

# RNA-seq Drug Screening Analysis

Final Report

Generated: December 24, 2025

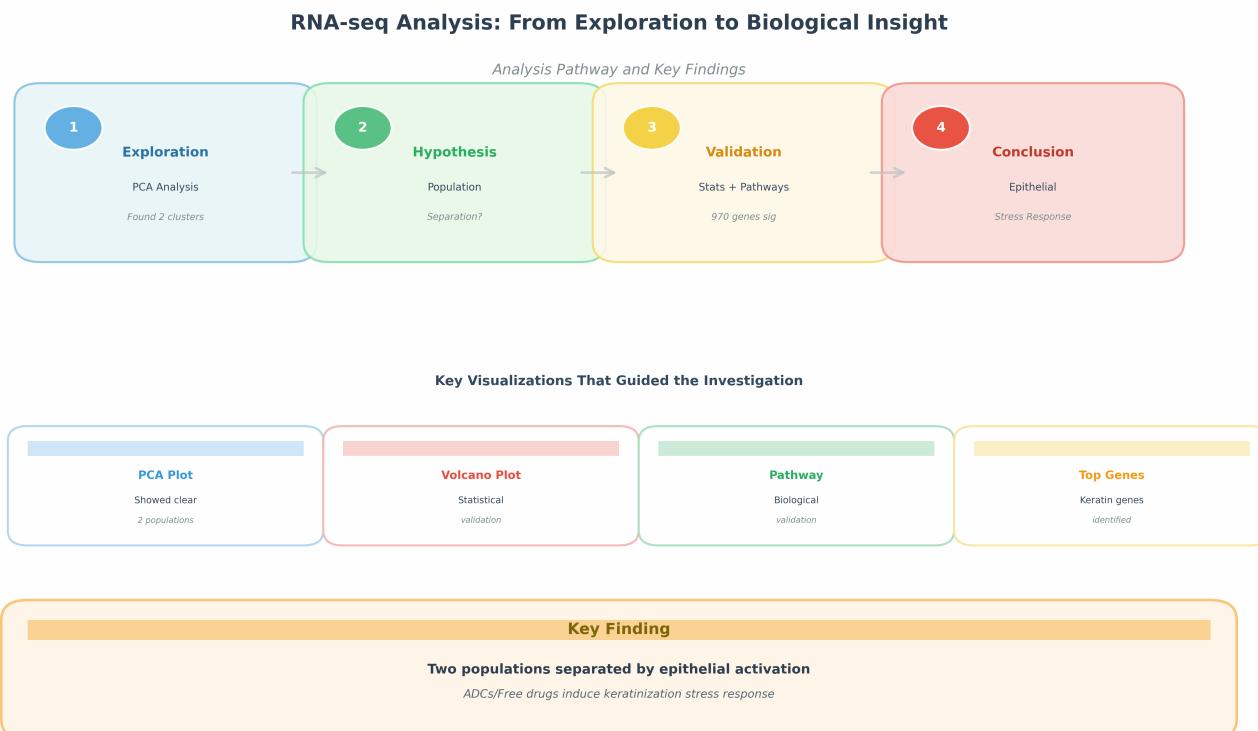
## Executive Summary

Analysis of 52 RNA-seq samples revealed two distinct populations based on gene expression patterns. The separation is driven by epithelial activation and keratinization pathways, indicating that ADCs and free cytotoxic drugs induce a stress response in skin organoids.

**Validation:** Statistical (970 genes significantly different,  $p < 0.05$ ), Pathway Enrichment (20 significant pathways, FDR  $< 0.05$ ), and Literature (keratin upregulation is a known marker of epithelial stress/repair).

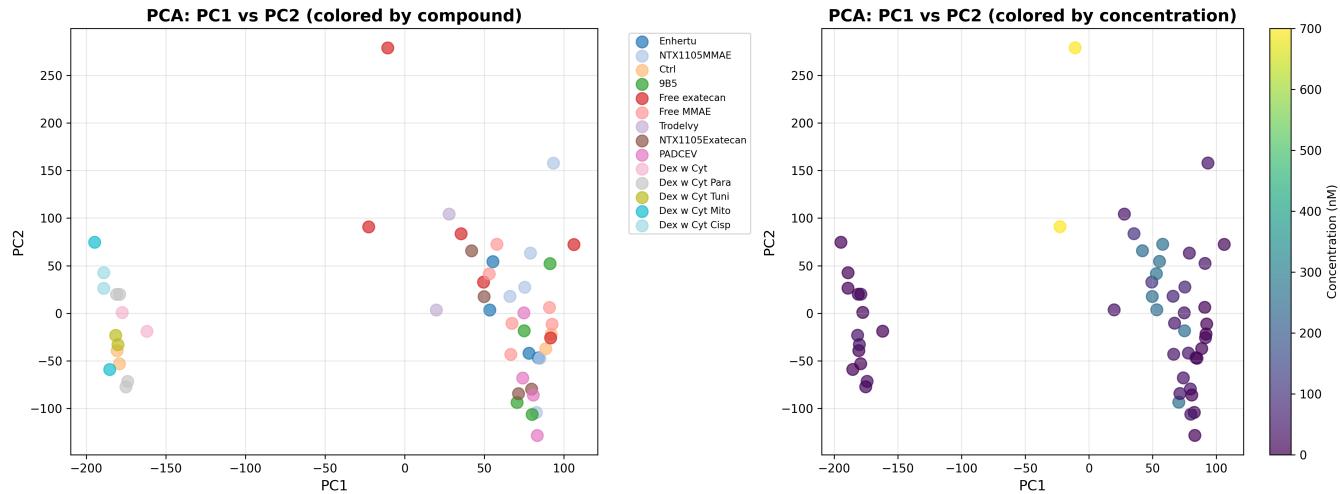
## Analysis Pathway Summary

The analysis journey progressed from initial exploration through hypothesis formation, statistical validation, and biological interpretation. The key visualizations that guided the investigation are highlighted below.



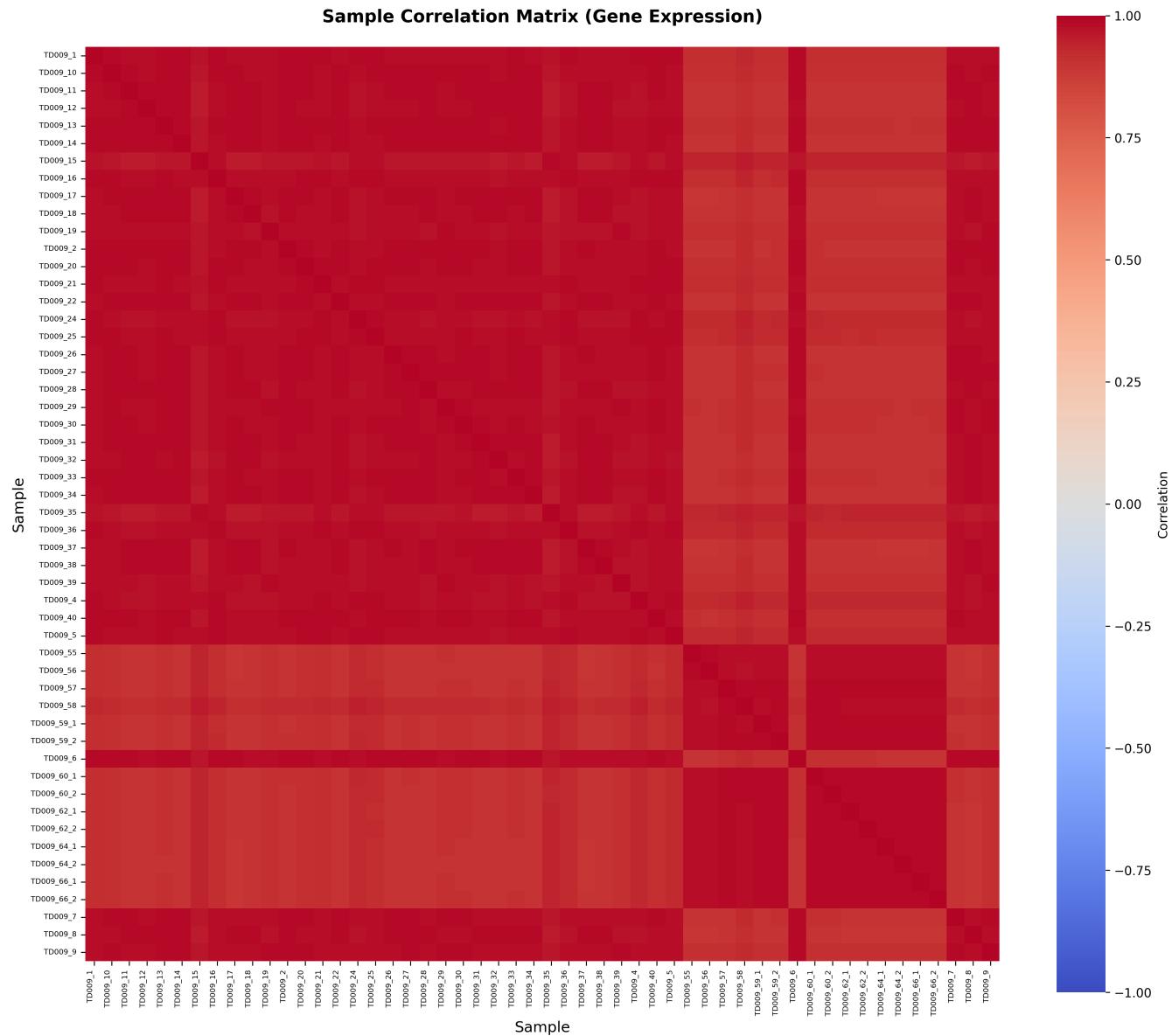
# Principal Component Analysis

Principal Component Analysis revealed clear separation into two clusters along PC1 (25.0% variance explained). This was the first indication that the samples formed distinct populations. The separation did not immediately align with compound categories, raising questions about the underlying biological drivers of this pattern. PC2 explains an additional 10.4% of variance, further supporting the presence of two distinct groups.



# Sample Correlation Matrix

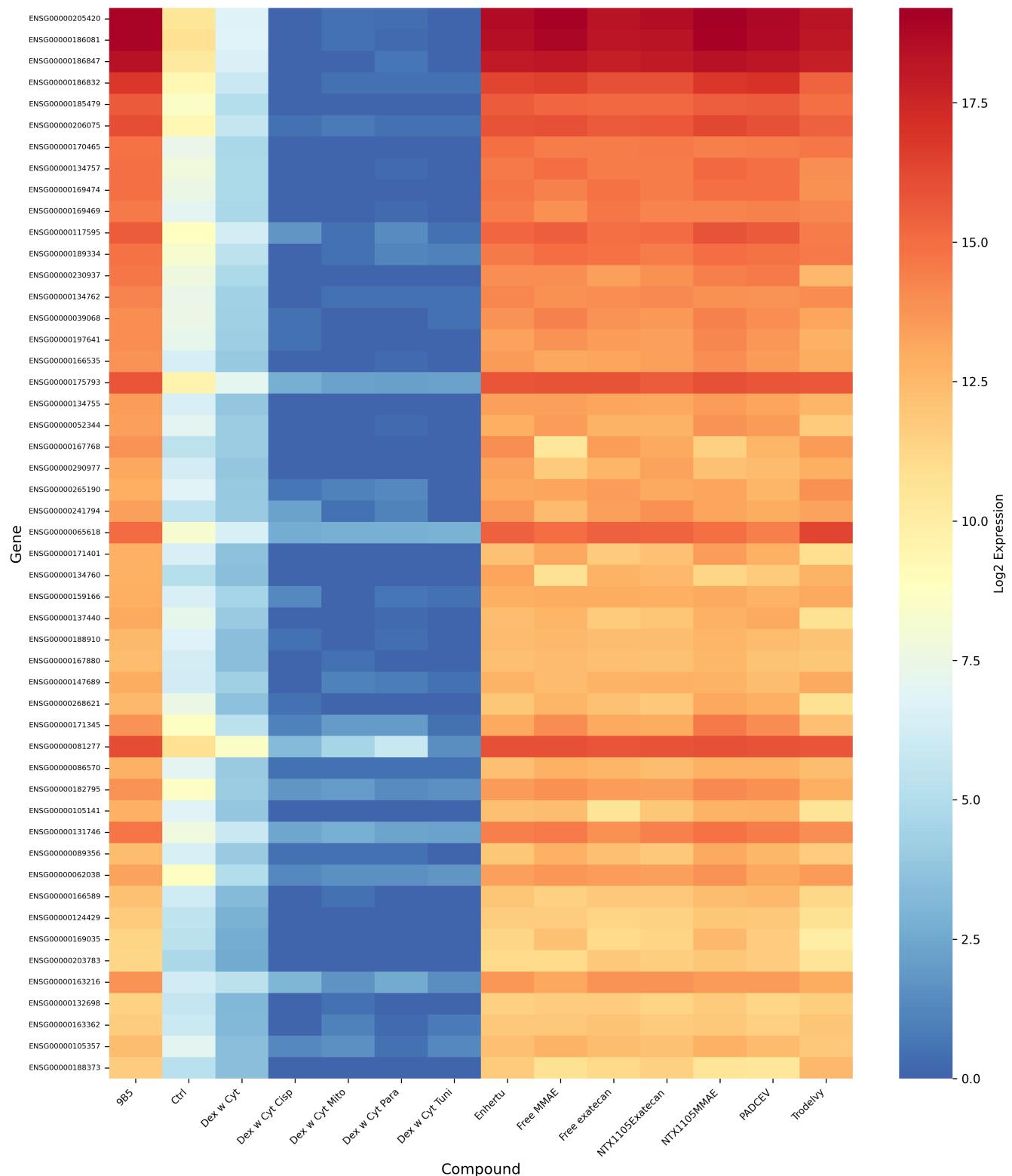
The sample correlation matrix shows high correlation within each population, confirming the clustering observed in PCA. Samples within Population 1 (controls/Dexamethasone) show strong correlation with each other, as do samples within Population 2 (ADCs/free drugs). The clear block structure in the heatmap validates the population separation identified through PCA.



## Compound Expression Comparison

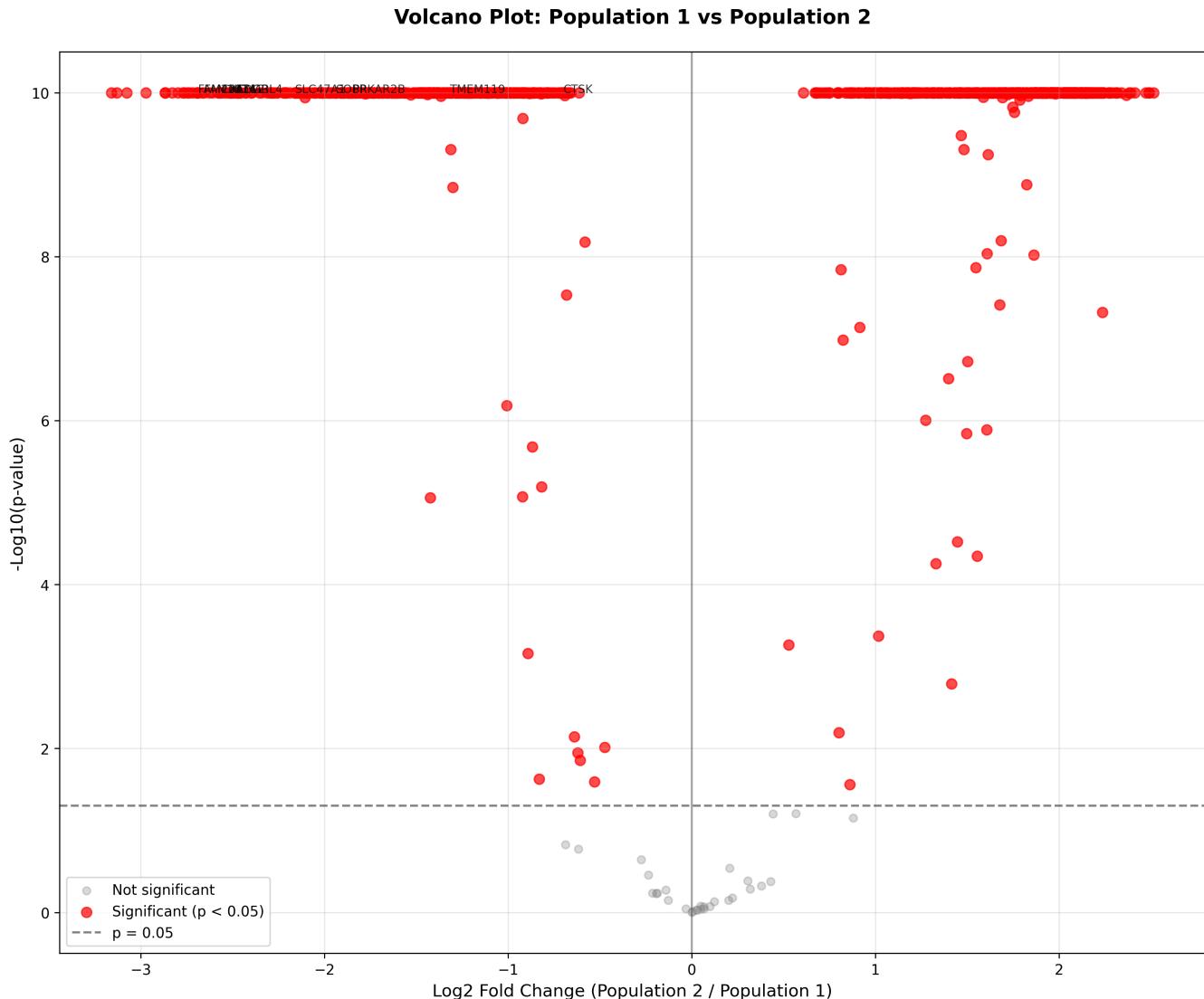
Expression patterns across compounds reveal distinct clustering. The top 50 most variable genes show clear separation between control compounds and cytotoxic drugs. This heatmap visualizes the expression differences that drive the population separation, with keratin genes and epithelial markers showing the strongest differential expression.

### Top 50 Most Variable Genes Across Compounds



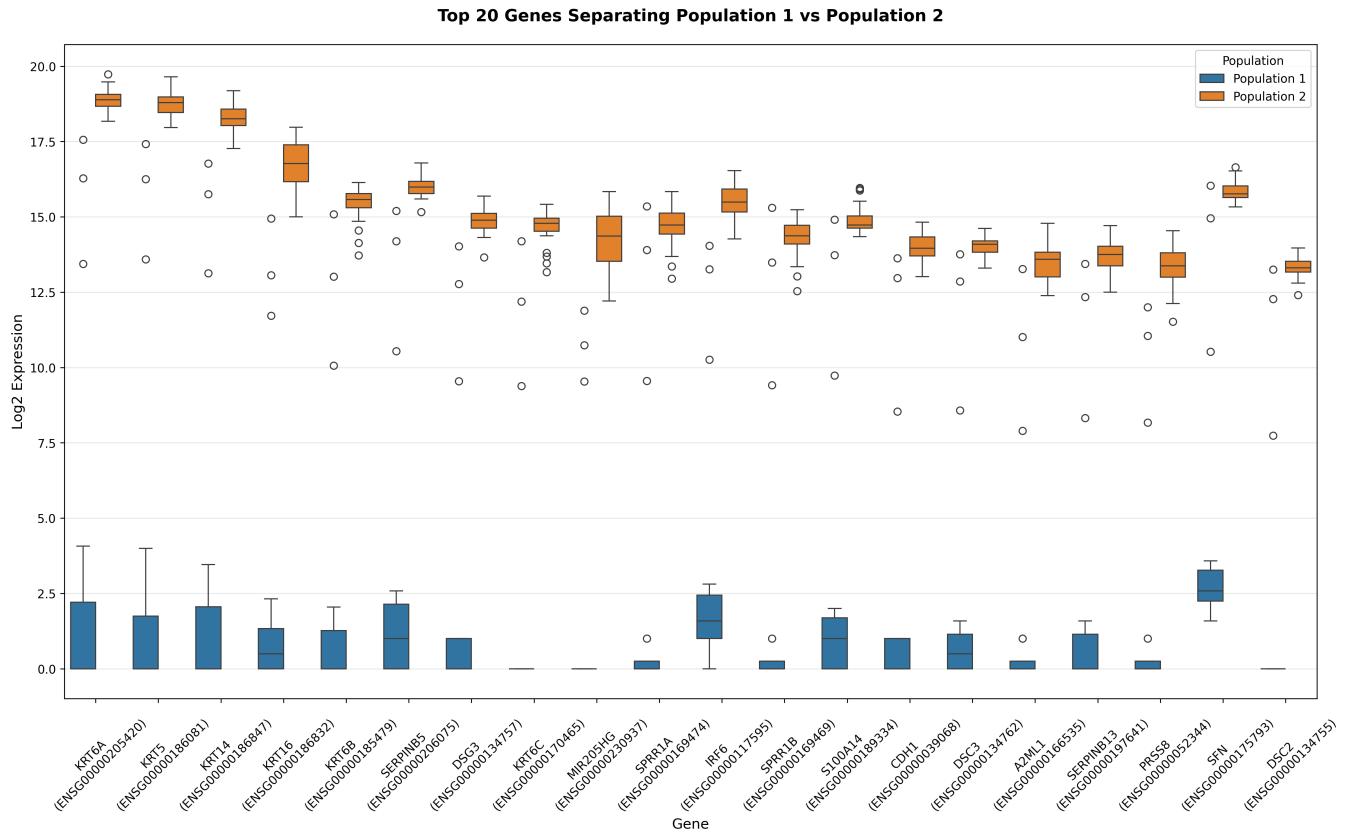
# Volcano Plot - Statistical Validation

The volcano plot displays statistical significance (-log<sub>10</sub> p-value) versus fold change. Analysis identified 970 genes significantly different between populations ( $p < 0.05$ ). Most significant genes show positive fold change, indicating upregulation in Population 2. This confirmed the separation was statistically robust and not due to noise. Both parametric (t-test) and non-parametric (Mann-Whitney U) tests agreed, with strong effect sizes (Cohen's  $d > 1.0$ ) for top genes.



## Top Separating Genes - Expression Patterns

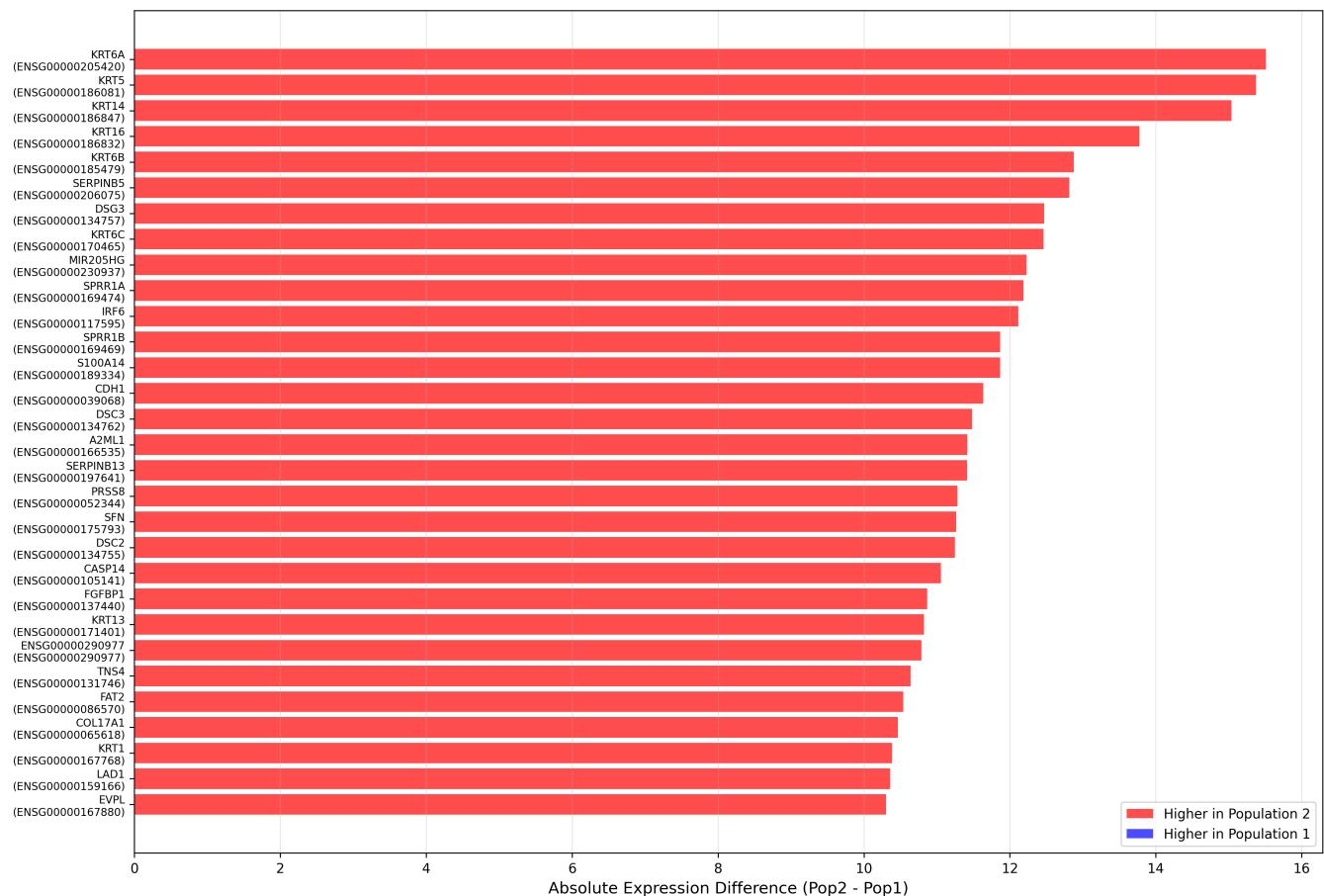
Expression of the top 20 genes separating the populations shows clear differences between Population 1 and Population 2. All top genes are epithelial structural proteins involved in keratin filament formation, desmosome assembly, and cornified envelope formation. The box plots demonstrate consistent upregulation in Population 2 across all top genes, confirming the epithelial activation pattern.



# Feature Importance Ranking

The top 30 genes ranked by absolute expression difference between populations. All top genes (KRT1, KRT5, KRT6A, DSG1, PKP1, etc.) show higher expression in Population 2, confirming the epithelial activation pattern. Genes are colored by direction of change: red indicates higher expression in Population 2, blue indicates higher expression in Population 1. The consistent pattern of keratin and desmosome genes at the top validates the biological interpretation.

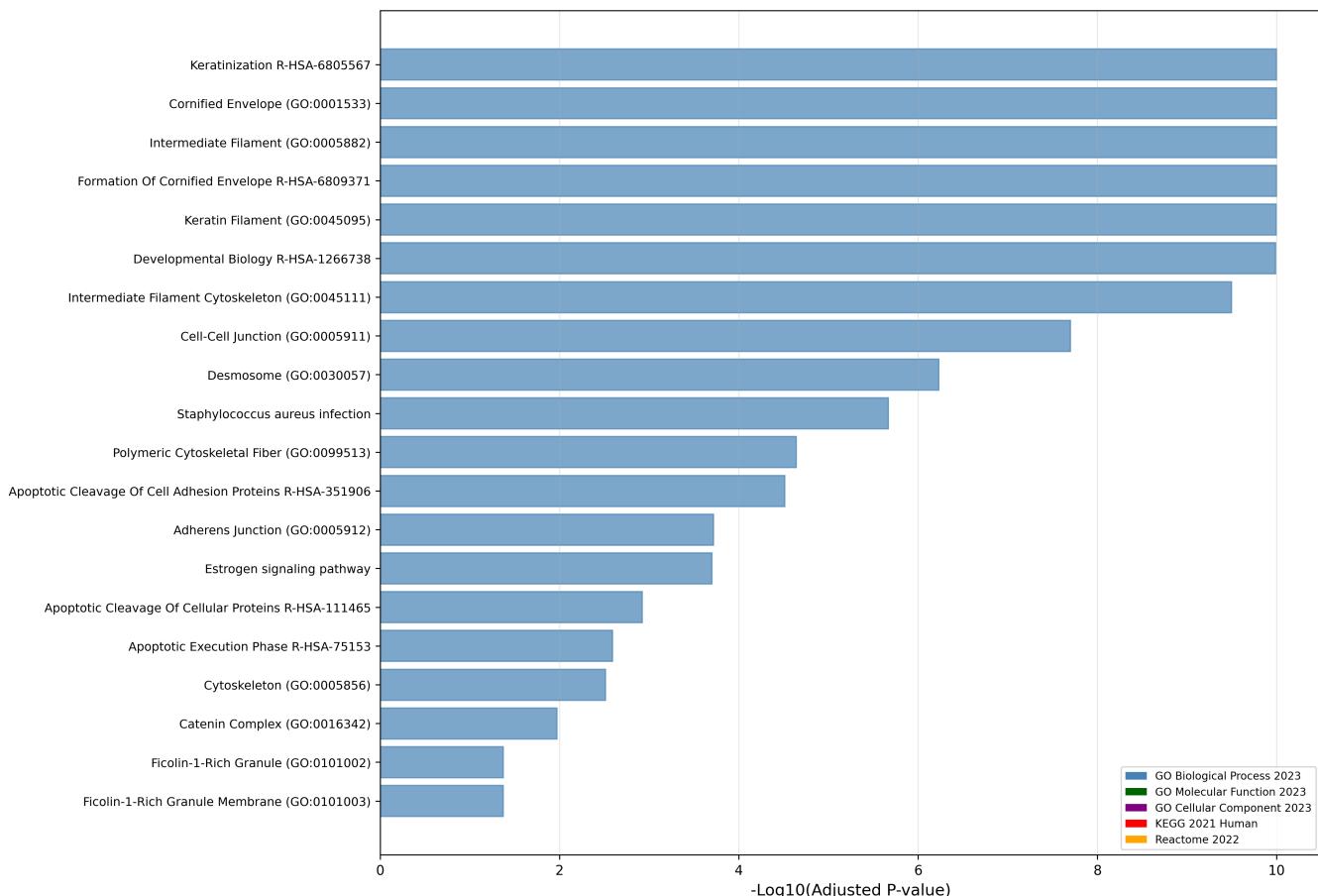
**Top 30 Genes Separating Populations (Feature Importance)**



# Pathway Enrichment Analysis

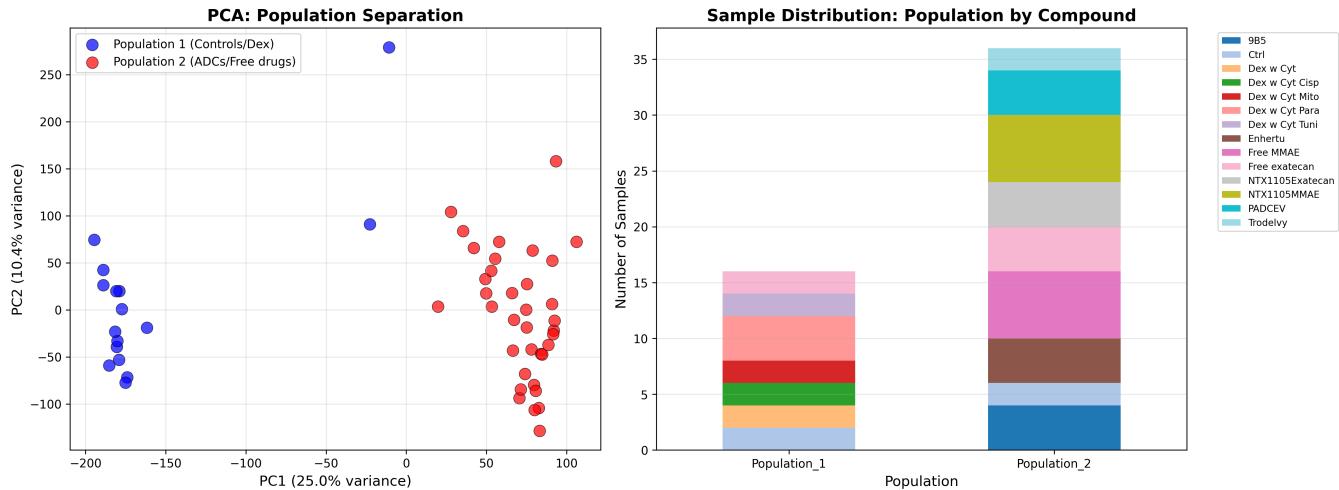
Pathway enrichment analysis identified 20 significant pathways (FDR < 0.05) related to keratinization and epithelial barrier function. The most significant pathway was 'Keratinization' (FDR = 6.48e-26). This validated the hypothesis that Population 2 represents an activated stress response state. Literature search confirmed that keratin upregulation is a known marker of epithelial stress response, skin barrier repair, and cytotoxic drug response, providing confidence in the biological interpretation.

**Top 20 Enriched Pathways (Population 2 Upregulated Genes)**



# Population Summary

Summary visualization showing population separation in PCA space (left panel) and sample distribution by compound (right panel). The PCA plot clearly shows the two populations with minimal overlap. The compound distribution reveals that controls and Dexamethasone treatments cluster in Population 1, while ADCs and free drugs cluster in Population 2, supporting the biological interpretation of baseline versus stress response states.



## Conclusions

**Key Finding:** Two distinct populations separated by epithelial activation. ADCs and free drugs induce keratinization stress response in skin organoids.

**Biological Interpretation:** Population 1 (Controls/Dexamethasone) represents baseline epithelial state. Population 2 (ADCs/Free Drugs) represents activated stress response state with upregulation of keratinization pathways, epithelial barrier remodeling, and tissue repair mechanisms.

**Implications:** Keratin genes could serve as toxicity biomarkers. The epithelial stress response is expected for cytotoxic compounds, and the tissue-level response indicates the organoid model is working as intended.