MBG 326 Project:
Systemic Immune Profiling of Alopecia Areata: A Bulk DESeq2 Validation of Single-Cell Findings

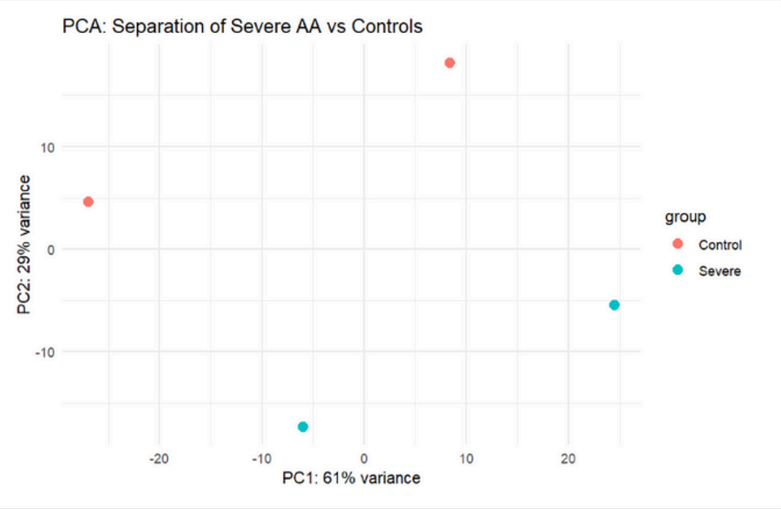
Gökdeniz Derelioğlu

December 22, 2025

Data Pipeline Used

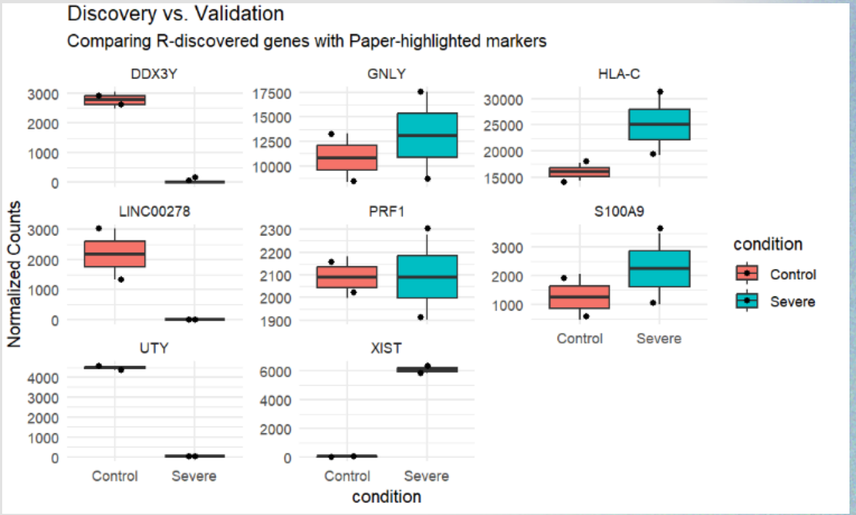
Data Acquisition:	Sourced raw single-cell feature-barcode matrices from PBMCs. Selected 2 Severe AA (ARE032, ARE040) and 2 Healthy Controls (CON205, CON209).
Preprocessing:	Utilized counts aligned to the GRCh38 human reference genome. Used Read10X to parse features and matrices.
Pseudobulk Aggregation:	Aggregated scRNA-seq UMI counts into patient-level profiles using rowSums. This collapses data from 32,453 cells to mimic bulk RNA-seq.
Differential Expression Analysis:	Constructed a DESeqDataSet to model differences between Severe and Control conditions. Filtered genes with < 10 total counts to reduce noise.
Results:	Applied Variance Stabilizing Transformation (VST) for PCA and clustering. Ranked genes by Adjusted P-value (padj < 0.05) to identify drivers from scratch.

Results



PC2 (29% Variance): Captures the primary disease split; all Controls cluster in the positive range, while all Severe AA samples cluster in the negative range.

PC1 (61% Variance): Dominated by non-disease biological variance, specifically sex-linked expression differences found in this subset.



Independently identified sex-linked genes (UTY, XIST, DDX3Y) as the most statistically significant drivers (padj < 10^-17).

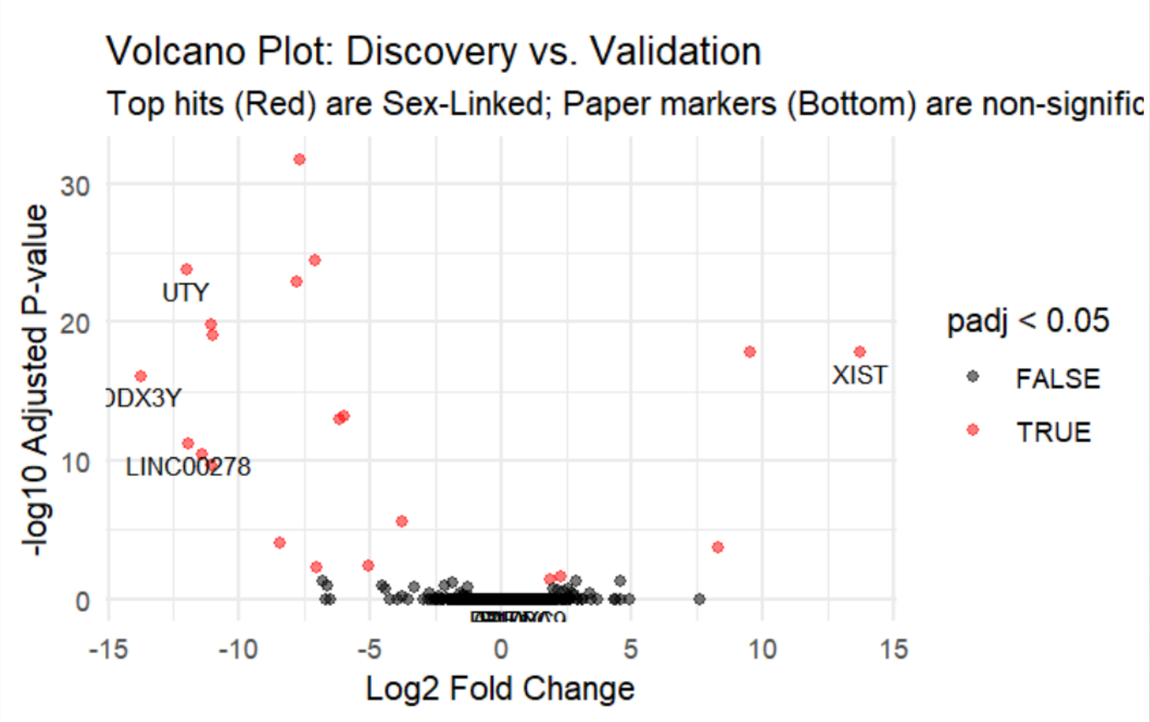
Represents a biological confounder (Male Controls vs. Female Severe) that the paper's authors mitigated via adjusted statistical models.

	log2FoldChange <dbl>	padj <dbl>
UTY	-1.197512e+01	1.883402e-24
XIST	1.373052e+01	1.657136e-18
DDX3Y	-1.372789e+01	8.373827e-17
LINC00278	-1.193477e+01	6.322850e-12
S100A9	8.407080e-01	9.996246e-01
GNLY	2.763536e-01	9.996246e-01
HLA-C	6.477441e-01	9.996246e-01
PRF1	-2.198162e-04	9.998709e-01

Markers such as S100A9 and HLA-C show visible upward trends in Severe patients, validating the paper's findings of systemic inflammation.

Statistical Limitation: These markers remain non-significant (padj approx 0.99) due to small sample size and signal dilution in the averaged pseudobulk pool.

Results Cont'd



Notice how UTY, XIST, and DDX3Y are far to the left and right and very high up. This visually confirms they have massive fold changes and extreme significance (padj < 10^{-15}).

The genes the paper cares about (S100A9, HLA-C) are buried in that bottom-center cluster. They have low -log10(padj) values because they are not statistically significant in this bulk-averaged view.

Conclusions and Comparisons

Systemic Nature of AA

Confirmed a measurable "systemic footprint" of follicular immune dysregulation in peripheral blood.

Visual upregulation of antigen-presentation (HLA-C) and monocyte activation (S100A9) even in bulk-averaged data.

Methodological Contrast

Bulk DESeq2 analysis masks the specific cytotoxic "hubs" (NK and CD8+ T cells) that the paper's single-cell map was able to isolate.

Validates the authors' premise: single-cell resolution is necessary to bypass transcriptional heterogeneity in AA.

Future Outlook

Systemic activation suggests a window for early intervention with JAK inhibitors before disease escalation.

Visual Validation

Antigen presentation (HLA-C) and monocyte activation (S100A9) are visible even in bulk-averaged data.

Thank **you!**