Immunologic Research (2025) 73:57
https://doi.org/10.1007/s12026-025-09609-6

RESEARCH

A comprehensive analysis of scRNA-Seq and RNA-Seq unveils B cell immune suppression in the AAV-loaded brain

Shunyu Wu[1,2,3,4] · Lu Xue[1,3,4] · Xiang Li[1,3,4] · Yaoxuan Wang[1,3,4] · Yuting Zhu[1,3,4] · Yuanbo Luo[1,3,4] · Jiayu Sun[1,3,4] · Tingting Jin[1,3,4] · Wenying Shu[1,3,4] · Zhaoyan Wang[1,2,3,4]

Received: 15 November 2024 / Accepted: 17 February 2025

Abstract

The use of AAV vectors for in vivo gene therapy has demonstrated the potential to permanently correct genetic diseases by delivering functional gene copies to the nuclei of affected tissues. AAV vectors, as tools for in vivo gene delivery, are particularly appealing and have shown safety and long-term efficacy in numerous organ-targeted experiments. Nevertheless, employing AAV vectors for gene therapy in the brain faces a notable hurdle in the shape of immune responses, chiefly instigated by the brain's resident immune cells, microglia. Additionally, lower levels of AAV vector-neutralizing antibodies have been detected in the cerebrospinal fluid compared to the circulatory system. This research, leveraging transcriptomic and single-cell RNA sequencing (scRNA-seq) data in conjunction with Mendelian randomization analysis, has identified the potential role of the XBP1 protein in mediating B-cell immunosuppression in the brain via the MDK-NCL ligand-receptor pair and associated genes. Furthermore, it paves the way for further investigation into the regulatory factors and pathways within the immune modulation network, as well as their prospective beneficial implications in immunotherapeutic treatments. By employing various innovative approaches, the study seeks to recreate the immune environment generated by AAV in the brain and preliminarily explore the immune suppression mechanisms induced by AAV vectors in the brain.

Keywords Immune suppression · The AAV-loaded brain · XBP1 · MDK-NCL ligand-receptor pair · B lymphocytes

Introduction

The recombinant adeno-associated virus (AAV) consists of a roughly 26-nm-diameter icosahedral protein capsid and a single-stranded DNA genome of approximately 4.7 kilobases [1]. In recent years, the use of AAV vectors for in vivo gene therapy has demonstrated the potential for permanent correction of genetic diseases by delivering copies of functional genes into the nuclei of affected tissue cells. The transferred genes, or transgenes, compensate for the genetic mutations underlying hereditary disorders. AAV vectors, as an especially attractive tool for in vivo gene delivery, are non-integrative and capable of transducing a variety of differentiated tissues to achieve long-term transgene expression. Adeno-associated virus AAV vector gene therapy has proven its safety and long-term efficacy in multiple organ-targeted experiments, including the eyes, liver, skeletal muscles, and central nervous system [2–6]. Since the initial evidence that AAV vectors can elicit human capsid T cell responses and impact the duration of transgene expression, significant progress has been made

Shunyu Wu, Lu Xue, Wenying Shu, and Zhaoyan Wang contributed equally to this work.

Shunyu Wu and Lu Xue share first authorship.

Wenying Shu and Zhaoyan Wang are co-corresponding authors.

✉ Wenying Shu
shuwenying1996@163.com

✉ Zhaoyan Wang
wzyent@sjtu.edu.cn

1 Department of Otolaryngology, Head and Neck Surgery, Shanghai Ninth People'S Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

2 College of Health Science and Technology, Shanghai Jiao Tong University School of Medicine, Shanghai, China

3 Ear Institute, Shanghai Jiao Tong University School of Medicine, Shanghai, China

4 Shanghai Key Laboratory of Translational Medicine on Ear and Nose Diseases, Shanghai, China

Published online: 05 March 2025

in understanding and modulating the immunogenicity of AAV vectors. However, the immune responses encountered in human gene transfer with AAV vectors have remained a significant barrier to advancements in the field [7]. The intricate interplay between innate and adaptive immunity appears to play a pivotal role in determining the outcomes of gene transfer experiments, particularly in the complex interactions between vector components and transgene products.

Several distinct natural serotypes of AAV have been isolated in nature, each characterized by variations in their capsid sequences [8]. The presence of distinct capsid serotypes and specific receptors on host cells determine the selectivity of each AAV serotype for different tissues [9]. AAV vectors exhibit broad adaptability, rendering them ideal vehicles for multifaceted therapeutic applications. These vectors can be prepared through various methods, with the most common involving the utilization of three distinct DNA plasmids encoding the vector genome, rep and cap genes sourced from specific AAV serotypes, and helper plasmids for transfection into HEK293 cell lines [10, 11]. The capsid, its genome, and transgenic products represent the primary potential immunogenic components of AAV vectors, about which our current understanding remains limited. In recent years, immune-mediated toxicity has been considered a significant factor that may be triggered during clinical trials, garnering increasing attention. However, comprehensive clinical evidence is still lacking to definitively establish direct causality. A key determinant of the anti-transgene immune response level is the target organ for gene delivery, determined by the tissue-specific combination of AAV capsids, vector delivery routes, and tissue-specific promoters driving gene expression [12–14].

Compared to systemic vector delivery, AAV vectors have several distinctive features when it comes to delivery to the brain [7]. Firstly, due to the physical separation mediated by the blood–brain barrier, cerebrospinal fluid typically contains lower levels of antibodies capable of neutralizing AAV vectors. Secondly, there exists a unique resident immune cell subset in the brain, primarily composed of specialized macrophages known as microglial cells. Although studying brain samples is challenging, the human central nervous system (CNS) appears to mount fewer immune responses following gene transfer [15].

The brain harbors specific innate immune cells whose response to AAV vectors requires further investigation. Microglial cells are the initial responders to pathogens or tissue damage, triggering inflammatory reactions in the brain. Additionally, when engulfing dying or dead cells, microglial cells prevent the release of pro-inflammatory signals from necrotic tissue, thereby limiting further brain damage [16]. Furthermore, microglial cells are capable of engulfing virus-infected live cells. For instance, disruptions in intracellular calcium ion balance and exposure of phosphatidylserine have been observed in cases of adenovirus infection [17]. Nevertheless, it remains uncertain whether AAV vectors can induce cell stress levels associated with neuronal transduction, leading to microglial clearance of transduced cells.

The complexity and diversity of CNS diseases have driven researchers to propose the concept of multi-antigen/multi-epitope targeting, which is regarded as a more efficacious immunotherapy approach. Studies [18] have demonstrated that the suppression and treatment of chronic experimental autoimmune encephalomyelitis (EAE) through tolerogenic administration of Y-MSPc are associated with a significant increase in a unique subset of dendritic cells (DCs), specifically CD11c + CD11b + Gr1 + myeloid-derived DCs, observed in the spleens and CNS of treated mice. These DCs, characterized by their potent immunomodulatory properties, play a critical role in regulating MS-like diseases. They exhibit increased production of IL-4, IL-10, and TGF-β, while simultaneously reducing IL-12 production. These findings underscore the effectiveness of multi-epitope targeting agents in modulating immune responses and suggest the potential role of these immune cells in maintaining peripheral tolerance and participating in the regulation of CNS diseases. With the deepening understanding of the immune system and advancements in novel technologies, we anticipate the emergence of more innovative immunomodulatory approaches in the future. This development promises to expand the potential applications of gene therapy within the CNS.

In current research, there are significant disparities in potential immunomodulatory approaches regulating CNS gene transfer. Many unanswered questions persist regarding the immunogenicity of AAV vectors. This study, based on transcriptomic and single-cell RNA sequencing (scRNA-seq) data, and employing Mendelian randomization (MR) analysis, aims to explore causal relationships between immune factors and the brain immune environment loaded with AAV. By identifying key genes and their immunological pathway effects, this study seeks to elucidate the immune response mechanisms triggered by AAV vectors in the brain, providing crucial insights for further optimization of gene transfer therapeutic strategies. Approaching the immunogenicity of AAV vectors from a novel perspective and methodological angle, this research aims to offer vital scientific foundations for the advancement and clinical application of gene transfer therapies. By delving into the interactions between immune factors and AAV vectors, subsequent studies hold promise for bringing new breakthroughs and advancements to the field of neurological genetic therapies.

Methods

Data sourcing, downloading, and organization

We retrieved relevant gene expression datasets from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) of the National Center for Biotechnology Information (NCBI) [19]. During the retrieval process, we employed broadly defined keywords such as "AAV," "virus," and "CNS" to maximize the comprehensiveness of our search. Subsequently, we conducted a rigorous manual screening of the retrieved datasets based on predefined inclusion criteria. The screening process prioritized datasets containing brain tissue samples while excluding those with incomplete information or that did not meet our analysis requirements. Emphasis was placed on datasets involving AAV viral vectors or including control groups, as these were critical for our research objectives. After meticulous evaluation and comparison, we ultimately selected 14 transcriptome datasets and 4 single-cell transcriptome datasets for further analysis.

For the transcriptome data, we selected two groups, comprising a total of 14 samples from the dataset GSE235515. All samples were derived from the brain tissue of wild-type C57 mice. The experimental group received an injection of AAV virus, whereas the control group did not undergo any injection. For the single-cell RNA sequencing (scRNA-seq) data, we selected two samples from each of two datasets (GSE230709 and GSE233363), also sourced from the brain tissue of wild-type C57 mice. In the experimental group, mice were intraperitoneally injected with 10^{10} AAV9 virus particles twice, at weeks 0 and 3; the control group did not receive this treatment. After acquiring these datasets, we downloaded the count and matrix files for the transcriptome data, as well as the barcode, feature, and matrix files for the scRNA-seq data. We then systematically organized and normalized these files to ensure the accuracy and reliability of subsequent analyses. The source code developed and utilized in this study is publicly accessible via our GitHub repository at the following URL: https://github.com/codeconnoisseur5/Immune-Suppression-in-the-AAV-Loaded-Brain.

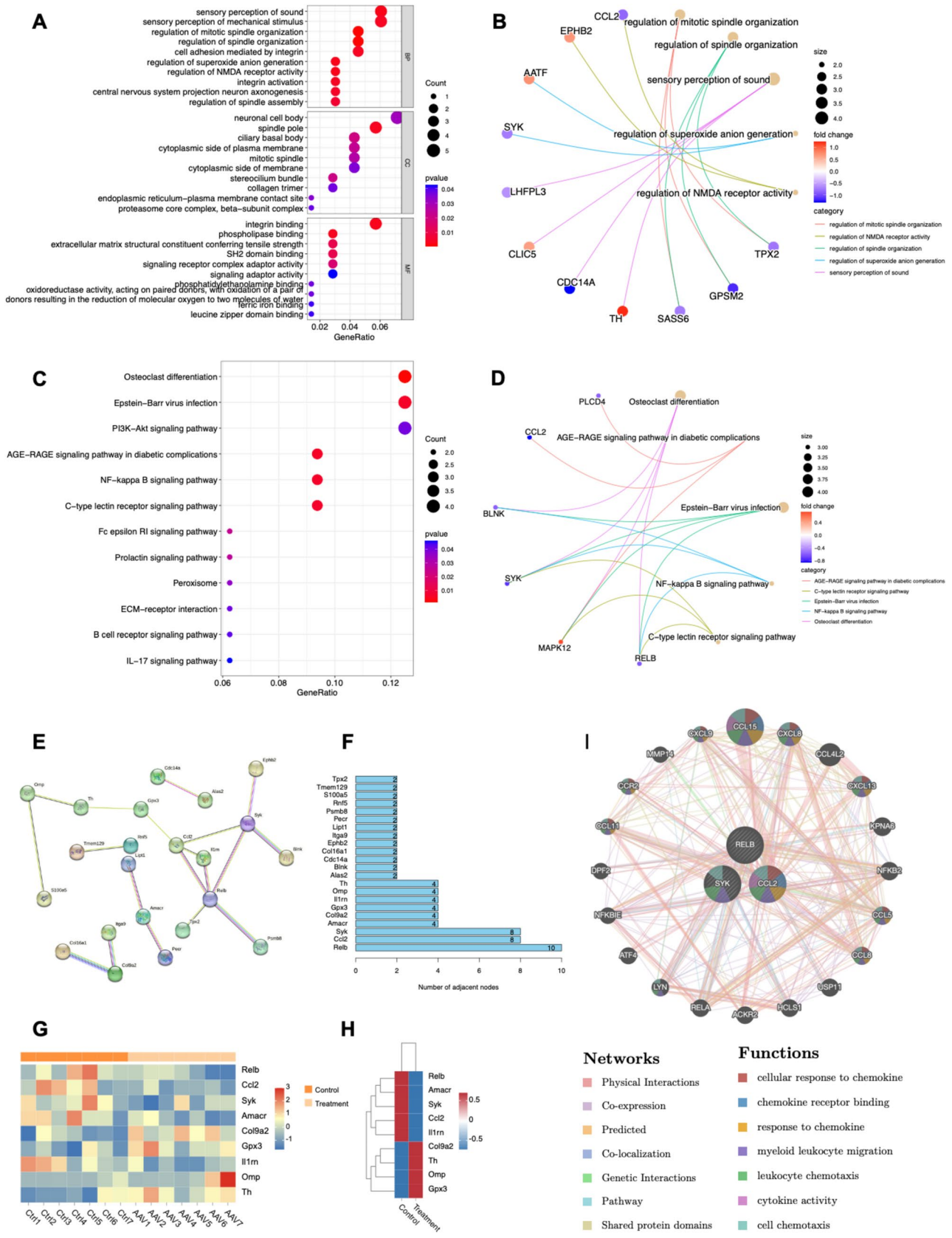
Processing of transcriptome data and functional enrichment analysis of differentially expressed genes

We performed a comprehensive analysis of mRNA-Seq expression data utilizing the limma [20] and pheatmap [21] packages in R. Initially, we imported the mRNA-Seq expression data as the foundation for our subsequent analyses. To enhance computational efficiency, we conducted preprocessing steps that included filtering out genes with an average expression level below 5 and selecting the top 100,000 most highly expressed genes to ensure both speed and data representativeness. Following this, we transformed the preprocessed data into a format compatible with the limma package. We then constructed linear models and performed differential expression analysis to identify genes significantly differentially expressed under varying conditions. For screening thresholds, we applied criteria of a P value less than 0.05 and an absolute log2 fold change greater than or equal to 1, ensuring the robustness and reliability of our results. Additionally, we incorporated gene interaction data from the STRING database (https://cn.string-db.org/cgi/input?sessionId=bLrKx9CmduN4). By analyzing these interactions, we quantified the frequency of each gene's occurrence within the network and visualized the results using horizontal bar charts. Finally, leveraging the list of significantly differentially expressed gene ID, we conducted Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis using the clusterProfiler package [22]. These analyses provided insights into the biological functions and pathways associated with the differentially expressed genes.

Cluster annotation of core cells

We employed the R programming language to load a series of essential software packages, including Seurat [23], SingleR [24], and Monocle [25], which provided robust support for our analysis of single-cell RNA sequencing data. Initially, we imported the raw count matrix of the single-cell RNA sequencing data as the foundation for our analysis. During the data preprocessing phase, we applied stringent criteria to filter out low-quality cells and lowly expressed genes. Specifically, we established reasonable thresholds and excluded cells with fewer than 50 features (i.e., number of expressed genes) and those with a mitochondrial gene proportion exceeding 15%. This process focused on evaluating the total RNA count, number of features, and proportion of mitochondrial genes for each cell to ensure the accuracy and reliability of the data. Following the completion of data preprocessing, we conducted normalization procedures, which included normalization, logarithmic transformation, and scaling steps. The normalized data enabled us to identify highly variable features, which played a critical role in subsequent clustering analysis and dimensionality reduction operations.

To correct for batch effects, we utilized the Harmony algorithm [26], which effectively mitigated noise arising from batch differences. Subsequently, we identified a set of highly variable genes through analysis of variance for further analysis, as these genes more accurately capture the biological variability between cells. During the principal component analysis (PCA) phase [27], we selected an optimal number of principal components ($n = 20$) based on data characteristics to maximize the retention of variation


A
sensory perception of sound
sensory perception of mechanical stimulus
regulation of mitotic spindle organization
regulation of spindle organization
cell adhesion mediated by integrin
regulation of superoxide anion generation
regulation of NMDA receptor activity
integrin activation
central nervous system projection neuron axonogenesis
regulation of spindle assembly
neuronal cell body
spindle pole
ciliary basal body
cytoplasmic side of plasma membrane
mitotic spindle
cytoplasmic side of membrane
stereocilium bundle
collagen trimer
endoplasmic reticulum–plasma membrane contact site
proteasome core complex, beta–subunit complex
integrin binding
phospholipase binding
extracellular matrix structural constituent conferring tensile strength
SH2 domain binding
signaling receptor complex adaptor activity
signaling adaptor activity
phosphatidylethanolamine binding
oxidoreductase activity, acting on paired donors, with oxidation of a pair of donors resulting in the reduction of molecular oxygen to two molecules of water
ferric iron binding
leucine zipper domain binding
BP
CC
MF
Count
pvalue
GeneRatio
B
CCL2
EPHB2
regulation of mitotic spindle organization
regulation of spindle organization
AATF
sensory perception of sound
SYK
regulation of superoxide anion generation
LHFPL3
regulation of NMDA receptor activity
CLIC5
TPX2
CDC14A
GPSM2
TH
SASS6
size
fold change
category
C
Osteoclast differentiation
Epstein–Barr virus infection
PI3K–Akt signaling pathway
AGE–RAGE signaling pathway in diabetic complications
NF–kappa B signaling pathway
C–type lectin receptor signaling pathway
Fc epsilon RI signaling pathway
Prolactin signaling pathway
Peroxisome
ECM–receptor interaction
B cell receptor signaling pathway
IL–17 signaling pathway
Count
pvalue
GeneRatio
D
PLCD4
Osteoclast differentiation
CCL2
AGE–RAGE signaling pathway in diabetic complications
BLNK
Epstein–Barr virus infection
SYK
NF–kappa B signaling pathway
MAPK12
C–type lectin receptor signaling pathway
RELB
size
fold change
category
E
F
Tpx2
Tmem129
S100a5
Rnf5
Psmb8
Pecr
Lipt1
Itga9
Ephb2
Col16a1
Cdc14a
Blnk
Alas2
Th
Omp
Il1rn
Gpx3
Col9a2
Amacr
Syk
Ccl2
Relb
Number of adjacent nodes
I
CXCL9
CCL15
CXCL8
MMP14
CCL4L2
CCR2
CXCL13
CCL11
KPNA6
RELB
DPF2
SYK
CCL2
NFKB2
NFKBIE
CCL5
ATF4
CCL8
LYN
USP11
RELA
HCLS1
ACKR2
G
Relb
Ccl2
Syk
Amacr
Col9a2
Gpx3
Il1rn
Omp
Th
Ctrl1
Ctrl2
Ctrl3
Ctrl4
Ctrl5
Ctrl6
Ctrl7
AAV1
AAV2
AAV3
AAV4
AAV5
AAV6
AAV7
Control
Treatment
H
Relb
Amacr
Syk
Ccl2
Il1rn
Col9a2
Th
Omp
Gpx3
Control
Treatment
Networks
Physical Interactions
Co-expression
Predicted
Co-localization
Genetic Interactions
Pathway
Shared protein domains
Functions
cellular response to chemokine
chemokine receptor binding
response to chemokine
myeloid leukocyte migration
leukocyte chemotaxis
cytokine activity
cell chemotaxis

◂**Fig. 1** Integrating the transcriptional landscape of adeno-associated virus in the brain. **A** Performing GO analysis based on differentially expressed genes identified from transcriptome sequencing. The size of the circles represents the number of differentially expressed genes, with the color of the circles ranging from red to blue indicating increasing significance. **B** Association diagram illustrating the relationship between key genes and the key biological activities identified through GO analysis. **C** Conducting KEGG analysis based on differentially expressed genes identified from transcriptome sequencing. The size of the circles corresponds to the number of differentially expressed genes, with colors ranging from red to blue indicating increasing significance. **D** Association diagram illustrating the relationship between key genes and the key pathways identified through KEGG analysis. **E** Protein–protein interaction network analysis constructed using the STRING database. **F** The top 12 differentially expressed genes based on the number of key nodes identified through a combined analysis of KEGG and GO. **G** Heatmap illustrating the expression levels of key differentially expressed genes. The horizontal axis represents individual samples, while the vertical axis represents gene names. **H** Heatmap illustrating the expression levels of key differentially expressed genes. The horizontal axis represents control and treatment groups, while the vertical axis represents gene names. **I** Prioritizing the relevant genes through GeneMANIA for functional analysis

information. We then conducted global dimensionality reduction using the t-SNE algorithm, enabling visualization of cell populations on a two-dimensional plane and facilitating intuitive observation and analysis of cell distribution and relationships. Additionally, we employed the SingleR package to assess cell similarity and performed clustering analysis. This process allowed us to identify marker genes with significant expression in each cluster, providing a robust foundation for subsequent research and enhancing our understanding of biological differences and interactions among different cell types.

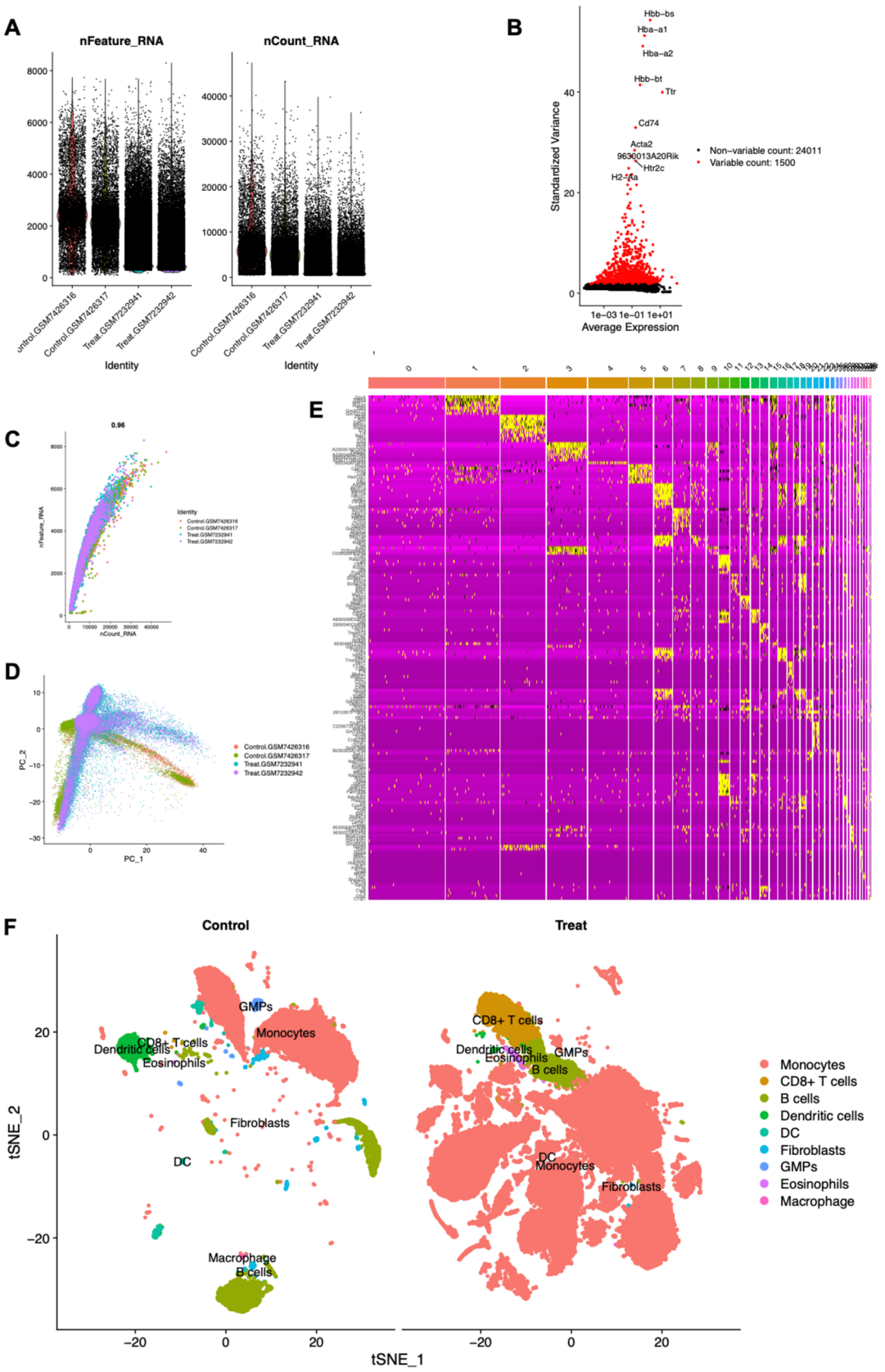
Cellular communication mechanisms implementation

To conduct an in-depth analysis of intercellular communication, we initially installed the CellChat package [28] and integrated essential auxiliary packages, including dplyr for data manipulation and ggplot2 for visualization. We then successfully constructed a CellChat object using the prepared expression matrix and metadata, which facilitated the enumeration of cells within each cell type and allowed us to evaluate the distribution of different cell types across the dataset. Following this, we incorporated the Human CellChat database, which encompasses comprehensive prior knowledge of human intercellular communication. To enhance the accuracy of our data analysis, we performed preprocessing by eliminating lowly expressed genes and irrelevant interactions, thereby minimizing data noise, and focusing on key communication pathways. Subsequently, we identified overexpressed genes specific to certain cell types and mapped these genes along with their interactions onto the human protein–protein interaction (PPI) network. This mapping provided an intuitive representation of potential pathways for intercellular communication.

To quantify the likelihood of intercellular communication, we computed communication probabilities and applied a stringent filtering criterion to exclude events occurring in fewer than 10 cells. This approach minimized data noise and enabled us to concentrate on significant communication events. Moreover, we evaluated communication probabilities between signaling pathways and integrated these interactions into a comprehensive network model. The resulting network facilitated a holistic understanding of intercellular communication dynamics. To visually represent the frequency and intensity of cell network interactions, we generated visualization images. Additionally, we constructed individual network diagrams for each cell type, delineating their specific interactions and offering a granular perspective on intercellular communication. Finally, we produced bubble plots that graphically illustrated the communication relationships and intensities between different cell types. The size and color coding of the bubbles corresponded to the frequency and intensity of communication, respectively, providing an intuitive and easily interpretable visualization of intercellular interactions.

Mendelian randomization analysis

In genomic analysis, we employed a suite of R packages specifically designed for genome-wide association studies (GWAS) [29] and MR [30]. Our focus was on datasets pertaining to differential gene expression in B cells and microglia, which included exposure data files, gene list files, and specific result data identifiers (e.g., prot-a-737). For each exposure variable, we conducted MR analyses using packages such as gwasglue, VariantAnnotation, TwoSampleMR, and other Bioconductor resources. Throughout this process, we applied stringent screening criteria to ensure the rigor and accuracy of our analysis. Upon completion, we meticulously documented the results for subsequent interpretation and visualization. To comprehensively evaluate the robustness of our findings, we generated a series of graphical representations, including scatter plots, funnel plots, and sensitivity analysis plots. Notably, we utilized the leave-one-out method to further validate the reliability of our results. Based on the MR analysis outcomes, we identified significant genes and constructed a forest plot to visually present estimated effects along with their confidence intervals, thereby aggregating data from multiple studies or experiments and providing strong support for understanding gene effects. Additionally, we used volcano plots to illustrate the relationship between gene effect sizes (beta values) and the negative logarithm of the corresponding *p* values. Volcano


A
nFeature_RNA
nCount_RNA
Identity
B
Standardized Variance
Average Expression
Non−variable count: 24011
Variable count: 1500
C
D
PC_1
PC_2
E
F
Control
Treat
tSNE_1
tSNE_2
Monocytes
CD8+ T cells
B cells
Dendritic cells
DC
Fibroblasts
GMPs
Eosinophils
Macrophage

◂**Fig. 2** Visual analysis of clustering for single-cell sequencing. **A** The violin plot illustrates the extraction of feature genes. The left chart compares the number of feature genes between the control group and the treatment group, while the right chart evaluates sequencing depth in both groups. The horizontal axis denotes the control and treatment groups ($n=4$), respectively. The left vertical axis represents the count of gene features relevant to the research objectives, whereas the right vertical axis indicates the number of differentially expressed genes. **B** The feature variance plot highlights the top 10 feature genes along with their corresponding variance values. The horizontal axis lists various genes within the dataset, with the top 10 feature genes labeled accordingly. The vertical axis measures the variance value of each feature, where a higher variance signifies greater variability or information content within the dataset. **C** The scatter plot demonstrates the correlation between sequencing depth and the number of detected genes, exhibiting a correlation coefficient of 0.96. The horizontal axis represents sequencing depth, a critical metric that quantifies the amount and coverage of sequencing data, calculated as the total number of bases read per sample or the product of average read length and the number of reads per sample. The vertical axis reflects the number of genes successfully identified and counted at a given sequencing depth. A high correlation near 1 suggests that an increase in sequencing depth leads to a proportional increase in the number of detected genes, consistent with the principles and expected outcomes of high-throughput sequencing. **D** In the principal component analysis (PCA) result plot, the *x*-axis and *y*-axis represent the first principal component (PC1) and the second principal component (PC2), respectively. Each point in the plot corresponds to an individual observation, positioned based on its scores along these principal components. The relative positions of these points reflect the degree of similarity among the samples within the dataset. **E** The PCA heatmap illustrates the results for the top 20 clusters. The *x*-axis denotes different sample groups categorized by their spatial distribution following PCA analysis, while the *y*-axis represents differentially expressed genes. The color gradient indicates the magnitude of each element in the data matrix. **F** The t-SNE clustering visualization plot depicts the distribution of differentially expressed genes across various cell groups. Tsne_1 and Tsne_2 represent two dimensions derived from dimensionality reduction using the t-SNE algorithm, corresponding to the horizontal and vertical axes on the two-dimensional plane, respectively

plots not only intuitively display significance levels but also clearly indicate the direction of effects, serving as a powerful tool for gaining deeper insights into gene expression differences.

Prediction of core genes

For the initially identified genes, we employed the STRING database (https://string-db.org/) and GeneMANIA (http://genemania.org) [31] to weight the predictive values of each functional gene set for the query, generating hypotheses related to gene function, analyzing gene lists, and determining the priority of gene function detection.

Implementation of technological platform

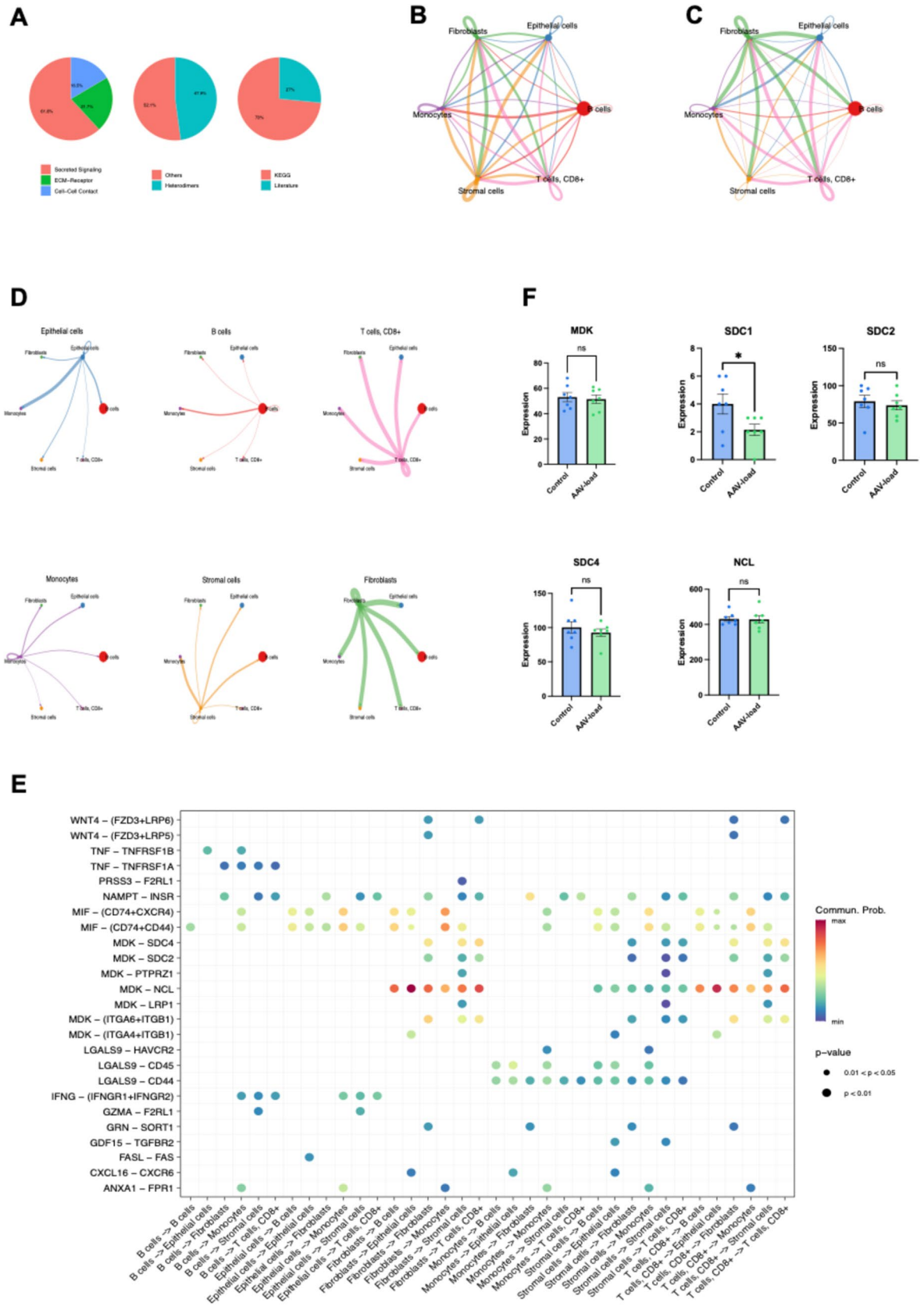
This study was conducted on the macOS 14.5 (23F79) operating system, utilizing the R 4.4.2 software environment. The software packages gwasglue, VariantAnnotation, and TwoSampleMR were installed via GitHub (https://github.com) and Bioconductor 1.30.25 [22, 32–35].

Results

Studying the transcriptomic landscape and key gene interaction networks of AAV in the integration process in the brain

To investigate the transcriptomic changes following the intracranial injection of adeno-associated virus (AAV), we conducted an in-depth analysis of mRNA datasets from both untreated mouse brains ($n=7$) and AAV-integrated mouse brains ($n=7$). Utilizing a linear model approach, we analyzed RNA-seq data covering 53,801 genes. Differential expression analysis identified a total of 11,980 differentially expressed genes (DEGs). Applying stringent p value and logFC filtering criteria (pFilter = 0.05, logFCfilter = 1), we ultimately identified 98 significantly differentially expressed genes. Specifically, compared to the control group, 47 genes were downregulated, and 51 genes were upregulated in the AAV-integrated group. We then extracted gene names and mapped them to Entrez ID, successfully mapping 73 genes. To elucidate the functional implications of these genes and proteins, we performed GO and KEGG pathway analyses. GO analysis revealed that DEGs were enriched in biological processes such as sound perception and mechanosensory perception, as well as cellular components like neuronal soma and spindle pole. The molecular functions enriched by DEGs included integrin binding and phospholipase binding (Fig. 1A, B). KEGG analysis indicated that DEGs were enriched in pathways such as osteoclast differentiation, Epstein-Barr virus infection, and PI3K-Akt signaling (Fig. 1C, D).

We constructed a protein–protein interaction network with a confidence score of 0.900 using the STRING database and identified 22 node genes from this network (Fig. 1E). Subsequently, we employed R language to quantify the interaction frequencies among these genes and visualized the results as a sorted horizontal bar chart (Fig. 1F). Based on higher interaction frequency, we selected nine genes, including RELB, CCL2, and SYK, as key targets for further investigation and generated a heatmap (Fig. 1G). The analysis revealed that following AAV treatment, the expression levels of RELB, AMACR, SYK, CCL2, and IL1RN were downregulated, whereas COL9A2, TH, OMP, and GPX3 showed upregulation within the treated group (Fig. 1H). Using the online tool GENEMANIA (Fig. 1I), we inferred the potential functions of RELB, CCL2, and SYK and prioritized their


A
Secreted Signaling
ECM–Receptor
Cell–Cell Contact
Others
Heterodimers
KEGG
Literature
B
Fibroblasts
Epithelial cells
B cells
Monocytes
Stromal cells
T cells, CD8+
C
Fibroblasts
Epithelial cells
B cells
Monocytes
Stromal cells
T cells, CD8+
D
Epithelial cells
B cells
T cells, CD8+
Monocytes
Stromal cells
Fibroblasts
F
MDK
SDC1
SDC2
SDC4
NCL
Expression
Control
AAV-load
ns
*
E
WNT4 – (FZD3+LRP6)
WNT4 – (FZD3+LRP5)
TNF – TNFRSF1B
TNF – TNFRSF1A
PRSS3 – F2RL1
NAMPT – INSR
MIF – (CD74+CXCR4)
MIF – (CD74+CD44)
MDK – SDC4
MDK – SDC2
MDK – PTPRZ1
MDK – NCL
MDK – LRP1
MDK – (ITGA6+ITGB1)
MDK – (ITGA4+ITGB1)
LGALS9 – HAVCR2
LGALS9 – CD45
LGALS9 – CD44
IFNG – (IFNGR1+IFNGR2)
GZMA – F2RL1
GRN – SORT1
GDF15 – TGFBR2
FASL – FAS
CXCL16 – CXCR6
ANXA1 – FPR1
Commun. Prob.
max
min
p–value
0.01 < p < 0.05
p < 0.01

◂**Fig. 3** Study on the interaction between receptors and ligands in cellular biology. **A** Pie chart vividly illustrates the sources of data on cellular interactions. **B** Cellular communication diagram representing interactions between different cell types. Nodes represent cell types, edges represent interactions between cells, and the thickness of the edges represents the frequency of interactions. **C** Weighted cellular communication network diagram. **D** Individual cellular communication diagrams for each cell. **E** Cell communication bubble chart. Colors represent the strength of the interaction between cells based on receptor-ligand binding affinity. **F** The gene expression levels of MDK, SDC1, SDC2, SDC4, and NCL were analyzed in the transcriptome sequencing data

functional analysis. The findings indicated that CCL2 plays a more significant role in the process of AAV vector integration into the brain compared to RELB and SYK.

Differences in cellular characteristics between B lymphocytes and microglia

To further evaluate the integration process of AAV vectors in the brain at the single-cell level, we selected four single-cell sequencing samples from the GEO database. The control group includes control samples of normal brain tissue, while the experimental group consists of brain tissues with integrated AAV9 vectors. Firstly, we conducted gene feature extraction and created violin plots (Fig. 2A). Correlation analysis revealed a high linear relationship of up to 0.96 between sequencing depth and the number of genes (Fig. 2C). Next, we generated a feature variance plot and annotated the top 10 feature genes (Fig. 2B). Finally, in the PCA dimensionality reduction analysis, we selected 1500 feature genes to streamline clustering time. Using the PCA results for clustering, we employed 20 principal components for t-SNE clustering analysis and visually annotated the grouped cells (Fig. 2D, Supplementary Fig. 1). Additionally, we identified differentially expressed genes for each cluster and created corresponding heatmaps to illustrate their individual expression profiles (Fig. 2E).

To investigate differential genes between cell clusters, we performed group visualization by plotting single-cell clustering maps for both the control and experimental groups (Fig. 2F). We observed significant variations in microglia, B lymphocytes, CD8-positive T cells, and fibroblasts. Through statistical analysis, we identified 6828 differential genes related to microglia, 6075 differential genes related to B lymphocytes, 241 differential genes related to dendritic cells, and 106 differential genes related to fibroblasts. Since the number of differential genes in other cell types was minimal and had minimal impact, they were not considered.

An MDK-NCL-dependent immune suppression mechanism exists between cells during the integration process of AAV vectors in the brain

To thoroughly investigate the mechanisms of intercellular interactions, we performed a comprehensive analysis of cellular communication based on ligand-receptor pairs. By integrating data from the KEGG database (comprising 73%) and literature sources (comprising 27%), we elucidated the intricate interactions among fibroblasts, epithelial cells, B lymphocytes, CD8 + T cells, stromal cells, and microglial cells during brain integration mediated by adeno-associated virus (AAV) vectors (refer to Fig. 3A–D). Further examination of the cellular communication network revealed that the ligand-receptor pair consisting of MDK, and neuronal cell adhesion molecule (NCL) plays a pivotal role in this process (Fig. 3E).

The MDK-NCL complex exhibits immunosuppressive properties, effectively modulating immune cell activity and influencing both the magnitude and direction of immune responses. During the integration of AAV vectors into the brain, the MDK-NCL complex may serve as a critical regulator, not only impacting immune system function and response but also potentially indirectly affecting AAV vector integration efficiency and immune tolerance by altering immune cell function and subsequent cellular interactions. To substantiate these observations, we analyzed the expression changes of key genes involved in the MDK-NCL ligand-receptor pair (including MDK, SDC1, SDC2, SDC4, and NCL) using transcriptome data (Fig. 3F). By correlating high-probability regions for the MDK-NCL pair in the bubble plot with gene expression levels from transcriptome data, we further validated the significance of this ligand-receptor pair in specific cellular communication processes. These findings enhance our understanding of the intricate nature of intercellular interactions and provide novel theoretical foundations and potential therapeutic targets for the treatment and prevention of neurological diseases.

Causal inference of AAV receptor integration based on Mendelian randomization

To further elucidate causal hypotheses, we employed Mendel's second law to infer the influence of biological factors on the target outcome by utilizing the impact of randomly distributed genotypes in nature on phenotypes. Initially, through conducting a GWAS study, we identified gene variant single nucleotide polymorphisms (SNPs) that were associated with the previous single-cell results, serving as "instrumental variables (IV)." Subsequently, we used these IVs to infer the impact of biological factors on AAV receptor

exposure	nsnp	method	pval	OR(95% CI)
ARHGAP31	3	MR Egger	0.451	1.374 (0.806 to 2.341)
	3	Weighted median	0.052	1.228 (0.998 to 1.511)
	3	Inverse variance weighted	**0.034**	1.236 (1.016 to 1.504)
	3	Simple mode	0.470	1.158 (0.836 to 1.603)
	3	Weighted mode	0.243	1.218 (0.962 to 1.541)
CMPK2	3	MR Egger	0.699	1.193 (0.607 to 2.343)
	3	Weighted median	0.176	1.297 (0.890 to 1.891)
	3	Inverse variance weighted	**0.038**	1.411 (1.019 to 1.955)
	3	Simple mode	0.464	1.268 (0.755 to 2.129)
	3	Weighted mode	0.386	1.282 (0.823 to 1.997)
DPY19L1	6	MR Egger	0.282	1.101 (0.946 to 1.281)
	6	Weighted median	**0.033**	1.108 (1.008 to 1.219)
	6	Inverse variance weighted	**0.013**	1.113 (1.023 to 1.212)
	6	Simple mode	0.164	1.112 (0.979 to 1.263)
	6	Weighted mode	0.071	1.107 (1.014 to 1.209)
NEIL2	4	MR Egger	0.432	0.900 (0.729 to 1.112)
	4	Weighted median	0.075	0.893 (0.788 to 1.011)
	4	Inverse variance weighted	**0.030**	0.878 (0.782 to 0.987)
	4	Simple mode	0.699	0.953 (0.766 to 1.187)
	4	Weighted mode	0.219	0.897 (0.781 to 1.029)
NPDC1	3	MR Egger	0.659	0.946 (0.787 to 1.137)
	3	Weighted median	0.088	0.897 (0.791 to 1.017)
	3	Inverse variance weighted	**0.046**	0.884 (0.783 to 0.998)
	3	Simple mode	0.350	0.890 (0.736 to 1.075)
	3	Weighted mode	0.245	0.899 (0.791 to 1.022)
NUP85	6	MR Egger	0.415	0.880 (0.669 to 1.159)
	6	Weighted median	**0.037**	0.877 (0.776 to 0.992)
	6	Inverse variance weighted	**0.010**	0.870 (0.784 to 0.967)
	6	Simple mode	0.354	0.907 (0.752 to 1.094)
	6	Weighted mode	0.098	0.884 (0.784 to 0.996)
OSGIN2	4	MR Egger	0.375	0.879 (0.704 to 1.099)
	4	Weighted median	**0.047**	0.874 (0.766 to 0.998)
	4	Inverse variance weighted	**0.024**	0.864 (0.760 to 0.981)
	4	Simple mode	0.352	0.859 (0.655 to 1.126)
	4	Weighted mode	0.139	0.876 (0.770 to 0.997)
SAMD12	5	MR Egger	0.161	1.145 (0.992 to 1.322)
	5	Weighted median	**0.046**	1.118 (1.002 to 1.247)
	5	Inverse variance weighted	**0.037**	1.120 (1.007 to 1.245)
	5	Simple mode	0.114	1.203 (1.005 to 1.439)
	5	Weighted mode	0.135	1.125 (0.994 to 1.272)
VIPR1	6	MR Egger	0.396	0.880 (0.676 to 1.146)
	6	Weighted median	**0.027**	0.813 (0.677 to 0.977)
	6	Inverse variance weighted	**0.043**	0.838 (0.707 to 0.994)
	6	Simple mode	0.165	0.793 (0.600 to 1.049)
	6	Weighted mode	0.092	0.809 (0.663 to 0.988)

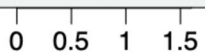


Fig. 4 A Mendelian randomization analysis forest plot for B lymphocytes

exposure	nsnp	method	pval	OR(95% CI)
AK9	4	MR Egger	0.689	1.057 (0.835 to 1.338)
	4	Weighted median	**0.047**	1.127 (1.002 to 1.268)
	4	Inverse variance weighted	**0.042**	1.126 (1.004 to 1.262)
	4	Simple mode	0.126	1.243 (1.015 to 1.521)
	4	Weighted mode	0.173	1.121 (0.988 to 1.271)
DPY19L1	6	MR Egger	0.282	1.101 (0.946 to 1.281)
	6	Weighted median	**0.033**	1.108 (1.009 to 1.218)
	6	Inverse variance weighted	**0.013**	1.113 (1.023 to 1.212)
	6	Simple mode	0.167	1.112 (0.978 to 1.264)
	6	Weighted mode	0.096	1.107 (1.004 to 1.221)
MAP3K5	3	MR Egger	0.857	1.039 (0.747 to 1.444)
	3	Weighted median	0.054	1.143 (0.998 to 1.308)
	3	Inverse variance weighted	**0.047**	1.146 (1.002 to 1.310)
	3	Simple mode	0.170	1.402 (1.024 to 1.919)
	3	Weighted mode	0.238	1.126 (0.979 to 1.296)
POLR1E	3	MR Egger	0.651	0.951 (0.808 to 1.119)
	3	Weighted median	**0.018**	0.881 (0.794 to 0.978)
	3	Inverse variance weighted	**0.012**	0.875 (0.788 to 0.971)
	3	Simple mode	0.215	0.819 (0.658 to 1.019)
	3	Weighted mode	0.167	0.890 (0.800 to 0.991)
SAMD12	5	MR Egger	0.161	1.145 (0.992 to 1.322)
	5	Weighted median	**0.045**	1.118 (1.002 to 1.246)
	5	Inverse variance weighted	**0.037**	1.120 (1.007 to 1.245)
	5	Simple mode	0.129	1.203 (0.995 to 1.453)
	5	Weighted mode	0.104	1.125 (1.008 to 1.255)
SLC22A5	6	MR Egger	0.744	1.030 (0.875 to 1.211)
	6	Weighted median	0.067	1.087 (0.994 to 1.188)
	6	Inverse variance weighted	**0.041**	1.085 (1.003 to 1.174)
	6	Simple mode	0.142	1.121 (0.986 to 1.274)
	6	Weighted mode	0.143	1.086 (0.990 to 1.192)
SOX4	4	MR Egger	0.852	0.904 (0.353 to 2.313)
	4	Weighted median	**0.014**	0.659 (0.472 to 0.919)
	4	Inverse variance weighted	**0.003**	0.652 (0.491 to 0.865)
	4	Simple mode	0.188	0.677 (0.432 to 1.062)
	4	Weighted mode	0.146	0.665 (0.441 to 1.002)
XBP1	10	MR Egger	0.218	0.876 (0.721 to 1.064)
	10	Weighted median	0.071	0.906 (0.814 to 1.008)
	10	Inverse variance weighted	**0.028**	0.904 (0.826 to 0.989)
	10	Simple mode	0.136	0.880 (0.755 to 1.025)
	10	Weighted mode	0.091	0.900 (0.807 to 1.004)

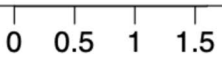
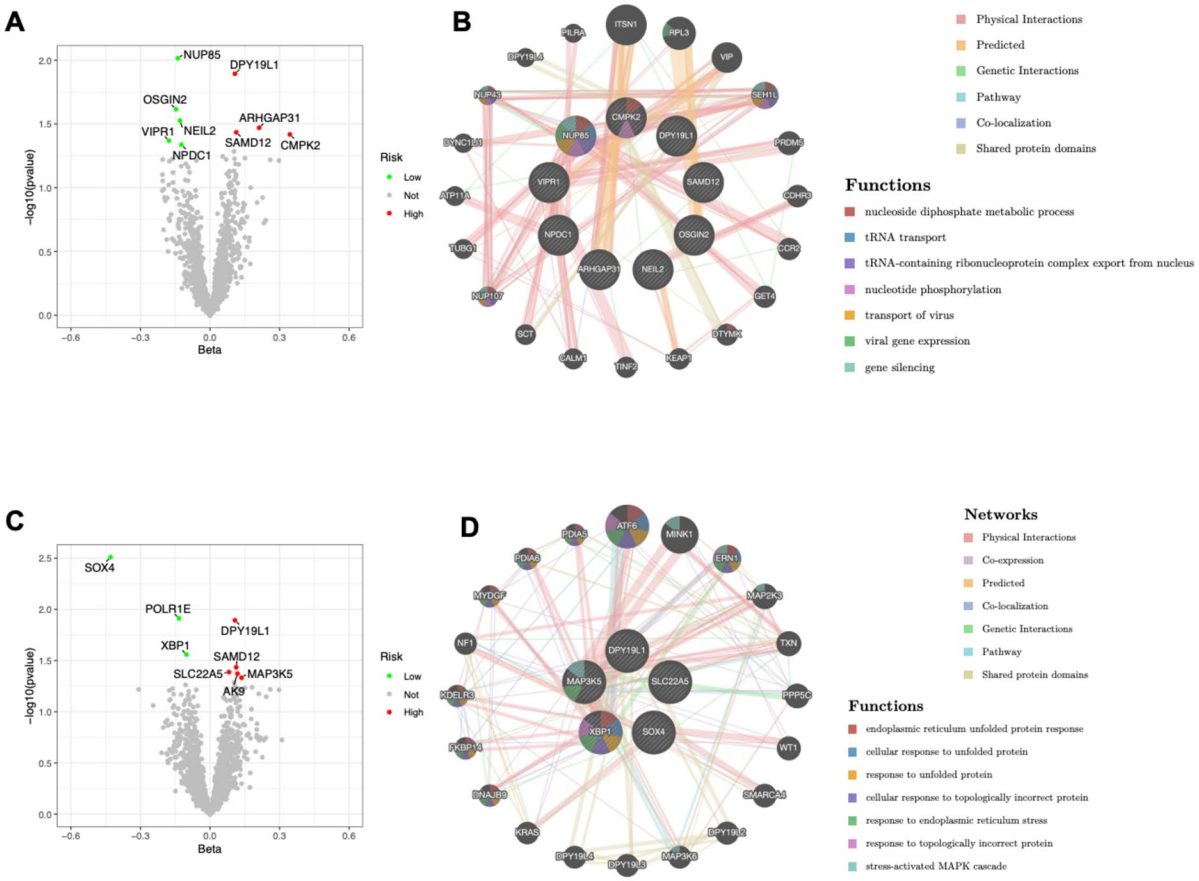


Fig. 5 A Mendelian randomization analysis forest plot for microglia


A
NUP85
DPY19L1
OSGIN2
ARHGAP31
VIPR1
NEIL2
SAMD12
CMPK2
NPDC1
−log10(pvalue)
Beta
Risk
Low
Not
High
B
Networks
Physical Interactions
Predicted
Genetic Interactions
Pathway
Co-localization
Shared protein domains
Functions
nucleoside diphosphate metabolic process
tRNA transport
tRNA-containing ribonucleoprotein complex export from nucleus
nucleotide phosphorylation
transport of virus
viral gene expression
gene silencing
C
SOX4
POLR1E
DPY19L1
XBP1
SAMD12
SLC22A5
MAP3K5
AK9
−log10(pvalue)
Beta
Risk
Low
Not
High
D
Networks
Physical Interactions
Co-expression
Predicted
Co-localization
Genetic Interactions
Pathway
Shared protein domains
Functions
endoplasmic reticulum unfolded protein response
cellular response to unfolded protein
response to unfolded protein
cellular response to topologically incorrect protein
response to endoplasmic reticulum stress
response to topologically incorrect protein
stress-activated MAPK cascade
E
Heterogeneity
Pleiotropy
leave-one-out sensitivity
NUP85
XBP1
MR Method
Inverse variance weighted
MR Egger
MR Test
Weighted median
Weighted mode
Simple mode
SNP effect on NUP85
SNP effect on XBP1
SNP effect on Adenovirus Receptor

◂**Fig. 6** Key gene interaction analysis for Mendelian randomization. **A** A volcano plot of exposure genes for B lymphocytes. Red indicates exposure genes with an odds ratio (OR) greater than 1, representing risk factors. Green indicates exposure genes with an OR less than 1, representing protective factors. **B** Prioritizing exposure genes for B lymphocytes using GeneMANIA for functional analysis. **C** A volcano plot of exposure genes for microglia. Red represents exposure genes with an odds ratio (OR) greater than 1, indicating risk factors. Green represents exposure genes with an OR less than 1, indicating protective factors. **D** Prioritizing exposure genes for microglia using GeneMANIA for functional analysis. **E** Heterogeneity analysis, pleiotropy analysis, and leave-one-out sensitivity analysis of the key exposure gene NUP85 in B lymphocytes and the key exposure gene XBP1 in microglia

integration. Since genes are randomly allocated and not influenced by confounding factors, utilizing genetic variants to study causal relationships can eliminate the influence of confounders on outcomes, thereby enhancing the reliability of causal inference. We employed five statistical methods for Mendelian randomization analysis, including MR Egger, weighted median, inverse variance weighted (IVW), simple mode, and weighted mode, and presented the results in forest plot format.

Based on IVW < 0.05, we identified nine relevant genes (ARHGAP31, CMPK2, DPY19L1, NEIL2, NPDC1, NUP85, OSGIN2, SAMD12, VIPR1) associated with B lymphocytes and AAV receptor integration outcomes (Fig. 4). Eight relevant genes (AK9, DPY19L1, MAP3K5, POLR1E, SAMD12, SLC22A5, SOX4, XBP1) associated with microglia and AAV receptor integration outcomes (Fig. 5). No relevant genes were found in the remaining cell types. Subsequently, we conducted heterogeneity analysis, pleiotropy analysis, and leave-one-out sensitivity analysis on the identified genes for further evaluation (Fig. 6E, Supplementary Fig. 2). The analysis suggests that the aforementioned genes may serve as potential factors influencing AAV receptor integration. Further research can delve into the specific mechanisms through which these genes operate in the process of AAV receptor integration, shedding light on their roles in modulating intracellular signaling and immune responses.

Analysis of the mechanisms and key factors of immune cell-related genes in AAV receptor integration

To further narrow down the scope of key genes, in the B lymphocyte-related genes, utilizing volcano plot visualization, it was revealed that ARHGAP31, CMPK2, DPY19L1, and SAMD12 were positively correlated with AAV receptor integration (OR > 1), while NEIL2, NPDC1, NUP85, OSGIN2, and VIPR1 were negatively correlated with AAV receptor integration (OR < 1) (Fig. 6A). Predicting the functions of these genes and analyzing their interaction networks, it was found that NUP85 played a particularly significant role. NUP85 is involved in crucial biological processes such as virus transport, gene silencing, and nucleoside diphosphate metabolism. Specifically, NUP85 may participate in regulating the process of virus entry into host cells, impacting its infectivity (Fig. 6B).

According to the research findings, it was observed that among the mononuclear cell-related genes, DPY19L1, MAP3K5, SAMD12, and AK9 showed a positive correlation (OR > 1) with the AAV receptor integration process. Additionally, POLR1E, SLC22A5, SOX4, and XBP1 were found to exhibit a negative correlation (OR < 1) with the AAV receptor integration process (Fig. 6C). Further gene function prediction and interaction network analysis revealed the distinct roles these genes play in regulating the AAV receptor integration process. Among them, XBP1 emerged as one of the most significant participants. Acting as a transcription factor, XBP1 is involved in multiple crucial biological processes, including endoplasmic reticulum protein folding, endoplasmic reticulum stress, and the unfolded protein response (Fig. 6D). In summary, the aforementioned results provide insights into the diverse roles and mechanisms of immune cell-related genes in the AAV receptor integration process, aiding in a deeper understanding of the biological events and interrelationships involved in this process.

XBP1 indirectly participates in the regulation of immune suppression in B cells

Based on key findings from transcriptome sequencing data, we identified CCL2 as the target gene for further investigation. Subsequently, through comprehensive analysis of single-cell transcriptome sequencing data, we successfully pinpointed B lymphocytes and monocytes as two critical immune cell types. In B lymphocytes, Mendelian randomization analysis confirmed NUP85 as a pivotal gene, while in monocytes, XBP1 emerged as an important candidate gene. The selection of CCL2, NUP85, and XBP1 for functional prediction and interaction analysis was based on their significant expression changes observed in both transcriptome and single-cell data, as well as their verified roles via Mendelian randomization analysis in specific cell types. To further elucidate the functions of these genes and the biological pathways they are involved in, we conducted KEGG and GO enrichment analyses (Fig. 7A). The KEGG analysis revealed that these genes participate in viral protein-cytokine interactions, chemokine signaling pathways, and cytokine-receptor interactions (Fig. 7B). According to the GO analysis results, these genes are involved in several key biological processes, including myeloid cell differentiation, cell chemotaxis, leukocyte migration, and chemotaxis, which play crucial roles in immune system regulation, infection defense, and tissue

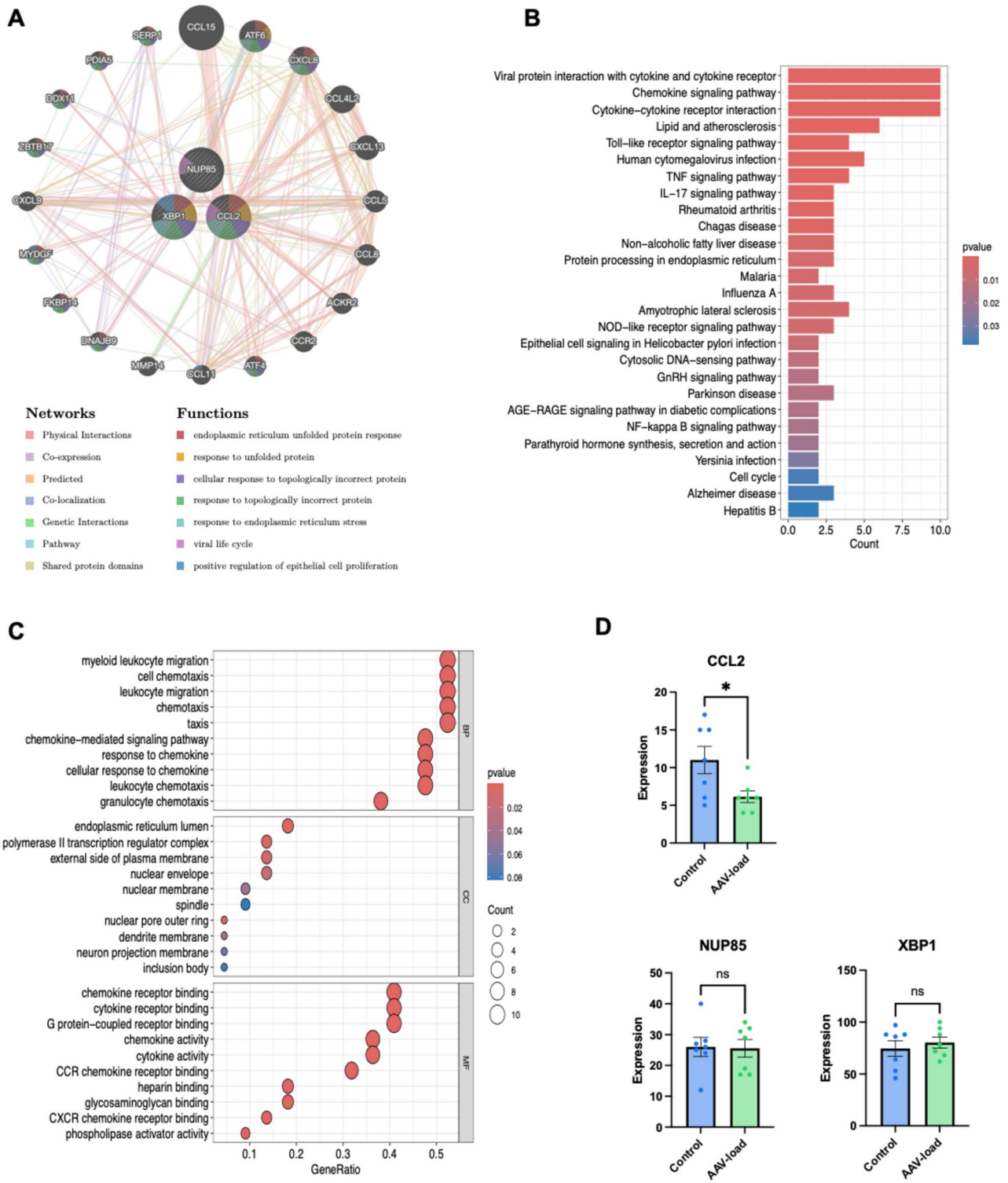


Fig. 7 Functional prediction and interactome analysis of CCL2, NUP85, and XBP1. **A** Prioritizing CCL2, NUP85, and XBP1 using GeneMANIA for functional analysis. **B** Bar graph representing the predicted genes' KEGG pathways following analysis using GeneMANIA. **C** Bubble chart representing the predicted genes' GO biological processes following analysis using GeneMANIA. **D** The gene expression levels of CCL2, NUP85, and XBP1 were analyzed in the transcriptome sequencing data

repair. Additionally, in terms of cellular components, the endoplasmic reticulum lumen emerged as an important structure, while in terms of molecular functions, these genes are implicated in chemokine receptor binding, cytokine receptor binding, and G-protein coupled receptor binding, among others (Fig. 7C).

While XBP1 is predominantly expressed in monocytes, it may indirectly influence B cell function by modulating the production of cytokines such as CCL2. As a chemokine, CCL2 plays a critical role in attracting and activating various immune cells, including B cells. Consequently, alterations in XBP1 expression within monocytes could impact B cell migration, activation, or differentiation through the regulation of CCL2 production. Moreover, XBP1 may also indirectly affect B cell function by influencing interactions between monocytes and B cells, such as antigen presentation and costimulatory signaling. This hypothesis was validated using transcriptome data analysis (Fig. 7D), which demonstrated decreased expression levels of CCL2 and NUP85, alongside increased expression of XBP1 following AAV injection. These changes in gene expression may significantly influence the brain's immune microenvironment. Specifically, reduced expression of CCL2 and NUP85 may collectively contribute to an immunosuppressive state within the brain, mitigating unnecessary inflammatory responses and immune cell activation, thereby protecting the brain from potential damage. Additionally, elevated XBP1 expression may further modulate the brain's immune microenvironment by regulating monocyte and other immune cell functions. Furthermore, increased XBP1 expression may suggest that cells are experiencing endoplasmic reticulum stress post-AAV injection, potentially related to the processing, trafficking, or expression of the AAV vector within cells, which could impact cell survival and proliferation.

Taken together, while there is currently insufficient direct evidence to demonstrate a direct interaction between XBP1 and B cells via the MDK-NCL signaling pathway, we propose a hypothetical mechanism: alterations in XBP1 expression in monocytes may indirectly influence B cell function by modulating pathways such as CCL2. This proposed mechanism offers a novel perspective on MDK-NCL-dependent immune suppression in the brain and suggests potential avenues for future research. Further experimental validation is essential to elucidate the precise nature and mechanisms of these interactions.

Discussion

The process of virus infection in cells is generally similar in nature, involving attachment, entry, uncoating, genome replication, protein synthesis, assembly, and release [36]. In general, the efficiency of infection is closely associated with the specific recognition between virus surface proteins and cell membrane receptor proteins, as well as polysaccharides and other molecules [37]. AAV vectors are widely utilized tools for in vivo gene therapy. However, the human immune system poses significant challenges to the use of this platform. Both innate and adaptive immunity appear to be crucial factors determining the outcomes of gene transfer experiments, due to the complex interactions between vector components and transgenic products. In this study, transcriptomic and scRNA-seq data are employed to investigate the causal relationships between immune factors and the brain immune environment associated with AAV delivery. Key genes and their impact on immune pathways are identified to elucidate the immunogenic response mechanisms triggered by AAV vectors in the brain.

The brain harbors specialized resident immune cells [38], and further investigation is needed to understand their response to AAV vectors. To investigate the transcriptional landscape following AAV injection into the brain, we initially analyzed mRNA databases of normal mouse brains and mouse brains integrated with AAV. A total of 11,980 genes were examined, and under conditions of fdrFilter = 0.05 and logFCfilter = 1, 73 differentially expressed genes were identified. Through GO, KEGG, and differential gene-based protein interaction network analysis, we discovered that, in the process of AAV vector integration in the brain, ccl2 played a crucial role compared to RELB and SYK. Subsequently, for a more in-depth assessment at the single-cell level, we selected four scRNA-seq samples from the GEO database. During the integration process of AAV vectors in the brain, B lymphocytes and microglia exhibited distinct cell characteristics, with an observed MDK-NCL-dependent immunosuppression.

To further clarify the causal hypothesis, we inferred the impact of scRNA-seq results on AAV infection based on MR analysis. Five statistical methods were employed for MR analysis, and based on IVW < 0.05, nine relevant genes (ARHGAP31, CMPK2, DPY19L1, NEIL2, NPDC1, NUP85, OSGIN2, SAMD12, VIPR1) were identified in B lymphocytes and AAV receptor integration outcomes, while eight relevant genes (AK9, DPY19L1, MAP3K5, POLR1E, SAMD12, SLC22A5, SOX4, XBP1) were identified in microglia and AAV receptor integration outcomes. Heterogeneity analysis, pleiotropy analysis, and leave-one-out sensitivity analysis conducted on the identified relevant genes suggest that these genes may serve as potential factors influencing AAV receptor integration.

The field of gene therapy is currently experiencing an exciting period, with several clinical trials demonstrating long-term efficacy. Candidate gene therapy drugs have also progressed to late-stage clinical development and market approval phases [39]. Diseases that were previously

incurable or had suboptimal treatment options are now being effectively cured, bringing unprecedented results for patients. However, challenges such as the immunogenicity of AAV vectors and emerging complexities persist. Our research innovatively employs a variety of analytical approaches to delve into the AAV immune environment within the brain, aiming to preliminarily uncover the immunosuppressive mechanisms induced by AAV vectors in the brain.

This study elucidates the B cell immune suppression induced by AAV vectors in the brain through a comprehensive analysis of scRNA-Seq and RNA-Seq. Furthermore, it posits that XBP1 may play an indirect role in mediating MDK-NCL-dependent B cell immune suppression. It is evident that this study has certain limitations. Currently, in AAV trials, circulating T cells are frequently utilized for immune monitoring, serving as a crucial surrogate marker to track vector-related immune responses [40]. Nevertheless, analyzing in situ immune responses may contribute to a better understanding of the mechanisms underlying transgene expression loss or maintenance. Preexisting B cell immunity poses a significant hurdle, as targeting therapies are more complex than merely inhibiting immune responses [41]. From a technical standpoint, there exists a significant paucity of suitable enrollment samples within the current databases. The transcriptomic sequencing data included in this study lack fundamental information regarding serotypes; thus, we supplemented this gap with data derived from single-cell sequencing. It is conceivable that the immune response initially targets the AAV capsid, followed subsequently by the transgene product. Consequently, the temporal aspect is of paramount importance, as the initiation of transgene expression necessitates a certain period. Additionally, the choice of promoter is critical, as it dictates the expression profile. Building upon this premise, future investigations should employ intervention experiments to validate the hypothesis that XBP1 indirectly contributes to MDK-NCL-dependent B cell immune suppression.

Conclusion

In summary, the intricate interplay between innate and adaptive immunity is a pivotal determinant of the outcomes of in vivo gene therapy utilizing AAV vectors. This study underscores the potential role of the XBP1 protein in mediating B-cell immunosuppression in the brain via the MDK-NCL ligand-receptor pair and associated genes, as evidenced by RNA-seq and scRNA-seq data. Moreover, through the application of MR analysis, this research lays the foundation for investigating regulatory factors and pathways within the immune modulation network, along with their potential benefits in immunotherapeutic interventions.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12026-025-09609-6.

Acknowledgements Acknowledgments for the support from the Shanghai Ninth People's Hospital and the School of Medicine at Shanghai Jiao Tong University.

Author contribution SY.W. was engaged in conceptualizing the experimental design, executing the experiments, conducting statistical analyses, and composing the manuscript. L.X. was instrumental in data analysis, result interpretation, and data visualization. X.L., YX.W., YT.Z., YB.L., JY.S., and TT.J. managed data organization, formatting, and validation. WY.S. and ZY.W. undertook tasks such as article proofreading, securing funding, and overseeing project administration.

Funding This research was financially supported by the collaborative initiative "Joint Efforts and Expanded Utilization of Key Technologies in Diagnostics and Treatment of Brainstem Implants for Artificial Auditory Perception (SHDC12020105)" and the National Natural Science Foundation of China (82201270). The funding sources were not involved in the study's design, data collection and analysis, or the decision to submit the work for publication.

Data availability The code produced or employed in the investigation is available on GitHub at the following link: https://github.com/codeconnoisseur5/Immune-Suppression-in-the-AAV-Loaded-Brain.git.

Declarations

Ethics approval and consent to participate This study is based on publicly available databases and therefore does not applicable ethical approval.

Competing interests The authors declare no competing interests.

References

1. Wu Z, Yang H, Colosi P. Effect of genome size on AAV vector packaging. Mol Ther. 2010;18(1):80–6.
2. Lovric J, et al. Terminal differentiation of cardiac and skeletal myocytes induces permissivity to AAV transduction by

relieving inhibition imposed by DNA damage response proteins. Mol Ther. 2012;20(11):2087–97.
3. Bennett J, et al. Safety and durability of effect of contralateral-eye administration of AAV2 gene therapy in patients with childhood-onset blindness caused by RPE65 mutations: a follow-on phase 1 trial. Lancet. 2016;388(10045):661–72.
4. Buchlis G, et al. Factor IX expression in skeletal muscle of a severe hemophilia B patient 10 years after AAV-mediated gene transfer. Blood. 2012;119(13):3038–41.
5. Mueller C, et al. 5 year expression and neutrophil defect repair after gene therapy in alpha-1 antitrypsin deficiency. Mol Ther. 2017;25(6):1387–94.
6. Nathwani AC, et al. Long-term safety and efficacy of factor IX gene therapy in hemophilia B. N Engl J Med. 2014;371(21):1994–2004.
7. Costa Verdera, H., K. Kuranda, and F. Mingozzi, AAV vector immunogenicity in humans: a long journey to successful gene transfer. Mol Ther, 2020. **28**(3): p. 723–746.
8. Balakrishnan B, Jayandharan GR. Basic biology of adeno-associated virus (AAV) vectors used in gene therapy. Curr Gene Ther. 2014;14(2):86–100.
9. Nonnenmacher M, Weber T. Intracellular transport of recombinant adeno-associated virus vectors. Gene Ther. 2012;19(6):649–58.
10. Ayuso E, Mingozzi F, Bosch F. Production, purification and characterization of adeno-associated vectors. Curr Gene Ther. 2010;10(6):423–36.
11. Matsushita T, et al. Adeno-associated virus vectors can be efficiently produced without helper virus. Gene Ther. 1998;5(7):938–45.
12. Herzog RW, et al. Regulatory T cells and TLR9 activation shape antibody formation to a secreted transgene product in AAV muscle gene transfer. Cell Immunol. 2019;342: 103682.
13. Ashley SN, et al. TLR9 signaling mediates adaptive immunity following systemic AAV gene therapy. Cell Immunol. 2019;346: 103997.
14. Hoffman BE, et al. Nonredundant roles of IL-10 and TGF-β in suppression of immune responses to hepatic AAV-factor IX gene transfer. Mol Ther. 2011;19(7):1263–72.
15. Piguet F, Alves S, Cartier N. Clinical gene therapy for neurodegenerative diseases: past, present, and future. Hum Gene Ther. 2017;28(11):988–1003.
16. Brown GC, Neher JJ. Microglial phagocytosis of live neurons. Nat Rev Neurosci. 2014;15(4):209–16.
17. Tufail Y, et al. Phosphatidylserine exposure controls viral innate immune responses by microglia. Neuron. 2017;93(3):574-586.e8.
18. Kaushansky N, et al. Modulation of MS-like disease by a multi epitope protein is mediated by induction of CD11c(+)CD11b(+) Gr1(+) myeloid-derived dendritic cells. J Neuroimmunol. 2019;333: 476953.
19. Barrett, T., et al., NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res, 2013. **41**(Database issue): p. D991–5.
20. Ritchie ME, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7): e47.
21. Zhang X, et al. Single-cell RNA and transcriptome sequencing profiles identify immune-associated key genes in the development of diabetic kidney disease. Front Immunol. 2023;14:1030198.
22. Wu T, et al clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb), 2021; **2**(3): p. 100141.
23. Satija R, et al. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol. 2015;33(5):495–502.
24. Aran D, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol. 2019;20(2):163–72.
25. Trapnell C, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 2014;32(4):381–6.
26. Fu Y, et al. Harmony loss for unbalanced prediction. IEEE J Biomed Health Inform. 2022;26(2):828–39.
27. Principal component analysis improves reliability of epigenetic aging biomarkers. Nat Aging, 2022; **2**(7): p. 578–579.
28. Jin S, Plikus MV, and Nie Q CellChat for systematic analysis of cell-cell communication from single-cell transcriptomics. Nat Protoc, 2024;
29. Visscher PM, et al. Five years of GWAS discovery. Am J Hum Genet. 2012;90(1):7–24.
30. Birney E, Mendelian randomization. Cold Spring Harb Perspect Med, 2022; **12**(4).
31. Franz M, et al. GeneMANIA update 2018. Nucleic Acids Res. 2018;46(W1):W60-w64.
32. Wang Q, et al. Exploring epigenomic datasets by ChIPseeker. Curr Protoc. 2022;2(10): e585.
33. Gu Z, et al. circlize Implements and enhances circular visualization in R. Bioinformatics. 2014;30(19):2811–2.
34. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32(18):2847–9.
35. Gu Z Complex heatmap visualization. Imeta 2022; **1**(3):e 43
36. Carty M, Guy C, Bowie AG. Detection of viral infections by innate immunity. Biochem Pharmacol. 2021;183: 114316.
37. Jang A, Lehtinen MK. Experimental approaches for manipulating choroid plexus epithelial cells. Fluids Barriers CNS. 2022;19(1):36.
38. Castellani G, et al Transforming the understanding of brain immunity. Science, 2023; **380**(6640): eabo7649.
39. Mendell JR, et al. Current clinical applications of in vivo gene therapy with AAVs. Mol Ther. 2021;29(2):464–88.
40. Kang L, et al. AAV vectors applied to the treatment of CNS disorders: Clinical status and challenges. J Control Release. 2023;355:458–73.
41. Harkins AL, Ambegaokar PP, Keeler AM. Immune responses to central nervous system directed adeno-associated virus gene therapy: does direct CNS delivery make a difference? Neurotherapeutics. 2024;21(4): e00435.