UNRAVELING T-CELL AND ENDOTHELIAL CELL INTERACTIONS THROUGH COMPUTATIONAL ANALYSES

by

Krithika Umesh

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Master of Science in Bioinformatics and Computational Biology

Fall 2023

UNRAVELING T-CELL AND ENDOTHELIAL CELL INTERACTIONS THROUGH COMPUTATIONAL ANALYSES

by

Krithika Umesh

Approved: _____
Jason P Gleghorn, Ph.D
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____
Cathy H Wu, Ph.D
Chair of the Department of Bioinformatics and Computational Biology

Approved: _____
Levi T Thompson, Ph.D
Dean of the College of Engineering

Approved: _____
Louis F. Rossi, Ph.D.
Vice Provost for Graduate and Professional Education and
Dean of the Graduate College

ACKNOWLEDGMENTS

First and foremost, I extend my deepest gratitude to my advisor, Dr. Jason Gleghorn, whose unwavering encouragement, support, and guidance have been instrumental in the success of this research project. Without his invaluable mentorship, this endeavor would not have been possible. I would also like to express my appreciation to my committee members, Dr. Shawn Polson, and Dr. Ryan Zurakowski, for their insightful contributions during the project.

I am indebted to all the members of our laboratory for their fresh perspectives and valuable suggestions, which proved to be pivotal in overcoming minor obstacles. Special thanks go to Amelia Harrison for her timely advice and assistance throughout the course of this project.

Last but certainly not least, I wish to acknowledge my husband, Abhiram Krishnakumar, for his unwavering love and support, without which this project would not have come to fruition. My heartfelt thanks also go out to my parents, brother, and in-laws for their constant love and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

A continuous challenge in drug delivery is to achieve tissue and organ-specific targeting, in particular to places like the lymph node and brain, where many difficult-to-treat diseases reside. This challenge is due to organ-specific endothelial barriers that possess unique cell membrane proteins and serve as restrictive barriers to molecules, particles, and cells moving from the systemic circulation into the tissue parenchyma. However, immune cells navigate these barriers regularly, particularly when patrolling lymph nodes and responding to brain cancer. As such, we sought to uncover unique interactions between local endothelial cells and immune counterparts to enable an analysis pipeline of tissue-specific receptor binding pairs that may ultimately advance the development of therapeutics finely tuned for precise tissue-specific targeting. This pipeline has generated preliminary data that uncovers distinct cell-cell interactions between local endothelium and T-lymphocytes across two key endothelial barriers i.e., lymph node and the blood-brain barrier, from publicly available single-cell RNA sequencing datasets through computational analyses. The differential expression analysis of HEV vs brain endothelial cells and naïve vs activated T lymphocytes led to the identification of various cell-type specific markers. The co-expression analysis revealed expression patterns of cell surface markers and identification of modules of co-expressing genes. The cell-cell communication analysis helped in identifying binding partners of markers expressed on HEV with that of naïve T lymphocytes and markers expressed on brain endothelial cells with that of activated T lymphocytes. This approach offers the opportunity to shed light on the signaling mechanisms involved in these interactions.

Chapter 1

INTRODUCTION

1.1 Immune system

The immune system is a complex network of cells and proteins that defend the body against a wide variety of pathogens. The immune response to a foreign body can be categorized into two primary types: innate and adaptive immunity.

Innate immunity is the first line of defense in the body. It acts by generating rapid, non-specific immune response to pathogens. Within this system, various types of white blood cells, known as leukocytes, play a crucial role. They either directly combat invaders or produce cytokines and collaborate with other leukocytes to neutralize threats before active infection takes hold. When necessary, innate immunity serves as the trigger for the adaptive immune response[1].

Adaptive immune response is a specialized mechanism involving recognition of specific antigens and producing antibodies or immune cells tailored for to target those antigens. It functions by producing memory cells after its first encounter with an antigen thereby enhancing the specificity and robustness of the immune response[2]. It can in turn be categorized into humoral and cellular immunity.

Humoral response primarily involves B cells or B lymphocytes that are produced in bone marrow. These cells produce antibodies to neutralize pathogens or label them for destruction by other immune cells.

Cellular immune response is mediated by T cells or T lymphocytes that are produced in thymus. T cells in thymus undergo positive selection to recognize major histocompatibility complex (MHC) molecules and are assigned to either CD4+ T helper cells or CD8+ T cytotoxic cells. Helper T cells assist other immune cells by releasing

signaling molecules called cytokines that activate B cells and cytotoxic T cells. Cytotoxic T cells are responsible for eliminating infected cells and tumor cells. T cells that exit the thymus form a pool of naïve cells that circulate in the peripheral lymphoid tissues[3]. Upon encountering an antigen, they become activated, undergo clonal expansion, and acquire effector functions to eliminate the pathogen. These activated cells are referred to as effector cells, with most of them having a short lifespan, typically lasting only a few weeks. A small population of these cells survive as memory cells in the body[4].

T cells that are autoreactive i.e., they react to the self are eliminated by negative selection. This mechanism, called central tolerance, ensures that the immune system does not attack self-peptides.

1.2 Endothelial barriers

Understanding endothelium, which forms the inner lining of blood and lymphatic vessels throughout the body, is essential for gaining insights into the immune response. One of its crucial functions is to act as a selective barrier between the blood/lymph and the surrounding tissues[5,6,7,8]. It accomplishes this by regulating the passage of small molecules, nutrients, and immune cells across the vasculature, thus maintaining tissue homeostasis. Endothelium also plays a significant role in immune surveillance and response by facilitating the trafficking of immune cells to sites of infection and inflammation. The structure, function, and gene expression of the endothelium exhibit remarkable diversity based on its tissue origin. Notably, the endothelial barriers found at lymph nodes and the blood-brain barrier (BBB) are of particular interest.

1.2.1 Lymph Node

Lymph nodes are small bean-shaped secondary lymphoid organs that are distributed throughout the body and serve as critical sites of lymphocyte activation. The structure of a lymph node (Fig.1-1) can be broadly divided into five distinct zones[9]:

a) **Capsule**: Each lymph node is surrounded by a dense fibrous capsule which provides shape and structural support to the node.
b) **Subcapsular sinus**: This region is present beneath the capsule and facilitates the transportation of incoming lymphatic fluid from afferent vessels.
c) **Cortex**: This region is present beneath the subcapsular sinus and mainly consists of B-cells arranged into follicles that are a site for proliferation of naïve B cells after encountering antigens[10].
d) **Paracortex**: This layer is also called T-cell layer as it is a homing region for naïve T-lymphocytes. Naïve T cells are presented with antigens in this zone where they get activated and subsequently undergo differentiation into effector cells[11].
e) **Medulla**: This is the innermost layer of the lymph node and is composed of large blood vessels, sinuses, and medullary cords. It drains the lymph out of the lymph node into efferent lymphatic vessels.

Blood vessels enter and exit the node through the hilum, a region on the concave side of the lymph node.

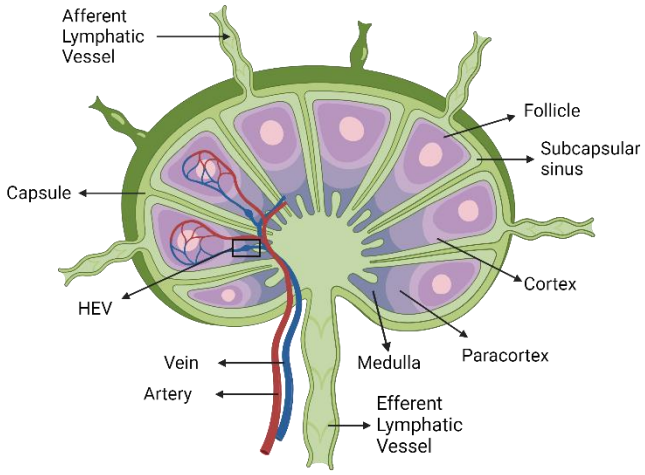


Figure 1-1: Structure of a lymph node
(Created with BioRender.com)

Lymph nodes contain specialized structures known as High Endothelial Venules (HEVs), which are crucial for effective immune surveillance against pathogens. HEVs create an interconnected network of post-capillary venules that seamlessly integrate into the regular bloodstream circulation in lymph nodes[12]. HEVs are lined with plump cuboidal endothelial cells that are surrounded by a basal lamina and fibroblast reticular cells (Fig.1-2). HEVs are responsible for recruiting naïve T cells from bloodstream into the lymph node[13]. Naïve T-cells enter HEVs through a multistep adhesion cascade which involves rolling, adhesion, crawling and transmigration.

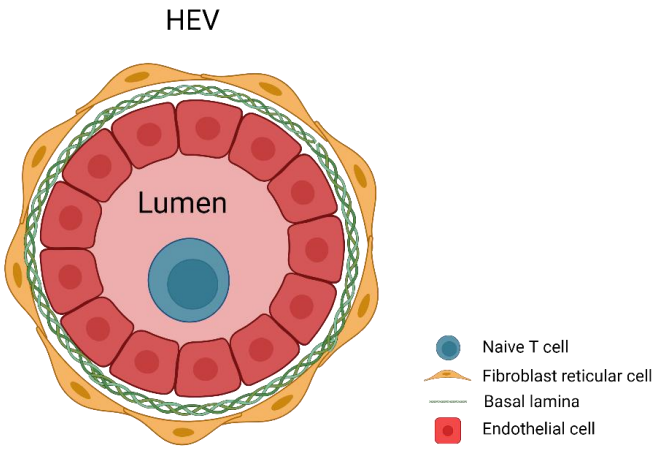


Figure 1-2: Anatomical structure of High Endothelial Venules (HEV)
(Created with BioRender.com)

1.2.2 Blood Brain Barrier

The endothelial cells of the brain are flattened, lack fenestration, and are closely linked to each other by tight junctions. They are supported by pericytes and astrocytes creating a highly selective physical barrier called the blood-brain barrier (BBB) as shown in Fig.1-3. The tight junctions effectively restrict the movement of pathogens, large molecules, and diffusion of solutes. Pericytes, which envelop the endothelial cells in venules, arterioles, and capillaries, not only provide structural support to the vessels but also play a crucial role in angiogenesis and the formation of tight junctions[14].

Meanwhile, astrocytes are indispensable for the maturation and maintenance of the BBB[15].

The BBB exhibits high efflux transporter activity. Efflux transporters like P-glycoprotein plays a pivotal role in restricting the entry of certain drugs and toxins into the brain by pumping them back into bloodstream contributing to the low permeability of the barrier[16]. Interestingly, it has been observed that during episodes of inflammation, activated lymphocytes can breach the BBB[17]. The mechanism of transmigration of activated T-cells mirrors that of naïve T-cell migration across high endothelial venules (HEVs).

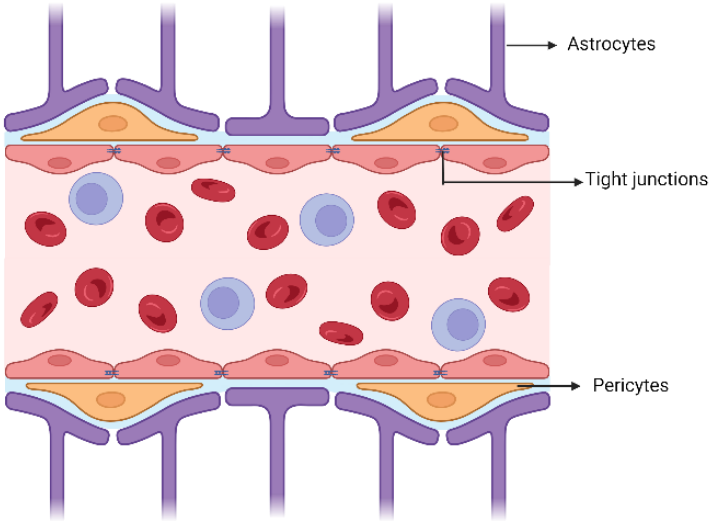


Figure 1-3: Structure of blood brain barrier showing endothelial cells held by tight junctions supported by astrocytes and pericytes
(Created with BioRender.com)

1.3 Extravasation of lymphocytes from lumen to tissue

Whereas there are multiple in vitro approaches to generate vascular models[18,19,20,21,22], unfortunately, due to the complexity of the lymph node HEV barrier and the endothelial blood-brain barrier there are no sufficient *in vitro* models that have been developed that sufficiently recapitulates *in vivo* function. Thus, our main understanding of lymphocyte

extravasation arises from some in vitro models and a large number of in vivo studies. Lymphocyte extravasation from the lumens of blood or lymphatic vessels consists of a multistep adhesion cascade encompassing the following stages:

Rolling and Tethering: The initial step in the extravasation process relies on the weak binding interactions between selectins expressed on lymphocytes and glycoproteins present on endothelial cells. In the context of HEVs during homeostasis, the rolling step commences with the binding of L-selectin expressed on naïve T cells with Peripheral Node Addressins (PNAds) expressed on HEVs. In case of inflamed BBB, the tethering and rolling of T cells are initiated by the interaction of P-selectin expressed on activated T cells with P-selectin glycoprotein-1 (Psgl-1) of endothelial cells[23,24].

Adhesion/Sticking: Following the initial weak interactions induced by the rolling of lymphocytes, chemokine receptors such as Ccr7 and Cxcr4 on naive T cells become activated. They subsequently bind to their respective ligands, including Ccl21, Ccl19, and Cxcl12, which are present on the surface of HEVs[11,25]. Notably, the chemokines Ccl19 and Ccl21, typically expressed in HEVs during homeostasis, have also been identified in the context of inflamed BBB[26]. The presence of Ccr7 is indicative of the recruitment of activated lymphocytes to the BBB. Moreover, another study mentions the binding of the Cxcr4 receptor on leukocytes to the chemokine Cxcl12 within the brain[23]. Firm adhesion is subsequently mediated by the activation of integrin receptors on the lymphocytes like Leukocyte Function-associated Antigen-1 (LFA-1), macrophage 1-antigen (Mac-1) and Very Late Antigen-4 (VLA-4). These integrins bind to ligands such as Icam-1 or Vcam-1, resulting in the firm attachment of T cells to the endothelial cell wall.

Crawling: After firm adhesion to the endothelial cells, lymphocytes exhibit a crawling behavior on the endothelial surface to find a suitable site for transmigration. This step is mediated by cell adhesion molecules like Icam-1 and Icam-2 in BBB. It has been observed that Icam-2 is constitutively expressed in both inflamed and non-inflamed BBB[27]. However, an interesting observation from a separate study indicates that an elevated expression of Icam-1 reduces the crawling distances of leukocytes[28].

Additionally, in lymph nodes, it has been noted that the downregulation of LFA-1 is a prerequisite for T cell crawling, while conversely, an upregulation in LFA-1 expression is essential for the firm arrest of lymphocytes[29].

Transmigration: Crawling leukocytes have been observed to exit into tissues through a process called diapedesis, utilizing two distinct routes:

1. **Paracellular diapedesis**: This route involves the migration of leukocytes through the junctions between the endothelial cells. It relies on the engagement of various adhesion molecules including Icam-1, Vcam-1, junctional adhesion molecules (Jam), Pecam-1, Cd99L, VE-cadherin, and the endothelial lateral border recycling compartment (LBRC)[25,31]. Paracellular route appears to be more prevalent in HEV[30], although transcellular migration has also been observed.

2. **Transcellular diapedesis**: This mechanism is still an area of active research. It commences with leukocytes scanning the endothelial surface to identify areas of least resistance, leading to the development of podosome-like protrusions[31]. These protrusions gradually become invasive and contribute to the formation of cup-like structures called caveolae. This process is mediated by caveolin 1 (Cav1) and occurs away from endothelial junctions. Most of the studies on this mechanism have been conducted on neutrophils. The molecular interactions involved share similarities with the paracellular route involving molecules like Icam, Jam, Cd99 and Pecam1 among other GTPases[32]. Several factors seem to favor this route like high expression of Cav1[33], level of leukocyte activation[31] and a high expression of Icam-1 on endothelium[28]. This mechanism is also more common in BBB[34] where the endothelial junctions are tight and when Pecam-1 levels are low[35] or when cells are unable to reach a junction[36]. Interestingly, some studies have indicated that cells may adopt this passage even when tight junctions in the BBB are compromised[30].

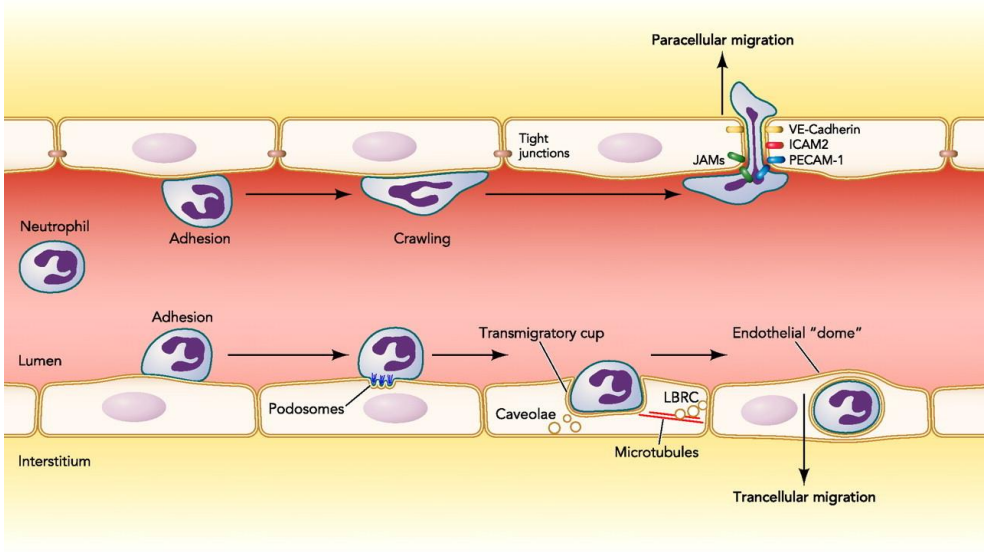


Figure 1-4: Diagram showing the paracellular and transcellular diapedesis
Figure adapted from Schmidt et.al[37]

1.4 Project objective

The lymphocyte extravasation process is being explored to fill the knowledge gaps in this field. For example, in the context of neutrophils adhesion, Cd2ap, an endothelial actin-binding protein, was reported to negatively regulate Icam-1 clustering. Its absence resulted in increased Icam-1 clustering dynamics, leading to enhanced neutrophil adhesion but reduced crawling behavior[38]. This suggests that a similar molecule might be involved in lymphocyte adhesion.

Additionally, Plvap, a cell-type specific marker for HEV is reported to participate in transcellular extravasation along with other proteins[39]. This protein is a key structural element for the formation of diaphragms that cover the openings of fenestrae, stomata. Plvap protein is not expressed in BBB which suggests that there are different proteins involved in transcellular diapedesis in these endothelial barriers. Furthermore, the role of the basement membrane in diapedesis, particularly in lymphocyte transmigration,

remains incompletely understood, with reported involvement of laminins and collagen in neutrophil transmigration but limited understanding in the case of lymphocytes.

The objective of this study is to employ computational analyses to elucidate the complex unique interactions between ligands and receptors of naïve T-cells with HEVs and activated T-cells with BBB, striving for a deep comprehension of their intricate dynamics. Unraveling the specific pairs of cell-cell binding can yield valuable insights into distinctive receptors that facilitate precise cell localization within these sites. Consequently, the investigation of critical barriers like HEV and BBB, along with the identification of unique receptor partners, holds the potential to inform the development of proteins or peptide therapeutics capable of efficiently traversing these barriers.

Chapter 2

LEVERAGING PUBLIC DATASETS FOR INITIAL COMPUTATIONAL ANALYSES

This emergence of single cell RNA sequencing (scRNA seq) has revolutionized transcriptomics enabling us to investigate the gene expression of a single cell providing deeper insights into the transcriptional variations between cells within a sample[40] making it ideal for characterizing cell-surface receptors and ligands across cell states and cell types and in disease[41,42]. This technology has led to discovery of novel cell types[43,44], identification of novel targets for drug delivery[45] and enabled profound understanding of variations within a cell population such as gene interaction, allelic expression, and gene co-expression patterns[46].

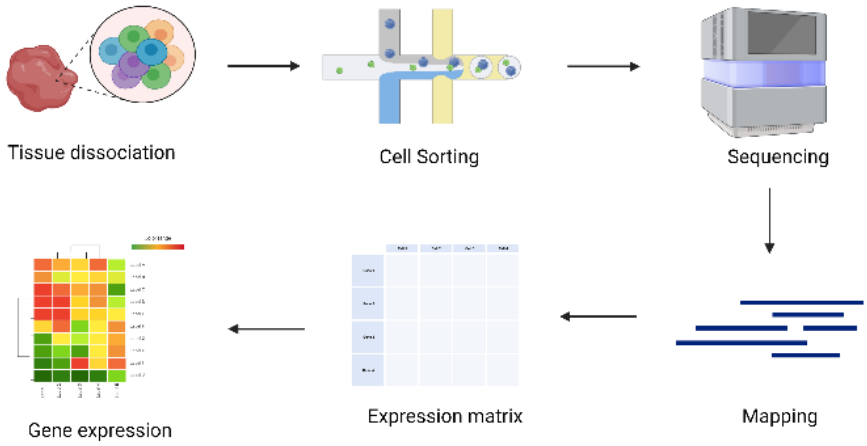


Figure 2-1: Single-cell RNA sequencing workflow
(Created with BioRender.com)

The scRNA-seq workflow involves isolating single cells using techniques like fluorescence-activated cell sorting (FACS), converting RNA to cDNA, amplifying the cDNA library, followed by high-throughput sequencing. This is followed by mapping the reads to the reference genome and generating a count matrix for analyses[47] (Fig.2-1). The key step that distinguishes scRNA seq from bulk RNA seq is the barcoding of transcriptome of each cell. One of the main drawbacks of this technology is the low gene retrieval yield i.e., only 1-5% transcripts per cell can be attributed to highly expressed genes leading to uncertainty in the observed results. The dropout phenomenon introduces notable variability between cells and results in a reduced signal-to-noise ratio (SNR)[48]. Various microfluidic based techniques are being developed and optimized to enhance the efficiency of capturing viable cells[49]. It is also important to note that this method is expensive and is a labor and time intensive process.

Concurrently, advancements in sequencing technology have spurred the development of a wide range of computational tools for the analysis of extensive single cell transcriptomic data to address its unique challenges like drop out effects and high technical noise. New methodologies have emerged to distinguish and quantify technical noise in transcriptional data[50,51]. Thus far, the field has seen the development of over 1600 tools for single-cell RNA sequencing data analysis, with R and Python being the primary programming languages of choice, representing approximately 55% and 43% of these tools, respectively[52].

This has led to an increasing number of scRNA seq datasets available on public repositories like NCBI GEO, allows for the integration of diverse datasets spanning cell types and experimental conditions and downstream analysis prior to performing in vivo or in vitro experiments. This approach not only proves to be cost and time-effective but also facilitates hypothesis validation, experimental condition optimization, and fosters collaborative opportunities among researchers. Considering these advantages, for our initial assessments, we opted to leverage the extensive pool of publicly available datasets.

Chapter 3

DATA EXTRACTION

The datasets utilized for our analyses were procured from the NCBI Gene Expression Omnibus (GEO) public data repository. GEO serves as an extensive repository for various forms of high-throughput functional genomics data, encompassing microarray, next-generation sequencing, and more, all contributed by the global research community. While we also explored databases such as DDBJ[53], a bioinformatic data repository in Japan, and the Chan Zuckerberg Initiative's Cell X Gene database for single cells[54], it's noteworthy that all resulting datasets retrieved during our search process possessed a GEO accession number. Consequently, we decided to primarily search and draw data from the GEO repository.

3.1 Method

In February 2023, we systematically searched the GEO repository using various keyword combinations. The datasets for the analyses originated from in vivo experiments conducted on mice. The search criteria provided below represent the specific terms we employed within the NCBI portal to retrieve results when used in different combinations.

<u>Search details for extracting HEV datasets</u>

- ((high[All Fields] AND ("endothelium"[MeSH Terms] OR endothelial[All Fields])) AND (("single person"[MeSH Terms] OR single[All Fields]) AND ("cells"[MeSH Terms] OR cell[All Fields]))) AND "Mus musculus"[porgn]
- ((high[All Fields] AND ("endothelium"[MeSH Terms] OR endothelial[All Fields])) AND scRNA seq[All Fields]) AND "Mus musculus"[porgn]
- ((high[All Fields] AND ("endothelium"[MeSH Terms] OR endothelial[All Fields]) AND ("venules"[MeSH Terms] OR venules[All Fields])) AND scRNA seq[All Fields]) AND "Mus musculus"[porgn]
- (hev[All Fields] AND scRNA seq[All Fields]) AND "Mus musculus"[porgn]

- (hev[All Fields] AND (("single person"[MeSH Terms] OR single[All Fields]) AND ("cells"[MeSH Terms] OR cell[All Fields]))) AND "Mus musculus"[porgn]
- ("lymph nodes"[MeSH Terms] OR lymph node[All Fields]) AND hev[All Fields] AND ("rna, small cytoplasmic"[MeSH Terms] OR scRNA[All Fields])

Search details for extracting BBB datasets

- (("blood-brain barrier"[MeSH Terms] OR blood brain barrier[All Fields]) AND scRNA seq[All Fields]) AND "Mus musculus"[porgn]
- (("blood-brain barrier"[MeSH Terms] OR blood brain barrier[All Fields]) AND (("single person"[MeSH Terms] OR single[All Fields]) AND ("cells"[MeSH Terms] OR cell[All Fields]) AND ("rna"[MeSH Terms] OR RNA[All Fields]))) AND "Mus musculus"[porgn]
- ((("brain"[MeSH Terms] OR brain[All Fields]) AND ("endothelium"[MeSH Terms] OR endothelial[All Fields])) AND (("single person"[MeSH Terms] OR single[All Fields]) AND ("cells"[MeSH Terms] OR cell[All Fields]) AND ("rna"[MeSH Terms] OR RNA[All Fields]))) AND "Mus musculus"[porgn]

Search details for extracting Naïve T cell datasets

- (naive[All Fields] AND ("t-lymphocytes"[MeSH Terms] OR T cell[All Fields]) AND scRNA seq[All Fields]) AND "Mus musculus"[porgn]
- ((naive[All Fields] AND ("t-lymphocytes"[MeSH Terms] OR T cell[All Fields])) AND (("single person"[MeSH Terms] OR single[All Fields]) AND ("cells"[MeSH Terms] OR cell[All Fields]) AND ("rna"[MeSH Terms] OR RNA[All Fields]))) AND "Mus musculus"[porgn]
- (("cd8-positive t-lymphocytes"[MeSH Terms] OR CD8 t cell[All Fields]) AND naive[All Fields] AND (("single person"[MeSH Terms] OR single[All Fields]) AND ("cells"[MeSH Terms] OR cell[All Fields]) AND ("rna"[MeSH Terms] OR RNA[All Fields]))) AND "Mus musculus"[porgn]

- (("cd4-positive t-lymphocytes"[MeSH Terms] OR CD4 t cell[All Fields]) AND naive[All Fields] AND (("single person"[MeSH Terms] OR single[All Fields]) AND ("cells"[MeSH Terms] OR cell[All Fields]) AND ("rna"[MeSH Terms] OR RNA[All Fields]))) AND "Mus musculus"[porgn]

<u>Search details for extracting Activated T cell datasets</u>

- ((activated[All Fields] AND ("t-lymphocytes"[MeSH Terms] OR t cell[All Fields])) AND scRNA seq[All Fields]) AND "Mus musculus"[porgn]
- (("cd4-positive t-lymphocytes"[MeSH Terms] OR CD4 t cell[All Fields]) AND scRNA seq[All Fields]) AND "Mus musculus"[porgn]
- ((activated[All Fields] AND ("t-lymphocytes"[MeSH Terms] OR t cell[All Fields])) AND (("single person"[MeSH Terms] OR single[All Fields]) AND ("cells"[MeSH Terms] OR cell[All Fields]) AND ("base sequence"[MeSH Terms] OR sequence[All Fields]))) AND "Mus musculus"[porgn]
- (("cd8-positive t-lymphocytes"[MeSH Terms] OR CD8 t cell[All Fields]) AND activated[All Fields] AND (("single person"[MeSH Terms] OR single[All Fields]) AND ("cells"[MeSH Terms] OR cell[All Fields]) AND ("rna"[MeSH Terms] OR RNA[All Fields]))) AND "Mus musculus"[porgn]
- (("cd4-positive t-lymphocytes"[MeSH Terms] OR CD4 t cell[All Fields]) AND activated[All Fields] AND (("single person"[MeSH Terms] OR single[All Fields]) AND ("cells"[MeSH Terms] OR cell[All Fields]) AND ("rna"[MeSH Terms] OR RNA[All Fields]))) AND "Mus musculus"[porgn]

We diligently curated the results through a thorough examination of the experimental designs. We prioritized minimally treated cells from the control group in our selection criteria, aiming to gain insights into the expression of marker genes under conditions of homeostasis.

Initially, we chose three HEV datasets for analysis; however, we had to exclude one of them from consideration due to an insufficient number of cells. From the remaining two datasets representing the control group, we included a total of 445 cells. To balance the

dataset sizes and mitigate potential bias resulting from variations in cell-type populations, we included 450 cells from each of the remaining three cell types.

Since the cell-type populations were relatively small, and datasets for naive T cells, activated T cells, and brain endothelial cells had a higher number of control group cells, we chose to limit the datasets from these three cell types to one each, thereby minimizing undesired variations.

3.2 Results

The details of the datasets are included in the table below:

Cell-Type	**GEO ID**
Brain Endothelial Cells	GSE134058
CD8+ T-Naïve Cells	GSE217656
CD8+ T-Activated Cells	GSE211602
Lymph node HEVs	GSE140348 and GSE198069

Table 3-1: Datasets used for analyses

Brain Endothelial Cells (GSE134058): This dataset contains expression profiling of brain endothelial cells in young (3 mon) healthy C57BL/6 mice. The experimental design included the study of transport of blood plasma proteins across BBB in young and aged mice[55].

CD8+ T- Naïve Cells (GSE217656): This dataset contains Naïve CD8+ T cells purified from spleen of Kmt2d WT and KO mice. Kmt2d mice were generated specifically for this experiment which studied the role of Lysine specific methyltransferase 2D (Kmt2d) gene in the naïve CD8+ T cell generation and survival[56]. Kmt2d WT mice expression profile was used for analysis.

CD8+ T-Activated Cells (GSE211602): This dataset contains in-vitro activated naïve CD8+ T cells isolated from the spleens of 8–13-week-old male OT-I transgenic mice. The cells were activated and differentiated using anti-mouse CD3, anti-mouse CD28

and recombinant murine IL-2 and cultured in physiological media[57] to study T-cell metabolism during immune response. Only activated T-cells were used for the analysis.

HEV-1 (GSE140348): This dataset contains transcriptomic profile of blood vascular endothelial cells from resting peripheral lymph nodes of Balb/c mice[58]. The expression values belonged to different cell-types. For this analysis only high endothelial cells were subset from the total population.

HEV-2 (GSE198069): This dataset contains the expression profile of peripheral lymph node HEV (homeostatic) and tumor HEV from C57BL/6, FVB/N mice. The effect of immunotherapies in endothelial fate mapping and differentiation of HEV[59]. Homeostatic LN-HEV cells were used for analysis.

Chapter 4

PIPELINE DEVELOPMENT

The increasing volume of datasets and the continual development of new analysis methods for single-cell data have resulted in a lack of standardization, primarily due to the evolving nature of this field[60]. One of the major challenges involves the necessity of identifying a shared analytical approach that can be applied across a wide array of biological data types, including cell lines, cancer cells, stem cells, and more. This has resulted in the formulation of general guidelines and best practices for data analysis[48,60]. Following the established principles, we developed a robust single cell RNA sequencing (scRNA seq) data analysis pipeline for identifying cell-type markers and unraveling cellular interactions across various cell-types and experimental conditions. One of the primary challenges we addressed in this endeavor was integrating heterogenous datasets from different sources before proceeding with downstream analyses. We additionally assessed the results of data integration for our datasets using six benchmarked algorithms, employing essential evaluation metrics. Importantly, this pipeline is compatible with different batches of expression matrices generated from sequencing experiments as well as from datasets sourced from public databases.

The data integration and evaluation are described in the next section and each of the downstream analyses is discussed separately in the subsequent chapters. We also tested the pipeline on a new dataset, the details of which are discussed in chapter 9.

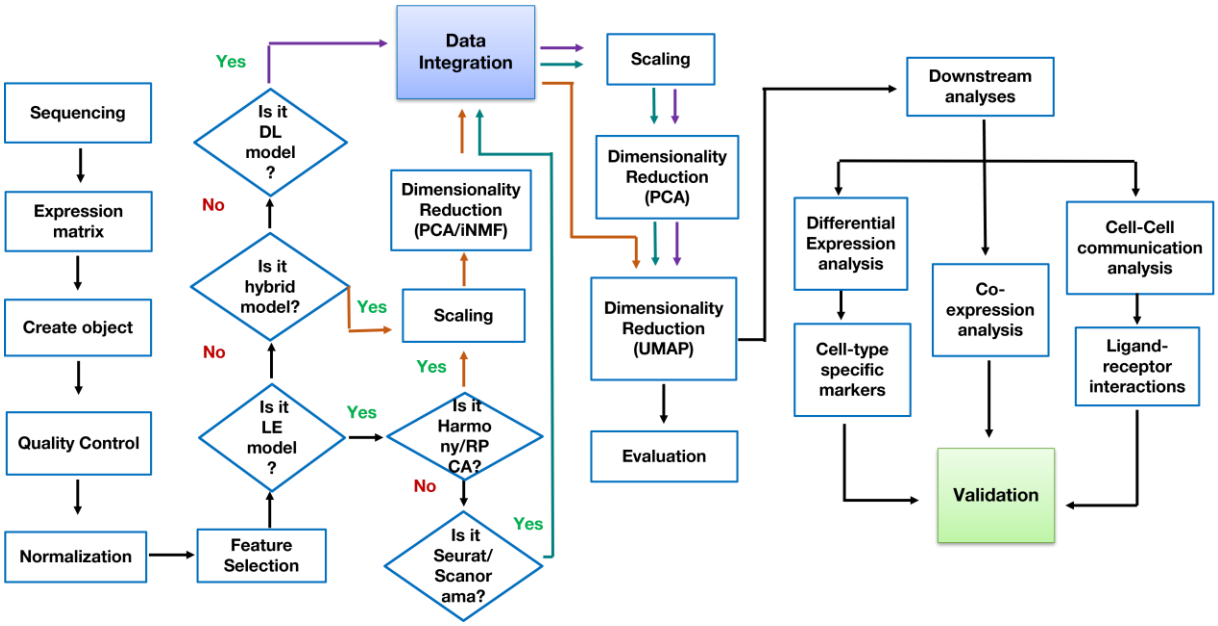


Figure 4-1: Single cell RNA sequencing data analysis pipeline

Chapter 5

DATA INTEGRATION AND EVALUATION

The capacity to investigate cellular diversity represents a significant benefit of scRNA-seq analyses, driving substantial progress in research and applications. However, this technique presents a significant challenge in the form of batch effects. Batch effects are alterations in expression level measurements resulting from variations in the handling conditions of cells across different batches. These variations can stem from factors such as distinct sampling times, library preparation methods, sequencing platforms, and experimental protocols[61]. Eliminating batch effects is a pivotal step to enable the integrated analysis of data. Striking the right balance between preserving intrinsic biological variation while effectively mitigating batch effects remains a central challenge in most scRNA-seq analyses.

Many models have been developed to address this issue. Batch effect removal strategies employ three steps of dimensionality reduction, modelling and eliminating batch effects and projection back into higher dimension space. While most approaches prioritize batch removal as the first step, some algorithms prefer to do it in reduced dimension space to improve signal to noise ratio[65]. A recent article on the best practices for single-cell analysis classifies the integration models into four categories[62]:

1) **Global models**: Global models, stemming from bulk transcriptomics, characterize the batch effect as a uniform influence (either additive and/or multiplicative) affecting all cells consistently. One of the popular choices using this approach is ComBat[63].
2) **Linear embedding models**: These techniques are specifically designed for batch correction in single-cell data. They frequently employ a modified form of singular value decomposition (SVD) to project the data, subsequently identifying clusters of similar cells across different batches in the projection. These clusters are then utilized to rectify the batch effect in a locally adaptive (non-linear) fashion. Most

prominent examples include Seurat integration[72], Harmony[73], MNN[64] and Scanorama[77].

3) **Graph-based models**: These methods employ a nearest-neighbor graph to depict the data within each batch. To mitigate batch effects, they establish connections between cells from distinct batches and subsequently adjust for variations in cell type compositions by selectively removing the newly introduced connections. BBKNN is a popular algorithm using this approach[65].

Deep learning based models: Deep learning (DL) techniques are the latest and most intricate strategies for addressing batch effects, usually demanding substantial data volumes for optimal results. These DL integration methods frequently rely on autoencoder networks. They either incorporate dimensionality reduction with consideration of the batch covariate using a conditional variational autoencoder (CVAE) or fit a locally linear adjustment within the embedded space. Popular approaches using this method include scVI[76], scANVI[66] and scGen[67].

The process of addressing batch effects is typically divided into two main tasks: batch effect correction and data integration[60]. Batch correction methods are used to mitigate batch effects within samples from the same experiment whereas data integration methods address intricate batch effects among datasets generated using different protocols, where cell identities may not overlap. Considering the complexity of this challenge and the variety of available tools, 19 methods have been benchmarked to identify optimal solutions for these tasks[72].

These benchmark studies focused on evaluation of the integration outcomes. Earlier studies prioritized batch effect removal and employes fewer metrics to assess the outcomes[68,70], the most popular being kBET. A recent study on benchmarking[72] used 14 evaluation metrics to measure batch effect removal and conservation of biological variance. Given the diversity of cell types in our datasets and the abundance of available tools, we opted to consolidate our data using six out of the 68 benchmarked algorithms and evaluate the results using four widely recognized evaluation metrics.

5.1 Materials & Methods

We performed all workflow steps using R version 4.2.2. For our analyses, we selected Seurat platform (v4.3.0), a widely acclaimed R package tailored for single-cell data exploration and downstream analysis, chosen for its seamless compatibility with a variety of other single-cell analysis tools. While two integration algorithms were Python-based, we utilized a Python interface called 'reticulate' throughout the analysis process. A comprehensive list of the libraries/packages employed in our analysis can be found in the table below:

1.	Seurat – 4.3.0	9.	Scvi – 0.14.6
2.	Matrix – 1.5-4	10.	Scanorama – 1.7.3
3.	Harmony – 0.1.1	11.	Sceasy – 0.0.7
4.	Patchwork – 1.1.2	12.	Anndata – 0.7.5.6
5.	Cluster – 2.1.4	13.	kBET – 0.99.6
6.	Liger – 2.0.1	14.	SeuratWrappers – 0.3.1
7.	Reticulate – 1.28	15.	Lisi – 1.0
8.	Scanpy – 1.9.3	16.	Cowplot – 1.1.1

Table 5-1:List of packages used for data integration

Our primary aim was to create a harmoniously integrated dataset that could support advanced analyses. To achieve this, we meticulously followed the steps outlined in the pipeline, which are elaborated upon below.

1) **Create object**: All single cell analysis tools require the creation of an object which serves as a container holding all necessary data associated with the scRNA seq experiment. This data includes gene expression values, metadata, and the results of various analyses.

   For the brain (GSE134058) and HEV-2 (GSE198069) datasets, the expression matrices were provided in .csv format and included gene expression values, cell annotations, and gene names. We processed these matrices by reading and ensuring the uniqueness of all values. Any duplicate values were removed, genes were

designated as row names, and Seurat objects were created using the 'CreateSeuratObject' function. Additionally, for the HEV-2 object, cell identities were manually matched to the provided cell annotations. Subsequently, we selected a subset of homeostatic LN-HEV cells from the HEV-2 object.

The remaining datasets, HEV-1 (GSE140348), CD8+ T-Naïve cells (GSE217656), and CD8+ Activated T cells (GSE211602), consisted of count matrices, separate barcode, and features files. We processed these datasets by associating features (genes) as row names and barcodes (cell identities) as column names. Seurat objects were then created using the 'CreateSeuratObject' function. Subsequently, we selected a subset of "High Endothelial Cells" from the HEV-1 object and "Act," representing activated T-cells, from the CD8+ Activated T cells object.

2) **Quality Control**: The general guidelines on best practices recommend considering lowest count depth and gene per barcode peak as non-viable cells for heterogenous datasets if no previous quality control have been performed[60]. All the studies from which the datasets were obtained confirmed the implementation of their own quality control procedures on the cells. However, there was no unanimous consensus regarding the filtering of mitochondrial transcripts. Elevated mitochondrial genes in a data is indicative of cellular stress and contributes to biological variability[60]. A thresholding value of 5% mitochondrial genes is accepted in the scientific community[69]. We performed this quality assessment using 'PercentageFeatureSet' function in Seurat.

3) **Normalization**: To address the unwanted bias arising out of differences in sample handling, library construction and sequencing we conducted normalization of the data. The widely adopted method for rendering gene expression values comparable is the use of Transcripts Per Million (TPM)[60]. This method is the default in Seurat, also called 'LogNormalize', where feature counts for each cell are divided by the total counts for that cell and then multiplied by a scale factor (default value: 10,000). Subsequently, the resulting expression values are subjected to log-transformation to

improve their fit to a normal distribution. For each dataset, we employed the 'NormalizeData' function in Seurat to perform this normalization.

4) **Feature Selection**: This step aims to identify genes that contribute strongly towards the cell-to-cell variations; in other words, genes that have the highest variance in expression across all the cells in the dataset. These genes are also called Highly Variable Genes (HVG). The uninformative genes i.e., genes with low basal expression levels or similar levels across all cells are filtered out to de-noise the data. This step has been reported to enhance data integration performance[70,71]. Typically, it is recommended to select a range of 1000 to 5000 HVGs. For our analysis, we opted for 5000 HVGs in each dataset.
5) **Scaling**: We scaled and centered the data by using 'ScaleData' function. This step is indicated to improve batch effect removal after integration[64]. Some algorithms like Harmony, LIGER and RPCA recommend scaling the data before integration and some like Seurat, Scanorama and scVI recommend the scaling step after integration.
6) **Principal Component Analysis (PCA)**: It's a statistical method widely utilized in machine learning for dimensionality reduction, aiming to retain most of the original data's variance. Some tools like Harmony and RPCA require dimension-reduced embeddings for data integration, while other algorithms recommend performing PCA after integration. In our case, we consistently reduced the data to 50 principal components.
7) **Data Integration**: We integrated the individual datasets using six different algorithms:
    a) **Seurat**: This algorithm leverages Canonical Correlation Analysis (CCA) to establish connections between individual cells across heterogeneous datasets that share a common set of genes. In essence, CCA identifies anchor points for integration by identifying common sources of variation between the datasets. It treats the datasets as multiple instances of a gene-to-gene covariance structure, with the goal of pinpointing shared patterns among them[72].

b) **Harmony**: Harmony takes principal component (PC) embeddings as input and employs soft k-means clustering to assign cells to multiple clusters. Clusters disproportionately containing cells from a small subset of datasets are penalized. It computes both global and dataset-specific centroids for each cluster and calculates correction factors for each dataset within clusters. Cells are corrected using cell-specific factors—a linear combination of dataset correction factors weighted by soft cluster assignments. These steps are iterated until convergence[73].

c) **LIGER**: Linked Inference of Genomic Experimental Relationships (LIGER) employs integrative non-negative matrix factorization (iNMF) to create a low-dimensional space where each cell is represented by dataset-specific factors (metagenes) and shared metagenes. Following iNMF it enhances joint clustering by assigning labels to cells based on maximum factor loadings and constructing a shared factor neighborhood graph that connects cells with similar factor loading patterns[74].

d) **RPCA**: Reciprocal PCA (RPCA) method is developed by Satija lad, developers of Seurat. They claim that this is a more conservative approach as compared to CCA and is suited for cells in different biological states or when a significant amount of cells in one dataset has no matching type in the other[75]. In this method when identifying anchors between two datasets with RPCA, each dataset is projected into the PCA space of the other and the anchors are constrained by the same mutual neighborhood requirement.

e) **scVI**: single cell Variational Interference (scVI) is a novel approach designed for the normalization and analysis of scRNA-seq data. It leverages a hierarchical Bayesian model with conditional distributions defined by deep neural networks. This model can efficiently handle even large datasets. Each cell's transcriptome is encoded into a low-dimensional latent vector of normal random variables using a nonlinear transformation. This latent representation is then decoded to estimate the distributional parameters of each gene in each cell, assuming a zero-

inflated negative binomial distribution to account for overdispersion and limited sensitivity in the data[76].

f) **Scanorama**: This approach extends the concept of mutual nearest neighbors matching to discover similar elements across multiple datasets. It reduces the dimensionality of gene expression profiles for each cell using a fast, randomized singular value decomposition (SVD), which improves the method's noise resistance. Additionally, it employs an approximate nearest neighbor search technique based on hyperplane locality-sensitive hashing and random projection trees which speeds up the query process[77].

8) **Uniform Manifold Approximation and Projection (UMAP)**: UMAP is a nonlinear dimensionality reduction method often utilized for visualization purposes. In our analysis pipeline, we employed UMAP in tandem with PCA or other dedicated dimension reduction techniques tailored to each specific tool. For instance, when working with LIGER, which incorporates its proprietary dimension reduction algorithm known as iNMF, we adapted our approach accordingly.

9) **Evaluation**: After integrating the data, it is important to evaluate the performance as some methods might overcorrect during batch effect removal[62,77]. We used four key metrics to assess how well the data had been integrated:

a) **Silhouette coefficient**: This parameter quantifies the clustering quality of similar cell types and serves as a common metric for evaluating single-cell data integration methods[70,71,78]. Its value ranges from -1 to 1 where a score of 1 indicates well-defined and distinct clusters. A score of 0 indicates unclear or overlapping cluster boundaries while -1 indicates misassignments of cells to clusters.

b) **Mixing Metric**: The "mixing metric" is a measure used to assess how effectively different datasets are combined in single-cell RNA sequencing (scRNA-seq) analysis. It evaluates the degree to which similar cells from various datasets are mixed together after integration. It quantifies the mixing by analyzing the distribution of nearest neighbors for each cell, aiming for a well-mixed

neighborhood with representation from multiple datasets. The value ranges from 0 to 300 and higher scores indicate better mixing[79].

c) **Local Inverse Simpson's Index (LISI)**: LISI is a diversity score assessing data integration accuracy, focusing on local cell diversity. It combines perplexity for local structure preservation and Inverse Simpson's Index for batch/cell type diversity. When applied to cell types, cLISI should ideally yield a score of 1, indicating well-mixed similar cell types while maintaining unique identities[73].

d) **Local structure preservation**: It is a metric to assess how effectively the original dataset structure is preserved post-integration. It compares the neighborhoods of cells based on the uncorrected data with those from the integrated dataset. By calculating the overlap fraction for each cell and averaging across all cells, an overall score is obtained indicating the degree of preservation. Higher score indicates better preservation[79].

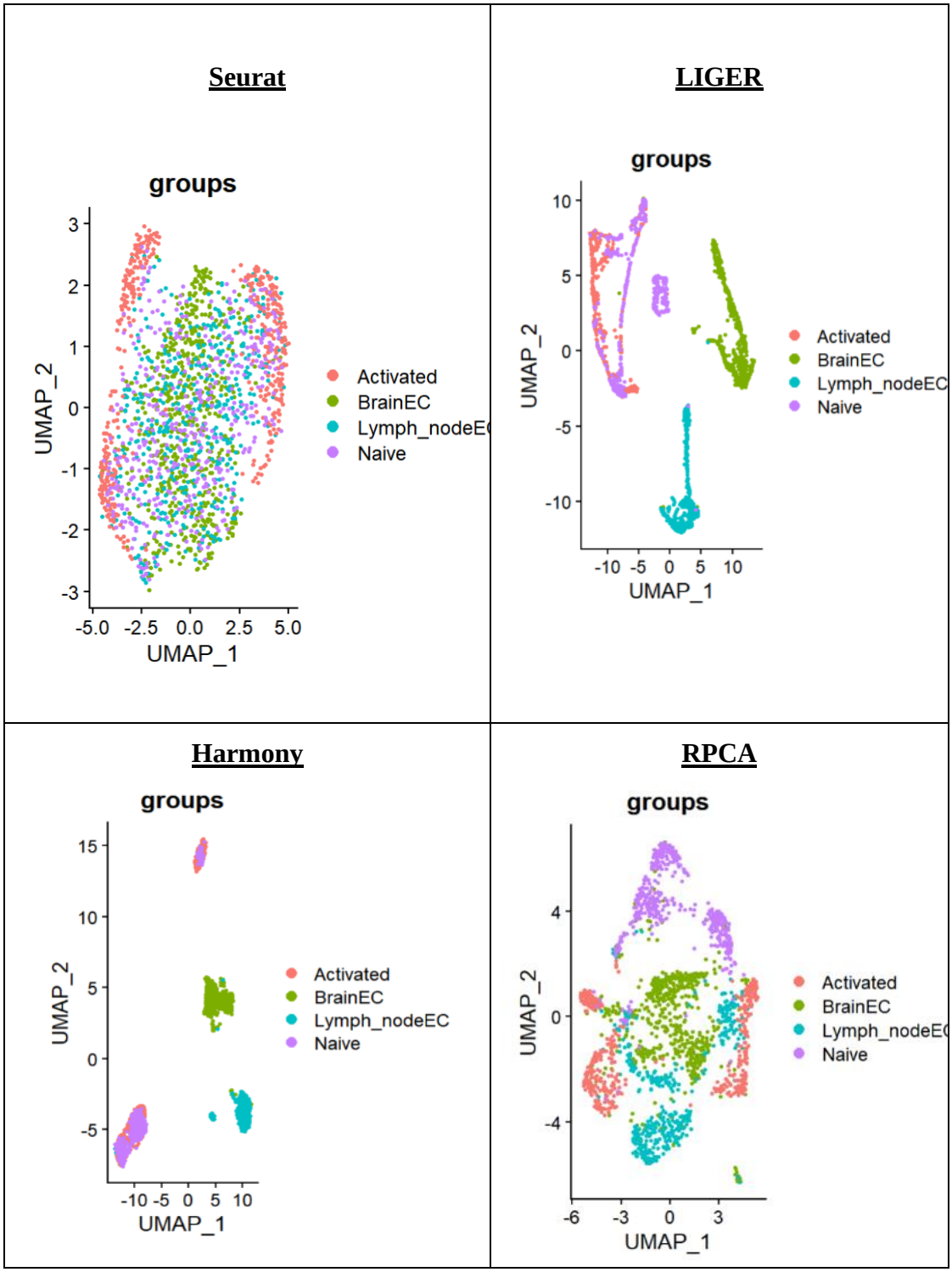
5.2 Results

The median scores obtained using different integration algorithms are summarized in the table below:

	Silhouette Score (-1 to 1)	**Mixing Metric (0 to 300)**	**cLISI (1 to 4)**	**Local structure preservation score (0 to 1)**
Seurat	-0.05	23.5	1.88	0.32
LIGER	0.16	58	1.001	0.38
Harmony	0.66	300	1.09	0.50
RPCA	0.169	27	1.13	0.48
scVI	-0.19	29.5	1.55	0.37
Scanorama	0.88	300	1.00	0.26

Table 5-2: Evaluation scores of different algorithms

Below table shows the UMAP visualization of the integration algorithms:



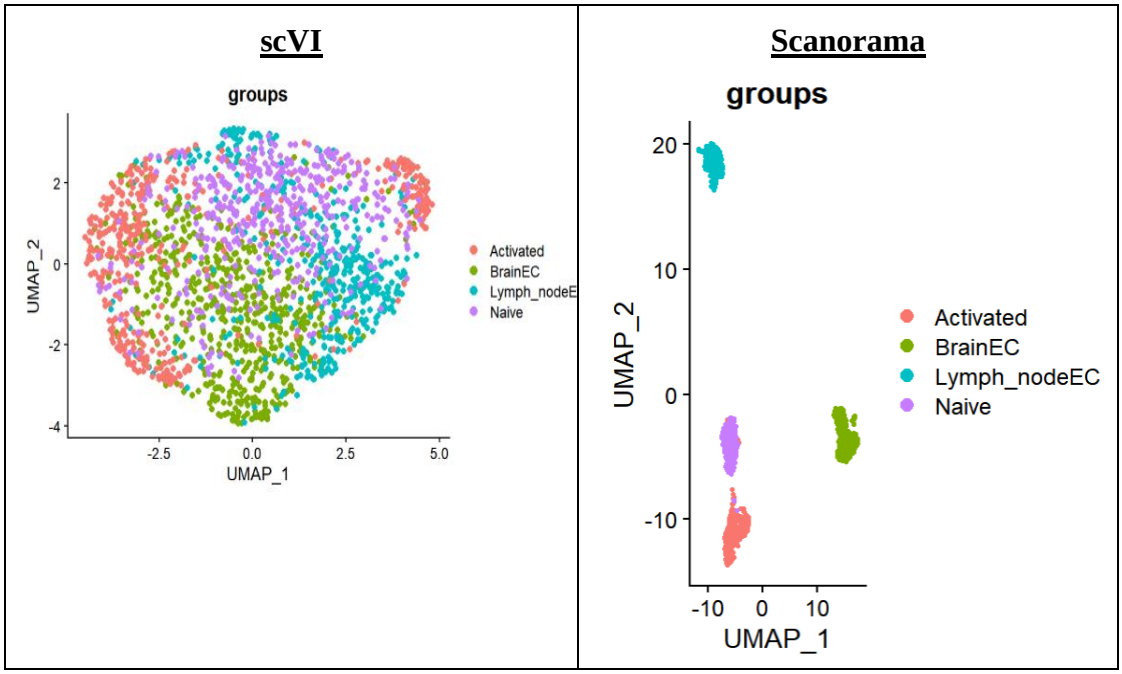


Figure 5-1: UMAP embeddings of integration of different algorithms

The table shows that both Seurat and scVI have the lowest scores for all integration metrics. This is consistent with the UMAP plot, where cells are highly mixed with no distinct clusters. Additionally, the silhouette coefficient for scVI is -0.19, and for Seurat, it's -0.05, indicating poor cluster separation. Both algorithms also exhibit a lower mixing metric, high cLISI, and low local structure preservation scores. Specifically, Seurat has scores of 23.5, 1.88, and 0.32 for mixing metric, cLISI, and local structure preservation, respectively, while scVI has scores of 29.5, 1.55, and 0.37 for the same metrics.

RPCA performs better in local structure preservation with a score of 0.48 but lags in all other metrics. It achieves scores of 0.169 for the silhouette coefficient, 27 for the mixing metric, and 1.13 for cLISI. While its UMAP plot displays improved cLISI scores compared to Seurat and scVI, cluster separation remains suboptimal.

LIGER performs slightly better than Seurat, scVI, and RPCA in the mixing metric with a score of 58 and boasts one of the best overall cLISI scores at 1.001. However, it records a lower silhouette coefficient of 0.16. The UMAP plot reveals distinct brain and

HEV clusters but indicates a slight overlap between naïve and activated T cell clusters. Unfortunately, LIGER also has a poor local structure preservation score.

Harmony stands out with a strong silhouette score of 0.66, reflecting its excellent performance. The UMAP plot further confirms this by revealing well-defined clusters. Harmony shares the top spot for the mixing metric, achieving the highest score of 300. It also excels in cLISI with a score of 1.06. It has the best structure preservation score at 0.50. The UMAP plot displays clear separation between clusters, particularly distinguishing brain and HEV clusters. However, it does show an overlap between activated and naïve T cells within one cluster.

Scanorama emerges as the top performer among all algorithms. It achieves the highest silhouette coefficient score of 0.88, a perfect score of 300 for the mixing metric, and a flawless cLISI score of 1.00. The UMAP plot showcases distinct clusters for brain, HEV, activated, and naïve T cells. However, Scanorama lags in local structure preservation, recording a score of 0.26.

5.3 Discussion

A wide range of performance variations is observed among different integration algorithms, each employing unique methodologies. A substantial study, encompassing 68 methods and over a million cells across 13 atlas-level integration tasks, was conducted to provide guidance in selecting an integration method[71]. This study revealed that Seurat and Harmony excelled in simpler integration tasks, while Scanorama and scVI demonstrated proficiency in handling more complex tasks.

In the context of our dataset, we noticed that the integration performance of Seurat and scVI, as evidenced by UMAP plots, fell short of expectations, suggesting a potential overcorrection of expression values during integration. To gain a broader perspective, the figures below (Fig.5-2, 5-3) display the performance of various algorithms across key metrics for tasks of varying complexities from the aforementioned study.

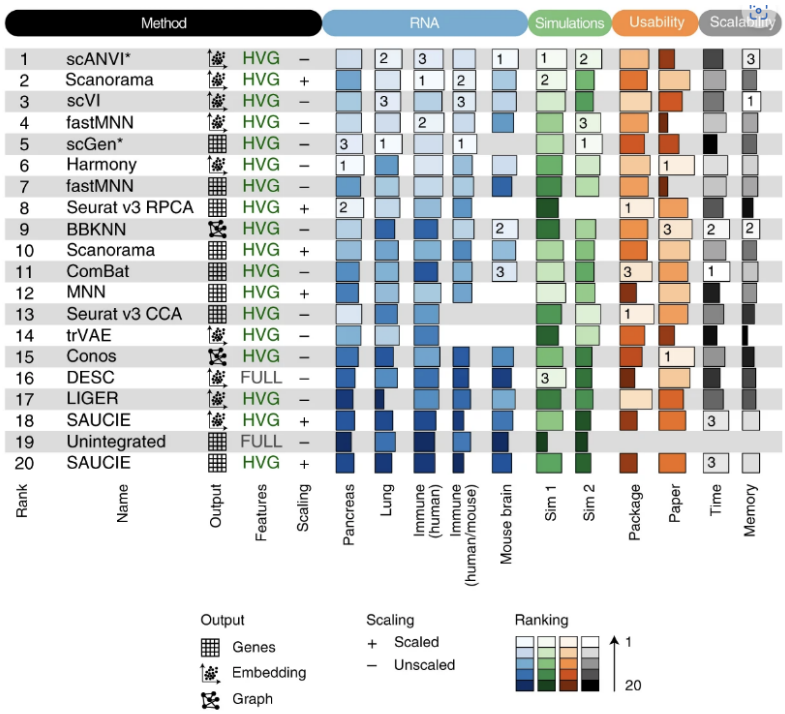


Figure 5-2: Comparison of data integration performances based on methodology, usability and scalability (Figure adapted from Leuken D et.al[71])

scVI exhibited strong performance when dealing with large datasets in the benchmark study, as seen in Fig.5-2 and 5-3. However, our dataset is relatively small, with fewer than 2000 cells and of moderate complexity due to sequencing on different platforms and originating from different tissues. It's worth noting that scVI, being a deep learning-based model, may underperform when dealing with limited data, as there might not be enough data for proper model training.

Seurat, on the other hand, did not meet the criteria for integration metrics, speed, and task details (Fig.5-3). The RPCA and original CCA methods of Seurat were ranked at 8 and 13, respectively, as shown in Fig.5-2, placing them lower than all the other algorithms assessed in our analysis, with the exception of LIGER. Our accuracy scores align with those of the study, with RPCA outperforming Seurat CCA, a trend evident in UMAP plots as well.

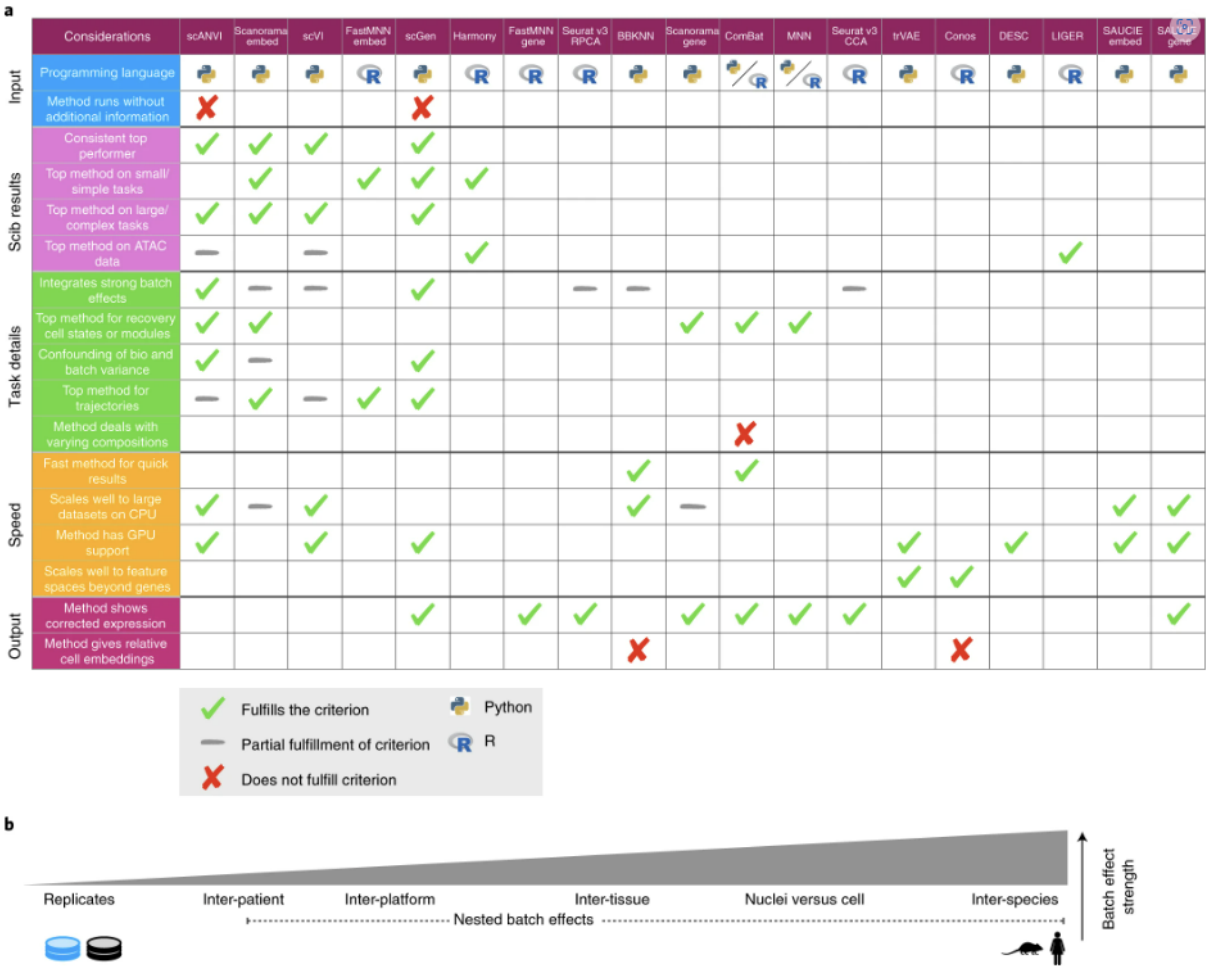


Figure 5-3: Table of criteria to consider when choosing an integration method, and which methods fulfill each criterion. Ticks show which methods fulfill each criterion and gray dashes indicate partial fulfillment.

(Figure adapted from Leuken D et.al[71])

Despite displaying relatively well-defined clusters in its UMAP plot, LIGER's performance metrics contradicted these visual results. In the benchmarked study, LIGER's performance was subpar, as indicated in Fig. 5-2, which corroborates our results.

Scanorama demonstrated remarkable performance in three crucial metrics within our dataset, aligning with its status as a top performer in the benchmarked study (Fig.5-2). However, it's important to note that its differential expression analysis resulted in the

identification of fewer than 10 markers, even when using a relaxed p-value threshold of 0.1. An examination of the integration output revealed that the integrated object contained embeddings of 100 Principal Components. These factors might explain the limited number of markers identified in the differential expression analysis. It's worth considering that having corrected counts in the integration object, rather than PC embeddings, might yield better results in the differential expression analysis.

Harmony showcased remarkable and consistent performance across all assessed metrics, which is also evident when examining the UMAP plots. Furthermore, Harmony stood out as the leading method for managing smaller and less complex integration tasks, as indicated in Fig. 5-3. In our performance ranking based on the criteria shown in Fig. 5-2, Harmony secured an impressive third position among the evaluated algorithms. Consequently, all subsequent analyses were carried out using the Harmony-integrated data.

Chapter 6

DECIPHERING CELL-TYPE DIVERSITY WITH DIFFERENTIAL EXPRESSION ANALYSIS OF SURFACE MARKERS

Differential gene expression analysis represents a fundamental use of transcriptomic data, frequently employed to identify genes that exhibit significant expression differences between two distinct biological conditions or cell types. This analysis offers valuable insights into the underlying pathways and mechanisms driving variations in phenotype within a population. Our objective is to identify specific surface markers associated with distinct cell types through differential gene expression analysis and gene set enrichment analysis. These markers may have significant implications in immune surveillance and immune responses, particularly concerning the endothelial barriers under investigation.

Statistical methods used for identifying differentially expressed (DE) genes for bulk RNA seq data include Fisher's exact test, DESeq2, Likelihood Ratio and edgeR. These methods often assume specific underlying distributions, such as Poisson or Negative Binomial[80]. However, single-cell transcriptomic data introduces unique challenges due to high drop-out rates, zero-inflation, and bimodal distribution. This renders the traditional bulk RNA-seq approaches ill-suited for the task.

To tackle these challenges, various methods have emerged, some adapted from bulk RNA-seq techniques, referred to as pseudo-bulk methods. These adapted methods encompass Negative Binomial models (e.g., DESeq2, edgeR, NBPseq), Poisson-based approaches (TSPM, DEGseq), Linear models (Limma), and Non-parametric methods (SAMSeq, NOIseq). Additionally, specialized methods tailored for single-cell RNA-seq data have been developed, such as Zero-Inflated Negative Binomial models (e.g., ZINB-Wave, DECENT), hurdle models (MAST), Linear models (Monocle, ZIAQ), and Non-parametric models based (Wilcoxon, Sincera)[81].

Model-based Analysis of Single-cell Transcriptomics (MAST) is one of the most popular tools used for DE analysis for single cell RNA seq data. It's a generalized linear

model that encompasses two essential components. First, it simultaneously models the rate of gene expression, accounting for both technical and biological factors that contribute to its variability. Second, it focuses on modeling the mean gene expression levels when genes are positively expressed. A recent study reported that MAST has the best performance for single-cell data among four other tools evaluated[82]. Another comparative study reported that it performed the best at extracting biologically relevant gene sets from the data[83]. Considering these advantages, we decided to conduct the differential gene expression analysis using MAST.

Following the differential expression analysis, we conducted Gene Set Enrichment Analysis (GSEA). This analysis identifies sets of genes or proteins that might exhibit an over-representation within the gene ontologies, thereby contributing to distinct phenotypic characteristics. Gene ontology is a standardized system of defining and categorizing genes and proteins based on their functions. It is categorized into three aspects[84]:

Biological Process (BP): This term describes gene products involved in large biological processes like cellular respiration.

Cellular Component (CC): This term defines gene products active in different cellular compartments, for example, cytoplasm.

Molecular function (MF): This term specifies biochemical activities or functions of gene products.

As our goal is unraveling cell-cell interactions we focused our analysis on the cellular component.

6.1 Materials & Methods

We conducted the analysis using R 4.2.2 version. Seurat (4.3.0) platform was used for differential gene expression analysis as it offers compatibility with MAST statistical test. The packages used for the analysis are listed in the table below:

1.	Seurat – 4.3.0	7.	MAST – 1.24.1
2.	Matrix – 1.5-4	8.	Metap – 1.8
3.	Harmony – 0.1.1	9.	Enrichplot – 1.18.4
4.	Patchwork – 1.1.2	10.	ggnewscle – 0.4.9
5.	Clusterprofiler – 4.6.2	11.	Org.Mm.eg.db – 3.16
6.	Multtest– 2.54.0	12.	dplyr – 1.1.2

Table 6-1: List of packages used for differential gene expression and enrichment analysis

We conducted differential gene expression analysis between HEV and brain endothelial cells using Seurat's 'FindMarkers()' function. The log-fold change threshold (logfc) parameter was set to the default value of 0.25 to include a broad range of differentially expressed genes. Additionally, we adjusted the 'min.diff.pct' parameter to 0.20 to ensure a higher confidence level in the identified markers. This analysis was performed twice, once with 'brain endothelial cells' as the reference and then with 'HEV' as the reference, with the 'only.pos' feature set to 'True' to identify upregulated genes in both cases. The results were saved to a file for further analysis.

We then conducted an enrichment analysis using the 'enrichGO()' function from the clusterprofiler package. We specified the use of the mouse database and selected 'SYMBOL' as the key type for the analysis. Although we performed enrichment analysis for all three ontologies, our primary focus was on the cellular component aspect. Consequently, we recorded the results of this specific analysis in a file for further examination.

We followed the same protocol to conduct differential expression analysis between Naïve and Activated T cells.

We then manually selected the gene sets active in the sub-cellular location of plasma membrane from the description column of the enrichment results. The terms selected for identifying unique markers from the enrichment results are shown in the table below:

GO Terms selected to filter genes active on plasma membrane	
apical plasma membrane	cell-substrate junction
basal plasma membrane	cell projection membrane
basolateral plasma membrane	bicellular tight junction
membrane raft	cell cortex
membrane microdomain	cell-cell contact zone
apicolateral plasma membrane	intercellular bridge
plasma membrane raft	T cell receptor complex
cell projection membrane	filopodium membrane
basement membrane	filopodium
lateral plasma membrane	podosome
anchored component of membrane	lamellipodium
plasma membrane signaling receptor complex	pseudopodium
extrinsic component of plasma membrane	adherens junction
extrinsic component of membrane	apical junction complex
outer membrane	tight junction
anchored component of plasma membrane	neuromuscular junction
anchored component of external side of plasma membrane	gap junction
membrane coat	caveola
cell body membrane	lumenal side of membrane
intrinsic component of external side of plasma membrane	uropod
coated membrane	filopodium tip
cell cortex region	

Table 6-2: List of terms used to identify unique cell-surface markers

Each GO term had many genes associated with it. We isolated the genes using the TEXTSPLIT() function in Excel and removed duplicates to identify unique markers. Subsequently, we manually evaluated the markers for cell-type specificity based on their expression levels and existing literature.

6.2 Results

The table below summarizes the number of upregulated genes obtained after differential expression analysis, number of GO terms associated with the cellular component and number of unique markers identified for each cell-type.

	Endothelial cells		**T-cells**	
	HEV	**Brain**	**Naïve T-cells**	**Activated T-cells**
Upregulated genes	743	320	285	934
GO: Cellular component	508	378	346	566
No. of unique cell-surface markers	109	85	54	73

Table 6-3: Summary of the differential expression and enrichment analysis results

Bar plot of top 20 terms of enrichment analysis for Cellular Component

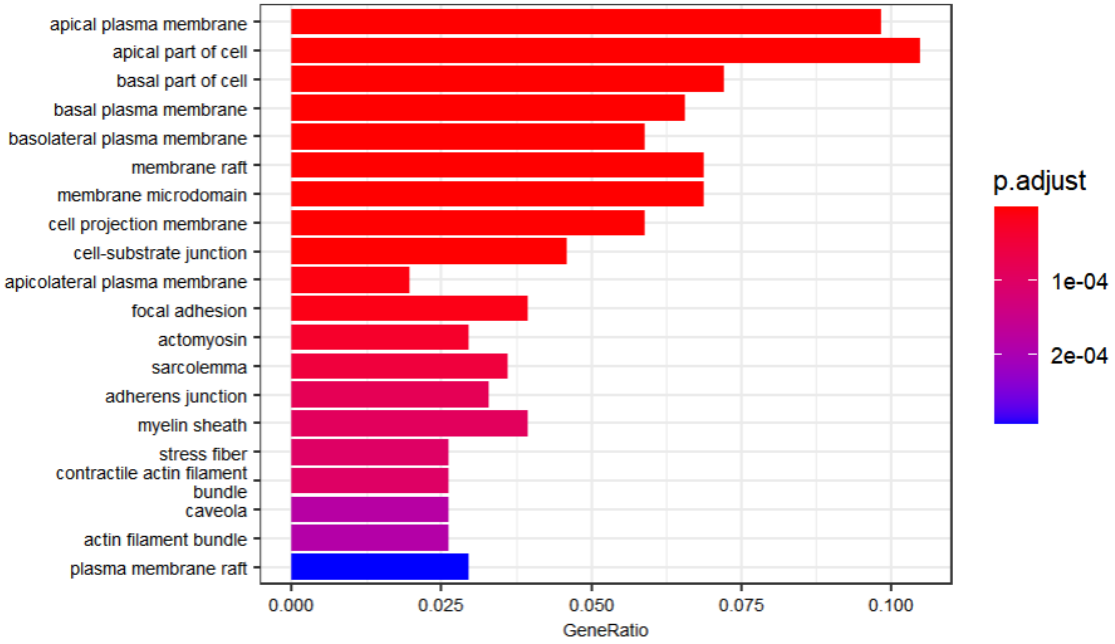


Figure 6-1: Barplot of top 20 enriched terms for CC in brain

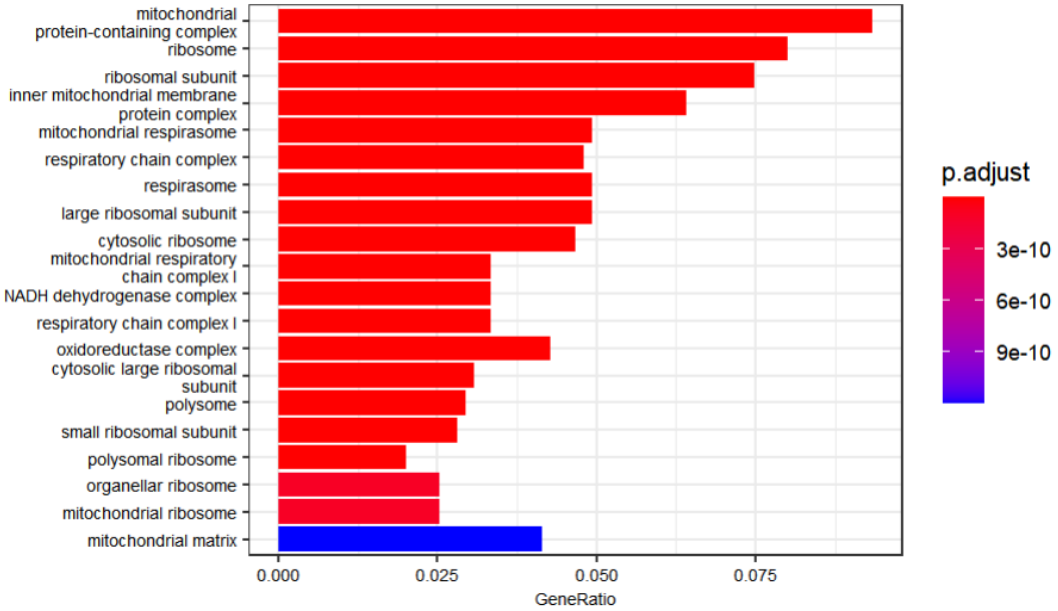


Figure 6-2: Barplot of top 20 enriched terms for CC in HEV

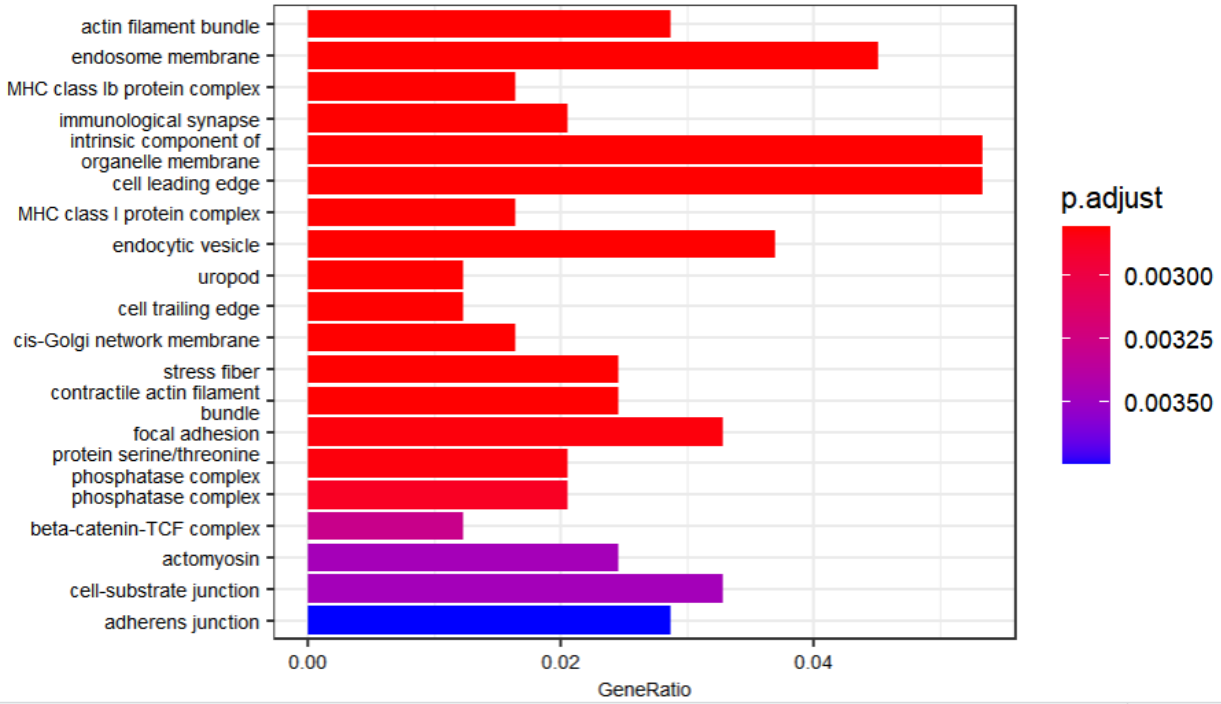


Figure 6-3: Barplot of top 20 enriched terms for CC in Naïve T cells

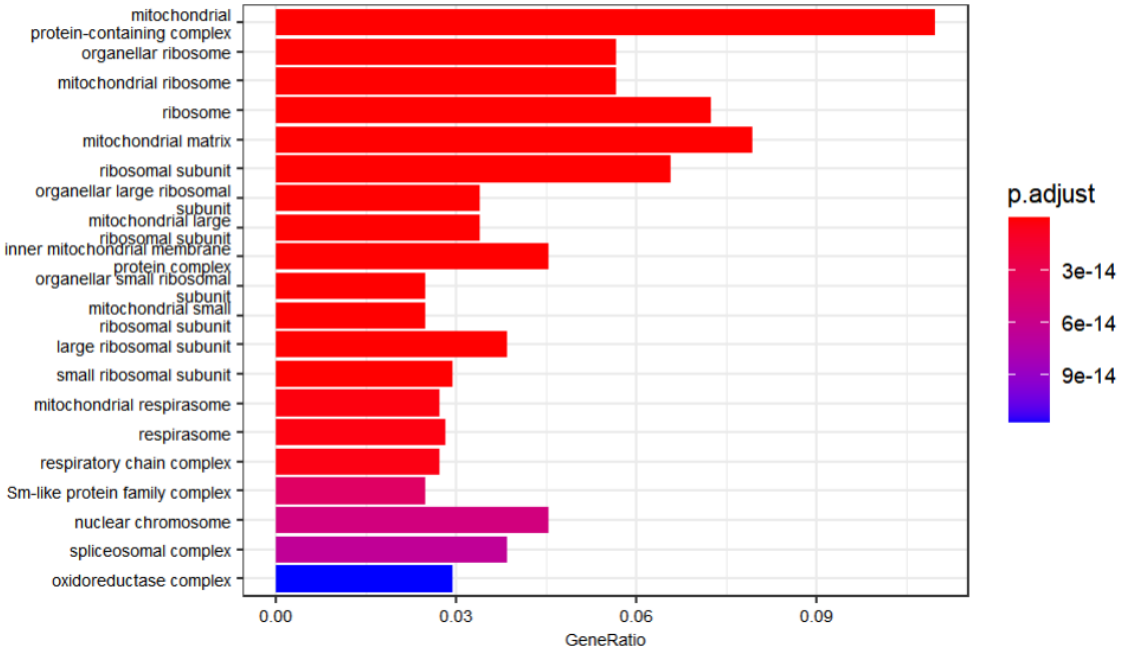


Figure 6-4: Barplot of top 20 enriched terms for CC in Activated T cells

Heatmap of cell-type specific markers

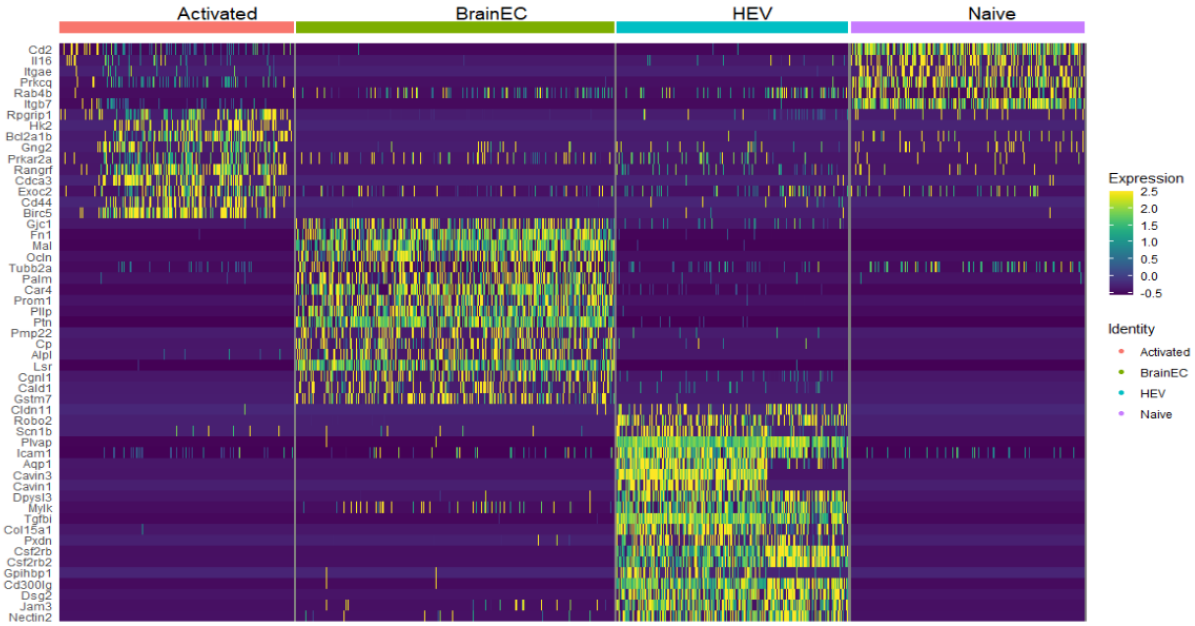


Figure 6-5: Heatmap of Cell-Type markers

We identified markers exhibiting cell-type specificity, associated with the cellular extravasation process. The violin plots indicating their expression levels are shown in Fig.6-6. The functions of these markers are described below according to cell-type:

Brain endothelial cell-type markers

- **Cxcl12**: It is a homeostatic chemokine that promotes cellular adhesion and migration. It is reported to play an important role in the T-cell recruitment across the BBB in CNS inflammation[85].
- **Fn1**: Fibronectin 1 is a crucial component of the extracellular matrix, playing a significant role in cell adhesion processes. It is reported to play anti-inflammatory role in the brain contributing to its neuroprotective effects[86].
- **Jup**: Junctional plaque proteins are expressed on cell junctions. They play a crucial role in stimulating VE-Cadherin which mediates cell adhesion[87].

HEV cell-type markers

- **Plvap**: Plasmalemma vesicle–associated protein, also known as MECA-32, plays crucial roles in regulating endothelial permeability and cellular extravasation. This protein is involved in controlling the passage of molecules and cells through the endothelial barrier, influencing processes such as immune cell migration and tissue homeostasis[88].
- **Cavin1**: Caveolae-associated protein 1, an integral part of caveolae, plays a pivotal role in stabilizing caveolin-1, the primary structural membrane protein of caveolae[89]. Caveolae play a crucial role in facilitating transcellular diapedesis on T cells.
- **Icam1**: Intercellular Adhesion Molecule-1 regulates T-cell rolling and adhesion on the endothelial wall during the extravasation process[11]. Its expression is known to be upregulated in inflamed blood brain barrier[90].

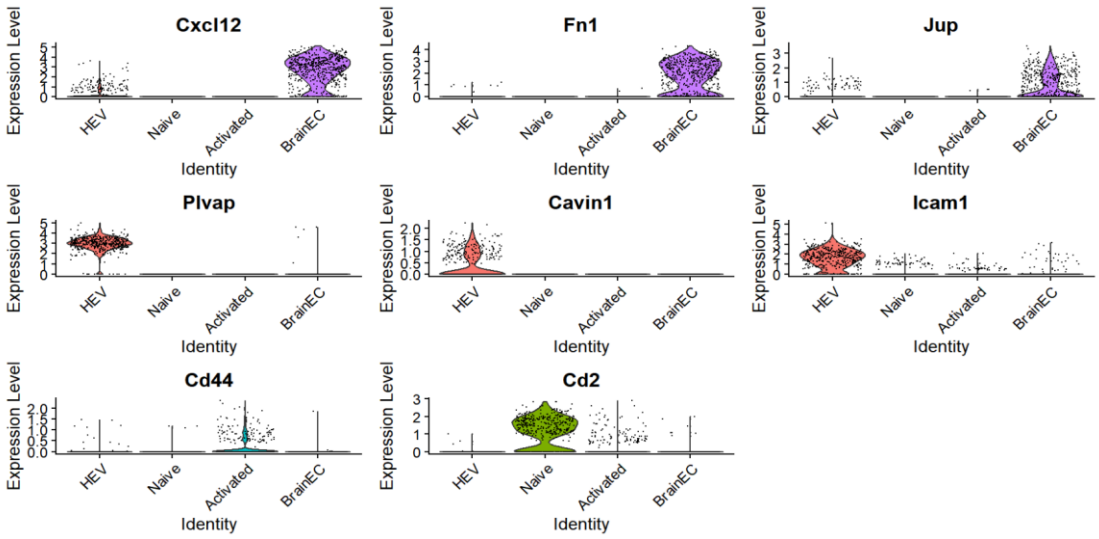


Figure 6-6: Violin plots showing expression levels of cell-type specific markers

Activated T-cell type marker

- **Cd44**: It is a cell surface receptor involved in cell-cell and cell-matrix adhesion, cell migration and signaling on the nervous system[91]. It is responsible for VLA-4 activation which is important for cell adhesion[92].

Naïve T-cell type marker

- **Cd2**: This is a transmembrane glycoprotein present on the surface of T-cells. It plays an important role in cell-cell adhesion and acts as co-stimulatory molecule to Cd58 (LFA-3) which is necessary for the generation of Th1 cells following T cell activation[93].

6.1 Discussion

In the enrichment bar plot of brain endothelial cells (Fig.6-1) one of the prominently enriched terms is 'caveola,' signifying that endothelial-mediated transcellular diapedesis is more pronounced in the blood-brain barrier (BBB) compared to HEV.

Previous research has elucidated the role of Major Facilitator Superfamily Domain containing 2a (Mfsd2a), a lipid transporter with high expression in brain endothelial cells, in suppressing caveolae formation in capillary endothelial cells, thus preserving barrier characteristics[94,95]. Conversely, arterial endothelial cells express lower levels of Mfsd2a, which contributes to increased caveolae formation[96].

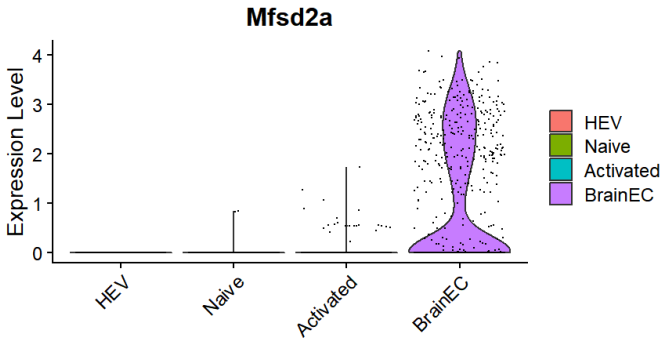


Figure 6-7: High expression of Mfsd2a in brain EC

Our findings corroborate the existing knowledge, showing that Mfsd2a is highly expressed in the brain but absent in HEV. The presence of 'caveola' as a significant GO term may seem contradictory, but it's important to note that our dataset comprises a mix of arterial, capillary, and venous endothelial cells, potentially explaining this observation.

Furthermore, an intriguing observation is that Mfsd2a, reported as a cell-surface protein and included among the highly expressed brain markers, did not exhibit enrichment for any terms within the cellular component ontology. This suggests that its cellular component associations may be complex and context-dependent, warranting further investigation.

Glycam1 (Glycosylation dependent cell adhesion molecule 1) and Ccl21a are well-established markers for high endothelial venules (HEV) and have been recognized for their crucial involvement in cell adhesion during the extravasation process, as documented in previous studies[97,11]. Our findings align with the established literature, demonstrating their robust expression in HEV, as evident in Fig. 6-8. It's noteworthy that despite their presence on the plasma membranes of cells, these markers did not exhibit enrichment within the Cellular Component ontology. This suggests that their cellular localization might be context dependent and further exploration is required.

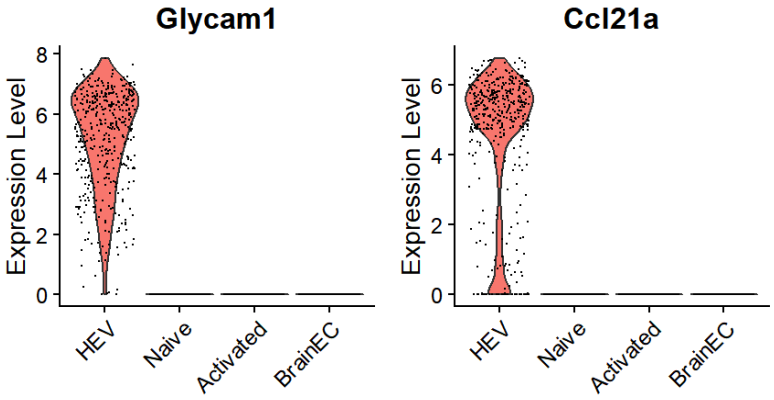


Figure 6-8: High expression of Glycam1 and Ccl21a in HEV

Consequently, we investigated the results of the Biological Process and Molecular Function ontology enrichment analyses. Within these analyses, Glycam1 displayed

enrichment in terms related to 'sulfur compound binding' and 'cell adhesion molecule binding,' while Ccl21a exhibited enrichment in terms associated with 'cytokine receptor binding' and 'cytokine activity' within the Molecular Function category.

Chapter 7

IDENTIFYING CO-EXPRESSION NETWORK PATTERNS USING CO-EXPRESSION ANALYSIS

Genes often collaborate closely to carry out specific functions, such as molecular transport or receptor binding, which can be indicative of tissue-specific expressions or biological conditions. For instance, a group of genes plays a vital role in wound healing, and these genes display co-expression patterns, meaning they are strongly correlated with one another. Examining these co-expression patterns in-depth can unveil the underlying dynamics of various functional pathways and might lead to the identification of unknown genes that might cluster together with known ones[98]. While differential gene expression analysis is instrumental in identifying markers for specific phenotypes, co-expression networks provide valuable insights into genes that exhibit coordinated expression patterns, shedding light on their collective functional significance.

A popular approach to identify co-expressing genes is through Weighted Gene Co-expression Analysis (WGCNA)[99]. It is based on differential network (or co-expression) analysis where the aim is to identify the changes in co-expression patterns of genes when their mean gene expression levels remain unaltered[100]. For example, a previous study reported that differential co-expression analysis was conducted on two bull varieties, one with a myostatin mutation and the other without. Despite no significant difference in average myostatin gene expression, it ranked highest among 920 transcriptional regulators in terms of a differential co-expression measure[101]. This has been reported in other studies as well[102,103]. This has been widely used for bulk RNA seq data in identifying networks in specific processes[104] and diseases[105].

hdWGCNA (high dimensional WGCNA), an extension of WGCNA, is an algorithm specifically tailored to identify co-expression networks in single cell and spatial transcriptomics data to address the sparsity and inherent noise in such data[106]. This approach first constructs 'metacells' to reduce dimensionality. Metacells are small groups of transcriptomically similar cells representing distinctive cell states. Then it

constructs a co-expression network representing relationships between genes based on their expression patterns across samples. Genes are assigned weights based on their contributions to the network. Genes that show similar expression patterns are grouped into modules, representing functional units within the cell. The process is optimized for efficiency and accuracy.

7.1 Materials & Methods

We conducted the analysis using R 4.2.2 version. Seurat (4.3.0) platform was used for co-expression analysis as hdWGCNA uses Seurat data structures. The packages used for the analysis are listed in the table below:

1.	Seurat – 4.3.0	6.	GeneOverlap – 1.34.0
2.	WGCNA – 1.72-1	7.	tidyverse – 2.0.0
3.	hdWGCNA – 0.2.18	8.	cowplot – 1.1.1
4.	igraph – 1.4.2	9.	patchwork – 1.1.2
5.	enrichR – 3.2		

Table 7-1: List of packages used for co-expression analysis

We conducted the co-expression network analysis on the integrated data object. hdWGCNA includes a function MetacellsByGroups to construct metacell expression matrices given a single-cell dataset. This function constructs a new Seurat object for the metacell dataset which is stored internally in the hdWGCNA experiment. We specified the Metacells to be constructed based on gene expression levels of ‘groups’. ‘groups’ metadata was created in Seurat where cell identities belonging to each cell-type was assigned to the following groups: ‘Naïve’, ‘Activated’, ‘BrainEC’ and ‘HEV’.

The initial step in network construction involves establishing a soft power threshold, which is a pivotal aspect of creating the gene-gene correlation adjacency matrix for deducing co-expression associations among genes. By exponentiating the correlations to a certain power, this procedure serves to diminish the impact of noise in the correlation matrix, preserving robust connections while filtering out weaker ones.

Consequently, selecting an appropriate value for the soft power threshold is of paramount importance in this process[107]. One of the hyperparameters requires to be specified is the type of co-expression network to be built. We chose ‘signed’ co-expression network for the analysis characterized by correlation values ranging from 0 to 1. In this setup, values below 0.5 signify negative correlations, while values exceeding 0.5 denote positive correlations. This approach is preferred because it generates networks that more effectively distinguish biologically meaningful modules, enhancing the interpretability of the results[108].

In the subsequent stage, Module Eigengenes are calculated to provide a concise representation of the gene expression patterns within individual co-expression modules. This is accomplished by applying principal component analysis (PCA) to a subset of the gene expression matrix specific to each module. The primary principal component (PC1) extracted from these PCA matrices serves as the Module Eigengene (ME). The next step involves computing pairwise correlations between genes and module eigengenes.

Following the establishment of module connectivity, we conducted module enrichment analysis using the enrichR package. Additionally, we performed marker gene overlap analysis, comparing the hdWGCNA modules with cluster or cell-type marker genes. This involved two steps: first, identifying marker genes in each cell type using Seurat's FindAllMarkers function, and subsequently employing the hdWGCNA function OverlapModulesDEGs to assess the overlap between the modules and the differentially expressed genes (DEGs).

7.2 Results

The algorithm classified our data into 7 modules of co-expression networks as shown in Fig.7-1.

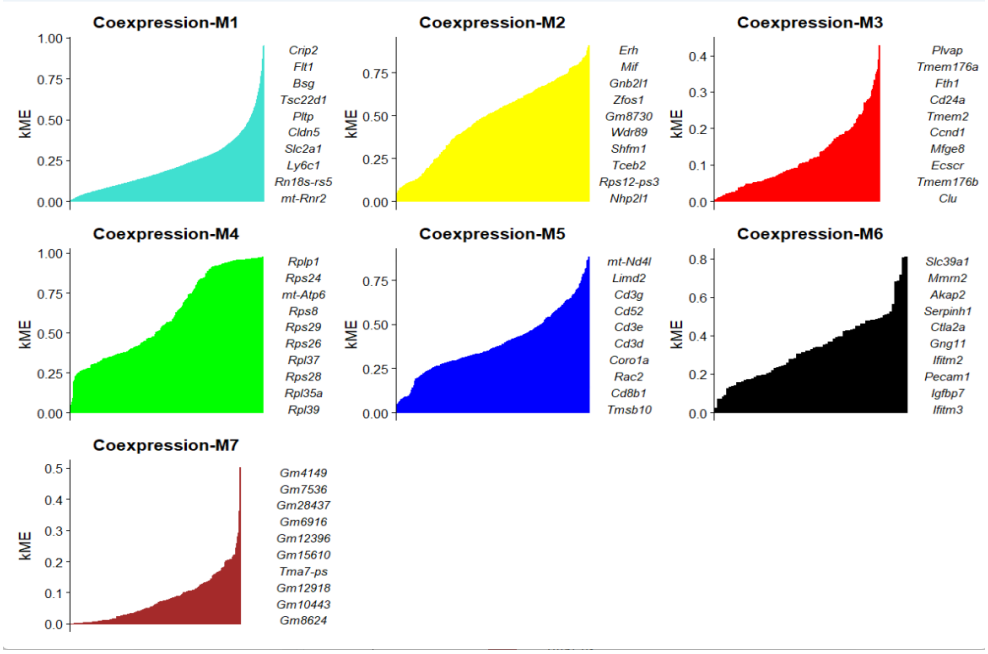


Figure 7-1: Assignment of co-expressing genes to modules

The mapping of modules to different cell-types and overlap of marker genes is shown in Fig.7-2.

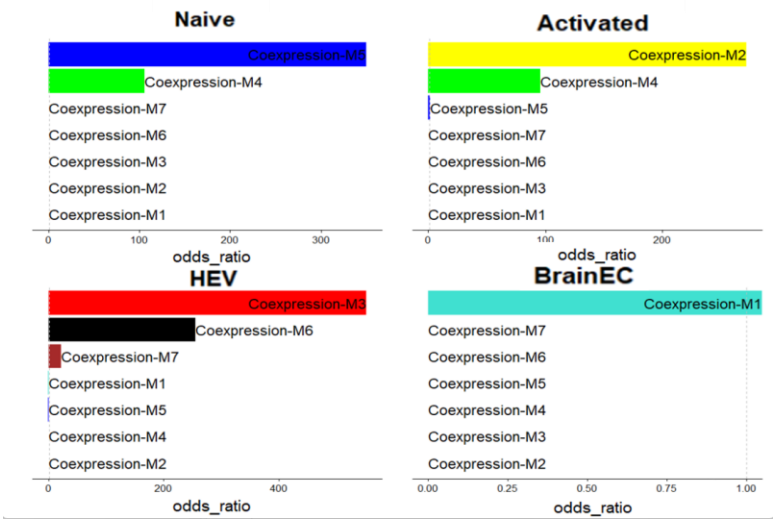


Figure 7-2: Cell-type classification of modules and their overlap

The below plot shows the correlation between different modules (Fig.7-3).

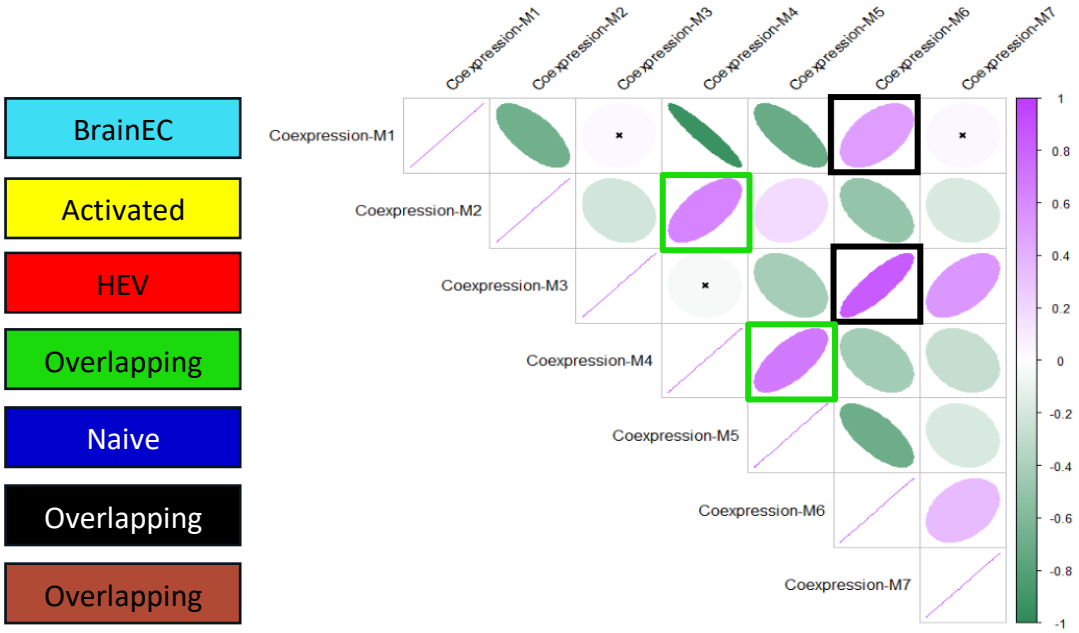


Figure 7-3: Module Correlogram of different modules and their corresponding cell-types

The top 10 enriched terms from the enrichment results did not yield much information. Hence functional annotation clustering was performed in DAVID, an online bioinformatics tool, for all the modules to examine the GO terms associated with all the ontologies.

The correlation plot shows that module 4 has overlapping genes between modules 2 and 5. Further investigation into this module yielded genes mainly responsible for ribosome formation, genes involved in mitochondria etc. These were not relevant to our study hence they were not pursued.

The functional annotation clustering results were searched for any terms associated with the extravasation process. The table below details the GO term, the associated module and the genes involved in the subcellular location or process.

GO Term	Module	Genes
Cell adhesion	HEV	Cd24a, Col15a1, Dsg2, Glycam1, Itga2, Icam1, Jam3, Mfge8, Nectin2, Pxdn, Pcdh12, Pcdh17, Pcdh7, Sulf1, Tgfbi
Cell-cell adhesion	HEV	Dsg2, Icam1, Jam3, Itga2, Tmem47
Regulation of extravasation	HEV	Icam1, Jam3, Plvap
Inflammatory response	HEV	Anxa1, Ackr1, Chst1, Chst4, Ccl21a, Cxcl1, F2r, Fut7, Ly96, Nrros, Nfkbiz, Pf4, Ptgs1, Serpinb1a, Sphk1
Caveola	HEV	Aqp1, Cavin1, Cavin3, F2r, Plvap, Tfpi
Caveola	Brain	Atp1a2, Atp1b3, Ehd2, Jak2, Tsc2, Adrb2, Bmpr2, Cdh13, Ctsb, Cav1, Cav2, Dlc, Flot1, Gnaq, Insr, Igf1r, Lipe, Ldlr, Lrp6, Lrp8, Mapk1, Mapk3, Myof, Nos3, Ptch1, Pld2, Plpp1, Pacsin2, Scarb1, Slc2a1, Slc27a1, Smpd2, S1pr1, Tgfbr2
Integrin binding	HEV	Itga2, Icam1, Jam3, Mmp14, Mfge8, Timp2, Tgfbi
Integrin binding	Brain	Cd151, Cd81, Cd9, F11r, Lyn, Ptk2, Vwf, Adam15, Adam17, App, Anxa7, Cxcl12, Cx3cl1, Dab2, Dst, Emp2, Fermt2, Fn1, Gsk3b, Itga1, Itga3, Itga6, Itgb1, Icam2, Jam2, Kdr, Lama5, Lamb2, Lgals8, Nf2, Nisch, Pxn, Ptn, P4hb, Pdia4, Sema7a, Tln1, Tln2, Tspan4, Utrn
Podosome	Brain	Asap1, Arhgef5, Sh3gl1, Wdr1, Afap1l1, Actb, Cttn, Hnrnpk, Ptpn12, Arhgef2, Svil, Tpm3, Tpm4, Vcl
Cell-cell junction	Overlapping module - 6	Cdh5, Ctnnd1, Gja1, Nck1, Pecam1
Cell surface	Overlapping module - 6	Cdh5, Ctnnd1, Pecam1, Hyal2, Pam, Ifitm3, Tgfbr3, Tnfrsf1a

Table 7-2: Cellular extravasation associated GO terms and respective modules

7.3 Discussion

The co-expression network analysis reveals plenty of insights about gene networks in different cell-types.

The markers involved in cell-cell adhesion (Fig.7-4), in regulation of extravasation (Fig.7-5) and in inflammatory response (Fig.7-6) exhibit HEV cell-type specificity.

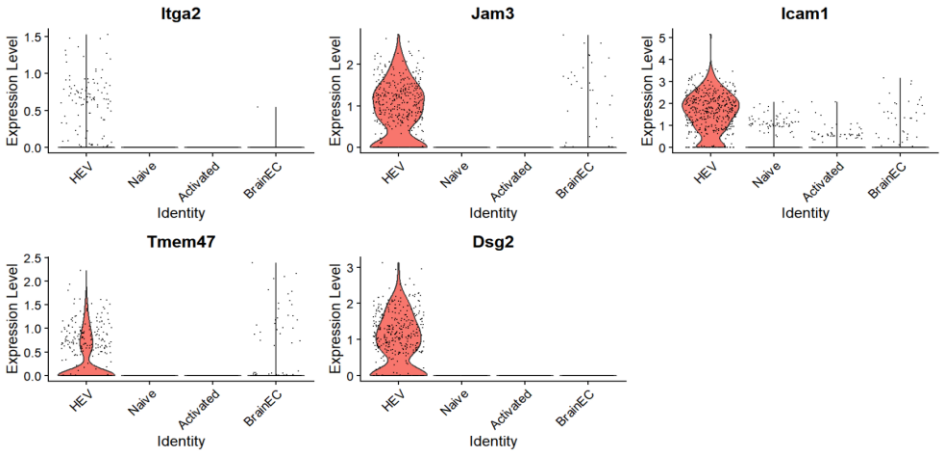


Figure 7-4: Genes involved in cell-cell adhesion

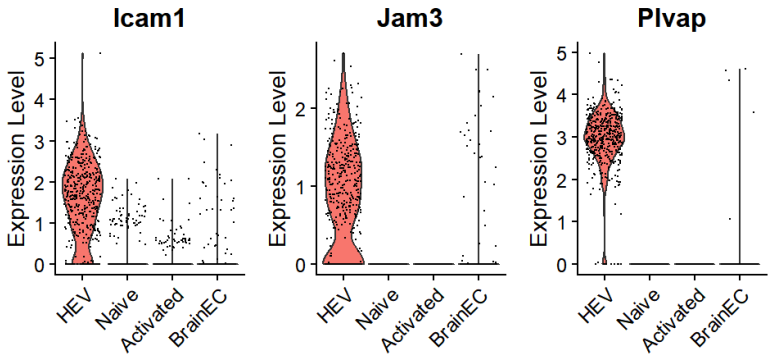


Figure 7-5: Genes involved in the regulation of extravasation

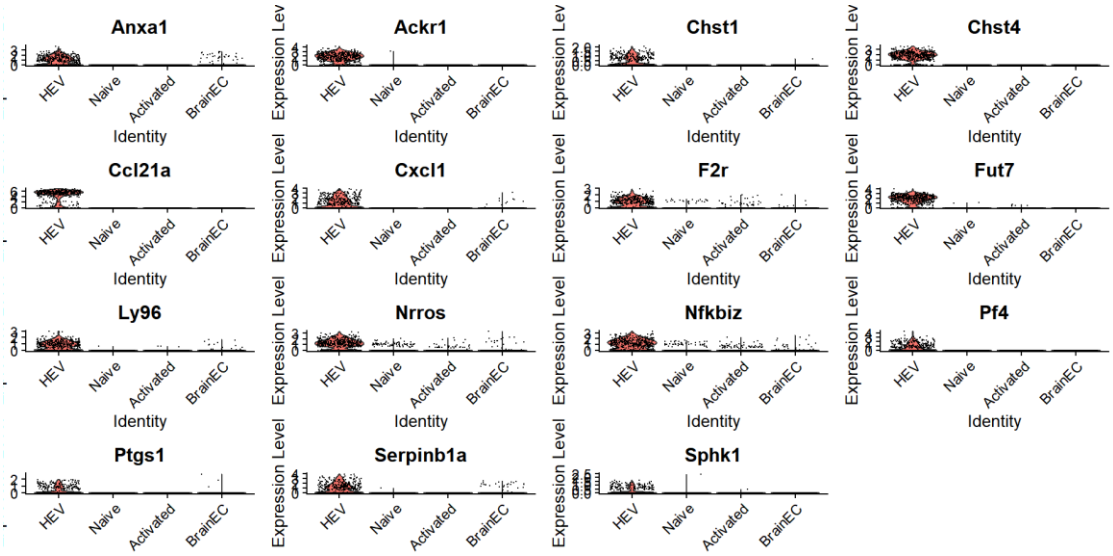


Figure 7-6: Genes involved in inflammatory response

Among the genes associated with the inflammatory response, it's worth noting that Fut7, Chst4, and Ccl21a are recognized for their roles in lymphocyte trafficking within HEV[97]. Interestingly, Ackr1, although traditionally not expressed in the brain, has recently been shown to exhibit increased expression of its protein in an in-vitro model involving primary mouse brain microvascular endothelial cells (pMBMECs). This upregulation of Ackr1 appears to promote transcellular diapedesis, highlighting its potential significance in certain brain-related processes[109].

The genes that play a crucial role in the transcellular diapedesis within the 'Caveola' structure are entirely distinct in HEV compared to the brain. Figure 7-7 illustrates the genes active within the Caveola in HEV, and notably, they display cell-type specificity. However, it's interesting to observe that the fundamental genes responsible for forming the Caveola, namely Cav1 and Cav2, are part of the brain module and exhibit similar expression patterns in both HEV and the brain, depicted in figure 7-8.

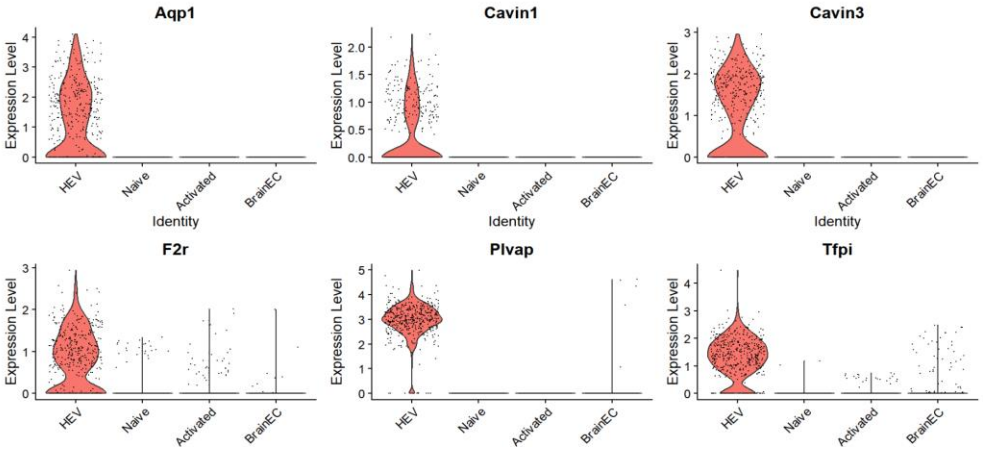


Figure 7-7: Genes active in caveola in HEV

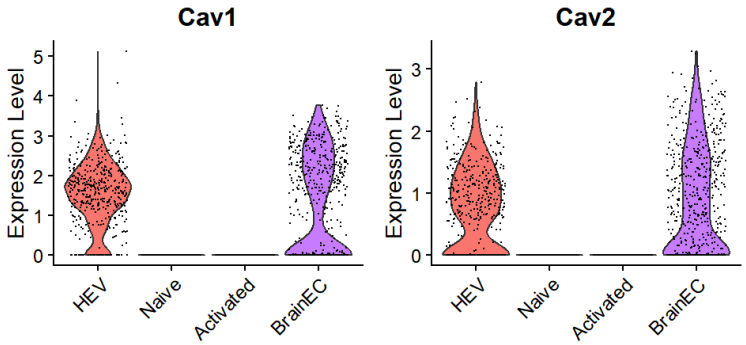


Figure 7-8: Caveola genes expressed similarly in HEV and Brain

Cavin1 and CAV1 have been identified as essential components that must work together to create a fundamental core system responsible for forming caveolae on the cell's outer membrane[89]. The Cavin protein family is known to assemble into tissue-specific caveolar complexes[110]. However, what makes the brain particularly intriguing is that none of the Cavin proteins appear to play a crucial role in caveola formation there. Remarkably, the cellular compartment in the brain hosts over 30 different genes, in contrast to the limited presence of just six genes in the context of HEV. This diversity in gene content within these cellular compartments adds to the complexity of their functional roles.

A parallel pattern is discernible concerning integrin binding function, wherein the brain demonstrates a more intricate landscape with the involvement of 40 genes, in stark contrast to the relatively simpler scenario in HEV, where only seven genes play a role in this capacity. This discrepancy underscores the heightened complexity and diversity of integrin-related processes within the brain's cellular environment.

Module 6, distinguished by its shared genes from both BrainEC and HEV modules and a strong correlation with them, encompasses Gene Ontology (GO) terms relevant to cellular extravasation. One notable example is the presence of genes expressed at cell-cell junctions, as demonstrated in Figure 7-9, where their similar expression patterns in both cell types are evident. Among these genes, the inclusion of Pecam1 within cell-junctions is noteworthy, as it plays a well-established role in paracellular diapedesis. Additionally, the Nck family of adaptor proteins is known to be recruited to cell-cell

junctions by Pecam1, particularly in response to oxidative stress[111]. Moreover, Cdh5 and Gja1, identified as VE-Cadherin and Gap junction alpha protein-1, respectively, are recognized components of junctions and hold pivotal roles in facilitating paracellular diapedesis and cellular transport processes[112].

Gene expression levels for cell-surface markers exhibit a similar pattern or consistency (Fig.7-10).

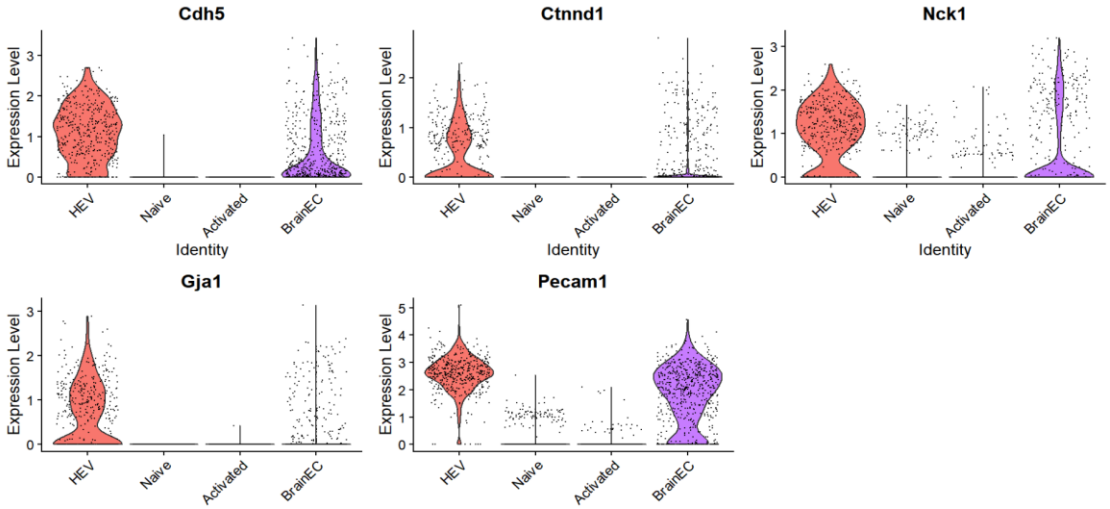


Figure 7-9: Genes expressed at cell-cell junctions similarly expressed in BrainEC and HEV

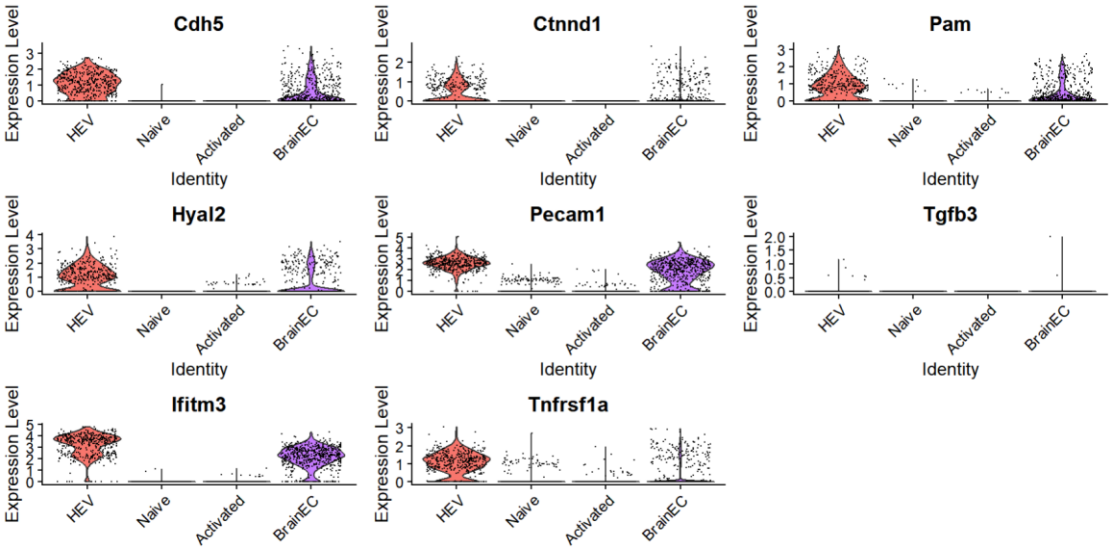


Figure 7-10: Genes expressed on cell-surface in BrainEC and HEV

Chapter 8

UNRAVELING CELL-CELL INTERACTIONS USING PREDICTIVE ALGORITHM

Understanding cell surface proteins offers valuable insights into intercellular communication among diverse cell types. This communication can manifest through direct ligand-receptor interactions, involving cell-to-cell contact, interactions with the extracellular matrix, or the secretion of proteins that bind to receptors on target cells. Previously, deciphering these interactions necessitated biochemical assays like proximity labeling, co-immunoprecipitation, and yeast two-hybrid screening[113]. However, the emergence of single-cell transcriptomics has led to the development of various tools for predicting ligand-receptor interactions.

Tools built to predict intercellular communication are based on existing literature knowledge as well as other databases including UniProt, KEGG, IntAct, STRING and Reactome[114]. This domain faces an additional challenge due to the absence of a definitive reference dataset[115,113] which can adequately represent the intricate and ever-changing interactions among numerous cells and molecules. Nevertheless, independent assessments have demonstrated that Cell-Cell Communication (CCC) approaches exhibit a notable resilience to noise introduction[116,117]. For our analyses we used three tools designed for CCC that contained curated database for mouse ligand receptor interactions: CellChat, Cellinker and CellTalkDB.

CellChat, one of the popular tools used for predicting CCC is developed to infer visualize and analyze intercellular communications from scRNA-seq data[118]. CellChat begins by creating a comprehensive database of known signaling molecule interactions, considering various structural aspects like ligand-receptor complexes and soluble molecules. It then uses this database to model the probability of CCC in the scRNA-seq data. This inference is based on mass action models, differential expression analysis, and statistical tests for different cell groups, which can be discrete or continuous states along a pseudotime cell trajectory. It quantitatively characterizes and compares these

inferred intercellular communications using social network analysis, pattern recognition methods, and manifold learning approaches. The majority of ligand–receptor interactions in CellChatDB were manually curated based on KEGG (Kyoto Encyclopedia of Genes and Genomes) signaling pathway database and contains 2021 curated ligand-receptor interactions in mouse and human.

Cellinker is a platform designed for the analysis of intercellular communication facilitated by ligand-receptor (L-R) interactions. It offers a manually curated database of L-R interactions, with over 3700 in humans and 3200 in mice, along with 400 endogenous small molecule-related L-R interactions. Additionally, Cellinker provides a webserver for conducting intercellular communication analysis using single-cell RNA sequencing (scRNA-seq) data[119].

CellTalkDB is a curated database housing 3,398 human and 2,033 mouse ligand-receptor (LR) pairs. These LR pairs were initially sourced from the STRING database and subsequently subjected to manual verification through extensive database searches and literature review[120].

8.1 Materials & Methods

CCC analysis using CellChat was conducted in R version 4.2.2. The packages used for the analysis are listed below:

1.	Seurat – 4.3.0
2.	CellChat – 1.6.1
3.	ggplot2 – 3.4.2
4.	igraph – 1.4.2
5.	patchwork – 1.1.2

Table 8-1: List of packages used for CCC analysis

For this analysis, we configured the ligand-receptor database to 'mouse.' We initiated the process by identifying overexpressed genes and their interactions through the 'identifyOverExpressedGenes()' and 'identifyOverExpressedInteractions()' functions

within the CellChat package. Subsequently, we inferred the cell-cell communication network utilizing the 'computeCommunProb()' function. Finally, we quantified the aggregated cell-cell communication network by either tallying the number of links or summarizing the communication probabilities.

Furthermore, we leveraged Cellinker's online portal to estimate communication probabilities for the cell-type markers identified during the differential expression analysis. Additionally, we downloaded CellTalkDB's database for mouse ligand-receptor interactions and meticulously verified interactions between distinct cell types.

8.2 Results

The communication probability was computed for all the cell-types. The pathway level interactions from endothelial cells to T-cells were classified as shown in Table 10 and the interactions from T-cells to endothelial cells are shown in Table 11 with the communication probability.

	source	target	pathway_name	probability
1	Activated	BrainEC	LT	0.000392
2	Activated	BrainEC	MIF	0.0156
3	Activated	BrainEC	VISFATIN	0.00086
4	Activated	HEV	CCL	0.009591
5	Activated	HEV	LT	0.007531
6	Activated	HEV	SEMA4	0.000899
7	Activated	HEV	TGFb	0.002933
8	Activated	HEV	VISFATIN	0.002698
9	Naive	BrainEC	ITGAL-ITGB2	0.009924
10	Naive	BrainEC	MIF	0.007578
11	Naive	BrainEC	SELL	0.124645
12	Naive	HEV	CD226	0.004877
13	Naive	HEV	ITGAL-ITGB2	0.021233
14	Naive	HEV	SELL	0.203917
15	Naive	HEV	SEMA4	0.001792
16	Naive	HEV	TGFb	0.001184
17	Naive	HEV	VCAM	0.006098

Table 8-2: Pathways mapped to cell interactions from T cells to endothelial cells

	source	target	pathway_name	probability
1	BrainEC	Activated	COLLAGEN	0.004079
2	BrainEC	Activated	FN1	0.004655
3	BrainEC	Activated	GALECTIN	0.053638
4	BrainEC	Activated	LAMININ	0.001173
5	BrainEC	Activated	MHC-I	0.347784
6	BrainEC	Activated	PTN	0.132747
7	BrainEC	Naive	CXCL	0.009496
8	BrainEC	Naive	FN1	0.03287
9	BrainEC	Naive	GALECTIN	0.104349
10	BrainEC	Naive	ICAM	0.009924
11	BrainEC	Naive	JAM	0.010805
12	BrainEC	Naive	MHC-I	0.731166
13	BrainEC	Naive	PTN	0.084995
14	HEV	Activated	CCL	0.092637
15	HEV	Activated	COLLAGEN	0.004543
16	HEV	Activated	GALECTIN	0.02811
17	HEV	Activated	ICAM	0.014793
18	HEV	Activated	LAMININ	0.001468
19	HEV	Activated	MHC-I	0.6841
20	HEV	Naive	CCL	0.299084
21	HEV	Naive	CXCL	0.000273
22	HEV	Naive	GALECTIN	0.054958
23	HEV	Naive	ICAM	0.04603
24	HEV	Naive	JAM	0.006629
25	HEV	Naive	MHC-I	1.441405
26	HEV	Naive	NECTIN	0.004877

Table 8-3: Pathways mapped to cell interactions from endothelial cells to T cells

The interactions between different proteins can be visualized in the plots shown from the below plots.

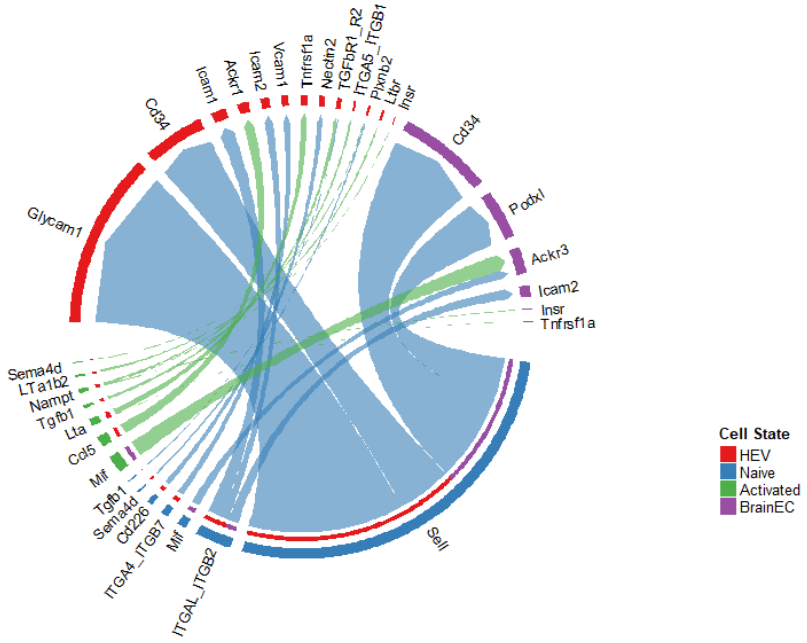


Figure 8-1: CCC from T-cells to endothelial cells

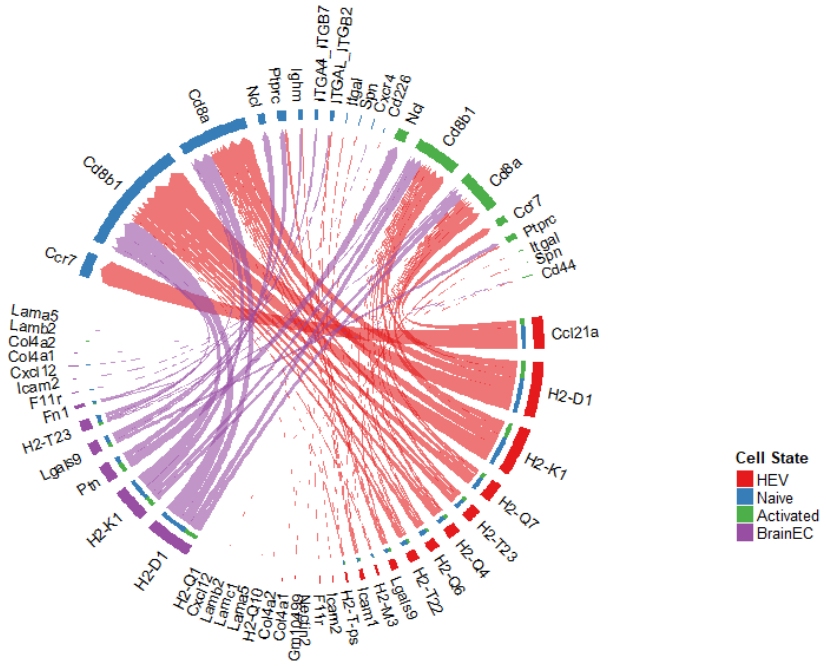


Figure 8-2: CCC from endothelial cells to T-cells

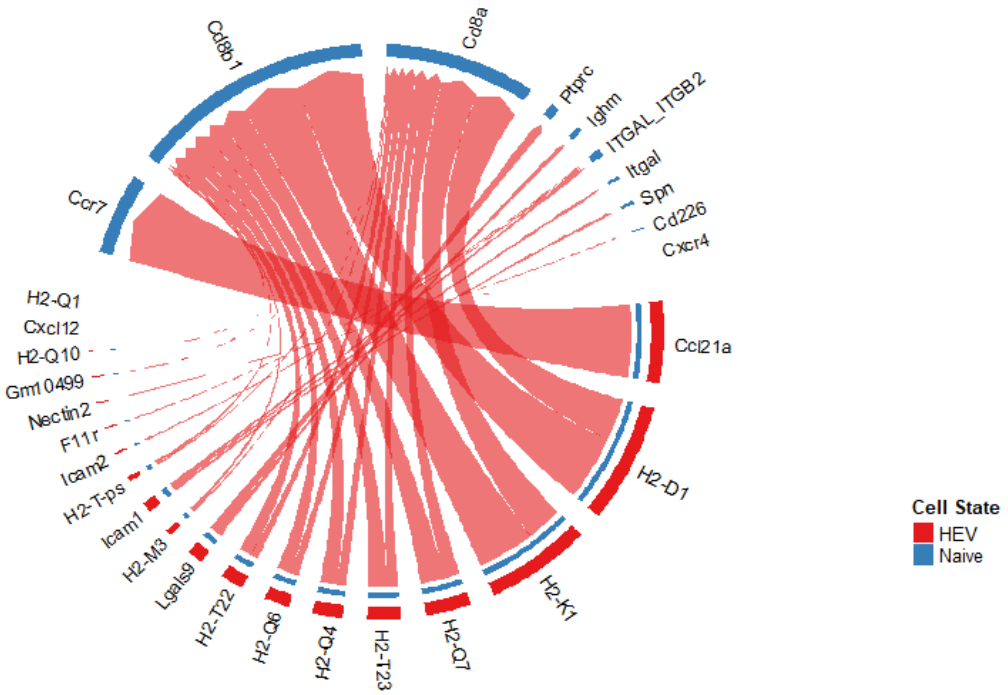


Figure 8-3: CCC from HEV to Naïve T cells

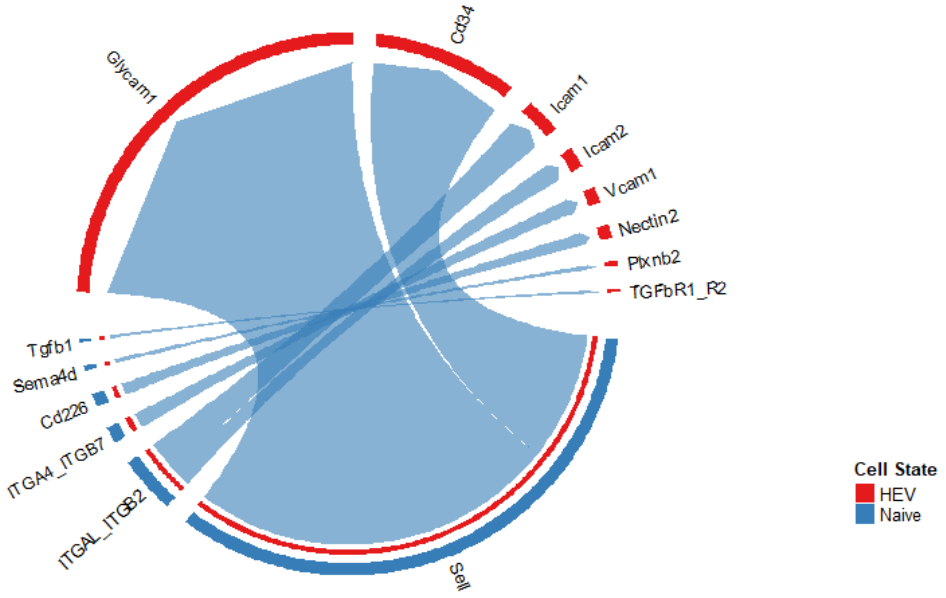


Figure 8-4: CCC from Naïve T cells to HEV

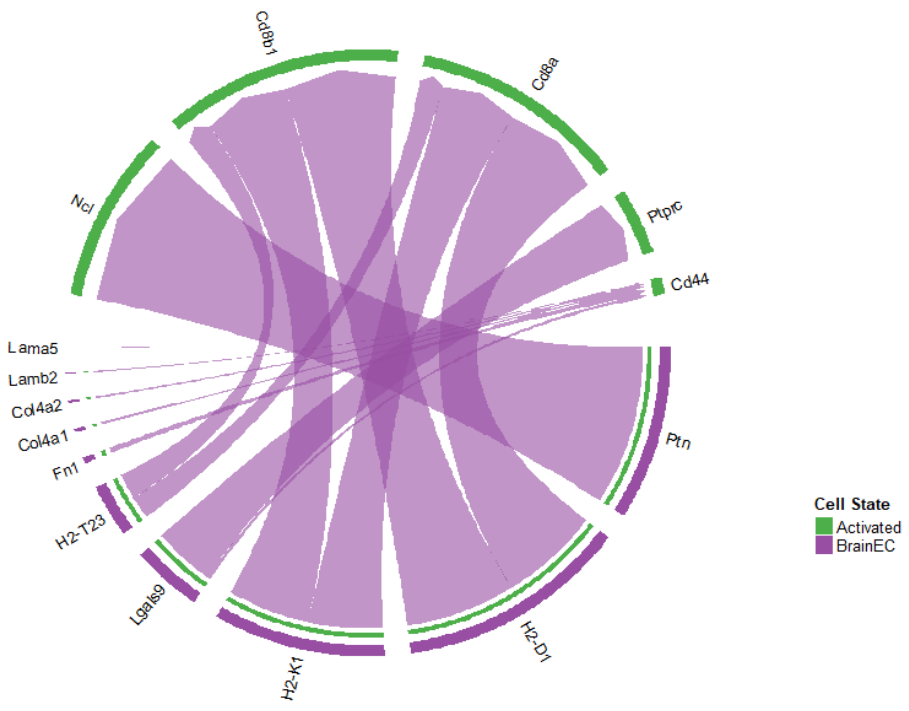


Figure 8-5: CCC from BrainEC to Activated T cells

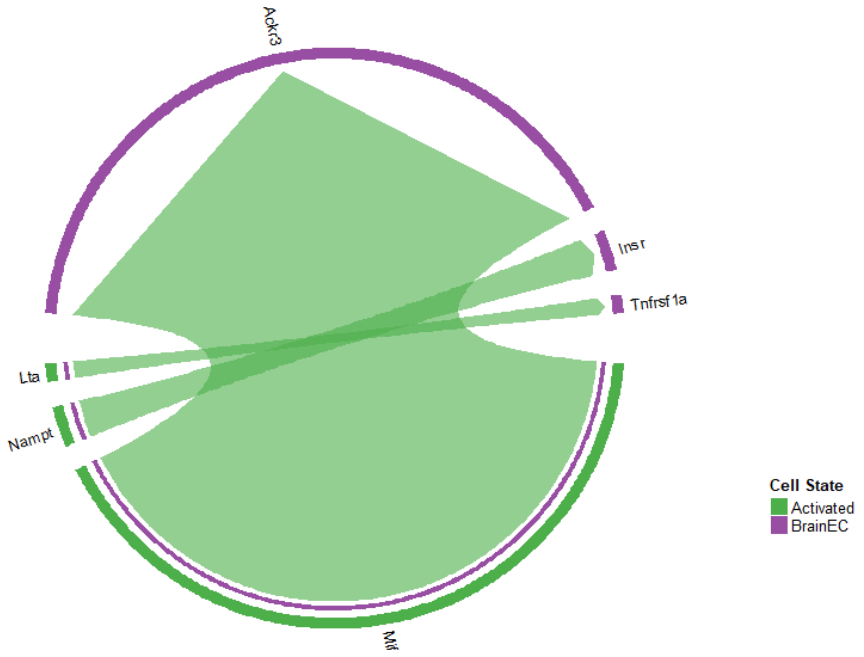


Figure 8-6: CCC from Activated T cells to BrainEC

The unique interactions were manually curated and are summarized in Fig 8-7. The figure also shows the type of signaling between the proteins.

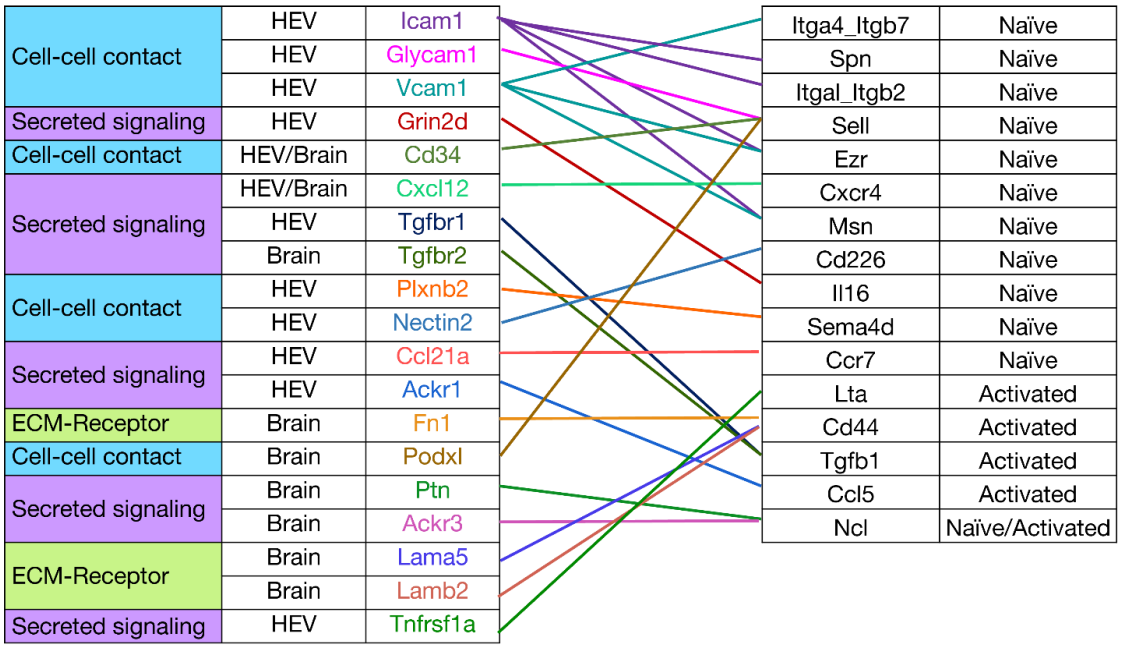


Figure 8-7: Mapping of genes (proteins) with their binding partners

8.3 Discussion

This analysis has shed light on intercellular interactions and their various modes involving numerous molecules. We now have a deep insight into many proteins involved in the cellular extravasation process. Fig 8-7 shows the endothelial cells on the left and their binding partners on the right.

While the binding of Glycam-1 with Sell (L-selectin) is well-established, our prediction of interactions between integrins Icam1 and Vcam1 and the genes Ezr and Msn prompted us to further investigate these genes. During our differential expression analysis, we observed significant upregulation of Ezr and Msn in naïve T cells. Additionally, these genes were enriched in Gene Ontology terms related to cellular components such as 'uropod,' 'cell-substrate junction,' 'cell-trailing edge,' and 'focal adhesion.' They are part of the ezrin, radixin, and moesin (ERM) family of closely related proteins. Specifically, Ezr (Ezrin) is known to be involved in protein binding

and cell-adhesion molecule binding and plays a role in leukocyte trans-endothelial migration[121]. These genes are also known to co-localize with Cd44[122], which was significantly upregulated in activated T cells, suggesting their potential involvement in immune responses in the brain. The violin plot (Fig.8-8) shows their expression levels in different cell-types.

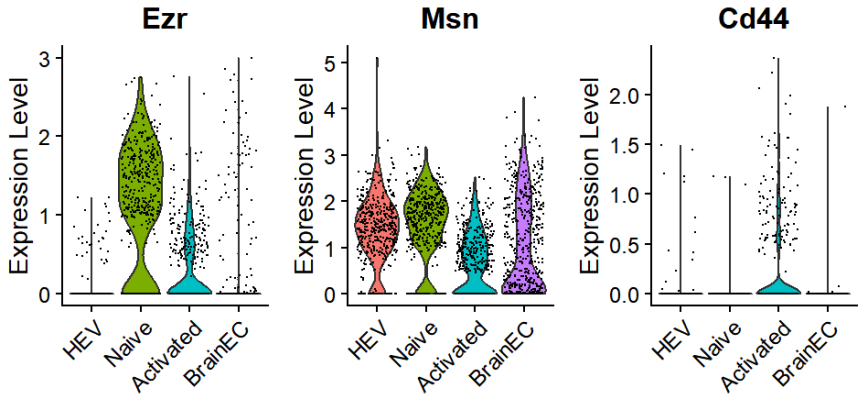


Figure 8-8: Expression levels of Ezr, Msn and Cd44

Ackr1 belongs to a family of Atypical Chemokine Receptors typically located in endothelial cell junctions, where they exert regulatory roles in immune and inflammatory responses[123]. Our co-expression analysis indicated its involvement in the inflammatory response in HEV, which aligns with existing literature. Ackr1 is known to bind chemokines from the CXC and CC groups[124], consistent with CCC analysis predictions regarding its interaction with Ccl5. Interestingly, Ackr1 is reported to promote transcellular diapedesis across the blood-brain barrier (BBB) during neuroinflammation in brain in vitro models[109]. Notably, Ackr1 exhibits predominant expression in HEV and is entirely absent in the brain (Fig. 8-9). This intriguing finding warrants further exploration.

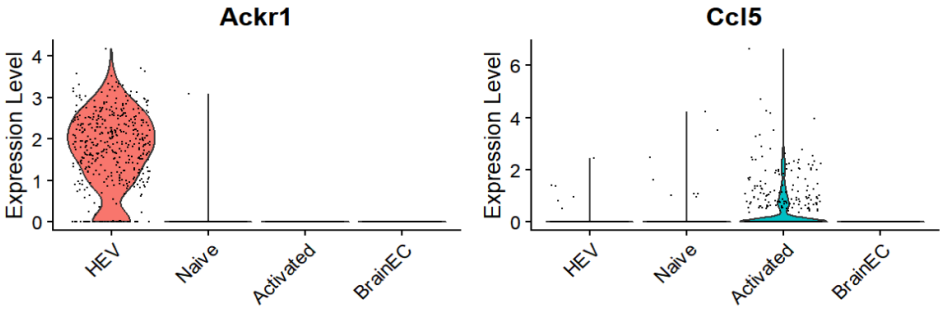


Figure 8-9: Expression levels of Ackr1 and Ccl5

Lama5 and Lamb2 are members of the Laminin family, integral components of the extracellular matrix. Alterations in their expression have been associated with inflammatory processes, and they play pivotal roles in modulating leukocyte activation and migration[125]. A recent study has has proposed that laminins may serve as reliable indicators of blood-brain barrier (BBB) structural integrity. This is because during neuroinflammation, the structural integrity of laminins may be compromised, consequently leading to increased BBB permeability[126]. The expression levels of laminins and its binding partner Cd44 are shown in Fig.8-10. Their interaction with Cd44 indicates that they play an active role in inflammatory response.

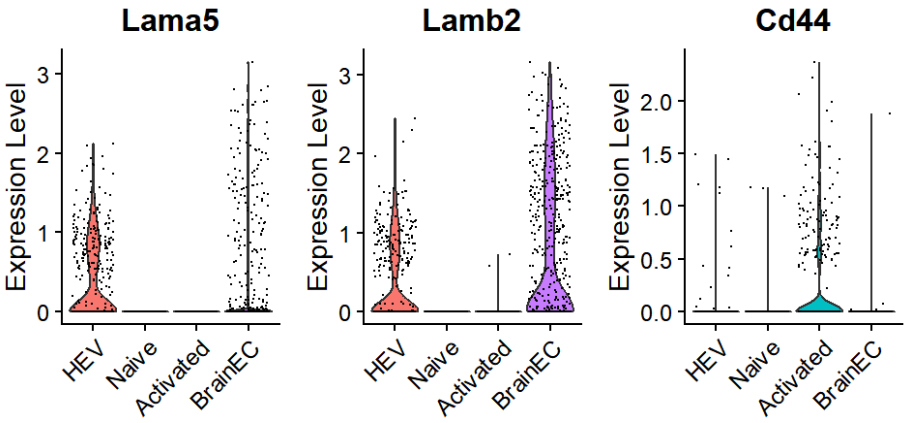


Figure 8-10: Expression levels of laminins and CD44

Chapter 9

PIPELINE APPLICATION: IN-DEPTH ANALYSIS OF INTERACTIONS BETWEEN ACTIVATED ENDOTHELIUM IN BBB AND ACTIVATED T CELLS

In our previous analysis, we used cells from the control group as we were conducting a comparative study with HEV. We wanted to understand the distinct interactions under homeostasis. After successfully identifying unique interactions between HEV and Naïve T cells and between BBB and Activated T cells, we wanted to further confirm our analysis in the BBB. It has been established that activated T cells are able to breach inflamed BBB. Therefore, for this analysis, we focused on analyzing the interactions between the inflamed BBB and activated T cells.

9.1 Method

In September 2023, we conducted a systematic search of the GEO repository using various keyword combinations. The datasets used for the analyses originated from in vivo experiments conducted on mice. The search criteria provided below represent the specific terms we employed within the NCBI portal to retrieve results when used in different combinations.

Search details for extracting injured brain endothelial cells:

- (("brain"[MeSH Terms] OR brain[All Fields]) AND ("endothelium"[MeSH Terms] OR endothelial[All Fields]) AND ("single person"[MeSH Terms] OR single[All Fields]) AND ("cells"[MeSH Terms] OR cell[All Fields]) AND ("mice"[MeSH Terms] OR "Mus musculus"[Organism] OR mus musculus[All Fields])) AND "Mus musculus"[porgn]
- scRNA seq[All Fields] AND ("brain"[MeSH Terms] OR brain[All Fields]) AND ("endothelium"[MeSH Terms] OR endothelial[All Fields])
- (("blood-brain barrier"[MeSH Terms] OR blood brain barrier[All Fields]) AND (("single person"[MeSH Terms] OR single[All Fields]) AND ("cells"[MeSH

Terms] OR cell[All Fields]) AND ("rna"[MeSH Terms] OR RNA[All Fields]))) AND "Mus musculus"[porgn]

- ("blood-brain barrier"[MeSH Terms] OR blood brain barrier[All Fields]) AND scRNA seq[All Fields]

Search details for extracting Activated T cells

- (("cd4-positive t-lymphocytes"[MeSH Terms] OR CD4 t cell[All Fields]) AND activated[All Fields] AND (("single person"[MeSH Terms] OR single[All Fields]) AND ("cells"[MeSH Terms] OR cell[All Fields]) AND ("rna"[MeSH Terms] OR RNA[All Fields]))) AND "Mus musculus"[porgn]
- (("cd8-positive t-lymphocytes"[MeSH Terms] OR CD8 t cell[All Fields]) AND activated[All Fields] AND (("single person"[MeSH Terms] OR single[All Fields]) AND ("cells"[MeSH Terms] OR cell[All Fields]) AND ("rna"[MeSH Terms] OR RNA[All Fields]))) AND "Mus musculus"[porgn]
- (("cd8-positive t-lymphocytes"[MeSH Terms] OR CD8 t cell[All Fields]) AND (("single person"[MeSH Terms] OR single[All Fields]) AND ("cells"[MeSH Terms] OR cell[All Fields]) AND ("rna"[MeSH Terms] OR RNA[All Fields]))) AND "Mus musculus"[porgn]

The results underwent meticulous curation, with a focus on scrutinizing the experimental designs. Our selection criteria prioritized datasets related to injured or inflamed BBB, allowing us to delve into the expression patterns of marker genes under inflammatory conditions and explore their interactions with activated T cells.

Initially, we assessed a total of 18 datasets, comprising 10 datasets derived from experiments involving brain endothelial cells and 7 from activated T cells. Notably, one dataset (GSE199460) encompassed both endothelial and activated T cells. This dataset was sourced from a single-cell transcriptomic analysis of the brain in an Experimental Autoimmune Encephalomyelitis (EAE) model, which serves as an animal model for Multiple Sclerosis. The study aimed to identify inflammatory networks within the endothelium. Brain tissue was collected from mice at the peak of the disease and from

control (CTRL) mice. Single cells were isolated using an optimized protocol based on the Miltenyi Brain Dissociation Kit and then subjected to single-cell RNA sequencing analysis[127]. The dataset included samples from 3 replicates of control and EAE mice, encompassing a total of 15 cell types.

The control population was characterized by a limited presence of T cells, with a total of 35 T cells across all replicates. In contrast, the experimental group exhibited a substantial increase, featuring more than 2500 T cells and over 1100 endothelial cells. This stark contrast presented an exceptional opportunity to investigate the interactions between these two distinct cell types.

After identifying the dataset for analysis, our initial step involved a thorough data quality check. Subsequently, we carefully selected a subset of endothelial cells from both the control and experimental groups. We also subsetted the activated T cells from the experimental group. Due to the limited number of T cells in the control population, we made the decision to utilize the naïve T cell dataset that had been employed in our previous analysis. We employed our pipeline to integrate all the cell-types.

After integration, we performed a differential expression analysis to identify markers in the inflamed and activated cell types, employing the MAST method. Initially, we executed this test using the same parameters as in the previous analysis, which led to just 2 positive control markers in the analysis between the control and inflamed BBB. As a result, we decided to adjust our parameters for this analysis. We set the log-fold change threshold (logfc) parameter to the default value of 0.25, allowing for a broader range of differentially expressed genes to be included. Furthermore, we modified the 'min.diff.pct' parameter to 0.10 to increase the number of identified markers. The ‘only.pos’ feature was set to ‘True’, leading to each analysis being conducted twice. Due to the limited number of differentially expressed genes observed between the control and EAE BBB, we performed the enrichment analysis online using DAVID.

Lastly, we conducted a comprehensive cell-cell communication analysis between inflamed endothelial cells and activated T cells. This analysis aimed to unveil novel insights while reinforcing our earlier findings in the context of intercellular interactions.

9.2 Results

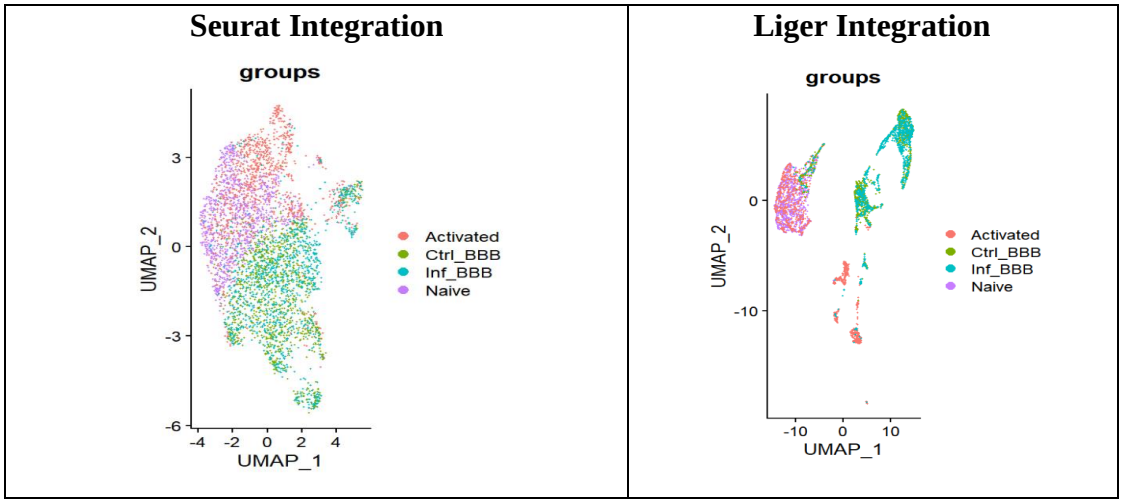
Data Integration

We evaluated the integration of 1058 cells in each cell-type (a total 4232 cells) across six algorithms. The median scores obtained using the algorithms are summarized in the table below:

	Silhouette Score (-1 to 1)	**Mixing Metric (0 to 300)**	**cLISI (1 to 4)**	**Local structure preservation score (0 to 1)**
Seurat	0.05	21	1.93	0.29
LIGER	0.05	60.5	1.66	0.33
Harmony	0.14	171.5	1.007	0.49
RPCA	0.041	25	1.46	0.58
scVI	-0.09	21	1.80	0.50
Scanorama	0.18	159.5	1.37	0.26

Table 9-1: Evaluation scores of different algorithms

The UMAP visualization of integration algorithms are shown in the figure below:



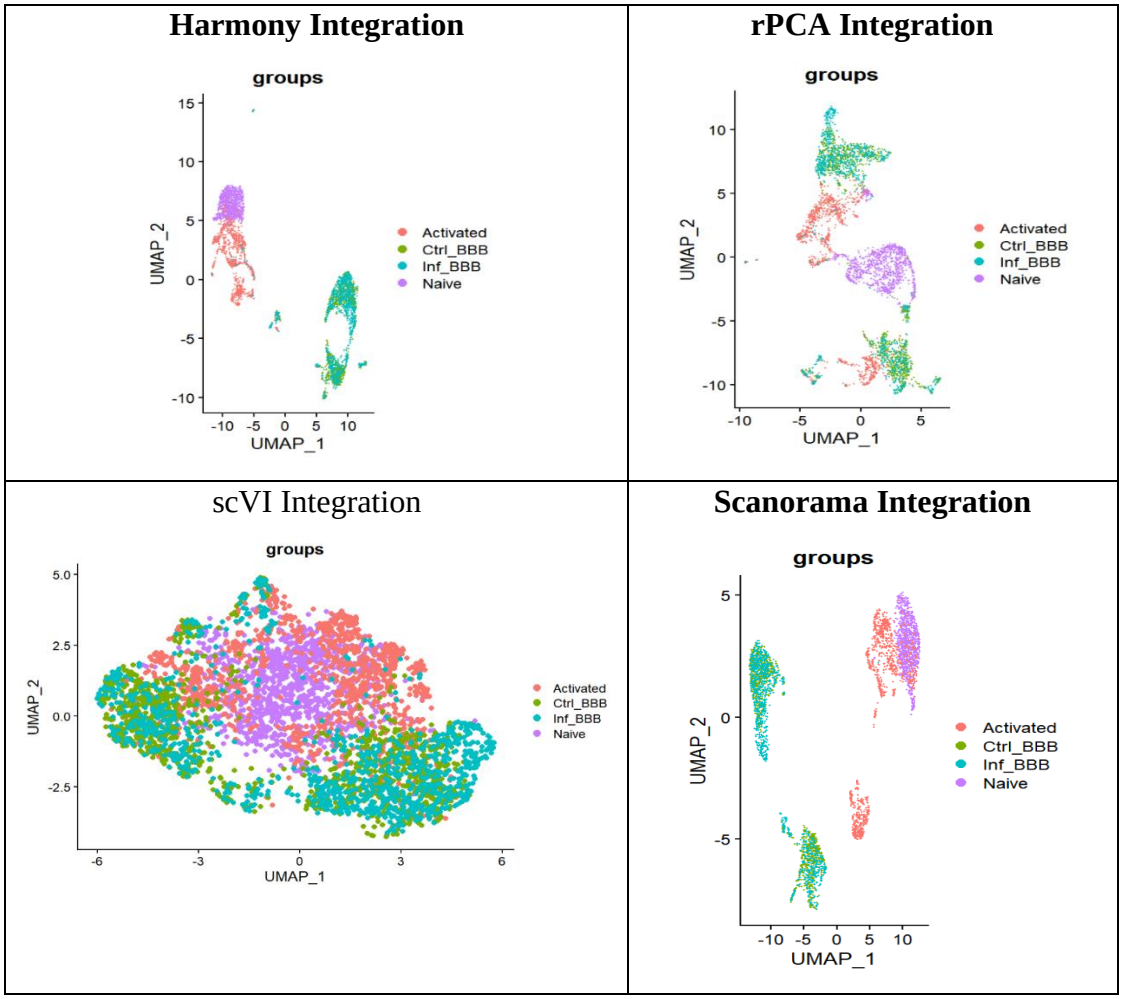


Figure 9-1: UMAP embeddings of integration of different algorithms

The table illustrates consistently low silhouette scores across all algorithms, indicating a suboptimal ability to form well-defined clusters, which aligns with the observations from the UMAP plot.

Harmony stands out as the highest performer in the mixing metric, achieving a score of 171.5, followed by Scanorama with a score of 159.5. In contrast, the performance of other tools is notably subpar. In terms of cLISI, Harmony once again excels as the top scorer with a score of 1.007. However, for local structure preservation, RPCA emerges as the leader with a score of 0.58, closely followed by scVI (0.50) and Harmony (0.49).

Summarily, Harmony demonstrates superior performance in integrating the data, making it the choice for downstream analyses. From the plot above, we can see that the control and inflamed BBB have formed one cluster indicating that they have similar gene expressions.

Differential Expression Analysis

The table below summarizes the number of upregulated genes obtained after differential expression analysis, number of GO terms associated with the cellular component and number of unique markers identified for each cell-type.

	Endothelial cells		**T-cells**	
	Control BBB	**Inflamed BBB**	**Naïve T-cells**	**Activated T-cells**
Upregulated genes	69	201	472	1409
GO: Cellular component	19	54	55	192
No. of unique cell-surface markers	13	111	67	620

Table 9-2: Summary of the differential expression and enrichment analysis results

The following violin plots display the upregulated markers within the inflamed BBB, shedding light on key findings. Notably, both Icam1 and Vcam1, pivotal in cell adhesion, exhibit significant upregulation in the inflamed BBB, thus affirming earlier research findings[19].

Moreover, a noteworthy observation relates to the heightened expression of Vwf, also known as von Willebrand factor. This glycoprotein is recognized for its role in reducing endothelium permeability, facilitating leukocyte recruitment[128], and weakening tight junctions[129] within the blood brain barrier during injury or inflammation.

Cd74, a regulator of macrophage inflammation and dendritic cell motility, and an active participant in the inflammatory response, displays upregulation in the context of activated T cells and the inflamed BBB, aligning with established scientific knowledge[130].

Ackr1, as mentioned in the co-expression analysis network section, is reported to be upregulated during inflammation in BBB, which is also reported in the study from where this dataset has been sourced[127].

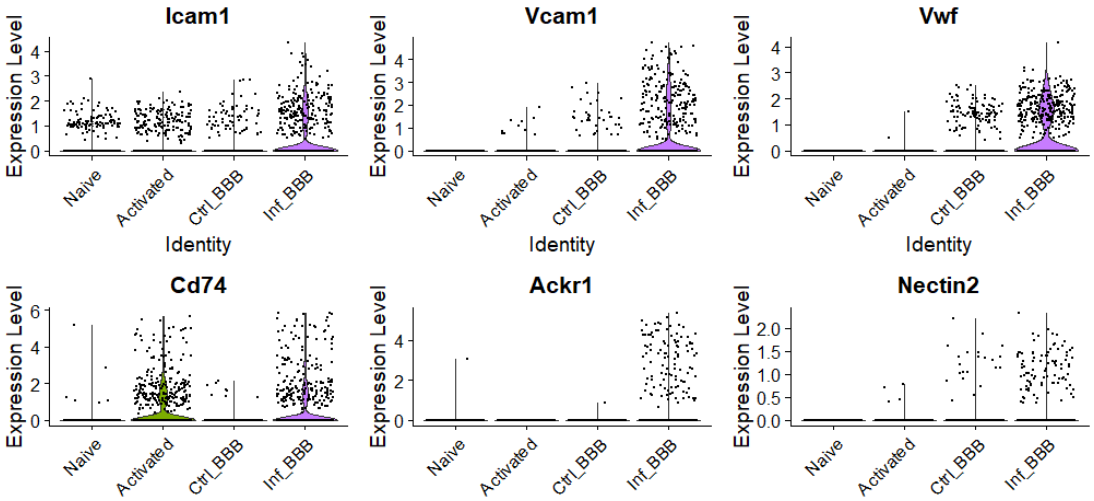


Figure 9-2: Upregulated markers in inflamed BBB

Nectin2, a cell-cell adhesion molecule residing in the adherens junction membrane, has been implicated in Alzheimer's disease and various cancer types[131,132]. It collaborates with or acts independently of cadherins and is known to participate in cellular signaling pathways[133]. A noteworthy aspect of Nectin2 is its binding to Cd226 on T-cells, a prediction that our cell-cell communication analysis has confirmed. Notably, our analysis reinforces the distinctive expression pattern of Nectin2, which exhibits high expression in HEV but minimal presence in the brain. This disparity is further underscored by examining the control BBB, which reveals a low expression level of this gene, validating our findings in depth. Furthermore, its expression levels exhibit an increase under inflammatory conditions, suggesting its potential utility as a valuable biomarker for future analyses.

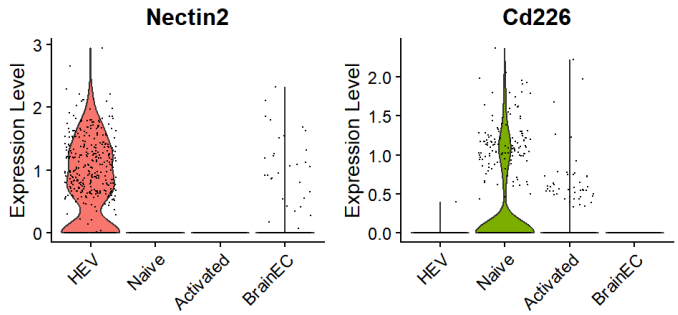


Figure 9-3: Expression of Nectin2 & Cd226 from previous comparative analysis

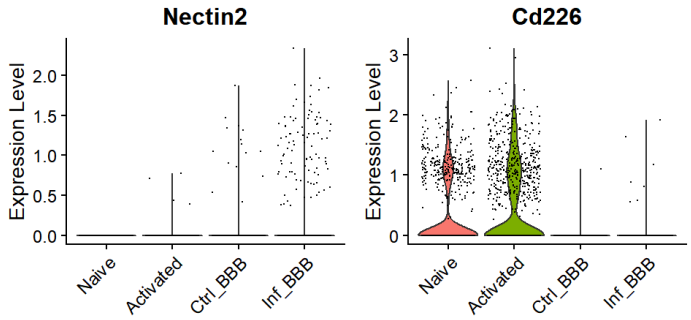


Figure 9-4: Expression of Nectin2 & Cd226 from the in-depth analysis

The below violin plot shows genes that are upregulated in activated T cells. These genes play an important role in immune regulation.

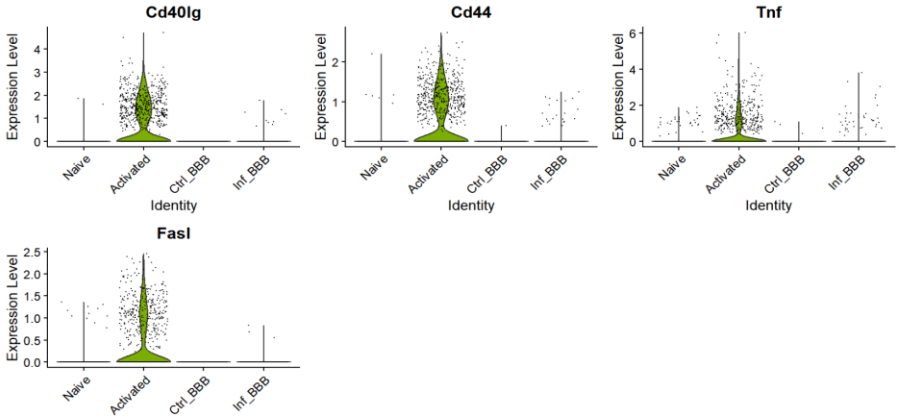


Figure 9-5: Expression of genes involved in immune response in activated T cells

Gene Co-expression Analysis

We performed a gene co-expression analysis to determine if the gene expressions in inflamed and control endothelial cells would be segregated into distinct modules. However, the algorithm assigned the gene expressions to just two modules, as depicted in the plots below. The modules were mapped to T cells and BBB cell-types.

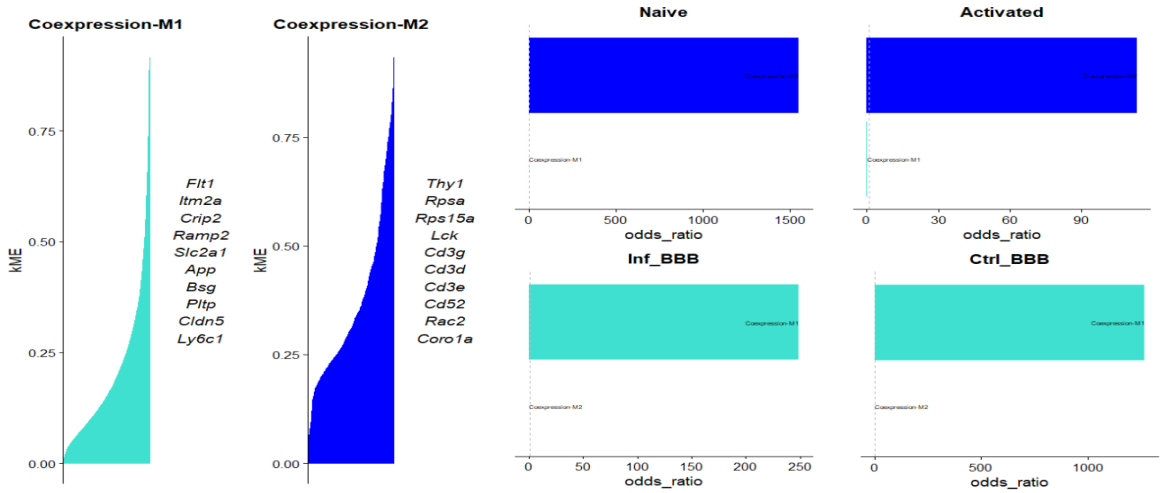


Figure 9-6: Co-expression modules and cell-type mapping of the modules

Interestingly, Icam1 was assigned to grey module in this analysis probably due to its varying expression levels between control and inflamed BBB.

Cell-Cell Communication Analysis

We performed the inter-cellular communication analyses between Inflamed BBB and Activated T cells to explore interactions that might not have been detected in our comparative study.

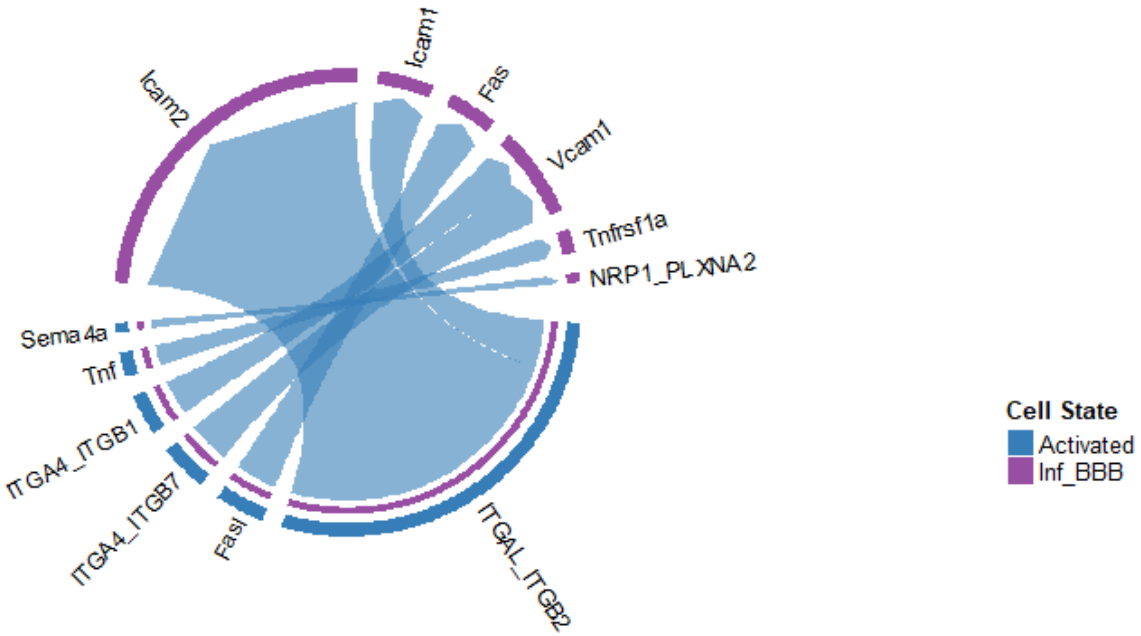


Figure 9-7: CCC from activated T cells to inflamed BBB

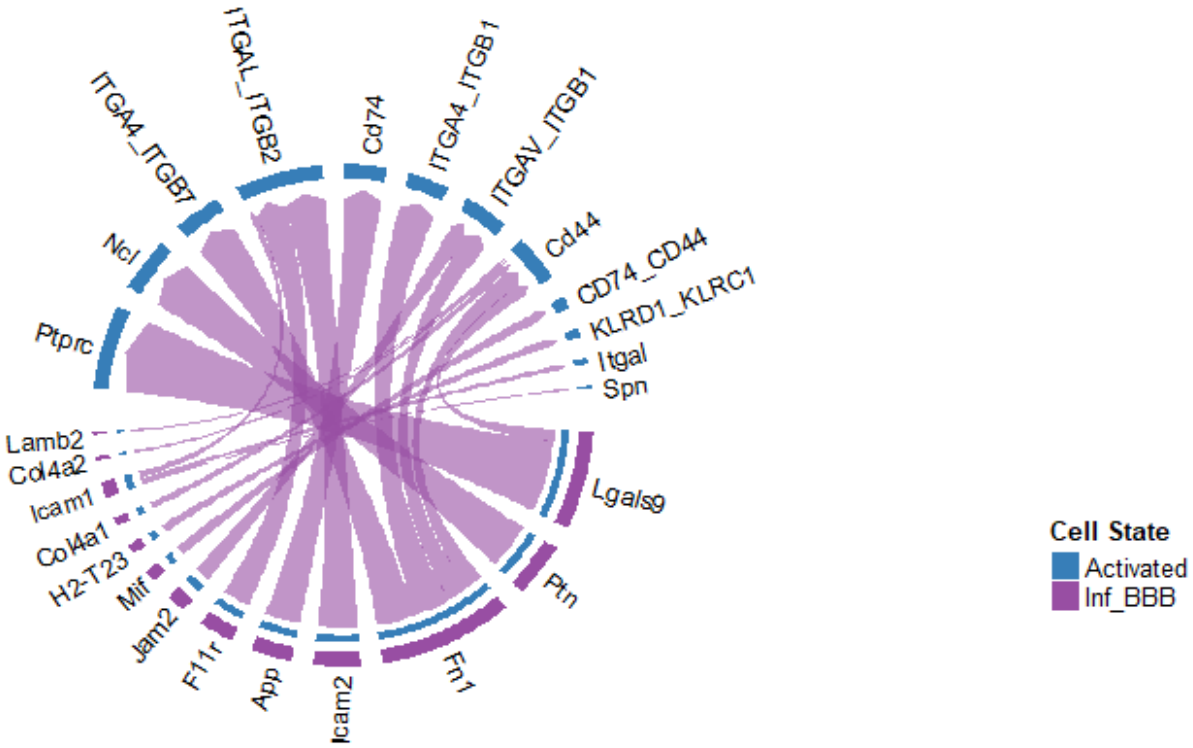


Figure 9-8: CCC from inflamed BBB to activated T cells

The unique interactions between inflamed BBB and activated T cells are summarized in the table below:

Signaling	Inf_BBB	Activated T cell
Secreted	Tnfrsf1a	Tnf
Secreted	Fas	Fasl
Cell-cell contact	Vcam1	Itga4_Itgb1
Cell-cell contact	NRP1_PLXNA2	Sema4a
Secreted	Mif	Cd74_Cd44
ECM - Receptor	Fn1	Itgav_Itgb1
Cell-cell contact	App	Cd74
Cell-cell contact	Jam2	Itgav_Itgb1
Cell-cell contact	H2-T23	KLRD1_KLRC1

Table 9-3: Mapping genes (proteins) with their binding partners

The algorithm has anticipated a higher number of interactions between inflamed BBB and activated T cells compared to the control BBB, as observed in our comparative study. Notably, the prediction of Icam1's interaction with Itgal, a connection previously observed only in HEV, is not unexpected, given its increased expression levels under inflammatory conditions. Additionally, we identified new interactions involving Fn1 with various proteins, such as Itga4_Itgb7, Itga4_Itgb1, and Itgav_Itgb1 in addition to Cd44, that was predicted earlier. Furthermore, we observed the activation of Jam2, Icam2, and App on the inflamed BBB.

Similarly, the elevated expression of Vcam1 results in interactions with Itga4_Itgb7 and Itga4_Itgb1. We also observed heightened expression levels of Fas ligand (Fasl), which is known to play a crucial role in activation-induced cell death (AICD) of T cells and cytotoxic T lymphocyte-induced cell death.

Semaphorins constitute a protein family known for their significant roles in both neuronal development and immune responses. They serve as crucial cues for guiding cell migration in immune processes[134]. Sema4a has gained recognition as a therapeutic target for the treatment of multiple myeloma, underscoring its clinical relevance[135]. Plxna2 has been reported to bind Sema3a[136]. Although existing evidence does not confirm a direct interaction between Plxna2 and Sema4a, this presents an intriguing opportunity to validate the potential interplay between these two proteins.

9.3 Discussion

There were some interesting observations during the in-depth analysis. The Cavin genes that are known to play a crucial supporting role in the caveola formation were completely absent in the brain in the comparative study as shown in fig. 9-9. However, the control population in the in-depth analysis dataset shows the contrary (fig 9-10). The brain dataset in the comparative study was from an experiment using mice that were 3 months old whereas the dataset from in-depth analysis does not mention the mice cell-line or age. This leads us to ponder if age plays a role in this gene expression.

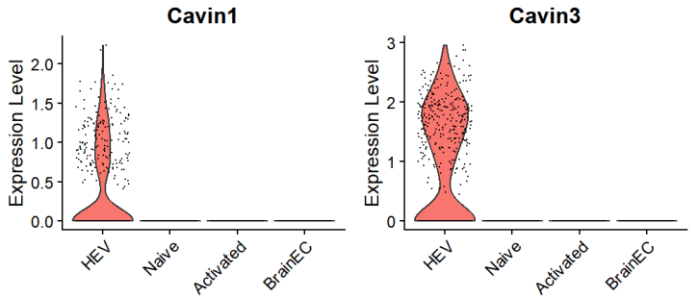


Figure 9-9: Cavin gene expression in the comparative study

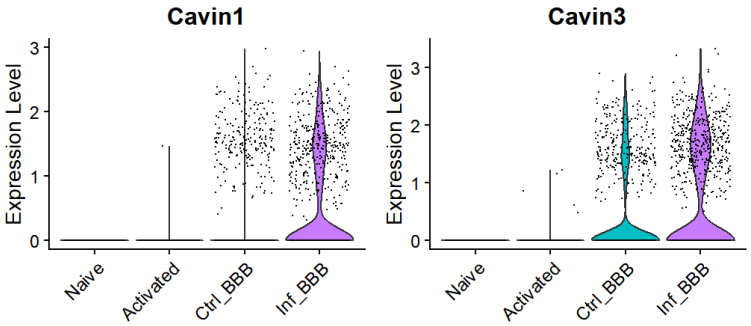


Figure 9-10: Cavin gene expression in the in-depth analysis

Similarly, the markers upregulated in the activated T-cells in this analysis (Fig.9-5), do not show similar expression levels in the activated dataset of the comparative study (Fig.9-11). In the dataset used in the comparative study, the T-cells were activated in-vitro as compared to the in-vivo activation of T-cells due to disease in the in-depth analysis. This discrepancy in the observed expression of the same genes could be attributed to different experimental conditions.

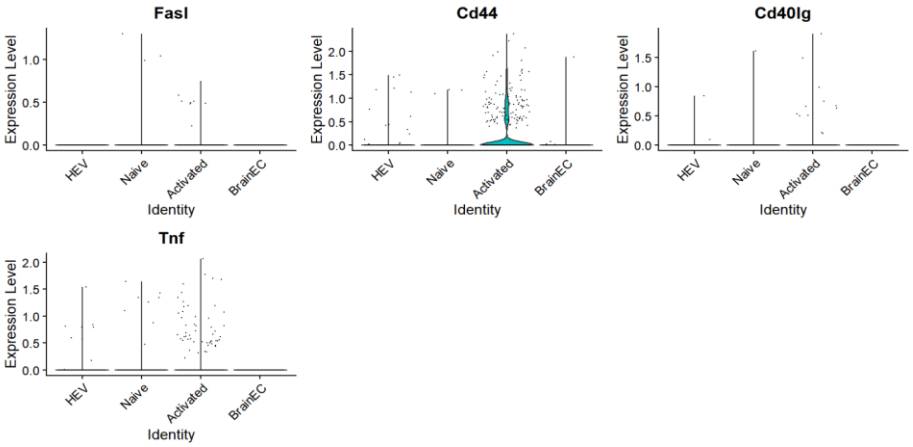


Figure 9-11: Gene expression levels of markers in activated T-cells in comparative study

Chapter 10

SUMMARY AND FUTURE DIRECTIONS

Our study was driven by the goal of unraveling the intricate molecular interactions that underlie lymphocyte trafficking during immune surveillance and immune responses across the endothelial barriers of HEV and BBB. It is well appreciated that cell-cell and protein-protein binding are complex processes[137,138,139,140] which are influenced by several factors[141,142,143]. To delve into this intricate process of cellular extravasation, we meticulously designed a robust pipeline that facilitated various analyses, enhancing our comprehension of this phenomenon within distinct tissues.

This comprehensive pipeline provided us with the capability to integrate data from diverse cell types and experimental conditions, fostering a holistic analysis. We evaluated six benchmarked algorithms for data integration, utilizing four key metrics to identify the most suitable integrated dataset for subsequent analyses. This pipeline is useful for integrating data from public repositories and harmonizing different batches generated from experiments.

Our differential expression analysis yielded the identification of well-established cell-type markers linked to various stages of the extravasation process. We confirmed the elevated expression of Mfsd2a in BrainEC, a key player in transcellular diapedesis. Despite its known plasma membrane localization, the absence of enrichment in the Cellular Component category is intriguing. Likewise, our analysis successfully pinpointed Glycam1 and Ccl21a, both essential for cell adhesion. Their lack of enrichment in the Cellular Component category suggests that their subcellular localization may extend beyond the cell surface.

Our co-expression analysis unveiled intricate gene networks associated with diverse functions, including cell adhesion and integrin binding. Furthermore, we effectively correlated these expression patterns with specific cell types. Notably, our analysis illuminated that processes occurring in the brain are more complex and distinct

involving the concerted action of numerous genes, in contrast to HEV, where fewer genes are involved in similar processes, as evidenced by caveola formation and integrin binding. Moreover, we successfully identified shared expression patterns between BBB and HEV by identifying overlapping modules.

Our cell-cell communication analysis resulted in identifying the unique binding partners of key players in the lymphocyte extravasation. We successfully validated the existing knowledge and unearthed novel candidates like Ezr and Msn, which play an important role in this process but were not previously mentioned in the context of cellular extravasation, opening new avenues for further research.

We aimed to extend the application of our pipeline to a distinct dataset, conducting a comprehensive analysis of brain endothelial cells under both control and inflamed conditions. Our objective was to gain a profound understanding of the unique molecular interactions governing the extravasation process, particularly between an activated endothelium and activated T-cells.

Our differential expression analysis of the inflamed and control endothelial cells of the brain resulted in the upregulation of cell-adhesion markers like Icam-1, and Vcam-1 which are established in the literature. Ackr1 has previously been reported to increase in expression in an in-vitro model mouse model for neuroinflammation. The upregulation of this gene in the inflamed BBB validates the previous finding. Another interesting observation was the upregulation of Nectin2, a cell-adhesion molecule, under inflammation. The co-expression analysis yielded two gene co-expressing modules mapped to BBB and T-cells indicating that the gene expression profiles were more similar between the control and diseased condition. The cell-cell communication analyses yielded 9 novel interactions between the inflamed endothelium and activated T-cell.

We observed contradictions in gene expressions for similar conditions between the current analysis and the previous one, strengthening our conviction that a larger dataset is necessary to validate our findings. Nevertheless, the establishment of this pipeline has

empowered us to conduct various analyses, facilitating the validation of these results both in the experiments conducted by us and in future data availability. This work has fundamental impact on the design of drug delivery systems for identification of targeting moieties[144,145,146,147,148,149,150], the design of cell capture technologies[151,152,153,154,155,156], and in aiding fundamental mechanistic interpretation of biological experiments[157,158,159,160,161].

Future directions for this research involve experimental validation of our findings. To further strengthen our results, we can explore the translation of gene expression into protein levels by conducting proteomics studies. This approach will provide a deeper understanding of the molecular mechanisms underlying lymphocyte extravasation.

REFERENCES

[1] Kay L.Medina. Chapter-4 Overview of the immune system. Handbook of Clinical Neurology, Vol 133, 2016, pages 61-76

[2] Francisco A. Bonilla, Hans C. Oettgen, Adaptive immunity, Journal of Allergy and Clinical Immunology, Volume 125, Issue 2, Supplement 2, 2010, Pages S33-S40, ISSN 0091-6749, doi.org/10.1016/j.jaci.2009.09.017.

[3] Ashby, K.M., Hogquist, K.A. A guide to thymic selection of T cells. *Nat Rev Immunol* (2023). doi.org/10.1038/s41577-023-00911-8.

[4] Charles D.Surh & Jonathan Sprent (2008). Homeostasis of naïve and memory T cells. Immunity Review, Cell Press. DOI 10.1016/j.immuni.2008.11.002

[5] Zoltán Jakus, Jason P Gleghorn, David R Enis, Aslihan Sen, Stephanie Chia, Xi Liu, David R Rawnsley, Yiqing Yang, Paul R Hess, Zhiying Zou, Jisheng Yang, Susan H Guttentag, Celeste M Nelson, Mark L Kahn. Lymphatic function is required prenatally for lung inflation at birth. Journal of Experimental Medicine, 2014/5/5, Vol 211, Issue 5, Pages 815-826.

[6] Joshua T Morgan, Jasmine Shirazi, Erica M Comber, Christian Eschenburg, Jason P Gleghorn. Fabrication of centimeter-scale and geometrically arbitrary vascular networks using in vitro self-assembly. 2019/1/1. Biomaterials. Vol 189, Pages 37-47.

[7] Yuhao Zhang, Xiaoyu Dong, Jasmine Shirazi, Jason P Gleghorn, Krithika Lingappan. Pulmonary endothelial cells exhibit sexual dimorphism in their response to hyperoxia.2018/11/1. American Journal of Physiology-Heart and Circulatory Physiology. Vol 315, Issue 5, Pages - H1287-H1292.

[8] Yuhao Zhang, Cristian Coarfa, Xiaoyu Dong, Weiwu Jiang, Brielle Hayward-Piatkovskyi, Jason P Gleghorn, Krithika Lingappan. MicroRNA-30a as a candidate underlying sex-specific differences in neonatal hyperoxic lung injury: implications for BPD. 2019/1/1. American Journal of Physiology-Lung Cellular and Molecular Physiology. Vol 316, Issue 1, pages – L144-L156.

[9] Bujoreanu I, Gupta V. Anatomy, Lymph Nodes. [Updated 2023 Jul 25]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK557717/

[10] Gabriel D. Victora and Michel C. Nussenzweig. Germinal Centers. Annual Review Immunology, 2022. 40:413–42.

[11] Girard J.P. et. al, (2012). HEV, lymphatics and homeostatic immune cell trafficking in lymph nodes. Nature Reviews Immunology, 12, 762-773.

[12] Ann Ager & Michael J May. Understanding high endothelial venules: Lessons for cancer immunology. OncoImmunology, Vol 4, 2015, Issue 6. doi: https://doi.org/10.1080/2162402X.2015.1008791.

[13] von Andrian, U. H., & Mempel, T. R. (2003). Homing and cellular traffic in lymph nodes. Nature Reviews Immunology, 3(11), 867-878.

[14] Praveen Ballabh, Alex Braun & Maiken Nedergaard (2004). The blood–brain barrier: an overview: Structure, regulation, and clinical implications. Neurobiology of Disease. Vol 16, Issue 1, June 2004, Pages 1-13. doi.org/10.1016/j.nbd.2003.12.016.

[15] Luca Marchetti & Britta Englehardt (2020). Immune cell trafficking across the blood-brain barrier in the absence and presence of neuroinflammation. Vascular Biology, 2:1, H1-H18. doi.org/10.1016/j.nbd.2003.12.016

[16] Schinkel, A. H., & Jonker, J. W. (2003). Mammalian drug efflux transporters of the ATP binding cassette (ABC) family: an overview. Advanced Drug Delivery Reviews, 55(1), 3-29.

[17] Kivisäkk, P., Mahad, D. J., Callahan, M. K., Trebst, C., Tucky, B., Wei, T., ... & Lassmann, H. (2003). Expression of CCR7 in multiple sclerosis: implications for CNS immunity. Annals of Neurology, 55(5), 627-638.

[18] Mario Cabodi, Nak Won Choi, Jason P Gleghorn, Christopher SD Lee, Lawrence J Bonassar, Abraham D Stroock. A microfluidic biomaterial. Journal of the American Chemical Society, 2005/10/12. Vol 127, Issue 40, 13788-13789.

[19] Nak Won Choi, Mario Cabodi, Brittany Held, Jason P Gleghorn, Lawrence J Bonassar, Abraham D Stroock. Microfluidic scaffolds for tissue engineering. Nature Materials, 2007/11, Vol 6, Issue 11, 908-915.

[20] Vonetta L Edwards, Elias McComb, Jason P Gleghorn, Larry Forney, Patrik M Bavoil, Jacques Ravel. Three-dimensional models of the cervicovaginal epithelia to study host–microbiome interactions and sexually transmitted infections. Pathogens and disease, 2022, Vol 80, Issue 1, Pages – ftac026.

[21] Laurel E Schappell, Daniel J Minahan, Jason P Gleghorn. A microfluidic system to measure neonatal lung compliance over late stage development as a functional measure

of lung tissue mechanics. Journal of Biomechanical Engineering. 2020/10/1, Vol 142, Issue 10, Pages – 100803.

22 Brea Chernokal, Cailin R Gonyea, Jason P Gleghorn. Lung Development in a Dish: Models to Interrogate the Cellular Niche and the Role of Mechanical Forces in Development. Engineering Translational Models of Lung Homeostasis and Disease. 2023/5/18. Pg 29-48.

23Takeshita Y, Ransohoff RM. Inflammatory cell trafficking across the blood-brain barrier: chemokine regulation and in vitro models. Immunol Rev. 2012 Jul;248(1):228-39. doi: 10.1111/j.1600-065X.2012.01127.x. PMID: 22725965.

24 J.Greenwood, S.J.Heasman, J.I.Alvarez, A.Prat, R.Lyck & B.Englehardt (2011). Review: Leucocyte–endothelial cell crosstalk at the blood–brain barrier: A prerequisite for successful immune cell entry to the brain. Neuropathology and Applied Neurobiology, 37, 24-39. doi.org/10.1111/j.1365-2990.2010.01140.x.

25 Veerman, K.M., Lafouresse, F., Girard, JP. (2014). Tumor High Endothelial Venules and Lymphocyte Trafficking. In: Feige, JJ., Pagès, G., Soncin, F. (eds) Molecular Mechanisms of Angiogenesis. Springer, Paris. https://doi.org/10.1007/978-2-8178-0466-8_16

26 Columba-Cabezas S, Serafini B, Ambrosini E, Aloisi F. Lymphoid chemokines CCL19 and CCL21 are expressed in the central nervous system during experimental autoimmune encephalomyelitis: implications for the maintenance of chronic neuroinflammation. Brain Pathol 2003; 13: 38–51

27 Michael Abadier, Neda Haghayegh Jahromi, Ludmila Cardoso Alves, Rémy Boscacci, Dietmar Vestweber, Scott Barnum, Urban Deutsch, Britta Engelhardt, Ruth Lyck; Cell surface levels of endothelial ICAM-1 influence the transcellular or paracellular T-cell diapedesis across the blood–brain barrier. European Journal of Immunology, 26th Dec 2014. doi: https://doi.org/10.1002/eji.201445125

28 Goswami D, Vestweber D. How leukocytes trigger opening and sealing of gaps in the endothelial barrier. F1000Res. 2016 Sep 14;5:F1000 Faculty Rev-2321. doi: 10.12688/f1000research.9185.1. PMID: 27703663; PMCID: PMC5031128.

29 Eun Jeong Park, António Peixoto, Yoichi Imai, Ahmad Goodarzi, Guiying Cheng, Christopher V. Carman, Ulrich H. von Andrian, Motomu Shimaoka; Distinct roles for LFA-1 affinity regulation during T-cell adhesion, diapedesis, and interstitial migration in lymph nodes. Blood 2010; 115 (8): 1572–1581. doi: https://doi.org/10.1182/blood-2009-08-237917

30 Yan SLS, Hwang IY, Kamenyeva O, Kehrl JH. In Vivo F-Actin Filament Organization during Lymphocyte Transendothelial and Interstitial Migration Revealed by Intravital Microscopy. iScience. 2019 Jun 28;16:283-297. doi:

10.1016/j.isci.2019.05.040. Epub 2019 May 30. PMID: 31203185; PMCID: PMC6581778.

[31] Carman CV, Sage PT, Sciuto TE, et al. Transcellular diapedesis is initiated by invasive podosomes. Immunity. 2007;26(6):784-797.

[32] Marie-Dominique Filippi, Neutrophil transendothelial migration: updates and new perspectives. Blood (2019) 133 (20): 2149–2158. 2019 May 16. doi.org/10.1182/blood-2018-12-844605.

[33] Marmon S, Hinchey J, Oh P, Cammer M, de Almeida CJ, Gunther L, Raine CS, Lisanti MP. Caveolin-1 expression determines the route of neutrophil extravasation through skin microvasculature. Am J Pathol. 2009 Feb;174(2):684-92. doi: 10.2353/ajpath.2009.080091. PMID: 19164603; PMCID: PMC2630575.

[34] Filippi MD. Mechanism of Diapedesis: Importance of the Transcellular Route. Adv Immunol. 2016;129:25-53. doi: 10.1016/bs.ai.2015.09.001. Epub 2015 Oct 14. PMID: 26791857; PMCID: PMC4889131.

[35]Wimmer I, Tietz S, Nishihara H, Deutsch U, Sallusto F, Gosselet F, Lyck R, Muller WA, Lassmann H, Engelhardt B. PECAM-1 Stabilizes Blood-Brain Barrier Integrity and Favors Paracellular T-Cell Diapedesis Across the Blood-Brain Barrier During Neuroinflammation. Front Immunol. 2019 Apr 5;10:711. doi: 10.3389/fimmu.2019.00711. PMID: 31024547; PMCID: PMC6460670.

[36] Filippi MD. Leukocyte transcellular diapedesis: Rap1b is in control. Tissue Barriers. 2015 May 21;3(3):e1052185. doi: 10.1080/21688370.2015.1052185. PMID: 26451346; PMCID: PMC4574898.

[37] Eric P. Schmidt, Warren L. Lee, Rachel L. Zemans, Cory Yamashita, and Gregory P. Downey. On, Around, and Through: Neutrophil-Endothelial Interactions in Innate Immunity. Physiology, 01 Oct, 2011. doi: https://doi.org/10.1152/physiol.00011.2011

[38] Schaefer A, van Duijn TJ, Majolee J, Burridge K, Hordijk PL. Endothelial CD2AP binds the receptor ICAM-1 to control mechanosignaling, leukocyte adhesion, and the route of leukocyte diapedesis in vitro. J Immunol. 2017;198(12):4823-4836.

[39] Vestweber, D. How leukocytes cross the vascular endothelium. Nat Rev Immunol 15, 692–704 (2015). https://doi.org/10.1038/nri3908

[40] Haque, A., Engel, J., Teichmann, S.A. et al. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. Genome Med 9, 75 (2017). https://doi.org/10.1186/s13073-017-0467-4

[41] Krithika Lingappan, Oluyinka O Olutoye 2nd, Abiud Cantu, Manuel Eliezer Cantu Gutierrez, Nahir Cortes-Santiago, JD Hammond, Jamie Gilley, Joselyn Rojas Quintero, Hui Li, Francesca Polverino, Jason P Gleghorn, Sundeep G Keswani. Molecular insights

using spatial transcriptomics of the distal lung in congenital diaphragmatic hernia. American Journal of Physiology-Lung Cellular and Molecular Physiology. 2023/10/1. Vol 325, Issue 4, L477-L486.

[42] Oluyinka O Olutoye II, Walker D Short, Jamie Gilley, JD Hammond II, Michael A Belfort, Timothy C Lee, Alice King, Jimmy Espinoza, Luc Joyeux, Krithika Lingappan, Jason P Gleghorn, Sundeep G Keswani. The Cellular and Molecular Effects of Fetoscopic Endoluminal Tracheal Occlusion in Congenital Diaphragmatic Hernia. Frontiers in Pediatrics. 2022, pages- 1046.

[43] Plasschaert, L.W., Žilionis, R., Choo-Wing, R. et al. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. Nature 560, 377–381 (2018). https://doi.org/10.1038/s41586-018-0394-6

[44] Suo S, Zhu Q, Saadatpour A et al (2018) Revealing the critical regulators of cell identity in the mouse cell atlas. Cell Rep 25:1436–1445.e3

[45] Nafiseh Erfanian et. al, Immunotherapy of cancer in single-cell RNA sequencing era: A precision medicine perspective. Biomedicine and pharmacotherapy, Vol 146, February 2022, 112558. https://doi.org/10.1016/j.biopha.2021.112558

[46] Liu S, Trapnell C. Single-cell transcriptome sequencing: recent advances and remaining challenges. F1000Res. 2016 Feb 17;5:F1000 Faculty Rev-182. doi: 10.12688/f1000research.7223.1. PMID: 26949524; PMCID: PMC4758375.

[47] Aleksandra A. Kolodziejczyk, Jong Kyoung Kim, Valentine Svensson, John C. Marioni, and Sarah A. Teichmann. The Technology and Biology of Single-Cell RNA Sequencing; Molecular Cell, 2015 May 21; Vol 58, Issue 4, 610-620. https://doi.org/10.1016/j.molcel.2015.04.005

[48] Slovin, S. et al. (2021). Single-Cell RNA Sequencing Analysis: A Step-by-Step Overview. In: Picardi, E. (eds) RNA Bioinformatics. Methods in Molecular Biology, vol 2284. Humana, New York, NY. https://doi.org/10.1007/978-1-0716-1307-8_19

[49] Jeanette Baran-Gale, Tamir Chandra, Kristina Kirschner, Experimental design for single-cell RNA sequencing, Briefings in Functional Genomics, Volume 17, Issue 4, July 2018, Pages 233–239, https://doi.org/10.1093/bfgp/elx035

[50] Brennecke P, Anders S, Kim JK, Kołodziejczyk AA, Zhang X, Proserpio V, et al. Accounting for technical noise in single-cell RNA-seq experiments. Nat Methods. 2013;10:1093–5.

[51] Kim JK, Kolodziejczyk AA, Ilicic T, Illicic T, Teichmann SA, Marioni JC. Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. Nat Commun. 2015;6:8687.

[52] Zappia, L., Phipson, B. & Oshlack, A. Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database. PLoS Comput. Biol. 14, e1006245 (2018). scRNA-tools. Accessed 30 September 2023

[53] https://www.ddbj.nig.ac.jp/index-e.html

[54] https://cellxgene.cziscience.com/collections

[55] Yang, A.C., Stevens, M.Y., Chen, M.B. et al. Physiological blood–brain transport is impaired with age by a shift in transcytosis. Nature 583, 425–430 (2020). https://doi.org/10.1038/s41586-020-2453-z

[56] Kim J, Nguyen T, Cifello J, Ahmad R, Zhang Y, Yang Q, Lee JE, Li X, Kai Y, De S, Peng W, Ge K, Weng NP. Lysine methyltransferase Kmt2d regulates naive CD8+ T cell activation-induced survival. Front Immunol. 2023 Jan 19;13:1095140. doi: 10.3389/fimmu.2022.1095140. PMID: 36741385; PMCID: PMC9892454.

[57] Fernández-García J, Franco F, Parik S, Altea-Manzano P et al. CD8+T cell metabolic rewiring defined by scRNA-seq identifies a critical role of ASNS expression dynamics in T cell differentiation. Cell Rep 2022 Nov 15;41(7):111639. PMID: 36384124

[58] Brulois K, Rajaraman A, Szade A, Nordling S, Bogoslowski A, Dermadi D, Rahman M, Kiefel H, O'Hara E, Koning JJ, Kawashima H, Zhou B, Vestweber D, Red-Horse K, Mebius RE, Adams RH, Kubes P, Pan J, Butcher EC. A molecular map of murine lymph node blood vascular endothelium at single cell resolution. Nat Commun. 2020 Jul 30;11(1):3798. doi: 10.1038/s41467-020-17291-5. PMID: 32732867; PMCID: PMC7393069.

[59] Hua Y, Vella G, Rambow F, Allen E et al. Cancer immunotherapies transition endothelial cells into HEVs that generate TCF1⁺ T lymphocyte niches through a feed-forward loop. Cancer Cell 2022 Dec 12;40(12):1600-1618.e10. PMID: 36423635

[60] Luecken MD, Theis FJ (2019) Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol 15

[61] Susanne C van den Brink, Fanny Sage, Ábel Vértesy, Bastiaan Spanjaard, Josi Peterson-Maduro, Chloé S Baron, Catherine Robin, and Alexander van Oudenaarden. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. Nature methods, 14(10):935–936, September 2017.

[62] Heumos, L., Schaar, A.C., Lance, C. et al. Best practices for single-cell analysis across modalities. Nat Rev Genet (2023). https://doi.org/10.1038/s41576-023-00586-w

[63] W Evan Johnson, Cheng Li, and Ariel Rabinovic. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics, 8(1):118–127, January 2007. doi:10.1093/biostatistics/kxj037.

[64] Laleh Haghverdi, Aaron T L Lun, Michael D Morgan, and John C Marioni. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nature biotechnology, April 2018. doi:10.1038/nbt.4091.

[65] Krzysztof Polański, Jong-Eun Park, Matthew D Young, Zhichao Miao, Kerstin B Meyer, and Sarah A Teichmann. BBKNN: Fast Batch Alignment of Single Cell Transcriptomes. Bioinformatics, August 2019. doi:10.1093/bioinformatics/btz625.

[66] Chenling Xu, Romain Lopez, Edouard Mehlman, Jeffrey Regier, Michael I Jordan, and Nir Yosef. Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. Molecular systems biology, 17(1):e9620, January 2021. doi:10.15252/msb.20209620.

[67] Mohammad Lotfollahi, F Alexander Wolf, and Fabian J Theis. scGen predicts single-cell perturbation responses. Nature methods, 16(8):715–721, August 2019. doi:10.1038/s41592-019-0494-8.

[68] Chen, W. et al. A multicenter study benchmarking single-cell RNA sequencing technologies using reference samples. Nat. Biotechnol. https://doi.org/10.1038/s41587-020-00748-9 (2020).

[69] Daniel Osorio, James J Cai, Systematic determination of the mitochondrial proportion in human and mice tissues for single-cell RNA-sequencing data quality control, Bioinformatics, Volume 37, Issue 7, March 2021, Pages 963–967, https://doi.org/10.1093/bioinformatics/btaa751

[70] 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). https://doi.org/10.1186/s13059-019-1850-9

[71] Luecken, M.D., Büttner, M., Chaichoompu, K. *et al.* Benchmarking atlas-level data integration in single-cell genomics. *Nat Methods* **19**, 41–50 (2022). https://doi.org/10.1038/s41592-021-01336-8

[72] Butler, A., Hoffman, P., Smibert, P. et al. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 36, 411–420 (2018). https://doi.org/10.1038/nbt.4096

[73] Korsunsky, I., Millard, N., Fan, J. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods 16, 1289–1296 (2019). https://doi.org/10.1038/s41592-019-0619-0

[74] Welch JD, Kozareva V, Ferreira A, Vanderburg C, Martin C, Macosko EZ. Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity. Cell. 2019 Jun 13;177(7):1873-1887.e17. doi: 10.1016/j.cell.2019.05.006. Epub 2019 Jun 6. PMID: 31178122; PMCID: PMC6716797.

[75] RPCA integration page. https://satijalab.org/seurat/articles/integration_rpca

[76] Lopez, R., Regier, J., Cole, M.B. et al. Deep generative modeling for single-cell transcriptomics. Nat Methods 15, 1053–1058 (2018). https://doi.org/10.1038/s41592-018-0229-2

[77] Hie, B., Bryson, B. & Berger, B. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. Nat. Biotechnol. 37, 685–691 (2019).

[78] Cao K, Gong Q, Hong Y, Wan L. A unified computational framework for single-cell data integration with optimal transport. Nat Commun. 2022 Dec 1;13(1):7419. doi: 10.1038/s41467-022-35094-8. PMID: 36456571; PMCID: PMC9715710.

[79] Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, Hao Y, Stoeckius M, Smibert P, Satija R. Comprehensive Integration of Single-Cell Data. Cell. 2019 Jun 13;177(7):1888-1902.e21. doi: 10.1016/j.cell.2019.05.031. Epub 2019 Jun 6. PMID: 31178118; PMCID: PMC6687398.

[80] Fang, Z., Martin, J. & Wang, Z. Statistical methods for identifying differentially expressed genes in RNA-Seq experiments. Cell Biosci 2, 26 (2012). https://doi.org/10.1186/2045-3701-2-26

[81] Das S, Rai A, Merchant ML, Cave MC, Rai SN. A Comprehensive Survey of Statistical Approaches for Differential Expression Analysis in Single-Cell RNA Sequencing Studies. Genes. 2021; 12(12):1947. https://doi.org/10.3390/genes12121947

[82] Li D, Zand M, Dye T, Goniewicz M, Rahman I, Xie Z. An evaluation of statistical differential analysis methods in single-cell RNA-seq data. Res Sq [Preprint]. 2023 Mar 23:rs.3.rs-2670717. doi: 10.21203/rs.3.rs-2670717/v1. PMID: 36993457; PMCID: PMC10055642.

[83] Wang, T., Li, B., Nelson, C.E. et al. Comparative analysis of differential gene expression analysis tools for single-cell RNA sequencing data. BMC Bioinformatics 20, 40 (2019). https://doi.org/10.1186/s12859-019-2599-6

[84] Gene ontology resource. Gene Ontology overview

[85] Kenneth King Yin Liu & Katerina Dorovini-Zis. Regulation of CXCL12 and CXCR4 expression by human brain endothelial cells and their role in CD4+ and CD8+ T cell adhesion and transendothelial migration; Journal of NeuroImmunology, Vol215, Issue 1-2, p49-64. 2009 Sep 22. DOI:https://doi.org/10.1016/j.jneuroim.2009.08.003

[86] Wang J, Yin L, Chen Z. Neuroprotective role of fibronectin in neuron-glial extrasynaptic transmission. Neural Regen Res. 2013 Feb 5;8(4):376-82. doi: 10.3969/j.issn.1673-5374.2013.04.010. PMID: 25206678; PMCID: PMC4107531.

[87] Nottebaum AF, Cagna G, Winderlich M, Gamp AC, Linnepe R, Polaschegg C, Filippova K, Lyck R, Engelhardt B, Kamenyeva O, Bixel MG, Butz S, Vestweber D. VE-PTP maintains the endothelial barrier via plakoglobin and becomes dissociated from VE-cadherin by leukocytes and by VEGF. J Exp Med. 2008 Nov 24;205(12):2929-45. doi: 10.1084/jem.20080406. Epub 2008 Nov 17. PMID: 19015309; PMCID: PMC2585844.

[88] Rantakari, P., Auvinen, K., Jäppinen, N. et al. The endothelial protein PLVAP in lymphatics controls the entry of lymphocytes and antigens into lymph nodes. Nat Immunol 16, 386–396 (2015). https://doi.org/10.1038/ni.3101

[89] Tillu, V.A., Rae, J., Gao, Y. et al. Cavin1 intrinsically disordered domains are essential for fuzzy electrostatic interactions and caveola formation. Nat Commun 12, 931 (2021). https://doi.org/10.1038/s41467-021-21035-4

[90] Gorina R, Lyck R, Vestweber D, Engelhardt B. β2 integrin-mediated crawling on endothelial ICAM-1 and ICAM-2 is a prerequisite for transcellular neutrophil diapedesis across the inflamed blood-brain barrier. J Immunol. 2014 Jan 1;192(1):324-37. doi: 10.4049/jimmunol.1300858. Epub 2013 Nov 20. PMID: 24259506.

[91] Dzwonek J, Wilczynski GM. CD44: molecular interactions, signaling and functions in the nervous system. Front Cell Neurosci. 2015 May 7;9:175. doi: 10.3389/fncel.2015.00175. PMID: 25999819; PMCID: PMC4423434.

[92] Sackstein, R., Schatton, T. & Barthel, S. T-lymphocyte homing: an underappreciated yet critical hurdle for successful cancer immunotherapy. Lab Invest 97, 669–697 (2017). https://doi.org/10.1038/labinvest.2017.25

[93] Binder C, Cvetkovski F, Sellberg F, Berg S, Paternina Visbal H, Sachs DH, Berglund E, Berglund D. CD2 Immunobiology. Front Immunol. 2020 Jun 9;11:1090. doi: 10.3389/fimmu.2020.01090. PMID: 32582179; PMCID: PMC7295915.

[94] Ben-Zvi A, Lacoste B, Kur E, Andreone BJ, Mayshar Y, Yan H, Gu C. Mfsd2a is critical for the formation and function of the blood–brain barrier. Nature. 2014;509(7501):507–511. doi: 10.1038/nature13324.

[95] Wang Z, Zheng Y, Wang F, Zhong J, Zhao T, Xie Q, Zhu T, Ma F, Tang Q, Zhou B, et al. Mfsd2a and Spns2 are essential for sphingosine-1-phosphate transport in the formation and maintenance of the blood-brain barrier.

[96] Kaya M, Ahishali B. Basic physiology of the blood-brain barrier in health and disease: a brief overview. Tissue Barriers. 2021 Jan 2;9(1):1840913. doi:

10.1080/21688370.2020.1840913. Epub 2020 Nov 15. PMID: 33190576; PMCID: PMC7849738.

[97] Veerman K, Tardiveau C, Martins F, Coudert J, Girard JP. Single-Cell Analysis Reveals Heterogeneity of High Endothelial Venules and Different Regulation of Genes Controlling Lymphocyte Entry to Lymph Nodes. Cell Rep. 2019 Mar 12;26(11):3116-3131.e5. doi: 10.1016/j.celrep.2019.02.042. PMID: 30865898.

[98] Bhar A, Haubrock M, Mukhopadhyay A, Maulik U, Bandyopadhyay S, Wingender E. Coexpression and coregulation analysis of time-series gene expression data in estrogen-induced breast cancer cell. Algorithms Mol Biol. 2013 Mar 23;8(1):9. doi: 10.1186/1748-7188-8-9. PMID: 23521829; PMCID: PMC3827943.

[99] Langfelder, P., Horvath, S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559 (2008). https://doi.org/10.1186/1471-2105-9-559

[100] Alberto de la Fuente. From ‘differential expression’ to ‘differential networking’ – identification of dysfunctional regulatory networks in diseases; Trends in Genetics, Vol 26, Issue 7, P326-333, July 2010. DOI:https://doi.org/10.1016/j.tig.2010.05.001

[101] Hudson N.J. et al. A differential wiring analysis of expression data correctly identifies the gene containing the causal mutation. PLoS Comput. Biol. 2009; 5: e1000382

[102] Carter SL, Brechbühler CM, Griffin M, Bond AT. Gene co-expression network topology provides a framework for molecular characterization of cellular state. Bioinformatics. 2004 Sep 22;20(14):2242-50. doi: 10.1093/bioinformatics/bth234. Epub 2004 May 6. PMID: 15130938.

[103] Elo LL, Järvenpää H, Oresic M, Lahesmaa R, Aittokallio T. Systematic construction of gene coexpression networks with applications to human T helper cell differentiation process. Bioinformatics. 2007 Aug 15;23(16):2096-103. doi: 10.1093/bioinformatics/btm309. Epub 2007 Jun 6. PMID: 17553854.

[104] Farhadian, M., Rafat, S.A., Panahi, B. et al. Weighted gene co-expression network analysis identifies modules and functionally enriched pathways in the lactation process. Sci Rep 11, 2367 (2021). https://doi.org/10.1038/s41598-021-81888-z

[105] Johnson ECB et.al. Large-scale proteomic analysis of Alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. Nat Med. 2020 May;26(5):769-780. doi: 10.1038/s41591-020-0815-6. Epub 2020 Apr 13. PMID: 32284590; PMCID: PMC7405761.

[106] Morabito S, Reese F, Rahimzadeh N, Miyoshi E, Swarup V. hdWGCNA identifies co-expression networks in high-dimensional transcriptomics data. Cell Rep Methods.

2023 Jun 12;3(6):100498. doi: 10.1016/j.crmeth.2023.100498. PMID: 37426759; PMCID: PMC10326379.

[107] hdWGNA tutorial. hdWGCNA in single-cell data • hdWGCNA (smorabit.github.io)

[108] Mason MJ, Fan G, Plath K, Zhou Q, Horvath S. Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. BMC Genomics. 2009 Jul 20;10:327. doi: 10.1186/1471-2164-10-327. PMID: 19619308; PMCID: PMC2727539.

[109] Luca Marchetti et.al., ACKR1 favors transcellular over paracellular T-cell diapedesis across the blood-brain barrier in neuroinflammation in vitro. European Journal of Immunology; 2021 September 15, doi: https://doi.org/10.1002/eji.202149238

[110] Bastiani M, Liu L, Hill MM, Jedrychowski MP, Nixon SJ, Lo HP, Abankwa D, Luetterforst R, Fernandez-Rojo M, Breen MR, Gygi SP, Vinten J, Walser PJ, North KN, Hancock JF, Pilch PF, Parton RG. MURC/Cavin-4 and cavin family members form tissue-specific caveolar complexes. J Cell Biol. 2009 Jun 29;185(7):1259-73. doi: 10.1083/jcb.200903053. Epub 2009 Jun 22. PMID: 19546242; PMCID: PMC2712963.

[111] Mabruka Alfaidi et.al, Sinner or Saint?: Nck Adaptor Proteins in Vascular Biology. Front. Cell Dev. Biol., 26 May 2021 Sec. Signaling. Volume 9 - 2021 | https://doi.org/10.3389/fcell.2021.688388

[112] Wittchen ES. Endothelial signaling in paracellular and transcellular leukocyte transmigration. Front Biosci (Landmark Ed). 2009 Jan 1;14(7):2522-45. doi: 10.2741/3395. PMID: 19273217; PMCID: PMC2654604.

[113] Erick Armingol, Adam Officer, Olivier Harismendy, and Nathan E Lewis. Deciphering cell-cell interactions and communication from gene expression. Nature Reviews. Genetics, 22(2):71–88, 2021.

[114] Xin Shao, Jie Liao, Chengyu Li, Xiaoyan Lu, Junyun Cheng, Xiaohui Fan, CellTalkDB: a manually curated database of ligand–receptor interactions in humans and mice, Briefings in Bioinformatics, Volume 22, Issue 4, July 2021, bbaa269, https://doi.org/10.1093/bib/bbaa269

[115] Axel A Almet, Zixuan Cang, Suoqin Jin, and Qing Nie. The landscape of cell-cell communication through single-cell transcriptomics. Current Opinion in Systems Biology, 26:12–23, jun 2021.

[116] Saidi Wang, Hansi Zheng, James S Choi, Jae K Lee, Xiaoman Li, and Haiyan Hu. A systematic evaluation of the computational tools for ligand-receptor-based cell-cell interaction inference. Briefings in functional genomics, 21(5):339–356, sep 2022.

[117] Daniel Dimitrov, Dénes Türei, Martin Garrido-Rodriguez, Paul L Burmedi, James S Nagai, Charlotte Boys, Ricardo O Ramirez Flores, Hyojin Kim, Bence Szalai, Ivan G Costa, Alberto Valdeolivas, Aurélien Dugourd, and Julio Saez-Rodriguez. Comparison of methods and resources for cell-cell communication inference from single-cell RNA-seq data. Nature Communications, 13(1):3224, jun 2022.

[118] Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, Myung P, Plikus MV, Nie Q. Inference and analysis of cell-cell communication using CellChat. Nat Commun. 2021 Feb 17;12(1):1088. doi: 10.1038/s41467-021-21246-9. PMID: 33597522; PMCID: PMC7889871.

[119] Zhang Y, Liu T, Wang J, Zou B, Li L, Yao L, Chen K, Ning L, Wu B, Zhao X, Wang D. Cellinker: a platform of ligand-receptor interactions for intercellular communication analysis. Bioinformatics. 2021 Jan 20:btab036. doi: 10.1093/bioinformatics/btab036. Epub ahead of print. PMID: 33471060; PMCID: PMC7929259.

[120] Xin Shao, Jie Liao, Chengyu Li, Xiaoyan Lu, Junyun Cheng, Xiaohui Fan, CellTalkDB: a manually curated database of ligand–receptor interactions in humans and mice, Briefings in Bioinformatics, Volume 22, Issue 4, July 2021, bbaa269, https://doi.org/10.1093/bib/bbaa269

[121] Zhang R, Zhang S, Xing R, Zhang Q. High expression of EZR (ezrin) gene is correlated with the poor overall survival of breast cancer patients. Thorac Cancer. 2019 Oct;10(10):1953-1961. doi: 10.1111/1759-7714.13174. Epub 2019 Aug 26. PMID: 31452341; PMCID: PMC6775014.

[122] Yonemura S, Hirao M, Doi Y, Takahashi N, Kondo T, Tsukita S, Tsukita S. Ezrin/radixin/moesin (ERM) proteins bind to a positively charged amino acid cluster in the juxta-membrane cytoplasmic domain of CD44, CD43, and ICAM-2. J Cell Biol. 1998 Feb 23;140(4):885-95. doi: 10.1083/jcb.140.4.885. PMID: 9472040; PMCID: PMC2141743.

[123] Nibbs, R., Graham, G. Immune regulation by atypical chemokine receptors. Nat Rev Immunol 13, 815–829 (2013). https://doi.org/10.1038/nri3544

[124] Girbl T, Lenn T, Perez L, Rolas L, Barkaway A, Thiriot A, Del Fresno C, Lynam E, Hub E, Thelen M, Graham G, Alon R, Sancho D, von Andrian UH, Voisin MB, Rot A, Nourshargh S. Distinct Compartmentalization of the Chemokines CXCL1 and CXCL2 and the Atypical Receptor ACKR1 Determine Discrete Stages of Neutrophil Diapedesis. Immunity. 2018 Dec 18;49(6):1062-1076.e6. doi: 10.1016/j.immuni.2018.09.018. Epub 2018 Nov 13. PMID: 30446388; PMCID: PMC6303217.

[125] Simon T, Bromberg JS. Regulation of the Immune System by Laminins. Trends Immunol. 2017 Nov;38(11):858-871. doi: 10.1016/j.it.2017.06.002. Epub 2017 Jul 3. PMID: 28684207; PMCID: PMC5669817.

[126] Zapata-Acevedo JF, García-Pérez V, Cabezas-Pérez R, Losada-Barragán M, Vargas-Sánchez K, González-Reyes RE. Laminin as a Biomarker of Blood-Brain Barrier Disruption under Neuroinflammation: A Systematic Review. Int J Mol Sci. 2022 Jun 17;23(12):6788. doi: 10.3390/ijms23126788. PMID: 35743229; PMCID: PMC9224176.

[127] Fournier AP, Tastet O, Charabati M, Hoornaert C et al. Single-Cell Transcriptomics Identifies Brain Endothelium Inflammatory Networks in Experimental Autoimmune Encephalomyelitis. Neurol Neuroimmunol Neuroinflamm 2023 Jan;10(1)

[128] Ayme G, Adam F, Legendre P, et al. A novel single-domain antibody against von Willebrand factor A1 domain resolves leukocyte recruitment and vascular leakage during inflammation-brief report. Arterioscler Thromb Vasc Biol 2017;37:1736–40

[129] Suidan GL, Brill A, De Meyer SF, et al. Endothelial von Willebrand factor promotes blood-brain barrier flexibility and provides protection from hypoxia and seizures in mice. *Arterioscler Thromb Vasc Biol* 2013;33:2112–20.

[130] Su H, Na N, Zhang X, Zhao Y. The biological function and significance of CD74 in immune diseases. Inflamm Res. 2017 Mar;66(3):209-216. doi: 10.1007/s00011-016-0995-1. Epub 2016 Oct 17. PMID: 27752708.

[131] Erturk K, Karaman S, Dagoglu N, Serilmez M, Duranyildiz D, Tas F. Serum nectin-2 and nectin-4 are diagnostic in lung cancer: which is superior? Wien Klin Wochenschr. 2019 Sep;131(17-18):419-426. doi: 10.1007/s00508-019-01537-4. Epub 2019 Aug 22. PMID: 31440821.

[132] Ho, D.WH., Tsui, YM., Chan, LK. et al. Single-cell RNA sequencing shows the immunosuppressive landscape and tumor heterogeneity of HBV-associated hepatocellular carcinoma. Nat Commun 12, 3684 (2021). https://doi.org/10.1038/s41467-021-24010-1

[133] Mizutani, K., Miyata, M., Shiotani, H. et al. Nectin-2 in general and in the brain. Mol Cell Biochem 477, 167–180 (2022). https://doi.org/10.1007/s11010-021-04241-y

[134] Shreya M. Kanth, Salina Garihe & Parizad Torabi-Parizi. The Role of Semaphorins and Their Receptors in Innate Immune Responses and Clinical Diseases of Acute Inflammation. Front. Immunol; 03 May 2021, Sec. Inflammation, Vol 12- 2021 | https://doi.org/10.3389/fimmu.2021.672441.

[135] Anderson GS, Ballester-Beltran J, Giotopoulos G, Guerrero JA, Surget S, Williamson JC, et al. (April 2022). "Unbiased cell surface proteomics identifies

SEMA4A as an effective immunotherapy target for myeloma". Blood. 139 (16): 2471–2482. doi:10.1182/blood.2021015161

[136] Hao-Long Zhou et al, Association between SEMA3A signaling pathway genes and BMD/OP risk: An epidemiological and experimental study. Front. Endocrinol., 08 November 2022, Sec. Bone Research. Volume 13 - 2022 | https://doi.org/10.3389/fendo.2022.1014431

[137] Saurabh Kartik Modi, Ryan Zurakowski, Jason P Gleghorn. Methodology for inference of intercellular gene interactions. bioRxiv, 2023. 2023.02. 26.530111.

[138] Logan Hallee, Jason P Gleghorn. Protein-Protein Interaction Prediction is Achievable with Large Language Models. bioRxiv. 2023, 2023.06. 07.544109.

[139] Catherine S Millar-Haskell, Allyson M Dang, Jason P Gleghorn. Coupling synthetic biology and programmable materials to construct complex tissue ecosystems. MRS Communications. 2019, pages: 1-12.

[140] Logan Hallee, Nikolaos Rafailidis, Jason P Gleghorn. cdsBERT-Extending Protein Language Models with Codon Awareness. bioRxiv. 2023. 2023.09. 15.558027.

[141] Rachel M Gilbert, Jason P Gleghorn. Connecting clinical, environmental, and genetic factors point to an essential role for vitamin A signaling in the pathogenesis of congenital diaphragmatic hernia. American Journal of Physiology-Lung Cellular and Molecular Physiology. 2023/4/1. Vol 324, Issue 4, Pages L456-L467.

[142] Brielle Hayward-Piatkovskyi, Cailin R Gonyea, Sienna C Pyle, Krithika Lingappan, Jason P Gleghorn. Sex-related external factors influence pulmonary vascular angiogenesis in a sex-dependent manner. American Journal of Physiology-Heart and Circulatory Physiology. 2023/1/1, Vol 324, Issue 1, H26-H32.

[143] Joshua T Morgan, Wade G Stewart, Robert A McKee, Jason P Gleghorn. The mechanosensitive ion channel TRPV4 is a regulator of lung development and pulmonary vasculature stabilization. Cellular and molecular bioengineering. 2018/10, Vol 11, 309-320.

[144] Rachel E Young, Katherine M Nelson, Samuel I Hofbauer, Tara Vijayakumar, Mohamad-Gabriel Alameh, Drew Weissman, Charalampos Papachristou, Jason P Gleghorn, Rachel S Riley. Lipid Nanoparticle Composition Drives Delivery of mRNA to the Placenta. bioRxiv. 2022.

[145] Catherine S Millar-Haskell, John L Sperduto, John H Slater, Colin Thorpe, Jason P Gleghorn. Secretion of the disulphide bond generating catalyst QSOX1 from pancreatic tumour cells into the extracellular matrix: Association with extracellular vesicles and matrix proteins. Journal of Extracellular Biology. 2022/7. Vol1, Issue 7, e48.

[146] Katherine M Nelson, Matthew K Hoffman, Jason P Gleghorn, Emily S Day. Diseases and conditions that impact maternal and fetal health and the potential for nanomedicine therapies. Advanced drug delivery reviews. 2021/3/1. Vol 170, 425-436.

[147] Suliman Khan, Majid Sharifi, Jason P Gleghorn, Mohammad Mahdi Nejadi Babadaei, Samir Haj Bloukh, Zehra Edis, Mohammadreza Amin, Qian Bai, Timo LM Ten Hagen, Mojtaba Falahati, William C Cho. Artificial engineering of the protein corona at bio-nano interfaces for improved cancer-targeted nanotherapy. Journal of Controlled Release. 2022/8/1. Vol 348, 127-147.

[148] S N'Dea, Katherine M Nelson, Megan N Dang, Jason P Gleghorn, Emily S Day. Gold nanoparticle biodistribution in pregnant mice following intravenous administration varies with gestational age. Nanomedicine: Nanotechnology, Biology and Medicine. 2021/8/1, Vol 36, 102412.

[149] S N'Dea, Katherine M Nelson, Jason P Gleghorn, Emily S Day. Design of nanomaterials for applications in maternal/fetal medicine. Journal of Materials Chemistry B. 2020. Vol 8, Issue 31, 6548-6561.

[150] Majid Sharifi, William C Cho, Asal Ansariesfahani, Rahil Tarharoudi, Hedyeh Malekisarvar, Soyar Sari, Samir Haj Bloukh, Zehra Edis, Mohamadreza Amin, Jason P Gleghorn, Timo LM ten Hagen, Mojtaba Falahati. An updated review on EPR-based solid tumor targeting nanocarriers for cancer treatment. Cancers. 2022/6/10. Vol 14, Issue 12. Pages 2868.

[151] Jason P Gleghorn, Erica D Pratt, Denise Denning, He Liu, Neil H Bander, Scott T Tagawa, David M Nanus, Paraskevi A Giannakakou, Brian J Kirby. Capture of circulating tumor cells from whole blood of prostate cancer patients using geometrically enhanced differential immunocapture (GEDI) and a prostate-specific antibody. Lab on a Chip. 2010, Vol 10, Issue 1, 27-29.

[152] Brian J Kirby, Mona Jodari, Matthew S Loftus, Gunjan Gakhar, Erica D Pratt, Chantal Chanel-Vos, Jason P Gleghorn, Steven M Santana, He Liu, James P Smith, Vicente N Navarro, Scott T Tagawa, Neil H Bander, David M Nanus, Paraskevi Giannakakou. Functional characterization of circulating tumor cells with a prostate-cancer-specific microfluidic device. PloS one. 2012/4/27.

[153] Erica D Pratt, Chao Huang, Benjamin G Hawkins, Jason P Gleghorn, Brian J Kirby. Rare cell capture in microfluidic devices. Chemical engineering science. 2011/4/1. Vol 66, Issue 7, 1508-1522.

[154] James P Smith, Alexander C Barbati, Steven M Santana, Jason P Gleghorn, Brian J Kirby. Microfluidic transport in microdevices for rare cell capture. Electrophoresis. 2012/11, Vol 33, Issue 21, 3133-3142.

[155] Jason P Gleghorn, James P Smith, Brian J Kirby. Transport and collision dynamics in periodic asymmetric obstacle arrays: Rational design of microfluidic rare-cell immunocapture devices. Physical Review E. 2013/9/26. Vol 88, Issue 3.

[156] Arnaldo Rodriguez-Gonzalez, Jason P Gleghorn, Brian J Kirby. Rational design protocols for size-based particle sorting microdevices using symmetry-induced cyclical dynamics. Physical Review E. 2020/3/18. Vol 101, Issue 3.

[157] Krithika Lingappan, Brielle Hayward-Piatkovskyi, Jason P Gleghorn. Neonatal lung disease: mechanisms driving sex differences. Sex-Based Differences in Lung Physiology. 2021, 115-144.

[158] Jasmine Shirazi, Joshua T Morgan, Erica M Comber, Jason P Gleghorn. Generation and morphological quantification of large scale, three-dimensional, self-assembled vascular networks. MethodsX. 2019/1/1. Vol 6, 1907-1918.

[159] Rachel M Gilbert, Laurel E Schappell, Jason P Gleghorn. Defective mesothelium and limited physical space are drivers of dysregulated lung development in a genetic model of congenital diaphragmatic hernia. Development. 2021/5/15. Vol 148, Issue 10.

[160] Jason P Gleghorn, Sriram Manivannan, Celeste M Nelson. Quantitative approaches to uncover physical mechanisms of tissue morphogenesis. Current opinion in biotechnology. 2013/10/1. Vol 24, Issue 5

[161] N'Dea S Irvin-Choy, Katherine M Nelson, Jason P Gleghorn, Emily S Day. Delivery and short-term maternal and fetal safety of vaginally administered PEG-PLGA nanoparticles. Drug Delivery and Translational Research. 2023/6/26. 1-11.