



# QUICK REFERENCE GUIDE FOR ANALYZING A DATASET USING GL4HS PREPARED SETS

## 1 SELECT A DATASET

- Navigate browser to *genelab.nasa.gov*
- Click on **Analyze Data**
- Sign in with Google account
- Click on **Shared Data** > Click on **Data Libraries** > Click on **GL4HS** > Choose & import a dataset\*

## 2 REVIEW METADATA

- Navigate browser to *genelab.nasa.gov*
- Click on **Data Repository**
- Use Search Data field by GLDS accession number
- Review Description, Protocols, Samples, Assays, Publications, etc

## 3 RUN NORMALIZATION AND DIFFERENTIAL EXPRESSION ANALYSIS

- In Galaxy platform, use **DESeq2\***
- **Normalization:** Raw data adjusted to account for technical variation and bias
- **Differential Gene Expression:** Estimates significance based on variable conditions

## 4 EXAMINE PLOTS

- **PCA:** Overall view of experimental covariants & batch effects
- **Heat Maps:** Overall similarities & dissimilarities of data
- **Dispersion Estimates:** Overall variability of data
- **Histogram:** Representation of range of data
- **MA Plot:** Relationship between expression & strength of test

## 5 VISUALIZE DIFFERENTIALLY EXPRESSED GENES

- Use results from normalization & differential expression analysis\*
- Filter data for **statistical significance**
- Filter data for **biological significance**
- Generate a **volcano plot** with labeled genes

## 6 CONDUCT FUNCTIONAL ANALYSIS

- **Gene Ontology Analysis\*:** Highlights over- or under-represented biological processes
- **Gene Set Enrichment Analysis\*:** Compares groups using signal to noise ratio for probes
- **Molecular Signature Databases:** Examines collective relatedness of genes across pathways

\*Follow instructions or parameters as described in the GeneLab for High Schools Bioinformatics Manual