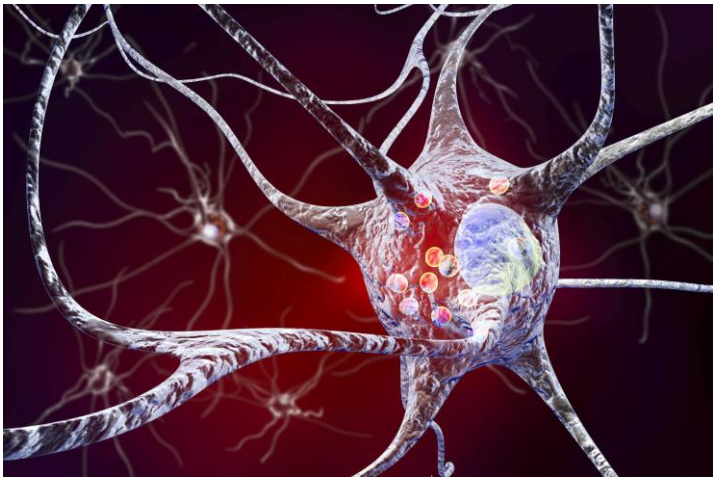

# Single-Cell RNA Sequencing of Peripheral Blood in Alzheimer's Disease

Murad Ektilat, Noam Wolf

Intro to Bioinformatics
236523, Technion

## Abstract

Alzheimer's disease (AD) is recognized by the World Health Organization as a global public health priority as it is the most common cause of dementia in the elderly and a major cause of death. However, we still don't have the full knowledge and understanding of the pathogenesis of the disease, the biological mechanisms involvement in its development and the different risk factors.

In our analysis, we mainly aim to better understand the involvement of the immune system in AD pathogenesis. To do so, we have performed SC RNA-seq analysis including Dimensionality reduction, Clustering and Cell type annotation to find the different immune cells distribution in AD compared with normal control (NC) group. In addition, we've performed differential gene expression and Functional and pathway enrichment analysis to better understand the characteristics and biological role of highly expressed genes in our data.

The main conclusion of our analysis is that the immune system reacts and is involved in AD pathogenesis. This conclusion is based on our results showing differences in the immune cells distribution in said group. We have found that numbers of CD8+ T cells and NK cells were significantly increased in the AD group (44.6% and 52.3%, respectively) compared with NC (25.8%, 22.4% respectively). while the proportions of B cells and CD4+ T cells were significantly lower in the AD group (2.2% and 0% respectively) than in NC (11.8% and 38.7% respectively).

# Literature review

## General Overview

Alzheimer's disease, the most common cause of dementia in the elderly (Okuzono et al., 2021), was first described by the physician Alois Alzheimer in 1907. Where in the case study of a 51-year-old woman with rapidly deteriorating memory and psychiatric disturbances was described.

Alzheimer disease accounts for 50% to 75% of progressive cognitive impairment in elderly patients. The prevalence of AD increases twofold every 5 years after 60, with 1% of 60-65 years suffering from the disease all the way up to 40% of those aged 85 years and older. The number of Alzheimer patients is set to triple in the next 50 years if current trends continue. It is a severe problem, with the annual cost of dementia in the USA alone set to cross 600 billion USD. It is also the major cause of death overall in England, causing 11% of all deaths in 2015, and causing death 8.5 years after presentation, on average (Lane et al., 2018).

Symptoms include progressive issues with episodic memory, spatial and topographical problems, and difficulties in multitasking and a loss of confidence. As the disease progresses so do symptoms, with patients becoming more dependant, increasing behavior changes and the emergence of hallucinations and seizures (Lane et al., 2018).

## Aetiology

It seems that AD occurs on an apparently random basis. The typical late onset AD seems to be caused by complex interactions between genetic and environmental factors, with around ~70% attributed to genetics (Lane et al., 2018). There are many immune related gene loci that have been suggested to be closely related to the pathogenesis of AD, including TREM2, CD33, and MS4A (Okuzono et al., 2021; Xu and Jia, 2021a). The dysregulation of the immune system play an important role in the development of AD (Xu and Jia, 2021a). In particular, the microglial activation in response to amyloid deposition. In general, it seems that each individual risk factor contributes only a very small increased risk to the subject, but the aggregated risk can be used to predict AD by more than double chance (Lane et al., 2018). Education and exercise seem to protect against AD, while hypertension and diabetes have a negative affect, although it remains unclear why vascular issues play a role in AD development.

## Pathology

The cardinal and most accepted features of AD pathology are amyloid plaques and neurofibrillary tangles (NFTs) (Lane et al., 2018; Xu and Jia, 2021a). However, the classic theory of the amyloid cascade does not seem to fully explain the pathogenesis. Amyloid plaques ae abnormally folded peptides, while NFTs are helical filaments made of tau. The accumulation of said plaques, driven by an imbalance of production and clearance is believed to be the primary pathological process, while the NFTs are thought to be resulting processes. the plaques appear to cause neurodegeneration and synaptic loss. However there are many cases where in individuals with a high burden of AD pathology display no symptoms, leaving the definite pathology still open to discussion (Lane et al., 2018). Genetic studies have shown that the loss of function of TREM2 receptors increases the risk of AD (Okuzono et al., 2021).

**Treatment**

There is currently no treatment that halts or restricts the disease progression, only providing symptomatic relief (Lane et al., 2018; Okuzono et al., 2021). Acetyl-cholinesterase inhibitors are used to increase acetylcholine availability, while Memantine has been shown to have a small but clinically appreciable benefit on cognition and functional decline in patients with moderate to severe AD. Both treatments, however, provide modest improvements at best.

## Still unknown and main challenges

The classic hypothesis of the amyloid cascade cannot fully explain the pathogenesis of AD. The failure of several major clinical trials targeting cerebral β-amyloid has prompted scepticism about the classic amyloid hypothesis and about prospects for disease modification in AD more generally (keeping in mind there might be other explanations to the failures) (Lane et al., 2018b). Hence, significant unresolved challenges are fully identifying AD's risk factors and further understanding the biological mechanisms involved in AD.

As mentioned above, studies have identified many genetic and environmental risk factors. Yet the role of some factors remains unclear. For example, obesity has long been considered a risk for dementia and AD, but this has recently been questioned.
Likewise, the mechanisms by which vascular risk factors might influence AD remain unclear as well. Vascular risk factors may increase the risk of clinical AD through a 'double-hit' with superimposed cerebrovascular damage, or vascular damage might influence the development of AD pathology directly (Lane et al., 2018b).

Another challenge is understanding the link between known key factors. Perhaps most significantly, whilst amyloid and tau pathology are clearly critical in the pathogenesis of AD, how the two are mechanistically linked is unclear (Lane et al., 2018b).

**Involvement of the immune system**

The immune system is a major mechanism whose involvement in the disease is yet to be fully understood. Many types of immune cells, involving innate and adaptive immunity, containing monocytes, macrophages, neutrophils, and T cells from peripheral blood may participate in the pathogenesis of AD. However, the current understanding of the distribution of peripheral blood immune cells in patients with AD is limited to flow cytometry study, and the cell type-specific functional status of immune cells, especially T and B cells, remains unclear. In addition, none of the studies has reported exploring the adaptive immune repertoire in the peripheral blood of patients with AD. (Xu and Jia, 2021b)

## How analysing genomic data can help to overcome those challenges?

Many of the challenges we've described above address the questions of which mechanisms are involved in the disease pathogenesis, and what is their part in it. Based on genomic data we can compare the presence of different genomic factors between different groups. With this comparison, we can better understand the involvement of proteins or wider mechanisms in the disease pathogenesis. For example, based on gene expression analysis for AD and NC groups, researchers have explored

1) The development of NFTs (Neurofibrillary tangles) and vivo molecular processes (Dunckley et al., 2006).

2) The involvement of the immune system based on GEA in peripheral blood (Xu & Jia, 2021).

## Biological question

We will explore how gene expression differs in peripheral immune cells between AD patients and cognitively normal individuals. We are looking for key genes and pathways that may differ between said groups. We will also be analysing the difference in overall composition of the peripheral immune system. These findings may help us address the challenges mentioned above in multiple ways. It has long been recognized that the conventional hypothesis for the pathogenesis of AD does not fully explain the disease, however it is thought that the immune system plays an important role, whether it be in clearing amyloid deposits or inflammation. The identification of relevant pathways and key genes may help solve the missing piece in the puzzle, as recognizing specific genes or pathways that are better understood can shed light on the problem. This can also contribute to developing treatment methods, as conventional treatment methods based on symptom relief have had little success. If however the mechanism is identified it may be possible to treat. Moreover, a key issue with AD is proper diagnosis. The presence of amyloid plaques and NFTs are required for a diagnosis of AD. These are not easily tested for. The findings may contribute to further research regarding the development of biomarkers or therapies, which could result in earlier diagnosis and thus improved treatment.

## Results

Researchers believe the peripheral immune system affects the pathology of the central nervous system in Alzheimer's disease. However, we don't have yet the adequate knowledge for understanding the characteristics of peripheral immune cells in AD. In our analysis we aim to explore the molecular basis of peripheral immune cells and the features of adaptive immune repertoire at a single cell level.

We used single-cell RNA sequencing data of Peripheral Blood from 5 samples (patients): 3 AD patients with amyloid-positive status and 2 normal control patients with amyloid-negative status (overall - 36,849 mononuclear cells). The data was taken from GSE181279.

Our main goal in this analysis is to find the distribution of cell-types in AD patients and NC patients. First, we preformed SC RNA-seq analysis using Seurat to find the most differently expressed genes in said groups. We preformed the analysis on the Count matrix and the corresponding barcodes and features.

Using this data, we preformed Dimensionality reduction (PCA, tSNE and UMAP) to visualize the cells in two-dimensional space (Figure 1). We then preformed Clustering and Cell type annotation (with SingleR) to identify the type of each cell. In the analysis we've identified Five major immune cell types: CD4+ T cells, CD8+ T cells, B cells, natural killer (NK) cells and monocyte–macrophage cells (Figure 2).

Note: as mentioned, we identified the cell types with the singleR library. However, it seems as if the researchers we took the data from (Xu & Jia, 2021), classified the clusters by these specific BioMarkers: **CD4+ T cells** with canonical marker genes CD3D, CD3E, CD3G, and CD4; **CD8+ T cells** with marker genes CD3D, CD3E, CD3G, CD8A, and CD8B; **B cells** with marker genes CD19, CD79A, and CD79B; **natural killer (NK) cells** with marker genes NKG7, GZMB, GNLY, and NCR1; **monocyte–macrophage cells** with marker genes CD14 and CD68. Later on, this may explain the difference between the distributions each found.

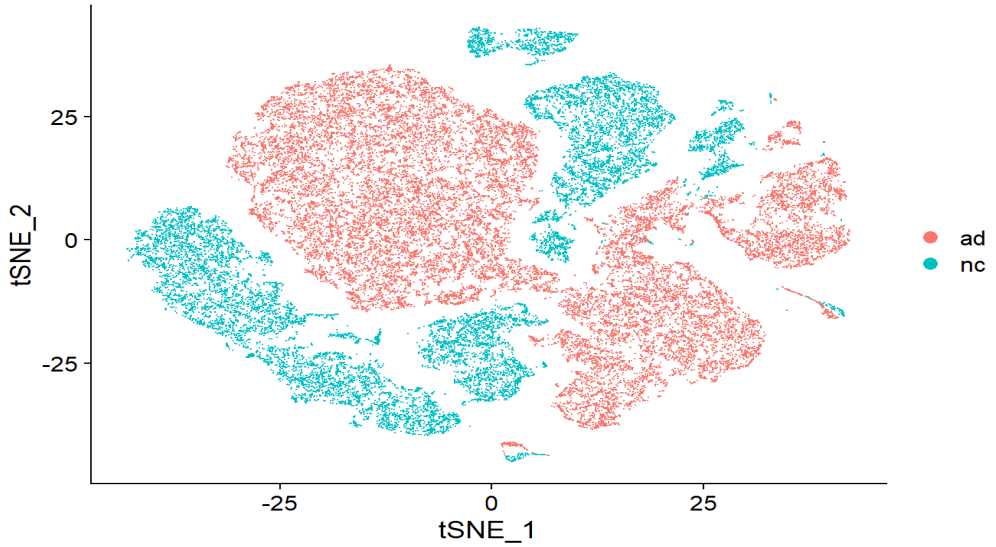


*Figure 1: tSNE plot displaying the clear distinction between the AD and NC groups.*

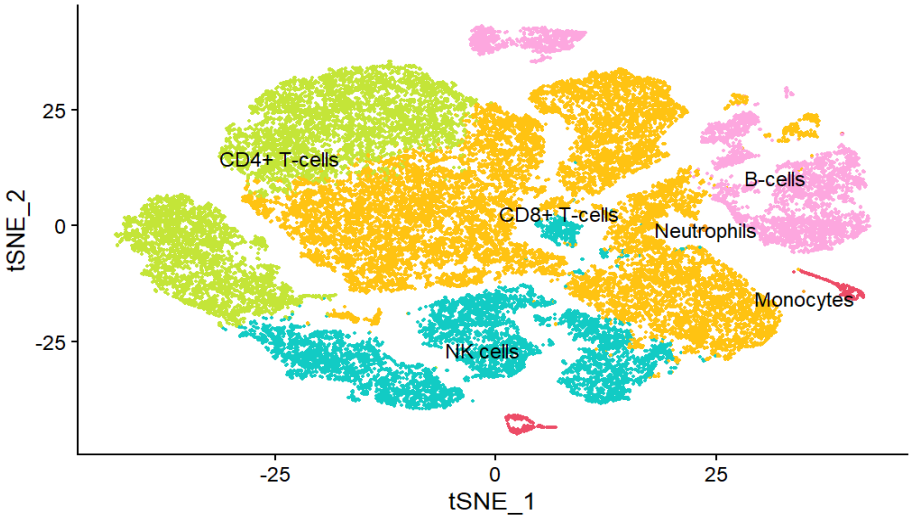


*Figure 2: tSNE plot displaying the samples clustering with the singleR cell-types annotation. This figure contains the data from all patients*

Eventually, we calculated the proportion of each cell-type in each group (Figures 3, 4). Based on these figures, we compared the proportion of the five types of immune cells between the AD group and NC. Numbers of CD8+ T cells and NK cells were significantly increased in the AD group (44.6% and 52.3%, respectively) compared with NC (25.8%, 22.4% respectively). In contrast, the proportions of B cells and CD4+ T cells were significantly lower in the AD group (2.2% and 0% respectively) than in NC (11.8% and 38.7% respectively) The proportion of Monocytes was significantly low in both groups.

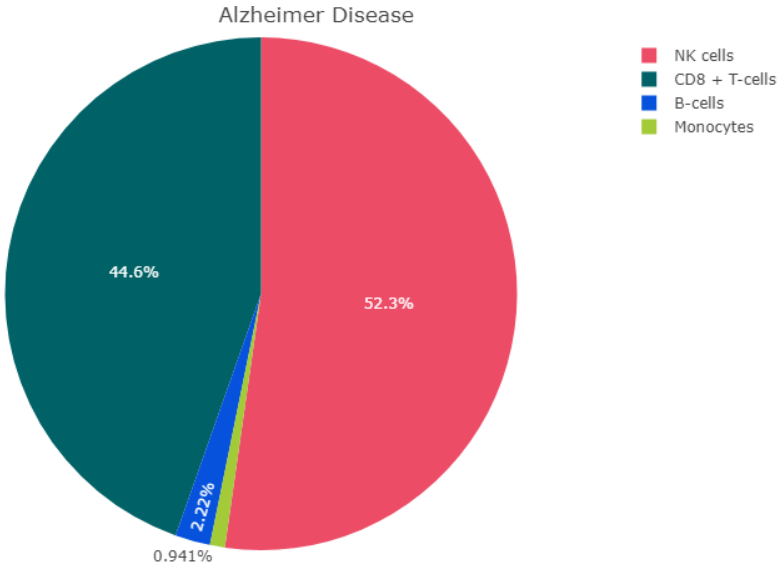


*Figure 3: a pie chart displaying the cell-types distribution in AD patients*

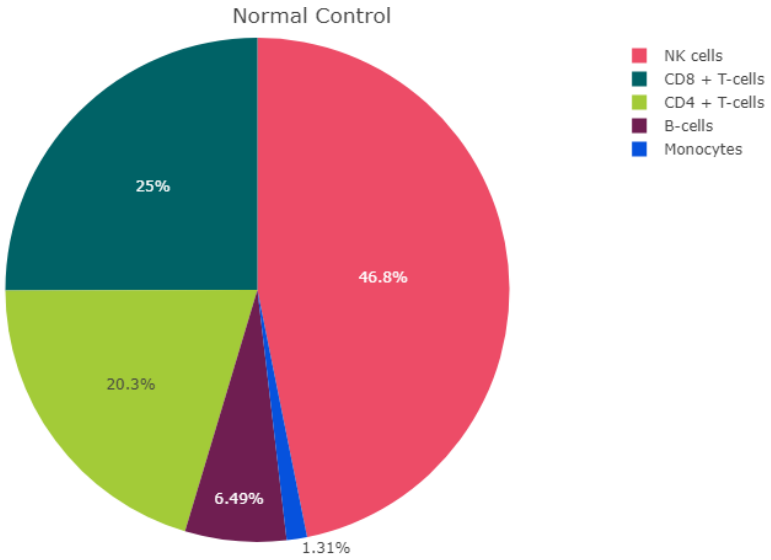


*Figure 4: a pie chart displaying the cell-types distribution in NC patients*

In addition, we aimed to identify genes that are highly expressed in our data. We preformed differential gene expression analysis per group (AD and NC), per cell type and over the entire data (GSE181279). With the differential gene expression analysis results, we then preformed Functional and pathway enrichment analysis for B cell, CD4+ T cell, CD8+ T cell, natural killer (NK) cell, and monocyte-macrophage cell subsets using Metascape.

We found that the most expressed genes in the data were GZMH, PRF1, GNLY, NKG7, FGFBP2, CST7 (p val=0, Figure 5).

| | p_val | avg_log2FC | pct.1 | pct.2 | p_val_adj | cluster |
|---|---|---|---|---|---|---|
| GZMH | 0 | 2.011916 | 0.525 | 0.068 | 0 | NK cells |
| PRF1 | 0 | 1.554531 | 0.995 | 0.744 | 0 | NK cells |
| GNLY | 0 | 1.409929 | 0.992 | 0.733 | 0 | NK cells |
| NKG7 | 0 | 1.344313 | 1.000 | 0.765 | 0 | NK cells |
| FGFBP2 | 0 | 1.252903 | 0.990 | 0.736 | 0 | NK cells |
| CST7 | 0 | 1.169820 | 0.999 | 0.775 | 0 | NK cells |

*Figure 5: table displaying the head of the most expressed genes sorted by p val.*

We also displayed the most expressed genes per cell type (Figure 6).

| | p_val | avg_log2FC | pct.1 | pct.2 | p_val_adj | cluster | gene |
|---|---|---|---|---|---|---|---|
| | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <fct> | <chr> |
| 0 | | 2.01 | 0.525 | 0.068 | 0 | NK cells | GZMH |
| 0 | | 1.55 | 0.995 | 0.744 | 0 | NK cells | PRF1 |
| 0 | | 1.11 | 0.924 | 0.408 | 0 | CD8+ T-cells | IL7R |
| 0 | | 0.830 | 0.857 | 0.653 | 0 | CD8+ T-cells | ZFP36L2 |
| 0 | | 1.22 | 0.724 | 0.394 | 0 | CD4+ T-cells | TCF7 |
| 0 | | 1.21 | 0.575 | 0.289 | 0 | CD4+ T-cells | PRKCQ-AS1 |
| 0 | | 3.54 | 0.994 | 0.201 | 0 | B-cells | HLA-DRA |
| 0 | | 3.22 | 0.994 | 0.106 | 0 | B-cells | CD79A |
| 0 | | 5.38 | 0.899 | 0.008 | 0 | Monocytes | S100A9 |
| 0 | | 5.23 | 0.937 | 0.004 | 0 | Monocytes | LYZ |
| 0 | | 7.95 | 1 | 0.008 | 0 | Neutrophils | G0S2 |
| 6.13e-297 | | 7.53 | 1 | 0.017 | 1.08e-292 | Neutrophils | S100A9 |
| 0 | | 11.5 | 0.833 | 0.002 | 0 | Erythrocytes | HBB |
| 5.37e-152 | | 9.72 | 0.833 | 0.006 | 9.42e-148 | Erythrocytes | HBA2 |

*Figure 6: table displaying the 2 most expressed genes for each cell type, sorted by avg_log2FC.*

We then compared gene expression in the AD group with the NC group. We found that there are differences in gene expression between the groups. For example, IL7R was highly expressed in AD patients but not in NC patients. This difference makes sense as IL7R is linked to NK cells, which we found were more common in AD patients than NC.
We plotted gene expression heatmap for each group to represent those differences (Figures 7, 8).

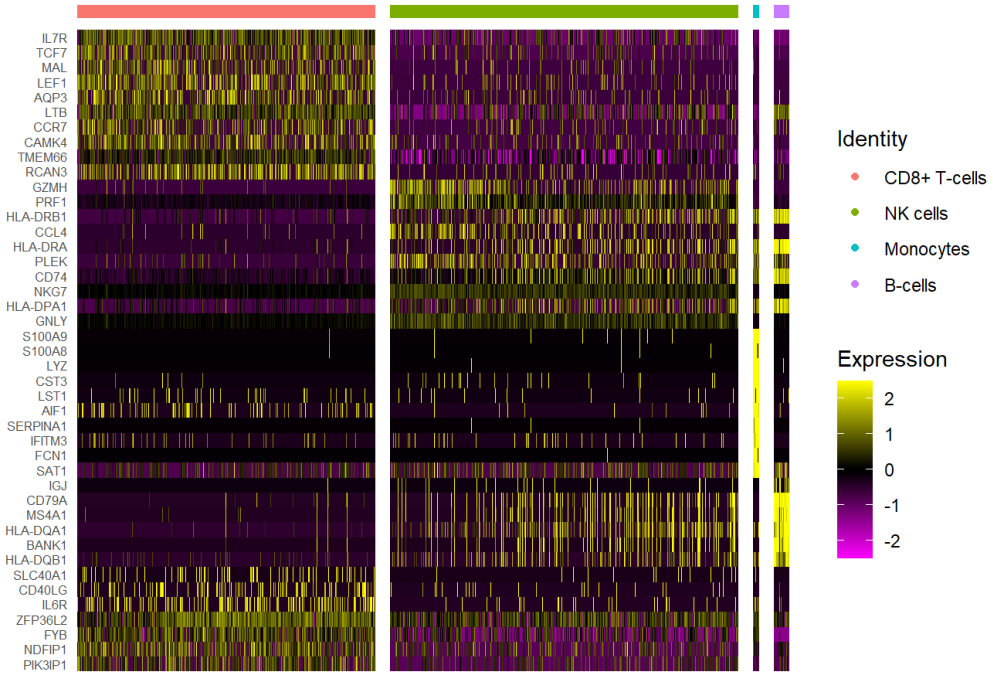


*Figure 7: heatmap displaying the expression level of genes in the AD group. Cell type annotations were taken from previous singleR analysis.*

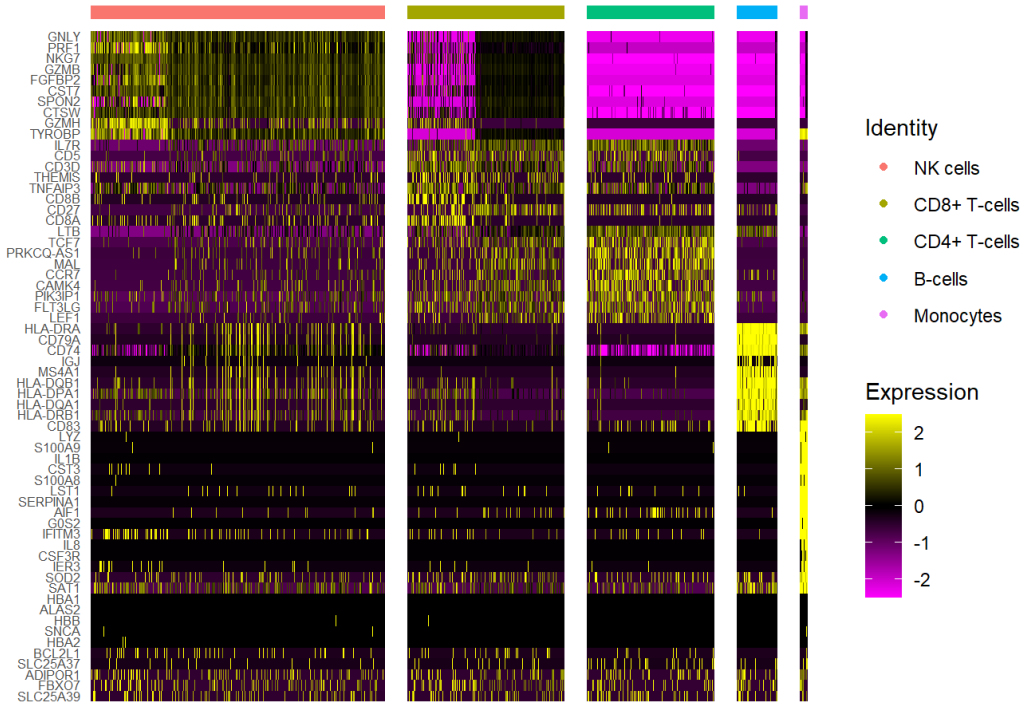


*Figure 8: heatmap displaying the expression level of genes in the NC group. Cell type annotations were taken from previous singleR analysis.*

**Additional results - Functional and pathway enrichment analysis using Metascape**

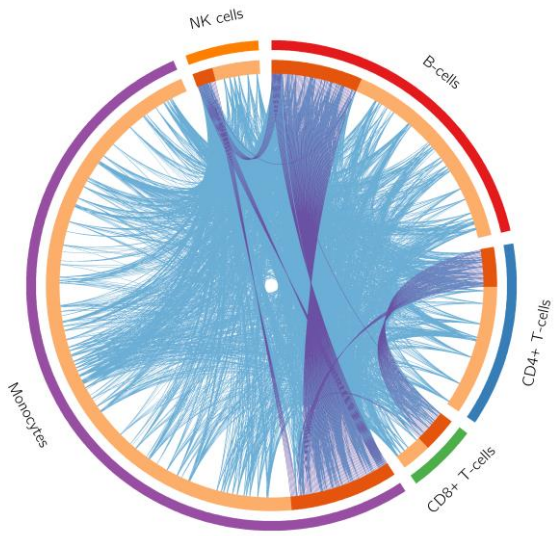


*Figure 9: AD group*

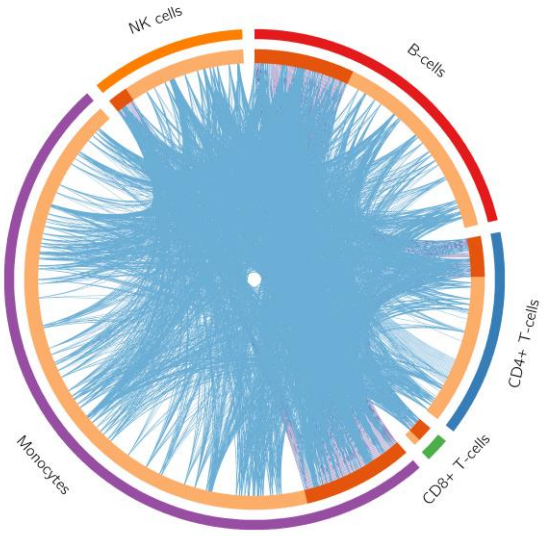


*Figure 10: NC group*

*Figures 9, 10: Circos plot showing overlaps of differentially expressed genes from the given immune cell subsets. Purple lines link the same gene that is shared by multiple gene lists. Blue lines link the different genes where they fall into the same ontology term.*

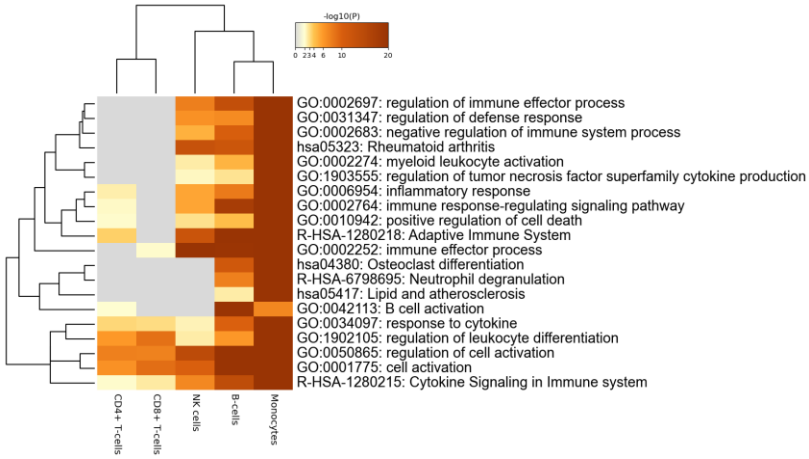


*Figure 11:* Heatmap displaying the enrichment level of each term in different cell types. Colors are based on the P-values of the enriched term (white indicate the lack of enrichment for that term). *AD group.*

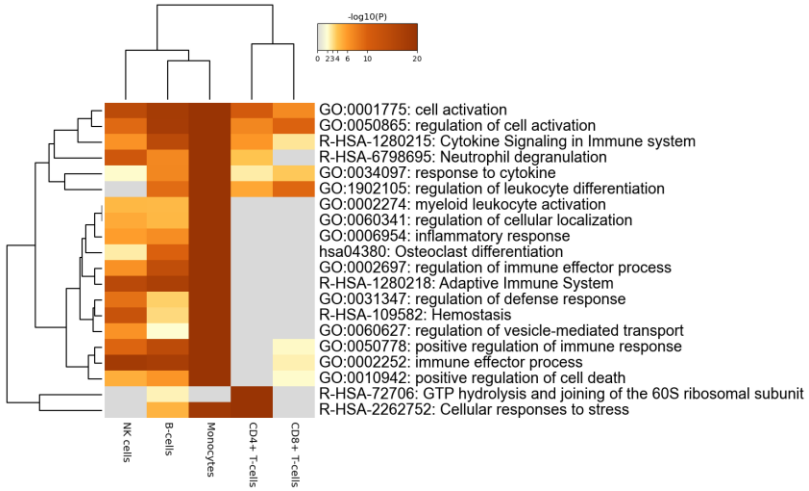


*Figure 12:* Heatmap displaying the enrichment level of each term in different cell types. Colors are based on the P-values of the enriched term (white indicate the lack of enrichment for that term). *NC group.*

**Enriched Ontology Clusters Colored by Cluster ID**

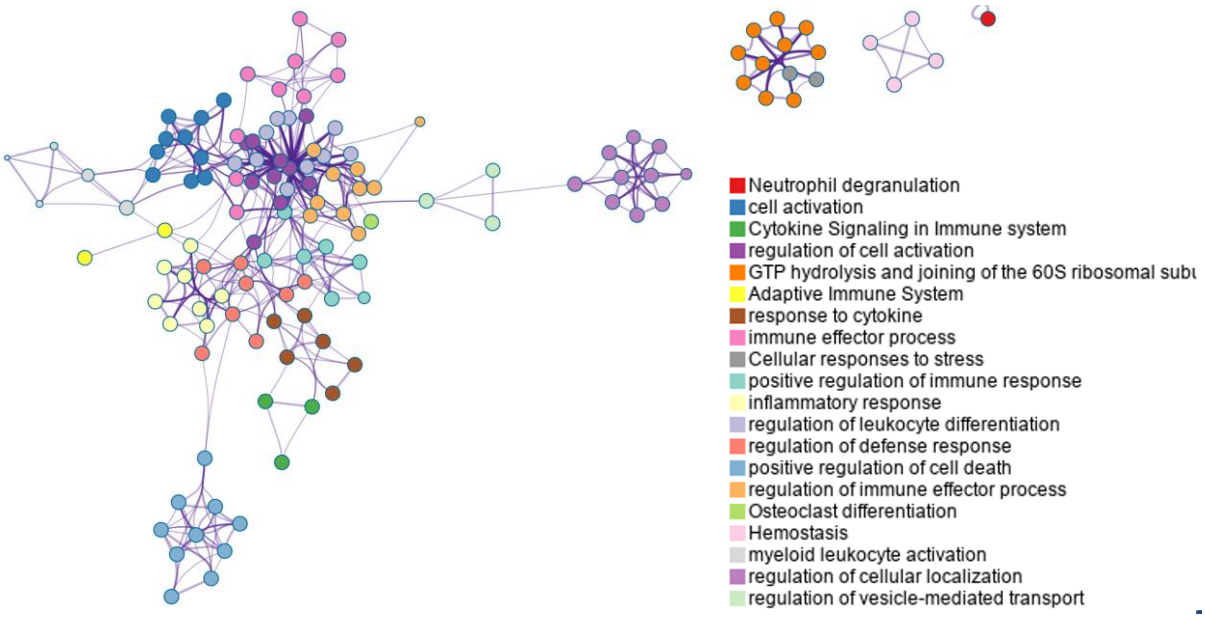


*Figure 13: network layout displaying similarity score of different terms. Each circle node is a term. colors represent the cluster identity. Terms with a high similarity score are linked by an edge.*

## Discussion

### Conclusions

The results of our analysis, based on 36,849 immune cells in patients with AD and NC at a single cell level revealed five major immune cell subsets (CD4+ T cells, CD8+ T cells, B cells, NK cells and monocyte–macrophage cells). We found that the proportions of these cell subsets in AD patients and NC patients were distinct. Moreover, the results also showed differences in gene expression in each cell type in AD compared with NC. Meaning we found changes both in the composition and in the functional state of immune cells in AD.

### Limitations

Our analysis had some major limitations. First, our data covers relatively small sample size (5 patients overall) hence, although we have identified immune changes in peripheral blood of AD, it is not clear that there is a direct causality connection. Second, although we have identified some key genes in AD, we did not cover these genes biological functions in our analysis. Third, most of our analysis used external data and tools whose correctness we couldn't verify (however, we've only used reliable sources and libraries).

### what's next?

First, we would like to recheck the distributions of cell types in the given data. As mentioned in the results, we found some differences between our results and the researchers (Xu & Jia, 2021) results. We assume that the differences derive from the annotation method (We used singleR while the researchers used bio markers).
Moreover, we would like to extend the analysis to a wider data set, such that we would get more accurate results, and a more reliable ground to our conclusions.

To overcome our limitations, we would have to collect more data and use more powerful tools and computer to able to analyze this data. In addition, we would need wider biological and immunological knowledge to better explain the practical meaning of our results.

As our results show distinct differences in proportions of the cell subsets in AD patients and NC patients, we would suggest examining the role of each immune cell in the disease pathogenesis. In addition, we would suggest a future experiment that would examine the immunological role of the proteins encoded by the genes with different expression levels in said groups.