

Gene Expression

Distinct transcriptomic profiles of early-onset atopic dermatitis in blood of pediatric patients

Final Project

Kavya Banerjee

Introduction & Background

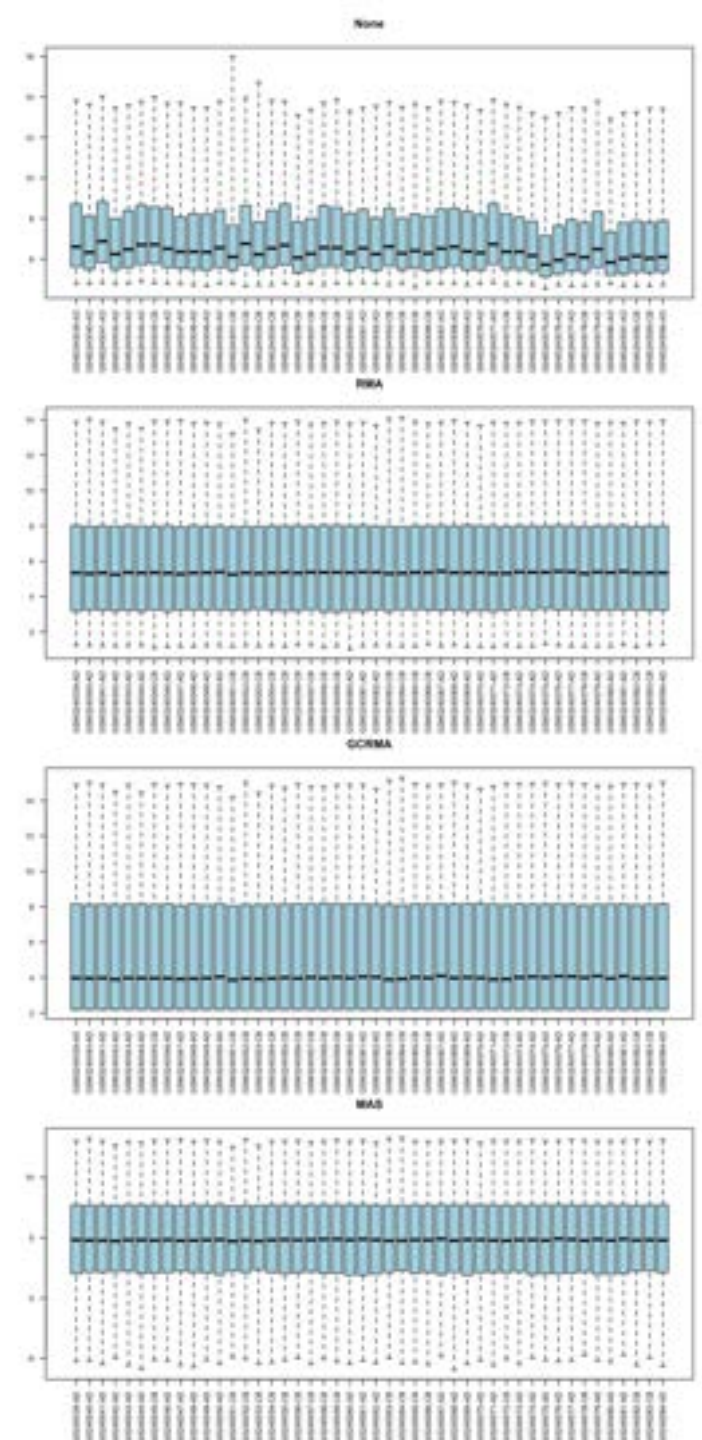
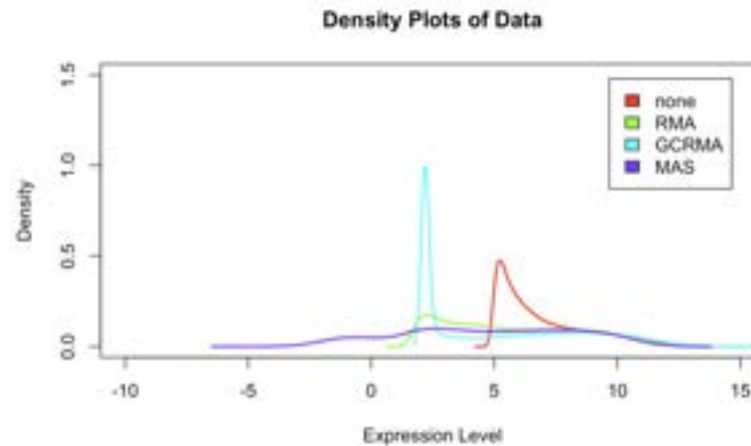
- GEO Dataset: GSE116486
- Title: Distinct transcriptomic profiles of early-onset atopic dermatitis in blood and skin of pediatric patients.
- Background
 - Atopic Dermatitis (AD): A condition predominantly affecting young children.
 - Current knowledge predominantly based on studies of adult AD using skin and blood samples from long-standing cases.
 - Recent genomic profiling of early pediatric AD biopsies revealed significant Th2 and Th17/Th22 skewing, distinct from the Th1 up-regulation seen in adults.
- Challenge: Pediatric skin biopsies are difficult to obtain, require alternative methods for understanding early AD pathogenesis.

Objective & Methods

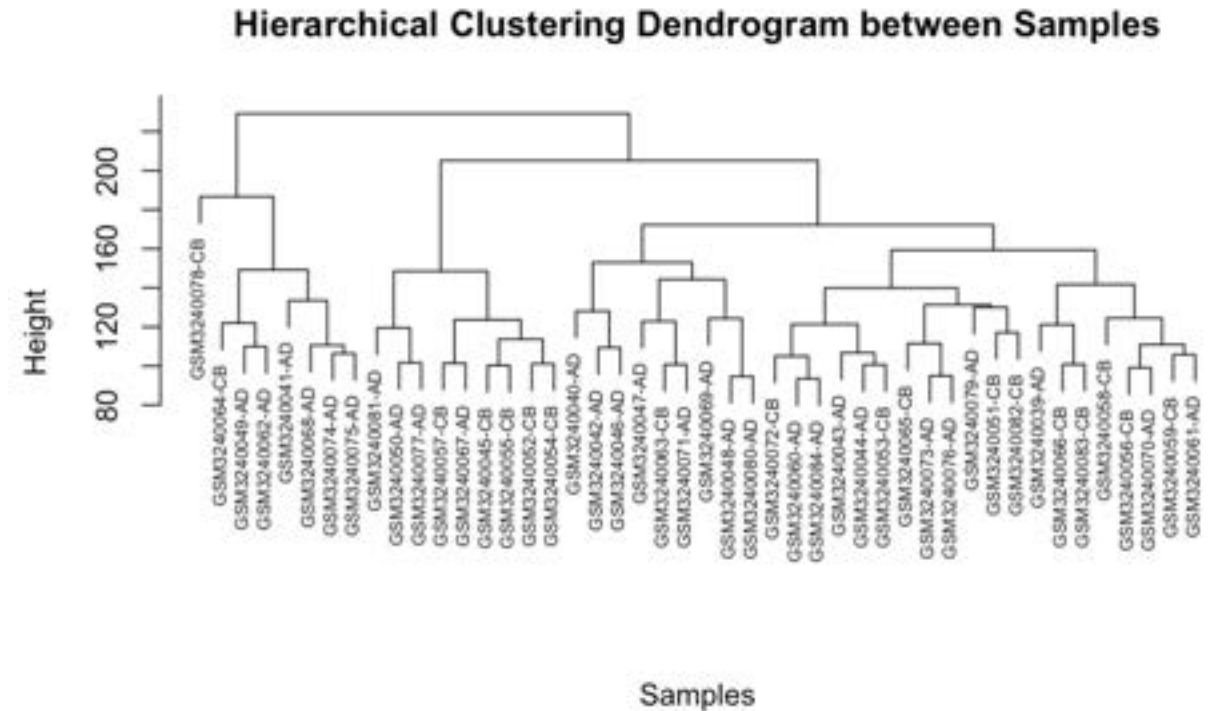
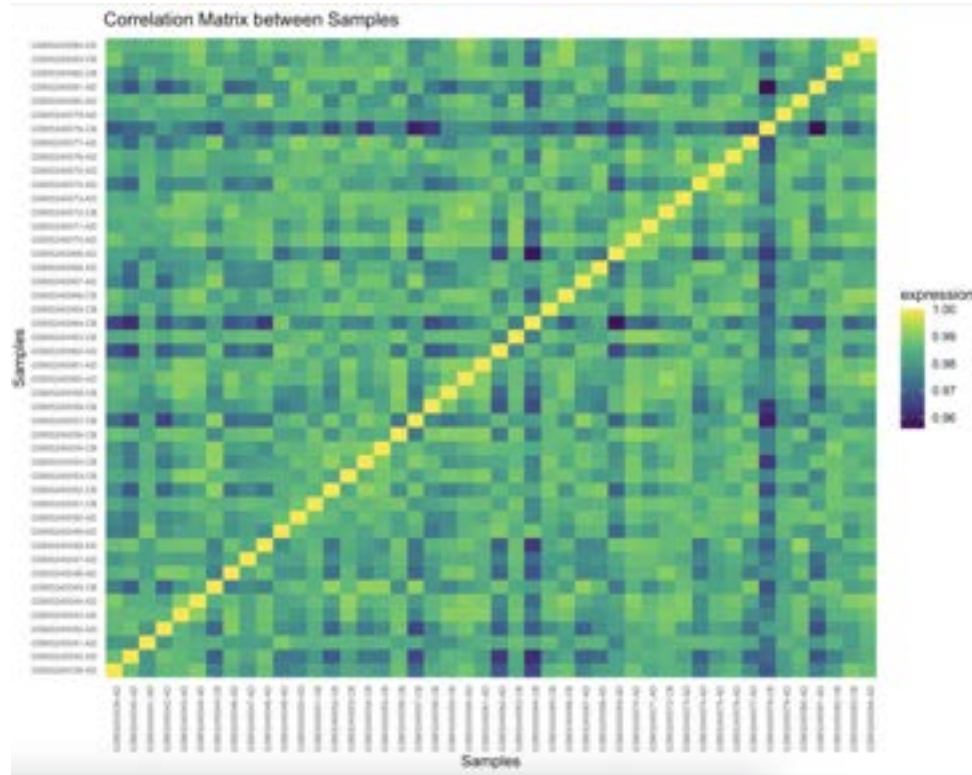
- Objective:
 - Primary Goal: To define the blood gene expression profile and identify associated biomarkers in early moderate-to-severe pediatric AD.
 - Significance: Insights into the underlying molecular mechanisms of AD in its early stages in children. Quantify systemic inflammation and contribute to unraveling new treatment targets in early AD.
- Methods:
 - Participants: Blood cells from 28 children with AD (under 5 years and within 6 months of disease onset) compared with healthy controls.
 - Sample Collection: Freshly drawn, unstimulated blood cells from the participants.
 - Approach:
 - Utilizing microarrays for gene expression analysis.
 - Identification of differentially expressed genes (DEGs) in blood (criteria: fold change >1.2, false discovery rate <0.05).
 - Microarray: [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array
 - Sample Size: n = 46 (AD = 28, CB (Healthy Control) = 18).

Normalization

- Analysis of Methods:
 - RMA & GCRMA: Similar peak locations suggesting effective normalization, maintaining a balance in data variance.
 - MAS: Exhibits the narrowest peak, potentially signifying over-normalization, which might suppress critical biological variability.
 - RMA and GCRMA are favorable choices as they demonstrate a balanced approach to normalization, maintaining central tendency and variance.
- RMA preferred over GCRMA despite higher correspondence since boxplots for RMA show a better normalization without impacting the distribution.



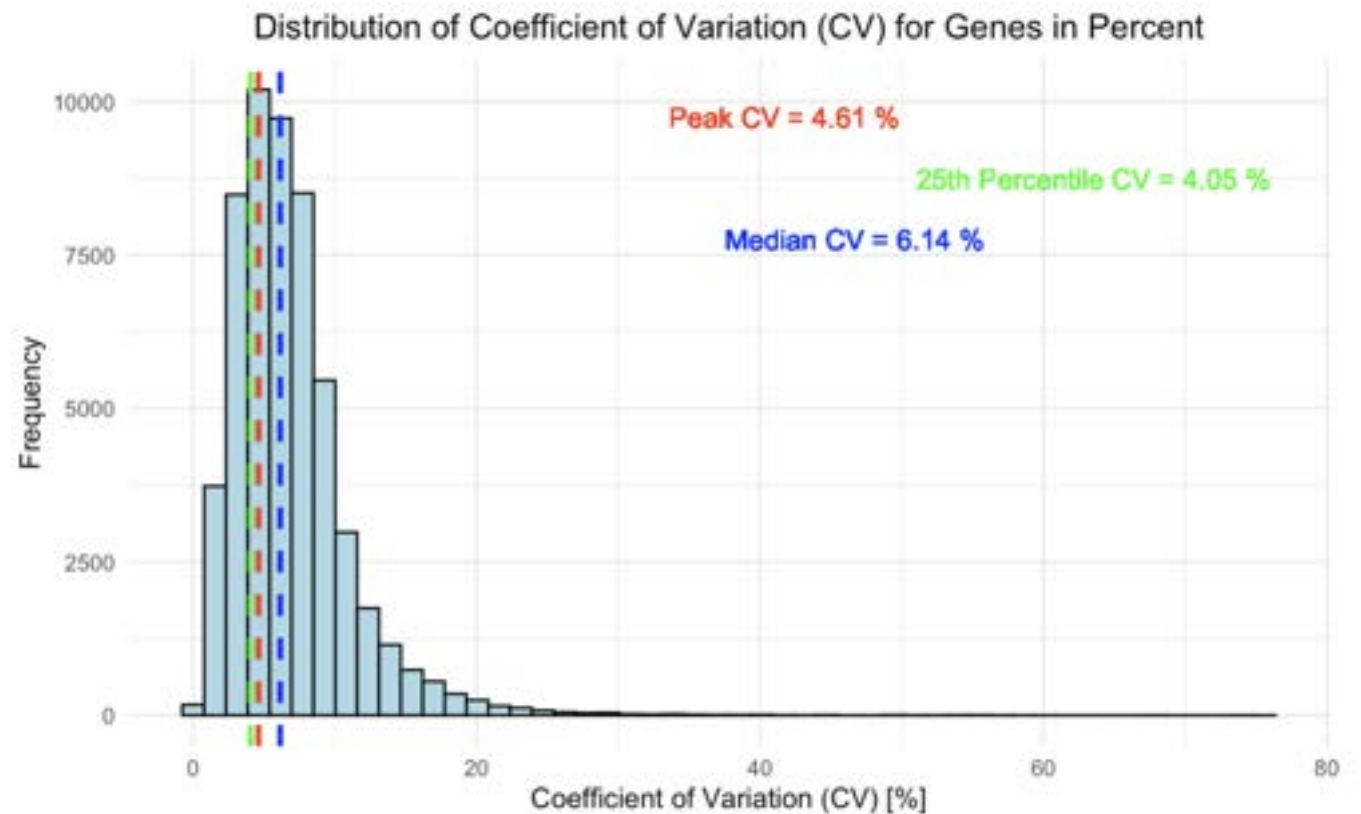
Outlier Assessment



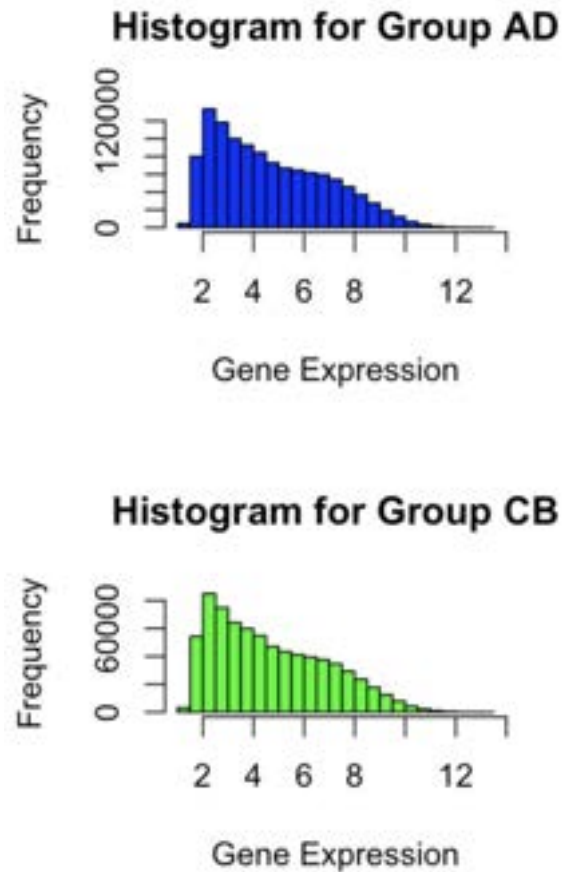
- Outlier assessment through correlation matrix and clustering of samples
- Both showed no obvious outliers; expression and clustering showed comparable grouping
- All samples considered for analysis

Filtering Low Expression Genes

- To refine gene expression data by discarding genes with minimal variation or low expression levels, which are less likely to be biologically significant.
- Reduces potential false positives, focusing on DEG.
- Method: Utilize the coefficient of variation (CV) to identify and remove genes with low variability.
- Criterion: Eliminate 25% of the low CV values



Feature Selection & Multiple Testing



- Fig 1: Data distribution follows non-normality confirmed by gene expression frequency histogram and Anderson-Darling Test p-values for each sample ($p = 3.7e-24 < 0.05$). Thus, non-parametric unpaired two-sample Wilcoxon-test was used for finding significant genes.

- Fig 2:
 - False Discovery Rate ($FDR < 0.05$) correction made to p-values to reduce false positives for multiple hypotheses testing in the analysis, conservative approach to find realistic false-positive estimate (Chen et al. 2021) .
 - Thus, the significance testing was done on $FDR (=p.adjusted < 0.05)$. Probesets with $FDR < 0.05 = 234$.

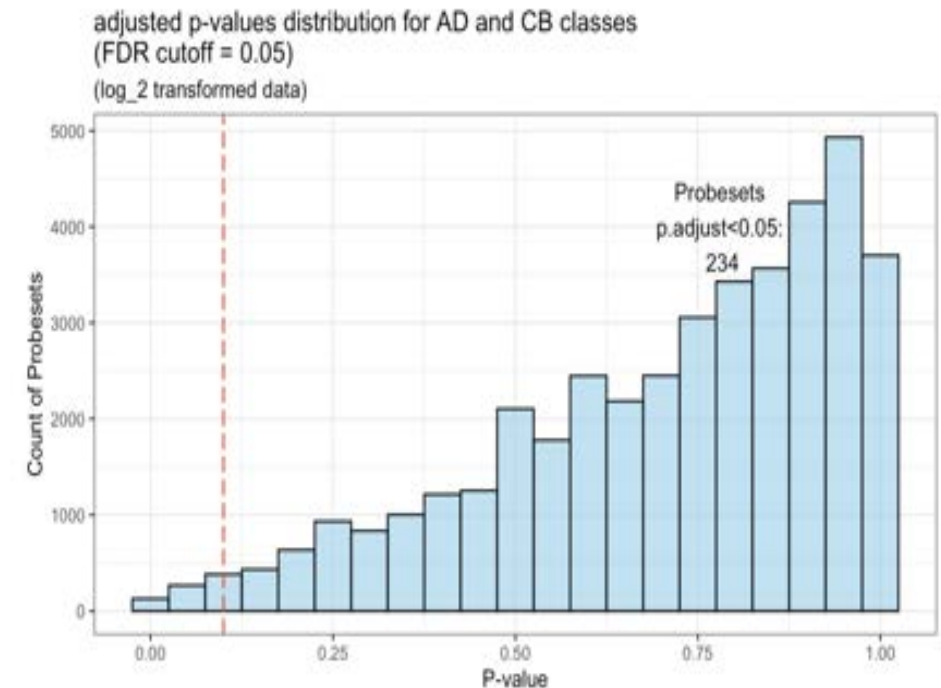


Fig 2. Adjusted p-values distribution for AD and CB groups.

Fig 1. Non-normality of data for groups AD and CB.

Feature Selection & Multiple Testing – eBayes (limma)

- For exploring other significant testing methods, eBayes from limma was used. Essentially, depends on moderated t-statistic and borrow information from ensemble of genes.
- Smoothing the variance, may change the exact p-values but often preserves the rank ordering, which is critical for gene selection.
- Same FDR correction was made to the eBayes p-values and plotted against the Wilcoxon test p-values.
- Dual significant genes ($n = 223$) was obtained (eBayes $p.adjust < 0.05$ & Wilcoxon $p.adjust < 0.05$) were obtained.

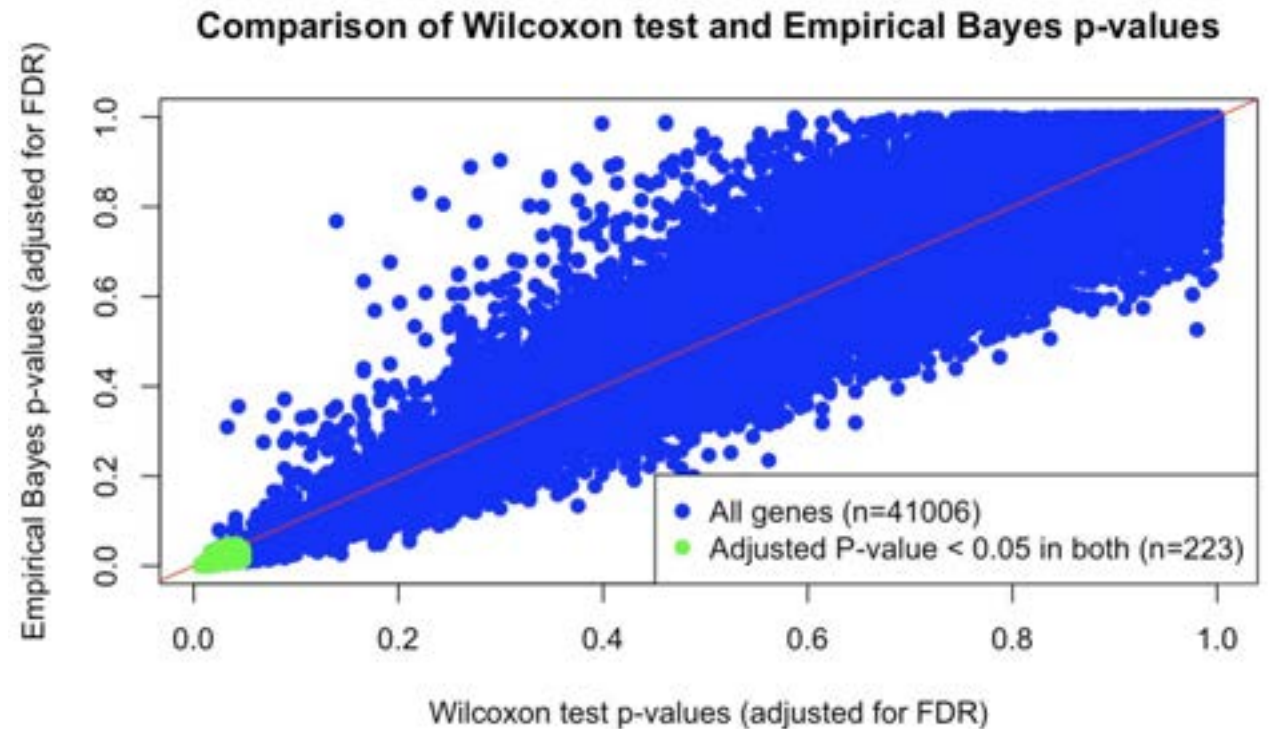


Fig 3. Adjusted p-values comparison for eBayes vs Wilcoxon p-values. Dual significant genes marked when $p.adjust < 0.05$

DEG Analysis – Volcano Plot

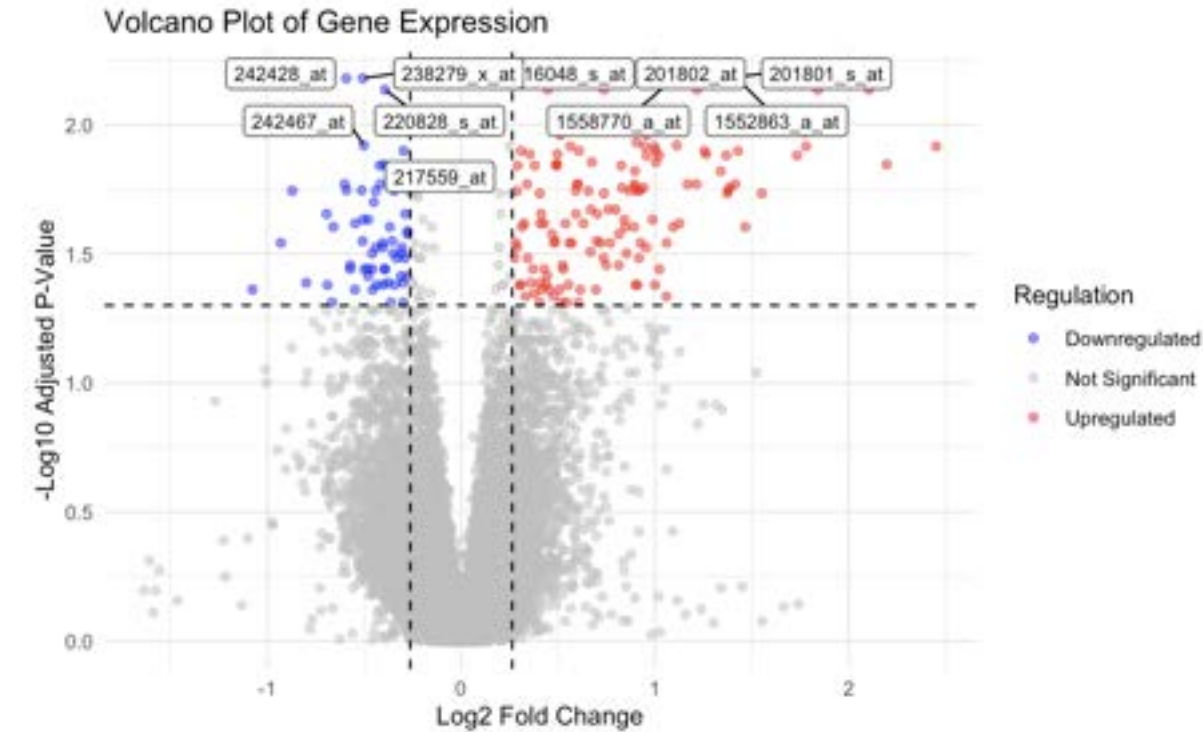


Fig 4. Volcano plot for genes and their regulation. Regulation and significance plotted.

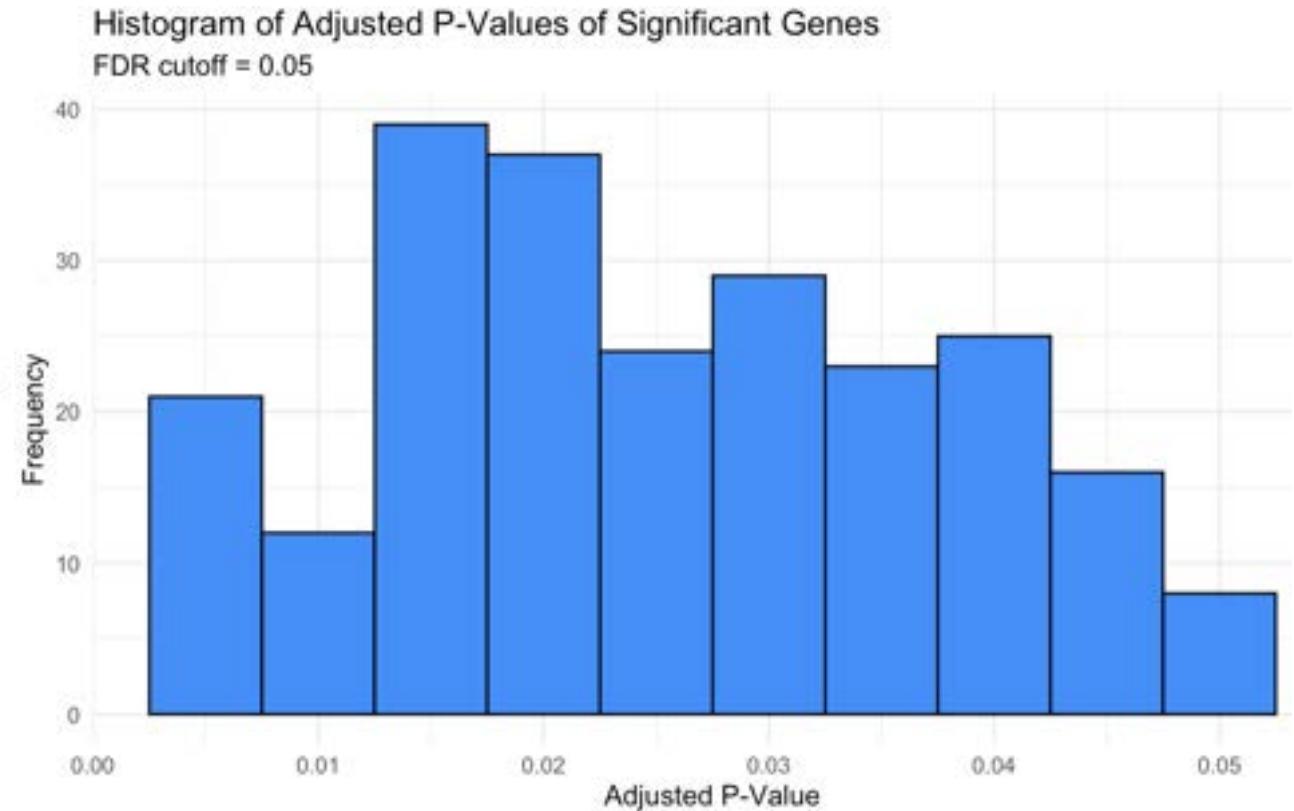


Fig 5. Histogram of significant genes with adjusted p-values.

- Volcano plot for the all genes to find DEGs. Criterion for DEGs: $p_{\text{adjusted}} \text{ (FDR)} < 0.05$ and absolute Fold Change (FC) > 1.2 . Total 204 DEGs were found (142 upregulated and 62 downregulated). Top 5 upregulated and downregulated genes (probesets) plotted.
- Histogram of the adjusted p-values of significant plotted majority of the genes fall below the threshold.

Dimensionality Reduction

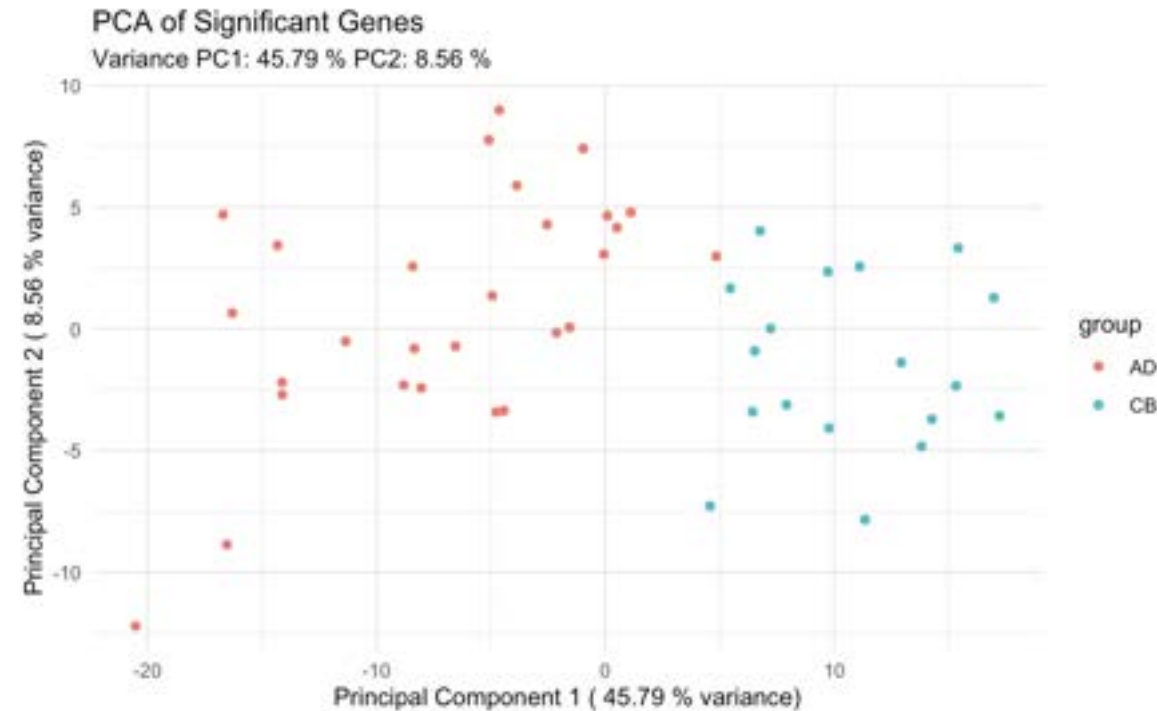


Fig 6. PCA plot showing the clustering and grouping of samples

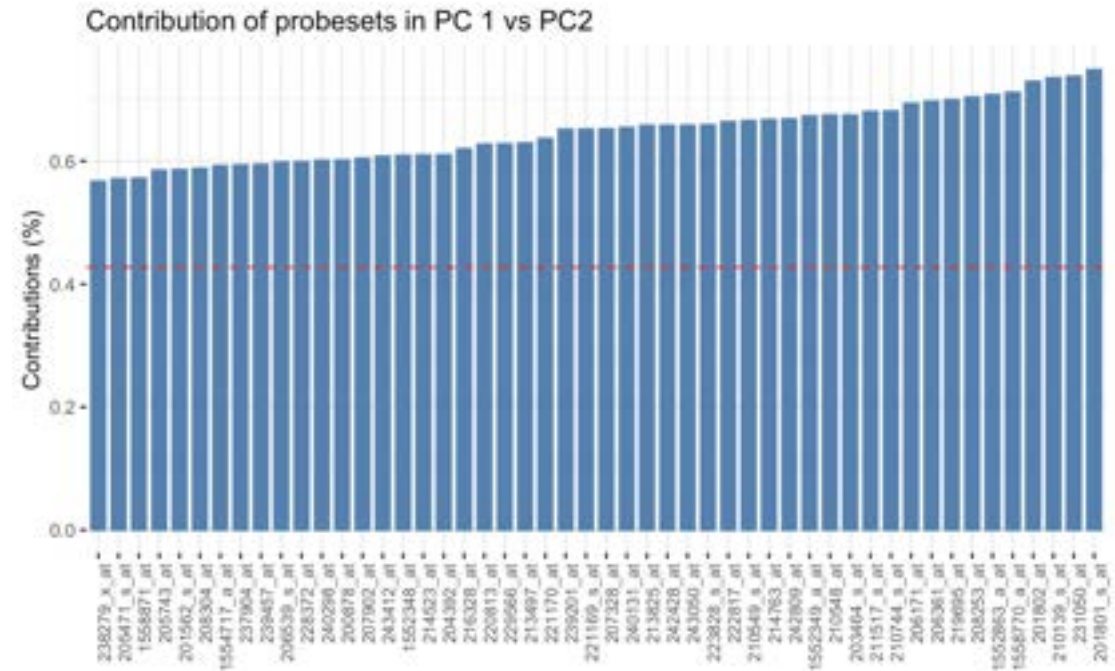


Fig 7. Contribution of probesets in PC 1 vs PC 2

- PCA is ideal for gene expression data analysis as it condenses high-dimensional data into principal components for easier visualization and interpretation, filters out noise, reveals patterns, prepares data for further statistical analysis, and efficiently handles large datasets with its scalable and computationally effective approach.
- Fig 6. PCA show some level of separation between the two groups. Visible overlap between the AD and CB samples, particularly around the center, implying significant genes differentiate to some extent, they do not create completely distinct clusters for AD and CB. Outliers in AD group from main cluster, likely due to biological validation or other underlying causes. Scree plot (not shown) shows subsequent components after PC1 explain variance lesser.
- Fig 7. Contribution of probesets in PC 1 vs PC 2. Top probesets such as 201801_s_at and 201802_at from DEG analysis show significant contribution.

Clustering

- Both CB and AD groups show similar expression in most significant genes but some show more pronounced expression such as probesets 201801_s_at, 223828_S_at and 201802_at.
- Euclidean and Complete Linkage done for its simplicity and less computational requirement.
- Doesn't require a predefined number of clusters (suitable for exploratory analysis)
- Sensitivity to outliers helps in identifying closely related gene expressions, reflecting possible shared biological pathways or functions (Eisen et al, 1998).

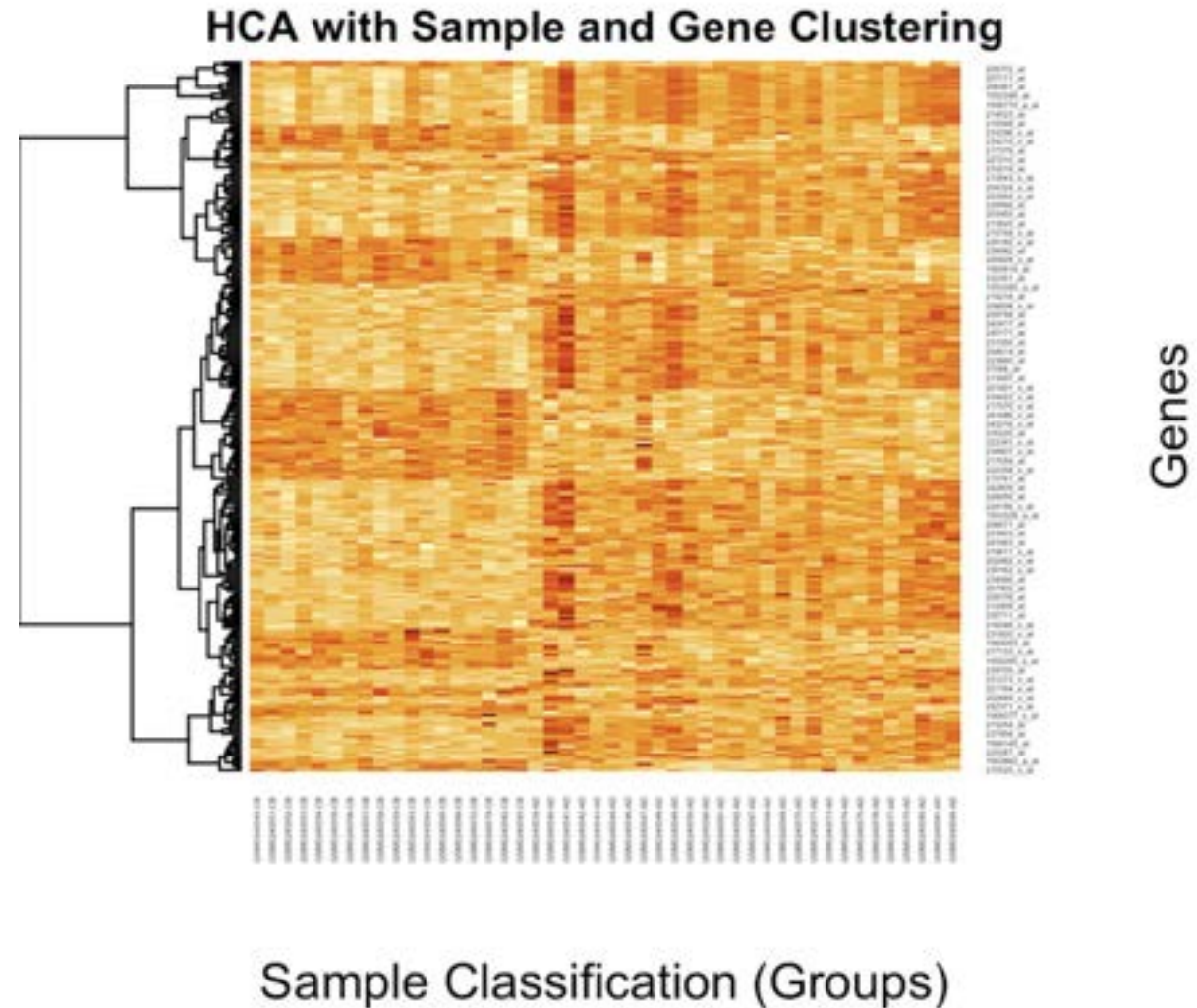


Fig 8. Clustering of Samples by groups and genes

Classification

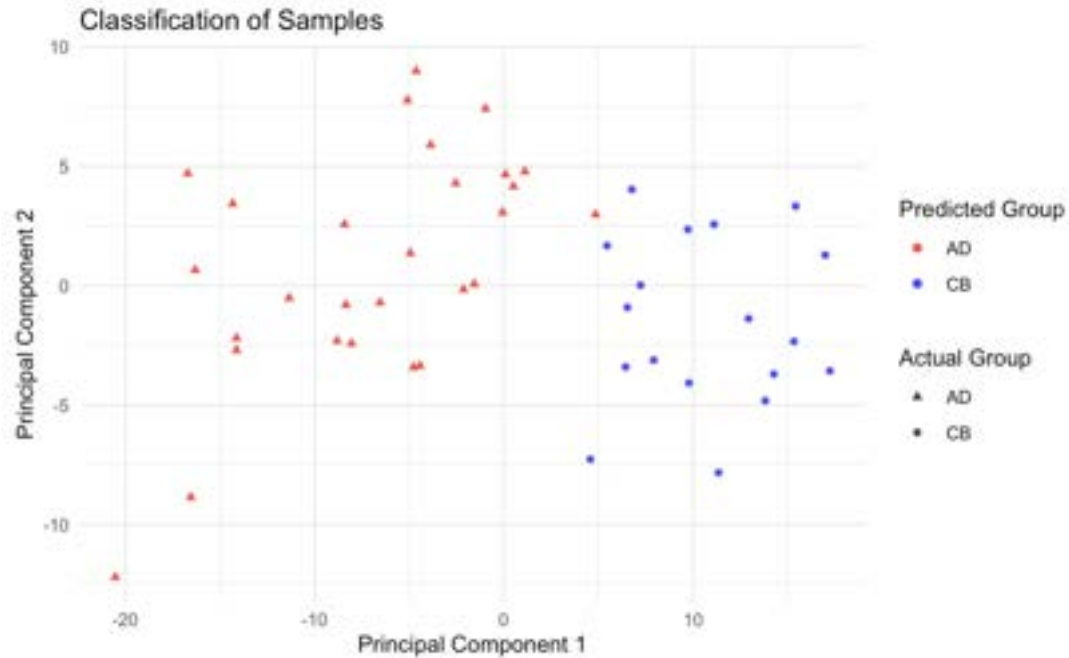


Fig 9. PCA plot of predicted and actual group classification on LRM.

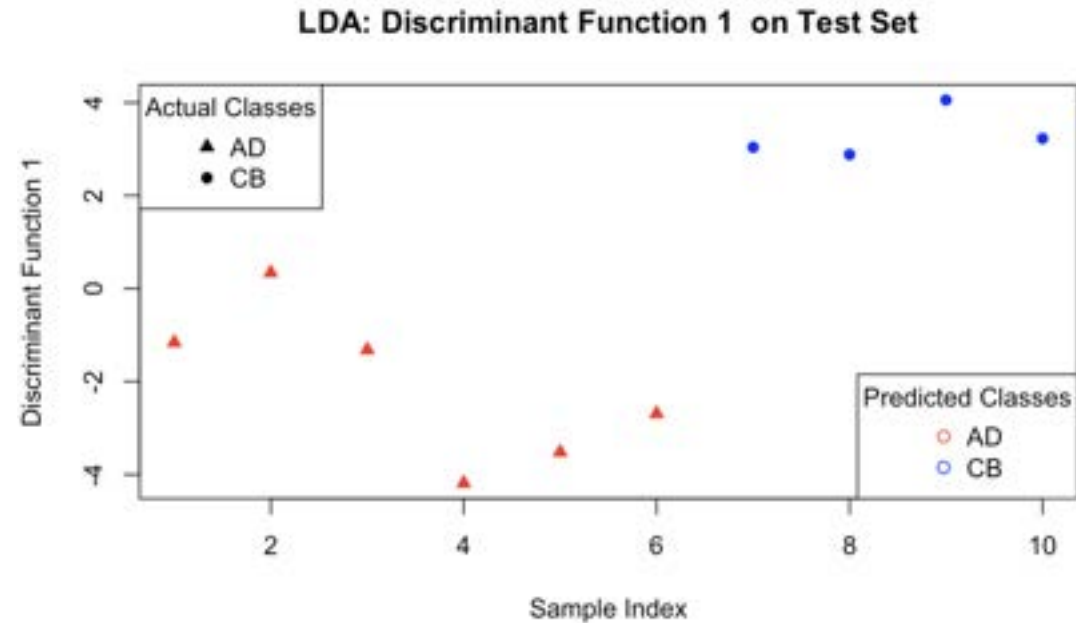


Fig 10. Plot of predicted and actual group LDA.

- Logistic Regression Model (LRM) chosen for its robustness in binary classification (Stoltzfus, 2011). However, misclassification 1 sample.
- LDA method used If the classes are well-separated (by PCA plot) since most samples are not overlapping and number of features is high relative to the number of observations. Based on 80:20 ratio, the LDA model classified all samples in test set correctly.

Gene Enrichment Analysis

Significant probes:

1552863_a_at

1558770_a_at

201801_s_at

201802_at

216048_s_at

238279_x_at

242428_at

220828_s_at

242467_at

217559_at

NCBI DAVID Output:

| 216048_s_at | Rho-related BTB domain containing 3 (RHORTB3) | Related Genes | Human sapiens |
|--------------------------|--|---------------|---------------|
| GOTERM_BP_DIRECT | actin filament organization, establishment or maintenance of cell polarity, small GTPase mediated signal transduction, regulation of cell shape, male gonad development, vesicle-mediated transport, cell migration, regulation of proteolysis, cortical cytoskeleton organization, regulation of actin cytoskeleton organization, retrograde transport, endosome to Golgi, proteasome-mediated ubiquitin-dependent protein catabolic process. | | |
| GOTERM_CC_DIRECT | cytoplasm, cytosol, cytoskeleton, plasma membrane, cell cortex, cytoplasmic vesicle, trans-Golgi network membrane, cell projection, extracellular exosome. | | |
| GOTERM_MF_DIRECT | GTPase activity, protein binding, ATP binding, GTP binding, ATPase activity, protein kinase binding, small GTPase binding, ubiquitin protein ligase binding. | | |
| INTERPRO | BTB/POZ-like, Small GTPase superfamily, Small GTPase superfamily, Rho type, BTB/POZ fold, P-loop containing nucleoside triphosphate hydrolase. | | |
| SMART | BTB | | |
| UP_KW_BIOLOGICAL_PROCESS | Transport. | | |
| UP_KW_CELLULAR_COMPONENT | Golgi apparatus. | | |
| UP_KW_DOMAIN | Repeat. | | |
| UP_KW_LIGAND | ATP-binding, Nucleotide-binding. | | |
| UP_KW_MOLECULAR_FUNCTION | Hydrolase. | | |
| UP_SEQ_FEATURE | DOMAIN:BTB, DOMAIN:BTB 1, DOMAIN:BTB 2, MUTAGEN:A->T: Abolishes interaction with RAB9A., MUTAGEN:D->E: Abolishes interaction with RAB9A., MUTAGEN:I->K: Abolishes interaction with RAB9A., MUTAGEN:Missing: Does not affect subcellular location, suggesting this protein is not prenylated., MUTAGEN:N->D: Abolishes ATP-binding., REGION:Disordered, REGION:Interaction with Rab9, REGION:Rho-like. | | |
| 1552863_a_at | calcium voltage-gated channel auxiliary subunit gamma 6(CACNG6) | Related Genes | Human sapiens |
| GOTERM_BP_DIRECT | calcium ion transport, regulation of ion transmembrane transport, calcium ion transmembrane transport, regulation of calcium ion transmembrane transport via high voltage-gated calcium channel. | | |
| GOTERM_CC_DIRECT | plasma membrane, voltage-gated calcium channel complex, integral component of membrane, L-type voltage-gated calcium channel complex. | | |
| GOTERM_MF_DIRECT | voltage-gated ion channel activity, voltage-gated calcium channel activity, calcium channel regulator activity, calcium channel activity. | | |
| INTERPRO | PMP-22/EMP/MP20/Claudin, Voltage-dependent calcium channel, gamma subunit, Voltage-dependent calcium channel, gamma-6 subunit. | | |
| KEGG_PATHWAY | MAPK signaling pathway, Cardiac muscle contraction, Adrenergic signaling in cardiomyocytes, Oxytocin signaling pathway, Hypertrophic cardiomyopathy, Arrhythmogenic right ventricular cardiomyopathy, Dilated cardiomyopathy. | | |
| UP_KW_BIOLOGICAL_PROCESS | Calcium transport, Ion transport, Transport. | | |
| UP_KW_CELLULAR_COMPONENT | Membrane, Cell membrane. | | |
| UP_KW_DOMAIN | Transmembrane, Transmembrane helix. | | |
| UP_KW_LIGAND | Calcium. | | |
| UP_KW_MOLECULAR_FUNCTION | Calcium channel, Ion channel, Voltage-gated channel. | | |
| UP_SEQ_FEATURE | REGION:Disordered, TRANSMEM:Helical. | | |
| 1558770_a_at | phosphoinositide 3-kinase regulatory subunit 6(PIK3R6) | Related Genes | Human sapiens |
| GOTERM_BP_DIRECT | angiogenesis, immune response, G-protein coupled receptor signaling pathway, regulation of natural killer cell mediated cytotoxicity, positive regulation of MAP kinase activity, positive regulation of T cell differentiation, positive regulation of angiogenesis. | | |
| GOTERM_CC_DIRECT | cytosol, plasma membrane, phosphatidylinositol 3-kinase complex, phosphatidylinositol 3-kinase complex, class 1A, phosphatidylinositol 3-kinase complex, class 1B, membrane. | | |
| GOTERM_MF_DIRECT | protein binding, kinase activity, phosphatidylinositol-4,5-bisphosphate 3-kinase activity, 1-phosphatidylinositol-3-kinase regulator activity. | | |
| INTERPRO | Phosphoinositide 3-kinase 1B, gamma adapter, p101 subunit. | | |
| KEGG_PATHWAY | cGMP-PKG signaling pathway, Chemokine signaling pathway, Phospholipase D signaling pathway, PI3K-Akt signaling pathway, Adrenergic signaling in cardiomyocytes, Aelin signaling pathway, Platelet activation, Cholinergic synapse, Oxytocin signaling pathway, Toxoplasmosis, Kaposi sarcoma-associated herpesvirus infection. | | |
| UP_KW_BIOLOGICAL_PROCESS | Angiogenesis. | | |
| UP_KW_CELLULAR_COMPONENT | Membrane, Cytoplasm, Cell membrane. | | |
| UP_KW_MOLECULAR_FUNCTION | Kinase, Transferase. | | |
| UP_SEQ_FEATURE | REGION:Disordered. | | |

Gene Enrichment Analysis

Significant probes:

1552863_a_at

1558770_a_at

201801_s_at

201802_at

216048_s_at

238279_x_at

242428_at

220828_s_at

242467_at

217559_at

NCBI DAVID Output (con't)

[illegible]

Gene Enrichment Analysis

Significant probes:

1552863_a_at
1558770_a_at
201801_s_at
201802_at
216048_s_at
238279_x_at
242428_at
220828_s_at
242467_at
217559_at

Affymetrix Output
(accessed through
<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL570>)

| ID | GB_ACC | SPOT_ID | Gene.Symbo | RefSeq.Transcript.ID | Gene.Ontology.Biological.Process | Gene.Ontology.Cellular.C | Gene.Ontology.Molecular.Function |
|--------------|-----------|---------|------------|--|---|----------------------------|---|
| 216048_s_at | AK023621 | | RHOBTB3 | NM_014899 | 0006200 // ATP catabolic process // inferred from | 0005794 // Golgi apparatus | 0000166 // nucleotide binding // inferred from |
| 1558770_a_at | AK091819 | | PIK3R6 | NM_001010855 /// NM_001290211 | 0001525 // angiogenesis // inferred from el | 0005737 // cytoplasm // | 0005515 // protein binding // inferred from phy |
| 238279_x_at | BF062155 | | | | | | |
| 242428_at | N58513 | | | | | | |
| 220828_s_at | NM_018382 | | FLJ11292 | NM_018382 | | | |
| 242467_at | BF433200 | | | | | | |
| 217559_at | AI001784 | | RPL10L | NM_080746 | 0006412 // translation // inferred from elec | 0005622 // intracellular / | 0003735 // structural constituent of ribosome / |
| 1552863_a_at | NM_145815 | | CACNG6 | NM_031897 /// NM_145814 /// NM_0006810 | // transport // inferred from electr | 0005891 // voltage-gated | 0005244 // voltage-gated ion channel activity / |
| 201802_at | NM_004955 | | SLC29A1 | NM_001078174 /// NM_001078175 | 0006139 // nucleobase-containing compoun | 0005886 // plasma mem | 0005337 // nucleoside transmembrane transpo |
| 201801_s_at | AF079117 | | SLC29A1 | NM_001078174 /// NM_001078175 | 0006139 // nucleobase-containing compoun | 0005886 // plasma mem | 0005337 // nucleoside transmembrane transpo |

- Majority of the enrichment show blood markers, ribosomal and transport protein markers.
- However, most show no major functional difference between AD and CB groups unlike in skin samples (Brunner et al., 2016)
- Immune Response Markers :
 - 1558770_a_at maps to phosphoinositide-3-kinase regulatory subunit 6(PIK3R6) involved in immune response. DEG significantly upregulated. Involved in MAPK signalling modulating Th2 expression. Interleukin-19 (IL-19) is implicated in the modulation of Th2 cytokine expression via MAPK signaling pathways.
 - IL-19 is a pro-inflammatory cytokine essential for the production of Th2 cytokines. Stimulated by IL-17, it serves as a bridge between the Th17 and Th2 cytokine pathways. Significant activation of both Th17 and Th2 cytokine pathways is observed in pediatric AD. In contrast, adult AD shows a dominant Th2 activation with less Th17 involvement (Esaki et al, 2016). IL-19 is potentially crucial for enhancing Th2 cytokine secretion in AD.
 - Suggested to play a similar role in other atopic diseases, such as asthma (Kannan et. al, 2012).
 - A strong Th2 signature in early-onset pediatric AD implicates IL-19 and related pathways as promising therapeutic targets.

Brunner, P. M., Israel, A., Zhang, N., ... Estrada, Y. D., Xu, H., Krueger, J. G., Paller, A. S., & Guttman-Yassky, E. (2018). Early-onset pediatric atopic dermatitis is characterized by TH2/TH17/TH22-centered inflammation and lipid alterations. *The Journal of allergy and clinical immunology*, 141(6), 2094–2106.

Esaki, H., Brunner, P. M., Renert-Yuval. (2016). Early-onset pediatric atopic dermatitis is TH2 but also TH17 polarized in skin. *The Journal of allergy and clinical immunology*, 138(6), 1639–1651. <https://doi.org/10.1016/j.jaci.2016.07.013>

Kannan, Y., & Wilson, M. S. (2012). TEC and MAPK Kinase Signalling Pathways in T helper (TH) cell Development, TH2 Differentiation and Allergic Asthma. *Journal of clinical & cellular immunology*, Suppl 12, 11. <https://doi.org/10.4172/2155-9899.S12-011>

Conclusions

- A limited array of genes were identified as being dysregulated in the blood of children with AD compared to healthy controls, a contrast to the broader gene dysregulation seen in the skin.
- The gene expression patterns in blood shed light on the initial systemic immune response characteristic of early-onset pediatric AD. However, these patterns show minimal correlation with the clinical and molecular indicators of the disease observed in the skin, excluding a select number of immune markers, such as IL-19 and Th1 markers.
- This information about systemic blood inflammation could enhance our understanding of the mechanisms that contribute to the progression of the atopic march and could be instrumental in developing treatments aimed at interrupting this progression.