NYCSEA

Journal of NYCSEA (JYEM)
2026, Volume 1, Issue 1

ISBN: 979-8-89238-262-5

Cross-disease Transcriptome Analysis of Astrocytic Pathways Regulating Neuroinflammation in Autism Spectrum Disorder

Valentina Zhang

Yale University

Abstract: Neuroinflammation has been implicated in autism spectrum disorder (ASD), but the underlying molecular mechanisms remain poorly understood. Given that astrocytes are crucial in neuroinflammatory processes and play a significant role in neurological disorders, exploring their involvement in ASD is of great value. Recent advancements in single-cell transcriptomics have allowed the identification of novel astrocytic pathways associated with neuroinflammatory disorders like multiple sclerosis (MS), presenting an opportunity to investigate the role of neuroinflammation in ASD pathology through cross-disease analysis. In this study, large-scale single-cell RNA-seq expression data was analyzed to characterize the shared astrocyte subpopulations in both ASD and MS. Batch correction using Harmony, alongside Seurat and SC3, was employed to identify common astrocyte clusters and their marker genes. Disease-specific differentially expressed gene (DEG) analysis was conducted using DESeq2, while Monocle and Enrichr facilitated trajectory and ingenuity pathway analysis (IPA). The research revealed a pronounced role of oxidative stress-induced ferroptosis common to ASD and MS. Consequently, it is hypothesized that key regulators of ferroptosis, including disordered iron metabolism, the oxidant system (ROS, MAPK), the antioxidant system (SLC7A11, Hippo pathway), and transcription factors (Nrf2, AhR), are potentially involved in ASD pathology and could serve as promising targets for disease intervention. These findings enhance our understanding of ASD pathology, suggesting a potential option for the development of novel therapeutic strategies for ASD.

Keywords: Neuroinflammation in autism spectrum disorder (ASD); Ferroptosis; Astrocytic signaling pathways; Cross-disease Transcriptome Analysis; Ingenuity Pathway Analysis (IPA); Gene set enrichment analysis (GSEA)

Introduction

Since its first clinical account in 1925 (Sher et al, 2023), autism spectrum disorder (ASD) is estimated to affects 1 in 44 children in the US (Maenner et al., 2021) and with a worldwide prevalence of around 1% (Zeidan et al., 2022) as of 2021. The disorder encapsulates a broad range of conditions characterized by challenges with social skills, repetitive behaviors, speech, and nonverbal communication (American Psychiatric Association, 2022). While the physiopathology of autism remains unclear, it likely involves several systems connectivity, neural, biochemical, neuroanatomical, cellular, and

molecular features. Accumulated evidence converges in highlighting changes in the white and grey matter, blood-brain barrier disruption, abnormal synaptic overgrowth, and synapses density. All these mechanisms indicate a possible involvement of neuroinflammation in the development of ASD (Matta et al., 2019; Hughes et al., 2018; Toscano et al., 2021; Eissa et al., 2020).

As the most common demyelinating disease of the central nervous system (CNS), multiple sclerosis (MS) is characterized by damage to the myelin sheath of neural axons in CNS ultimately resulting in axonal degradation and neuronal loss. Chronically persistent neuroinflammation has been recognized as a chief pathological component of MS, and the disease is now understood as an autoimmune inflammatory disease of the CNS driven by the inflammatory activity of peripheral immune cells recruited to the CNS and by CNS-resident glial cells (Kadowaki & Quintana, 2020). MS affects approximately 2.8 million people worldwide. It is the most common disabling neurological disorder of young adults (20–40 years) (Reich, 2018). Both MS and ASD put debilitating burdens on the affected individuals, their families, and the health care systems.

Astrocytes, the most abundant glial cell type in the CNS, have recently gained attention for their roles as primary immune effector cells (Lee et al., 2022) and in neuroinflammatory processes (Wheeler et al., 2020). Astrocytes respond to inflammatory signals and can themselves promote inflammation, making them essential players in neurological diseases (Linnerbauer et al., 2020; Gzielo & Nikiforuk, 2021). However, the inherent heterogeneity of astrocytes presents challenges in understanding their functional characteristics fully. Recent breakthroughs have shed light on the heterogeneity and regulatory mechanisms of astrocytes in the pathogenesis of MS (Lee et al., 2022; Miller, 2018). However, the specific dysregulation of cerebral astrocytes in ASD has not been thoroughly studied.

To address this gap, our research takes a comparative approach by exploring the gene expression profiles of cerebral astrocytes in both ASD and MS. We integrated multiple single-cell RNA sequencing datasets of human astrocytes (Speir et al., 2021) and analyzed their functional and phenotypic characteristics. Our unbiased analysis revealed two common astrocyte subpopulations shared between ASD and MS, indicating potential converging mechanisms of neuroinflammation. Notably, these subpopulations exhibited dysregulations of key regulators of ferroptosis, suggesting its pronounced role in the pathogenesis of both ASD and MS.

ASD's treatment is challenging due to its unknown cause and varied symptoms (Vashisth et al., 2023). No existing medication can fully cure ASD. Some FDA-approved antipsychotics can reduce repetitive behavior but don't address other ASD behaviors or comorbid conditions like autoimmune disorders (Comi et al., 1999). Our new finding of a neuroinflammatory mechanism in astrocytes offers a potential therapeutic approach for ASD through anti-neuroinflammatory candidates.

Materials and Methods

This research was motivated by the yet incomplete understanding of the pathology underlying ASD despite the growing evidence in recent years highlighting the important roles of dysregulated neuroinflammation in ASD (Eissa et al., 2020; Rahnama et al., 2021; Zürcher, et al., 2021) and its connection to autoimmune disorders (Trifonova et al., 2021; Whiteley et al., 2021). Our interest in astrocytes is due to their functional and phenotypic heterogeneity in the context of neurological diseases, particularly those involved in CNS inflammation. To identify candidate regulators of astrocyte pathogenic activities in ASD, we carried out a cross-disease analysis between ASD and MS, a neuroimmune inflammatory disease, to uncover the shared inflammatory mechanisms.

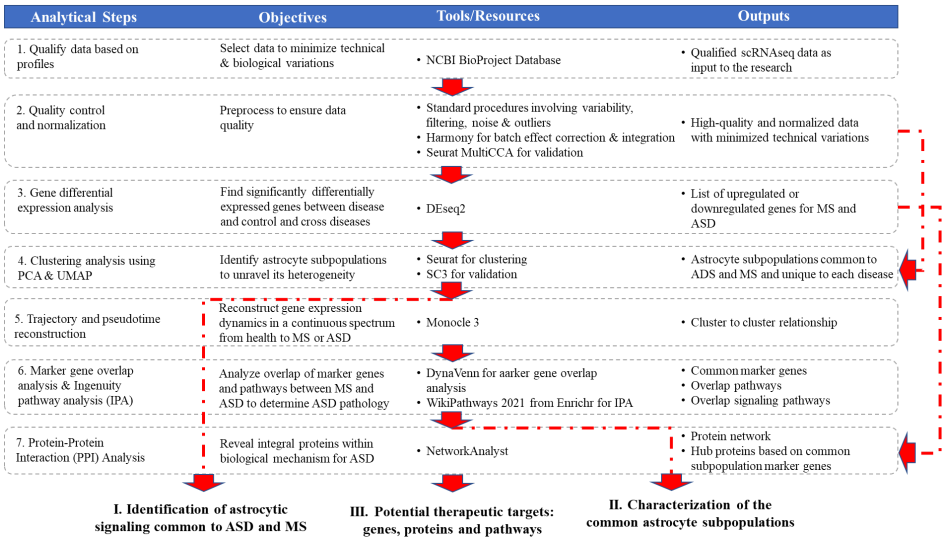

Figure 1. Schematic representation of the cross-disease transcriptome analysis workflow. The entire workflow consists of seven pipelined stages 1-7, arriving at the results I-III. Red dash lines indicate data dependencies across the otherwise sequential stages.

1. Data Mining

The transcriptome analysis workflow used to study the molecular gene signatures, pathways, and regulatory networks (Figure. 1) is constructed according to a best practice guideline (Luecken & Theis, 2019). All data processing was conducted using programming language R (version 4.2.1).

This research used the gene expression profiles of scRNA-seq datasets PRJNA544731 and PRJNA434002 gathered from the NCBI (National Center for Biotechnology Information) BioProject database, a public repository that offers free access to microarray, next-generation sequencing, and other types of high-throughput functional genome data. In summary, a total of 53 samples were considered in this study as shown in Table 1. These datasets were selected to minimize technical and biological variability. Both datasets were assayed with the same scRNA-seq technique and further restricted to three specific brain regions: the prefrontal cortex (PFC), anterior cingulate cortex (ACC) and insular cortex (IC) regions.

Table 1. Overview of datasets used in this study. The PRJNA544731 dataset collected 12 samples from 10 MS patients (7 female and 3 male, 1 primary progressive MS, 9 secondary progressive MS) with ages ranging from 34 to 55 years, and 9 samples from 9 control patients (4 female and 5 male). All samples were collected from snap-frozen brain tissue obtained from autopsies. The PRJNA434002 dataset collected 16 samples of ASD patients and 16 samples of control patients. Chronic inactive and acute/chronic active plaques were labeled across samples as well. Both datasets identified neuronal and glial types and labeled astrocyte cells.

SL No.	Accession	ID	Sample Size	Control	Case
1	PRJNA544731	544731	21	9	12
2	PRJNA434002	434002	32	16	16
	Combined Total		53	25	28

To further characterize the biological process of the pro-inflammatory astrocyte subpopulations identified by our analysis, we conducted Gene Set Enrichment Analysis (GSEA) using the bulk-RNA sequencing data for the secondary progressive stage

transcriptional signatures of MS. This dataset consists of induced pluripotent stem cell (IPSC) astrocytes from the skin cells of MS patients.

2. Pre-processing

There were 6,215 control cells sequenced, 3,810 cells from MS patients with 33,741 genes sequenced, and 7,573 cells from ASD patients with 36,501 genes sequenced for this research. Between the two datasets, there were 33,592 common genes, and the genes unique to each dataset were removed. After resolving the inconsistent format of the two datasets, preprocessing was conducted according to the standard quality control and normalization procedures (Koch et al., 2018). We used a structured approach that includes determining intra and inter-group sample variability, setting up low count filtering, and establishing a noise threshold.

In addition, since our single-cell transcriptomic data were compiled from multiple experiments with differences in capturing times, handling personnel, reagent lots, equipment, and technology platforms, the batch effects can be highly nonlinear, making it difficult to correctly align different datasets while preserving key biological variations (Tran et al., 2020).

To address this potential issue, our analysis pipeline selected Harmony (Korsunsky et al., 2019) as the primary tool for batch correction and integration, which was subsequently validated with the Seurat MultiCCA method. This setup was based on the recommendations from a well-known evaluation of batch correction methods (Tran et al., 2020). The evaluation recommends only Harmony, LIGER and Seurat 3 for batch integration, and names Harmony as the first method to try, with the other methods as viable alternatives. In this research, we used entropy regularization for soft k-means (1) in Harmony and observed neglectable differences from the results from Seurat MultiCCA, which could reflect that the datasets used for this research are limited to a single cell type data assayed with the same RNA-seq technology.

$$min_{R,Y} \sum_{i,k} R_{ik} \|Z_i - Y_k\|^2 \qquad (1)$$

Where Z is certain feature space of the data, shared by centroids Y. Rk,i, can takes values 0 or 1, denoting membership of cell i in cluster k.

3. Screening of Differentially Expressed Genes (DEGs)

Differential gene expression is important to understand the biological differences between health and disease states. DEG analyses were performed with DESeq2 (version 3.16) separately for each disease against its controls. To account for the sequencing depth and RNA composition differences between and within the samples, we first used the median of ratios method for count normalization. Afterward, we assessed if there were statistically significant DEGs between disease cells and control cells using an empirical Bayesian approach for testing gene expression fold change, assuming dispersion function follows a log normal prior distribution (2). Significant DEGs were filtered based on q-value < 0.05 and |log2FC| > 0.5. The DEGs resulted were subsequently used downstream to characterize the astrocyte subpopulations common to ASD and MS.

$$\emptyset_g \sim N(log\emptyset_{trend}(\mu_g), \sigma_d^2) \qquad (2)$$

Where ∅_trend (μ_g) is a function of the mean of normalized count for the gth gene, and σ_d^2 is the true dispersion around the trend that is common for all genes.

4. Clustering Analysis

For cells with high heterogeneity and functional diversity such as astrocytes, clustering scRNA-seq data provides exciting new opportunities to detect transcriptionally coherent cell subpopulations. Using Seurat, we first performed principal component analysis (PCA), reducing the high dimensionality to 30 principal components. UMAP was then applied for further dimension reduction. For this research, we used Seurat's K nearest neighbors (KNN) based on Euclidean distance in PCA space and applied the Louvain algorithm using (3) as the modularity quality function.

$$Q = \frac{1}{2m}\sum_{i,j}[A_{ij} - \frac{k_i k_j}{2m}]\delta(C_i, C_j) \qquad (3)$$

Where Ki refers to the degree of node i; Ci is the community where node i is; m is the weighted sum of all edges in the community network; Aij is the weight of the edges connected between node i and node j; δ(C_i,C_j) represents whether node i and node j are in the same community. The closer Q is to 1, the more obvious community structure is.

To validate results, we repeated the analysis with SC3 (version 1.24.0), another R-based clustering tool using consensus for unsupervised clustering. The algorithm conducts PCA, followed by K-means-based consensus clustering. In this study, because more than 5,000 cells were used, which required higher computational power, we implemented a hybrid approach via SC3: SC3 first clustered a sample of 5,000 cells with unsupervised clustering whose result was subsequently used to train a supervised support vector machine (SVM) to optimize computational capability. SC3 requires an input parameter k for the suggested number of clusters, for which we used 14 based on visualization. In the end, we mapped the clusters identified by SC3 onto those identified by Seurat and validated the cell IDs between the clusters of the two algorithms following a practice used in a publication (Duò et al., 2018). This analysis also revealed differences in the composition of the clusters, validating Seurat's clustering result.

5. Trajectory Analysis and Trajectory-guided Common Marker Gene Analysis

While single-cell expression profiling exploits cell-to-cell variation in gene expression to reveal regulatory networks governing cell differentiation and other biological processes, trajectory analysis captures dynamic changes in transcriptional configuration along a trajectory, allowing for a better understanding of cell states and their proximity to each other (Trapnell et al., 2014).

This research used Monocle 3, an R package for trajectory analysis among the cell clusters identified in the previous step. After fitting cells onto a low-dimensional space with UMAP, we developed the principal graph embedding algorithm via Monocle 3 to reconstruct possible trajectories across the clusters. Once trajectories were established, we studied the cluster-to-cluster relationship along selected trajectories. Marker gene modules rather than individual genes were extracted to improve the signal-to-noise ratio and facilitate the downstream signaling pathways analysis and biological interpretation. Additionally, our analysis relies on the pseudo-time generated by Monocle as directional reference of the cell state changes.

6. Marker Genes and Pathway Analysis

Overlapping gene lists can reveal biological meanings and may lead to novel insight. Using the marker genes identified for the clusters from Seurat and the disease-specific marker genes produced in Step 3, we performed overlap analysis to infer biological similarities with Benjamini-Hochberg adjusted p value in (4).

$$p\_value_{adjusted} = \sum_{l=1}^{|A[1:i]\cap B[1:j]|-1} \frac{(i\ l)(|A\cup B|-i\ j-l)}{(|A\cup B|\ j)} \quad (4)$$

Where A and B are two ordered lists of genes, while i and j are indexes into the lists.

Ingenuity pathway analysis (IPA) determines the enrichment of molecular pathways through a gene set. To the gene sets found through an earlier overlap analysis, we applied WikiPathways 2021 from Enrichr to capture the biological processes potentially implicated in the ASD pathology. We compared the pathways identified across the two common subpopulations to remove pathways that are statistically insignificant for this analysis.

Furthermore, we also applied IPA and pathway overlap analysis on the marker genes of the trajectories identified in Step 5. The goal was to characterize the trajectories connecting the common astrocyte subpopulations for both diseases to the ASD-specific and the MS-specific subpopulations and the trajectories connecting controls to those disease astrocyte subpopulations.

7. Protein-Protein Interaction (PPI) Analysis

PPI controls a variety of biological functions, including cell-to-cell connections, metabolic control, and developmental control. PPI analysis reveals integral proteins within biological mechanisms. Using NetworkAnalyst, a web tool for meta-analysis and visualization of protein interactions, we generated a protein network and screened for hub proteins based on the gene list from the common subpopulation marker genes. We implemented PPI, using a confidence score cutoff of 900 from STRING interactome, on the marker genes of the ASD cells within the common subpopulation. Hub proteins were screened based on the node degree of the proteins.

8. Characterization with GSEA

To further characterize the common astrocyte subpopulations, we obtained a bulk-RNA seq dataset from a set of Induced Pluripotent Stem Cells (IPSCs), in which skin cells from MS patients were reprogrammed into an embryonic-like pluripotent state and developed into astrocytes. Using the dataset, we created rank ordered gene lists for each stage of MS where the primary progressive list has 14,964 ranked genes, and the secondary progressive list has 17,497 ranked genes.

Using the bulk-RNAseq dataset as a pathological reference, we employed GSEA, a software created by the Broad Institute for evaluating the statistical significance of a set of genes between phenotypes, as well as the corresponding MSigDB which contains 32,284 reference gene sets. With the tool GSEAPreranked within the software for gene enrichment analysis using both Wikipathway and Gene Ontology of Biological Processes with 1,000 permutations, we were able to validate the marker genes and pathways of the common astrocyte subpopulations studied in the earlier clustering and trajectory analysis.

Results

1. Common Astrocyte Subpopulations

DEG analysis was conducted for the ASD and MS datasets (Figure 2.a). Subsequently, the scRNA-seq analysis based on the integrated datasets using Seurat identified 14 distinct astrocyte subpopulations as shown in Figure 2.c. Each subpopulation has a range of 268 to 2,992 cells. Our goal was to identify potential common astrocyte subpopulations between ASD and MS samples, in which there was significant (>1.25) expansion in both MS and

ASD samples, and a significant number of cells in both MS and ASD datasets (>500) to ensure statistical significance. Based on this methodology, we identified two common subpopulations: Cluster 1 and Cluster 5. Cluster 1 had a 4.47-fold and 1.89-fold expansion in MS and ASD, respectively, and Cluster 5 had a 5.68-fold and 1.41-fold expansion in MS and ASD, respectively. Figure 2.b displays the expansion analysis results of each subpopulation.

Using Monocle's hierarchical clustering algorithm, we further validated that the Cluster 1 and Cluster 5 identified through Seurat's graph-based clustering were indeed common states shared by ASD and MS in the trajectories as shown in Figure 3.a. While Cluster 1 was split into two clusters based on Monocle's clustering algorithm, we determined through cell ID tracking that the area of high density of MS and ASD cells in Cluster 1 was concentrated in a single cluster based on Monocle's algorithm, further validating Cluster 1 as a valid common subpopulation.

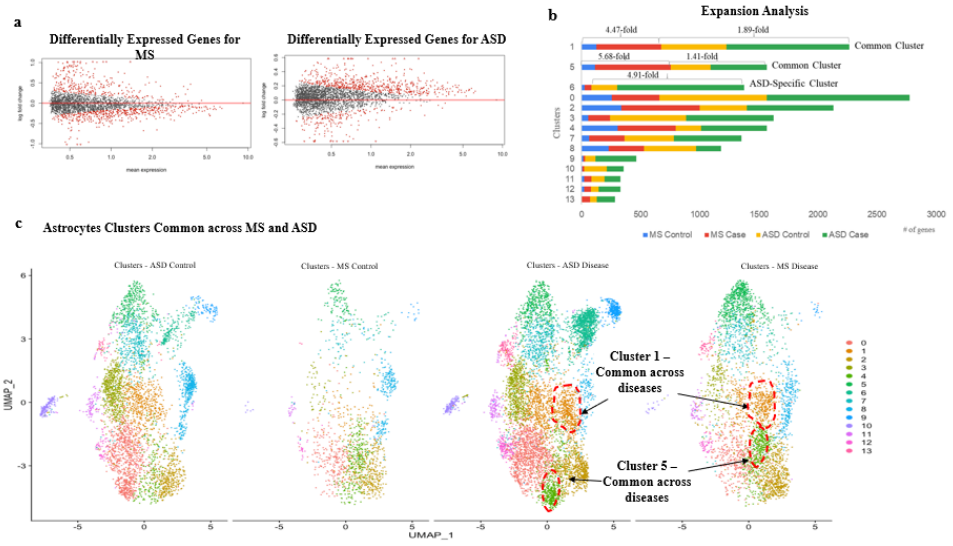

Figure 2. Common Astrocyte Subpopulation across ASD and MS. a. Disease-specific DEG analysis. Gene expression variability amongst the samples is shown with MA plots, an application of a Bland-Altman plot, for MS DEGs and for ASD DEGs, showing logarithmically transformed counts (log fold change) and mean expression. Significant DEGs are colored red. b. Expansion analysis shows Cluster 1 had a 4.47-fold and 1.89-fold expansion in MS and ASD, respectively, and Cluster 5 had a 5.68-fold and 1.41-fold expansion in MS and ASD, respectively. c. Common astrocytes subpopulations identified through K-nearest neighbor clustering. There are 14 identified clusters (0-13) as shown in different colors. Highlighted are Cluster 1 and Cluster 5, which are common clusters to ASD and MS, and Cluster 6, which is an ASD-specific cluster.

2. Common subpopulation characterization

Marker Genes. We identified marker genes for the two common clusters using Monocle and Seurat, which was the first step in characterizing the common astrocyte subpopulations. Additionally, we identified marker genes from disease-specific individual cells within the common clusters using Monocle to conduct overlap analysis. The overlap analysis identifies the marker genes that are involved in both disease-specific and common clusters. Table 2 below has top 15 marker genes for the ASD cells within the common clusters. Refer to Supplementary Table 2 for a complete list of marker genes that overlap ASD disease and common clusters for both ASD and MS. The pathways identified from the marker genes by Enrichr are consistent with the trajectory specific pathways that are discussed in the Discussion section.

Gene	Average Log2FC	Q Value	Marker Score	Mean Expression
ENSG00000179915-NRXN1	-0.636942561	1.43968312844771E-230	0.246324876	5.512946061
ENSG00000110436-SLC1A2	-0.854530444	0	0.241786985	5.927996919
ENSG00000219507-FTH1P8	0.272026487	7.51338507496952E-53	0.277254526	1.408110765
ENSG00000175161-CADM2	-0.358036548	2.40006317518309E-97	0.233371795	4.463217237
ENSG00000079215-SLC1A3	-0.718468527	7.55265457322383E-285	0.231028757	4.732560276
ENSG00000069667-RORA	-0.343553977	4.81631207731956E-102	0.230253108	4.480928624
ENSG00000230876-LINC00486	-0.383482321	5.65600851445833E-25	0.226622232	4.605999424
ENSG00000239268-RP11-384F7.2	-0.260990359	1.86364954078684E-33	0.213378189	3.699364561
ENSG00000182985-CADM1	-0.858895119	2.16291000280672E-247	0.199318525	3.612418269
ENSG00000245532-NEAT1	0.317171381	2.96084676028346E-74	0.190174013	3.041512311
ENSG00000109472-CPE	-0.361961898	8.02489839917901E-36	0.182230304	2.930487819
ENSG00000175497-DPP10	1.438427847	0	0.179476466	3.403643947
ENSG00000125462-C1orf61	-0.368078131	9.77160068510367E-28	0.178015934	2.705690387
ENSG00000080573-COL5A3	-1.381064848	1.22852119981953E-231	0.159532395	2.615653732
ENSG00000130203-APOE	-0.343638039	1.48812741437734E-13	0.154877692	2.713141247

Table 2. Marker Genes in ASD Cells in The Common Clusters

Transition Trajectory. In response to stimuli or throughout life, cells transition from one functional "state" to another through transcriptional reconfiguration. To further characterize the dynamic biological process of the common astrocyte subpopulations, we conducted trajectory analysis to reconstruct the dynamic gene expression in a continuous spectrum from control clusters to common, ASD-specific and MS-specific clusters. For the trajectory analysis, principal parts are labelled as one of the followings: 1. control, 2. ASD-specific, 3. MS-specific, 4. common between ASD and MS, and 5. unclear. The principal part selection was informed by the clusters defined by Seurat and the clusters' composition. Excluding the unclear principal parts, we created two trajectory paths for every pairwise combination of the three disease states (common, ASD-specific, and MS-specific) and control. As shown in Figure 3.c, the pseudotime-annotated trajectory created using Monocle enables us to establish Cluster 1 and Cluster 5 as intermediate state between control and the disease-specific states.

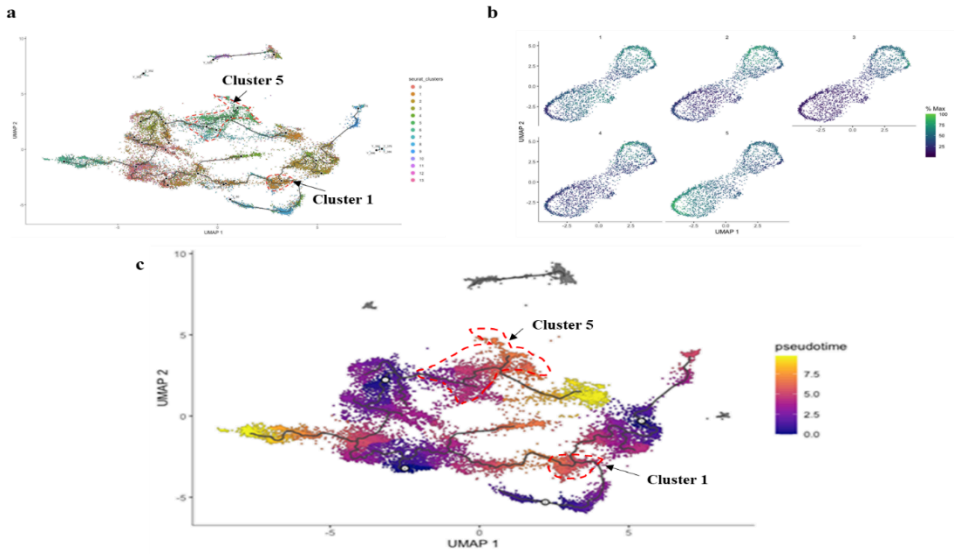

Figure 3. Trajectory Analysis. a. Monocle plot of gene expression using UMAP color coded with the clusters identified by Seurat. Clusters circled are Cluster 1 and Cluster 5. Monocle maps cells along a trajectory based on differential gene expression among the cells. Genes that are key players in transitions between states can be found through this analysis. b. 5 co-expression gene modules identified through trajectory analysis along the path from control to common. Each gene module is comprised of genes with similar expression patterns. Each module is colored from purple to green based on expression level. c. Astrocyte state trajectories annotated with pseudotime. Starting points were chosen based on clusters of the control cells along the trajectory, and pseudotime was plotted from there, as shown by the black outline. Pseudotime progresses by color from purple to yellow.

Marker Pathways. Based on the trajectories identified, we analyzed genes that are differentially expressed as a function of that trajectory path for downstream pathway analysis. Figure 3.b shows the changes in gene expression of 5 gene modules as the cell states progress along a trajectory path from the control cells to the cells common to ASD and MS. Subsequently, we identified the signaling pathways potentially involved in the pathological mechanism transitioning from the control state to the common state and between the common state and ASD or MS state. Figure 4. highlights the trajectory-specific pathways and demonstrates the dynamic relationships across all the states: control, ASD-specific, and common states. There are 56 trajectory specific pathways between the control and common stages, 32 between the common and ASD-specific states and 11 between the control and ASD-specific states. Examples of these pathways are oxidative stress, mitogen-activated protein kinase (MAPK) signaling pathway, Aryl Hydrocarbon Receptor (AhR) pathway, nuclear factor-erythroid 2 related factor 2 (Nrf2) pathway, and brain-derived neurotrophic factor (BDNF) signaling pathway. All these pathways have been implicated in neuroinflammation based on existing research (Ou et al., 2022; Barbosa et al., 2020). Some of these pathways will be further discussed in the Discussion section.o introduce targeted gene knockdowns (KDs) into organoids and cell culture tissues, lentiviral vectors can be used to deliver the CRISPR guide RNAs that aid Cas9 nucleases in DNA cleavage (Milone, 2018). Lentiviral transduction was used to introduce gene KDs, such as those targeting tumor suppressor genes (Moses, 2020). Thus, cell proliferation indicative of tumor growth can be induced in the SEAM organoid through the inhibition of crucial cell regulator genes. Ultimately, the UM replication potential of a model created through the KD of commonly mutated UM tumor suppressor genes was evaluated and doxycycline's effects on the UM metastasis phenotype organoids were compared to those of the non-mutated organoid through immunofluorescence and RT-qPCR.

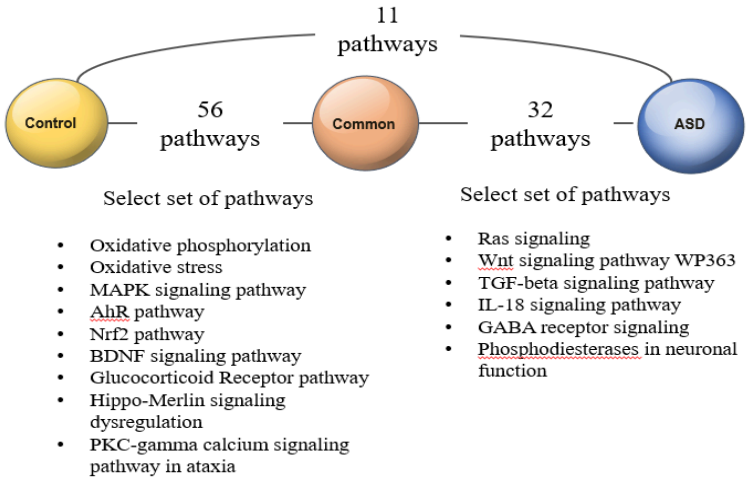

Figure 4. Trajectory-specific Pathway Analysis. Signaling pathways along the trajectories between control, common and ASD-specific states. Examples of pathways between control and common, and common and ASD-specific states are listed.

Hub Proteins. Using NetworkAnalyst and the STRING database, we built a protein network using marker genes from Cluster 1 and Cluster 5 (Figure 5). Cluster 1 marker genes returned a network with 548 nodes, 917 edges, and 18 seeds. Cluster 5 marker genes returned a network with 3,372 nodes, 7,504 edges, and 456 seeds. We used the node degree to identify hub proteins. Degree shows the number of connected nodes with the individual node. Therefore, higher degree indicates a characteristic of hub proteins. The overlap analysis between ASD and MS PPI analysis highlights heat shock proteins, ribosomal proteins, and phospholipases. Table 3. shows the 15 hub proteins. Supplementary Table 4 includes the ontology for all the proteins. Many of these proteins have been identified to play various roles in the pathology of ASD and MS in the existing literature. In this cross-disease research and common astrocyte subpopulation analysis, they are also found to be connected to neuroinflammatory pathways. Therefore, those proteins could be further researched as therapeutic target candidates.

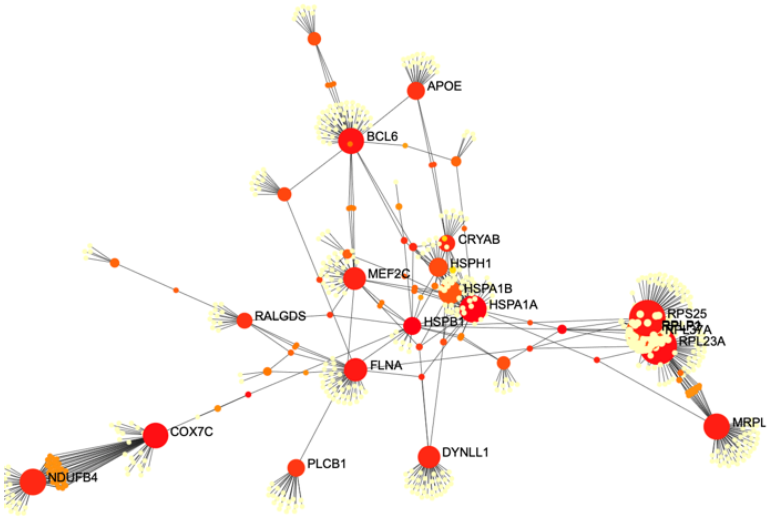

Figure 5. Common Astrocyte Subpopulations PPI Analysis. NetworkAnalyst was used for PPI network analysis and visualization. Proteins with a red symbol in the diagram are hub proteins, which are categorized by the number of interactions each protein has with other proteins.

Integrated Molecular Machinery of Ferroptosis. We used WikiPathways, a database of biological pathways, to reveal the proteins, transcription factors, other biological molecules, and pathway ontology terms involved in the pathways within the astrocytes of the common clusters. The pathway analysis revealed ferroptosis as a primary biological mechanism involved in all interactions between states. Figure 6. provides a summary of the potential mechanisms of the neuroinflammation implicating ferroptosis, indicated by the two common astrocyte subpopulations. Figure 7 also indicates that ferroptosis-related genes are dysregulated.

Ferroptosis is a ROS-dependent form of cell death associated with two main biochemical characteristics, namely iron accumulation and lipid peroxidation. (Tang et al., 2021). Ferroptosis is caused by a redox imbalance between the production of oxidants and antioxidants, which is driven by the abnormal expression and activity of multiple redox-active enzymes that produce or detoxify free radicals and lipid oxidation products. Oxidative stress pathways involving MAPK are the main catalyst for ferroptosis. Pro-inflammatory cytokine, such as IL-6, and anti-inflammatory cytokine, such as TGF-β, regulate the production of ROS (Tang et al., 2021). Thyroid stimulating hormones (TSH) have also been shown to induce ROS (Morshed & Davies, 2020). Downstream, ferroptosis activates NF-κB through 4HNE, a pro-inflammatory mediator, which further leads to the neurotoxicity involved in neuropathology.

As for anti-inflammatory processes, activation of the Hippo pathway promotes cadherin 1-dependent cell adhesion, resulting in ferroptosis resistance (Tang et al., 2021). PKC-gamma signaling limits iron intake and ferroptosis through mediation of heat shock protein family B (HSPB), which was found upregulated in the MS and ASD data. Type I Interferons (IFN-I) signaling in astrocytes reduces inflammation via the ligand-activated transcription factor AhR (Rothhammer et al., 2016). Nrf2 is a well-known transcription factor that plays a key role in antioxidation and is released from Kelch-like ECH-associated protein 1 (KEAP1) binding sites during states of oxidative stress to reestablish redox homeostasis (Tang et al., 2021). mTOR is an important negative regulator of autophagy (Pagani et al., 2021).

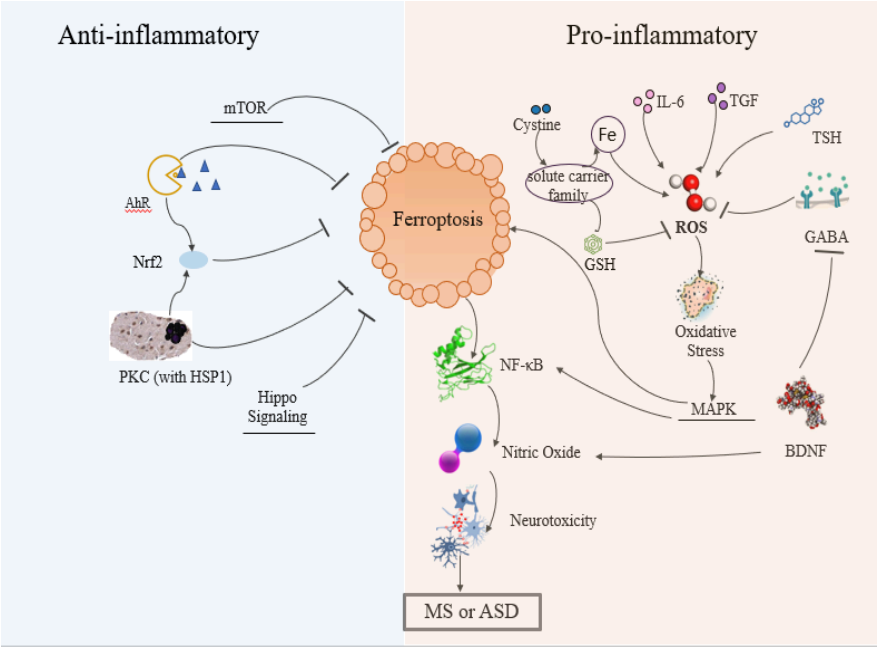

Figure 6. Ferroptosis: a cell death connecting oxidative stress, inflammation and MS or ASD

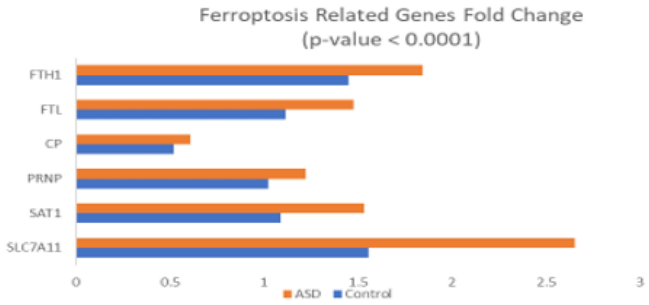

Figure 7. Ferroptosis-Related Genes Fold Change (ASD vs Control): The genes indicative of this pathway that were expressed along all trajectory paths are genes involved in iron metabolism: PRNP, CP, FTH1 and FTL, and genes involved in lipid metabolism and (anti)oxidant metabolism: SAT1 and SLC7A11

Discussion

1. Ferroptosis Mechanism

Ferroptosis is a type of cell death that was discovered in recent years. Increasing evidence shows that ferroptosis can be accompanied by inflammation (Li et al., 2020). In this section, we will discuss the key regulators of ferroptosis, including disordered iron homeostasis, the oxidant system, the antioxidant system, and transcription factors.

Disordered Iron Homeostasis. Disordered iron homeostasis plays a key role in inducing ferroptosis (Figure 8.a). After Tf-TfR1-mediated import and STEAP3-mediated conversion of Fe3+ to Fe2+, free Fe2+ can catalyze the production of ROS through the Fenton reaction, followed by lipid peroxidation. Overproduction of ROS results in oxidative stress and its consequent adverse effect on cells (Zhang et al., 2022). Ferritin

Heavy Chain 1 (FTH1), Ferritin Light Chain (FTL), HSPB Member 1 (HSPB1), six-transmembrane epithelial antigen of prostate 3 (STEAP3) and Ceruloplasmin (CP) are associated with iron metabolism and are overexpressed in our ASD disease cells (Figure 8.b).

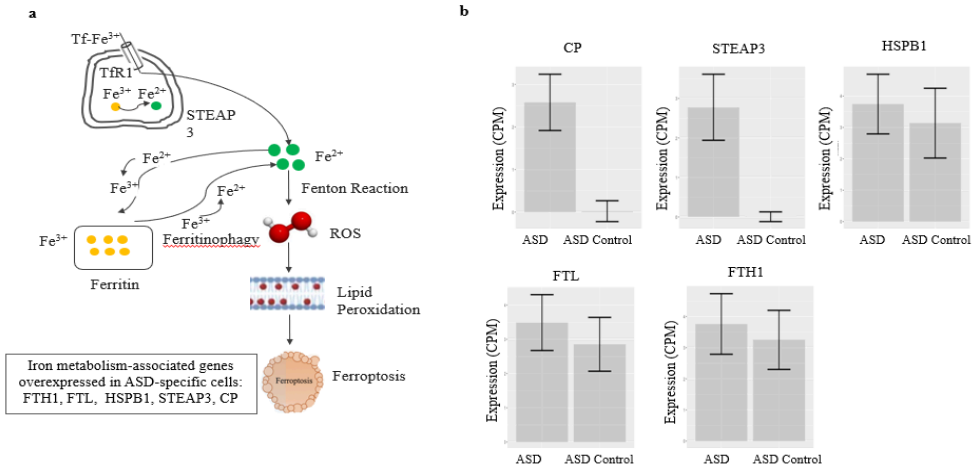

Figure 8. Iron Homeostasis: a. Iron metabolic routes contributing to ferroptosis. b. Genes associated with iron metabolism that are overexpressed in the ASD-specific cells.

MAPK-NF-κB Pathway Activated by ROS. MAPKs family, including extracellular signal-regulated kinase (ERK), p38 MAPK, and c-Jun N-terminal kinase, are a group of signaling molecules that play an important role in the expression of pro-inflammatory cytokine (Liu et al., 2020). As shown in Figure 9.a, antioxidants and inhibitors of ROS producing enzymatic systems block MAPK activation. If the production of ROS exceeds the capacity of the antioxidant proteins, it may cause oxidative stress. Oxidative stress may induce MAPKs signaling protein (e.g., RTKs and MAP3Ks), thereby leading to MAPK activation and subsequently triggering NF-κB to express pro-inflammatory cytokines (Liu et al., 2020). Our analysis indicates MAPK and c-Jun are overexpressed in the ASD disease cells (Figure 9.b).

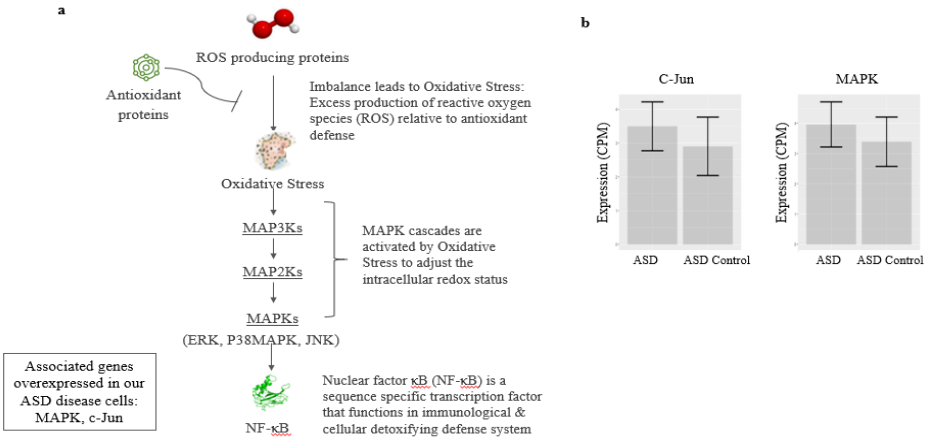

Figure 9. MAPK-NF-κB Pathway: a. MAPK-NF-κB Pathway activated by ROS. b. MAPK and c-Jun are overexpressed in the ASD-specific cells.

KEAP1/Nrf2/ARE Signaling Pathway. Nrf2 is a transcription factor that upregulates expression of genes crucial for defending against ROS and reactive nitrogen species (Schrier et al., 2022). As shown in Figure 10.a, KEAP1 inhibits the transcriptional activity of Nrf2 via ubiquitination. Under oxidative stress, Nrf2 disassociates with KEAP1 binding and then translocates into the nucleus, heterodimerizes with small musculoaponeurotic fibrosarcoma (Maf) proteins and binds to antioxidant response elements (AREs) to activate antioxidant enzymes, such as heme oxygenase-1 (HO-1). Upregulated HO-1 catalyzes the heme into CO, bilirubin and iron (Ahmed et al., 2017). They exert beneficial protective effects against oxidative injury. HO-1 also directly inhibits the proinflammatory cytokines and activates the anti-inflammatory cytokines, leading to balancing of the inflammatory process (Ahmed et al., 2017). As indicated in Figure 10.b, Nrf2 levels and Nrf2/KEAP1 activity are dysregulated in ASD. Nrf2 and KEAP1 levels were higher while HO-1 levels were lower in our ASD-specific cells. Our hypothesis is that oxidative stress is higher in ASD, and HO-1 levels are insufficient to achieve oxidative balance.

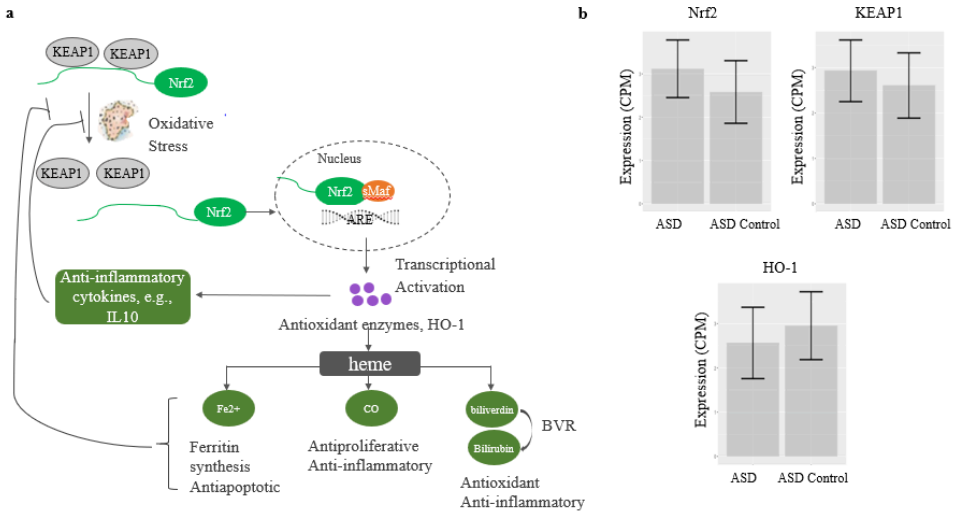

Figure 10 KEAP1/Nrf2/ARE Signaling Pathway. a. KEAP1/Nrf2/ARE Signaling Pathway promotes expression of antioxidant genes. b. Nrf2 and KEAP1 are overexpressed and HO-1 is under-expressed in the ASD-specific cells.

AhR Pathway. AhR is a receptor for multiple physiological ligands, with important roles in inflammation. The inactive form of AHR is localized in cytosolic multiprotein complex consisting of HSP90, tyrosine kinase c-src, and other cochaperones. Upon ligand binding, this complex dissociates with AhR, which results in its translocation to the nucleus. In the

nucleus, AhR dimerizes with aryl hydrocarbon receptor nuclear translocator (ARNT), and the heterodimer is responsible for the transcription of xenobiotic-responsive elements (XRE) – containing genes (Grishanova & Perepechaeva, 2022). Existing research identifies AhR as a negative regulator of NF-κB transcriptional responses that promote neurotoxic peripheral monocyte recruitment to the CNS and astrocyte-intrinsic neurotoxic activities (Barroso et al., 2021). As shown in Figure 11.a, AhR plays an important role in the cellular defense against ROS through its target genes and activation of AhR-Nrf2 pathway. The AhR-ARNT complex initiates the transcription of cytochrome P450 family 1 (CYP1A1, CYP1A2, CYP1B1) and AhR repressor (AHRR), which are overexpressed in our ASD disease cells (Figure 11.b).

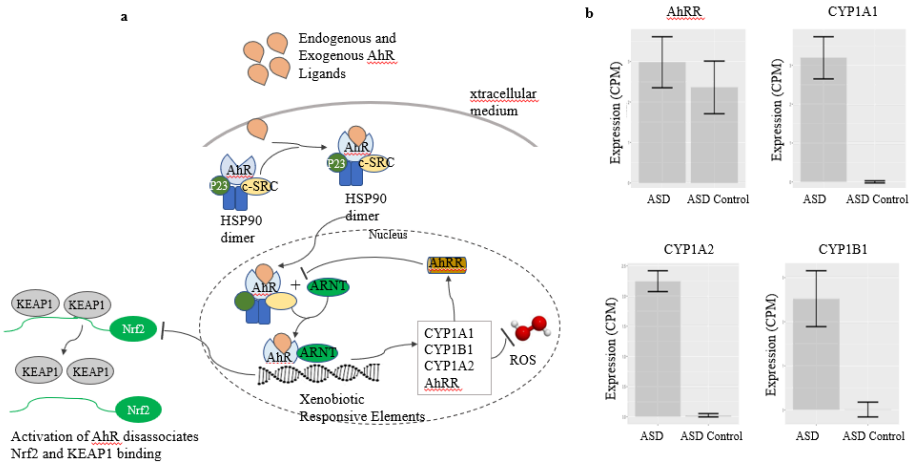

Figure 11 AhR Signaling Pathway. a. The inactive form of AhR is localized in the cytosol in a complex composed of HSP90, AIP, p23, and XAP2. AhR agonists induce conformational changes in AhR that result in its translocation to the nucleus. In the nucleus, AhR interacts with ARNT and activates the transcription of XRE-containing genes. Simultaneously, activation of AhR initiates dissociation of the Nrf2–Keap1 complex. b. Associated genes overexpressed in the ASD disease cells.

Hippo Signaling Pathway. Hippo signaling pathway is an evolutionarily conserved pathway that responds to various environmental cues and controls organ size, cell proliferation, death, and self-renewal capacity (Sun & Chi, 2020). As shown in Figure 12.a, the regulatory endpoints of the Hippo pathway are the two homologous transcriptional co-activators, yes-associated protein (YAP) and transcriptional co-activator with PDZ-binding motif (TAZ). The regulation of YAP and TAZ is governed by two major protein kinase complexes, the mammalian Sterile 20-like kinases 1 and 2 (MST1/2), and the large tumor suppressor homolog LATS1/2 and their direct interaction partners SAV1 and MOB1A/MOB1B (Stepan et al., 2018). Activation of MST1/2 and LATS1/2 causes phosphorylation of YAP/TAZ. Phospho-YAP/TAZ is degraded in the cytoplasm by the 14-3-3 protein, whereas after inactivation of the upstream kinase cascade, dephosphorylated YAP/TAZ translocates to the nucleus (Sun & Chi, 2020). Nuclear translocation of YAP/TAZ results in a reduction in NF-κB and pro-inflammatory cytokine production (Sun & Chi, 2020). Degraded YAP/TAZ are retained cytosolically and leads to ferroptosis inhibition (Sun & Chi, 2020). YAP1, SAV1, MOB1A and TEAD1 are overexpressed in the ASD disease cells. (Figure 12.b).

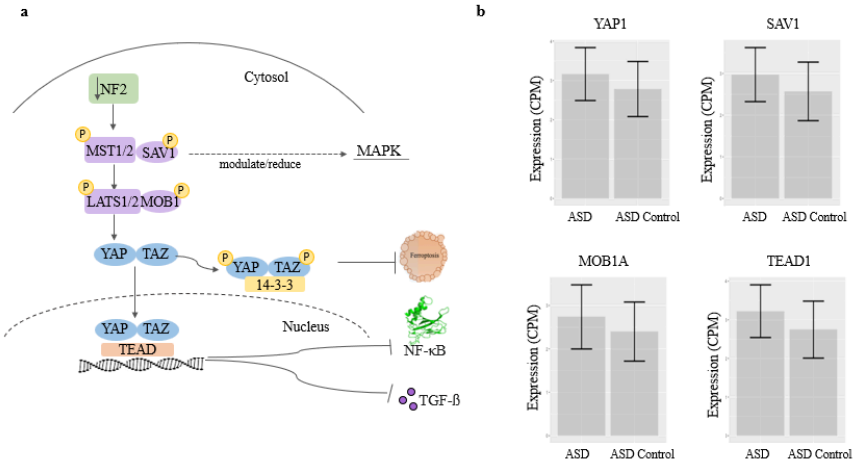

Figure 12 Hippo Signaling Pathway. a. NF2 activates Hippo pathways. The degradation of YAP/TAZ leads to ferroptosis inhibition, whereas translocation of YAP/TAZ to nucleus results in a reduction in NF-κB and pro-inflammatory cytokine production. **b.** Associated genes overexpressed in ASD disease cells.

2. Inflammatory markers reflecting disease activity in MS and ASD

We noted that Cluster 5 was predominantly composed of chronic inactive MS plaques, while Cluster 1 shows higher percentage of acute/chronic active MS plaques. Chronic active lesions have macrophages at their rim, leaving the center fully demyelinated. In chronic inactive lesions, there is complete demyelination, and microglia and macrophages are no longer present (Davoli-Ferreira et al., 2021). In ASD, current research believes that microglia play a key role in the immunosurveillance of the CNS during embryogenesis, in which they modulate neurogenesis through neurotoxic or neurotrophic factors. The main symptoms in ASD have been linked to defects in neurogenesis. However, post-mortem brains indicate pro- and anti-inflammatory factors downstream of such inflammation in adult ASD patients, suggesting that the state of the CNS in adult ASD patients is more similar to MS patients with chronic inactive lesions. The common astrocyte state connecting ASD with chronic inactive MS stage could be reflective of this hypothesis, and the neuroinflammatory conditions of adult ASD patients more closely resembles the chronic inactive MS stage instead of other MS stages such as acute/chronic active. This research therefore suggests new evidence sustaining this belief and indicates that chronic inactive lesions in MS could be explored as a model for studying ASD in the future.

3. Unclear roles of CCK in ASD and MS

Our analysis of common DEGs revealed CCK as being upregulated in MS and downregulated in ASD. Related proteins indicated the involvement of Cell-type Dependent Selectivity of CCKR Signaling. CCK receptor (CCKR) is divided into CCK1 receptor (CCK1R) and CCK2 receptor (CCK2R). CCK1R is mainly distributed in the gastrointestinal tract, peripheral nervous system, and some regions of the brain. CCK2R is primarily expressed in the brain, particularly in the cortex and the limbic structures and selected regions in the gastrointestinal tract. CCK receptors regulate a variety of physiological functions including digestion, satiety, emotion regulation, pain sensation, and memory process (Ding et al., 2022). While it has been widely reported that GI inflammation is one of the most common comorbidities in both MS and ASD, this research suggests that the dual function of CCK and CCKR in the gut and brain may cause such comorbidity through the gut-brain axis. While research has shown its involvement in

Gamma-aminobutyric acid (GABA) release, we were unable to be conclusive about its roles in the brain. Additional research in this direction could help elucidate the neurological mechanism of CCK in mediating social behaviors.

4. Additional validation of the marker genes and pathways identified

Compared with scRNA-seq, bulk-RNA-seq produces less noisy and less variable data which leads to more robust and reproducible results. Using the gene set 2.cp.wikipathways.v7.5.1.symbols.gmt with GSEA, we validated the pathways detected by our scRNA-seq analysis. Supplementary Table 5 contains the pathways of the common clusters found with scRNA-seq analysis that have an FDR q-val < 0.05. A few inconsistencies and additional pathways are revealed through GSEA of the IPSC dataset, which may be due to the cells having been reprogrammed rather than developing naturally.

Conclusion

In conclusion, recent advancements in high-throughput single-cell RNA sequencing have presented a valuable opportunity to investigate the heterogeneity of astrocytes and its implications in both healthy and diseased states. Linking transcriptionally defined astrocyte subpopulations cross-disease between ASD and MS has led us to propose that inflammation-promoting oxidative stress-induced ferroptosis plays a pronounced role in the pathological astrocyte subpopulations in ASD. This discovery sheds light on the underlying mechanisms of neuroinflammation in ASD and underscores the importance of further research into astrocyte-mediated neuroinflammation. Such research holds the potential to uncover novel therapeutic strategies for ASD.

Supplementary Materials

- Supplementary Table 1: Disease specific DEGs obtained from PRJNA544731 and PRJNA434002
- Supplementary Table 2: Complete list of overlapped Seurat and Monocle marker genes for MS and ASD in Cluster 5
- Supplementary Table 3: Complete listing of the genes defined on the trajectories involving the common clusters
- Supplementary Table 4: The ontology for all the proteins in PPI based on the marker genes of common clusters
- Supplementary Table 5: The pathways of the common clusters found with scRNA-seq analysis that have a NOM p-val < 0.05

Acknowledgments

The author is deeply grateful to the members of the Quintana Lab at Brigham and Women's Hospital headed by Professor Francisco J. Quintana who provided generous support for this research. Dr. Michael Wheeler supervised the entire study with extraordinary insights and inspirations. Dr. Zhaorong Li provided wonderful technical advice. A special thank-you also goes to Dr. Summer Morrill at Phillips Exeter Academy and Professor James Coffman and Dr. Elisabeth Marnik at MDI Biological Laboratory for enlightening discussions.

References

1. Sher, D.A., Gibson, J.L. Pioneering, prodigious and perspicacious: Grunya Efimovna Sukhareva's life and contribution to conceptualising autism and schizophrenia. Eur Child Adolesc Psychiatry 32, 475–490 (2023).

2. American Psychiatric Association. (2022). Neurodevelopmental disorders. In Diagnostic and statistical manual of mental disorders (5th ed., text rev.).

3. Maenner, M.J., Shaw, K.A., Bakian, A.V. et al. Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2018. MMWR Surveill Summ 2021;70(No. SS-11):1–16.

4. Zeidan, J., Fombonne, E., Scorah, J. et al. Global prevalence of autism: A systematic review update. Autism Research 2022 March.

5. Kadowaki A., Quintana F.J. The Gut-CNS Axis in Multiple Sclerosis. Trends Neurosci. 2020 Aug; 43(8):622-634

6. Reich, D.S., Lucchinetti C.F., Calabresi, P.A. Multiple Sclerosis. N Engl J Med. 2018 Jan 11; 378(2):169-180.

7. Lee, H.G., Wheeler, M.A,, Quintana, F.J. Function and therapeutic value of astrocytes in neurological diseases. Nat Rev Drug Discov. 2022 May;21(5):339-358.

8. Wheeler, M.A,, Clark, I.C,, Tjon, E.C. et al. MAFG-driven astrocytes promote CNS inflammation. Nature. 2020 Feb;578(7796):593-599.

9. Linnerbauer, M., Wheeler, M.A., Quintana, F.J. Astrocyte Crosstalk in CNS Inflammation, Neuron, Volume 108, Issue 4, 2020, Pages 608-622.

10. Matta, S.M., Hill-Yardin, E.L., Crack, P.J. The influence of neuroinflammation in Autism Spectrum Disorder. Brain Behav Immun. 2019 Jul;79:75-90.

11. Gzielo, K., Nikiforuk, A. Astroglia in Autism Spectrum Disorder. Int J Mol Sci. 2021 Oct 26;22(21):11544.

12. Speir, M.L., Bhaduri, A., Markov, N.S. et al. UCSC Cell Browser: visualize your single-cell data, Bioinformatics, Volume 37, Issue 23, 1 December 2021, Pages 4578–4580.

13. Vashisth, S., & Chahrour, M. H. (2023). Genomic strategies to untangle the etiology of autism: A primer. Autism Research, 16(1), 31– 39.

14. Toscano, C.V.A., Barros, L., Lima, A.B. et al. Neuroinflammation in autism spectrum disorders: Exercise as a "pharmacological" tool. Neurosci Biobehav Rev. 2021 Oct;129:63-74.

15. Eissa, N., Sadeq, A., Sasse, A., Sadek, B. Role of Neuroinflammation in Autism Spectrum Disorder and the Emergence of Brain Histaminergic System. Lessons Also for BPSD? Front Pharmacol. 2020 Jun 16;11:886.

16. Zürcher, N.R., Loggia, M.L., Mullett, J.E. et al. [11C] PBR28 MR–PET imaging reveals lower regional brain expression of translocator protein (TSPO) in young adult males with autism spectrum disorder. Mol Psychiatry 26, 1659–1669 (2021).

17. Rahnama, M., Tehrani, H.A., Mirzaie, M., Ziaee, V. Identification of key genes and convergent pathways disrupted in autism spectrum disorder via comprehensive bioinformatic analysis, Informatics in Medicine Unlocked, Volume 24, 2021, 100589, ISSN 2352-9148.

18. Trifonova, E.A., Klimenko, A.I., Mustafin, Z.S. et al. Do Autism Spectrum and Autoimmune Disorders Share Predisposition Gene Signature Due to mTOR Signaling Pathway Controlling Expression? Int J Mol Sci. 2021 May 16;22(10):5248.

19. Whiteley, P., Marlow, B., Ritika, K.R. et al. Autoimmune Encephalitis and Autism Spectrum Disorder. Frontiers in Psychiatry, 12/2021.

20. Luecken, M.D., Theis, F.J. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol. 2019 Jun 19;15(6):e8746.

21. Koch, C.M., Chiu, S.F., Akbarpour, M., A Beginner's Guide to Analysis of RNA Sequencing Data. Am J Respir Cell Mol Biol. 2018 Aug;59(2):145-157.

22. Korsunsky, I., Millard, N., Fan, J. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019 Dec;16(12):1289-1296.

23. Tran, H.T.N., Ang, K.S., Chevrier, M. et al. A benchmark of batch-effect correction methods for single-cell RNA sequencing data. Genome Biol 21, 12 (2020).

24. Duò, A, Robinson, M.D., Soneson, C. A systematic performance evaluation of clustering methods for single-cell RNA-seq data. F1000Res. 2018 Jul 26;7:1141.

25. Comi, A.M., Zimmerman A.W., Frye, V.H. et al. Familial Clustering of Autoimmune Disorders and Evaluation of Medical Risk Factors in Autism. Journal of Child Neurology. 1999;14(6):388-394.

26. Trapnell, C., Cacchiarelli, D., Grimsby, J. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 2014 Apr;32(4):381-386.

27. Li, J., Cao, F., Yin, H. et al. Ferroptosis: past, present and future. Cell Death Dis 11, 88 (2020).

28. Miller, S.J. Astrocyte Heterogeneity in the Adult Central Nervous System. Front Cell Neurosci. 2018 Nov 15;12:401.

29. Tang, D., Chen, X., Kang, R. et al. Ferroptosis: molecular mechanisms and health implications. Cell Res 31, 107–125 (2021).

30. Davoli-Ferreira, M., Thomson, C.A., McCoy, K.D. Microbiota and Microglia Interactions in ASD. Front Immunol. 2021 May 25;12:676255.

31. Rothhammer, V., Mascanfroni, I.D., Bunse, L. et al. Type I interferons and microbial metabolites of tryptophan modulate astrocyte activity and central nervous system inflammation via the aryl hydrocarbon receptor. Nat Med. 2016 Jun;22(6):586-97.

32. Morshed, S.A., Davies, T. Understanding Thyroid Cell Stress. J Clin Endocrinol Metab. 2020 Mar 1;105(3):e66–9.

33. Ding, Y., Zhang, H., Liao, Y.Y. et al. Structural insights into human brain–gut peptide cholecystokinin receptors. Cell Discov 8, 55 (2022).

34. Ou, M., Jiang, Y., Ji, Y. et al. Role and mechanism of ferroptosis in neurological diseases, Molecular Metabolism,Volume 61,2022, 101502, ISSN 2212-8778.

35. Pagani, M., Barsotti, N., Bertero, A. et al. mTOR-related synaptic pathology causes autism spectrum disorder-associated functional hyperconnectivity. Nat Commun 12, 6084 (2021).

36. Barbosa, A.G., Pratesi, R., Paz, G.S.C. et al. Assessment of BDNF serum levels as a diagnostic marker in children with autism spectrum disorder. Sci Rep 10, 17348 (2020).

37. Liu, Z., Yao, X., Jiang, W. et al. Advanced oxidation protein products induce microglia-mediated neuroinflammation via MAPKs-NF-κB signaling pathway and pyroptosis after secondary spinal cord injury. J Neuroinflammation 17, 90 (2020).

38. Schrier, M.S., Zhang, Y., Trivedi, M.S., Deth, R.C. Decreased cortical Nrf2 gene expression in autism and its relationship to thiol and cobalamin status. Biochimie. 2022 Jan;192:1-12.

39. Ahmed, S.M., Luo, L., Namani, A. et al. Nrf2 signaling pathway: Pivotal roles in inflammation. Biochim Biophys Acta Mol Basis Dis. 2017 Feb;1863(2):585-597.

40. Grishanova, A.Y., Perepechaeva, M.L. Aryl Hydrocarbon Receptor in Oxidative Stress as a Double Agent and Its Biological and Therapeutic Significance. Int J Mol Sci. 2022 Jun 16;23(12):6719.

41. Barroso, A., Mahler, J.V., Fonseca-Castro, P.H., Quintana, F.J. The aryl hydrocarbon receptor and the gut-brain axis. Cell Mol Immunol. 2021 Feb;18(2):259-268.

42. Sun, T., Chi, J.T. Regulation of ferroptosis in cancer cells by YAP/TAZ and Hippo pathways: The therapeutic implications. Genes Dis. 2020 May 18;8(3):241-249.

43. Stepan, J., Anderzhanova, E., Gassen, N.C. Hippo Signaling: Emerging Pathway in Stress-Related Psychiatric Disorders? Front Psychiatry. 2018 Dec 21;9:715.

44. Zhang, S., Xin, W., Anderson, G.J. et al. Double-edge sword roles of iron in driving energy production versus instigating ferroptosis. Cell Death Dis 13, 40 (2022).