Answer ALS Epigenomics Signal Reproducibility

Yogindra Raghav
November 9th 2021
Fraenkel Lab

Samples Not Included

- Exclusions:
  - **Failed QC (Aneesh investigating):**
    - CTRL-NEUCV136DHM-9152-E
    - CASE-NEUGK693CJZ-8649-E
  - **Original sample is sequestered:**
    - CTRL-NEUYM011FER-8963-E
  - **Original data missing:**
    - CTRL-NEUGM959DE2-8684-E
    - CASE-NEUBD062TJV-8614-E
  - **Unsure of AE8 to compare to:**
    - CTRL-NEUEU392AE8-5230-E
    - CTRL-NEUEU392AE8-9187-E
    - CTRL-NEUEU392AE8-8984-E

- **Total samples: 24**

Workflow

- Take peak regions where there is overlap in **at least** 10% of samples

- Retrieve counts under those peaks using DiffBind

- Normalize counts using TMM method from EdgeR

- Standardize all features

- 48 samples x 116K features counts matrix
  - Normalized & Standardized

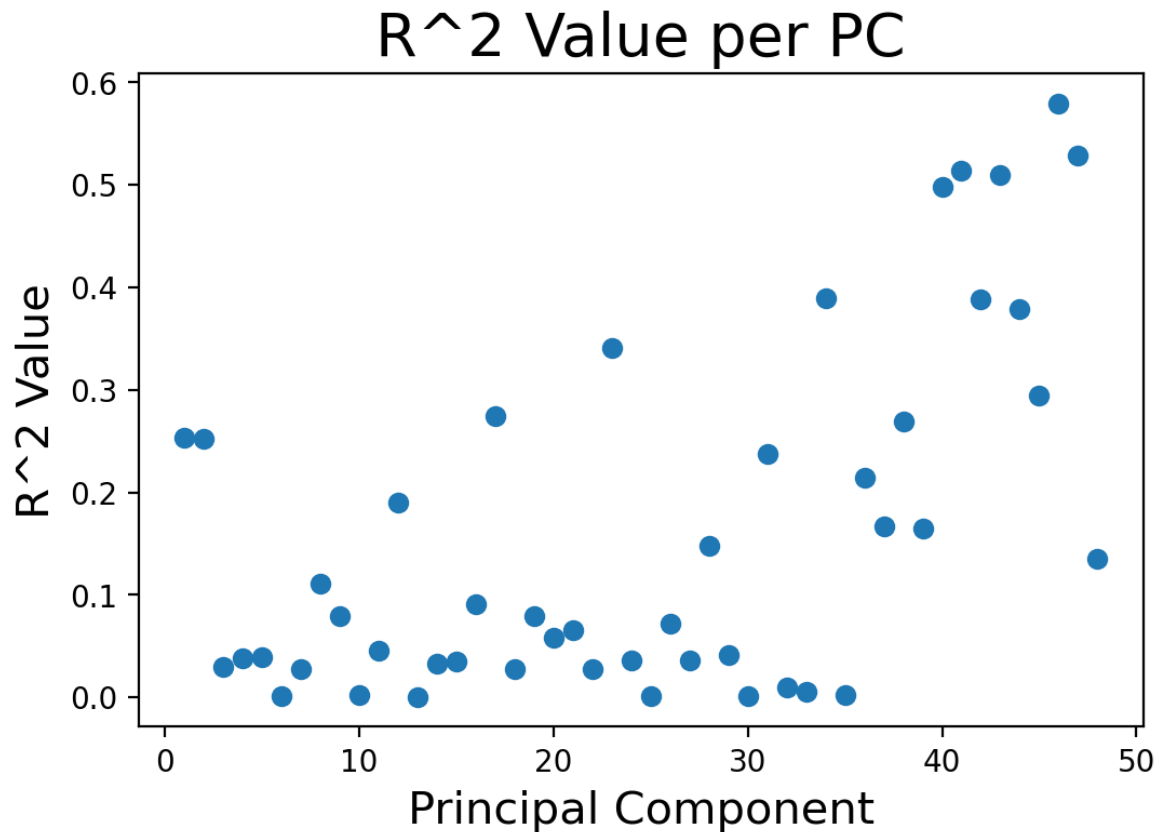# PCA does not Cluster by Condition

Original & New Lines PCA
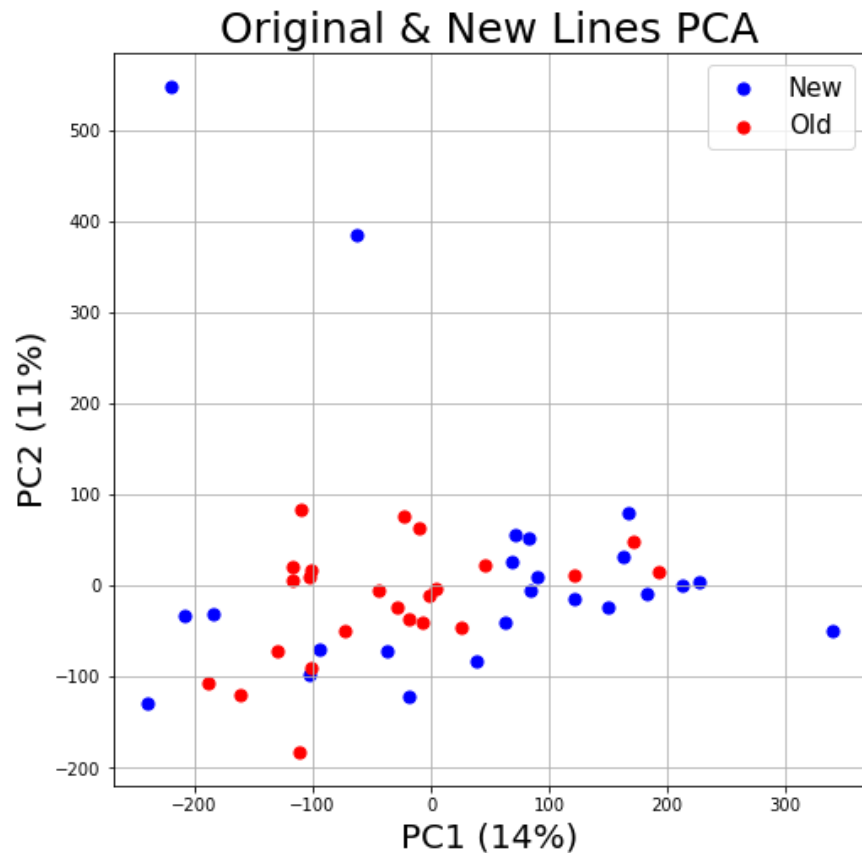
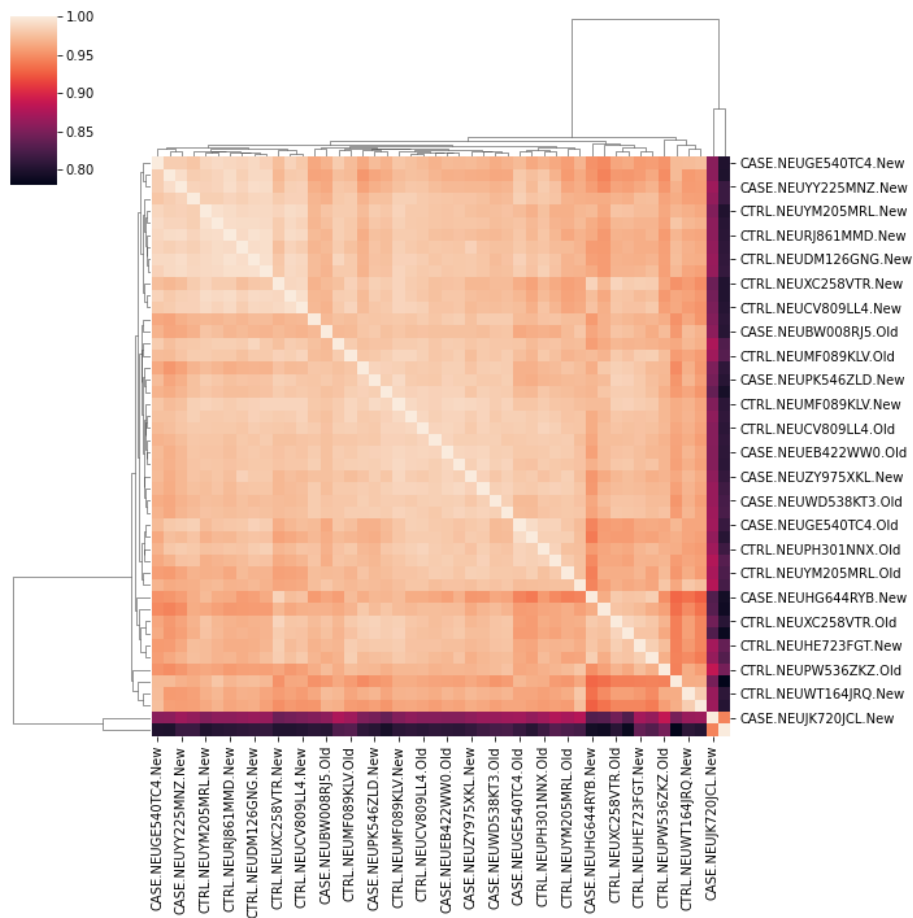# Low R^2 Value Between PC1 Coordinates (Old vs New)

PC1 Old vs New
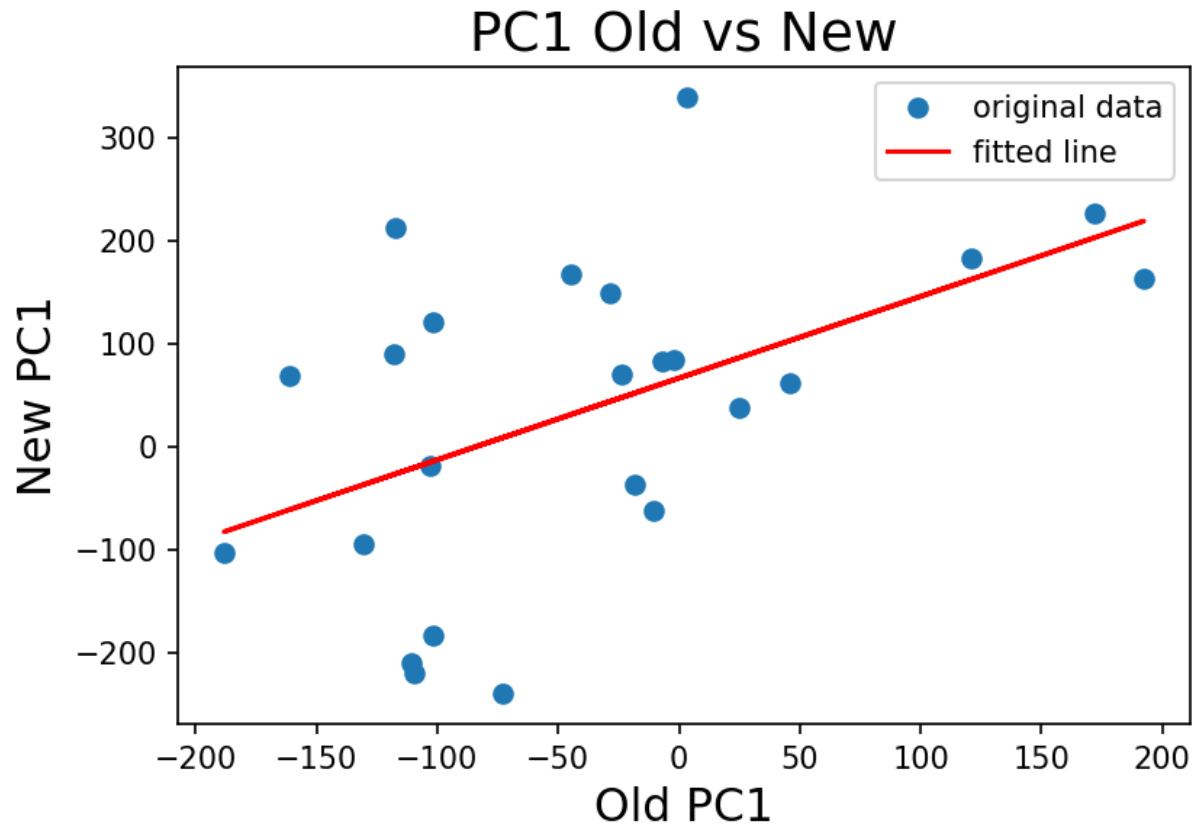
R² = .253
p = .01

R^2 Values Across PCs Increase

Correlation Heatmap & Hierarchical Clustering

Future Work

- CASE vs CTRL differential analysis using both groups

- Similarity of differential gene/peak regions

- Similarity in Gene Ontology results