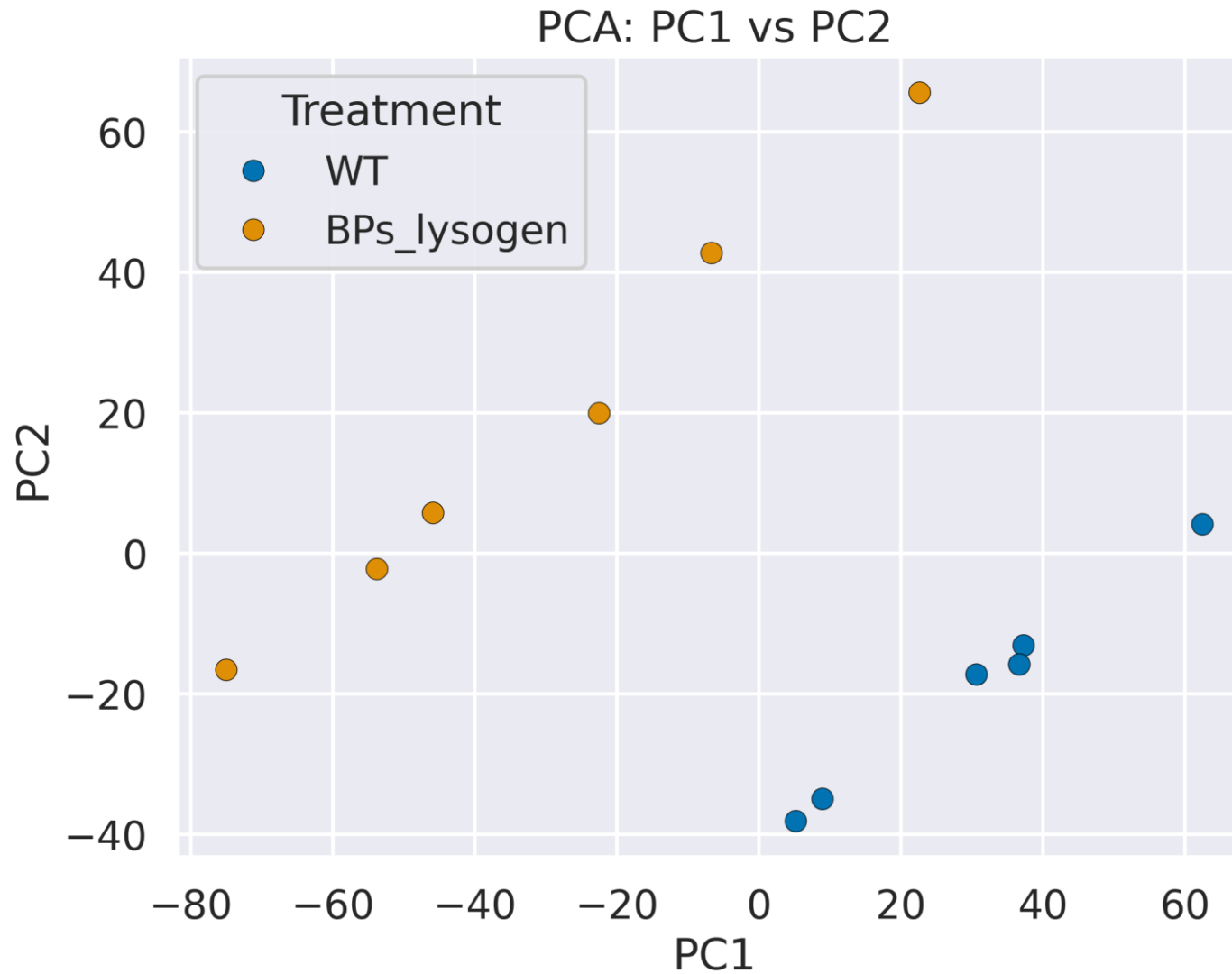


AI-Generated RNA-Seq Figures: Sample x Sample correlation matrix

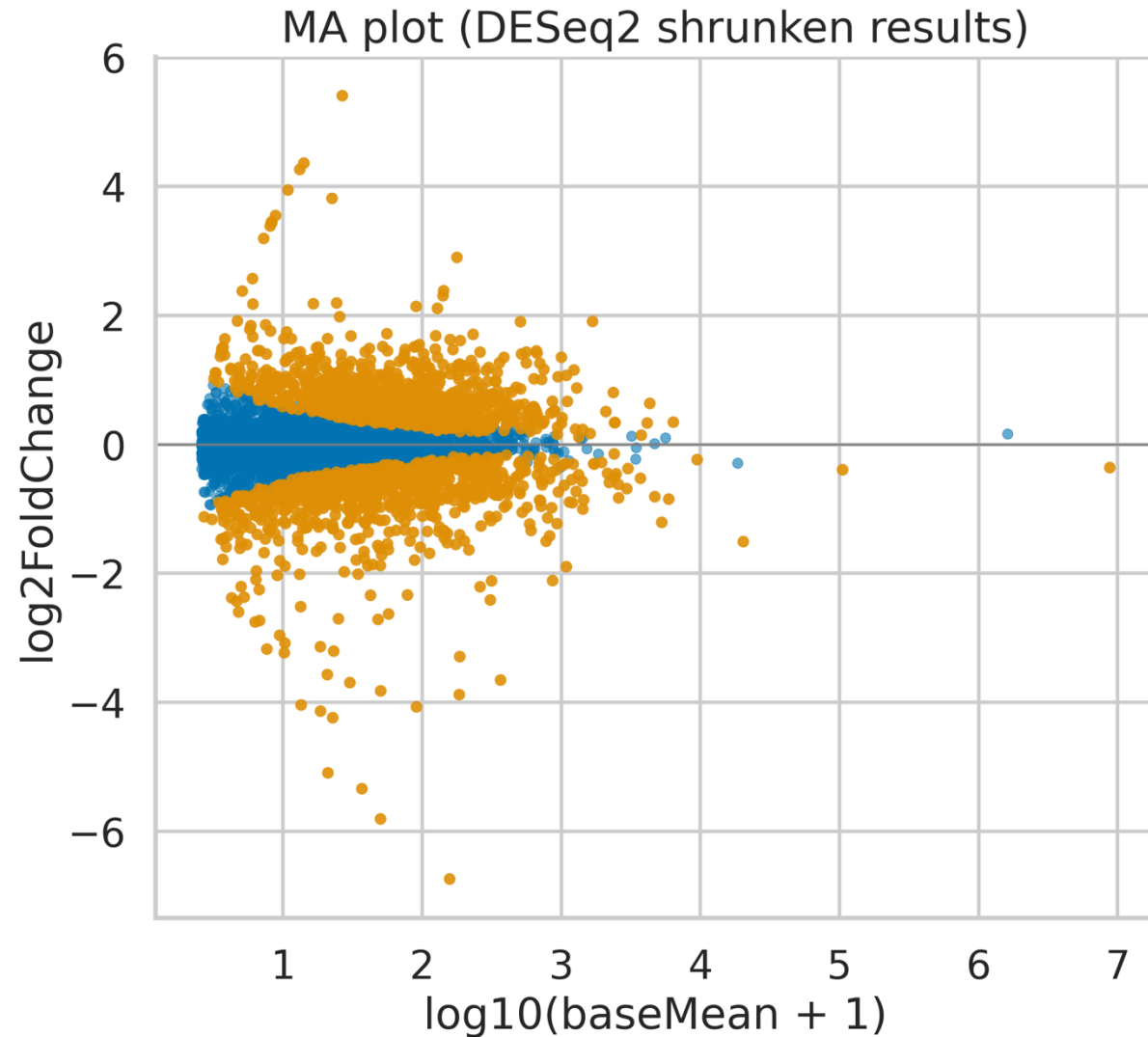
Sample distance heatmap (1 - Pearson correlation)



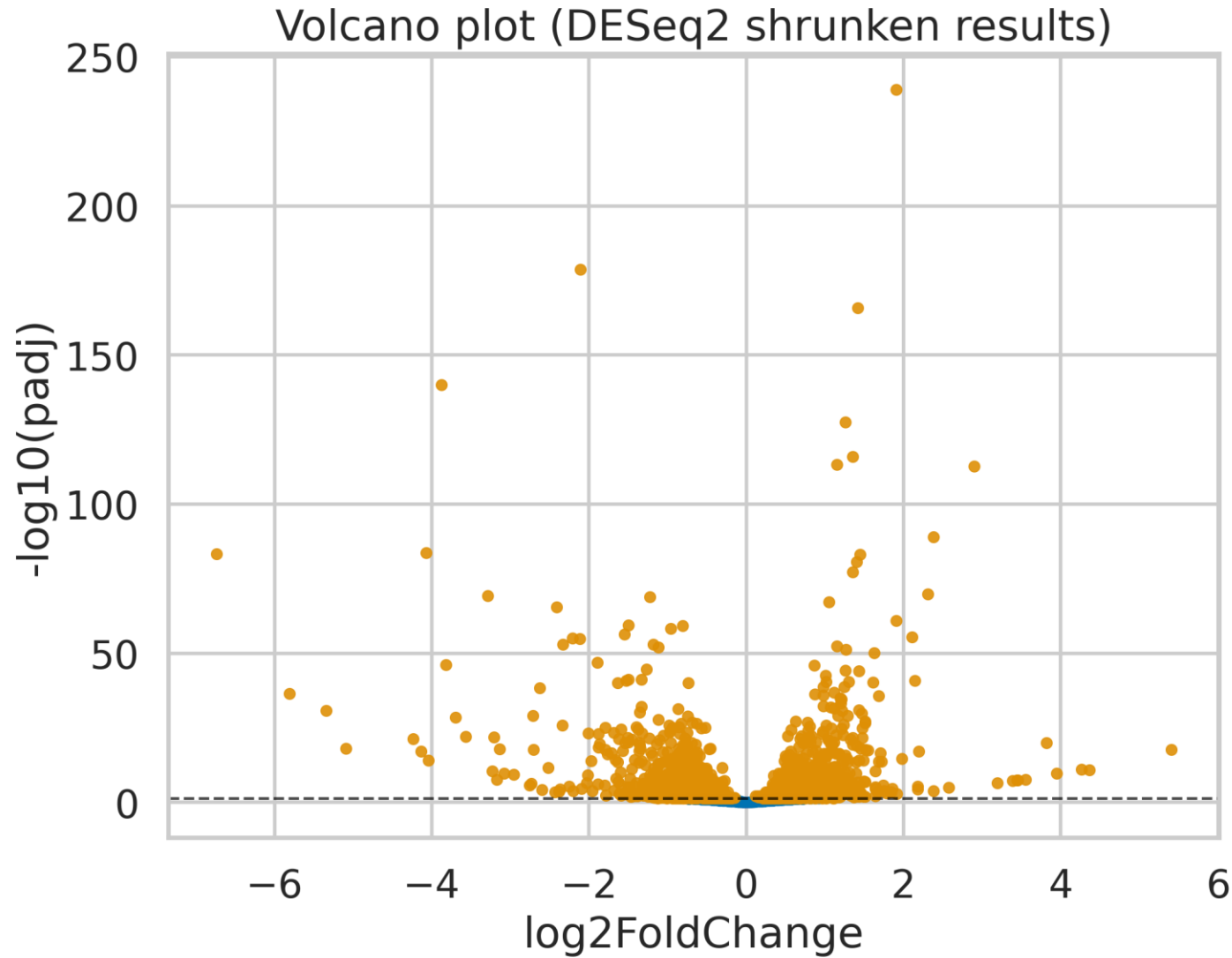
AI-Generated RNA-Seq Figures: PCA plot



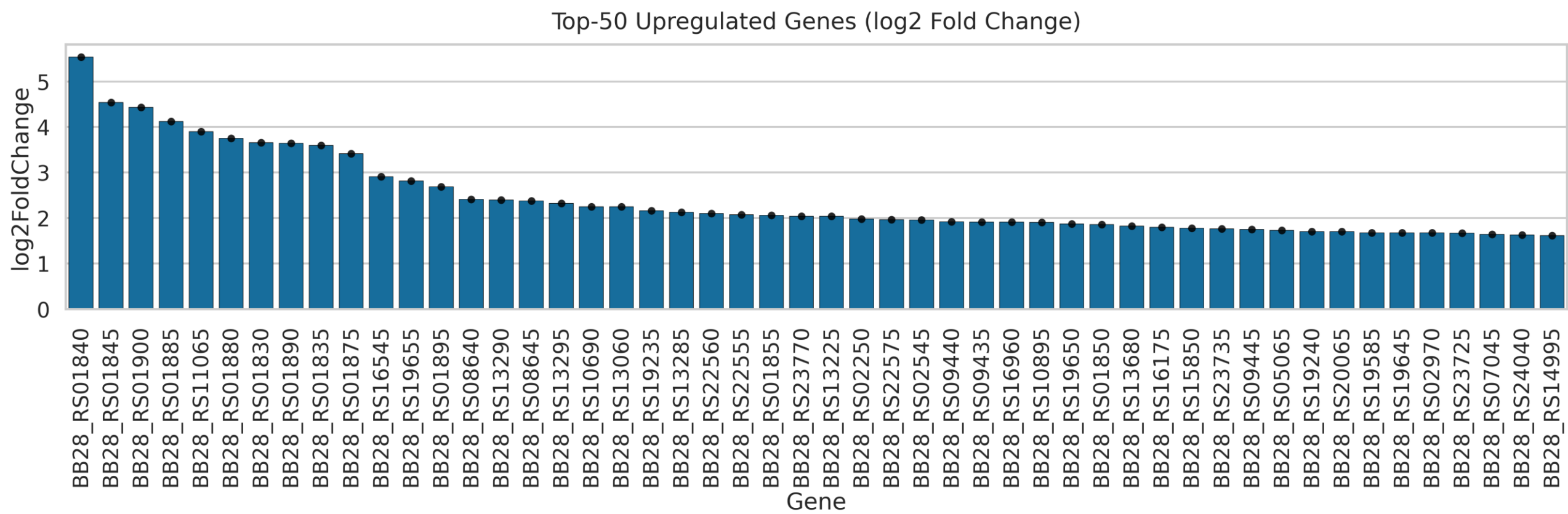
AI-Generated RNA-Seq Figures: MA plot



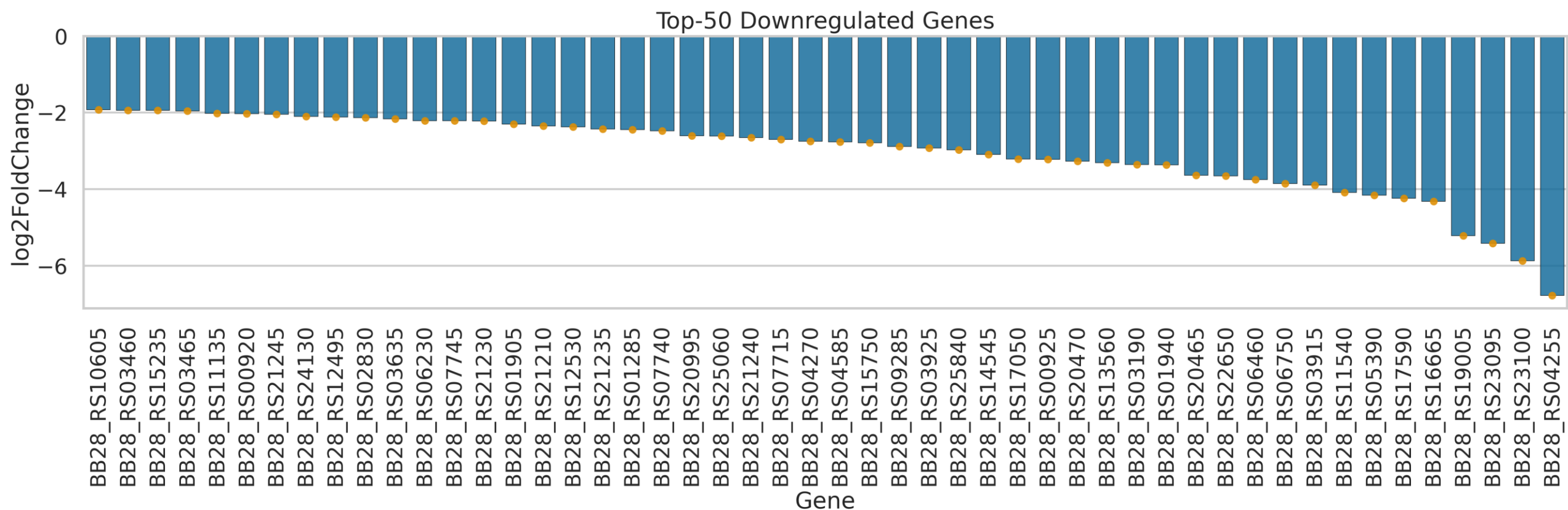
AI-Generated RNA-Seq Figures: Volcano Plot



AI-Generated RNA-Seq Figures: Top 50 Upregulated Genes



AI-Generated RNA-Seq Figures: Top 50 downregulated genes



Links to Aid in Enabling Visualizations through AI

1. https://github.com/STRIDES/NIHCloudLabAzure/blob/main/notebooks/GenAI/azure_infra_setup/README.md
2. https://github.com/STRIDES/NIHCloudLabAzure/blob/main/notebooks/GenAI/initiate_jupyter_notebook/InitiateJupyterNotebook.md
3. [NIHCloudLabAzure/notebooks/GenAI/notebooks/AI_enabled_visualization.ipynb at main · STRIDES/NIHCloudLabAzure](#)

NIH STRIDES

Accelerating biomedical research

For more information, please visit cloud.nih.gov
or send an email to STRIDES@nih.gov.
