Aus dem Bereich Klinische Bioinformatik
Klinische Medizin
der Medizinischen Fakultät
der Universität des Saarlandes, Homburg/Saar

Systems Biology of Immune Adaptations in Aging, Disease and Spaceflight

Dissertation zur Erlangung des Grades eines Doktors der Naturwissenschaften der Medizinischen Fakultät

der UNIVERSITÄT DES SAARLANDES

2025

vorgelegt von Friederike Grandke
geb. am 07.08.1996 in Worms, Deutschland

Tag der Promotion:	23. Februar 2026
Dekan:	Prof. Dr. Matthias Hannig
Berichterstatter:	Prof. Dr. Andreas Keller
	Prof. Dr. Bergita Ganse
	Prof. Dr. Christian Brander

"The role of the infinitely small in nature is infinitely great."

- Louis Pasteur

ABSTRACT

Understanding how the immune system adapts across various tissues and conditions is essential for uncovering the underlying mechanisms of tissue adaptations and pathological inflammatory processes. This thesis studies the immune system as a network interconnected with tissues responding to local and global signals. Using high-throughput sequencing techniques, including single-cell RNA sequencing and bulk small non-coding RNA sequencing, we investigate immune responses and tissue adaptations in two main contexts: neurodegenerative diseases and physiological stress induced by spaceflight.

We analyzed single-cell RNA sequencing data from over 909k peripheral blood mononuclear cells from both healthy individuals and patients with Alzheimer's disease, Parkinson's disease, mild cognitive impairment, and overlapping conditions. Our findings reveal disease- and sex-specific changes in immune cell abundance and gene expression, highlighting peripheral immune signatures corresponding to transcriptomic changes within the brain. This supports the idea of neurodegeneration being influenced by immune dysregulation and chronic inflammation from a systemic view.

We used single-cell and bulk small non-coding RNA sequencing data from mice exposed to spaceflight to study tissue and immune adaptations under physiological stress. We identified tissue-specific and systemic changes, especially in processes such as extracellular matrix remodeling, membrane and cytoskeletal reorganization, and immune regulation. Comparisons between age groups indicated an impaired immune system in the older group. The miRNA data further strengthens our findings by revealing regulatory mechanisms involved in tissue adaptations.

We developed and implemented extensive bioinformatics pipelines to process, analyze, and interpret the complex and multi-modal datasets generated in this study. These ensured reproducible and scalable processing across

a large number of samples, cell types, and conditions. Our analysis framework integrated state-of-the-art analysis methods, including differential expression analysis, pathway enrichment, RNA velocity, and cell-cell communication analysis, enabling a deep understanding of the molecular processes in the cells. We performed cross-tissue and cross-condition comparisons that revealed common or connected molecular signatures through both early integration (at the count data level) and late integration (comparative analysis of downstream results). These analyses demonstrated the immune system's central role in maintaining systemic and tissue-specific homeostasis under diverse physiological and pathological conditions.

This thesis shows the value of combining single-cell and small RNA sequencing data with computational methods to reveal immune-related adaptations in health, disease, and other conditions. By showing how immune regulation connects tissue-specific responses to systemic physiology, we enhance our understanding of the complexity and adaptability of the mammalian immune system.

ZUSAMMENFASSUNG

Zu verstehen, wie sich das Immunsystem an verschiedene Gewebe und Bedingungen anpasst, ist wichtig, um die zugrunde liegenden Mechanismen von Gewebeanpassungen und pathologischen Entzündungsprozessen aufzudecken. In dieser Arbeit betrachten wir das Immunsystem als ein systemisches Netzwerk, das mit Geweben verbunden ist und auf lokale und systemische Signale reagieren. Mit Hilfe von Hochdurchsatz-Sequenzierung, wie der Einzelzell-RNA-Sequenzierung und der Bulk-Sequenzierung kleiner nicht kodierender RNAs, untersuchen wir Immunreaktionen und Gewebeanpassungen im Hintergrund von neurodegenerativen Erkrankungen und physiologischem Stress durch einen Weltraumflug.

Wir analysierten Einzelzell-RNA-Sequenzierungs-Daten von über 909 Tausend peripheren mononukleären Blutzellen von gesunden Personen und Patienten mit Alzheimer, Parkinson, leichter kognitiver Beeinträchtigung und überlappenden Erkrankungen. Unsere Ergebnisse zeigen krankheits- und geschlechtsspezifische Veränderungen in der Anzahl der Immunzellen und der Genexpression, wobei periphere Immunsignaturen hervorgehoben werden, die mit transkriptomischen Veränderungen im Gehirn übereinstimmen. Dies unterstützt die Vorstellung, dass die Neurodegeneration durch eine Dysregulation des Immunsystems und chronische Entzündungen aus systemischer Sicht beeinflusst wird.

Wir haben Einzelzell-RNA- (280k Zellen) und Bulk-Sequenzierung-Daten nicht kodierender RNAs (686 Proben), mit besonderem Augenmerk auf miRNAs, von Mäusen verwendet, die einem Weltraumflug ausgesetzt waren, um Gewebe- und Immunanpassungen unter physiologischem Stress zu untersuchen. Wir identifizierten gewebespezifische und systemische Veränderungen, insbesondere bei Prozessen wie dem Umbau der extrazellulären Matrix, der Reorganisation von Membranen und Zytoskelett und der Immunregulation.

Vergleiche zwischen den Altersgruppen deuteten auf ein geschwächtes Immunsystem in der älteren Gruppe hin. Die miRNA-Daten untermauerten unsere Ergebnisse, indem sie Regulationsmechanismen aufzeigten, die an Gewebeanpassungen beteiligt sind.

Um die komplexen und multimodalen Datensätze die in dieser Studie generiert wurden zu verarbeiten, zu analysieren und zu interpretieren, haben wir umfangreiche Bioinformatik-Pipelines entwickelt und implementiert. Diese gewährleisteten eine reproduzierbare und skalierbare Daten-Verarbeitung über viele Proben, Zelltypen und Bedingungen hinweg. Wir integrierten moderne Analysemethoden, wie die differenzielle Expressionsanalyse, die Anreicherung von Signalwegen, die RNA-Geschwindigkeit und die Analyse der Zell-Zell-Kommunikation, um ein tiefes Verständnis der molekularen Prozesse in den Zellen zu erlangen. Wir führten gewebe- und bedingungsübergreifende Vergleiche durch, die gemeinsame oder zusammenhängende molekulare Signaturen sowohl durch frühe Integration (auf den Expressions-Daten) als auch durch späte Integration (vergleichende Analyse der Analyse-Ergebnisse) aufzeigten. Diese Analysen betonen die zentrale Rolle des Immunsystems bei der Aufrechterhaltung der systemischen und gewebespezifischen Homöostase unter verschiedenen physiologischen und pathologischen Bedingungen.

Diese Arbeit zeigt, wie wertvoll die Kombination von Einzelzell- und anderen RNA-Sequenzierungsdaten mit computergestützten Methoden ist, um immunbezogene Anpassungen bei Gesundheit, Krankheit und anderen Bedingungen aufzudecken. Indem wir untersuchen, wie die Immunregulation gewebespezifische Reaktionen mit der systemischen Physiologie verbindet, verbessern wir unser Verständnis für die Komplexität und Anpassungsfähigkeit des Immunsystems von Säugetieren.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AD	**A**lzheimer's **D**isease
ADRC	**A**lzheimer's **D**isease **R**esearch **C**enter
BAT	**B**rown **A**dipose **T**issue
BBB	**B**lood **B**rain **B**arrier
BCR	**B**-**c**ell **R**eceptor
BH	**B**enjamini-**H**ochberg
CCA	**C**anonical **C**orrelation **A**nalysis
CNS	**C**entral **N**ervous **S**ystem
CSF	**c**erebrospinal **f**luid
DC	**D**endritic **C**ell
DEG	**D**ifferentially **E**xpressed **G**ene
ECM	**E**xtra**c**ellular **M**atrix
FACS	**f**luorescence-**a**ctivated **c**ell **s**orting
FBS	**f**etal **b**ovine **s**erum
FC	**F**old-**C**hange
FL	**I**SS **F**light group
GAT	**G**onadal **A**dipose **T**issue
GO	**G**ene **O**ntology
GSEA	**G**ene-**S**et **E**nrichment **A**nalysis
HGC	**H**abitat **G**round **C**ontrol
HSC	**H**ematopoietic **S**tem **C**ells
HVG	**h**ighly **v**ariable **g**ene
ILC	**I**nnate **L**ymphoid **C**ell
ISS	**I**nternational **S**pace **S**tation

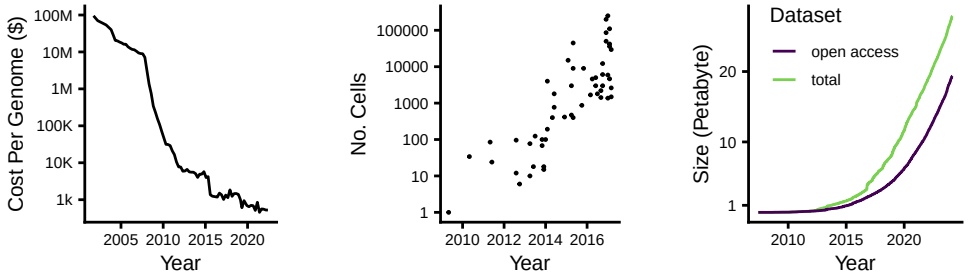
LAR	**L**ife **A**nimal **R**eturn
MAT	**M**arrow **A**dipose **T**issue
miRNA	**mi**cro **RNA**
mRNA	**m**essenger **RNA**
NK	**N**atural **K**iller
ORA	**O**ver-**R**epresentation **A**nalysis
PBMC	**P**eripheral **B**lood **M**ononuclear **C**ell
PBS	**P**hosphate-**B**uffered **S**aline
PCA	**P**rincipal **C**omponent **A**nalysis
PCR	**p**olymerase **c**hain **r**eaction
PD	**P**arkinson's **D**isease
PRR	**P**attern **R**ecognition **R**eceptor
PVCA	**P**rincipal **V**ariance **C**omponent **A**nalysis
RNA	**R**ibo**n**ucleic **A**cid
RNA-seq	**RNA seq**uencing
rpm	**r**eads-**p**er-million
rpmm	**r**eads-**p**er-**m**apped-**m**illion
rRNA	**r**ibosomal **RNA**
SCAT	**S**ubcutaneous **A**dipose **T**issue
scRNA-seq	**s**ingle-**c**ell m**RNA seq**uencing
sncRNA-seq	**s**mall **n**on-**c**oding **RNA seq**uencing
SOMA	**S**pace **O**mics and **M**edical **A**tlas
SRA	**S**equence **R**ead **A**rchive
TCR	**T**-**c**ell **R**eceptor
TERM	**I**SS **T**erminated
tRNA	**t**ransfer **RNA**
UMI	**U**nique **M**olecular **I**dentifier

VCA **V**ariance **C**omponent **A**nalysis
VGC **V**ivarium **G**round **C**ontrol

1

THE ROLE OF BIOINFORMATICS IN BIOLOGICAL AND MEDICAL RESEARCH

Since the development of the Sanger sequencing in 1977 [1], DNA sequencing has allowed us direct insights into the genomes of organisms. One of the most significant achievements was the Human Genome Project (1990-2003), which aimed to complete the first human genome [2]. This international project lasted 13 years and was estimated to have cost about 3 billion $. Since then, the landscape of sequencing has changed a lot. The introduction of second generation-sequencing (also called next-generation-sequencing) allowed parallel sequencing of millions of short reads and made sequencing a high-throughput and far more cost-efficient technology [3] (Figure 1.1a, 1.1b).



(A) Sequencing cost per genome (September 2001 - May 2022) [3].

(B) Number of cells captured by single-cell technologies (based on [4]).

(C) Size of SRA (June 2007 - February 2024) [5].

FIGURE 1.1: Development of sequencing technology over the years.

.

This development greatly impacted modern medicine and biology, leading to the era of data-driven research and bioinformatics. Since then, the number and complexity of the sequenced datasets has grown exponentially. The Sequence Read Archive (SRA) [6], for example, grew from 27.12 TB in January 2010 to 27.21 petabytes in January 2024 [5] (Figure 1.1c).

National and international consortia have established large-scale bioinformatics infrastructure to manage and leverage this data explosion. Projects

such as ELIXIR (EU) [7], NCBI & NIH (USA) [8], the DDBJ & NBDC (japan) [9], NGDC (China) [10] and the H3Africa (Africa) play a critical role in organizing, analyzing, and distributing biological data. These initiatives provide training, develop bioinformatics tools and pipelines, build open-access databases, and support data interoperability across disciplines and countries. For example, the European Life-Sciences Infrastructure for Biological Information (ELIXIR) [7] brings together over 240 institutes from more than 24 countries to coordinate life science data across Europe to support researchers by providing standardized pipelines, databases, and access to cloud computing resources through services such as the de.NBI cloud in Germany.

Even though decentralized pipelines like the ones provided by ELIXIR address some of the major challenges in modern bioinformatics, such as data fragmentation across experiments and labs, reproducibility of results, accessibility to the tools and results, and facilitation of cross-border collaboration, the usage of decentralized solutions and research is still necessary to drive research forward. Complementary to driving interoperability standards and a centralized system forward, custom decentralized solutions remain irreplaceable for cutting-edge, large-scale, or highly specific research questions. One example of these applications is single-cell RNA sequencing.

Modern DNA sequencing has enabled the exploration of genomes from thousands of species, cancer genomes, and microbiomes. The development of RNA sequencing, first on bulk and then at the single-cell level, allowed researchers to directly study transcriptomic changes in the cells. Today, studies often involve sequencing thousands to millions of cells, each characterized by thousands of features [4, 11, 12]. New technologies, such as single-cell multi-omics, further increase data complexity.

This explosion in data volume and dimensionality presents some major challenges [13]. Efficient storing, processing, and interpreting such datasets required the development of scalable bioinformatic pipelines, algorithms as well as specialized visualizations [14, 15]. These tools are essential both in biological research and in modern medicine.

In particular, bioinformatics allows a system-level understanding of gene expression across different cells and tissues, which is especially interesting when studying the dynamic and heterogeneous immune system. It involves various cell types with unique transcriptional and epigenetic profiles

shaped by their local tissue environments. Studying the immune system using single-cell technologies allows us to identify immune cell types and cell states, analyze cell-cell communication networks, and explore cellular trajectories such as differentiation of activation of cells [16]. This allows direct insights into the molecular pathways underlying immune responses, tissue homeostasis, and pathological conditions.

2

THE IMMUNE SYSTEM IN AGE AND DISEASE

The immune system plays a critical role in maintaining tissue homeostasis and defending against pathogens and infections. Beyond host defense, the immune system contributes to tissue repair, development, and organ homeostasis, including metabolism and regeneration [17]. These functions, though important to maintain organ function, can become dysregulated with age or disease, leading to decreased immune efficiency (immunosenescence) and chronic low-grade inflammation ("inflammaging"). This state can disrupt organ homeostasis and cause or accelerate age-related diseases, including metabolic disorders such as insulin resistance, neurodegenerative diseases, fibrosis, and organ dysfunction. Understanding the immune system's interactions with organs is essential for developing suitable therapies that treat age-related and chronic diseases.

2.1 CELLS OF THE IMMUNE SYSTEM

The immune system is a network of organs, cells, molecules, and processes that protect the organism from foreign and potentially harmful molecules and cells [18] such as microbes, viruses, toxins, and cancer cells [18].

The first line of defense of the immune system is physical barriers like the skin, the intestine, the lung epithelium, and the mucous membrane, complemented by antimicrobial proteins like C3, Defensine, and REGIII-γ. If a pathogen still crosses these barriers, the immune cells get activated. [19]

2.1.1 *Adaptive and innate immune cells*

These immune cells fall into two broad categories: innate and adaptive. Innate Immune responses happen in minutes to hours [18] and involve cells that eliminate pathogens, trigger inflammation, and present antigens to the

Cell type	Lineage	Circulating	Tissue resident	Function
Macrophages	Myeloid	✗(rarely)	✓	Phagocytosis; Release cytokines; tissue repair; antigen presentation
Monocytes	Myeloid	✓	✗	Differentiate into macrophages or DCs
Dendritic Cells	Myeloid	✓	✓	antigen presentation
Neutrophils	Myeloid	✓	✗	Phagocytosis; Degranulation
Mast Cells	Myeloid	✗	✓	Release histamines, enzymes and cytokines; Degranulation
Basophils	Myeloid	✓	✗	Release histamines, enzymes and cytokines; Degranulation
Eosinophils	Myeloid	✓	✓ / ✗	Degranulation; Release enzymes, growth factors and cytokines
NK Cells	Lymphoid	✓	✓ / ✗	Kills infected and tumor cells; Induces apoptosis
ILC	Lymphoid	✗	✓	Regulate immunity and tissue homeostasis; Release cytokines

TABLE 2.1: Innate immune cells and their function [18, 22].

adaptive immune system [18]. Adaptive responses take several days to initiate [18] but are highly specific and create a memory of previously seen antigens [20].

Innate immune cells

The key innate cell types include Macrophages, Dendritic Cells (DCs), Neutrophils, Natural Killer (NK) Cells, Innate Lymphoid Cells (ILCs), Mast Cells & Basophils (Table 2.1). Each cell type has a distinct function, like initializing inflammation, recruiting other immune cells, and triggering the adaptive immune response.

After pathogen detection by Pattern Recognition Receptors (PRR) such as Toll-like receptors on macrophages or dendritic cells [21], innate immune cells internalize and digest the pathogens and release cytokines and chemokines to alert and recruit more immune cells and activate inflammation. Dendritic cells migrate to the lymph nodes carrying the antigens and present these to T cells, initiating the adaptive response. [18]

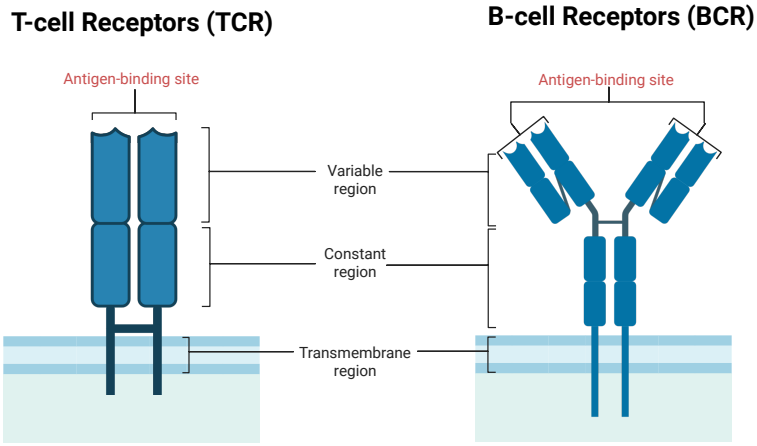


FIGURE 2.1: T-cell and B-cell Receptors in Comparison. Created with BioRender.com.

Adaptive immune cells

The adaptive immune cells include B and T cells. While B cells are responsible for antigen production, T cells have several different tasks, from coordinating the immune response (CD4+ helper T cells) to killing infected cells (CD8+ cytotoxic T cells) and even suppressing the immune response (Regulatory T cells). Both B and T cells rely on highly specific receptors, the T-cell receptors (TCR) and B-cell receptors (BCR), to recognize pathogens using antigens [18] (Figure 2.1).

Upon activation, B cells differentiate into plasma cells that segregate antibodies and memory B cells [23]. T cells undergo thymic selection for tolerance and specificity [24].

Interactions between innate and adaptive immune cells

Innate immune cells trigger the adaptive immune response by presenting antigens and secreting cytokines [25, 26]. Cytokines shape the development of T cells [27, 26], and the antibodies produced by B cells enhance the phagocytosis by macrophages and neutrophils [28], resulting in a bi-directional flow of information [26].

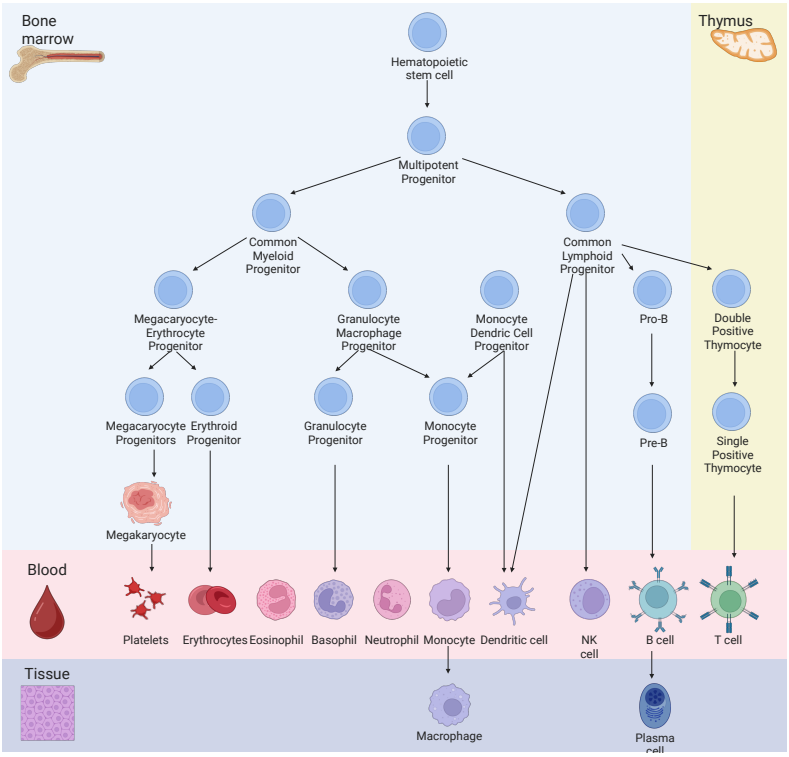


FIGURE 2.2: Development of immune cells from the bone marrow (Figure based on [29]). Created with BioRender.com.

2.1.2 *Origin and destination of immune cells*

Most immune cells originate from the hematopoietic stem cells (HSCs) in the bone marrow [19] (Figure 2.2). They divide into myeloid and lymphoid lineages and can be part of the innate immune system (macrophages, neutrophils, and dendritic cells), the adaptive immune system (T and B lymphocytes), or have a hybrid function.

However, some macrophage-like cells such as microglia, Kuffer, Langerhans, Alveolar, Peritoneal, and cardiac-resident macrophages are seeded during embryogenesis and persist into adulthood [30] without requiring continual input from the bone marrow [31]. The distinction between bone marrow-derived adult-origin immune cells and embryonic-origin immune cells is increasingly important in tissue-specific immunity [31], aging, neurodegeneration [32] and fibrotic diseases [33].

	Tissue-Resident Cells	Peripheral Immune Cells
Location	Stay within Tissues	Circulate in blood and Lymph
Mobility	Mostly stationary	Highly mobile
Longevity	Often long-lived (months to years)	Short-lived (innate: minutes to hours) or recirculating (pu to months)
Function	Local surveillance, homeostasis	Systemic defense, rapid development

TABLE 2.2: Properties of Tissue-Resident and Peripheral Immune Cells.

Immune cells either remain within specific tissues [34], performing local functions like debris clearance (e.g., microglia in the brain, Kupffer cells in the liver), or circulate through blood and lymph, providing constant systemic immunosurveillance and defense (Table 2.2).

2.2 REGULATION OF IMMUNE RESPONSES

To maintain immune balance and prevent pathology, the immune system is regulated at molecular, cellular, tissue, and systemic levels [35].

MOLECULAR REGULATION Molecular mechanisms of immune regulation refer to molecules that are involved in the signaling and detection of pathogens. Both innate and adaptive immune cells are activated using pattern recognition and antigen receptors, such as the PRRs in innate immune cells and the antigen receptors (TCR, BCR) in adaptive immune cells. Pro- and anti-inflammatory cytokines regulate the inflammation, and chemokines direct the immune cell trafficking into the affected tissues [36]. Dendritic cells secrete immune checkpoint molecules that activate T cell responses (positive checkpoint) or express inhibitory receptors that suppress these responses (negative checkpoint). This activity is essential under chronic stimulation or to maintain tolerance [18]. In addition to these molecules, a variety of transcription factors [37, 38] and epigenetic mechanisms such as histone modification [39, 40] and DNA methylation [41] influence immune cell differentiation and function.

CELLULAR REGULATION In addition to the immune regulation at the molecular level, other immune cells and the surrounding tissues also have

a regulatory role. Regulatory Immune cells such as T_{reg}, which suppress T cell activation, and B_{reg}, which modulate responses via IL-10, or the Myeloid-derived suppressor cells, that inhibit both innate and adaptive immune responses, play a key role in guiding the immune response [42, 43, 44]. Tissue-resident immune cells adapt to the local tissue signals and maintain tissue-specific tolerances.

TISSUE-LEVEL REGULATION The tissue environment generally has a significant influence on the immune cells, with cells like the stromal and epithelial cells [45] and the local tissue environment (e.g., pH, temperature, and hypoxia) further shaping the local immune system [34]. During the resolution of inflammation, the immune system shifts from a pro-inflammatory state to one promoting tissue repair and return to homeostasis. Lipid mediators such as resolvins and lipoxins help terminate the immune response actively, promoting the return to homeostasis after an inflammation [46].

SYSTEMIC REGULATION The systemic regulation of the immune system through neuro-immune interactions, like the modulation of the immune response through the vagus nerve or the influence of neural signals on cytokine production, introduces an additional layer of regulation [47]. The immune system is furthermore influenced by hormones such as cortisol (suppress immune response) or sex steroids (influence immune development and activity) [48, 49]. A feedback mechanism actively resolves inflammation and promotes tissue repair in the later stages of the inflammatory process. System-wide checks prevent the development of chronic inflammation or autoimmunity [50].

This multi-layer regulatory system ensures that immune responses are appropriate and targeted to prevent over or under-activation. This balance also depends on how immune cells interact with different organs, which help guide and support immune responses in specific body parts.

2.3 INTERACTIONS BETWEEN IMMUNE CELLS AND ORGANS

The immune system is integrated throughout the body, functioning as a decentralized yet coordinated surveillance network [51].

2.3.1 *Organ-specific immune landscapes*

Each tissue in the mammalian body creates its unique immune microenvironment. Resident immune cells, structural cells (such as epithelial, stromal, and endothelial cells), the specific function of the tissue, and exposure to external stressors all influence this environment [52]. Tissue-resident immune cells help maintain homeostasis by monitoring for stress or damage, modulating local inflammation, and supporting tissue-specific tasks like maintaining barrier integrity or regulating metabolism. Tissue-derived signals such as metabolites, cytokines, and neurotransmitters regulate these immune cells and drive local adaptations [53].

For example, the immune system in barrier tissues such as the lungs, gut, and skin is specialized to provide a fast defense against pathogens while maintaining a tolerance to harmless stimuli. In the lung, cells tolerate inhaled particles but should still provide a fast response to airborne pathogens [54]. The gut is constantly exposed to food and microbiota and promotes a regulatory response (oral tolerance) while preserving the defense mechanisms against pathogens [55]. As a physical and first line of defense, the skin is a fast-response barrier rich in resident immune cells such as dendritic and memory T cells [56].

The immune activity in the brain is carefully regulated to protect the sensitive structure of the brain from damage. The brain-resident microglia play a key role in the neurodevelopment, surveillance, and modulation responses to injury and disease [57].

On the other hand, the liver is a tolerogenic organ that regulates immune activity to prevent an over-reaction due to the constant exposure to antigens, for example, from the gut. As the organ that filters the blood, it uses the Kupffer cells and innate lymphocytes to balance immunity and tolerance [58].

In addition, some tissues have key roles in metabolism. Adipose tissues, for example, are central to metabolic homeostasis and regulate the immune responses context-dependently, such as shifting the immune state from homeostasis to inflammation. In health, the immune cells regulate insulin sensitivity but contribute to chronic low-grade inflammation in obesity, impairing metabolic regulation [59, 60].

2.3.2 *Functional contributions to tissue physiology*

Apart from the defense against pathogens, the immune system fulfills various tasks in the body that are central to maintaining tissue health. These range from development, homeostasis, and regeneration to adaptation to environmental and metabolic cues [17]. These functions are strongly tissue-dependent and tailored to supporting local tissue function [61].

DEVELOPMENT AND ORGANOGENESIS During tissue development in both embryonic and postnatal development, immune cells contribute to tissue formation. Particularly, macrophages contribute by cleaning up apoptotic cells, pruning synapses during brain development (microglia), and guiding endothelial tip cells to fuse during blood vessel formation (angiogenesis) [17, 62].

HOMEOSTASIS, TISSUE-SURVEILLANCE AND -REMODELING In order to maintain tissue homeostasis, tissue-resident immune cells constantly sense and respond to physiological changes. This includes maintaining the barrier function in the gut and skin by dendritic and T cells that reinforce the epithelial integrity and support tissue repair, cytokine production by resident macrophages and ILCs (e.g., in the gut to prevent over-activation) and the phagocytosis of both cellular and environmental debris for example by microglia in the brain, kupffer cells in the liver, alveolar macrophages in the lungs [31].

In tissue remodeling, macrophages regulate extracellular matric (ECM) turnover, aiding tissue remodeling and fibrosis [33]. Monocytes and macrophages also contribute to developing new blood vessels during wound healing and ischemia [62, 63].

REGENERATION AND REPAIR After a tissue injury, for example, in the skin, muscle, and liver, immune cells orchestrate the repair process in three main steps. Neutrophils and monocytes first respond to the injury and remove damaged cells and pathogens (initial inflammation). Macrophages then switch to the M2-like phenotype and secrete growth factors to promote

tissue remodeling. Regulatory T cells and ILC2s resolve the local inflammation and further activate tissue progenitor cells [64].

MODULATION OF LOCAL CELL BEHAVIOR Communication with epithelial, neuronal, and muscle cells significantly impacts immune cells' function. For example, microglia modulate neuronal activity and synaptic pruning in the brain based on environmental cues [65]. In the gut, ILC3s stimulate the production of protective mucus and antimicrobial peptides by epithelial cells [66], and immune-derived cytokines regulate the quiescence and activation of tissue stem cells [67].

STRESS RESPONSE AND ADAPTATION In general, immune cells act as sensors and respond to physical, metabolic, and microbial changes [61]. For example, tissue-resident macrophages can sense low oxygen and initiate adaptive angiogenic responses. In general stress situations, such as systemic inflammation, nutrient deprivation, and hormonal shifts, immune cells use cytokine and endocrine signaling to help tissues modulate their function [68].

METABOLIC ADAPTATION AND NUTRIENT SENSING In metabolic organs such as the liver, gut, and adipose tissue, immune cells help tissues adapt to nutrition and energy requirements. For example, macrophages and T cells are involved in nutrient sensing and respond to fatty acids, glucose, and microbial metabolites to adjust the baseline level of immune activation or immune readiness (inflammatory tone) [69]. In response to cold exposure and higher energy requirements, adipose tissue ILC2s and eosinophils promote the conversion to thermogenic fat [70]. Immune cells also influence insulin sensitivity [59]. In lean adipose tissue, anti-inflammatory, immune cells (e.g., T_{reg}, M2 macrophage) maintain the metabolic balance, while pro-inflammatory states contribute to insulin resistance in obesity [60].

2.3.3 *Systemic communication across organs*

Beyond the local immune function, the immune system forms a body-wide network that allows tissues to influence each other through immune-mediated pathways. This inter-organ communication synchronizes immune processes and supports systemic homeostasis. Immune activity in one tissue can also influence the physiology of others, positioning the immune system as both a regulator and communicator between organs, integrating local immune signals into systemic responses.

IMMUNE CELL CIRCULATION Immune cells distribute local information through circulation. Memory and effector cells are generated in response to local infections and can circulate through the body, providing broad immune protection and memory [71]. Monocyte-derived and dendritic cells can traffic to lymph nodes, inducing systemic T and B cell activation and thus have an indirect effect on other organs. Even tissue-specialized cells, such as gut-specialized T cells, may re-enter circulation and affect immunity in other barrier sites [72]. This dynamic redistribution connects immune memory and adaptations across tissues, translating local immune experiences into systemic preparedness.

EXTRACELLULAR VESICLES Tissues also release extracellular vesicles, including exosomes, which circulate immune-modulating molecules like cytokines and microRNAs [73, 74]. These vesicles shape immune responses in distant tissues, modulating inflammation, tolerance, and regeneration [74].

FUNCTIONAL IMMUNE AXES Specific tissues form recurrent immunological relationships known as functional immune axes. Examples include the gut–liver axis, the lung–gut axis, the skin–brain axis, and the adipose–liver–muscle axis. For instance, gut-derived metabolites influence liver-resident immune cells [75], disruptions in intestinal immunity or microbiome composition alter the pulmonary immune response [76], and chronic skin inflammation affects neuroinflammation and behavior [77]. Adipose inflammation can impair insulin sensitivity and affect immune function in the liver and muscle

[78]. These axes illustrate how immune activity in one tissue can influence distant organs.

INFLUENCE OF METABOLISM Microbiota-derived metabolites circulate to organs such as the lung, brain, and bone marrow, modulating the immune tone [79]. Inflammation can trigger emergency hematopoiesis, shifting bone marrow output toward specific immune lineages such as neutrophils and monocytes [80]. On the other hand, immune responses can lead to higher energy requirements, which subsequently trigger changes in the metabolic activity in the liver, fat, and muscle tissues [81].

The immune system thus plays a central role, translating local immune activity into systemic effects.

2.4 THE IMMUNE SYSTEM THROUGHOUT LIFE

The immune system is highly dynamic and adaptable throughout the human lifespan (Figure 2.3). From fetal development until old age, the immune system undergoes several structural and functional transformations [82]. This adaptability is essential because it provides lifelong protection across various physiological stages and allows immune responses to a constantly changing environment, including pathogens, allergens, and beneficial microbes [82]. As the organs age, the immune system must adapt to counter declining function and rising inflammatory signals (inflammation), contributing to age-related diseases such as cancer, cardiovascular disease, and neurodegeneration [83].

Understanding how the immune system changes throughout life is essential to understanding the fundamentals of human immunology and addressing challenges such as vaccines and age-related disorders.

2.4.1 *Immune development and adaptation*

From conception, the immune system develops and matures parallel to the body's growth. During fetal development, immune organs such as the bone marrow, thymus, and spleen begin to form. However, the immune system

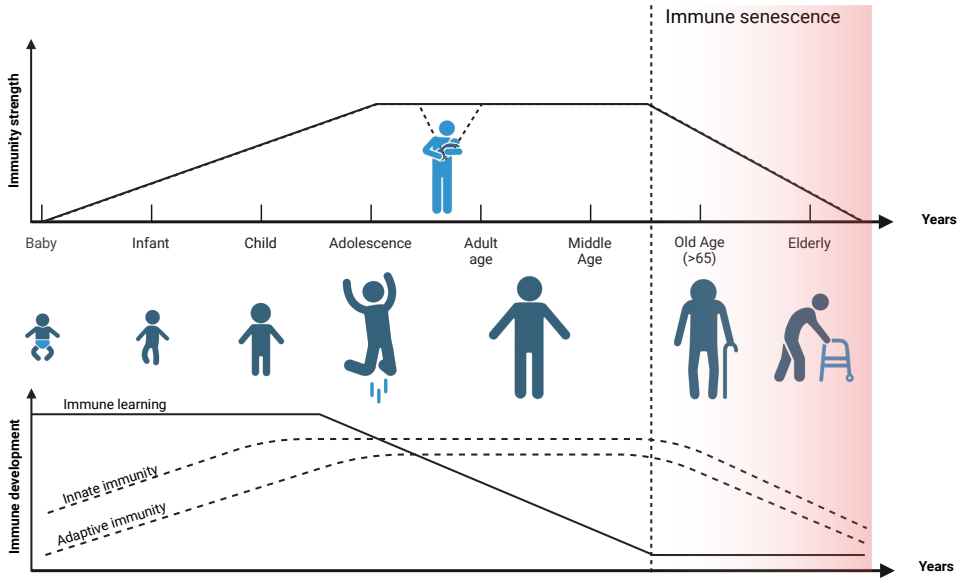


FIGURE 2.3: The Immune System throughout life (based on [84]). Created with BioRender.com.

at this stage is still skewed toward tolerance and is less likely to react to pathogens. After birth, the adaptive immune system is still immature, and the immune response mainly relies on innate immunity, supported by maternal antibodies provided via the placenta or breast milk. Once the child progresses into childhood, the adaptive immunity gradually matures and builds an immunological memory [85].

Sex hormones like estrogen and testosterone shape immunity, leading to generally stronger adaptive responses in females and higher innate cytokine activity in males [86, 87].

While infants have a higher risk for infections, older adults show a decreased efficacy of vaccines and an increased risk of chronic diseases. Older adults especially show a rising probability of developing cancer, chronic infections, and neurodegenerative diseases [83].

2.4.2 *The aging immune system*

As the immune system ages, it undergoes structural and functional changes that reduce its ability to effectively respond to infections, clear pathogens, and maintain immune homeostasis [88].

As people age, their immune function gradually declines, known as immunosenescence. This decline is characterized by a reduced ability to produce naive T and B cells, along with changes in the thymus, decreased immune cell diversity (resulting in fewer lymphocytes and more myeloid cells), and the accumulation of memory and effector cells. Over time, these changes impair the body's ability to clear pathogens and respond effectively to vaccines.

Alongside immunosenescence, another hallmark of aging is inflammation. It refers to chronic low-grade inflammation driven by immune cell aging, senescence, and changes in the microbiome and is associated with diseases like arteriosclerosis, neurodegeneration, and osteoarthritis. [89, 90, 91]

Aging also significantly impacts the structure of the Extracellular Matrix. Aging stiffens the extracellular matrix by damaging collagen and elastin glycation, cross-linking, and fragmentation, impairing immune cell migration and tissue repair [92]. These changes lead to increased ECM stiffness, impairing its ability to support immune cell migration, signaling, and tissue repair [92]. Immunosenescence and ECM remodeling reinforce each other, resulting in compromised tissue integrity, increased disease vulnerability, and impaired regeneration in aged individuals [92].

All twelve hallmarks of aging affect or are affected by the immune system, from genomic instability and stem cell exhaustion to mitochondrial dysfunction and chronic inflammation [93] (Figure 2.4).

2.4.3 *Neurodegenerative diseases*

Neurodegenerative disorders such as Alzheimer's disease (AD) and Parkinson's disease (PD) share overlapping mechanisms, including protein aggregation, immune dysregulation, and neuronal dysfunction. [94, 95]

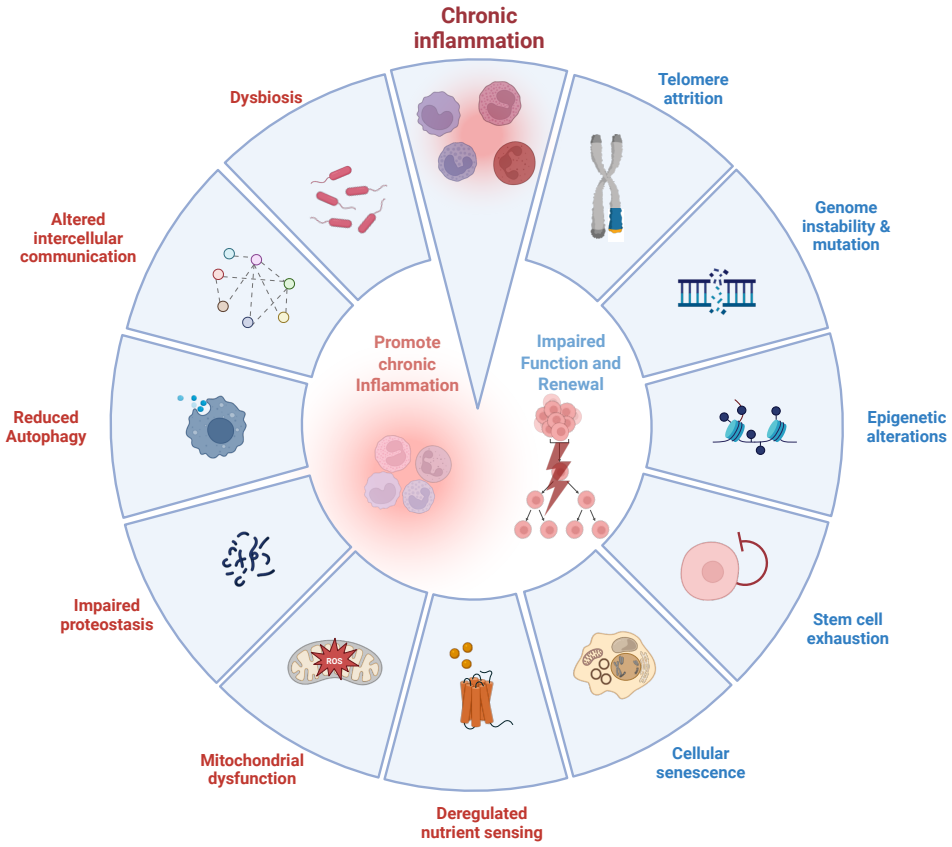


FIGURE 2.4: Hallmarks of aging (Figure based on [93]). Created with BioRender.com.

ALZHEIMER'S DISEASE Alzheimer's disease (AD) is the most common cause of dementia worldwide and results in progressive memory loss and cognitive decline. It is associated with extracellular amyloid-β (Aβ) plaques, intracellular tau tangles, and widespread synaptic and neuronal loss [96]. Alzheimer patients show increased microglial clustering around plaques, indicating both clearance and chronic inflammation [97]. Peripheral inflammation and the leakage of cytokines through the blood-brain barrier (BBB) may further amplify neuroinflammation. Women have a higher risk and more severe cognitive decline in Alzheimer's disease [98, 99].

PARKINSON'S DISEASE In contrast to that, Parkinson's Disease (PD), the second most common neurodegenerative disorder, is more prevalent in men [100]. It results in both motor symptoms, such as bradykinesia and tremor, and non-motor symptoms, such as apatite, sleep problems, depression, and cognitive decline [101]. Parkinson show loss of dopaminergic neurons and α-synuclein accumulation, with mounting evidence of immune contribution such as microglial activation, MHC-I expression, T cell involvement, and peripheral immune changes [100].

Aging is the most significant risk factor for most neurodegenerative diseases and is associated with widespread immune remodeling [102, 103]. The chronic low-grade inflammation ("inflammaging") promotes neurodegeneration, the aged microglia become less efficient in debris clearance and more pro-inflammatory, and the increased cytokine and immune cell translocation across a weakened BBB further promote inflammation [104]. Under homeostasis, the blood-brain barrier limits the immune entry to the brain. However, the BBB breakdown during disease allows T cells and monocytes to infiltrate the Central Nervous System (CNS). Peripheral immune activations could thus indirectly influence CNS inflammation [105].

Understanding these regulatory mechanisms is crucial, given the immune system's complexity, adaptability, and involvement in diverse physiological processes. This requires methods to capture dynamic gene expression changes across different cell types, tissues, and conditions. Transcriptomic technologies like RNA sequencing provide a robust framework to dissect these regulatory mechanisms and uncover immune cell states and functions at high resolution.

2.5 RNA SEQUENCING IN IMMUNE SYSTEM RESEARCH

RNA sequencing (RNA-seq) offers a way to study gene expression in immune cells under different conditions and across tissues. Providing a snapshot of gene expression in a given cell or tissue allows the identification of condition-specific shifts in the transcriptomic profiles and the detection of disease-associated pathways and signals. In addition, miRNA profiling adds insights into the regulatory mechanisms underlying the transcriptomic changes.

The development of single-cell sequencing enabled us to separate the RNA profiles of single cells. This massive increase in resolution allows the detection of rare and transient cell types and states and the mapping of pathways and immune responses to specific cell types. For example, Peripheral Blood Mononuclear Cells (PBMCs) are cells that can be extracted from blood samples of patients, providing a low-invasive access point to study disease-related changes in circulating immune cells on the cell type level.

3

ANALYSIS OF RNA SEQUENCING DATA

Studying Ribonucleic Acid (RNA) in cells and tissues has fundamentally changed our understanding of gene expression and regulation. RNAs play a key role in the cell and are involved in processes from coding genetic information to regulation of gene expression. The various types of RNA, such as mRNAs, tRNAs, rRNAs, and a variety of small non-coding RNAs, all have different functions in a cell. High-throughput sequencing technologies allow us to sequence a big part of the transcriptome from a sample or cell simultaneously, providing insights into gene expression, splicing patterns, and RNA modifications. This technology allows us to capture the mRNA profiles of tissues at the single-cell level (single-cell sequencing) or the small non-coding RNA fingerprints at the sample level (bulk sequencing).

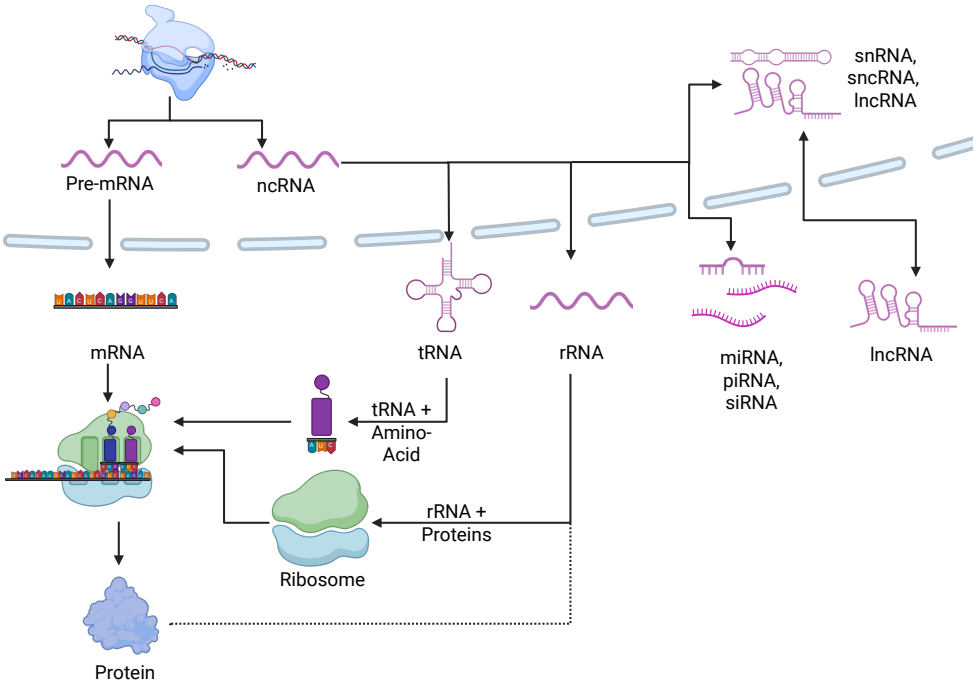


FIGURE 3.1: Overview of the Types of RNA in the cell. Created with BioRender.com.

3.1 INTRODUCTION INTO RNA SEQUENCING

Ribonucleic Acids (RNA) are molecules in a cell that fulfill various roles, mainly depending on the type of RNA. These include information transfer from the nucleus to the ribosomes, transporting amino acids, and acting as an important building block.

3.1.1 *Types of RNA and its role in a cell*

The three most frequent RNAs in a cell are messenger RNAs (mRNA), transfer RNAs (tRNA), and ribosomal RNAs (rRNA). Other RNAs are non-coding RNAs (ncRNA), like lncRNAs, miRNAs, or piRNAs.

RNA-family	**Abbr.**	**Size (nt.)**
Messenger RNA	mRNA	40k
Ribosomal RNA	rRNA	121-5k
long non-coding RNA	lncRNA	>200
Transfer RNA	tRNA	74-95
small nuclear RNA	snRNA	100-300
small nucleolar RNA	snoRNA	60-300
small Cajal body-specific RNA	scaRNA	120-300
micro RNA	miRNA	22
PIWI-interacting RNA	piRNA	26-30
transfer RNA fragment	tRF	14-32

TABLE 3.1: Types or RNA in Mammals (numbers from [106, 107, 108, 109]).

Transfer RNAs play a crucial role in translation, as they deliver amino acids but are usually not captured in RNA sequencing. Ribosomal RNAs form the structural and functional core of ribosomes and are often depleted in RNA sequencing due to their high abundance. Single-cell RNA sequencing usually aims to capture the messenger RNAs in a cell, which allows us to infer the expression profile in the cells and draw conclusions about the cellular functions. Messenger RNAs carry the genetic information from the DNA (nucleus) to the ribosomes. They are the product of gene transcription and are the template for the translation into proteins [110]. Non-coding RNAs such as miRNA, piRNAs, tRFs, snoRNAs, and snRNAs play an important

role in regulating gene expression, translation, and RNA processing. Understanding this set of RNAs thus provides insights into the regulation of biological processes in the tissues and organisms. In combination with mRNAs, the class of miRNAs is of special interest. These about 20 nucleotide-long RNAs can down-regulate their target mRNAs by binding to their 3′ UTR, suppressing their translation and speeding up the degradation. They regulate mRNAs in a tissue-specific manner and thus play a critical role in the gene expression patterns determining tissue functions.

3.1.2 *Regulation of mRNAs in the cell*

During the life of a mRNA, countless factors influence the RNA molecule. In the cell nucleus, the RNA polymerase transcribes the DNA into mRNA [111]. The binding of promotors, enhancers, and transcription factors regulate the transcription of RNAs [112]. Transcription factors are proteins that bind to the DNA sequence near a gene and inhibit its translation. We can observe this targeted suppression in single-cell data by linking the expression level of the mRNA encoding the transcription factor to the down-regulation of the targeted gene. After transcription, the cell processes the mRNA through capping, splicing, and polyadenylation [110]. This poly(A) tail is crucial for capturing the mRNA during scRNA-seq library preparation. The mature mRNA then gets exported into the cytoplasm, where ribosomes translate it into proteins [111]. At this stage, miRNAs and other non-coding RNAs can regulate the mRNA level and translation [110]. Eventually, the cell degrades the mRNAs. Therefore, RNA sequencing can only provide a snapshot of the mRNA level at a specific time.

3.1.3 *RNA sequencing*

mRNA abundance does serve as a proxy for gene expression but is not always predictive of protein levels. Other factors strongly influence protein levels, including the regulation of translation and the stability of RNA and proteins. So, what does RNA sequencing tell us about a cell? It provides information about the regulation of the mRNA levels to detect cellular states

that reflect cell differentiation, activation, stress responses, etc. It also allows us to infer disease mechanisms and classify subtypes that reflect disease states and serve as biomarkers. High-throughput sequencing allows us to measure the RNA abundance of thousands of RNAs at once, thus giving us an unbiased view of the current transcriptional profile in a cell or tissue.

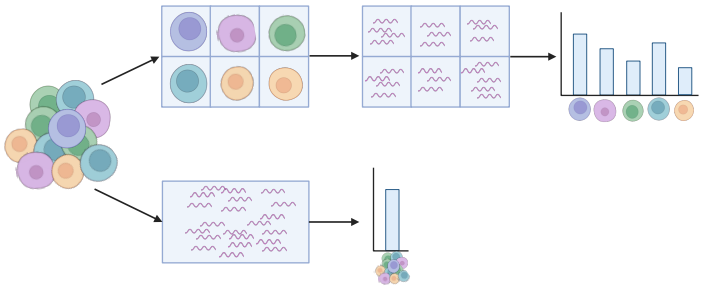


FIGURE 3.2: Comparison between single-cell and bulk RNA sequencing (Figure based on [113]). Created with BioRender.com.

SINGLE-CELL VS. BULK RNA SEQUENCING RNA sequencing can be performed at different resolutions: at bulk level or single-cell level resolution (Figure 3.2). Bulk RNA sequencing means we take all the RNA from a sample (solid tissue, body fluids, or even FACS-sorted cells) and simultaneously measure the RNA abundance in the whole sample, averaging the expression across many cells. This technology thus provides a high coverage but does not distinguish between different cell types. Single-cell mRNA sequencing captures the RNA expression for each cell separately, enabling the identification and characterization of rare cell types and dynamic states, but with a much lower coverage [14].

SINGLE-CELL MRNA SEQUENCING To perform single-cell mRNA sequencing (scRNA-seq) on human or animal samples, the first step is to isolate single cells from a tissue (Figure 3.3). Solid tissue samples usually require tissue digestion before the cells are separated. One method to separate the cells is the droplet-based method. These platforms use a microfluidic chip to separate individual cells and the surrounding solutions. They use oil to encapsulate each cell, a bead containing the cell barcodes and UMIs,

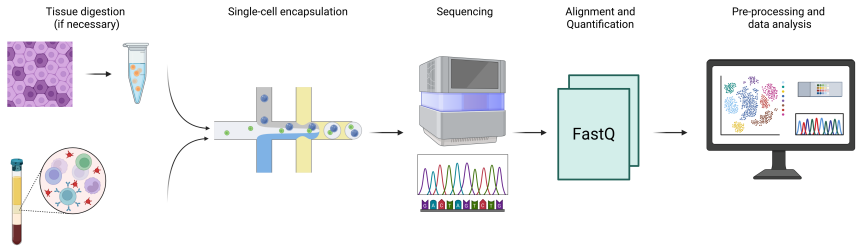


FIGURE 3.3: Single-cell RNA sequencing workflow with droplet-based cell isolation. Created with BioRender.com.

and the solution containing the required chemicals in one droplet. These droplets thus create a separate reaction chamber for each cell, parallelizing the high-throughput processing of thousands of cells. After separating and lysing the cells, the mRNA is captured using poly(t) primers that bind to the mRNAs' poly(A) tail. These primers have attached barcodes and unique molecular identifiers (UMIs). The primed mRNAs are converted into cDNAs via reverse transcription, amplified (usually with polymerase chain reaction (PCR)), and sequenced. The previously attached cell barcodes and UMIs allow us to separate later and group sequences from the different cells and RNA molecules. [114]

BULK SMALL NON-CODING RNA SEQUENCING The bulk small non-coding RNA sequencing targets the cell RNAs' small non-coding RNA fraction. As the library preparation for our small non-coding RNA sequencing (sncRNA-seq) focuses on RNA molecules and fragments in the typical size range of a miRNA (usually 50 nucleotides), we focus the analysis mainly on miRNAs. As this method filters the RNAs using size selection, this also means that we capture other small non-coding RNAs such as piRNAs and tRFs, but also RNA fragments from tRNAs, rRNAs, snoRNAs, scaRNAs, and lncRNAs (Table 3.1).

3.2 PROCESSING OF RNA SEQUENCING DATA

After the sequencing, the raw sequencing data are mapped against a reference, and the number of RNA molecules is counted. Given the count matrices, quality control is performed to remove low-quality cells and uninformative RNAs and correct for technical noise. The data is then clustered, and cell types are annotated to identify meaningful cell populations to prepare the data for the downstream analysis.

3.2.1 *Alignment and quantification*

The first step in this process, after the sequencer generates the raw sequencing files (.fastq files), is to determine which RNA molecules were captured (Alignment) and how many of them were captured (Quantification).

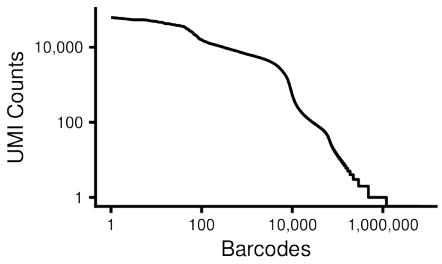


FIGURE 3.4: Barcode Rank Plot by CellRanger.

A frequently used pipeline for aligning and quantifying single-cell data is `Cellranger` (10X). It performs multiple steps for pre-processing [115]. To align the reads to the transcriptome, read trimming removes the TSO sequence (5' end) and the sequences' poly-A tail (3' end). The splice-aware alignment algorithm STAR aligns the reads to the genome, and the alignment is adjusted to prefer the mapping to exonic loci (MAPQ adjustment). The reads are then aligned to the transcriptome and classified as exonic, intronic, sense, or antisense. The 10X barcodes are compared against a whitelist and corrected in case of a mismatch. The reads are grouped by barcode, UMI, and gene annotation, and sequencing errors are corrected. Each gene, UMI, and barcode combination is one count in the count matrix. After the alignment, Cellranger removes cells with a very high RNA content (OrdMag algorithm) and a very low RNA content that are probably empty droplets (EmptyDrops algorithm)(Figure 3.4). This step is called the cell calling.

For the bulk small non-coding RNA data, the sequences are mapped against a reference using methods like `Bowtie` or `STAR`, and the reads are counted per RNA and sample after filtering out low-quality mappings [116].

3.2.2 *Removing low-quality cells and genes*

Quality control ensures that all samples are sufficiently high quality and comparable across conditions. Pre-processing in single-cell data starts with removing doublets and ambient RNA contamination to ensure that each signal represents only the content of one cell. To ensure high-quality data, different pre-processing steps were performed to orient themselves to the state-of-the-art best practices in single-cell data analysis [114, 15].

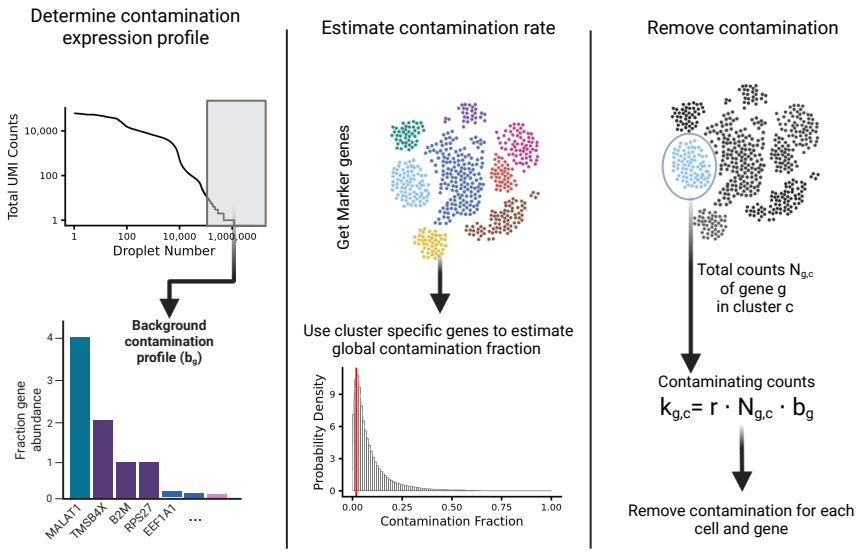


FIGURE 3.5: Removal of ambient RNA contamination using SoupX (based on [117]). Created with BioRender.com.

AMBIENT RNA CONTAMINATION The first pre-processing step for the single-cell RNA-seq data is the removal of ambient or cell-free RNA contamination using `SoupX` [117]. Ambient (or cell-free) RNA is part of the input solution and is captured along with a cell [117]. Once inside a droplet, the ambient RNA is barcoded and amplified as one of the cell's mRNAs [118]. This contamination usually comes from stressed, broken, or dead cells [118] and can confound the biological interpretation [117, 118]. The amount of contamination varies between experiments and samples, depending on factors like the tissue type [119], and is especially high in single-nucleus samples

[120] and necrotic tissues [117]. Tools such as `SoupX` [117], `FastCAR` [121], and `DecontX` [118] address this issue and aim to identify and remove the background contamination. SoupX automatically estimates the contamination in each sample using empty droplets to characterize the surrounding fluid's expression profile and adjusts the cells' gene counts by subtracting the estimated contamination (Figure 3.5).

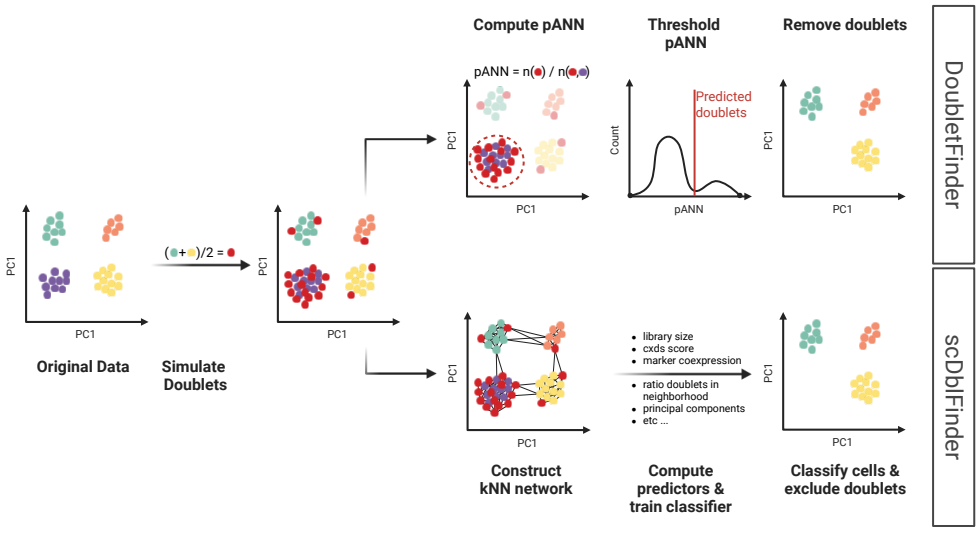


FIGURE 3.6: Identification of Doublets and Multiplets using Doublet-Finder [122] and scBdlFinder [123]. Created with BioRender.com, based on [122].

DOUBLETS AND MULTIPLETS Another important processing step is identifying and removing doublets (droplets with two cells) and multiplets (droplets with more than two cells), a common artifact in droplet-based platforms. These droplets mix the transcription profiles from two or more distinct cells (possibly different cell types) and confound the downstream data analysis [124, 114, 125]. To address this issue, a growing number of methods have been developed [126]. For our data, we used `DoubletFinder` [122] and `scDblFinder` [123]. `DoubletFinder` (Figure 3.6) simulates doublets by averaging the transcription profile of two randomly selected cells. It then determines how many simulated doublets fall into the k-nearest neighbors for each cell. Cells with a proportion higher than a threshold are predicted to be doublets. `scDblFinder` uses a similar approach, simulating doublets and

constructing a neighborhood graph. In a neighborhood graph, each node is a data point (e.g., a cell), and the edges connect nodes that are nearest neighbors. In contrast to `DoubletFinder`, `scDblFinder` does not solely rely on a fixed neighborhood size and uses a combination of predictors (e.g., library size, marker co-expression, Doublet ratio in neighborhoods) to train a classifier. Once it identifies the doublets, they can be visualized and filtered out.

CELL FILTERING Doublets, multiplets, and empty droplets can also be identified using the mRNA content. Doublets show a very high mRNA content and empty droplets have a very low one. Usually, filters use the number of captured reads or the number of found genes. Another filtering criterion is the proportion of mitochondrial RNA. A high proportion of mitochondrial RNA in a cell can indicate cell stress [127], often coming from dying, stressed, or broken cells. Removing these cells is an important step in ensuring sound data quality and avoiding bias in the analysis. An early-established threshold for filtering out cells is 5% of mitochondrial reads in a cell. However, the mitochondrial RNA proportions differ depending on the tissue and organism of origin, showing higher values in cells with a high metabolic activity [128]. Osorio and Cai (2021) [127] highlighted the variance between the proportion of mitochondrial RNA in different healthy tissues and organisms and proposed a threshold of 5% for mice and 10% for human tissues.

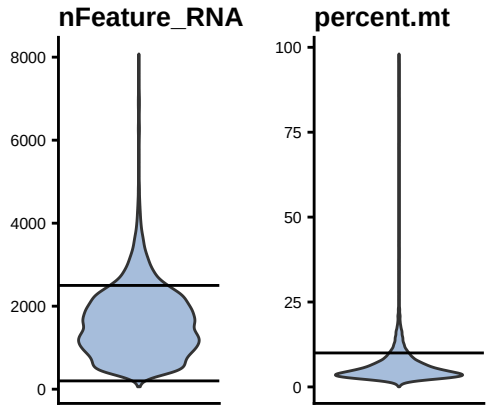


FIGURE 3.7: Number of features (genes), counts (mRNA molecules), and the percentage of mitochondrial reads in a single-cell dataset and the thresholds used for filtering

3.2.3 *Normalization and batch-correction*

Another major challenge in single-cell data is the biological and technical noise in the data [129, 130]. Technical noise is variation in the data introduced by sample and library preparations, sequencing, the environment, or other

technical variables. Due to the low amount of starting material in single-cell sequencing [130], these factors can strongly influence the data and need to be corrected. In contrast, variation in single-cell data can also stem from biological variations [129]. These come from stochastic processes like transcriptional "bursts" [131, 132] or non-stochastic processes such as the cell cycle, signaling, and stress responses [129]. Because these biological signals might define partially unknown cell states, metabolic functions, or cell identities [129], our primary goal is to distinguish biological variations from the technical noise [129].

NORMALIZATION To correct variations in the number of RNA molecules captured for each cell (or sample), we perform a Normalization [114].

In the single-cell data, we used log normalization. Given the counts $\text{count}_{g,c}$ of gene g and cell c, the log normalization is defined as:

$$\text{log-norm}(\text{count}_{g,c}) = \ln\left(\left(\frac{\text{count}_{g,c}}{\sum_g \text{count}_{g,c}} \cdot f\right) + 1\right) \tag{3.1}$$

where f is a scaling factor (usually 10,000). Using the logarithm of expression values means that distances between two genes can be interpreted as fold-changes, helps to reduce the variance-mean dependency in the data [130], and helps to approximate a normal distribution of the data, a frequent assumption of many downstream analysis methods [114]. After normalization, the single-cell data is scaled so that each gene has a mean expression of 0 across cells and a variance of 1 to reduce the influence of highly expressed genes on the downstream analysis.

For the bulk data, the counts are normalized using a reads-per-mapped-million (rpmm) normalization. For sample s and gene g, is defines as:

$$\text{rpmm}(\text{count}_{g,s}) = \frac{\text{count}_{g,s}}{\sum_{g,s} \text{count}_{g,c}} \cdot 10^6 \tag{3.2}$$

In contrast to the reads-per-million (rpm) normalization, which uses the total number of sequenced reads, we use the total reads mapped to the reference genome here. For most downstream analyses, the counts are $\log_2$ normalized.

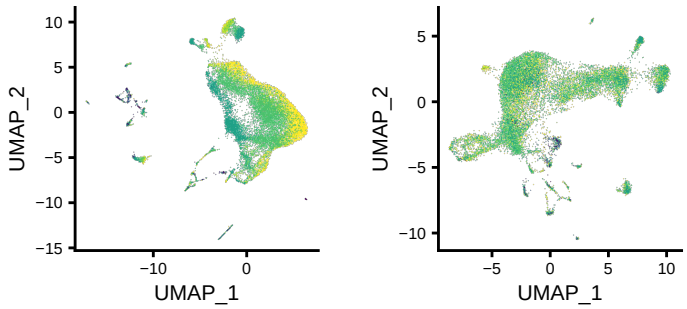


FIGURE 3.8: Visualization of Batch effects in a UMAP embedding and the corrected embedding (Thymus from [F1])

INTEGRATION AND BATCH CORRECTION Variations caused by differences in experimental conditions can not always be avoided when performing large sequencing experiments and can not be corrected using normalization. These differences can cause so-called Batch Effects. Batch Effects describe differences between the gene expression of a cell (or sample) in one batch and from a cell (or sample) in another batch caused by differences in the way the data was generated (e.g., different time-point, different operators, different laboratories, different methods or materials or different sequencing platforms) [133, 134]. These differences between batched can introduce spurious results, especially with methods such as dimensionality-reduction and clustering that rely on similarities in the gene expression of cells [135] (Figure 3.8).

We handle batch effects in our single-cell data by using Data Integration, thus performing batch correction by projecting it into a shared space. Methods that implement data integration are, for example, `Harmony` [136] and Seurat's canonical correlation analysis (CCA) based Integration [137]. Other methods for batch correction, such as `ComBat` [138], only correct the count data of the dataset without combining the datasets in a shared space.

Harmony uses a low-dimensional embedding of the cells as input (Figure 3.9), for example, using Principal Component Analysis (PCA). It uses soft clustering to group cells into clusters containing multiple datasets. Each dataset's cluster-specific centroid is used to determine a linear correction factor for each cluster, which is then used for each cell to correct it. These steps

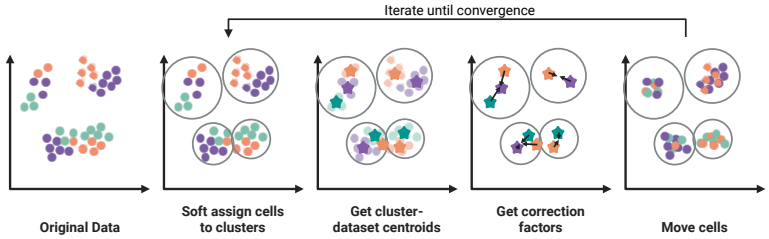


FIGURE 3.9: Batch-Correction using `Harmony`. Created with BioRender.com.

are repeated until the clustering of cells remains stable, and the embedding can then be used for downstream analysis [136].

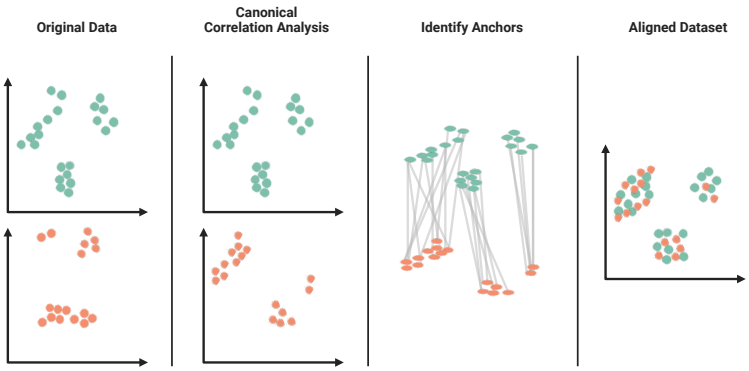


FIGURE 3.10: Data integration using Seurat's CCA-based integration method. Created with BioRender.com.

In contrast, Seurat's CCA-based Integration method uses the gene expression count matrix as input (Figure 3.10). It identifies conserved gene expression patterns across datasets or batches by learning a correlation structure using canonical correlation analysis. It identifies anchor cells using mutual nearest neighbors to find corresponding cells across datasets and uses these to transform and align these datasets in a common feature space [137].

DROPOUTS Another characteristic of single-cell data is so-called dropout events. A dropout event occurs when, in the same cell type, a gene is detected in one cell but not in a different cell [139], caused by the low amount of mRNA in the cell or the stochasticity of the gene expression [140]. These dropouts are frequently reported as noise in the data, but correcting it

with data imputation or normalization can introduce additional noise [141]. Dropout patterns are cell type specific, and most dropouts disappear when splitting the dataset by cell-types [141, 140]. This signal can thus be used to identify cell types [140].

3.2.4 *Reducing the dimensionality of the datasets*

Even though single-cell technologies can capture thousands of cells and tens of thousands of genes in one experiment, sequencing usually captures no more than 60% and, on average, less than 10% of the transcriptome [142, 143]. This causes dropouts and zero counts in the data, resulting in a sparse count matrix. Combining the counts of up to 25,000 genes [114] of thousands to millions of cells, the dimensionality of the data can become a problem when analyzing and handling the data. One reason for this is the so-called Curse of dimensionality [144]. It describes the problem of high dimensional data where a larger feature space results in an explosion of possible combinations of features. This results in a low (or even zero) number of data points covering the pattern of interest [145], making it difficult to identify hidden structures in the data. To reduce the noise in the data [146] and enable better visualization [114], we reduce the dimensionality of the data. Even though the dimensionality of bulk RN-seq data is not as high as that of single-cell data, dimensionality reduction can help visualize the data.

Reducing the dimensionality of single-cell data can be achieved by either reducing the number of genes used as input for subsequent analysis (Feature Selection) or by using a low-dimensional embedding (Dimensionality Reduction) of the data to define similarities and relationships in the data [147].

FEATURE SELECTION The goal of feature selection is to identify a subset of useful features (i.e., genes) from the captured transcriptome to reduce the dimensionality, remove redundancies, and prevent issues such as overfitting [148]. Feature selection methods can rely on the coefficient of variation, the average expression, or even high dropout rates [147]. For the single-cell data, we mainly use the highly variable genes (HVGs) as implemented in the `FindVariableFeatures` function in Seurat. This function performs three major steps: variance calculation, gene ranking, and selection of the top features

[149]. For this, the method first corrects for the mean-variance relationship (Figure 3.11).

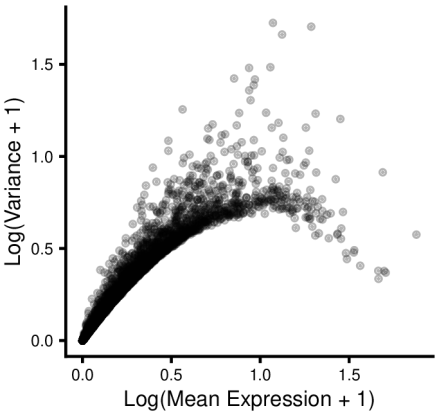


FIGURE 3.11: The mean-variance relationship in single-cell data.

Using the raw counts, it performs a $\log_{10}$ transformations and fits a curve (polynomial of degree 2) to predict a feature's variance based on its mean. Using this estimator for the expected standard deviation σ_i of a feature g, it calculates a standardized value $\tilde{e}_{g,c}$ of feature g in cell c:

$$\tilde{e}_{g,c} = \frac{e_{g,c} - \bar{e}_g}{\sigma_g}, \tag{3.3}$$

with $e_{g,c}$ the raw expression value of feature g in cell c and $\bar{e}_g$ the mean raw expression for feature g. The standardized values are clipped at $\sqrt{N}$ (N is the total number of cells). Using these standardized values, the method calculates the variance for each gene across all cells and uses these to rank and filter for the highest variable genes [149].

A frequently used method to filter for the most meaningful RNAs in the bulk data is to consider only RNAs captured in a large proportion (usually 50%) of samples in each group. This reduces the number of RNAs that must be considered, especially in cross-tissue and cross-condition comparison, leaving only those found in all conditions.

DIMENSIONALITY REDUCTION Dimensionality Reduction aims to project the still high-dimensional data into a lower-dimensional space using algorithms such as Principal Component Analysis (PCA) [150], t-distributed stochastic neighbor embedding (t-SNE) [151], and Uniform Manifold Approximation and Projection (UMAP) [152] (Table 3.2). This reduces the noise, helps find descriptive components, and enables better visualization and interpretation of the data. It furthermore makes the downstream analysis more computationally efficient, given that most analysis would be difficult or even impossible in the high-dimensional space [153].

PCA uses normalized and scaled data to identify linear combinations of the original features that explain most variance in the data [154, 153]. This is

PCA	t-SNE	UMAP
Linear	Non-linear	Non-linear
Matrix factorization	Uses pairwise similarities of data points (cells)	Uses pairwise similarities of data points (cells)
Preserves global structure	Preserves local structure	Preserves local and global structure
Highly scalable	Computationally expensive	Scalable
Interpretable components	Not directly interpretable	Not directly interpretable

TABLE 3.2: Comparison of dimensionality-reduction methods.

achieved by performing an eigen-decomposition using the covariance matrix to calculate the eigenvectors and eigenvalues. Using the eigenvectors with the largest eigenvalues as principal components, the data is projected into a lower dimensional space.

In contrast, t-SNE uses a non-linear projection [154]. It computes pairwise similarities between the cells in the high-dimensional space by converting distances into probabilities of neighborhood memberships. A similar probability distribution is then created in the lower-dimensional space. The algorithm minimizes the difference (Kullback–Leibler divergence) between the two distributions to produce a layout that conserves the local structure.

UMAP, also a non-linear projection, preserves the data's local and global structure and is less computationally intensive than t-SNE [155], even though the approach is very similar. UMAP constructs a neighborhood graph for the cells in the high-dimensional space. It then optimizes the low-dimensional embedding by minimizing the difference between the original and the low-dimensional graph by trying to recreate the neighborhood structure.

A commonly used approach for visualizing single-cell data is UMAP or t-SNE on PCA initialization. That means using PCA to project the full count matrix into a lower-dimensional space while preserving the majority of variance in the data (Figure 3.12) and then using t-SNE or UMAP to project this matrix into the 2-dimensional space for visualization. [156]

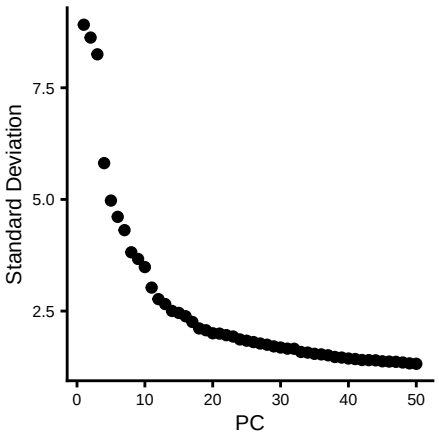


FIGURE 3.12: Elbow Plot used to determine the number of principal components.

Although these methods project similar cells close together in an embedding, they do not cluster them. Cell clusters, usually visualized in a 2-dimensional embedding, are determined using separate clustering algorithms and, in most cases, using the higher-dimensional PCA embedding [154]. This is the case for building the neighborhood graph, clustering, and other downstream analysis [154].

3.2.5 *Clustering algorithms*

Clustering cells, genes, or samples reduces the complexity of the data and structures it to facilitate interpretation. Examples are k-means-based, hierarchical, graph-based, density based, or deep-learning-based clustering methods (Table 3.3) [157].

Method	Examples
K-means	`RaceID` [158], `SAIC` [159], `pcaReduce` [160], `SC3` [161]
Hierarchical	`CIDR` [162], `BackSPIN` [163], `pcaReduce` [160], `SINCERA` [164]
Graph-based	`Louvain` [165], `Leiden` [166]
Density-based	`Monocle2` [167], `GiniClust` [168], `DBSCAN` [169]
Deep learning–based	`DESC` [170], `scziDesk` [171], `scVAE` [172]

TABLE 3.3: Clustering algorithms and example methods.

Hierarchical clustering

Hierarchical clustering is frequently used to group values for visualization. It can uncover the hierarchical relations of the clustered features and does not require the definition of the number of clusters. As hierarchical clustering has a high time complexity (agglomerative: $O(N^3)$, divisive: $O(2^N)$), it is usually only used for small input matrices.

Both methods for hierarchical clustering use a similarity/distance function to determine the similarity between the different data points.

Agglomerative clustering (Bottom-up) is the most commonly used type. It starts with each data point as its cluster. The two most similar clusters are merged after determining the pairwise distances between each cluster. It repeats this step until all points form a single cluster. Divisive clustering (Top-down), on the other hand, starts with all points in one cluster and splits

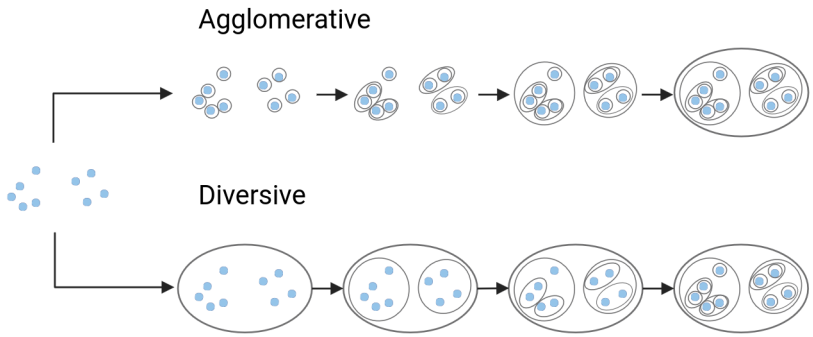


FIGURE 3.13: Agglomerative and divisive hierarchical Clustering. Created with BioRender.com.

the clusters by maximizing the distance between the two groups. It repeats this step until all points are in their own cluster (Figure 3.13). Both clustering algorithms result in a dendrogram, thus allowing the clustering at different resolutions.

Graph-based clustering

When clustering cells in single-cell data, the most commonly used methods are graph-based clustering methods [157]. They are usually based on a neighborhood graph built in the lower-dimensional PCA embedding, thus capturing the cell-cell similarities.

This enables them to identify sub-populations in the data as they capture the graph's local structure, making them more robust to noise. They do not make any assumptions about the shape of the clusters and are very efficient and scalable, an important feature when analyzing high-dimensional single-cell data.

Like the hierarchical clustering, it does not estimate the number of clusters in the data. Thus, it relies on a user-defined resolution for the clustering.

The two most commonly used methods are Louvain and Leiden clustering. Both algorithms aim to optimize quality functions such as modularity, or the Constant Potts Model [166]. Modularity is the difference between the observed and the expected number of edges in a cluster [173].

LOUVAIN CLUSTERING Louvain clustering [165] usually starts with each node being in its own community. It then merges nodes into the clusters that cause the highest increase in the quality function. For modularity, that means it moves the node into the community, showing the strongest increase in connections inside the community. In the aggregation phase, each community is aggregated into a single node. The algorithm repeats these two steps until the quality function does not improve anymore [166].

One significant disadvantage of Louvain clustering is that it can result in clusters of cells that are not fully connected.

LEIDEN CLUSTERING Leiden clustering [166] is based on the Louvain clustering but guarantees well-connected communities that are more stable across resolutions. It also implements the local moving and aggregation phase but adds refinement of the partition before the aggregation. The algorithm creates an initial partition in the local moving phase. Using this partition, the algorithm takes each community and refines it. Each node starts as its own community again, and the nodes are recursively merged with other communities (if both are well connected within the original partition and are in the same community). After the refinement, the algorithm continues with the aggregation. The nodes are aggregated based on the refined partition, using the original partition as a community starting point. It repeats the steps until it does not improve anymore.

One of the main goals of single-cell analysis is the Identification and Characterization of cell types within samples. This method's single-cell resolution allows us to discover new cell types and cell states. This usually involves a clustering step to group similar cells based on their gene expression profiles and an annotation step that assigns the clusters' biological meaning, annotating them as specific cell types.

3.2.6 *Cell type annotation in single-cell data*

Annotation of cell clusters can be done manually or automatically. Characterizing cell types within a tissue or sample is crucial for downstream analysis, such as Differential gene expression, Trajectory, and cell-cell communication analysis.

Cell types are groups of cells with similar structure and function, which they do not share with cells from other cell types [174, 175]. Before the rise of single-cell sequencing, cells were characterized using morphological features, developmental origins, and functional profiles [176]. The introduction of antibody labeling enabled the identification of cell types via surface markers. RNA sequencing then enabled the detection and classification based on gene expression profiles and allowed the refinement of cell type definition by revealing new cell types and states.

Cell type annotation is an important step in identifying cell types, as it reduces the complexity of datasets and allows us to understand the organization of tissues and organisms. Cell types that are defined using the transcriptome often follow a hierarchical organization [175]. The main challenge in single-cell data is distinguishing between cell types and states that may reflect variations such as circadian rhythm, metabolic or developmental changes, aging, or disease. That means that not all transcriptomic clusters represent stable cell types [175] that distinguishing those states from cell types is an ongoing challenge.

EXPERT MANUAL ANNOTATION One way to assign cell types is by performing a manual annotation. For this, the dataset is clustered, and cell-specific marker genes are identified. Cell type labels are then assigned based on known marker genes (Figure 3.14 for PBMC cell type markers). Manual annotation of datasets allows for identifying new cell types and cell populations and is common in large-scale projects. Usually, this requires biological knowledge about the annotated tissue or sample and is time- and labor-intensive. It furthermore results in non-standardized annotations and is often not reproducible across studies [185].

AUTOMATIC ANNOTATION The development of automatic annotation methods improved standardization and reproducibility across studies. Most methods are either Marker-based Annotation and Reference-based Annotation methods [154].

Marker-based annotation methods use known marker genes to assign cell types. They use databases such as PanglaoDB [186] and CellMarker [187] as

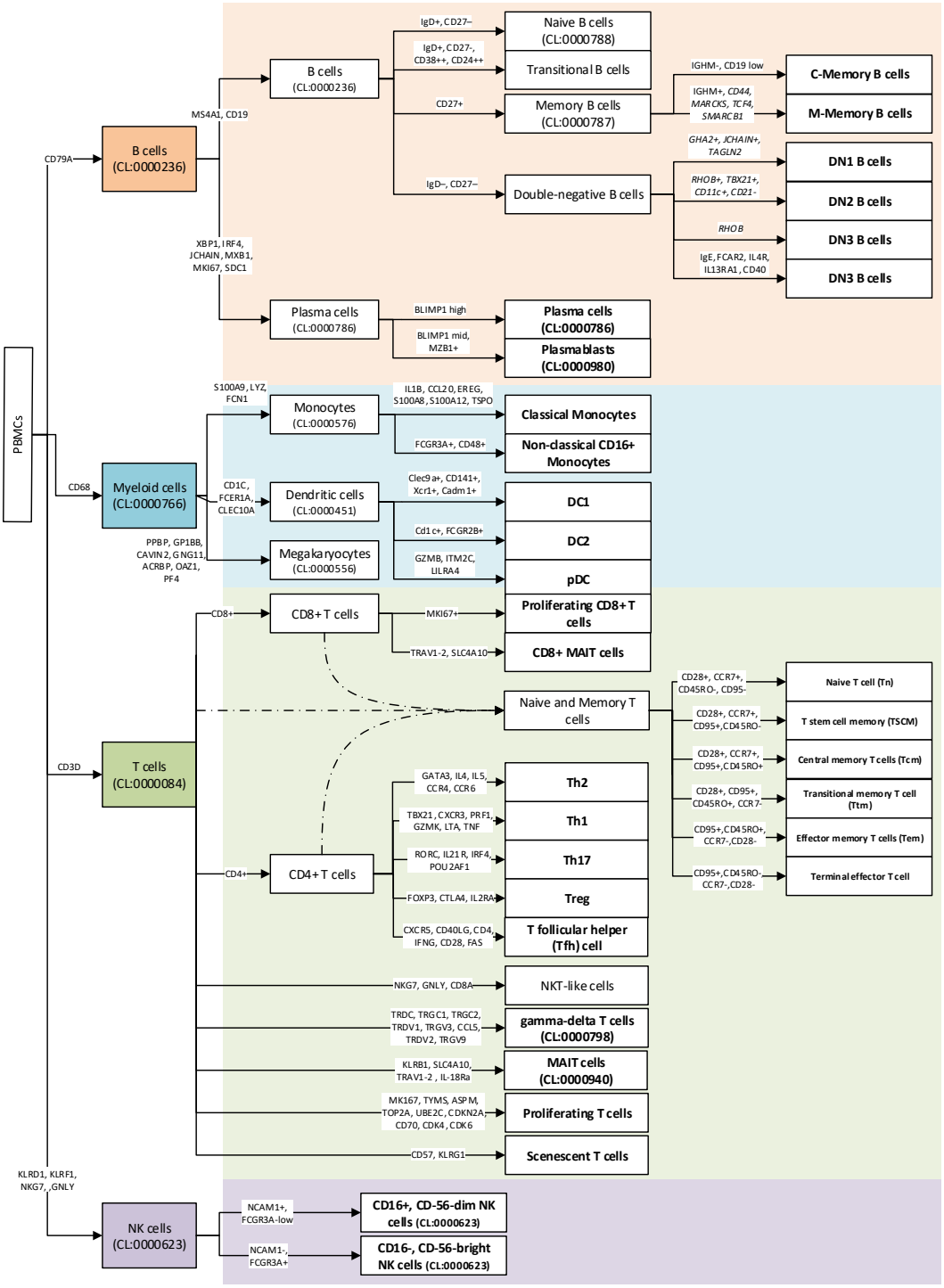


FIGURE 3.14: Cell Type Markers for PBMC single-cell Data, based on [177, 178, 179, 180, 181, 182, 183, 184].

resources for cell type marker genes. Examples of marker-based annotation methods are `AUCell` [188] and `SCINA` [189].

Reference-based annotation uses an annotated reference dataset and transfers the cell type labels to the query data based on similarities in the transcription profile of cells. As references, it usually uses data from the Gene Expression Omnibus (GEO) [190] or cell atlas datasets such as the Single Cell Expression Atlas [191], the Human Cell Atlas [192] or the Human Immune Health Atlas [12]. Examples of reference-based annotation methods are `singleR` [193], `scrap-cell` [194], and Integration methods such as Seurat's CCA [195] and `Harmony` [136].

In addition, modern machine learning and neural network–based methods (e.g., `scNym` [196], `ACTINN` [197], `CellTypist` [198]) are trained on millions of cells from diverse tissues and offer fast annotations in multiple resolutions.

Automatic annotation has the advantage of being fast and scalable, which makes it helpful in annotating large datasets and well-known datasets. On the other hand, it is not applicable for annotating rare and novel cell types and might provide inaccurate labeling if markers are missing [154]. Thus, it can be helpful to identify major cell types and guide further exploration.

SEMI-AUTOMATIC ANNOTATION Combining automatic and manual annotation can speed up the analysis while ensuring that rare cell types are accurately labeled. Semi-automatic annotation combines automated predictions with manual validation by a tissue expert [154]. Automatic annotation usually creates the initial labeling, which an expert manually refined and corrected. This approach balances the annotation's speed, accuracy, and interpretability, thus making it suitable for high-resolution studies with many cells.

3.3 PRIMARY ANALYSIS OF SINGLE-CELL AND BULK DATASETS

Both single-cell mRNA sequencing and bulk small non-coding RNA sequencing allow us to perform general analysis that aims to detect a change in the expression profile of our samples (Table 3.4). Single-cell data furthermore allows us insights into the cell type-specific changes, cell-cell communications and cell dynamics, and the changes in the cell type abundances. Providing a

Analysis	scRNA-seq	sncRNA-seq
PVCA	$\checkmark_{pb}$	✓
Differential abundance	✓	
Differential expression & Pathways	✓	✓
Cell-cell communication	✓	
Trajectory/RNA velocity	✓	

TABLE 3.4: Overview about the primary analysis on scRNA-seq and sncRNA-seq data. $\checkmark_{pb}$ means that the analysis is performed on a pseudo-bulk level.

much more detailed view of the biological processes in a cell, this resolution also has its downsides. The high computational requirements and the introduction of false positive hits led to the development of pseudo-bulk methods. These methods still use the single-cell resolution in the sense that they incorporate the cell type annotation into the bulk samples, but also allow much quicker processing of the data and reduce the false positive rate [199]. Pseudo-bulk means that the expression counts from a group of cells (e.g., cell type) from the same individual or sample are aggregated, for example, by summing up the counts or taking the median.

The following section briefly introduces the state-of-the-art analysis methods used to analyze RNA sequencing data.

3.3.1 *Principal Variance Component Analysis (PVCA)*

In experiments with multiple groups, conditions, and factors, or even in experiments with information about batches, donors, and other technical variables, it can be useful to determine how much each factor contributes to the overall variance in the dataset. PVCA is a method that combines PCA with Variance Component Analysis (VCA) to determine the contribution of different factors to the total variability [200].

The method consists of three steps: the PCA, VCA, and the aggregation of the results. The PCA uses the gene x sample (or pseudobulb gene x cell type/sample) expression matrix to determine the principal components that explain most of the total variance (usually 80 - 90 %). For each

principal component, the model performs a Variance component analysis. For this, a linear mixed model is used to model the variance of the component by the given factors (e.g., batch, donor, or treatment) and determine the proportion of variance that each factor explains for this principal component. In the last step, these numbers are summed up, using the variance explained by the principal component as a weight for the variance explained by the different factors in each principal component. This information can help us determine whether additional quality control steps, such as batch correction, are necessary and make it easy to see which experimental factor significantly influences the data's variation.

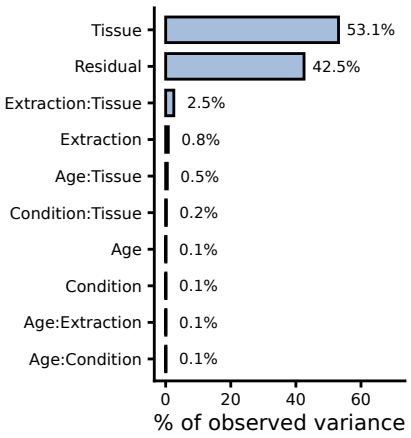


FIGURE 3.15: Principal Variance Component Analysis (PVCA) from [F2].

3.3.2 *DEG analysis*

Differentially expressed genes (DEGs) show a change in expression between two conditions that are statistically significant [201].

These tools operate either on a bulk or pseudo-bulk level, such as `DESeq` [202], `edgeR` [203], `limma` [204] or the Wilcox-rank-sum test, or on a single-cell level such as `MAST` [205] or `SCDE` [139].

Single-cell data has its own challenges, such as differences in read coverage and depth, dropout events, or cell volumes. DEG methods developed for single-cell data are designed to deal with these variations. Pseudo-bulk or bulk-based methods have the advantage of a much lower run-time and produce fewer false positives than single-cell methods [199].

When performing a differential expression analysis, we usually report the fold change, an estimate of the size of the difference in the expression level, the p-value, the statistical significance of the change, and sometimes the Cohen'd, the effect size of the change.

FOLD-CHANGE The fold-change (FC) of a gene is usually defined as

$$FC = \frac{x_a}{x_b}, \tag{3.4}$$

where x_c is the Expression in condition c. As this results in a value between 0 and 1 for down-regulated and 1 and infinite for up-regulated genes, we usually report the $\log_2 FC$ to achieve a symmetrical value around 0 for up- and down-regulation. A $\log_2 FC > 0$ thus means that the expression in condition a is bigger than in condition b, and $\log_2 FC < 0$ means that the expression in condition b is bigger than in condition a.

COHENS'D Cohens'd is a measure for the effect size. According to Kelly and Preacher in 1012, the "Effect size is defined as a quantitative reflection of the magnitude of some phenomenon that is used to address a question of interest" [206]. The Cohens'd effect size is defined as

$$d = \frac{\bar{x}_a - \bar{x}_b}{S_{ab}}, \tag{3.5}$$

where x_c is the Expression in condition c and S_{ab} is the pooled standard deviation [207]. Effect sizes are usually interpreted as follows: no effect (<0.2), small (0.2–0.5), medium (0.5–0.8), and large effect (>0.8) [208].

P-VALUE To calculate the p-value, we must first define the H_0 hypothesis. The H_0 hypothesis in our case is that there is no difference in the gene expression between the two conditions. When looking at the difference between the gene expression distributions in two conditions, we calculate the probability that we, by chance, observe a difference as large or larger than what we see in our data. This probability is then called the p-value. Suppose the probability of the gene expression distribution being equal (and still observing a difference this big) is below 5% (0.05). In that case, we reject this hypothesis and assume that there is a difference in the gene expression distributions [209]. How this probability is calculated depends heavily on the method and model used and can thus be different depending on the analysis. To account for the multiple hypothesis testing (independent testing for thousands of genes),

we adjust the p-value using methods such as Benjamini-Hochberg (BH) or Bonferroni.

3.3.3 *Comparison between RNA-expression changes*

When comparing gene expression changes, for example, when studying the effect of a disease on male and female patients, we need to compare the changes in the gene expression patterns. One way to do this is by using similarity measurements between the DEG lists or correlation of the fold-changes.

Correlations

The idea behind correlating fold changes in two different conditions is the assumption that both conditions have the same cause (or at least a causal relationship) and thus should show similar changes. Correlation means that if one variable changes, the other variable changes in the same (positive correlation) or the opposite direction (negative correlation). Of note, this does not mean that the change in variable 1 causes the change in variable 2. Instead, correlation measures how one variable changes in relation to another variable. In our case, we compare the fold changes in one condition to those in the other. The two most commonly used correlation methods are Pearson and Spearman correlation. Both methods have correlation values between -1 and 1, where 0 means no linear or monotonic relationship and 1 means a straight line (Pearson) or a monotonic curve (Spearman). The Pearson correlation assumes a linear relationship between two normally distributed random variables [210]. It is more sensitive to outliers but more powerful than Spearman. The Pearson Correlation Coefficient r is defined as

$$r = \frac{\sum (fc_{g,1} - \bar{fc}_1)(fc_{g,2} - \bar{fc}_2)}{\sqrt{\sum (fc_{g,1} - \bar{fc}_1)^2 \sum (fc_{g,2} - \bar{fc}_2)^2}}, \tag{3.6}$$

where $fc_{g,c}$ is the fold-change of gene g in condition $c \in \{1, 2\}$ and $\bar{fc}_c$ is the mean fold-change in condition c.

A Spearman rank correlation assumes a monotonic relationship between two variables and is calculated based on the ranks in the list [210]. In our

case, gene lists ordered by fold-changes are used as inputs for the correlation. The Spearman coefficient ρ is defined as

$$\rho = 1 - \frac{6 \sum (rk_1(g) - rk_2(g))^2}{n(n^2 - 1)}, \tag{3.7}$$

where $rk_c(g)$ is the rank of gene g in condition $c \in \{1, 2\}$ and n is the number of genes.

Cosine similarity

Similar to Spearman's coefficient, we can also compare the similarity between two DEG lists using the cosine similarity. Usually used to compare the similarities of two vectors, this method does not assume the distribution of the data but instead gives a relative similarity of the fold-change lists.

The cosine similarity $S_C(fc_1, fc_2)$ of the fold-change lists fc_c in condition $c \in \{1, 2\}$ is defined as

$$S_C(fc_1, fc_2) = \frac{\sum_g fc_{g,1} fc_{g,2}}{\sqrt{\sum_g fc_{g,1}^2} \cdot \sqrt{\sum_g fc_{g,2}^2}}, \tag{3.8}$$

where $fc_{c,g}$ is the fold-change of gene g in condition c.

These methods allow us to access the correlation and similarity between two changes and thus give us an estimate of the gene expression profiles. So far, these methods do not estimate which biological functions are affected and how similar these are in the different conditions.

3.3.4 *Pathway analysis*

As solely relying on identifying individual genes can be very time-consuming and does not provide an overview of the affected biological pathways, the next step is to perform a pathway analysis. Pathways are actions, interactions, or functional relationships between molecules that lead to a biological process or signal [211]. This can result in the production of another molecule or a change in the cell, such as the regulation of genes or movements of the cell.

Using pathway analysis allows us to put the different genes into context. We can consider the interactions of genes and proteins and interpret the results in the biological context. This system's biological view of gene expression changes allows us to identify the genes' roles in the cell or tissue. It makes it easier to determine similarities and differences between conditions.

PATHWAY DATABASES Pathways are collected and stored in pathway databases. Examples are the Kyoto Encyclopedia of Genes and Genomes (KEGG) [212], the Gene Ontology (GO) database [213] and the Reactome Pathway database [214]. The components of the pathways are either stored as gene lists, which are unstructured and do not include information about the interactions, or as pathway topologies, which include the components of a pathway and the interactions between them.

Pathways itself are usually stored as trees, representing the relationships between pathways. This allows us to determine distances and clusters of pathways based on the overall topology. As the pathway analysis often results in sets of similar pathways (not necessarily with equal names), grouping the pathways based on topology makes interpreting the results easier. It enables the comparison of pathways across conditions.

ALGORITHMS FOR PATHWAY ANALYSIS To perform the pathway analysis, two frequently used methods are Over-representation Analysis (ORA) and Gene-Set Enrichment Analysis (GSEA) [215]. Both methods require gene-lists as input. For ORA, we use a filtered gene list to contain only the relevant genes, and we use this unordered gene list as input. GSEA requires an ordered gene list as input but allows us to work with the unfiltered gene list, usually ordered by fold-change and/or p-value.

ORA is part of the first-generation pathway analysis [216]. It tests if the proportion of genes from the input set that also occurs within a given pathway exceeds the number of genes that would randomly be expected.

We usually use a Fisher's Test to determine the p-values for an ORA. Given the size of the pathway (i.e., number of genes) K, the number of input genes (e.g., DEGs) n, the size of the overlap between the two $k = n \cap K$, and

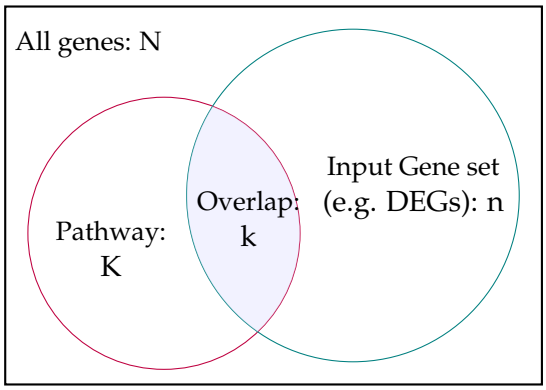


FIGURE 3.16: Gene sets in the Over-Representation Analysis (ORA).

the size of the overall gene set N (e.g., all genes in the transcriptome), the probability is defined as

$$P(X = k) = \frac{\binom{K}{k}\binom{N_K}{n-k}}{\binom{N}{n}} \tag{3.9}$$

This p-value then has to be adjusted for multiple hypothesis testing. This method has the disadvantage of requiring the user to define a threshold for the input features. Second-generation pathway analysis methods such as GSEA do not require this; instead, they use the unfiltered but sorted gene list.

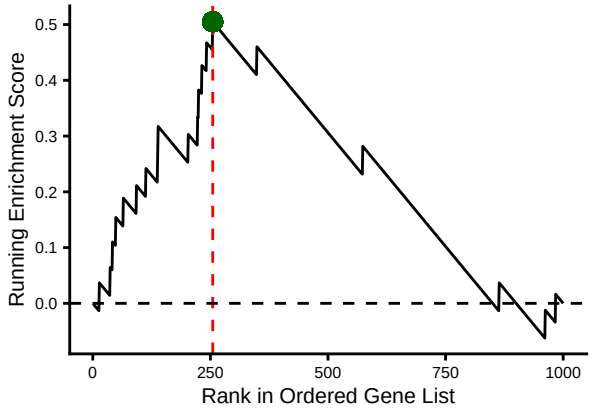


FIGURE 3.17: Running Sum of a Gene Set Enrichment Analysis (GSEA) and the Enrichment Score (green dot).

GSEA consists of three main steps. The first one is to calculate the enrichment score that represents the enrichment of the gene set (i.e., pathway) at the beginning or end of the list. It is determined using running-sum statistics. That means that while walking down the ranked list of genes, a running sum increases if the gene is in the pathway set and decreases if it is not. The enrichment score is the value with the largest deviation from zero that the running sum achieves during its walk. Using this enrichment score, a permutation test is then used to determine the statistical significance of the score. This raw p-value then needs to be adjusted for multiple hypothesis testing. [217]

3.3.5 *Cell-cell communication*

Single-cell data not only allows us to put changes in the transcriptomic profile in the context of biological pathways but also allows us to detect possible changes in cell-cell communication between cell types. Tools such as `CellChat` [218], `CellPhoneDB` [219] or `LIANA` [220] allow the automatic detection of cell-cell communication changes in single-cell data. `CellChat` uses its database (CellChatDB), which contains known ligand-receptor pairs, agonists and antagonists, co-stimulators, and inhibitors. Using these interactions, a mass action model in combination with the DEGs is used to infer the communication between cell groups. This enables us to examine how cell communication changes between two conditions.

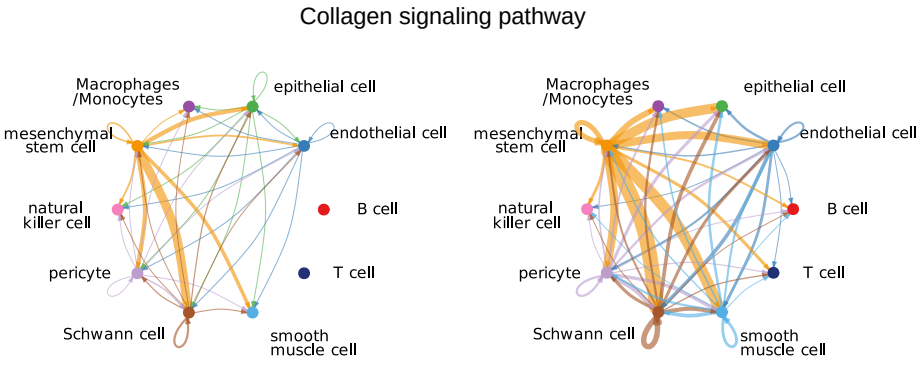


FIGURE 3.18: Cell-cell communication analysis showing the change in the Collagen signaling pathway from [F1].

3.3.6 *mRNA regulation*

Single-cell data allows us to link changes in the expression levels of transcription factors to deregulation or target mRNAs to detect general changes in mRNA regulation. Databases such as TFLink [221] and Jaspar [222] provide information about the transcription factor targets. Even though the mRNA levels of transcription factors do not provide direct information about the protein level, they allow us to detect key mRNA regulators in combination with the transcription profile of cells. Similarly, combining mRNA with miRNA data provides a general insight into the miRNA-mRNA regulation. By using databases such as TargetScan [223], miRDB [224] or miRTargetLink [225], that provide the target mRNAs of each miRNA, we can link changes in the miRNA level with deregulation in the transcriptome.

3.4 DETECTING CHANGES IN CELL TYPES AND CELL STATES

Single-cell data provides information about the transcriptomic profiles of different groups of cells and gives us insights into the cell type composition of the tissue or sample and the relationships between cells. Methods such as RNA velocity or trajectory analysis allow us insights into cell development and differentiation.

3.4.1 *Differential abundance test*

For uncovering differential abundant cell types in single-cell data, which means cell types are significantly more or less frequent in the test compared to the control case, a wide variety of tools have been developed [226]. Examples are `Milo` [227], `scCODA` [228], `DA-seq` [229], `Meld` [230] or `Cna` [231]. Differential abundance testing in scRNA-seq is challenging as the cell counts are relative and not absolute. The increase in the number of cells in one cell type might be caused by the decreasing number in different cell types. Because single-cell data always represents a random sample of cells, sampling noise mainly influences small cell type clusters. `scCODA` is based on the fact that the total number of samples is random and depends on the sequencing depth

and the capture efficiency. It uses a Bayesian multinomial logistic regression model to model how each cell type changes based on a reference cell type [228].

3.4.2 *RNA velocity and trajectory analysis*

Using differential abundance tests has the disadvantage that it can only identify changes in the cell type composition and does not capture changes in cell states or gene expression dynamics within the cell type clusters. Furthermore, it provides no information about ongoing or future changes in cell types or states.

Two groups of methods have been developed to capture the dynamics of cell types and states. RNA velocity-based methods estimate the future transcription profile of individual cells using information about RNA splicing. Trajectory analysis assigns each cell a pseudo time, a relative ordering along the developmental process within a cell population.

Trajectory Analysis Methods such as Monocle [232] use the gene expression profile of each cell to model transitions between cell states based on their similarity. It creates a neighborhood graph, capturing these similarities and ordering the cells along the longest path in this graph. It then assigns each cell a pseudotime value that describes the position of each cell along this graph, basically describing its position along the developmental process.

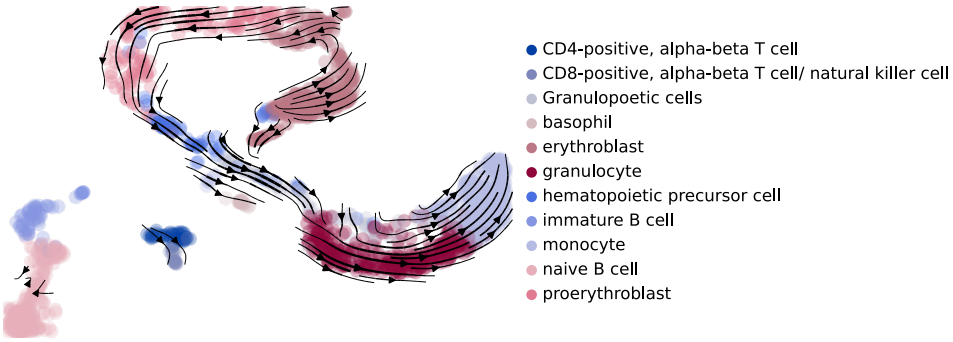


FIGURE 3.19: Visualization of an RNA velocity analysis (from [F1]).

RNA velocity uses RNA processing to predict the future expression profile of a cell [233]. RNA undergoes a few processing steps during its lifetime. It is transcribed from the DNA, spliced, and then degraded. In RNA velocity, each of these reactions is modeled using reaction kinetics. Although most analyses only use spliced mRNA, the sequenced data also contains information about the unspliced mRNA molecules. Using the amount of unspliced and spliced RNA, as well as transcription and degradation rates, the model uncovers dynamics in the gene expression of single genes and the gene expression profiles of cells, identifying steady and dynamic cell states [233, 234].

3.5 WORKFLOW MANAGEMENT AND REPRODUCIBILITY

These diverse processing and analysis methods require the coordinated use of numerous tools and packages. Managing these in a reproducible and scalable manner is critical, especially when working across many tissues, conditions, and datasets. To address this, we applied dedicated analysis frameworks and workflow management systems.

3.5.1 *Analysis frameworks*

Frameworks such as Seurat [149] and Scanpy [235] can help create reproducible pipelines as they include a large variety of standard functionality, helpful in analyzing single-cell data. This includes a data structure that enables easy storing and access to the data but also functions for filtering, normalization, dimensionality reduction, DEG and pathway analysis, and a variety of visualizations. Even though they include a large variety of functionality and extensibility to tools such as Monocle, many tools used for pre-processing and analysis of single-cell data still have to be used independently.

3.5.2 *Workflow management*

To ensure reproducibility, scalability, and modularity of the pre-processing and analysis workflow, we used snakemake [236] as a pipeline management tool. This rule-based system enables the definition of reproducible pipelines, allows parallelization of jobs, and easy integration of multiple programming languages (e.g., R and Python). Integrating conda [237] environments allowed software versions to be controlled.

This minimizes the need for manual intervention and allows for a reproducible execution of the same pipeline across tissues, experiments, and computational environments.

4

GOAL OF THIS THESIS

This thesis aimed to apply and develop scalable and reproducible bioinformatics pipelines to study how the immune system functions as a systemic, tissue-integrated network that adapts to physiological stress states and diseases. These methods allow us to study scRNA-seq and sncRNA-seq data in a multi-omics analysis and investigate miRNA-mediated regulation, immune remodeling, and inter-organ coordination. The identification of transcriptome patterns in immune cells across tissues and conditions allows us to identify markers and evaluate the potential of peripheral blood immune profiles as biomarkers in diseases that affect inaccessible tissues such as the brain.

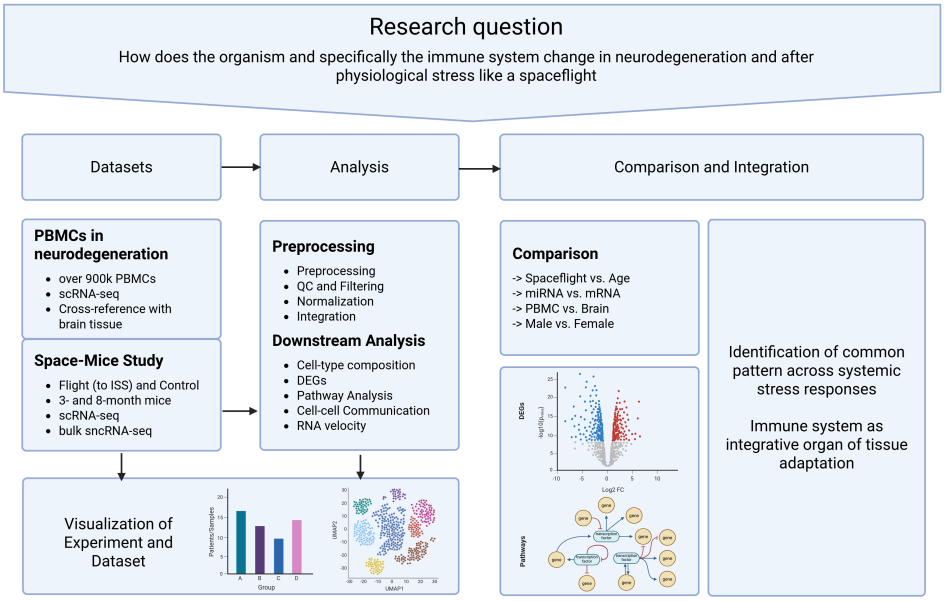


FIGURE 4.1: Goal and workflow of this thesis. Created with BioRender.com.

Highly complex data from single-cell RNA sequencing require applying bioinformatics methods and integrating these results to compare diseases,

conditions, tissues, or cell types. This thesis describes three studies that investigate the changes after physiological stress and during neurodegeneration to link cell type-specific to systemic changes.

The first study, which is described in chapter 5, has already been published under the title "A single-cell atlas to map sex-specific gene-expression changes in blood upon neurodegeneration" (Grandke et al., 2025) [F3] and includes single-cell RNA sequencing data from 136 patients with neurodegenerative diseases and 121 controls. There, we study how the transcriptome of PBMCs reflects the disease-specific immune regulation in Alzheimer's, Parkinson's, and Mild Cognitive Impairment in a sex-specific manner.

The other two studies are described in chapter 6 and aim to uncover how systemic stress, such as spaceflight, affects the immune system and tissue remodeling across organs. 40 female mice were send to the International Space Station (ISS) and compared against two ground control groups. We used single-cell RNA sequencing data to identify cell type-specific changes as a response to spaceflight in teo age-groups (3-month old and 8-month old) [F1]. We then used the pseudo-bulk of these data in combination with miRNA data from bulk sncRNA-seq to uncover miRNA-mRNA regulation patterns in a tissue- and age-specific context [F2].

PUBLICATIONS AND PRE-PRINTS During the course of my PhD, I was involved in the preparation of six manuscripts. The following list contains both the first-author publications and pre-prints, as well as the co-author publications till this date[1]. The thesis will focus on the first-author projects [F1, F2, F3].

[F3] **Friederike Grandke**, Tobias Fehlmann, Fabian Kern, David M Gate, Tobias William Wolff, Olivia Leventhal, Divya Channappa, Pascal Hirsch, Edward N. Wilson, Eckart Meese, et al. A single-cell atlas to map sex-specific gene-expression changes in blood upon neurodegeneration. Nature communications, 16(1):1965, 2025.

[F1] **Friederike Grandke**, Nicholas Schaum, Fabian Kern, Jérémy Amand, Thomas H. Ambrosi, Teni Anbarchian, Karl Annusver, Reza Ashrafi,

[1] For pre-prints, the complete author lists are given. For already published papers, please see the publication.

Kruti Calcuttawala, Shubhangi Das Barman, Antoine de Morree, Angela Detweiler, Hala S. Dhowre, Annika Engel, Tobias Fehlmann, Virginia Ferguson, Matthias Flotho, Astrid Gillich, Michael Haney, Yan Hang, Pascal Hirsch, Taichi Isobe, Joanna Kalucka, Maria Kasper, Seung Kim, Thomas A. Kluiver, Davis Lee, Song Lee, Olivia Leventhal, Ling Liu, Catriona Y. Logan, Maria F. Lugo-Fagundo, Luca Mannino, Maxim Markovic, Maurizio Morri, Shravani Mukherjee, Norma F. Neff, Patricia K. Nguyen, Róbert Pálovics, Weng Chuan Peng, Maksim V. Plikus, Ermelinda Porpiglia, Thomas A. Rando, Jayesh Salvi, Nil Sanz, Ina Maria Schiessl, Stephanie A. Schubert, Arun Sharma, Shaheen S. Sikandar, Rahul Sinha, Daniel Staehli, Louis Stodieck, Aditi Swarup, Michelle Tan, Aris Tay, Krissie Tellez, Stefan Veizades, Bruce Wang, Tobias W. Wolff, Mike Wosczyna, Albert Y. Wu, Rose Yan, Andrew C. Yang, Macy E. Zardeneta, Weichen Zhao, Tony Wyss-Coray, Stephen R. Quake, and Andreas Keller. "Spaceflight induces systemic effects on extracellular matrix and immune system in different age-stages." pre-print.

[F2] **Friederike Grandke***, Shusruto Rishik*, Viktoria Wagner, Annika Engel, Nicole Ludwig, Kruti Calcuttawala, Fabian Kern, Verena Keller, Marcin Krawczyk, Louis Stodieck, Virginia Ferguson, Amanda Roberts, Eckart Meese, Nicholas Schaum, Steven Quake, Tony Wyss-Coray, and Andreas Keller. Mirnas shape mouse age-independent tissue adaptation to spaceflight via ecm and developmental pathways. pre-print. *shared co-first authors

[F4] Matthias Flotho, Jérémy Amand, Pascal Hirsch, **Friederike Grandke**, Tony Wyss-Coray, Andreas Keller, and Fabian Kern. Zebra: a hierarchically integrated gene expression atlas of the murine and human brain at single-cell resolution. Nucleic Acids Research, 52(D1):D1089–D1096, 2024.

[F5] Annika Engel, Nicole Ludwig, **Friederike Grandke**, Viktoria Wagner, Fabian Kern, Tobias Fehlmann, Georges P Schmartz, Ernesto Aparicio-Puerta, Dominic Henn, Barbara Walch-Rückheim, et al. Skin treatment with non-thermal plasma modulates the immune system through mir-223-3p and its target genes. RNA biology, 21(1):31–44, 2024.

[F6] Shusruto Rishik, Pascal Hirsch, **Friederike Grandke**, Tobias Fehlmann, and Andreas Keller. mirnatissueatlas 2025: an update to the uniformly processed and annotated human and mouse non-coding rna tissue atlas. Nucleic Acids Research, 53(D1):D129–D137, 2025.

AUTHOR CONTRIBUTIONS The author contributions, as published or (for pre-prints) agreed and used for the submission, for [F1, F2, F3] were the following:

[F3] "**F.G.**, T.F., F.K., M.F., and T.W.W. performed computational analysis. **F.G.**, T.F., and F.K. analyzed the data and assembled the figures. J.S.G. performed the literature research. **F.G.**, T.F., F.K., and A.K. wrote the manuscript with input from all authors. T.F. and Q.S. performed primary data processing. E.Mass and D.M.G. assisted with celltype annotation. D.M.G., O.L., D.C. provided and organized samples. P.H. developed the webserver. E.N.W. performed the CSF biomarker measurement experiments and K.I.A. supervised this part of the study. E.Meese, D.M.G. and K.P. provided field-expertise and guided clinical interpretation. C.L., Y.L., C.C., Y.Y., J.X., M.J., Z.W., T.W. performed the sequencing experiments. L.L. and Y.H. supervised the sequencing experiments. A.K. and T.W-C. designed and supervised the study."

[F1] "**F.G.**, T.F., T.W.W and M.F. performed computational analysis and/or primary data processing. **F.G.**, N.Sc, A.K. F.K. and A.E. wrote the manuscript with input from all authors. N.Sc., V.F., M,Mo., N.F.N, L.S., S.R.Q., T.W.-C. and A.K. organized the study. N.Sc., K.C., D.L., S.L., M.E.Z., O.L., M.F.L.-F., A.D., M.Ma., M.T. and R.Y. were involved in the Sequencing preparation and the Sequencing. N.Sc. (Adipose/Skin), P.K.N. and S.V. (Aorta/Thymus), T.H.A. and C.K.F.C (Bone), R.S. (Bone Marrow), J.A., A.C.Y. and D.S. (Brain), S.D.B. (Diaphragm), R.A., H.S.D., S.M., A.Sw. and A.Y.W. (Eye), A.Sh. (Heart/Blood), T.I. (Intestine), J.K. (Kidney), T.A., T.A.K., C.Y.L., W.C.P., S.A.S., A.T. and B.W. (Liver), A.G. (Lung), S.S.S. (mammary gland), L.L., T.A.R. and J.S. (Muscle), M.W. (Muscle/Adipose), A.D.M. (Muscle/Diaphragm), Y.H., S.K., K.T. and W.Z. (Pancreas), K.A., M.K., M.P. and N.Sa. (Skin),

M.H. (Spleen, Kidney), L.M., and I.M.S. (Kidney) and E.P. (Spleen/-Diaphragm) were involved in the tissue extraction, cell preparation and/or the cell type annotation."

[F2] "**F.G.**, S.R., A.E. and A.K. performed computational analysis and/or primary data processing. **F.G.**, S.R., A.E., V.W., N.S., V.K. and A.K. wrote the manuscript with input from all authors. F.K., L.S., V.F., E.M., N.S., S.Q., T.W.-C. and A.K. organized the study. A.R., N.S. and K.C. were involved in sample collection. V.W., N.L. and M.K. performed the Sequencing preparation and the Sequencing."

5

CHANGES IN THE IMMUNE SYSTEM WITH NEURODEGENERATIVE DISEASES

This chapter contains results, figures, and descriptions from the publication [F3]. See chapter 4 for the author contributions.

Neurodegenerative disorders like Alzheimer's disease (AD) and Parkinson's disease (PD) are becoming more prevalent in the global population, significantly impacting the quality of life for the affected patients [238, 239]. Risk factors include genetic predispositions, lifestyle, environmental influences, and age [103]. While the clinical symptoms, diagnosis, and progression differ between AD and PD, researchers have identified shared cellular pathways that may underlie both diseases [240]. These conditions can emerge from unique precursor stages of varying lengths and may present with overlapping symptoms, such as mild cognitive decline [241, 242].

5.1 BACKGROUND OF PBMC STUDIES IN NEURODEGENERATION

Traditionally, neurodegenerative diseases have been considered disorders primarily affecting the brain itself; however, the interplay between the adaptive immune system and brain inflammation in these conditions is increasingly gaining recognition [243, 244]. To this day, research has identified key cell types in the central nervous system (CNS) and important genetic risk factors that contribute to the underlying disease mechanisms, such as APOE and TREM2 in microglia for Alzheimer's disease (AD) and alpha-synuclein (SNCA) found in Lewy bodies within dopaminergic neurons for Parkinson's disease (PD). Additionally, the role of the human brain's vascular system in the development of neurodegenerative disorders is of increasing interest [245, 246].

The specific brain areas impacted by these diseases have been the focus of extensive, though mostly small-scale, research due to challenges in obtaining fresh-frozen human tissue samples and the associated costs. Consequently, researchers have turned their attention to model organisms, such as genetically modified mice and in vitro cell systems, as well as more accessible human peripheral systems, like blood or cerebrospinal fluid (CSF), and the interfaces connecting them, including the blood-brain barrier (BBB) [247, 248, 249]. Current evidence suggests that the BBB may experience a partial breakdown with aging and during age-related diseases, indicating an underestimated involvement of the peripheral immune system in neurodegenerative disorders. This involvement may be mediated through communication between these otherwise isolated physiological environments [250, 251, 252, 253].

These findings prompt further investigation into the role of the peripheral immune system during the progression of the disease, particularly regarding the contribution of systemic inflammation [254, 255, 256, 257]. A significant aspect that is gaining attention in the context of Alzheimer's Disease (AD) and Parkinson's Disease (PD) is the biological sex of the patients. For example, research using single-cell analysis in AD indicates that female cells are more prominent in disease-associated subpopulations and that transcriptional responses in oligodendrocytes and other cell types vary between the sexes [258].

We performed a literature review and found sex-specific findings both in Alzheimer's and Parkinson's (Table 5.1). Previous studies found sex-dependent changes in both the brain and blood of AD patients [259, 260, 261], and in the peripheral blood in monocytes in PD patients [262].

Broadly accessible and minimally invasive blood samples are a promising way to identify circulating biomarkers related to neurodegeneration, particularly when analyzing peripheral blood mononuclear cells (PBMC) in large-scale studies [291, 292, 293, 294, 295]. Clinical studies have demonstrated that blood markers can help in partially predicting or monitoring the progression of neurodegenerative diseases, although this approach does come with its own set of challenges [296, 297, 298, 299, 300]. Consequently, a logical next step is to thoroughly explore the molecular pathways associated with, or potentially driving, these disease biomarker signatures using high-resolution, high-throughput techniques like single-cell RNA sequencing.

Paper	**Year**		**Tissue** (Detailed)	**Tissue**	**Disease**	**Sex-specific?**
Xiong et al.	2021	[263]	Blood/PBMC	Blood	AD	✗
Hu et al.	1995	[264]	Peripheral blood lymphocytes (PBL)	Blood	AD & others	✗
Pirttilä et al.	1992	[265]	Peripheral blood	Blood	AD	✗
Gate et al.	2020	[248]	PBMC	Blood	AD	✗
Shad et al.	2013	[266]	Blood	Blood	AD	✗
Wang et al.	2019	[267]	Blood	Blood	AD	✗
Amin et al.	2020	[254]	PBMC	Blood	AD & LBD	✗
Qian et al.	2022	[268]	Peripheral blood	Blood	AD & MCI	✗
Li et al.	2022	[269]	Peripheral blood	Blood	PD	✗
Chen et al.	2021	[270]	PBL	Blood	PD	✗
Rocha et al.	2018	[271]	Peripheral blood	Blood	PD	✗
Garfias et al.	2022	[272]	PBL	Blood	AD & PD	✗
Kustrimovic et al.	2018	[273]	Peripheral blood	Blood	PD	✗
Sochocka et al.	2022	[259]	PBL	Blood	AD	✓
Ji et al.	2022	[260]	Peripheral blood	Blood	AD	✓
Coales et al.	2022	[261]	Microglia, monocyte, macrophages	Blood & Brain	AD	✓
Patel et al.	2020	[274]	Brain	Brain	AD	✗
Felsky et al.	2022	[275]	Brain	Brain	AD	✗
Zhao et al.	2015	[276]	Parietal lobe tissues	Brain	AD	✗
Moradifard et al.	2018	[277]	Brain	Brain	AD	✗
Carlisle et al.	2021	[262]	Peripheral blood monocytes	Blood	PD	✓
Sommer et al.	2018	[278]	Blood	Blood	PD	✗
Mogi et al.	2000	[279]	Substantia nigra	Brain	PD	✗
Arif et al.	2010	[280]			PD	✗
Lang et al.	2019	[281]	Human iPSC-Based Model of Parkinson's Disease	Model	PD-related	✗
Varghese et al.	2009	[282]	Blood	Blood	PD	✗
Holmes et al.	2016	[283]	Rat Dopaminergic Neuronal Cell Line	Model		✗
Maki et al.	2002	[284]	CSF	CSF	AD	✗
Tsugu et al.	1998	[285]	CSF	CSF	AD	✗
Nilsson et al.	2021	[286]	Hippocampus	Brain	AD	✗
Kish et al.	1998	[287]	Brain	Brain	AD	✗
Butterfield et al.	2010	[288]	Brain	Brain	AD	✗
Tsai et al.	2020	[289]	Blood	Blood	AD	✗
Dang et al.	2022	[290]	Astrocytes	Brain	AD	✗

TABLE 5.1: Literature research on sex-specificity of AD and PD studies (Lewy Body Dementia = LBD, Peripheral blood lymphocytes = PBL).

As recently demonstrated for parenchymal brain tissue, brain vasculature, or PBMCs [258, 301, 245, 246, 302, 303, 304, 305, 306, 307], this technology allows us to uncover changes in the cell type-specific transcriptomics profiles in neurodegenerative diseases. With datasets containing up to one million cells per publication, the amount of transcriptomic data produced is increasing exponentially and is expected to continue growing [308, 309, 310, 311].

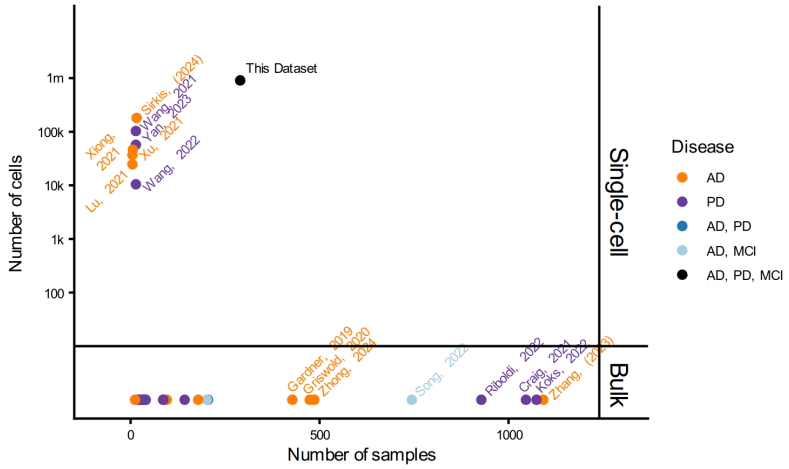


FIGURE 5.1: Previously published PBMC datasets of AD and PD.

For instance, literature research on this topic revealed that the single-cell RNA-sequencing datasets PBMCs from Alzheimer's disease (AD) and Parkinson's disease (PD) (Figure 5.1, Table 5.2) grew from around 20k cells in 2021 to over 180k in 2024.

5.2 STUDY DESIGN

In our study, we generated a single-cell atlas of peripheral immune responses in neurodegeneration by profiling over 909,000 PBMCs from 290 blood samples. The cohort included 155 individuals diagnosed with mild cognitive impairment (individual patients n = 48, total samples from the patients n_t = 55; MCI), Alzheimer's disease (n = 27, n_t = 34; AD), Parkinson's disease (n = 46, nt = 48; PD), PD with MCI (n = 15, n_t = 18; PD-MCI), and healthy individuals (n = 121, n_t = 135; HC) (Figure 5.2, Table A.1).

Paper	Year		Single-cell/bulk	#Samples	#Cells	Disease
Sirkis et al.	(2024)	[312]	Single-cell	16	182000	AD
Wang et al.	2022	[313]	Single-cell	14	10466	PD
Wang et al.	2021	[314]	Single-cell	14	103365	PD
Yan et al.	2023	[315]	Single-cell	14	57126	PD
Xiong et al.	2021	[263]	Single-cell	6	46244	AD
Xu et al.	2021	[316]	Single-cell	5	36849	AD
Lu et al.	2021	[317]	Single-cell	5	24679	AD
Zhang et al.	(2023)	[318]	Bulk	1092		AD
Koks et al.	2022	[319]	Bulk	1074		PD
Craig et al.	2021	[320]	Bulk	1046		PD
Riboldi et al.	2022	[321]	Bulk	928		PD
Song et al.	2022	[322]	Bulk	744		AD, MCI
Zhong et al.	2024	[323]	Bulk	486		AD
Griswold et al.	2020	[324]	Bulk	474		AD
Gardner et al.	2019	[325]	Bulk	428		AD
Hooshmand et al.	2022	[326]	Bulk	205		AD, PD
Song et al.	2022	[327]	Bulk	202		AD, MCI
Panitch et al.	2022	[328]	Bulk	179		AD
Hemmings et al.	2022	[329]	Bulk	143		PD
Dressman et al.	2022	[330]	Bulk	96		AD
Infante et al.	2015	[331]	Bulk	87		PD
Dhanwani et al.	2022	[332]	Bulk	86		PD
Infante et al.	2016	[333]	Bulk	40		PD
Annesley et al.	2022	[334]	Bulk	40		PD
Garofalo et al.	2020	[335]	Bulk	36		AD, PD
Carlisle et al.	2021	[262]	Bulk	34		PD
Henderson et al.	2021	[336]	Bulk	30		PD
Kurvits et al.	2021	[337]	Bulk	24		PD
Hu et al.	2020	[338]	Bulk	21		PD
Cardona et al.	2021	[339]	Bulk	20		AD
Schlachetzki et al.	2018	[293]	Bulk	20		PD
Wang et al.	2022	[340]	Bulk	12		AD

TABLE 5.2: RNA-sequencing datasets of AD and PD.

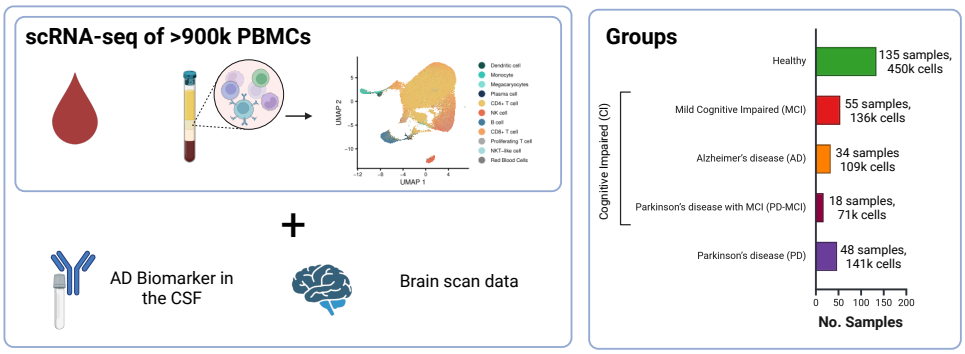


FIGURE 5.2: Overview of the PBMC neurodegeneration project from the Alzheimer's disease research center (ADRC) [F3]. Created with BioRender.com.

In addition to transcriptional profiles, brain volume information was received from MRI and PET imaging data for 158 patients, and protein-level CSF AD biomarker data was collected from 40 patients. We furthermore collected longitudinal samples for a subset of patients (n = 33), allowing assessment of immune changes over time. This dataset exceeds previously published PBMC studies in patient number and cellular resolution (Figure 5.1). It enables cell type-specific comparisons of immune responses among different diseases and patient subgroups.

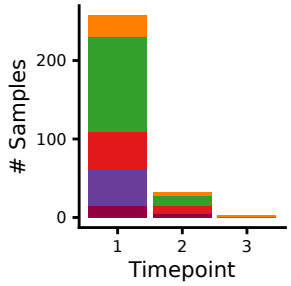


FIGURE 5.3: Number of samples per time point.

5.3 MATERIALS AND METHODS

The patients were enrolled by the Stanford Alzheimer's Disease Research Center (ADRC) and provided written informed consent[1]. Patients were selected to ensure a similar ratio of sex and age in the different groups.

SAMPLE COLLECTION, PROCESSING, AND SEQUENCING Frozen PBMCs were thawed at 37 °C, washed with pre-warmed PBS + 10% FBS (Thermo Fisher Scientific, #10010031 and #A3160601; 37 °C), centrifuged (300 x g, 10 minutes), filtered (40 µm), centrifuged again (300 x g, 10 minutes) and purified via magnetic bead-based viability selection (Miltenyi Biotech, #130-090-101). Cells were resuspended at 1000 viable cells/µL and processed using the DNBelab C Series Single-Cell Library Prep Kit (MGI, #1000021082) as described in [341]. Sequencing was performed on a DIPSEQ T1 (paired-end) at CNGB.

ALIGNMENT AND QUANTIFICATION Reads were demultiplexed and filtered using `PISA` (version 0.2). STAR (version 2.7.4a) [342] was used for alignment to the hg38 reference, the alignments were sorted with `Sambamba` (version 0.7.0) [343], genes annotated via GENCODE [344] and count matrices were generated using `PISA` [345].

QUALITY CONTROL AND DATA PROCESSING The ambient RNA contamination was corrected using `SoupX` (version 1.4.8); samples with >10% contamination were excluded. The data was processed using `Seurat` (version 4.0.0) [149]. Filtering excluded cells with >7% mitochondrial reads, cells with <300 or >4000 genes, and samples with <300 cells. Doublets were removed with `scDblFinder` (version 1.2), assuming 1% per 1000 cells. Data was normalized and scaled using Seurat. The top 2000 variable genes were selected. PCA was performed (10 PCs; elbow plot), followed by UMAP for 2D embedding. Clustering used Seurat's `FindClusters` (resolution = 0.1). Initial clusters were annotated manually (B cells, T/NK cells, myeloid cells, plasma cells). Subclustering was done per major lineage. T/NK cells were first filtered and clustered per batch, then merged. Subclusters with >50%

[1] Study approval by: Stanford University IRB

of cells from a single sample were iteratively filtered and re-clustered. Final annotations were based on known marker genes.

ANALYSIS We performed 2D kernel density estimation per sample and cell type using `MASS::kde2d` (MASS, version 7.3-55). Results were averaged per patient group. Changes in cell composition were tested using an unpaired Student's t-test (BH correction) and in a second analysis using `scCODA` (version 0.1.9) with Age and ApoE as covariates (reference: CD4 + T cell for broad annotation and Treg CD4+ cell for fine annotatio). We computed separate models for males and females. Significant differences were defined based on a reported p-value < 0.05. Differential expression was performed using `muscat, BH correction` (version 1.6.0) with pseudobulk aggregation by sample and cell type (summing SoupX-corrected counts). `pbDS` with `limma-voom` were used for testing. Models included diagnosis, sex, and batch. Genes that were expressed in at least 5% of all cells in at least one group were kept. Genes with adj. $p < 0.05$ and $|\log_2 \text{FC}| > 0.5$ were considered significant. Data processing for the ZEBRA Brain Atlas followed Flotho et al. [346]. We removed cells from overlapping studies (i.e., 78 and 79). DEGs were computed with a pseudo-bulk approach using `edgeR` (version 3.36.0), including study ID as a latent variable if needed. Significance used `glmQLFit`; p-values adjusted with BH correction (R stats package, v 4.1.3). Data from the Religious Order and Aging Project (ROSMAP, PMID: 39351424) was collected from Synapse (syn52293433, syn18681734, syn51062116, and syn2580853) using `synapseclient` (version 4.5.1). Alignments used `CellRanger` (version 8.0.1, 2T2 genome). Duplicates and donor overlaps were removed. QC steps included that the genes must be detected in $\geqslant$3 cells, and cells must express at least 100 genes. Doublets were removed using `scrublet` (version 0.2.2), and ambient RNA correction was done using `CellBender` (version 0.3.2). Cell types were labeled using Allen Brain Institute's `MapMyCells` (RRID: SCR_024672). Only types with $\geqslant$100 cells were retained. DEG computation followed the ZEBRA pipeline. Pathway analysis for all datasets using GSEA was performed using `GeneTrail` [347] on genes ranked by $-\log_{10}(\text{adj.p})$ and $|\log_2 \text{FC}|$. Databases included GO (BP, CC, MF), KEGG, Reactome, WikiPathways, and Pfam. ORA was also performed using

the same tool and databases [347]. Pathways with adj. $p < 0.05$ were considered significant. Changes in the Cell-cell communication were inferred using `CellChat` (version 1.5.0) [218]. Pearson's correlation coefficients were computed with `cor.test`. Significance was set at BH-adjusted $p < 0.05$. The similarity of fold-change patterns was calculated as cosine similarity between unfiltered gene lists ranked by fold-change. All statistical tests were two-sided unless stated. Fixed seeds were used where relevant. P-values were BH-corrected unless noted. Unless specified, all analyses used baseline samples (first time point per patient).

LITERATURE SEARCH Literature was manually curated via PubMed using combinations of keywords: "Alzheimer's" and "Parkinson's" with "blood", "peripheral immunity", "peripheral blood", "PBMC" or "CSF marker" or "brain", with and without "sex". For the single-cell and bulk PBMC datasets, we performed a manually curated literature research with the keywords: "transcriptomics", "single-cell", "PBMC", "human", "RNAseq", "blood", "Parkinson", "Alzheimer" and "neurodegeneration".

5.4 EFFECT OF NEURODEGENERATION ON PBMCS

For the study of neurodegeneration (an age-related disease) and the sex differences herein, we ensured that the subgroups were well-balanced in regards to sex and age (Figure 5.4, A.1a). PD-MCI was excluded from most of the sex-specific analysis, as the low number of female patients does not provide sufficient statistical power.

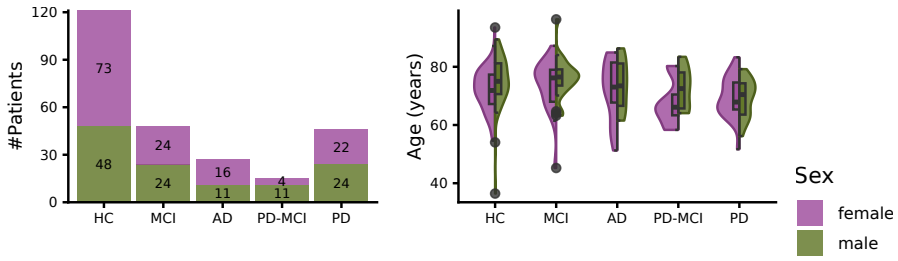
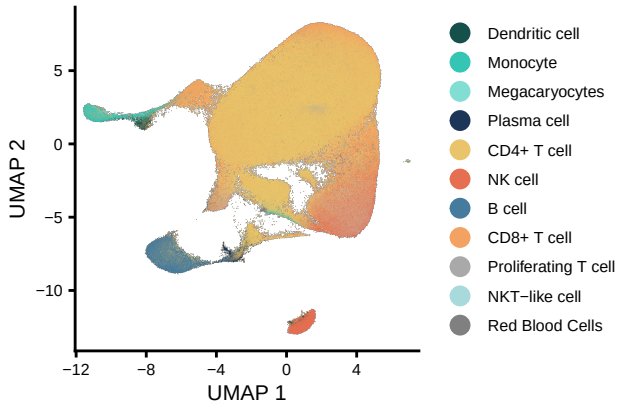
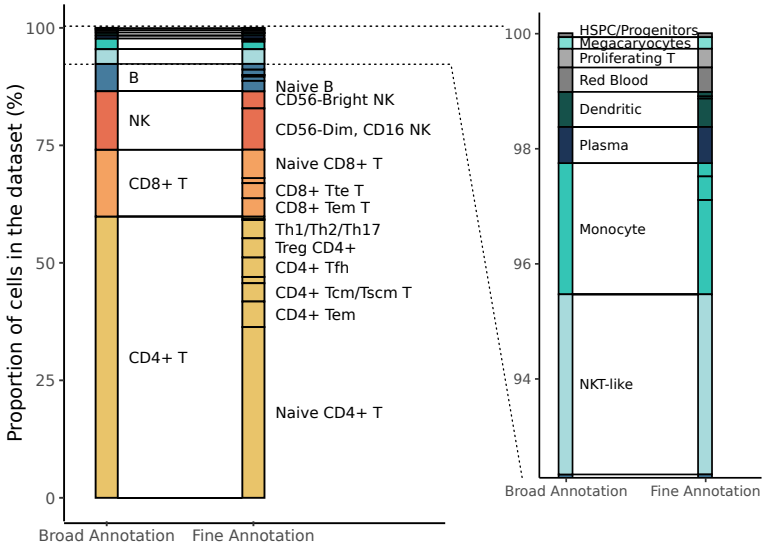


FIGURE 5.4: Distribution of age and sex in the different diagnosis groups.



(A) UMAP showing the broad cell type annotation



(B) Proportion of the broad and fine cell type annotation in the whole dataset

FIGURE 5.5: Broad and fine cell type annotation

For each sample, we conducted droplet-based single-cell RNA sequencing [341] on the collected PBMCs, resulting in 1,374,714 cellular expression profiles. 909,322 high-quality single cells (66%) successfully passed the quality control (Figure A.1b).

To identify cell type specific changes, we classified the PBMCs into 13 broad and 33 fine sub-cell types (Figure 5.5, A.2). These include CD4+ and CD8+ T cells, NK cells, Monocytes, and Magacaryocytes, as well as B cells

and Plasma cells. Changes in neurodegenerative diseases have been shown to appear both as changes in the cell type composition in PBMC cells and as changes in the transcriptomic profiles within those.

5.4.1 *Neurodegeneration causes sex-specific shifts in cell type proportions*

Visualizing large cell populations in a 2-dimensional embedding can become challenging due to a phenomenon known as overplotting. To address this, we assessed cell density to show better each diagnostic group's distribution (Figure A.3a). By interpreting densities, we identified subtle variations in cell distribution that suggest a disease-related shift in the overall abundance of different cell type populations. Given prior research indicating that cell type proportions in PBMC samples vary between males and females [348], we separated the data by sex for a more detailed examination of disease-related variations. Overall, we noted distinct shifts in cell type proportions among men and women (Figure A.3b). A density analysis further shows these effects in the integrated data space (Figure 5.6).

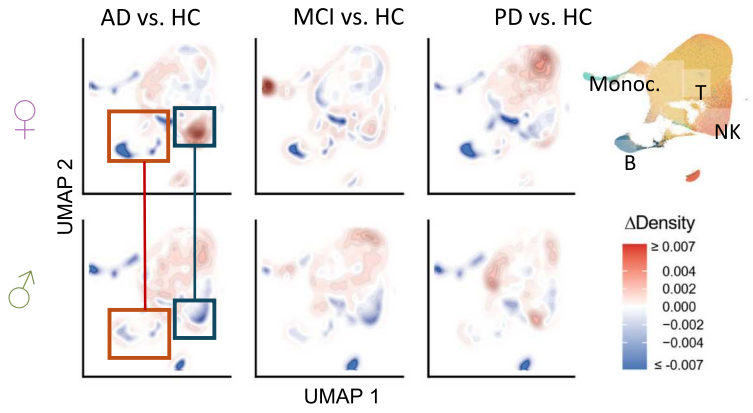


FIGURE 5.6: Differential density UMAP embedding between healthy and disease. The color indicates the highest density in disease (red) or healthy (blue).

Differential abundance testing with scCODA further supported sex-specific changes as it found no significant changes in cell type distributions

(abs. $\log_2$ FC > 0.35, significant according to scCODA) when sex was excluded as a covariate (Figure A.3c).

When separating the data by male and female patients, we observed sex-specific significant differences (Figure 5.7). CD8 + T cells and plasma cells showed a positive fold-change in males, while showing a negative fold-change in females, in Parkinson's disease. In addition, B cells were more prevalent in females with PD-MCI but less so in males with the same condition.

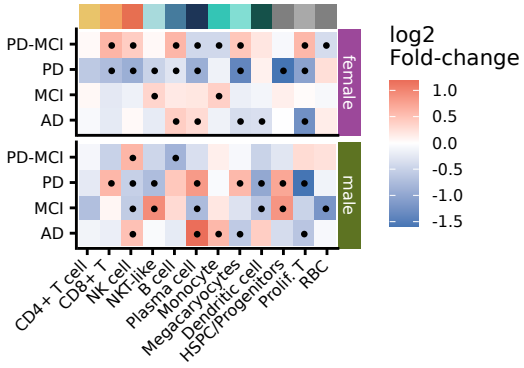


FIGURE 5.7: Differential abundance test using scCODA (covariates: age & ApoE) on the broad annotation. Significant cell types (abs. FC $\geqslant$ 0.5, significant according to scCODA) are marked with a dot.

While the density embedding suggested a sex-dependent change in B and NK cells among AD patients, scCODA did not result in significant differences. In the finer cell annotation, we found significant changes in only 3 cases for males and 9 cases for females when analyzing the raw cell type proportions (Figure 5.8).

The significant changes in AD, MCI, and PD in the whole dataset align with those observed in female patients (Figure A.3d). Employing scCODA on the finer annotation revealed a reduced percentage of T_{fh} cells in AD (Figure 5.9).

When we compared these results with earlier reported changes, we noted some that had been identified by others, including the increased proportion of monocytes in AD and variations in B cells in PD (Table 5.3) .

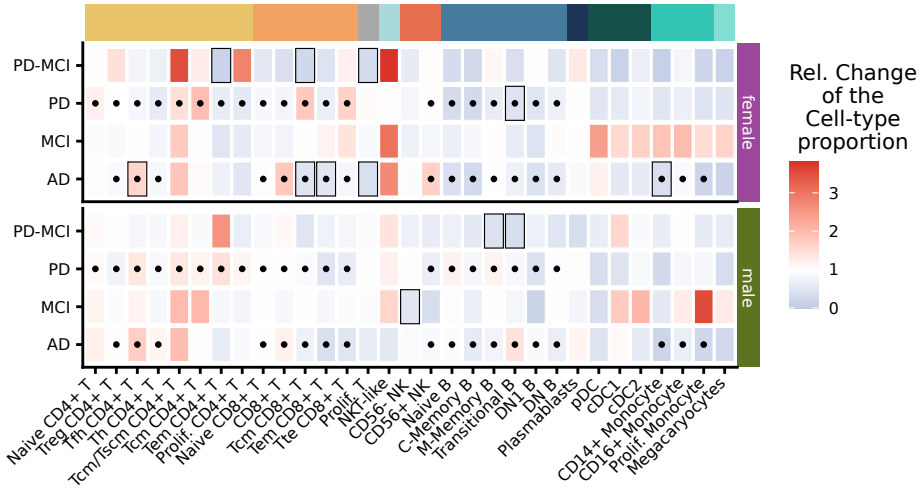


FIGURE 5.8: Relative change in the cell type-proportion (in %) between healthy and disease in the fine annotation. A frame marks significant values (adj. $p < 0.05$), and previously reported changes are marked with a dot.

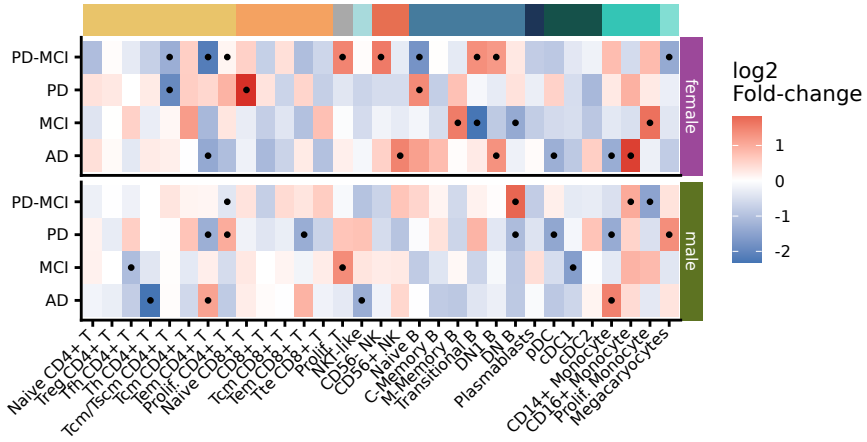


FIGURE 5.9: Result of differential abundance test using scCODA (covariates: age & ApoE) on the fine annotation. Significant cell types (abs. FC $\geqslant$, significant according to scCODA) are marked with a dot.

Our findings generally indicate a significant sex-specific change in the cell abundance of male and female patients, highlighting disease-specific variations. In addition to cell type abundances, we furthermore explore the specific gene and widespread transcriptomic disruptions in neurodegeneration.

	Literature					scCODA
AD	Xiong et al.	2021	[263]	↓	B	↑f
	Hu et al.	1995	[264]	↓	CD8+ T	ns
	Pirttilä et al.	1992	[265]	↓	CD8+ T	
	Gate et al.	2020	[248]	↑	T_{erma} CD8+	-
	Shad et al.	2013	[266]	↑	Monocytes	↑m
	Wang et al.	2019	[267]	↑	T-Helper	↓m
	Amin et al.	2020	[254]	↑	T-Helper	
	Qian et al.	2022	[268]	↓	MAIT	-
				↓	T_{reg} cells	ns
				↓	T_{fh}	↓m
PD	Li et al.	2022	[269]	↓	B	↓f↓
	Chen et al.	2021	[270]	↑	CD4+ T cell	↓f
	Rocha et al.	2018	[271]	↓	CD4+ T cell	↓f
	Garfias et al.	2019	[272]	↓	CD4+ T cell	↓f
	Garfias et al.	2019	[272]	↓	CD8+ T cell	↓f↑m
	Kustrimovic et al.	2018	[273]	↓	Th2, Th17, T_{reg}	ns

TABLE 5.3: Previously reported changes of cell type proportions and the significant results of scCODA on our PBMC data. If no sex is provided, the test was conducted without taking sex into account.

5.4.2 *Sex- and disease-specific transcriptomics changes*

We thus performed a DEG analysis for male and female patients with MCI, AD, or PD. Due to the small sample sizes of PD-MCI, we decided to exclude it from our analysis. We observed a difference in the number of DEGs between sexes, especially among AD patients, a finding that aligns with previous research by Mathys et al. [258] (Figure 5.10). We found DEGs in only 12 cell types in male patients, while female patients showed deregulation in 20. A similar, though less pronounced sex difference was found in PD vs. HC and MCI vs. HC.

We compared the fold-changes between males and females, resulting in a positive correlation for MCI vs. HC (corr = 0.35, $p < 2.2 \cdot 10^{-16}$) and for AD vs. HC (corr = 0.32, $p < 2.2 \cdot 10^{-16}$). Fold-changes especially correlated in cell types with a higher number of de-regulated genes (Figure 5.11). In contrast, PD vs. HC the cell types showed much lower correlations (corr = -0.0064, p = 0.098).

In PD vs. HC, most cell types were either negatively correlated (9 cell types) or showed no significant correlation. In contrast, in both AD and

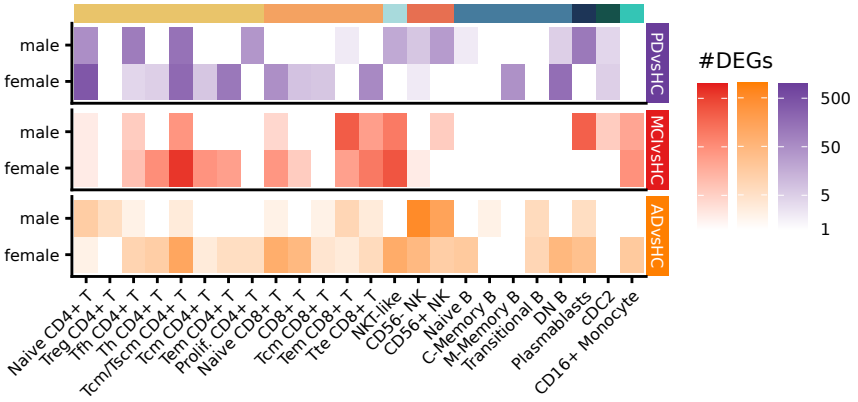


FIGURE 5.10: Number of DEGs per cell type, sex, and disease.

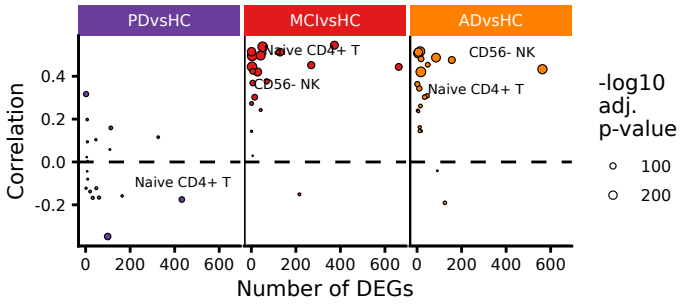


FIGURE 5.11: Comparison of the number of DEGs and the correlation of fold-changes using all genes.

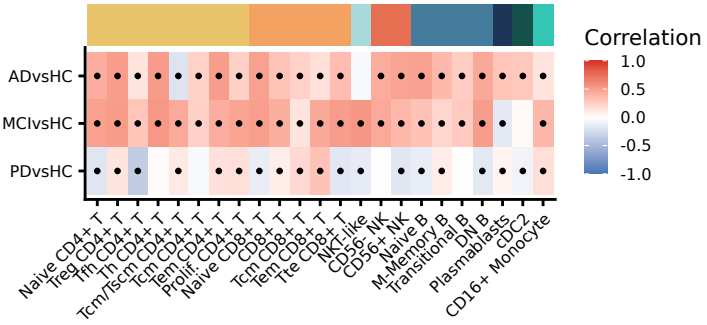


FIGURE 5.12: Correlation of fold-changes using all genes per disease and cell type.

MCI, most cell types showed a positive correlation (22 cell types each) (Figure 5.12).

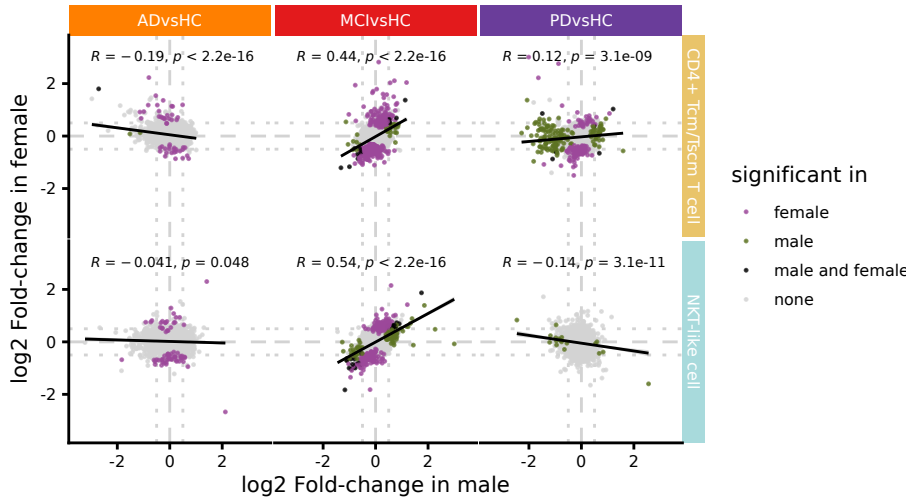


FIGURE 5.13: Comparison of fold-changes in males and females in CD4+ Tcm/Tscm T cells and NKT-like cells.

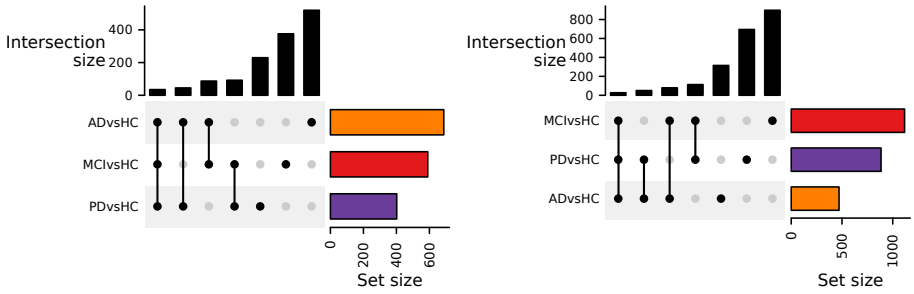
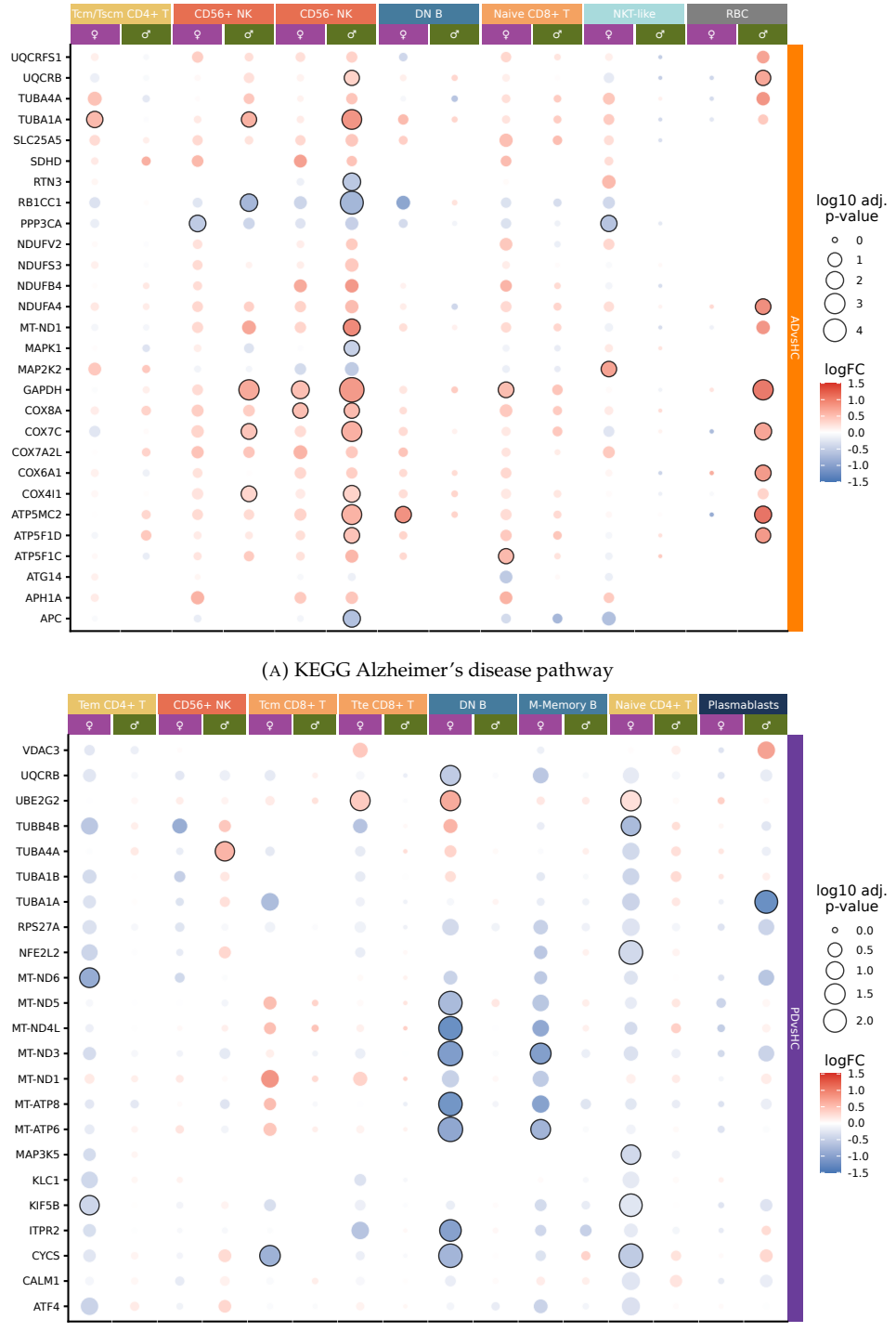


FIGURE 5.14: Number of DEGs shared between the different comparisons in males (left) and females (right).

When comparing the genes, we found 84 genes that showed the same deregulation in males and females. Most were found in CD4+ Tcm/Tscm T cells and NKT-like cells (Figure 5.13).

When comparing the deregulations between diseases, we found similar sizes of DEG overlaps in males and females, with DEGs mainly shared between MCI vs. HC and PD vs. HC (Figure 5.14). Specifically, 35 differentially expressed genes are common to all three comparisons for male patients and 27 for female patients.

These findings indicate generally similar expression patterns across different diseases regardless of sex. However, sex-specific changes can appear



(A) KEGG Alzheimer's disease pathway

(B) KEGG Parkinson's disease pathway

FIGURE 5.15: Deregulation of genes from the KEGG AD and PD pathways

depending on disease and cell type. When we examined the gene expression changes across various cell types, we noticed that the changes in subtypes of the same cell type show a high similarity (Figure A.4).

5.4.3 *Pathways affected in neurodegeneration*

In order to identify if the disease signatures reflect the dysregulation of known disease biomarkers, we analyzed gene deregulation in the KEGG Alzheimer's disease and Parkinson's disease pathways. Genes from the KEGG AD pathway appear largely sex- and cell type-independent. This could be attributed to a bias in the identification of genes in previous research, which has often not accounted for sex-specific effects. While some cell types exhibit different fold changes across sexes, the overall direction of the dysregulation remains consistent (Figure 5.15a). In contrast, genes from the KEGG PD pathway showed differences between males and females and the different cell types (Figure 5.15b).

This raises the question of which pathways specifically impact males and females. The results of GSEA pathway analysis indicated that the pathways affected in AD and MCI are similar for both sexes (Figure 5.16), while PD pathways are more sex-specific.

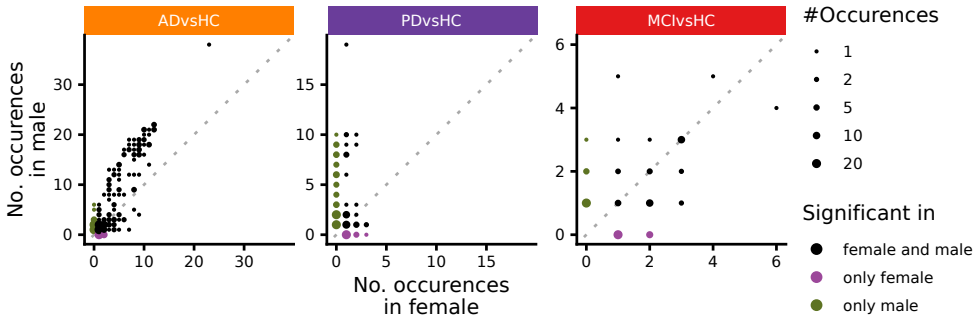


FIGURE 5.16: Frequency of pathways in males and females in the different comparisons.

The pathway "SRP-dependent cotranslational protein targeting to the membrane" is most commonly enriched in AD, in females with MCI, and

in males with PD (Figure 5.17). In PD, the "Asthma" pathway is notably impacted in females, while in MCI, males experience significant effects from the "Oxidative phosphorylation" pathway.

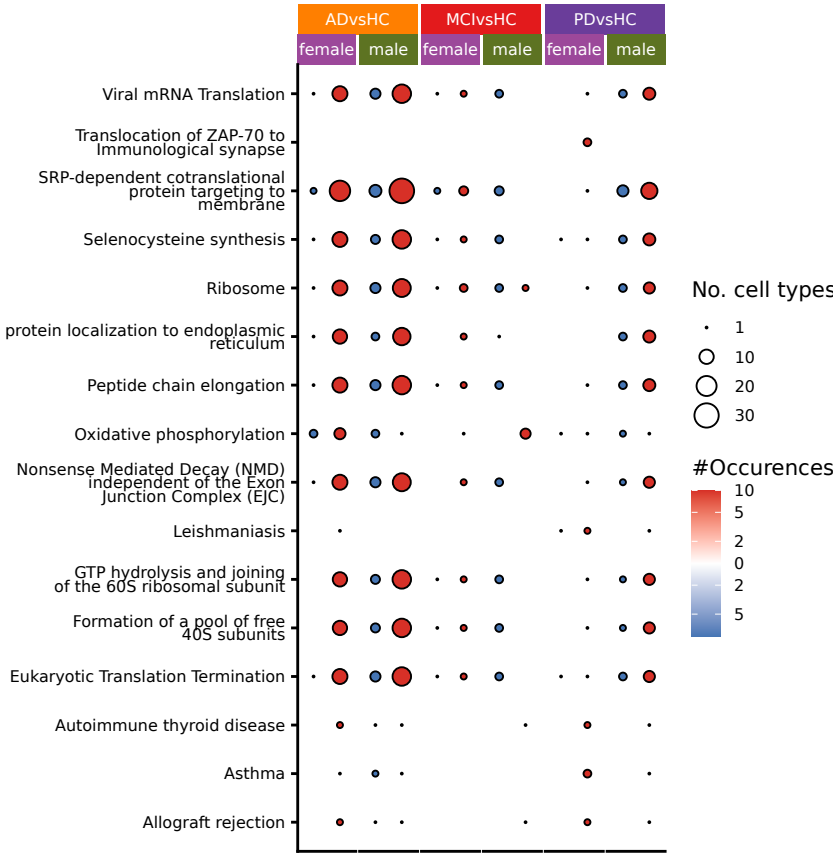


FIGURE 5.17: Top 5 most frequent pathways in each comparison and sex.

The pathways in PD, MCI, and males in AD are similar across all cell types (Figure A.5). However, in females with AD, B cells exhibit a similar cluster of pathways that are associated with junctions (e.g., "adherens junctions", "cell-substrate junction", and "cell-substrate adherens junction") as well as pathways linked to catabolic and biosynthetic processes (e.g., "amide biosynthetic process" and "aromatic compound catabolic process"). In males, these pathways occurred across all cell type groups.

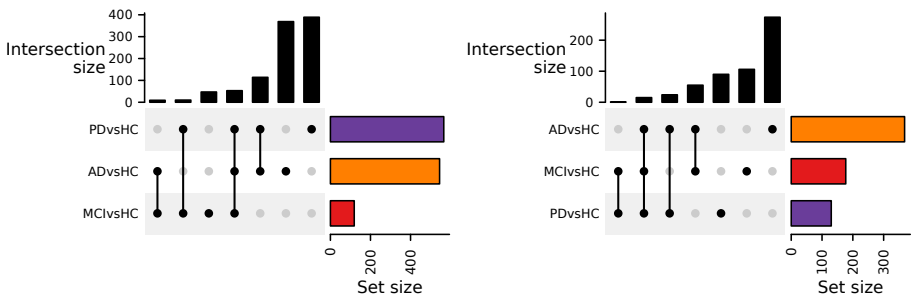


FIGURE 5.18: Overlap of pathways, independent of cell types, in the different comparisons in males (left) and females (right).

Females only share 15 pathways among PD, MCI, and AD, while males share 53. Additionally, 114 pathways are common between AD and PD in males, compared to 24 in females (Figure 5.18).

5.4.4 *Comparison of PBMC and brain cell signatures*

After identifying differences between PBMCs from AD and healthy controls, we aimed to compare the transcriptomes of peripheral cells with those from the brain. Numerous studies have analyzed single-cell gene expression data across various brain regions in different diseases and physiological states.

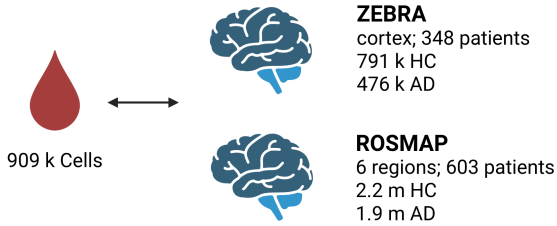


FIGURE 5.19: Brain datasets used for comparison.

ZEBRA is a comprehensive database that combines single-cell data from 21 different studies, enabling detailed insights into gene regulation across various cell types, age, sex, and disease (Figure 5.19) [346]. As few common cell types exist in the brain and blood, our analysis focused on overlapping gene expression patterns and pathways to identify similarities in neurodegenerative diseases.

We discovered genes that show differing expression patterns AD vs. HC across both males and females in PBMCs. We then examined whether these genes showed similar expression changes in the brains of Alzheimer's patients (Figure A.6a). Our results indicate that changes in gene expression in peripheral cells may not necessarily mirror those in the brain, highlighting the differences in cell types between blood and the brain. This underscores the importance of conducting further in-depth studies, especially considering the immune system's role in Alzheimer's disease, the associated brain inflammation, and the infiltration of specifically activated immune cells.

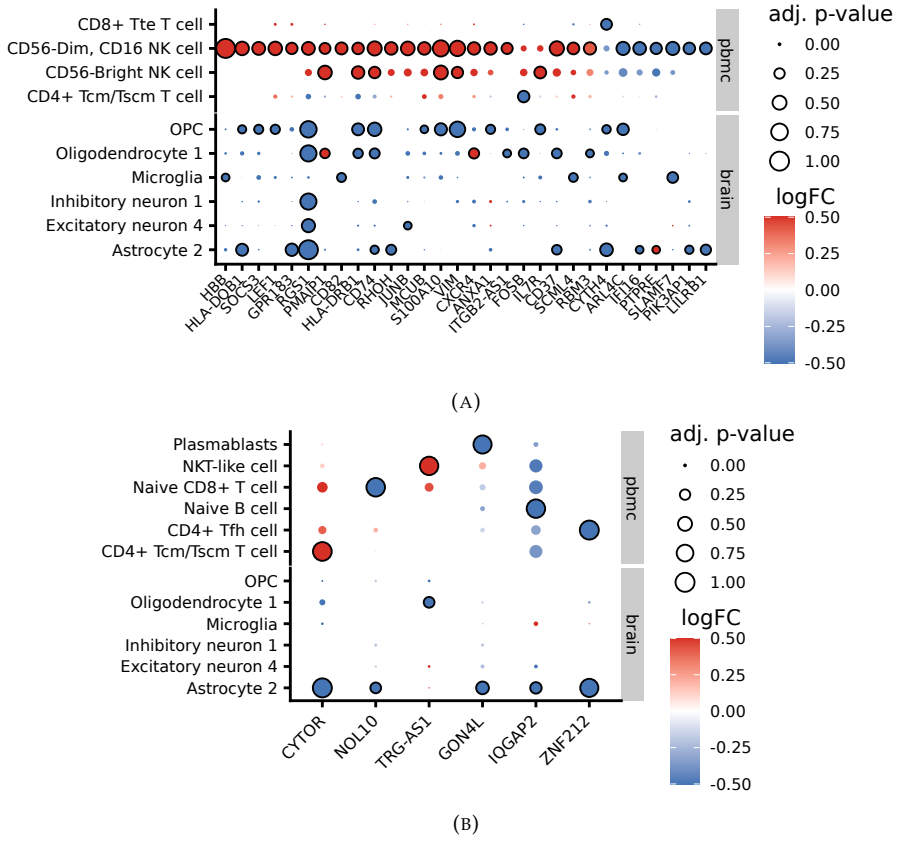


FIGURE 5.20: Genes with adj. $p < 0.05$ in PBMCs and brain in males and with adj. $p < 0.05$ in PBMCs and the brain and abs. $\log_2$ FC > 0.6 in the PBMCs in females.

In total, 36 genes in males and 7 genes in females were significantly deregulated in both PBMCs and the brain across the main cell types (Figure 5.20). The male genes (adj. $p < 0.05$ and abs. $\log_2$ FC > 0.5 in both datasets)

showed enrichment in pathways related to immune system regulation and membrane functions (Figure 5.21, A.6b). In contrast, the female genes (adj. $p < 0.05$ in both datasets) were associated with the Herpes Simplex Virus 1 pathway. Previous research has linked herpes simplex to an elevated risk of developing Alzheimer's disease, positioning it as a potential factor in the progression of the illness [349, 350].

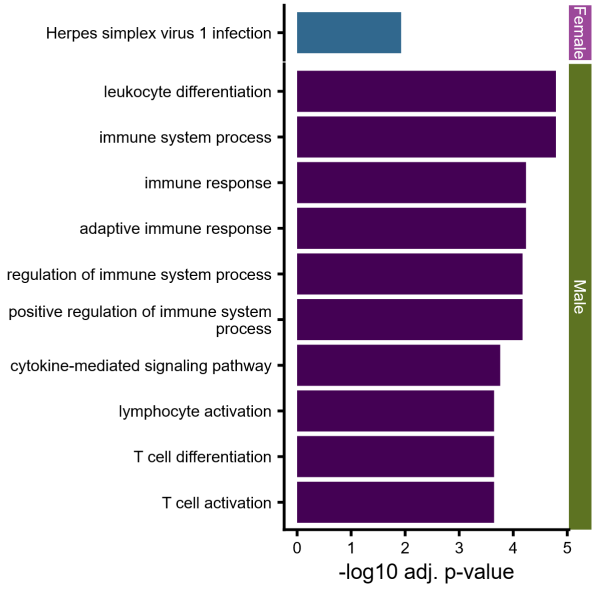
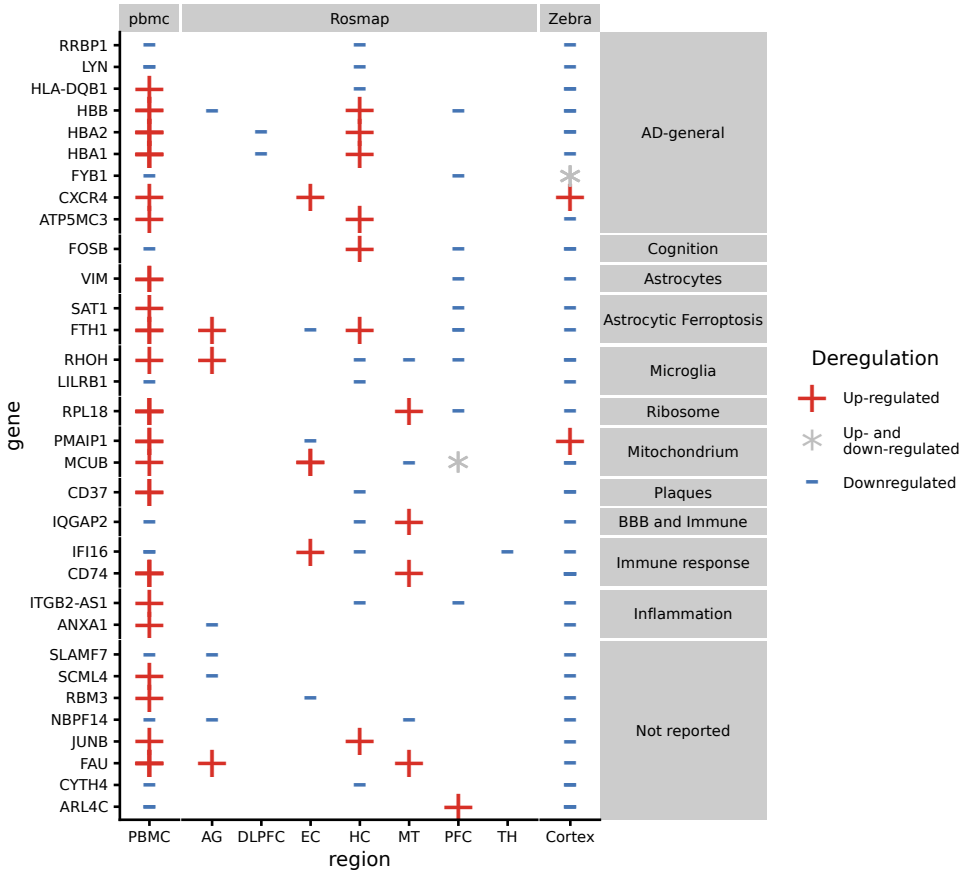


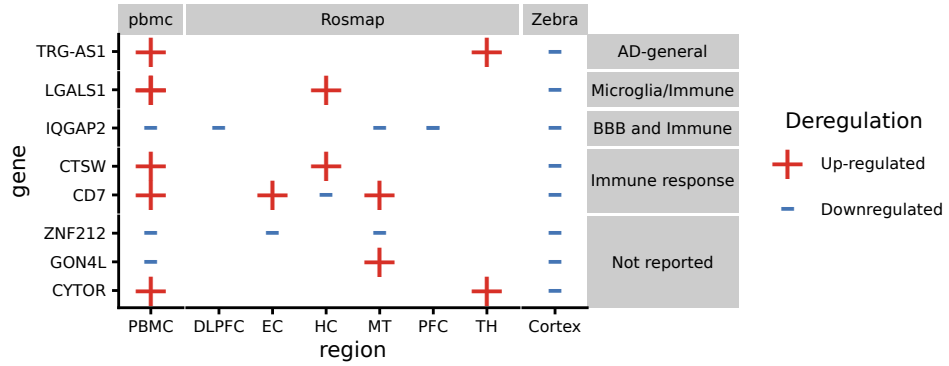
FIGURE 5.21: Top 10 enriched pathways (by p-value) of the PBMCs and brain DEG overlap.

To verify these findings and assess whether expression levels vary across different brain regions, we examined the overlap of DEGs in the PBMC dataset, the ZEBRA dataset (which includes 21 studies), and an additional dataset from the Religious Orders Study and Memory and Aging Project (ROSMAP), which contained 4 studies across 6 brain regions. It is important to note that two studies from the prefrontal cortex within the ROSMAP dataset are part of the ZEBRA dataset, making them non-independent. The other brain regions, however, are distinct between the datasets. Our analysis revealed an overlap of 32 genes in male and 8 in female patients. Of these, 24 out of the 32 male genes and 5 out of the 8 female genes have been previously reported in the context of AD, primarily in the immune system, blood-brain barrier, astrocytes, and microglia (Figure 5.22, Table A.2).

To better understand potential interactions, we compared the DEGs with the CellChat signaling database [351]. We identified genes related to the CCL (Figure A.6c) and the MHC-I signaling pathway in the deregulated genes in females, while males showed changes in the CD45 and CCL signaling pathways. These genes have been previously reported in the context of Alzheimer's [352, 353, 354]. This suggests that changes in the blood may



(A)



(B)

FIGURE 5.22: Direction of deregulation for the common DEGs between all three datasets and the direction of deregulation (+: Consistent upregulation in all cell types, -: Consistent downregulation, *: Mixed). Previously reported genes are annotated (Table A.2)

affect the brain or vice versa; however, because of the limited number of DEGs in these signaling pathways, additional experiments are necessary to validate this.

Using a pathway analysis, we found a general enrichment of membrane, ribosome, and adherens junction-related pathways in both blood and brain (Figure 5.23). Especially the "SRP-dependent cotranslational protein targeting to membrane" pathway is frequently enriched in all three datasets. This suggests that the observed symptoms might be caused by a systemic change that affects both blood and the brain.

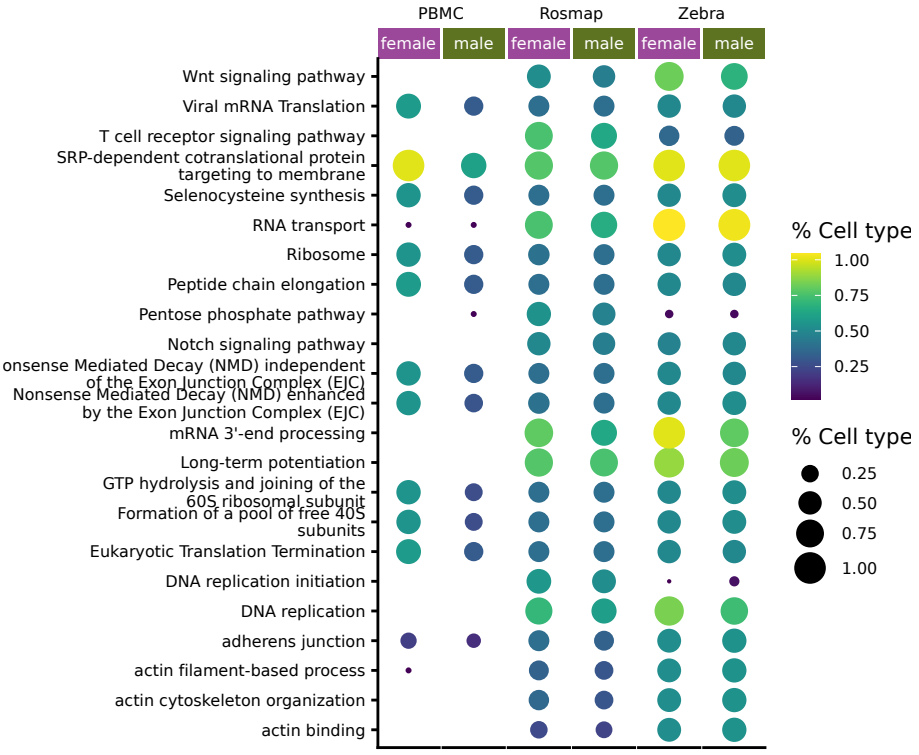
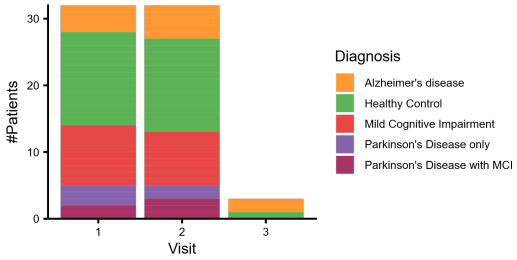
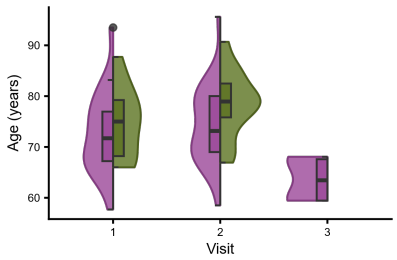


FIGURE 5.23: Top 10 pathways in the three datasets and their frequency among the cell types.

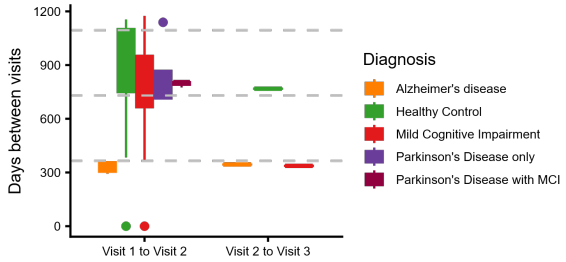
To further explore the relationship between changes in gene expression across both tissues, we analyzed an additional bulk RNA-sequencing dataset from the ROSMAP cohort (Synapse: syn3388564). While the fold-changes observed in the brain single-cell and bulk data showed a significant positive



(A) Number of patients at the first, second, and third visit (for patients with >1 visit). Of note, diagnosis can change between visits.



(B) Age distribution of longitudinal patients



(C) Duration between visits (in days).

FIGURE 5.24: Overview longitudinal data.

correlation (corr = 0.14, p = 0.00025), we did not observe a significant correlation between the PBMC single-cell and the brain bulk data (Figure A.6e).

5.4.5 *CSF and brain data*

The dataset included scRNA-seq data used for the analysis, longitudinal data, brain volume measurements, and CSF-marker values.

This included more than one time point for 32 patients and more than two for 3 patients (Figure 5.24). All patients with three visits were female, and we noted variations in age distribution between male and female patients within this subgroup. The average interval between the first and second visits was under a year for Alzheimer's disease (AD) patients, while it ranged from two to three years for those with other conditions.

We found no significant correlations when analyzing the relationship between cell-type proportions and visit numbers (Figure A.7a). Overall, the cell

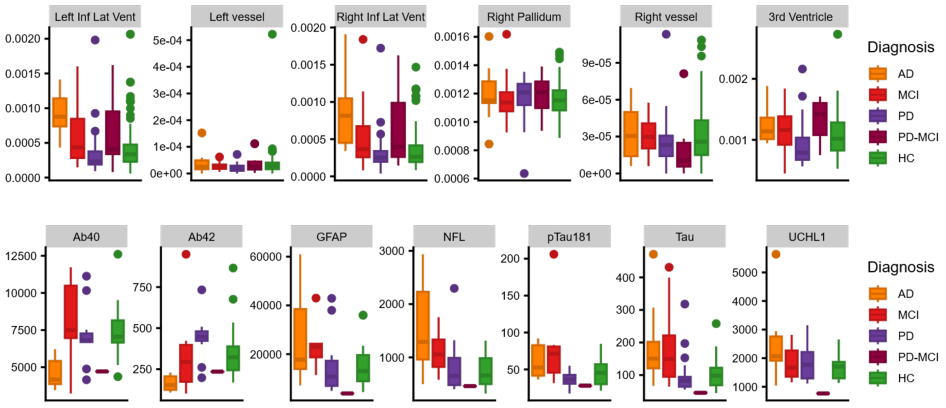


FIGURE 5.25: Comparison of known AD biomarkers in the CSF (top) and brain volume measurements (bottom).

type proportions within the same patient across different visits and among various patients showed large variations (Figure A.7b).

Similarly, we correlated brain volume measurements and CSF marker values with the cell type proportions (Figure 5.25). We found significant correlations (adj. $p < 0.05$) primarily in healthy controls.

Overall, our dataset revealed differences based on sex in the alterations of cell type composition and gene expression patterns associated with neurodegenerative diseases, particularly Parkinson's disease.

5.5 DISCUSSION

Single-cell -omics is crucial for understanding the systemic changes that occur before and alongside neurodegeneration, a field essential for identifying early biomarkers as well as for developing individualized markers for risk and disease progression. Our dataset includes 290 gene-expression profiles of PBMCs from individuals with MCI, AD, PD, PD-MCI, and healthy controls (HC).

Even though this dataset exceeds published PBMC datasets for neurodegeneration, the study of the exact biological mechanisms would require larger and more specialized datasets. This not only means larger sample

sizes [311], but also additional demographic and clinical data (e.g., medication use, other diseases, lifestyle factors). Especially the changes occurring in sub-cell types require further investigation through targeted experiments, as the number of cells in each cluster remains limited.

When examining the composition of cell types, we observed significant differences between men and women, often specific to particular sub-cell types. The changes in B cells and monocytes in Alzheimer's patients were inconsistent between male and female subjects. Previous research has indicated that the depletion of B cells in mice is associated with cognitive deficits and an increase in amyloid-beta plaques [263]. Furthermore, B cells play an important role in modulating local immune responses, potentially serving as indicators of inflammatory states in Alzheimer's. The increased prevalence of monocytes in AD could indicate chronic inflammation, which may negatively affect structures like the blood-brain barrier (BBB) [266]. In contrast to the pronounced sex-specific differences in PD cell type compositions that we noted, earlier studies mainly reported changes in the T-cell populations of Parkinson's patients. It's important to note that previous studies occasionally had conflicting results, which may partially stem from the substantial variation in cell type compositions observed in the same patient [355], leading to significant variability in cell type composition between samples. Additionally, some of our findings have not been documented previously or only partially aligned with earlier research. This inconsistency could be due to cohort biases (such as variations in ethnicity, disease severity, or other factors), technological limitations, or other biases.

Our data also suggest sex-specific differences in gene-expression signatures for PD, as well as more smilar but still sex-specific differences in AD. Previous findings have identified sex-specific differences in monocytes relevant to PD [262], while T-memory cells exhibited sex-independent expression changes [332]. We discovered an anti-correlation in gene-expression changes between males and females, highlighting substantial differences in disease-specific changes in PD; however, the sex-specific differences in AD were primarily found in the magnitude of gene-expression changes. Notably, we identified a larger number of DEGs in females compared to the males. Although this trend was consistent across all diseases examined in

this study, it was especially noticeable in AD, an observation previously reported by Mathys et al.[258]. In Parkinson's, this phenomenon has been noted in blood monocytes[262], while an opposing observation has been made in brain tissues[356].

The gene changes observed in Alzheimer's disease, independent of sex, mainly affect the "SRP-dependent cotranslational protein targeting to membrane" pathway. This pathway has been previously linked to both Alzheimer's [357] and Parkinson's diseases [358, 359], suggesting similar alterations in male and female AD and male PD patients. The genes involved in this pathway showed upregulation during microglia activation, a state associated with Alzheimer's [357]. This "SRP-dependent cotranslational protein targeting to membrane" pathway has also been identified in brain samples from both the ZEBRA and ROSMAP datasets. The genes in this pathway thus have shown changes not only in PBMCs but also within the brain. Interestingly, these genes have also been discussed in relation to periodontitis, a condition primarily attributed to Porphyromonas gingivalis [357, 360]. This bacterium is known to produce gingipains, which have been associated with Alzheimer's pathology in both mouse models and human studies [361, 362].

In our investigation of the changes observed in both blood and brain tissues, we identified 30 genes in males and 7 genes in females that are significantly deregulated in both. Most of these genes have previously been linked to Alzheimer's disease across various tissues and cell types (Supplementary Table 3). Notably, these include genes associated with immune system changes and the blood-brain barrier, particularly in females, as well as those related to astrocytes and microglia. This suggests general changes within the immune system, potentially linked to infections (such as herpes simplex) or age-related inflammatory processes. Infection with herpes simplex is known to increase the risk of Alzheimer's for APOE4 carriers [349, 350], and is discussed to significantly contribute to the development of the disease.

As the patients in the ADRC cohort are continuously monitored, two key limitations of this dataset will be addressed in future research. First, investigating longitudinal data will be crucial to understanding the temporal dynamics at the molecular level. Second, conducting a patient-matched analysis of both brain and blood tissues will help validate the effects reported here.

In conclusion, our dataset suggests sex-related differences in the neurodegenerative diseases. While the precise relationship between the immune system and the brain remains unclear, this research highlights the need to consider the potential impacts of the immune system and sex on disease progression.

Building on the findings, the second project explored how the mouse organism and the immune cells react to a specific systemic disturbance in a controlled environment. While neurodegeneration represents a clinically diverse and complex issue, spaceflight acts as a well-defined physiological stressor, with factors like microgravity and radiation that affect multiple organ systems. The two projects both study changes in immune cells, tissue-resident cells, and extracellular matrix remodeling but differ in their contexts. One explores age-related diseases in humans, while the other explores induced adaptations in mice. Mice were selected as the model organism because they allow us to study organ-specific changes within an individual. The analysis methods used in the neurodegeneration project to integrate results and enhance data visualization were expanded in the spaceflight study, allowing for comparisons across tissues, age groups, and RNA types. Overall, both studies provide insights into how systemic changes and diseases influence immune and organ function, aiming at a systems-level understanding of immune adaptation.

6

EFFECT OF SPACEFLIGHT ON MURINE MRNA AND MIRNA LANDSCAPE

This chapter contains results, figures, and descriptions from the preprints [F1, F2]. See chapter 4 for the author contributions for each of the preprints and publications. Specifically, section 6.3.2 describes the methods, and section 6.4 the results of [F1]. Section 6.3.3 contains descriptions of the methods, and section 6.5 the results of [F2].

Since 1960, space travel has led to more and more people visiting space in missions that get more frequent and last longer. This development is partially due to the improvements in technology and the decreasing costs of spaceflight. The costs of reaching low earth orbit dropped from $54,500 / kg to $2,720 /kg in the recent decades - a 20-fold decrease - and continues to fall [363]. This enables a fast expansion in human activity and long-duration missions in space. Prolonged exposure to spaceflight induces physiological effects that get stronger with mission duration [364, 365, 366], especially problematic with long missions beyond the earth's orbit, and continue temporarily after returning to earth [367]. Current ISS missions typically last a few weeks to months. When aiming at missions to Mars, these future missions will require much longer durations in space [368]. To mitigate these effects, molecular studies are required to identify biomarkers and therapeutic targets.

6.1 BACKGROUND OF SPACE STUDIES

Previous studies, such as the NASA Twin Study [369], demonstrated spaceflight-induced changes in the cardiovascular, musculoskeletal, and neural function [370]. Databases such as NASA's GeneLab integrate microbiome profiles, scRNA-seq, scATAC-seq, and bulk gene expression across multiple

tissues and organisms, facilitating access to these data for cross-study analysis [371] Building upon this collection and integrating it with data from the JAXA CFE study [372, 373], the SpaceX Inspiration4 crew [374, 375, 376] and the Axiom and Polaris missions, the Space Omics and Medical Atlas (SOMA) [377] provides a growing resource of space-related multi-omics datasets [378].

Previous research analyzed RNA-seq data from the liver, kidney, adrenal gland, thymus, mammary gland, skin, and skeletal muscle from NASA's Genelab and identified miRNA changes that regulate adaptation to microgravity and radiation through the TGFB1 and p53 genes [379], a signature that has been further explored on the circulating miRNA profiles [380].

Spaceflight caused changes similar to age-related diseases [93, 381] (e.g., muscle atrophy and bone loss), suggesting that both affect common or similar pathways. As miRNAs also show deregulation in cardiovascular diseases, cancer, and aging [382, 383, 384] across multiple tissue types [383], system-wide analysis of spaceflight effects is necessary. Although previous studies have examined the deregulation in different tissues, none have sampled them from the same mice.

Human spaceflight studies are restricted to minimally invasive sampling such as body fluids [377] and thus lack systemic insights. Mice, a well-established model organism, allow us to study the systemic effect of spaceflight on the body. Simulated microgravity and parabolic flights [385] are frequently used to study the effect on mice and humans but do not allow long-term exposure to real microgravity. Microgravity studies in space [386] that allow the long-term exposure of organisms [387] are rare.

To provide a comprehensive view of spaceflight effects across organ systems, we profiled multiple tissues from mice that were sent to the ISS using single-cell RNA sequencing and bulk small non-coding RNA sequencing. Building on previous work from the Tabula Muris and Tabula Muris Senis datasets [388, 389, 390], as well as prior miRNA datasets [391, 383], we expanded the mouse tissue atlas that allows us to explore the age-specific spaceflight effects at the mRNA and miRNA level.

6.2 STUDY DESIGN

We performed scRNA-seq and bulk sncRNA sequencing in female Balb/C mice to study the effect of spaceflight on different cells, tissues, and organs (Figure 6.1).

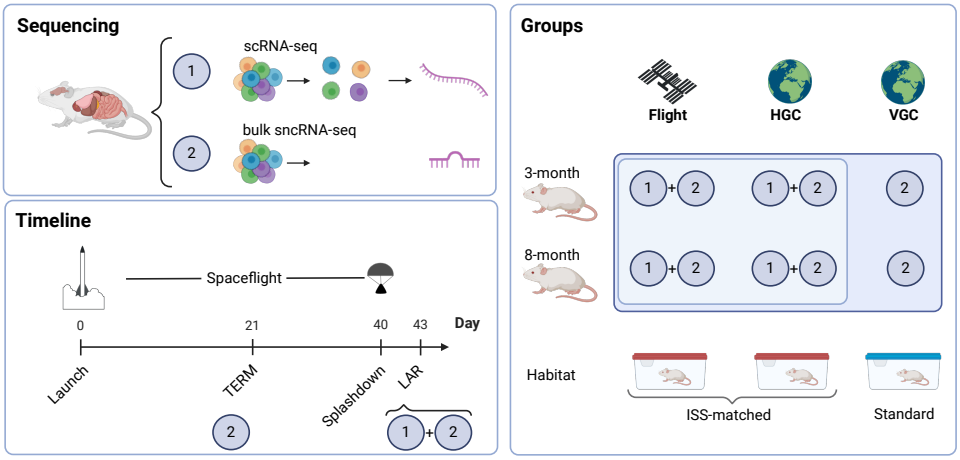


FIGURE 6.1: Overview of the spaceflight project [F1, F2]. Created with BioRender.com.

Two age groups were included to capture the typical age range of astronauts in order to study age-dependent changes: one 3-month-old group, equivalent to young adults, and one 8-month-old group, equivalent to mature adults.

Besides the flight (FL) group, which was exposed to spaceflight, we included two control groups. The first one, Vivarium Ground Control (VGC), was housed under standard conditions on earth. Even though double-density housing only minimally affects mice [392], other parameters such as CO2 concentration, temperature, and humidity [393] impact mice physiology and behavior [393]. The second control group, Habitat Ground Control (HGC), was thus included, and the environmental conditions of the ISS were matched to control these factors.

To identify spaceflight-related signals from the mice's return stress, the mice were taken down at two different time points. The first group called ISS Terminated (TERM), was taken down directly on the ISS after 21 days of spaceflight. The second group called Life Animal Return (LAR), was taken

down three days after returning to earth from a 40-day spaceflight. The time points were matched for the earth control groups.

Single-cell RNA sequencing was only performed for the Flight and HGC mice from the LAR group, including both age groups.

Spaceflights affect the musculoskeletal, pulmonology, and cardiovascular systems, as well as neurology, immunology, gastroenterology, and dermatology (Figure 6.2). It can cause a variety of symptoms, ranging from muscle atrophy, bone loss, and cardiovascular deconditioning to immune system dysregulation [394, 395, 396, 397, 398, 399].

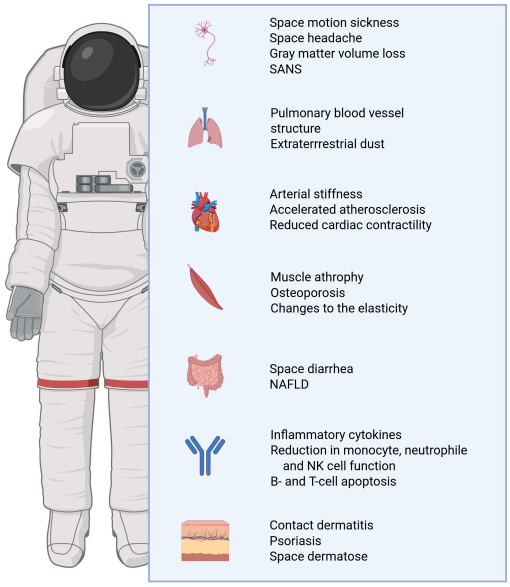


FIGURE 6.2: Effects a spaceflight can have on different organs, based on Figure 1 in [400]. Created with BioRender.com.

The change in fluid distribution and cardiovascular load in astronauts can lead to cardiac remodeling and vascular changes ([401]). We thus sequenced samples from **aorta** and **heart** (**ventricle & atrium**) to study the changes in heart function and structure at the molecular level. Fat tissues such as Marrow Adipose Tissue (**MAT**), Brown Adipose Tissue (**BAT**), Subcutaneous Adipose Tissue (**SCAT**), Gonadal Adipose Tissue (**GAT**) play an important role in metabolism and energy homeostasis, processes that are affected by spaceflight. In combination with sequencing data from **liver** (affected lipid metabolism in space) and **pancreas** (affected insulin signaling and glucose metabolism), these tissues provide insights into the molecular changes underlying metabolic disruptions in these tissues. Muscle Atrophy, bone loss, and reduced osteogenesis are well-known effects of spaceflight. We thus sequenced skeletal muscle (**diaphragm**, **limb muscle**) and **bone** and **skeletal stem cells** to study the effect spaceflights have on muscle structure and function as well

Organ group	scRNA-seq	sncRNA-seq
Cardiovascular	Aorta	-
	Heart (Ventricle)	Heart
	Heart (Atria)	
Bone	Bone	-
	Skeletal Stem Cells	-
Skeletal Muscle	Diaphragm	Diaphragm
	Limb Muscle	-
Fat Tissue	BAT	BAT
	MAT	MAT
	GAT	GAT
	SCAT	SCAT
Digestive organs	Liver	Liver
	Pancreas	Pancreas
	Large Intestine	-
Brain Tissue	Brain - Cerebellum (Myeloid)	Brain
	Brain - Forebrain (Myeloid)	
	Brain - Hippocampus (Myeloid)	
	Brain - Cerebellum (Non-myeloid)	
	Brain - Forebrain (Non-myeloid)	
	Brain - Hippocampus (Non-myeloid)	
Immune system & Lymphoid	Blood	-
	Bone Marrow	-
	Thymus	Thymus
	Spleen	Spleen
Other organs	Eye	-
	Skin	-
	Mammary Gland	-
	Kidney	Kidney
	Lung	Lung

TABLE 6.1: Overview of organs profiled using single-cell RNA-seq and small RNA sequencing in the spaceflight experiment [F1, F2]

as bone remodeling. Sequencing the **brain** (**hippocampus, forebrain, cerebellum; myeloid and non-myeloid cells**) provides insights into the molecular mechanisms impacting cognitive function, neuroinflammation, and intracranial pressure in space. To study changes related to the immune system, we sequenced the primary lymphoid organs **bone marrow** and **thymus**, the secondary lymphoid organ **spleen**, and the **blood** to track changes in hematopoiesis and T-cell development. Tissue-resident immune cells from other organs complete this picture. Other sequenced tissues include **skin**, **eyes**, **mammary gland**, **kidney**, **large intestine** and **lung**.

We sequenced 28 tissues using single-cell RNA sequencing and 13 tissues using bulk small RNA sequencing (Table 6.1)

6.3 MATERIALS AND METHODS

For this experiment, forty mice were sent to the ISS on 12/05/2018 (Fl group), while two control groups, HGC and VGC, with 40 mice each, remained on earth. The mice were divided into two age groups (3 and 8 months). From each group, two mice were selected (total $n = 8$) for single-cell RNA-seq profiling[1].

6.3.1 *Housing conditions*

VGC mice were housed under standard vivarium conditions (4 mice/cage, 20–22.2 °C, 12h light/dark cycle). HGC mice were housed in conditions matching the Flight group's environment, with the same Transporter and Habitat Units at double density, replicating flight conditions for humidity, CO_2, and temperature.

Each group was split into two subgroups:

- **Terminal Group (TERM)**: Sacrificed 22–24 days after launch, either on the ISS (Flight) or on Earth (VGC, HGC).

[1] Ethics approval by: NASA Flight Animal Care and Use Committee (Protocol #FLT-18-116), Scripps Research Institute (Protocol #09-0004) and Stanford University IACUC (Protocol #WYS1591)

- **Live Animal Return (LAR)**: Returned to Earth after 39.5 days in space, taken down 2–4 days post-splashdown on Earth.

Feeding with NASA Nutrient Food Bars was started four weeks before the launch. Food was replenished weekly, and food and water were available ad libitum.

Two notable deviations from the plan affected the study design:

1. **Contaminated food bars:** Within 24 hours of launch, mold on the food bars was discovered in both the Flight and HGC groups. Replacement bars were mounted using tape instead of epoxy, resulting in unsecured bars during launch and reentry. Mold recurrence on the ISS led to early termination of the TERM group (after 3 instead of 8 weeks), which was matched on ground controls.

2. **Delayed return and housing stress:** Inclement weather delayed the return of the LAR mice, extending confinement in double-density housing from 2–3 to 7 days. During this period, food bars floated freely, and monitoring was impossible. Similar conditions were replicated for HGC mice. Additionally, rough sea conditions during splashdown may have further stressed the Flight group, an effect not mirrored in the ground controls.

6.3.2 *Single-cell mRNA Data*

To generate the single-cell dataset, sample collection, preparation, sequencing, and analysis were performed as follows [F1]:

Tissue dissociation and sample preparation

All mice in the single-cell experiment belonged to the Life Animal Return (LAR) Group and thus returned to earth alive before being let down.

After euthanization via CO_2 inhalation and cervical dislocation, cardiac puncture was used for blood collection (0.5 - 1 mL). The blood was collected into K2/EDTA-coated syringes and MiniCollect tubes to prevent clotting. The red blood cells were depleted, and the remaining cell suspension was

pelleted (300xg, 4 °C, 5 min), filtered (70 μm strainer), and resuspended. The exact reagents can be found in Table B.1.

The eyes were collected with forceps in cold Phosphate-Buffered Saline (PBS). Subsequently, extraocular muscles, retina, cornea, lens, and optic nerve tissue components were separated, and each mouse's tissues were digested separately for single cell preparation (Table B.2), mechanically dissociated using pipetting and pooled together as an eye cell suspension. Ppooled cells were resuspended in fetal bovine serum (FBS) (1%) with PBS (106 cells/ml).

For the bone tissue, the femurs, hip bones, and vertebral columns were dissected and cleaned of soft tissue. Bones were processed as previously described [402, 403]; bones were crushed and digested (37 °C, 60 min, constant movement). Cells were strained (100 μm nylon filter), washed in staining medium (10% FBS in PBS), pelleted (200xg, 4°C), and finally resuspended in staining medium. Red blood cells were depleted (ACK lysis, 3 min), and the cells were washed and pelleted again (200 xg, 4°C). Fluorescence-activated cell sorting (FACS) was prepared by staining the cells with fluorochrome-conjugated antibodies (CD45, TER119, CD51, Tie2, CD31, Th1.1, Thy1.2, 6c3, CD105 and streptavidin–PE–Cy7 conjugate). Cells were sorted on a FACS Aria II Instrument (BD BioSciences) (70–μm nozzle), using appropriate controls and propidium iodide to exclude dead cells, with sorted cells directly processed for 10X Genomics scRNA-seq. Information about the Reagents can be found in Table B.3

All other tissues were processed as previously described [389].

Microfluidic droplet single-cell sequencing

Single cells were captured in droplet emulsions, and scRNA-seq libraries were constructed following the 10x Genomics protocol, targeting up to 5,000 cells per sample. Enzymatic reactions were performed in a Thermal cycler with 12 cycles for cDNA amplification and sample index PCR. Library preparation was conducted one organ type at a time, normalizing cDNA and running 13-20 PCR cycles based on cDNA amount. Fragment length was assessed using a fragment analyzer and qPCR. Libraries were diluted to 4nM,

pooled equally, and sequenced on the NovaSeq 6000 with a PhiX control library (0.2 to 1%). Detailed kit and instrument information is available in Table B.4

Data pre-processing

Sequencing reads were aligned using `cellranger` (version 5.0.0) [115] with the mm10 reference genome. Ambient RNA contamination was removed using `SoupX` (version 1.5.2) [117]; samples with >10% estimated contamination were excluded. Gene counts were processed with `Seurat` (version 4.0.3) [404], filtering cells with <200 or >2500 detected genes, or >5% mitochondrial reads. Samples with <50 cells were discarded. Doublets were identified and removed using `DoubletFinder` (version 2.0.3) [122]. Highly variable genes (n=2000) were selected per sample. Data were normalized, scaled, and integrated across batches using Seurat's integration workflow. Dimensionality reduction was performed via PCA (30 components; elbow plot selection) followed by UMAP for 2D embedding. Clustering was performed using `FindNeighbors` and `FindClusters` with resolution = 0.8. Clusters were annotated manually based on canonical marker genes. When needed, clusters were sub-clustered to improve resolution. Annotated cell types were mapped to Cell Ontology terms. The skin tissue was excluded due to insufficient cell type resolution. To evaluate sample group balance (Space 3M, Space 8M, Control 3M, Control 8M), a group imbalance score was computed per tissue t as:

$$B(t) = \sum_{g \in G} \left(\frac{|g_t|}{\sum_{h \in G} |h_t|} - \frac{1}{|G|} \right) \tag{6.1}$$

where $G = \{\text{Fl 3M}, \text{Fl 8M}, \text{HGC 3M}, \text{HGC 8M}\}$ and $|g_t|$ and $|h_t|$ denotes the number of cells of the group $g, h \in G$ for the tissue t. Tissues with $B(t) > 0.125$ were considered unbalanced.

Analysis

Differential gene expression was performed using Seurat's `FindMarkers` with the Wilcoxon rank-sum test, including the sample ID as a latent variable. Genes with Bonferroni-adjusted $p < 0.05$ and absolute $\log_2$ fold-change >

0.5 were considered significant. GSEA was performed with `GeneTrail` (version 3.2) [347] using gene lists ranked by $-\log_{10}(\text{adj. p}) \cdot \text{sign}(\log_2 \text{FC})$. Pathway databases included GO (BP, CC, MF), KEGG, Reactome, WikiPathways, and Pfam. The Kolmogorov-Smirnov test was used with default parameters. ORA was performed using `clusterProfiler` (version 4.6.0) [405], with GO terms (BP, CC, MF). Pathways with adj. $p < 0.05$ were considered enriched or depleted. Pseudo-bulk expression matrices were generated using Seurat's `AverageExpression` function and analyzed with `pvcaBatchAssess` (`pvca` version 1.30.0, threshold = 0.75). PVCA was applied separately by tissue and by cell type. Spliced/unspliced matrices were generated using `velocyto` (version 0.17.17) [233], Seurat objects were converted to AnnData using `SeuratDisk` (version 0.0.9019) and used for RNA velocity analysis with `scvelo` (version 0.2.2) [234]. Cell-cell signaling was inferred using `CellChat` (version 1.5.0) [218], using only shared cell types between samples. The `stats::cor.test` function (version 4.2.3) was used to calculate Pearson's correlation coefficients, and BH-adjusted $p < 0.05$ was considered significant.

We used the following formulas to determine the specificity scores. For any cell type $\hat{k} \in K$ and tissue $\hat{t} \in T$, we define the similarity between two clusters $c_{\hat{k},\hat{t}}$ and $c_{\tilde{k},\tilde{t}}$ in C as follows:

$$\text{similar}(c_{\hat{k},\hat{t}}, c_{\tilde{k},\tilde{t}}) = \begin{cases} 1, & \text{if } s(c_{\hat{k},\hat{t}}, c_{\tilde{k},\tilde{t}}) > 0.6; \\ 0, & \text{otherwise} \end{cases} \tag{6.2}$$

$s(c_{\hat{k},\hat{t}}, c_{\tilde{k},\tilde{t}})$ is the cosine similarity of two gene lists, ordered by fold-change.

Cell type similarity is defined as

$$\text{CellSim}(c_{\hat{k},\hat{t}}) = \frac{1}{|C_{\hat{k}}|} \sum_{t \in T} \text{similar}(c_{\hat{k},\hat{t}}, c_{\hat{k},t}), \tag{6.3}$$

while tissue similarity is defined as

$$\text{TissueSim}(c_{\hat{k},\hat{t}}) = \frac{1}{|C_{\hat{t}}|} \sum_{k \in K} \text{similar}(c_{\hat{k},\hat{t}}, c_{k,\hat{t}}) \tag{6.4}$$

where $C_{\hat{k}} = \{c_{k,t} \in C | k = \hat{k}\}$ and $C_{\hat{t}} = \{c_{k,t} \in C | t = \hat{t}\}$.

The specificity score is then determined as follows:

$$\text{specificity}(c_{\hat{k},\hat{t}}) = \begin{cases} \text{CellSim}(c_{\hat{k},\hat{t}}), & \text{if CellSim}(c_{\hat{k},\hat{t}}) > \text{TissueSim}(c_{\hat{k},\hat{t}}); \\ -\text{TissueSim}(c_{\hat{k},\hat{t}}), & \text{if CellSim}(c_{\hat{k},\hat{t}}) < \text{TissueSim}(c_{\hat{k},\hat{t}}); \\ 0, & \text{otherwise.} \end{cases} \tag{6.5}$$

It is positive when the cluster exhibits more similar connections within the same cell type and negative when it demonstrates stronger similarities with clusters from the same tissue.

6.3.3 *Bulk sncRNA data*

In addition to the single-cell data, an additional sncRNA-seq dataset was generated [F2]. It contained LAR and TERM mice, as well as VGC, as an additional housing control group.

Tissue dissociation and sample preparation

LAR mice were anesthetized (gas) and euthanized by exsanguination and double thoracotomy. Livers were dissected into lobes, weighed, and snap-frozen in liquid nitrogen within 10–16 minutes post-thoracotomy. All samples were stored at $-80\,^{\circ}\text{C}$ or colder throughout processing and transport.

TERM mice were anesthetized with ketamine/xylazine/acepromazine (120/15/3 mg/kg) and euthanized via exsanguination and double thoracotomy. Spleens were dissected into RNAlater; carcasses were foil-wrapped and frozen aboard the ISS in MELFI or ground freezers at $-80\,^{\circ}\text{C}$. Of note, the free-floating of the carcasses in microgravity prevented snap-freezing as they were in direct contact with air. Upon return, carcasses were thawed (60–90 min, staggered) to allow the dissection of multiple tissues, including the liver, which was processed analogously to LAR samples.

Bulk micro RNA sequencing

Total RNA was extracted using a modified miRNeasy-96 protocol: Tissues were trimmed as follows:

- Liver, spleen, kidney: ~25 mg
- Pancreas: ⩽25 mg (if >45 mg)
- GAT, MAT, SCAT: ~100 mg (if >200 mg)
- BAT, lung: ~50 mg (if >100 mg)
- Thymus: ~25 mg (if >50 mg)
- Heart, hemi-brain: homogenized whole with TissueRuptor (1ml QIAzol per 50mgs tissue)
- Diaphragm: used in full

All samples were randomized in 96-well plates. The 96-well collection microtube plate was chilled, and trimmed tissue samples were placed in each well. One 5mm chilled steel ball and 700 µL pre-chilled QIAzol (or homogenate for heart/brain) were added. Plates were homogenized on a TissueLyser II (4×5 min at 25 Hz, alternating arms), rested for 5 min at room temperature, and centrifuged (6000×g, 1 min, 4 °C).

Debris- and bead-free homogenate was then transferred to a new plate, to which 140µl chloroform was added to each well, and plates were shaken for 3min (room temperature) and centrifuged (6000 x g, 4min, 4°C). 300µl aqueous phase was collected and transferred to a new S-Block. 525µl 100% EtOH was added, mixed by pipetting, and spun at room temperature (6000 x g, 30s). It was then transferred to a miRNeasy-96 plate on an S-plate, sealed (AirPore tape sheet), and spun again at room temperature (5600 x g, 4min). After discarding the flow and a washing step (800µl RWT), it was spun again (5600 x g, 4min), washed twice with 800µl RPE, and spun for 10min. Each column was finally eluted twice with 50µl RNase-free water, followed by a 1 min incubation at room temperature and centrifugation (5600 x g, 4min). All RNA was stored at −80 °C.

Small RNA libraries were prepared using the MGIEasy Small RNA Library Prep Kit on the SP-960 system. After adapter ligation and reverse transcription (barcodes 1–4, 13–16, 25–32), libraries were amplified via 21-cycle PCR and purified via size selection through AMPure XP beads (Beckman Coulter, Germany). Library quality was assessed via the Agilent DNA 1000 Kit (Agilent Technologies) and Qubit HS quantification (Thermo Fisher Scientific).

Sixteen samples were pooled per library. BGI (Hong Kong) circularized and sequenced 45 libraries on the BGISEQ-500RS (SE50, single-end).

Data pre-processing

Sequencing reads were mapped to the GRCm39 reference and miRBase (version 22.1) [406] using `miRmaster` (version 2.0) [116]. MiRNA family annotations were retrieved from miRGeneDB 3.0 [407]. MiRNAs were filtered to have at least 5 reads in 10% of samples within each group (condition-tissue) and data was normalized with reads per mapped million (rpmm).

Analysis

PVCA was performed using `pvca::pvcaBatchAssess` (version 1.30.0) on log2-transformed, filtered rpmm values. The top principal components were chosen to explain 90% of the variance. Factors tested included Age, Tissue, Space (Flight vs. Controls), Environment (VGC vs. non-VGC), and their combinations. The brain was excluded from the Age × Condition comparison due to having only one age group in LAR.

UMAP was applied using the `umap` R package on log2-transformed, filtered rpmm data for two-dimensional embedding.

Differential Expression Analysis of mRNA and miRNA was performed as defined in Table 6.2.

Dataset	**Input**	**Statistical Test**	**Significance Criteria**
miRNA	Filtered rpm counts (log2-transformed)	Wilcoxon Rank-Sum Test (BH-adjusted)	abs(log2FC) > 1 Cohen's d > 0.5
mRNA (pseudobulb)	Aggregated raw counts (`aggregateData`; fun = "sum"; `muscat` (version 1.12.0) [408])	Wilcoxon Rank-Sum Test (BH-adjusted)	abs(log2FC) > 1 Cohen's d > 0.5

TABLE 6.2: Differential Expression Analysis for mRNA and miRNA data.

Top deregulated mRNAs for downstream pathway analysis were selected based on fold-change rank. For tissues with >100 hits, the top 10% were used; otherwise, all were retained. Directional agreement between deregulated miRNAs and mRNAs was defined as matching log2 fold-change signs.

Pathway analysis on mRNA and miRNA lists were performed as defined in Table 6.3.

All pathway p-values were BH-adjusted; significance was defined as adj. $p < 0.05$. Redundant GO terms were removed using `minimal_set` from the

Analysis	**Input**	**Tool**	Database
GSEA (miRNA)	Ranked by $\log_{10}(p) \cdot \text{sign}(\text{log2FC})$	`miEAA` (default) [409]	GO via miRTarBase [410, 411] & miRWalk [412]
ORA (miRNA)	Unranked miRNA sets	`miEAA` (default) [409]	GO via miRTarBase [410, 411] & miRWalk [412]
GSEA (mRNA)	Ranked list	`GeneTrail` [347] (see Section 6.3.2)	
ORA (mRNA)	Unranked mRNA sets	`enrichGO` `clusterProfiler` (version 4.6.0) [413]	GO-BP or full GO

TABLE 6.3: Pathway analysis for mRNA and miRNA data.

`OntologyIndex` package [414] with `org.Mm.eg.db` (version 3.16.0). Related terms were grouped with `get_term_descendancy_matrix` using ancestor relationships. Clustering was performed via the `GO_similarity` and `binary_cut` functions from the `simplifyEnrichment` package (version 1.8.0) [415]. Pathway clusters were summarized by most frequent words ($\geqslant$2 occurrences).

Predicted miRNA-mRNA target pairs were obtained for M. musculus from TargetScan (version 8.0) [416]. Only targets with a weighted context++ score above the 75th percentile were retained. We intersected these with deregulated miRNAs and their significantly deregulated mRNA targets to focus on likely functional interactions.

6.4 SINGLE-CELL ANALYSIS OF SPACEFLIGHT EFFECTS

The single-cell mRNA sequencing dataset included data from four spaceflight (Fl) and four HGC control mice, all part of the LAR group that returned to earth before being taken down. Each group included two 3-month-old and two 8-month-old mice. Of the 232 sequenced samples from 29 tissues, 16 did not pass the quality control.

The final dataset thus consisted of 216 samples from 28 tissues (Figure 6.3) and a total of 280,745 cells. The removed samples included all samples from skin, as key cell type markers were not expressed in the spaceflight group. The samples had between 406 and 30,963 cells each. SCAT had the most cells, with a median of 3,692 (sd = 2610.2), and skeletal stem cells had the fewest cells per sample, with a median of 67 (sd = 8.7). We saw differences in several tissues when comparing the number of cells in the different groups, i.e., in Fl

3-month, Fl 8-month, HGC 3-month, and HGC 8-months. To quantify these differences, we defined an imbalance score and used a threshold of 0.125 to identify tissues whose cell numbers show a considerable deviation from the expected number. Nine tissues were unbalanced across conditions: Non-myeloid forebrain, limb muscle, and thymus have fewer cells in spaceflight, and the large intestine, bone, BAT, mammary gland, and aorta have more cells. As imbalances in single-cell datasets can lead to differences in clustering and annotation, differential expression analysis, and trajectory inference [417], these should be considered when interpreting the results.

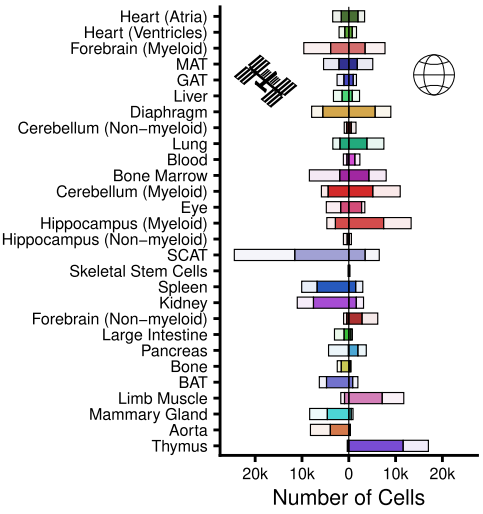


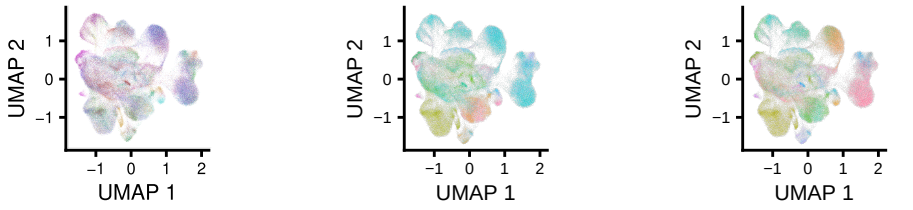
FIGURE 6.3: The number of cells from 28 tissue locations from the space group (on the left), and the control group (on the right).

CellOntology Name	**CellOntology ID**	**Cell type Cluster (fine)**	**Cell type Cluster (broad)**
tendon cell	CL:0000388	Fibroblast	Connective tissue cell
pericyte	CL:0000669	Pericyte	Connective tissue cell
capillary endothelial cell	CL:0002144	Endothelial cell	Endothelial cell
endothelial cell of artery	CL:1000413	Endothelial cell	Endothelial cell
terminal Schwann cell	CL:0000692	Schwann cell	Glia cell
myelinating Schwann cell	CL:0000218	Schwann cell	Glia cell
B cell	CL:0000236	B cell	Immune Cell
macrophage	CL:0000235	Macrophage	Immune Cell
mature T cell	CL:0002419	T cell	Immune Cell
smooth muscle cell	CL:0000192	Muscle cell	Muscle cell
slow muscle cell	CL:0000189	Muscle cell	Muscle cell
⋮	⋮	⋮	⋮

TABLE 6.4: Examples of the Cell Ontology classification of cell types in the diaphragm.

After expert manual annotation, we assigned each cell type a Cell Ontology ID and classified the cell types using Cell Ontology categories (Example see Table 6.4). We grouped cell types across tissues using these categories based on their primary function. The top-level groups cells into Bone tissue, Ciliated cell, Connective tissue cell, Dendritic cell, Endothelial cell, Epithelial

cell, Erythrocyte, Glia cell, Immune Cell, Muscle cell, Neuron, Plasma cell, Platelet, Stem/Progenitor cell.



(A) Tissue (B) Cell type Cluster (broad) (C) Cell type Cluster (fine)

FIGURE 6.4: UMAP embedding of the whole dataset, colored by tissue and cell types groups.

When grouping all cells by gene expression in a joint UMAP embedding, the cells cluster by cell type group and not by tissue (Figure 6.4). For example, 85.8% of cells in cluster 1 were endothelial cells, 92.2% of cells in cluster 3 were T cells and cluster 9 mainly contained monocytes (32.3%), granulocytes (36.6%), macrophages (12.4%) and neutrophils (9.8%) (Figure B.1). The clusters all contained cells across the different age and condition groups and tissues (Figure B.2). The cell types thus had a greater influence on gene expression than the tissue from which they came.

32.9% of cells in the dataset were immune cells, including circulating immune cells such as T cells (29,573 cells), B cells (24,009 cells) and thymocytes (11,974 cells), and tissue-resident immune cells such as macrophages (4,234 cells) (Figure 6.5). The second most represented cell type group were glial cells (21.0%), such as microglia (54,188 cells) from the brain, followed by endothelial cells with 15.7% of cells. Of note, even though microglia are tissue-resident immune cells, they were classified as glial cells in this dataset.

6.4.1 *Impact of spaceflight and age on the gene-expression profile*

In order to identify which tissues and cell types are most affected by spaceflight and to what extent age influences the gene expression profiles, we performed a PVCA analysis to calculate the amount of variance explained by spaceflight and age.

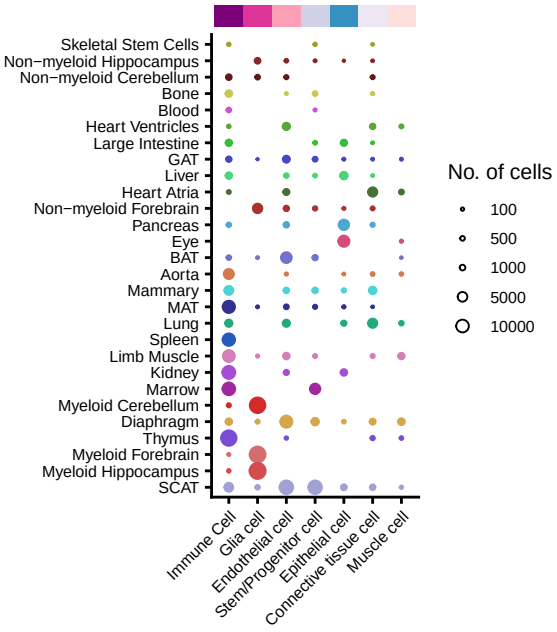
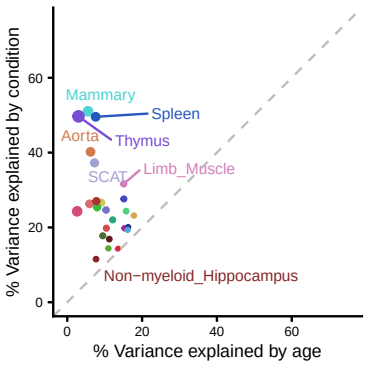


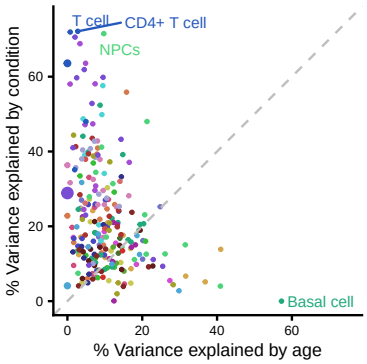
FIGURE 6.5: Number of cells in the main Cell Ontology groups in the different organs.



Tissue	% Age	% Space
Mammary	5.54	**51.03**
Thymus	2.99	**49.68**
Spleen	7.59	**49.56**
Aorta	6.22	**40.19**
SCAT	7.29	**37.25**
Diaphragm	**17.80**	23.16
MAT	**16.27**	20.08
Pancreas	**16.15**	19.32
Liver	**15.74**	24.35
Marrow	**15.29**	19.77

FIGURE 6.6 & TABLE 6.5: PVCA values for Condition (space/control) and Age (3/8 months) in the different tissues. The colors represent the tissues, and the dot sizes represent the ratio of condition-explained variance to age-explained variance. The table shows the top values in space (top) and in age (bottom).

On the tissue level, spaceflight always resulted in a higher percentage of variance explained than age (Figure 6.6). It showed the strongest influence on the mammary, thymus, and spleen, whereas age showed the strongest influence on the diaphragm, MAT, and pancreas.



Cell type	% Age	% Space
CD4+ T cell (spleen)	2.76	**72.12**
T cell (spleen)	0.74	71.92
NPCs (liver)	9.73	71.49
T cell (thymus)	2.04	70.56
Leukocyte (kidney)	3.36	68.82
Basal cell (lung)	**57.22**	< 0.01
B cell (diaphragm)	40.97	13.85
Dendritic cell (liver)	40.89	4.01
Macrophage (diaphragm)	36.83	5.15
Mesothelial cell (diaphragm)	31.62	6.68

FIGURE 6.7 & TABLE 6.6: PVCA values in the different cell types as in Figure 6.6

Similarly, it explained a higher proportion of variance on the cell type level than age in 191 of 248 cell types (Figure 6.7). T cells such as the CD4+ T cells, the general T cells in the spleen, and the T cells in the thymus showed a particularly high variance explained by spaceflight.

To see if there are age-specific changes in the spaceflight effect or if spaceflight changes the age differences, we performed a correlation analysis on fold changes. We expect a high correlation for gene expression changes that affect the same genes and pathways.

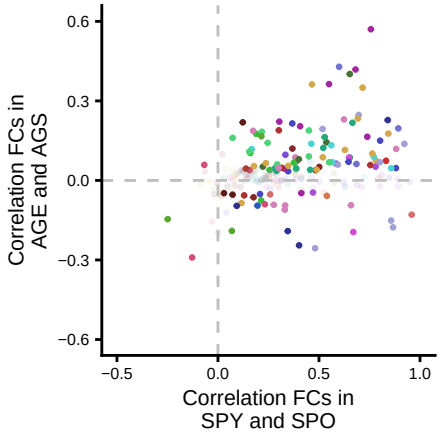


FIGURE 6.8: Correlation of the $\log_2$ fold-changes in FL vs. HGC in 3M (SPY) and FL vs. HGC in 8M (SPO) and of the fold-changes in 8M vs. 3M in HGC (AGE) and in 8M vs. 3M in FL (AGS). Dots are colored by tissue, and pale dots are not significant (adj. $p > 0.05$) in both comparisons.

We found a positive correlation (adj. p-value < 0.05) of fold-changes in spaceflight in 3-month (FL vs. HGC in 3M) and 8-month mice (FL vs. HGC in 8M) (Figure 6.8) for 202 out of 212 cell types. In contrast, in the age-related comparisons, only 87 cell types showed positive correlation values between the earth (8M vs. 3M in HGC) and spaceflight (8M vs. 3M in FL) group.

	Tissue	Cell type	correlation	adj. p-value
Strongest	Forebrain (Myeloid)	microglial cell	0.95	$< 10^{-23}$
	Kidney	kidney collecting duct epithelial cell	0.95	$< 10^{-23}$
	SCAT	mesenchymal stem cell	0.92	$< 10^{-23}$
Lowest	Heart (Ventricles)	Immune Cell	-0.24	$2.49 \cdot 10^{-40}$
	Eye	retinal pigment epithelial cell	-0.12	$3.45 \cdot 10^{-11}$
	Diaphragm	slow muscle cell	-0.08	$6.00 \cdot 10^{-6}$

TABLE 6.7: Top three and bottom three cell types by the correlation between spaceflight-effect on 3- and 8-month mice.

In the comparison between the spaceflight-effect on 3-month (FL vs. HGC in 3M) and 8-month mice (FL vs. HGC in 8M), microglia cells from the myeloid forebrain, kidney collecting duct epithelial cell, and mesenchymal stem cell from SCAT showed the highest correlations (Table 6.7). Of note, nine of the top 20 cell types with the highest correlation came from adipose tissues (BAT, GAT, MAT, and SCAT).

	Tissue	Cell type	correlation	adj. p-value
Strongest	Bone Marrow	erythroblast	0.57	$3.41 \cdot 10^{-84}$
	BAT	Macrophages/Monocytes	0.42	$4.92 \cdot 10^{-199}$
	Bone Marrow	proerythroblast	0.41	$9.03 \cdot 10^{-152}$
Lowest	Eye	retinal pigment epithelial cell	-0.29	$1.51 \cdot 10^{-50}$
	SCAT	pericyte	-0.25	$3.26 \cdot 10^{-48}$
	MAT	endothelial cell	-0.24	$7.28 \cdot 10^{-48}$

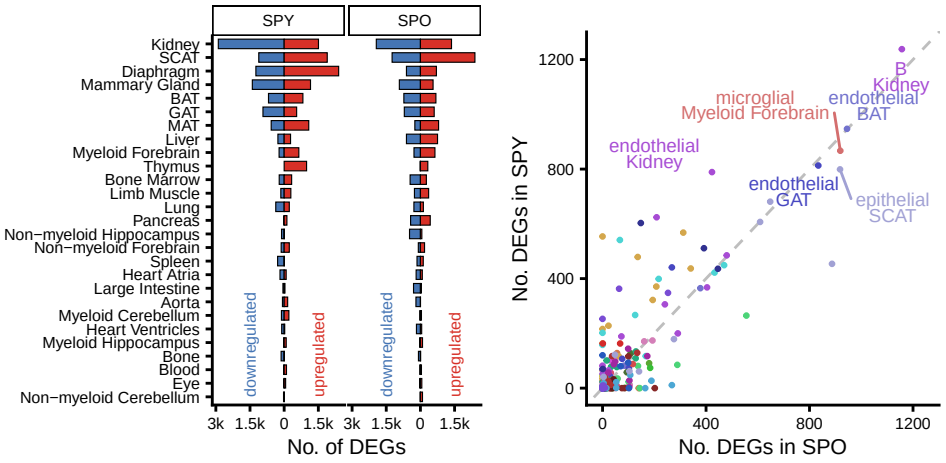
TABLE 6.8: Top three and bottom three cell types by correlation between the age-related comparisons in the earth and the spaceflight group.

The age-related comparisons in the earth (8M vs. 3M in HGC) and spaceflight (8M vs. 3M in FL) groups showed the highest correlation in erythroblasts from bone marrow, macrophages/monocytes from BAT, and proerythroblasts from bone marrow and the lowest correlation in retinal pigment epithelial cell and pericyte from SCAT (Table 6.8).

These results suggest that spaceflight causes similar changes in the mice transcriptome in both age groups but that age has a different impact on control and flight mice.

6.4.2 *Systemic spaceflight effects*

We thus analyzed the deregulated genes across tissues in more detail to identify how the different factors influence gene expression and how many genes are affected.



(A) Sum of DEGs for all cell types in each tissue.

(B) Number of DEGs per cell type.

FIGURE 6.9: Number of DEGs in the FL vs. HGC in 3M (SPY) and FL vs. HGC in 8M (SPO) comparisons.

When looking at the effect of spaceflight on the different age groups (FL vs. HGC in 3M and FL vs. HGC in 8M), the two comparisons displayed similar numbers of significantly up- and down-regulated genes (abs $\log_2$FC > 0.5 & adj. p < 0.05). Overall, the kidney (4399 and 3302 genes), SCAT (3006 and 3636), diaphragm (3635 and 1315), and mammary gland (2555 and 1480) had the most deregulated genes in both comparisons (Figure 6.9a). When looking at the cell types, B cells from the kidney, endothelial cells in BAT, and microglial cells in the brain (myeloid forebrain) had the most deregulated genes. Endothelial and epithelial cells like the endothelial cells from BAT,

GAT, and SCAT, and the kidney, as well as epithelial cells from SCAT and kidney collecting duct epithelial cells, were prominent among the top 10 cell types (Figure 6.9b).

Observing these similar numbers of de-deregulated genes and the positive correlation indicates that 3—and 8-month-old mice are affected by the same amount and similar changes in their gene expression profiles. To study these age-independent spaceflight effects, we first identified the systemic changes that can be observed across most tissues and cell types.

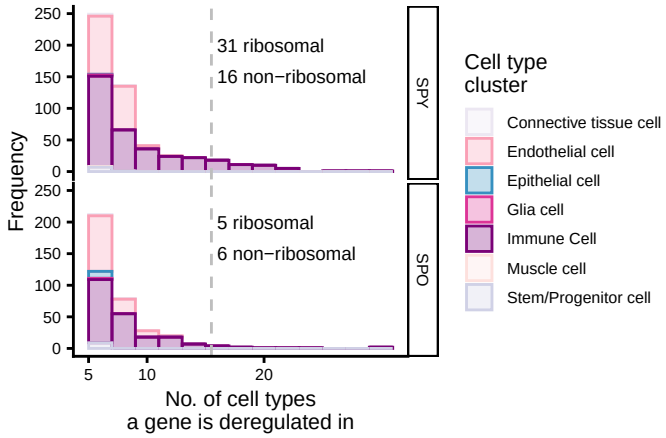


FIGURE 6.10: Frequency of deregulation per gene (for genes in >5 cell types) for FL vs. HGC in 3M (SPY) and FL vs. HGC in 8M (SPO), split by Cell Ontology group.

We found frequent deregulation of both ribosomal protein and non-ribosomal genes across multiple cell types and tissues. Within the Cell Ontology groups, immune and endothelial cells displayed the most systematically deregulated genes (Figure 6.10). Based on the frequency of deregulation, the top non-ribosomal genes were *AY036118*, *Ahnak*, *Gm42418*, *Tpt1*, *Lars2*, and *Malat1*. *AY036118*, *Ahnak*, and *TPT1* are involved in regulating cell growth and proliferation [418, 419], while *Lars2* is a tRNA synthetase, and *Malat1* has a role in gene regulation [420]. Ribosomal protein genes comprise a substantial part of the systemically deregulated genes. 31 of the 44 most frequently deregulated genes (>15 cell types) in 3-month-old and 5 out of 11 genes in 8-month-old mice were ribosomal protein genes. The ribosomal protein genes that were most commonly deregulated in young adult mice included *Rpl9*,

Rps29, *Rps20*, *Rps7*, and *Rps23*, while mature mice showed frequent deregulation of *Rps29*, *Rps28*, and *Rpl38*. The systemic deregulation of ribosomal protein genes indicates possible stress responses, metabolic changes, and modified cellular signaling, which may be caused by microgravity or increased radiation exposure.

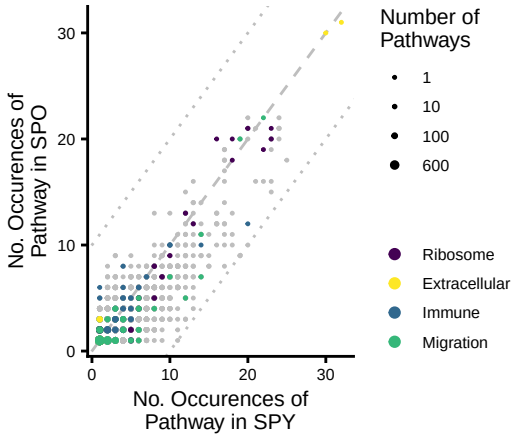


FIGURE 6.11: Frequencies of pathways in FL vs. HGC in 3M (SPY) and FL vs. HGC in 8M (SPO).

A pathway enrichment analysis (GSEA) supports the age-independent systemic change in gene expression after spaceflight (Figure 6.11). The pathway frequency (number of cell types in which a pathway was enriched or depleted) was similar between 3-month-old and 8-month-old mice, with a difference of no more than ten, suggesting a widespread, largely age-independent effect.

In both age groups, the most frequently affected pathways were associated with the extracellular matrix (e.g., "extracellular space", "extracellular matrix") (Figure 6.12), reflecting alterations in cell and tissue structure. Other frequently affected pathways were related to membrane functions (e.g., "plasma membrane"), the exon junction complex (e.g., "Nonsense-mediated decay independent of the exon junction complex"), the ribosomes (e.g., "GTP hydrolysis and joining of 60S ribosomal subunits") and general metabolic processes (e.g., "peptide biosynthetic process"). Immune, endothelial, and epithelial cells across various tissues were most frequently involved.

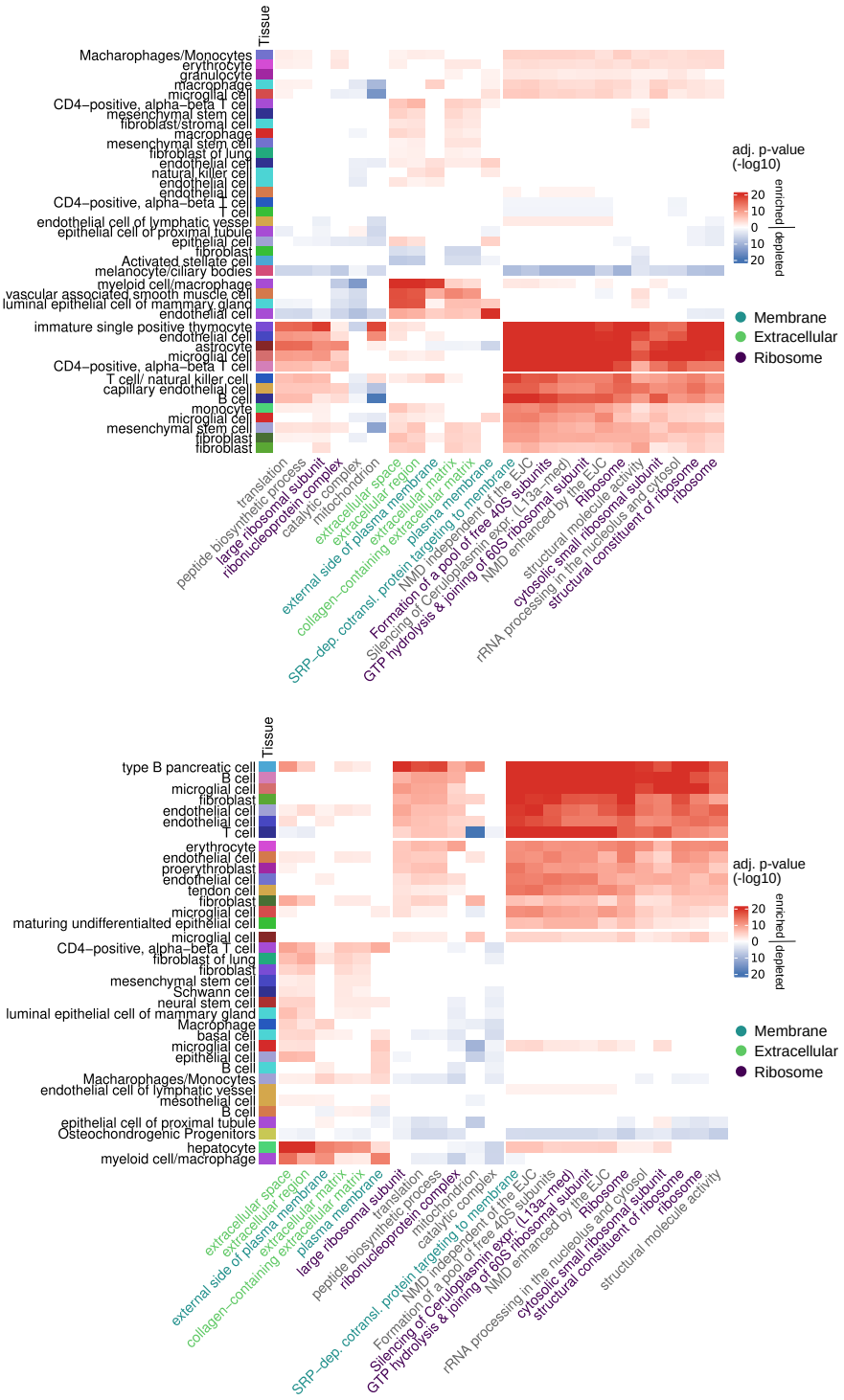
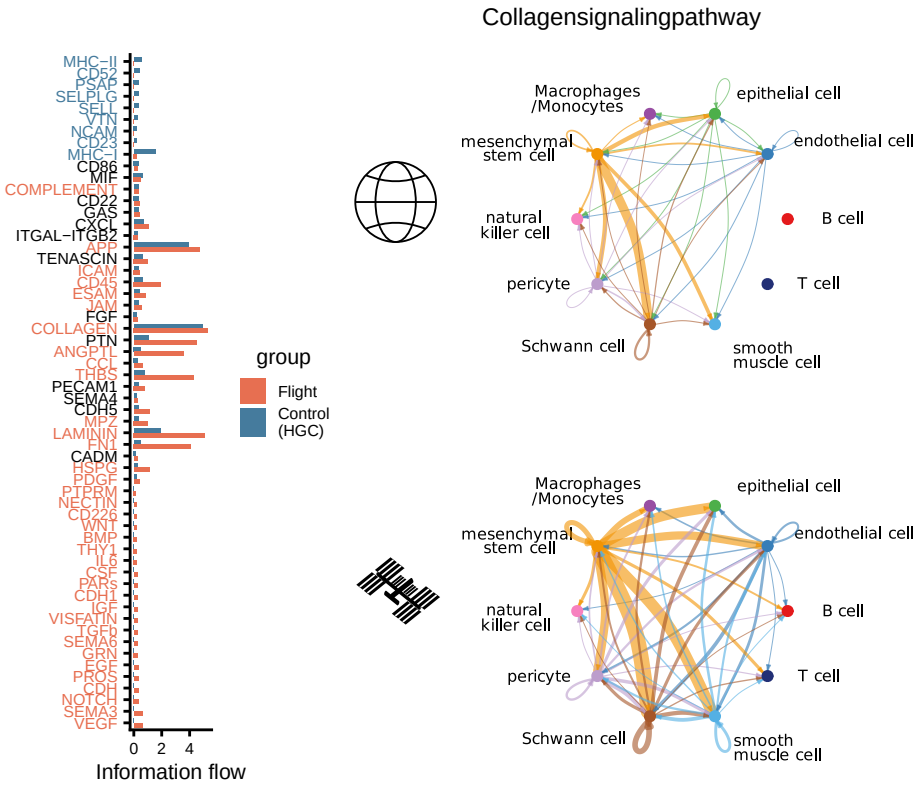


FIGURE 6.12: Most frequently enriched and depleted pathways (>3 cell types) within the most frequent cell types for FL vs. HGC in 3M (top) and FL vs. HGC in 8M (bottom).



(A) Affected signaling pathways. (B) Change in the COLLAGEN pathway.

FIGURE 6.13: Cell-Cell Communication analysis for SCAT in FL vs. HGC.

Given the strong correlation of the spaceflight effects observed in adipose tissues, we performed a cell-cell communication analysis in SCAT (Figure 6.13a). We observed changes in signaling pathways related to immune responses, tissue remodeling, and metabolism. Immune-related pathways, such as MHC-I, MHC-II, Cd23, Cd52, Sell, and Selplg, showed reduced activity, while increased expression of Tgfb, Sema6, and Grn further indicates immune modulation. Downregulated pathways involved in tissue remodeling and cell adhesion include Vtn, Ncam, Klk, and Plau. In contrast, pathways related to wound healing and structural adaptation, like Vegf, Sema3, and Notch, exhibit increased activation, suggesting ongoing tissue remodeling. Notably, increased Vegf and Sema3 signaling indicates changes in vascular development. At the same time, upregulation of Cldn, Cholesterol, and Cdh

reflects adjustments in tight junctions and cell adhesion, with growth factors like Vegf, Egf, and Tgfb promoting cell proliferation and differentiation. Consistent with this shift, mice exposed to spaceflight showed increased activity in signaling pathways related to Pecam, Laminin, Icam, Collagen, and Thbs compared to controls, suggesting ongoing tissue remodeling (Figure 6.13b,B.3). Additionally, the downregulation of Psap, a key regulator of lipid metabolism, indicates potential metabolic disruptions affecting energy balance and lipid processing in adipose tissue.

These results suggest that spaceflight may alter the organization of extracellular structures and affect tissue remodeling, which could be associated with the muscle and bone loss [421], as well as the blood coagulation and cardiovascular issues observed in astronauts [401, 422]. To investigate further how these systemic changes relate to specific changes in tissues, we identified tissue and cell type-specific gene expression patterns.

6.4.3 *Cell type-specific changes in spaceflight*

To differentiate systemic from specific gene expression changes in response to spaceflight, we classified the deregulated genes into four categories: cell type-specific, cell type cluster-specific, tissue-specific, and non-specific (Figure B.4). Genes deregulated in one cell type and one tissue across the whole dataset were considered cell type-specific. If a gene appears in multiple cell types belonging to the same cluster and the gene is not deregulated in any other cell type belonging to a different cell type group, it is considered cell type cluster-specific. Genes deregulated in multiple cell types but only in one tissue are called tissue-specific. Among the 3,510 deregulated genes, 1,718 were cell type-, 251 cluster-, and 260 tissue-specific; the rest were non-specific.

The kidney (131 genes), mammary gland (60 genes), and SCAT (45 genes) had the highest number of tissue-specific genes (Figure 6.14). Other tissues with specific gene sets included MAT, GAT, BAT, limb muscle, non-myeloid forebrain, lung, pancreas, and bone marrow. In terms of cell types, immune cells had the most deregulated genes (107), followed by endothelial (99), stem/progenitor (17), epithelial (14), and glial cells (10) (Figure 6.14).

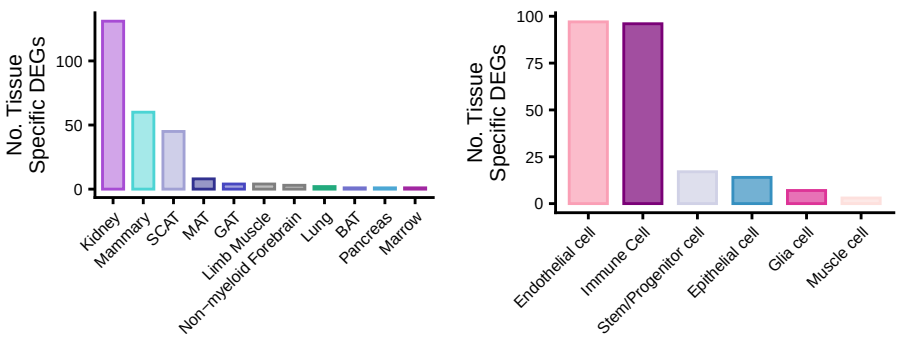


FIGURE 6.14: Number of tissue- and cell type cluster-specific genes across groups.

Of note, there were more cell type-specific genes than those specific to cell type clusters or tissues, indicating that many genes are unique to specific cell types and tissues.

Tissue-specific changes could be caused by the functions of specific cell type clusters, the diversity of cell types, or general differences between tissues. We thus examined the tissue-specific gene sets more closely (Figure B.5). Although, by definition, each gene was deregulated in just one tissue, we noticed similar patterns across different tissues. These included ECM remodeling, immune regulation, and metabolism. For instance, genes involved in ECM remodeling included *Adamtsl1* in SCAT, *Apbb1ip* in the kidney, and *Acp5* in the kidney. Immune regulation was evident in SCAT (*Ccr7*, *Cd24a*, *C1s1*), bone marrow (*S100a8*), and lung (*Icam1*). Changes in metabolism were observed in lipid transport in the mammary gland (*Adh1*, *Akr1c12*, *Atp10b*, *Atp2a3*), glucose metabolism in the pancreas (*Gcg*), and energy homeostasis in the non-myeloid forebrain (*Mcr1*). These findings indicate that most changes were likely driven by cell type or cell type cluster-specific gene expression, leading us to investigate the cell type-specific changes during spaceflight further.

To explore the cell type cluster-specific changes, we performed an Over-Representation Analysis (ORA) to explore their biological functions (Figure 6.15 6.16,B.6). Among the top five genes in each category, we discovered genes associated with immune regulation in the immune and glial cell clusters and changes related to the cytoskeleton and extracellular matrix in endothelial, epithelial, stem/progenitor, and muscle cells.

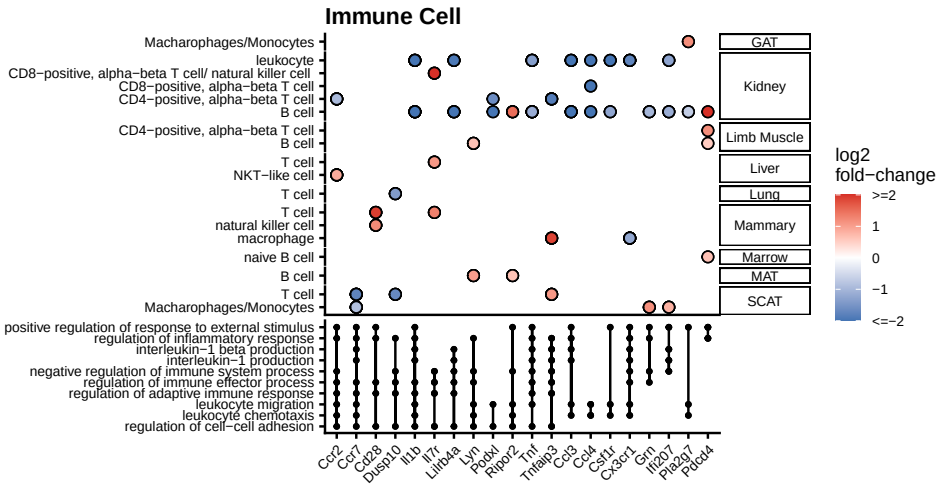


FIGURE 6.15: ORA pathway analysis of the Immune cell-specific genes and the deregulation the top 20 involved genes.

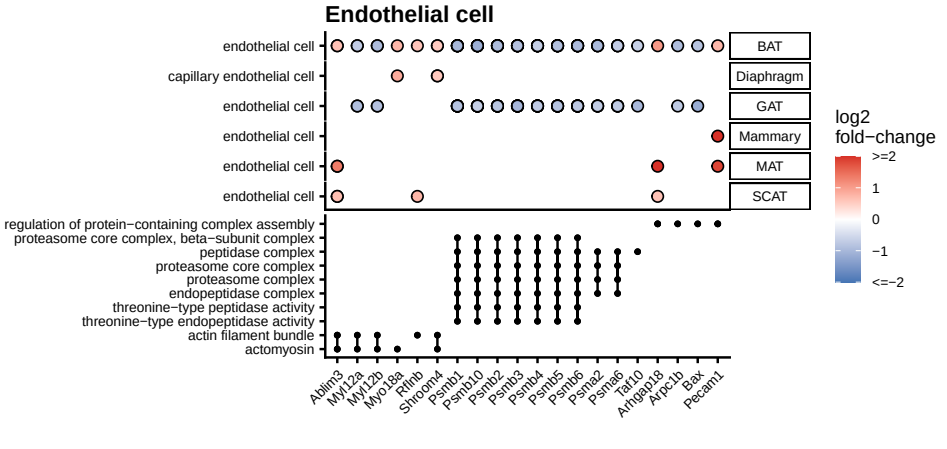


FIGURE 6.16: ORA pathway analysis of the Endothelial cell-specific genes and the deregulation the top 20 involved genes.

Transcription factors such as Fosl2 and Pdcd4 were noted within the immune cells, while glial cells exhibited genes like Akna and Lag3, contributing to immune regulation and neural development. Additionally, platelets deregulated Pf4, which is involved in cytokine and chemokine signaling. The pathway analysis further supports these findings, revealing an enrichment for immune activation and leukocyte migration, including terms like

"leukocyte chemotaxis" and "regulation of inflammatory response," as well as interleukin production (e.g., "interleukin-1 production") in immune cells.

Genes related to tissue organization and cell adhesion were identified in endothelial and epithelial cells, such as *Ablim3* and *Arhgap18* in endothelial cells and *Alcam* and *Aldh1a1* in epithelial cells. Meanwhile, stem/progenitor cells displayed genes like *Pdgfra* and *Dapk1*, critical for cell signaling and apoptosis regulation. Pathways related to cytoskeleton or extracellular matrix can be observed in endothelial cells ("actin filament bundles", "actomyosin"), Stem/Progenitor cells ("extracellular matrix organization", "collagen binding"), and muscle cells ("actin binding"). Muscle cells additionally show deregulation of *Dmd* and *Tpm1*, genes critical for muscle fiber integrity and contraction. While these findings explain some gene expression changes, they highlight the need to connect them to the physiological changes and diseases astronauts face.

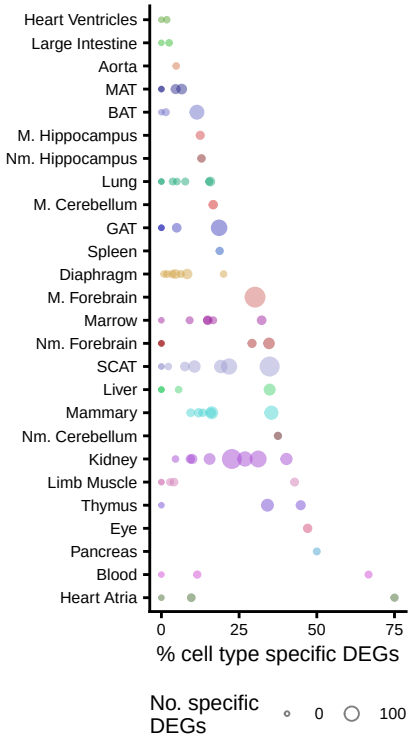


FIGURE 6.17: Ratio cell type-specific to all deregulated genes in each cell type. Labels mark cell types with $>$ 30% cell type-specific genes and $<$ 10 genes.

With 1718 genes, most deregulated genes were cell type-specific, only occurring in one cell type cluster and tissue. Focusing on changes mainly driven by cell type-specific genes, we selected cell types with fewer than ten cell type-specific genes but representing over 30% of their deregulated genes (Figure 6.17). Cardiac muscle cells showed deregulation of genes such as *Actc1* and *Myl* (related to muscle contractions) and *Ankrd1* (involved in muscular stress signaling). Likewise, fast muscle cells found in the Limb Muscle exhibited deregulated expression of *Neb*, *Mybpc2*, *Tcap*, and *Cmya5*, which are crucial for muscle structural integrity and sarcomere organization, along with *Jph1*, which connects the sarcoplasmic reticulum to the plasma membrane. Type B pancreatic cells displayed deregulation of *Pcsk2* and *Sst*, impacting the regulation and processing of hormones like insulin. Additionally, retinal rod cells presented altered expression of *Cited2*, *Hmgn1*, and *Nr2e3* (linked to

gene regulation), *Plekhb1* (involved in signal transduction), as well as *Rcvrn* and *Crybb2* (associated with phototransduction and vision). These specific alterations in cell types might contribute to health issues experienced by astronauts, such as Spaceflight-Associated Neuro-Ocular Syndrome [423], insulin resistance [424], muscle atrophy [425], heart atrophy [401], and irregular cardiac rhythms [426]. Gaining insights into these changes is crucial for assessing the effects of spaceflight on mammals.

6.4.4 *Age-dependent deregulation of genes in spaceflight*

Most genes exhibited deregulation in a consistent direction between spaceflight in 3-month- (FL vs. HGC in 3M) and 8-month-old mice (FL vs. HGC in 8M) (Figure 6.18). However, 27 non-ribosomal genes and three ribosomal protein genes (Rps12, Rps27, Rps28) showed an age-dependent deregulation.

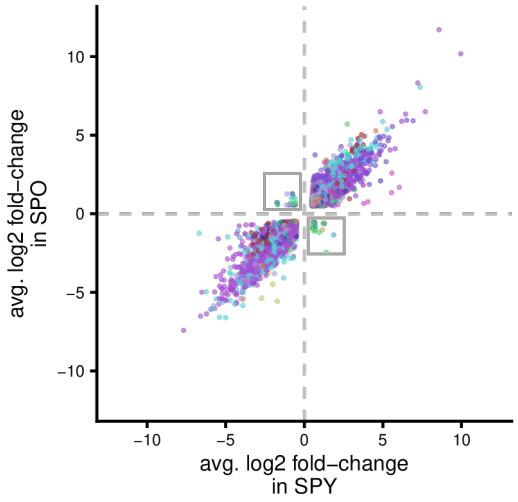


FIGURE 6.18: Comparison of fold-changes in FL vs. HGC in 3M (SPY) and FL vs. HGC in 8M (SPO).

The primary age-dependent effect observed in spaceflight was the deregulation of the immune system, found in both immune and non-immune cells. For example, *Tyropb* expression increased in macrophages and monocytes in SCAT from 8-months- but decreased in 3-month-old mice. In contrast, the bone marrow showed the opposite trend for *Oas3* in monocytes and *Ifi27l2a*

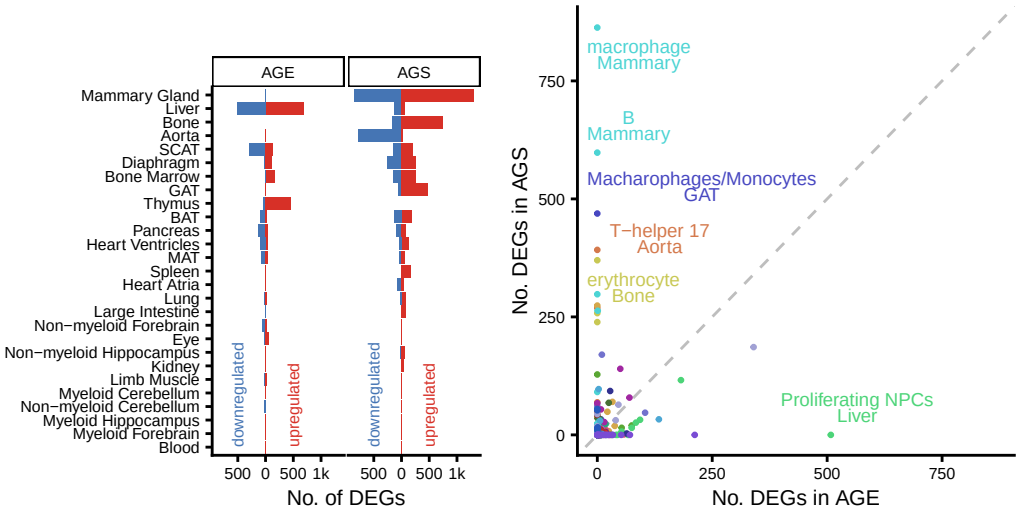
gene	**celltype**	**Tissue**	**avg log2FC** 3M	**avg log2FC 8M**
Gm26917	mesenchymal stem cell	BAT	-1.75	0.72
Hspa1a	tendon cell	Diaphragm	1.15	-1.03
Rhob	mesenchymal stem cell	Diaphragm	0.67	-0.97
Cd163	Macharophages/Monocytes	GAT	-0.75	1.28
Cidec	endothelial cell	GAT	-0.95	0.56
Ifi27l2a	fibroblast	Heart Atria	0.52	-0.91
mt-Nd4	endothelial cell of vascular tree	Heart Ventricles	-0.59	0.56
Bhmt	hepatocyte	Liver	-0.71	1.07
Scd1	hepatocyte	Liver	1.40	-2.45
Scp2	hepatocyte	Liver	-0.55	0.73
Sod1	hepatocyte	Liver	1.22	-0.62
Tdo2	hepatocyte	Liver	-0.88	0.67
Ubb	hepatocyte	Liver	-0.69	0.65
Uba52	T cell	Lung	-1.69	0.58
Zbtb16	myofibroblast cell	Lung	0.71	-0.80
Mknk2	natural killer cell	Mammary	-0.82	1.00
Pmepa1	natural killer cell	Mammary	-0.60	0.99
Ifi27l2a	granulocyte	Marrow	1.27	-0.62
Oas3	monocyte	Marrow	0.56	-0.52
Gm26917	T cell	MAT	-0.72	0.56
mt-Atp8	T cell	MAT	0.67	-0.56
Iapp	type B pancreatic cell	Pancreas	-1.72	0.72
Lars2	type B pancreatic cell	Pancreas	1.87	-1.35
Car3	endothelial cell	SCAT	0.57	-0.73
H2-K1	epithelial cell	SCAT	-1.13	0.56
Rarres2	endothelial cell	SCAT	-1.10	1.26
Rnase4	endothelial cell	SCAT	-0.72	0.67
Tyrobp	Macharophages/Monocytes	SCAT	-0.92	0.61

TABLE 6.9: DEGs in (FL vs. HGC in 3M) and (FL vs. HGC in 8M) with different fold-changes directions ($\mathtt{sign}(\mathrm{FC})$).

in granulocytes. Similarly, GAT macrophages and monocytes showed deregulation of *Cd163*, which is related to inflammation, and epithelial cells in SCAT and heart fibroblasts also showed immune-related changes, indicating systemic effects. These findings suggest that age affects immune regulation, possibly due to changes in endothelial and epithelial cells. Other age-dependent effects of spaceflight included stress response, metabolism, mitochondrial function, and tissue homeostasis. Most ribosomal protein genes showed similar deregulation in both spaceflight conditions, except the liver, where 8-month mice had decreased *Rps27* and *Rps28* expression while 3-month mice showed the opposite.

6.4.5 *Age-related differences in space reveal systematic changes in the immune system*

Following up on these age-dependent signals, we decided to study the transcriptome changes related to age on earth (8M vs. 3M in HGC) and in space (8M vs. 3M in FL).



(A) Sum of DEGs for all cell types in each tissue.

(B) Number of DEGs per cell type.

FIGURE 6.19: Number of DEGs in the 8M vs. 3M in HGC (AGE) and 8M vs. 3M in FL (AGS) comparisons.

In contrast to Fl vs. HGC in 3M and in 8M, which had a median of 386 deregulated genes (sd = 1043.4), both 8M vs. 3M in FL and in HGC showed a lower number of DEGs with a median of 79 (sd = 373, Figure 6.19a). For the age differences on earth, the most deregulated genes were found in the liver (1198 genes), thymus (512), and SCAT (432). In spaceflight, the mammary gland showed the highest number of affected genes (2165), followed by bone (913) and aorta (813).

Four of the five clusters with the most DEGs in 8M vs. 3M in FL (abs $\log_2$FC $>$ 0.5 & adj. p $<$ 0.05) were related to immune cells, including macrophages/monocytes from the mammary gland and GAT, as well as B cells from the mammary gland and T cells from the aorta (Figure 6.19b).

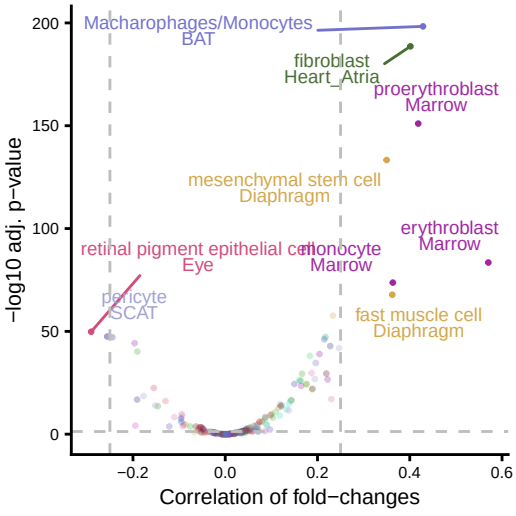


FIGURE 6.20: Correlation of $\log_2$ fold-changes between 8M vs. 3M in HGC and in FL. Labels show cell type and tissue; non-significant correlations (p> 0.05) are shown in pale color.

To differentiate between age-related changes on earth and those after a spaceflight, we conducted a correlation analysis of fold changes of 8M vs. 3M in FL and HGC. The comparisons revealed similar gene expression changes in only seven cell types across four tissues (Figure 6.20). Notably, three of the nine most correlated cell types came from the bone marrow, with correlation values of 0.57 (erythroblast, adj. $p = 3.41 \cdot 10^{-84}$), 0.41 (proerythroblast, adj. $p = 9.03 \cdot 10^{-152}$), and 0.36 (monocyte, adj. $p = 1.91 \cdot 10^{-74}$). The correlation never exceeded 0.6, indicating that spaceflight impacts age-related changes, even in those cell types that show a positive correlation.

When performing an RNA velocity analysis, changes in the bone marrow indicated a disturbance in the differentiation of granulocytes and monocytes (Figure 6.21), especially in older space mice. On earth, we observed a decrease in the intensity of velocities for mice from 3 to 8 months. Notably, the RNA velocity patterns observed in 8-month-old HGC mice were similar to those found in 3-month-old flight mice, indicating that spaceflight may influence immune aging processes.

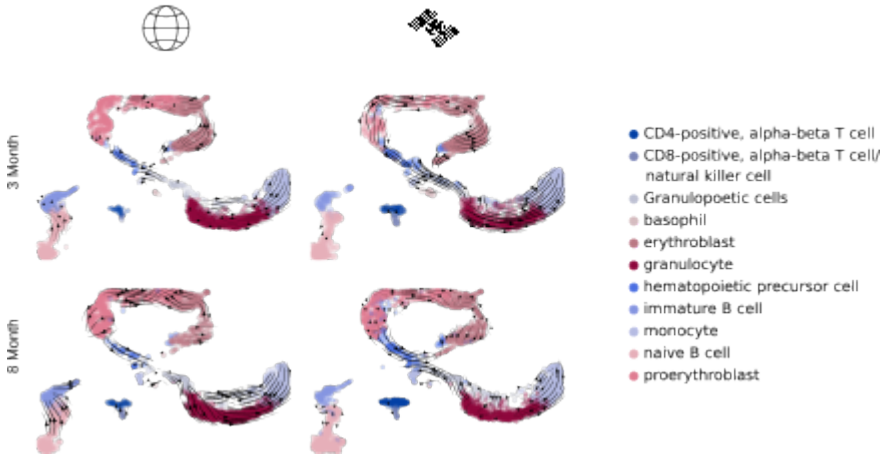


FIGURE 6.21: RNA velocity analysis in bone marrow.

Supporting this, cell-cell communication (Figure 6.22) showed differences in age-related changes between space and earth, suggesting that younger mice exhibited greater information flow in immune-related pathways such as MHC-II, SEMA4, BTLA, VISFATIN, CD6, and ALCAM.

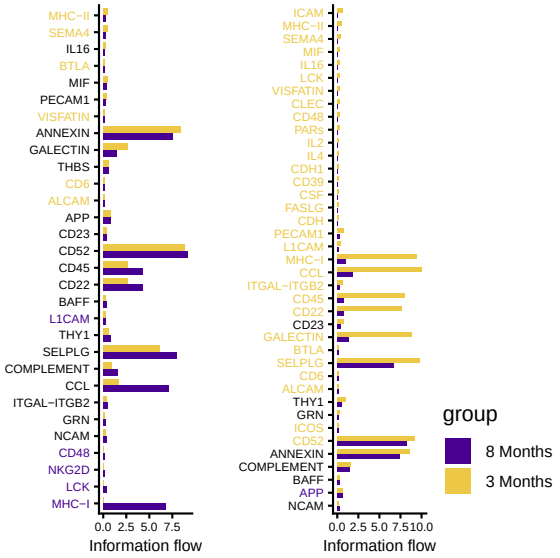


FIGURE 6.22: Cell-cell communication analysis of bone marrow for HGC (left) and FL mice (right).

Multiple signaling pathways showed changes unique to spaceflight (Figure 6.22). For instance, cell adhesion (ICAM) and immune signaling (CLEC),

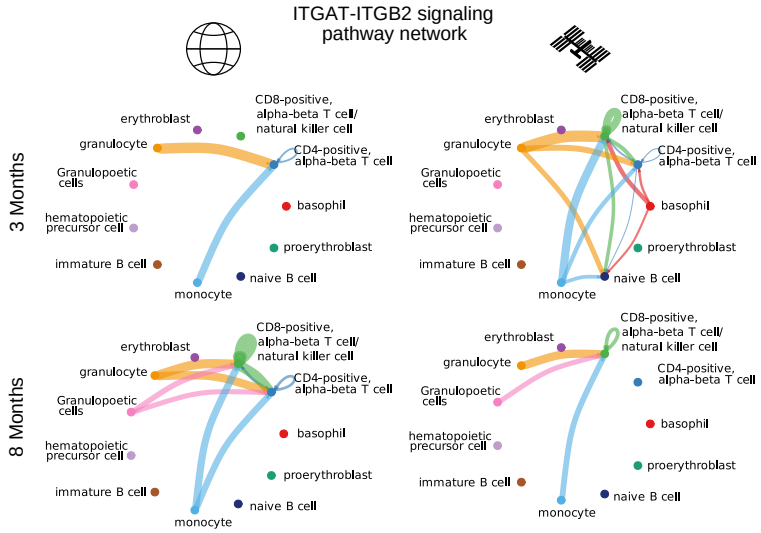


FIGURE 6.23: ITGAT-ITGB2 signaling pathway in bone marrow.

along with markers for inflammation and tissue repair (PARs) and immune response (IL2 and IL4), were elevated in 3-month spaceflight mice. LCK and CD48 showed stronger information flow in these mice, while 8-month control mice showed a similar increase. In 8-month spaceflight mice, increased APP (Amyloid Precursor Protein) indicated a shift in cellular activities related to neural or stress responses. Contrarily, the ITGAT-ITGB2 signaling pathways showed the opposite pattern stronger; the control group of 8-month-old mice and 3-month-old mice in space showed stronger signaling (Figure 6.23). This suggests that extracellular matrix variations could affect immune cells. These observations led us to a broader investigation into age-related differences among various cell types and tissues in space and on earth.

A GSEA pathway analysis showed the "extracellular region" and "extracellular space" pathways as most frequently affected across the cell types and tissues in both 8M vs. 3M in FL and in HGC (Figure 6.24). Unlike the FL vs. HGC comparisons, where pathways appeared with similar frequency in both age groups, the "regulation of locomotion," "regulation of developmental process," "regulation of cellular component movement", "cell adhesion", "cell surface" and "immune system process" pathways occurred more often in 8M vs. 3M in FL than in 8M vs. 3M in HGC, i.e., in at least 10 additional

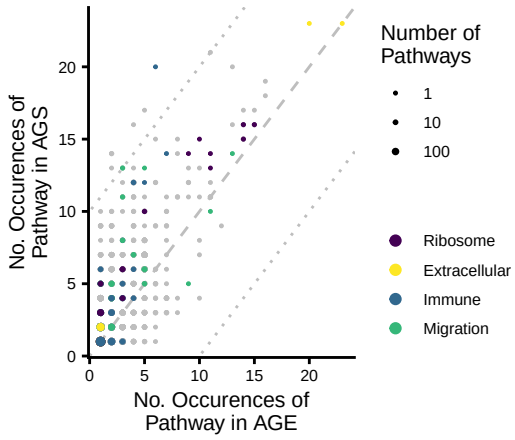


FIGURE 6.24: Frequency of pathways in 8M vs. 3M in FL (AGS) and 8M vs. 3M in HGC (AGE).

cell types. Within the most enriched and depleted pathways, extracellular matrix (e.g., "extracellular matrix", "extracellular space"), ribosomes (e.g., "ribosome", "large ribosomal subunit"), and membranes (e.g., "plasma membrane", "integral component of membrane") appeared both in space and on earth (Figure 6.25). In contrast to that, the immune regulation pathways ("immune system regulation" and "immune system"), among the top enriched and depleted pathways in 8M vs. 3M in FL, showed no general enrichment in 8M vs. 3M in HGC. These pathways were specific to certain cell types; for example, extracellular matrix and region pathways were deregulated in fibroblasts and mesenchymal stem cells, while immune regulation pathways occurred in endothelial cells, B cells, and macrophages.

We calculated specificity scores for both within-cell type and within-tissue similarities to determine whether the observed changes were more consistent within a particular tissue or among cell type groups. We mainly observed higher similarities within cell types in immune cells such as T cells, NK cells, macrophages/monocytes, and B cells, as well as in endothelial cells and mesenchymal stem cells (Figure 6.26). Clusters belonging to the same cell type group showed strong similarities in gene expression changes (measured by cosine similarity) across different tissues, suggesting a general shift within the immune system (Figure B.7, B.8). Additionally, the diaphragm,

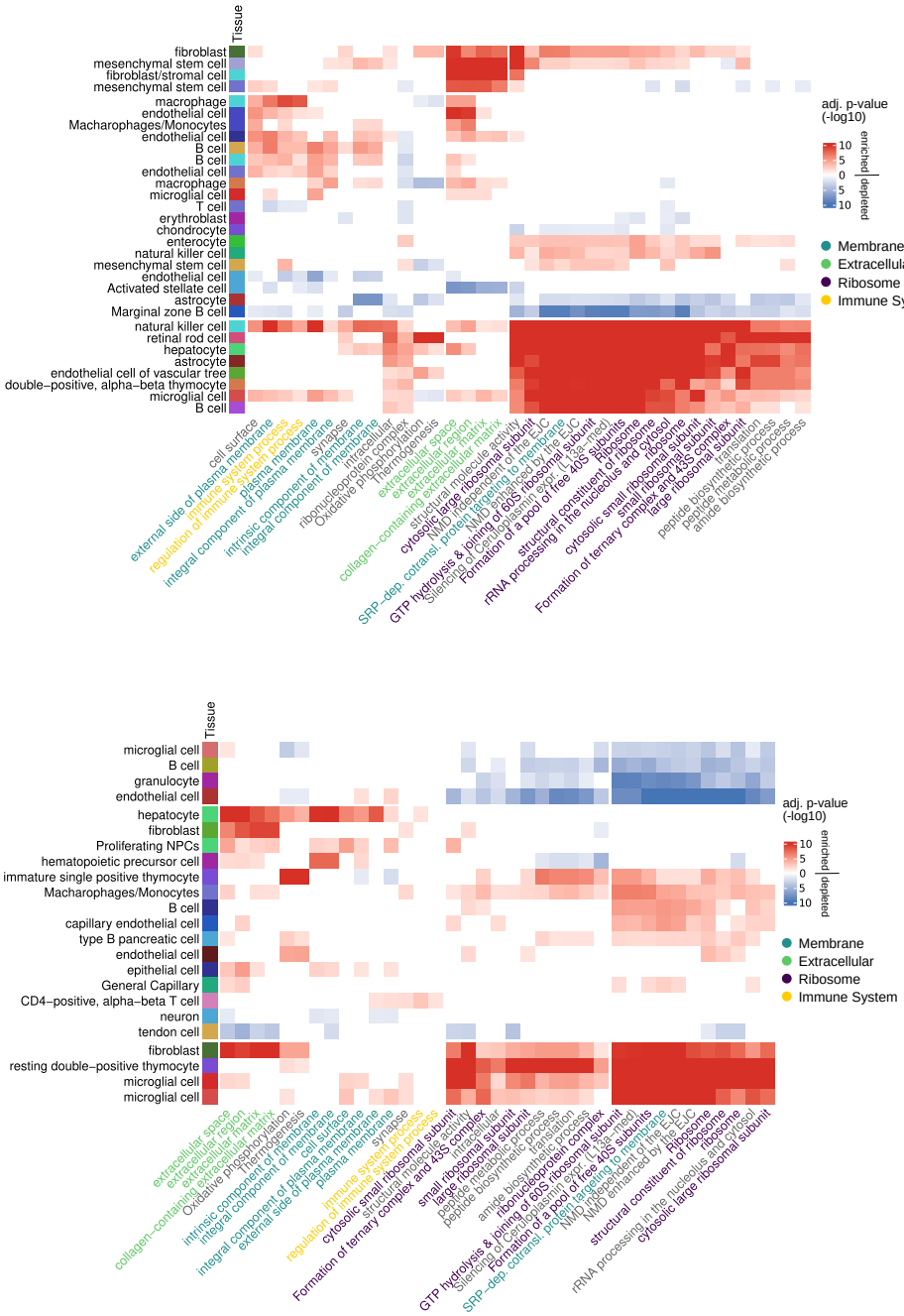


FIGURE 6.25: Most frequently enriched and depleted pathways within the most frequent cell types for 8M vs. 3M in FL (top) and 8M vs. 3M in HGC (bottom).

SCAT, and mammary gland showed tissue-specific changes. Therefore, we decided to analyze the gene expression patterns within these specific groups.

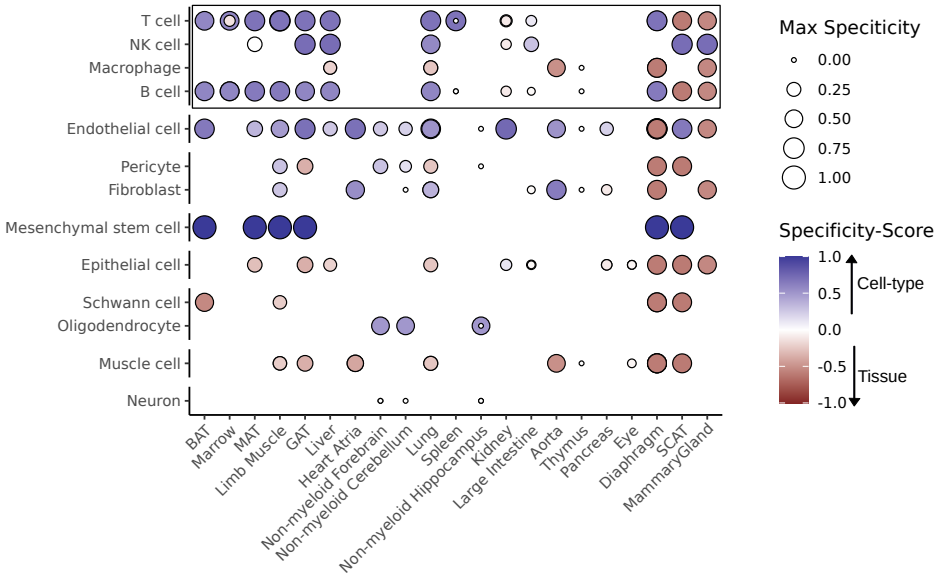


FIGURE 6.26: Specificity score for cell types and tissues. Multiple dots indicate different cell types belonging to the same broader cell type.

To focus on age-related differences affected by space, we excluded any genes that showed similar changes in both 8M vs. 3M in FL and 8M vs. 3M in HGC, considering both the direction of deregulation and their significance. We kept only those genes that were deregulated exclusively in 8M vs. 3M in FL or that exhibited different deregulation directions in 8M vs. 3M in FL compared to 8M vs. 3M in HGC. This process resulted in a selection of 3,297 genes across 26 tissues and 74 cell types. Similar to the earlier analysis, we created gene sets and identified 677 cell type cluster-specific and 184 tissue-specific genes.

The tissue-specific genes primarily included 68 genes related to the mammary gland and 42 genes specific to SCAT (as shown in Figure 6.27). A pathway analysis revealed that mammary gland-specific genes enrich for leucocyte activation (e.g., "myeloid leukocyte activation", "lymphocyte proliferation"), tissue development (e.g., "endothelial cell migration", "regulation of vasculature development"), and cell adhesion (e.g., "cell-matrix adhesion",

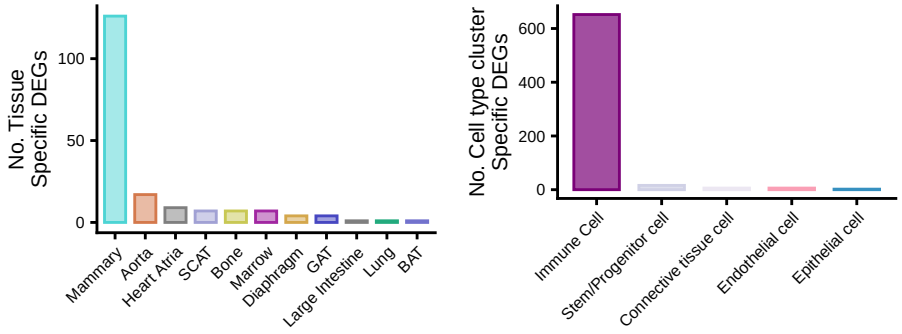


FIGURE 6.27: Distribution of tissue-specific (B) cell type cluster specific (C) genes across the groups.

"cell-substrate adhesion"), suggesting a potential change in how immune cells interact with tissue cells (Figure 6.28).

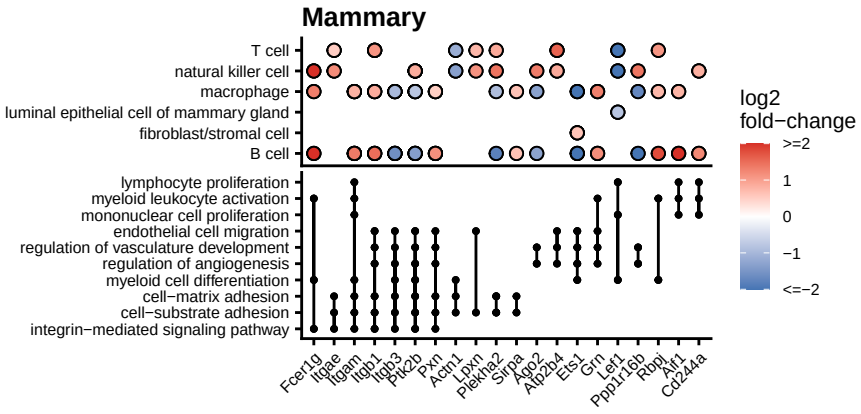


FIGURE 6.28: ORA pathway analysis of the Mammary-specific genes and the deregulation of the top 20 involved genes.

Immune cells, with 652 genes, contributed most of the cell type cluster-specific genes and enriched for pathways related to leukocyte differentiation and migration (e.g., "myeloid leukocyte migration", "leukocyte differentiation") and cell-cell adhesion (e.g., "leukocyte cell-cell adhesion") (Figure 6.29). This further supports the idea that alterations in the cytoskeleton, membrane, and extracellular matrix (ECM) may lead to age-related changes in the regulation of the immune system.

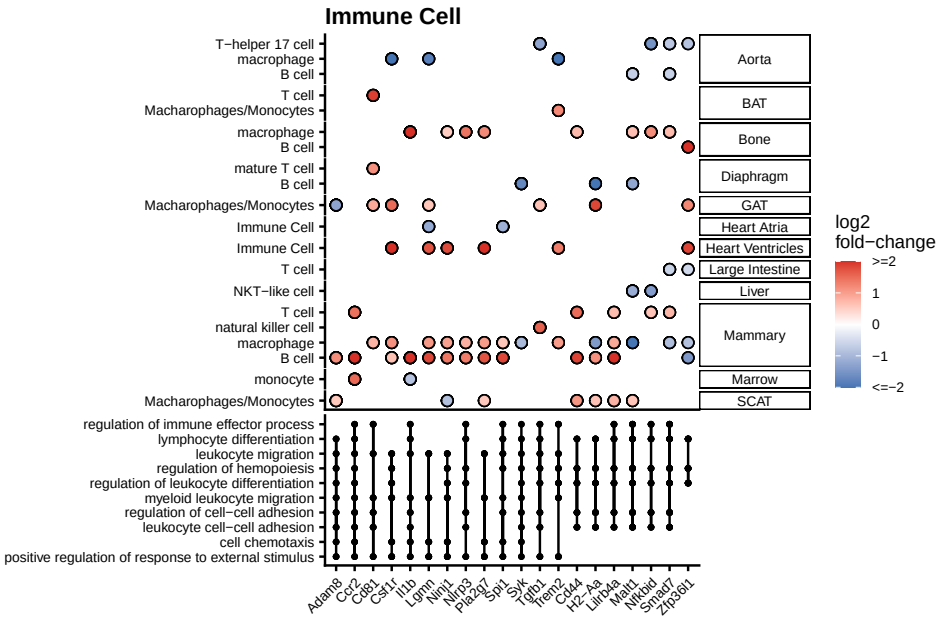


FIGURE 6.29: ORA pathway analysis of the immune cell-specific genes and the deregulation of the top 20 involved genes.

After analyzing these transcriptomic changes, we analyzed the miRNA data to determine whether and which changes were driven by miRNA regulation.

6.5 MURINE MIRNA REGULATION IN SPACEFLIGHT

To investigate how spaceflight affects the sncRNA profile of M. Musculus, we sequenced 686 samples, including 13 organs of mice sent to the ISS and ground controls. In addition to the FL and HGC groups used in the scRNA-seq data, the sncRNA-seq dataset comprised an additional VGC control group kept under standard conditions. Half of all mice were dissected after 21 days of spaceflight (TERM group) directly on the ISS; the other half, the LAR mice, were dissected on earth after returning from a 40-day spaceflight.

The samples comprised 348 from the Live Animal Return (LAR) group and 338 from the ISS Terminated (TERM) group (Figure 6.30). Overall, we

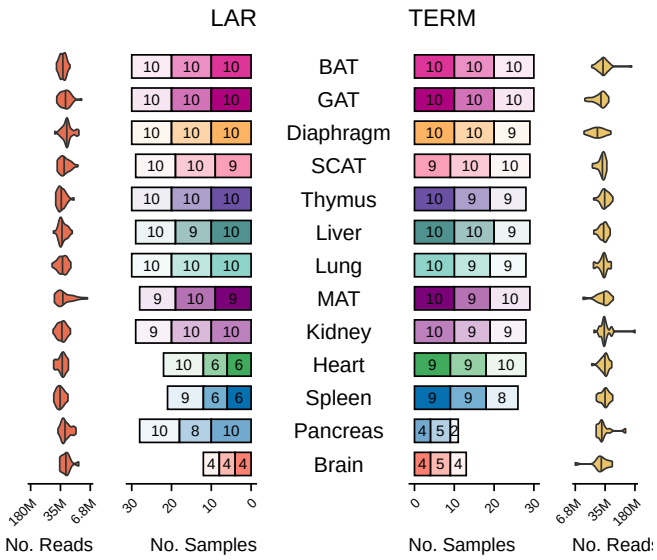
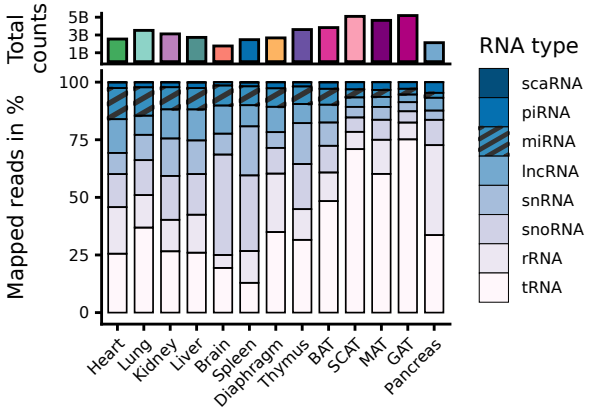
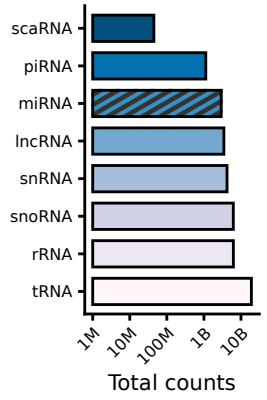


FIGURE 6.30: Number of samples in FL (dark), HGC (medium), and VGC (light) in the LAR and TERM groups for each tissue and the number of reads per sample. Lines in the violin plots show the median. Figure created by S. Rishik.

sequenced 39.8 billion short reads, with a median of 28.4 million per sample (interquartile range of 10.3 million), successfully mapping 70% to the mouse genome (GRCm39).



(A) Total counts of aligned reads and percentage of reads that aligned to each type of non-coding RNA.



(B) Total number of reads for each type of non-coding RNA.

FIGURE 6.31: Number of Reads per tissue and RNA type.

GAT, MAT, SCAT, and BAT showed the highest percentage of mapped reads. At the same time, the heart, lung, kidney, and liver had the highest proportion of miRNA-mapped reads (Figure 6.31a). Notably, the majority of reads mapped to tRNA (44%) and rRNA (14%), which are essential for mRNA translation (Figure 6.31b).

By performing a PVCA analysis, we found a median of 2.05% across tissues of the variance to be explained by the spaceflight (Figure 6.32). Given that the small RNA sequencing technology we used is optimized for miRNA detection, it provides the most reliable data. We thus focused our analysis on miRNAs.

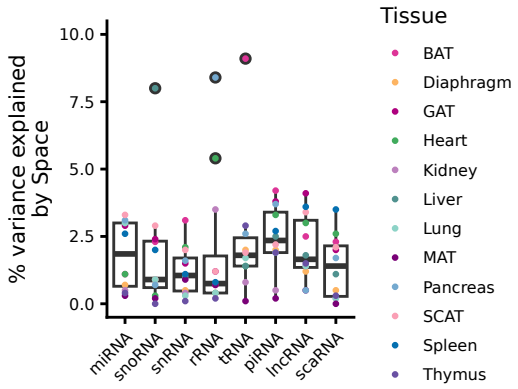
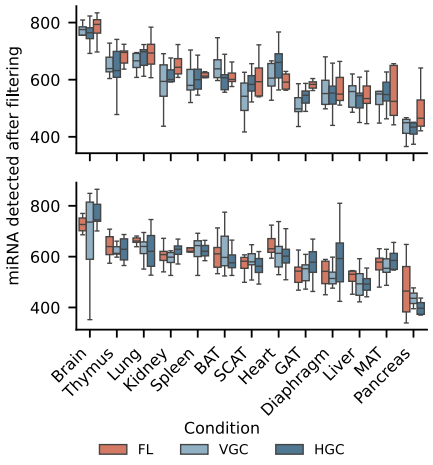
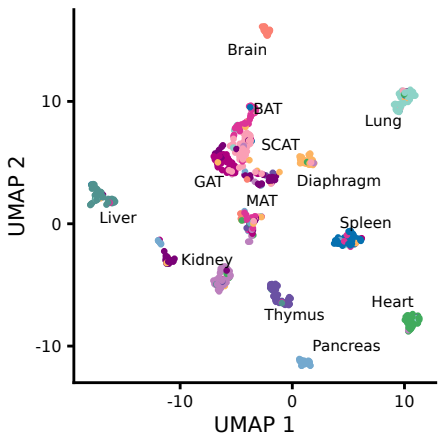


FIGURE 6.32: PVCA analysis for space/control in LAR per tissue and RNA type.

After mirBase quantification and filtering (see Methods), we kept 1,148 miRNAs in the LAR group and 1,133 in the TERM group (Figure 6.33a).



(A) Number of detected miRNAs in LAR (top) and TERM (bottom) per tissue and condition. Figure created by S. Rishik.



(B) Two-dimensional UMAP embedding of the 686 samples colored by tissue of origin.

FIGURE 6.33: Overview of miRNA sequences in the dataset.

For visualization, we projected all miRNA samples into a 2D UMAP space, revealing distinct clusters for each organ, except for the adipose tissues (BAT, MAT, GAT, and SCAT), which clustered together (Figure 6.33b).

6.5.1 *Impact of different factors on RNA expression*

Tissue type was the most significant factor, explaining 53.1% of the variance in the data based on Principal Variance Component Analysis (PVCA; Figure 6.34). Age, condition, and extraction method did not form distinct clusters in the embedding (Figure C.1). However, the factors Extraction-Tissue interaction and Extraction contributed 2.5% and 0.8% of the total variance, respectively, highlighting the impact of extraction timing (TERM: day 21, LAR: day 43) and preservation methods on the miRNA profiles. Consequently, we focused our analysis on the LAR group, using the TERM group to validate key findings.

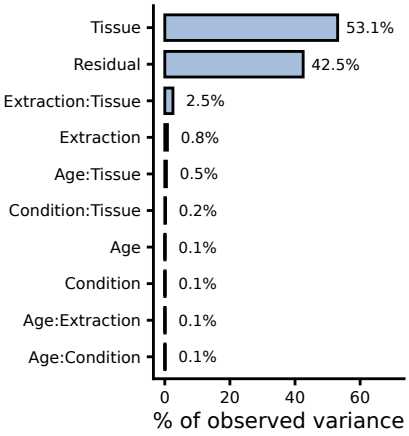
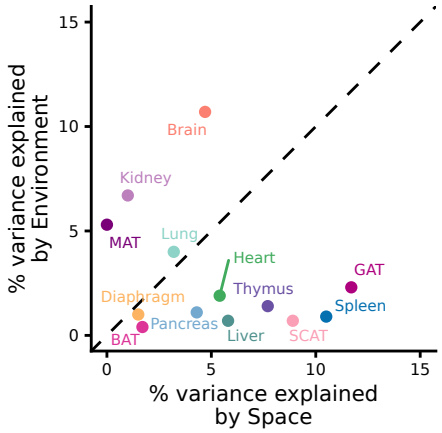


FIGURE 6.34: PVCA analysis of the variance explained by tissue, extraction method, age, condition, and their interactions. The residual variance is the variance in the data that can not be explained by these variables.

During spaceflight, mice are exposed to microgravity, increased radiation, and environmental changes, including modified housing and dietary conditions. To distinguish changes caused by spaceflight from those caused by environmental factors, we established two ground control groups: one that simulated spaceflight conditions and another under standard laboratory



Tissue	% Space	% Environment
BAT	1.7	0.4
Brain	4.7	10.7
Diaphragm	1.5	1
GAT	11.7	2.3
Heart	5.4	1.9
Kidney	1	6.7
Liver	5.8	0.7
Lung	3.2	4
MAT	0	5.3
Pancreas	4.3	1.1
SCAT	8.9	0.7
Spleen	10.5	0.9
Thymus	7.7	1.4

FIGURE 6.35 & TABLE 6.10: PVCA results of environment and spaceflight for each tissue. Environment corresponds to standard conditions (VGC) versus ISS/ISS-matched conditions (HGC and FL). Space corresponds to spaceflight (FL) or earth (HGC and VGC).

conditions. To determine the factors influencing expression variation, we performed a PVCA analysis calculating the proportion of miRNA variance attributable to spaceflight compared to environmental influences. This analysis indicated an organ-specific response: the muscle, kidney, and brain showed a greater portion of variance linked to housing conditions (Figure 6.35), while the GAT, spleen, SCAT, thymus, liver, and pancreas displayed variance primarily associated with spaceflight. Among these, the GAT, Spleen, and SCAT exhibited the highest levels of spaceflight-induced variance with 11.7%, 10.5%, and 8.9%, respectively.

6.5.2 *Distinguishing housing and spaceflight effects*

The number of deregulated miRNAs further supported this trend (Figure 6.36). GAT and SCAT were among the top three tissues with the highest counts of deregulated miRNAs. Specifically, GAT (66), heart (50), SCAT (33), and BAT (31) showed the most deregulated miRNAs in the context of spaceflight (FL vs. HGC), while diaphragm (1), kidney (2), MAT (2), and brain (7) had the lowest numbers. Interestingly, although the kidney, MAT, and brain had few deregulated miRNAs in the FL vs. HGC comparison, they showed

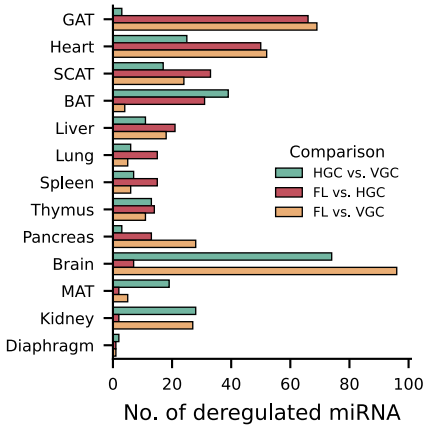


FIGURE 6.36: Number of deregulated miRNA (abs. $\log_2$FC > 1 & abs. cohen's D > 0.5) per tissue and comparison. Figure created by S. Rishik.

a considerably higher number in the FL vs. VGC and HGC vs. VGC comparisons, indicating that housing conditions have a greater impact on these organs. For instance, the brain comparison revealed 96 deregulated miRNAs for FL vs. VGC and 74 for HGC vs. VGC.

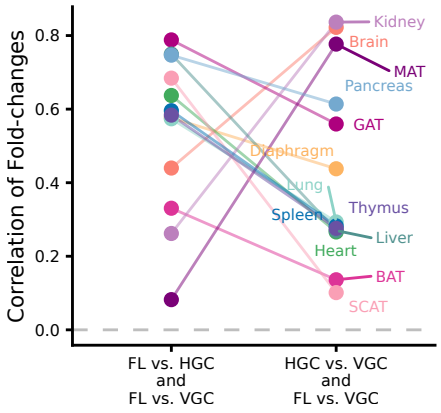


FIGURE 6.37: Correlation of $\log_2$ fold-changes in FL vs. HGC and FL vs. VGC (spaceflight-effects) and in HGC vs. VGC and FL vs. VGC (housing-effects). All correlations were significant (adj. p < 0.05).

To identify spaceflight-related changes, we compared FL vs. HGC (which highlights the effects of spaceflight) with FL vs. VGC (which reflects both spaceflight and housing influences). Moreover, we evaluated the impact of housing by comparing FL vs. VGC and HGC vs. VGC, with the second comparison focusing only on housing effects.

Correlation analysis showed that the fold changes between FL vs. VGC, as well as HGC vs. VGC were most similar in the kidney, MAT, and brain (Figure 6.37), suggesting that these organs were mainly affected by housing conditions. Conversely, GAT, spleen, and SCAT showed stronger correlations between FL vs. HGC and FL vs. VGC, displaying fold-change correlations exceeding 0.7.

We studied the commonly deregulated miRNAs between the three comparisons to refine our analysis at the individual

miRNA level. The diaphragm, thymus, and liver exhibited the highest percentages of commonly deregulated miRNAs in the spaceflight comparisons FL vs. HGC and FL vs. VGC (diaphragm: 100%, thymus: 63%, liver: 55%). Interestingly, neither the diaphragm nor the thymus showed any overlap in the housing comparisons FL vs. VGC and HGC vs. VGC (Figure 6.38). Contrarily, while MAT and kidney did not share any deregulated miRNAs across the two spaceflight comparisons, they showed a considerable overlap in the housing comparisons (MAT: 60%, kidney: 48%).

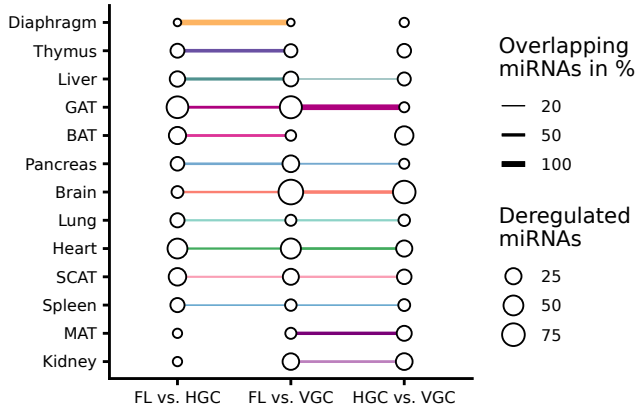
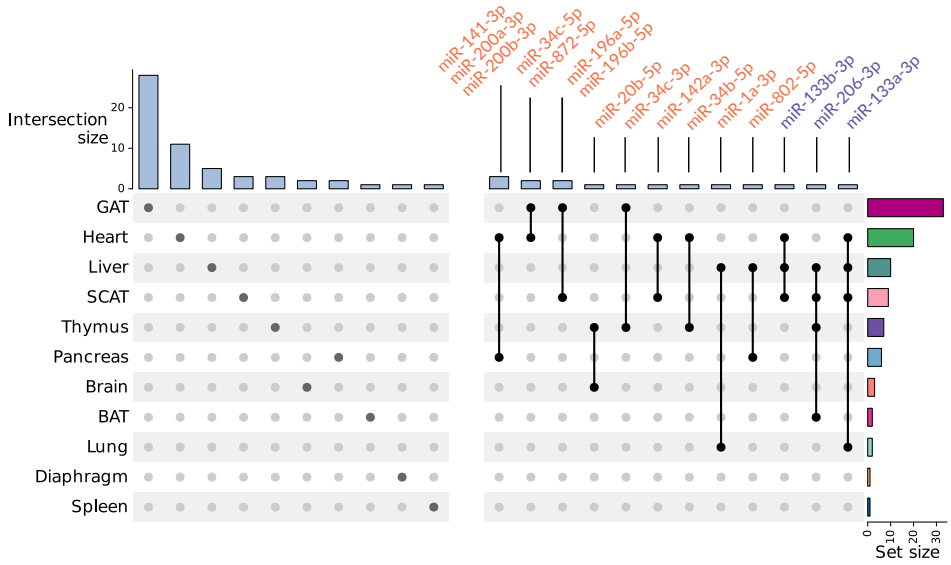
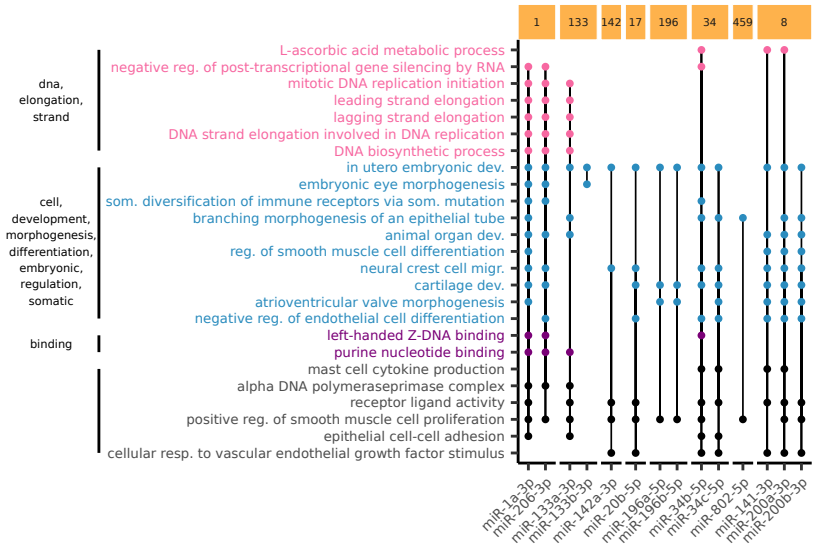


FIGURE 6.38: Number of deregulated miRNAs per tissue and comparison and the overlap between comparisons. The percentage is calculated as the proportion of the smaller set.

We identified 73 miRNAs that were consistently deregulated in both FL vs. HGC and FL vs. VGC, indicating they are affected by spaceflight (Figure 6.39a). Among these, six miRNAs were from the MIR-8 family (miR-141-3p/5p, miR-200a/b/c-3p, and miR-429-3p) and were found to be deregulated in the heart and pancreas. Additionally, five miRNAs belong to the MIR-17 family (miR-106a-5p, miR-20b-5p, miR-17-5p, miR-20a-5p, and miR-18b-5p), detected in the brain, GAT, heart, and thymus. Notably, 78% of these miRNAs exhibited tissue-specific deregulation, with the highest counts in GAT (28 miRNAs), heart (11 miRNAs), liver (5 miRNAs), and SCAT (3 miRNAs). While no single miRNA was deregulated across all tissues, 16 miRNAs showed deregulation in multiple tissues, indicating the systemic effects of spaceflight. These include members of the MIR-8 family, along with the MIR-196 family, where miR-141-3p, miR-200a-3p, and miR-200b-3p impact



(A) Spaceflight-related miRNAs within the different tissues. Spaceflight-related miRNAs show deregulation in FL vs. VGC and FL vs. HGC with the same direction of deregulation.



(B) ORA pathway analysis of the spaceflight-related miRNAs from (A) that occur in more than one tissue. Pathway clustering was based on GO similarity and labeled with overlapping terms.

FIGURE 6.39: Spaceflight-related miRNAs and their biological function

the heart and pancreas, while miR-196a-5p and miR-196b-5p influence GAT and SCAT.

To clarify the roles of these 16 systemic miRNAs, we performed a pathway analysis (ORA using miEAA; Figure 6.39b). The identified pathways clustered into three main categories: DNA synthesis and repair, developmental and structural changes, and binding (including Z-DNA and purines). The remaining pathways were associated with development and cell structure. MiRNAs from the MIR-1 family and miR-133a-3p impacted almost all the pathways, while others, such as MIR-34 or MIR-196, mainly affected developmental and structural processes.

A Gene Set Enrichment Analysis using all miRNAs (GSEA with miEAA) and filtered by significance in both FL vs. HGC and FL vs. VGC, found the most frequent pathways to be primarily enriched in SCAT, spleen, pancreas, and GAT (Figure 6.40). Pathway clustering showed pathway groups related to development, cellular structure, extracellular matrix, and nuclear remodeling processes.

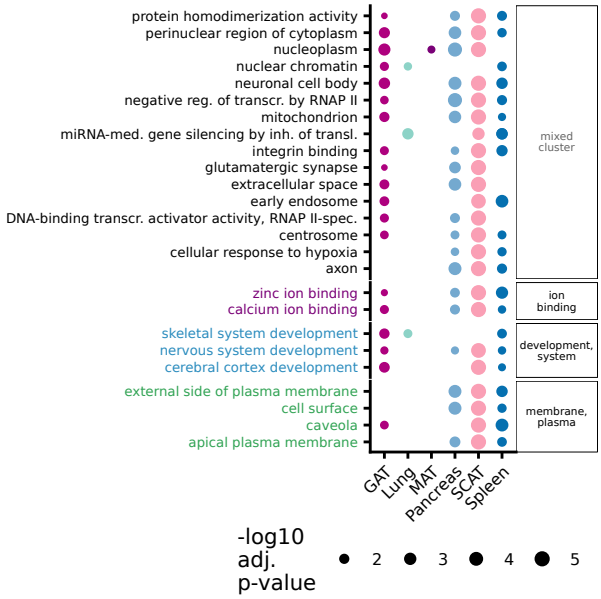


FIGURE 6.40: Top 25 pathways (GSEA; miEAA) that are significant in both FL vs. HGC and FL vs. VGC in most tissues. All top 25 pathways were enriched. Pathway clustering was based on GO similarity and labeled with overlapping terms.

6.5.3 *Effect of re-entry stress and spaceflight duration*

To assess whether the effects of spaceflight remain after a shorter spaceflight duration and in the absence of reentry-related stress in mice, we compared the LAR and TERM groups. Overall, the magnitude of miRNA deregulation was lower in the TERM group (Figure C.2). However, the patterns of deregulation showed a correlation between the FL vs. VGC and FL vs. HGC for both the LAR (Pearson's corr = 0.54, p-value $< 2.2 \cdot 10^{-16}$) and TERM groups (Pearson's corr = 0.58, p-value $< 2.2 \cdot 10^{-16}$) (Figure 6.41).

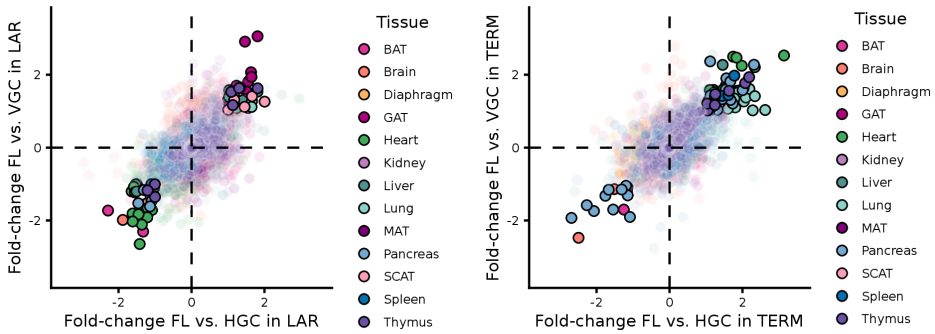


FIGURE 6.41: Comparison of spaceflight-related effects in LAR (left) and TERM (right). The opaque dots were deregulated in both comparisons, with matching fold change directions.

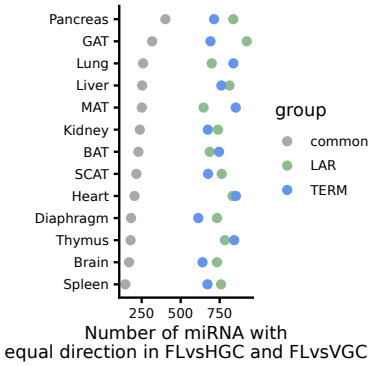


FIGURE 6.42: Number of miRNAs with matching fold-change direction between FL vs. HGC and FL vs. VGC in LAR, in TERM and that are common between LAR and TERM.

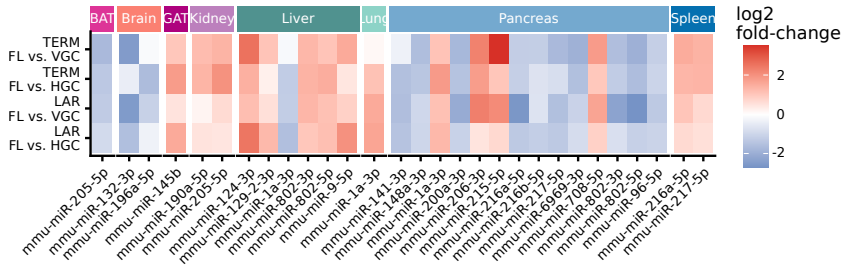


FIGURE 6.43: Overlapping miRNA from Figure 6.42 that are significant at least once in LAR and in TERM.

Filtering for miRNAs that were consistently up- or down-regulation in both the LAR and TERM groups revealed that the pancreas (with 402 miRNAs), GAT (317 miRNAs), and lung (259 miRNAs) had the biggest overlap (Figure 6.42). Among these, 22 miRNAs were significantly deregulated in one of the comparisons within both LAR and TERM (Figure 6.43), while four miRNAs showed consistent deregulation across all comparisons: miR-802-3p (found in the liver), miR-200a-3p and miR-96-5p (in the pancreas), and miR-802-5p (in both liver and pancreas).

To connect these alterations to biological function, we conducted a pathway analysis (GSEA with miEAA) for TERM (Figure 6.44). The top 25 pathways were notably similar to those observed in LAR, particularly regarding binding processes and modifications of the membrane and nuclear surroundings. A comparison of the top 25 pathways in the LAR and TERM groups indicated that all were enriched under both conditions, with specific differences in the affected tissues. For instance, the pancreas, GAT, and lung exhibited similar pathway enrichment in both LAR and TERM, while the spleen, thymus, and diaphragm demonstrated

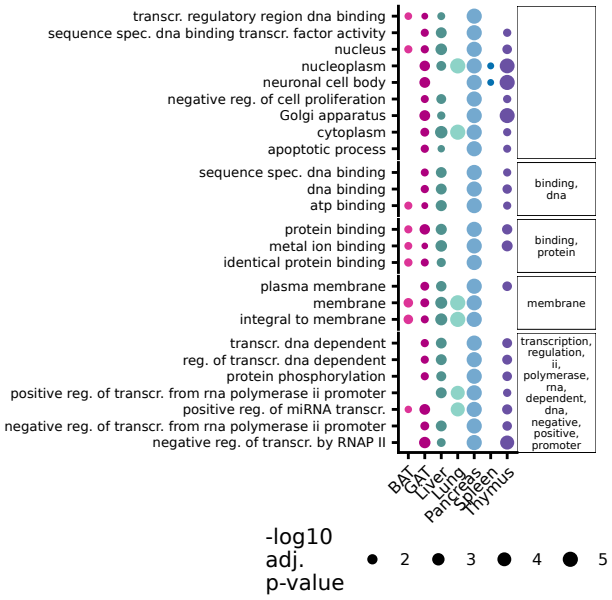


FIGURE 6.44: Top 25 pathways in TERM as in Figure 6.40.

more marked differences (Figure C.3). These observed similarities corresponded with overlaps in the pancreas, GAT, and lung, which showed comparable pathway activation. In contrast, the thymus and spleen displayed patterns specific to the involved conditions and tissues.

We also compared the pathways identified in LAR with the findings from the corresponding mRNA data using Gene Set Enrichment Analysis (GSEA via Genetrail). The most frequently overlapping pathways in both datasets were found in cell types from SCAT, spleen, and pancreas (Figure C.4). The predominantly affected pathways involved extracellular matrix organization and tissue remodeling, impacting endothelial cells in the kidney, epithelial cells from SCAT, and immune-related cells such as macrophages in the spleen and kidney. These similarities in pathways suggest a direct connection between the deregulated mRNAs and the target genes influenced by the deregulated miRNAs.

6.5.4 *Integrational analysis of mRNA and miRNA reveals systemic and tissue-specific regulation*

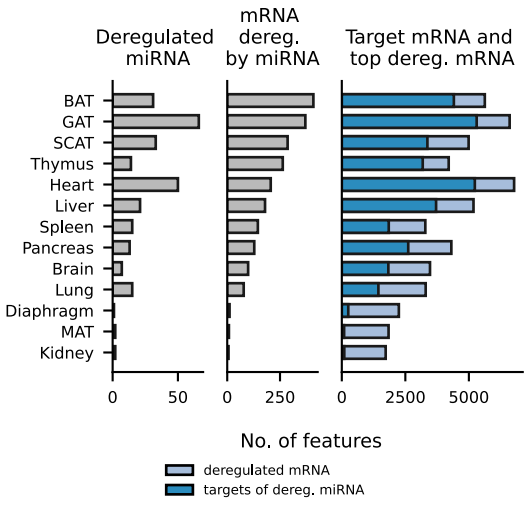


FIGURE 6.45: Number of deregulated miRNA, top deregulated mRNA, mRNA that targeted by deregulated miRNA and the overlap between targeted mRNAs and deregulated mRNAs (mRNA deregulated by miRNA). Figure created by S. Rishik.

MicroRNAs' primary role is regulating gene expression by targeting mRNAs for degradation or inhibiting their translation. Integration of single-cell RNA sequencing data allows us to directly assess whether spaceflight-induced alterations in miRNA levels correspond to changes in gene expression. Given that the mRNA dataset encompasses all 13 tissues and the LAR-FL and LAR-HGC groups, we identified miRNA targets that were also altered at the mRNA level using tissue-level pseudo-bulk gene expression analysis. Importantly, we did not filter the results based on the direction of alteration in either mRNA or

miRNA, as the interaction between these two does not always produce a negative correlation due to factors such as indirect regulation, co-regulation with other miRNAs, and the influence of transcription factors [427]. Our analysis revealed that the BAT had the highest number of deregulated mRNAs targeted by deregulated miRNAs, with a total of 408 different mRNAs (Figure 6.45). The miRNA-mRNA pairs miR-132-3p - *Rorb*, miR-206-3p - *Smad9*, and miR-378a-3p - *Tfcp2l1* were altered across most tissues (4 each). Notably, these three targets are all transcription factors, suggesting they amplify miRNA regulatory signals in various tissues. Interestingly, we noted that the deregulation of mRNAs did not always align with the deregulation of miRNAs. For instance, the diaphragm, MAT, and kidney tissues exhibited 2001, 1757, and 1635 deregulated mRNAs, respectively, but only had 1, 2, and 2 deregulated miRNAs, leading to a correspondingly low number of miRNA-mRNA pairs. These observations imply that factors beyond miRNA deregulation likely contribute to the effects observed with spaceflight.

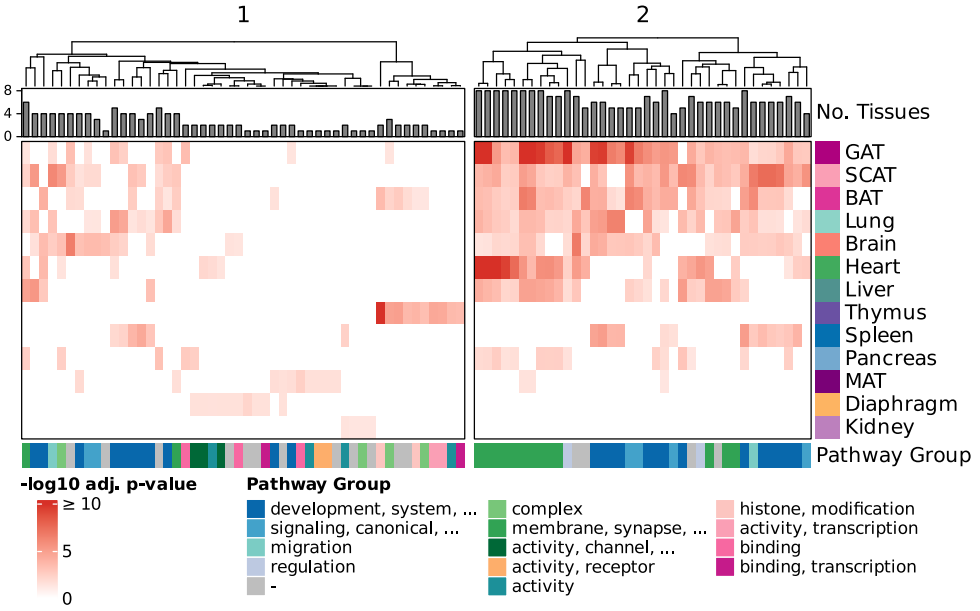


FIGURE 6.46: Top 5 most enriched pathways for each tissue (ORA; clusterProfiler; GO: Biological Process) using the top deregulated mRNA as input. Pathway groups were determined using clustering by GO similarity and labeled by the most frequent terms in the cluster.

We performed a pathway analysis (ORA with clusterProfiler) on the targeted and deregulated mRNAs to further explore these changes. From initially 3,455 pathways, we identified the top 10 most enriched pathways for each tissue (Figure 6.46). Hierarchical clustering revealed two distinct categories: pathways that affect only a limited number of tissues (cluster 1) and pathways that display systemic enrichment across multiple tissues (cluster 2).

Disruption of systemic pathways primarily impacted SCAT, GAT, BAT, liver, lung, heart, brain, and spleen. The most significant pathways within this cluster were related to development and processes associated with cell membranes, including "lung development", "epithelium migration", and "presynaptic membrane" (Figure C.5).

The thymus, MAT, and diaphragm showed tissue-specific patterns of pathway deregulation. In MAT, diaphragm, and kidney, the affected pathways focused on membrane function, such as "cadherin binding" and "adherens junctions". The diaphragm showed disruption in several ion channel-related pathways, including "ion channel inhibitor activity" and "potassium ion leak channel activity". In contrast, the thymus displayed disturbances in pathways linked to the nuclear environment, including "regulation of histone modification" and "heterochromatin organization". Of note, biological pathways consist of subcomponents, so while pathways like "bone development" and "lung development" may seem similar, they have both common and distinctive elements.

To determine whether common genes caused the pathway enrichment of related pathways, we examined the distribution of mRNAs enriched in these pathways across various tissues from cluster 2 (Figure 6.47).

While the selected pathways exhibited similar functions, the participating mRNAs were vastly different, with most appearing in fewer than five pathways. However, four miRNAs (miR-124-3p, miR-142a-5p, miR-1a-3p, and miR-206-3p) regulated mRNAs involved in all pathways. As a result, we concentrated on the top 20 miRNAs involved in most pathways. In line with the systemic pathways, several miRNAs, including miR-20b-5p and miR-132-3p, participated in up to 18 distinct tissue-specific pathways from cluster 1 (Figure C.6). The MIR-17 family of miRNAs was particularly noteworthy due

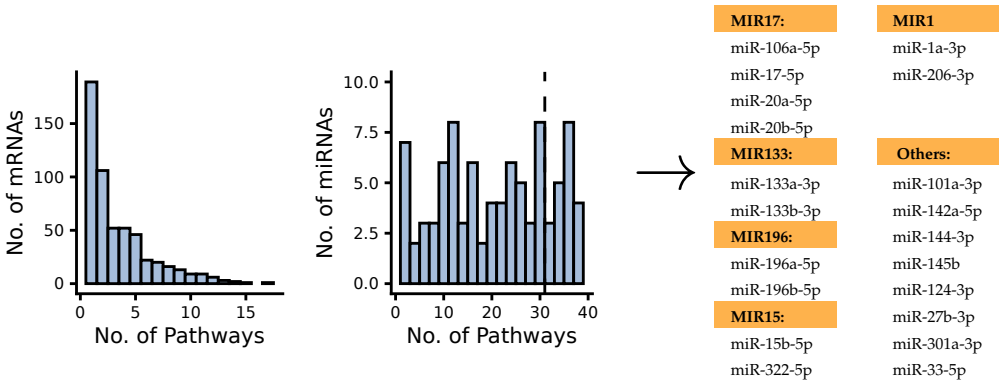


FIGURE 6.47 & TABLE 6.11: Frequency of each mRNA across pathways and the number of miRNA targeting these mRNAs for the systemic pathways from Figure 6.46. miRNAs contributing to at least 30 pathways are shown on the right and grouped by the MIR family.

to its role in structural changes and nuclear reorganization through histone modification in BAT and the thymus.

The similarities observed between the pathways led us to identify target genes of miRNAs involved in both tissue-specific and broadly dysregulated pathways from clusters 1 and 2. We specifically identified the mRNAs targeted by most miRNAs (Figure 6.48). In SCAT, BAT, and GAT, the most commonly targeted genes included *Itgb8*, *Flt1*, *Rarb*, *Mecom*, and *Eya4*, crucial for focal adhesion, blood vessel formation, and tissue structure. Among these, several genes are targeted by miRNAs from the MIR-1, MIR-133, and MIR-17 families. Interestingly, four out of the 35 most frequently targeted mRNAs encode transcription factors (*Lcor*, *Rarb*, *Smad9*, and *Mecom*), which likely enhance the regulatory signals to other genes (Figure 6.49).

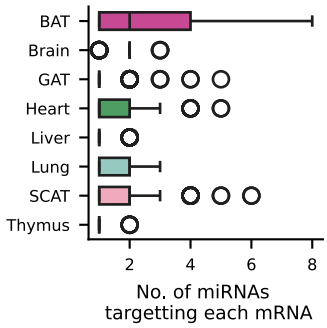


FIGURE 6.48: Number of deregulated miRNAs that are targeting each mRNA per tissue. Figure created by S. Rishik.

The deregulation of miRNAs led to decreased mRNA levels in the thymus, liver, and lung and increasing levels in BAT, SCAT, GAT, and the brain. To analyze miRNA expression variations across different tissues and conditions, we visualized members of the MIR-17 and MIR-1 families (Figure 6.50).

Members of the MIR-17 family were most prevalent in the thymus. However, their target genes exhibited significant dysregulation, mainly in SCAT,

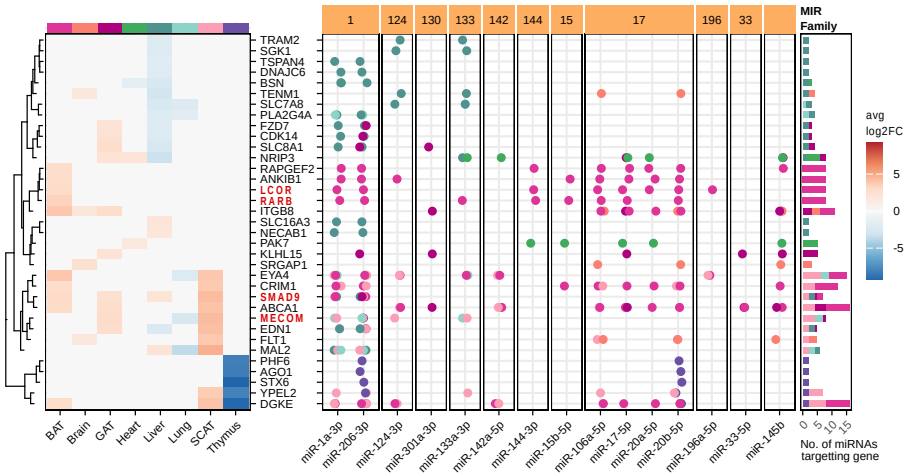


FIGURE 6.49: Deregulation of the mRNA set with the highest number of miRNAs targeting them in each tissue from Figure 6.48. Transcription factors are marked red. The dot plot shows miRNA-mRNA targeting, colored by the tissue in which the pair was found, and the barplot shows the frequency of miRNA-mRNA targeting per mRNA and tissue. Figure created by S. Rishik.

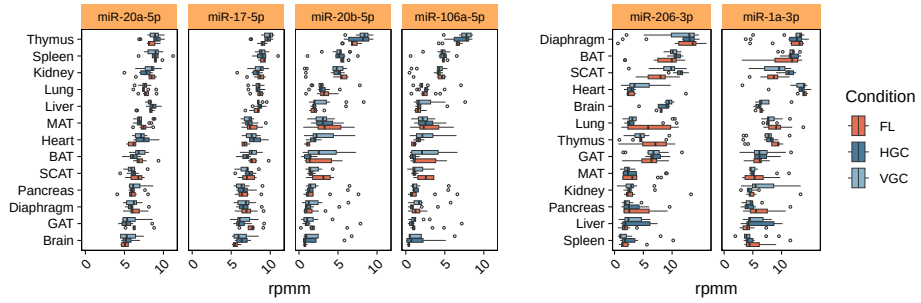


FIGURE 6.50: Normalized expression of miRNAs from MIR-17 and MIR-1 cluster by condition and tissue. Figure created by S. Rishik.

BAT, and GAT. Specifically, miR-20b-5p targeted the downregulated mRNAs Phf6, Ago1, Stx6, Ypel2, and Dgke in the thymus. On the other hand, MIR-1 family members were most abundant in the diaphragm but also regulated genes in SCAT, GAT, and BAT. Interestingly, miR-206-3p targeted the same

set of downregulated genes in the thymus as miR-20b-5p. Our findings suggest that spaceflight alters the specific expression patterns of miRNAs in different tissues, which seems to correlate with gene dysregulation. The MIR-17 family, primarily expressed in the thymus, influences mRNAs in adipose tissues, while the MIR-1 family, primarily found in the diaphragm, also regulates genes across SCAT, GAT, and BAT. Moreover, the joint targeting of a cluster of downregulated genes in the thymus by miR-20b-5p and miR-206-3p indicates a potential coordinated regulation.

Given that miRNA expression can change with age and that the effects of spaceflight mirror aspects of aging, an important question arises: Are the changes induced by spaceflight different between young (3-month-old) and middle-aged (8-month-old) mice and can those differences explain the age-differences observed in the single-cell data?

6.5.5 *Age-dependent and -independent miRNA-mRNA regulation in spaceflight*

Having observed age-dependent spaceflight changes in the single-cell data, we investigated whether miRNA regulations cause those changes. Due to insufficient samples, we excluded the brain from this analysis (Figure 6.51).

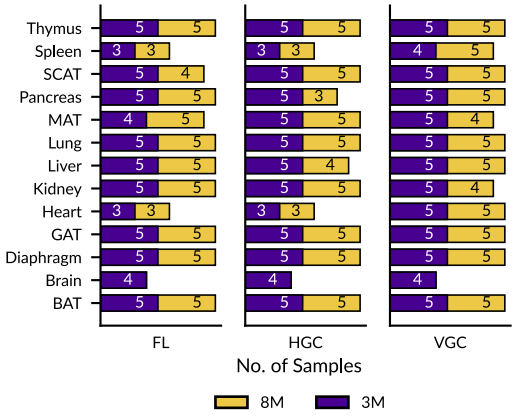
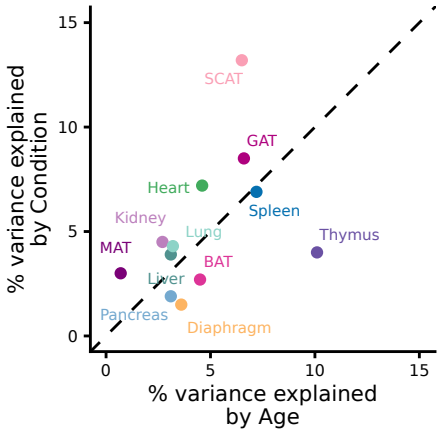


FIGURE 6.51: Age-distribution per tissue and condition (LAR only). Figure created by S. Rishik.

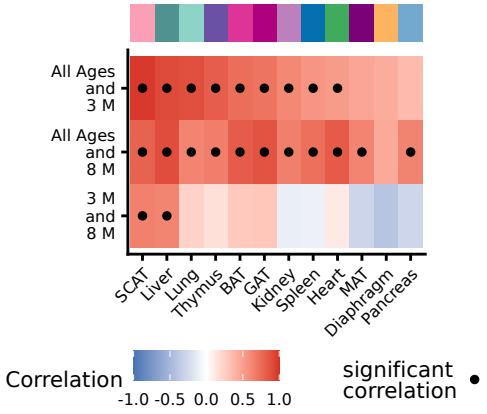
Using PVCA for the variables age and condition (FL, HGC, VGC), we found that the thymus exhibited the highest age-related effect, with results

showing 10.1% of variance explained by age and 4% by condition (Figure 6.52).



Tissue	% Age	% Condition
Thymus	**10.1**	4
Spleen	**7.2**	6.9
GAT	**6.6**	**8.5**
SCAT	6.5	**13.2**
Heart	4.6	**7.2**
BAT	4.5	2.7
Diaphragm	3.6	1.5
Lung	3.2	4.3
Liver	3.1	3.9
Pancreas	3.1	1.9
Kidney	2.7	4.5
MAT	0.7	3

FIGURE 6.52 & TABLE 6.12: PVCA for condition and age.



Tissue	Correlation	adj. p-value
SCAT	0.65	$1.73 \cdot 10^{-139}$
Liver	0.62	$1.53 \cdot 10^{-126}$
GAT	0.29	$3.03 \cdot 10^{-24}$
BAT	0.27	$2.60 \cdot 10^{-21}$
Lung	0.23	$6.44 \cdot 10^{-15}$
Thymus	0.16	$8.26 \cdot 10^{-8}$
Heart	0.11	$1.84 \cdot 10^{-4}$
Spleen	-0.10	$1.32 \cdot 10^{-3}$
Kidney	-0.11	$1.60 \cdot 10^{-4}$
Pancreas	-0.28	$5.10 \cdot 10^{-22}$
MAT	-0.29	$5.25 \cdot 10^{-23}$
Diaphragm	-0.40	$1.60 \cdot 10^{-45}$

FIGURE 6.53 & TABLE 6.13: Correlation of $\log_2$ fold-changes in FL vs. HGC in 3M, 8M and both. Significant correlations (adj. $p < 0.05$, abs. corr > 0.5) are marked with a dot.

When examining the fold-change correlations between FL and HGC, both the 3-month and 8-month groups showed significant positive correlations with the overall group across both age brackets (average Pearson's corr = 0.66, standard deviation = 0.20; average corr = 0.71, standard deviation = 0.13; Figure 6.53). However, the correlation diminished when directly comparing the 3-month and 8-month groups (average corr = 0.09, standard deviation =

0.34), suggesting a shared response to spaceflight influenced by tissue specificity and age-related factors.

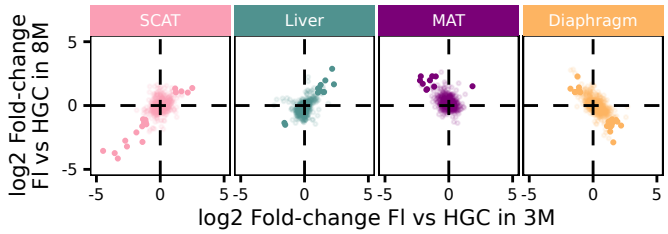
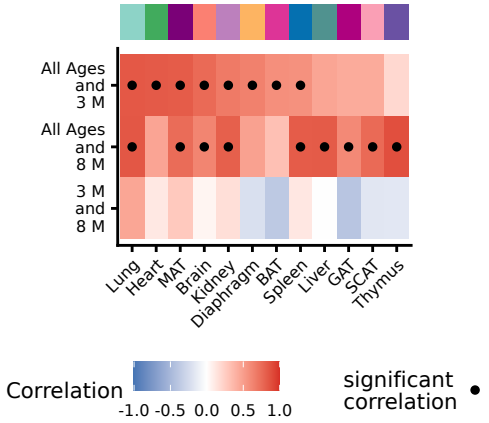


FIGURE 6.54: Comparison of fold-change for the two tissues with the highest and the lowest correlation.

SCAT (corr = 0.65, $1.72 \cdot 10^{-}139$) and liver (corr = 0.62, $1.53 \cdot 10^{-}126$) demonstrated a strong correlation, suggesting minimal age dependence. In contrast, MAT (corr = -0.29, $5.25 \cdot 10^{-}23$) and diaphragm (corr = -0.40, $1.59 \cdot 10^{-}45$) exhibited a weak correlation (Figure 6.54).

We repeated this analysis with the TERM group to ensure that these differences are not caused by re-entry stress or the effects of extended spaceflight (Figure 6.55).



Tissue	Correlation	adj. p-value
Spleen	0.12	$7.69 \cdot 10^{-5}$
SCAT	-0.17	$2.31 \cdot 10^{-8}$
Brain	0.05	$1.51 \cdot 10^{-1}$
BAT	-0.37	$5.85 \cdot 10^{-38}$
Heart	0.12	$1.15 \cdot 10^{-4}$
MAT	0.28	$2.68 \cdot 10^{-21}$
GAT	-0.40	$1.15 \cdot 10^{-43}$
Kidney	0.16	$9.62 \cdot 10^{-8}$
Diaphragm	-0.21	$1.89 \cdot 10^{-12}$
Lung	0.46	$8.70 \cdot 10^{-60}$
Liver	-0.007	1
Thymus	-0.16	$1.23 \cdot 10^{-7}$

FIGURE 6.55 & TABLE 6.14: Correlation as in Figure 6.53 for the TERM group.

LAR and TERM showed differences in the tissues exhibiting the strongest positive and negative correlations. In contrast to LAR, which showed significant effects in the MAT, diaphragm, and pancreas tissues, TERM displayed the weakest correlation in the diaphragm (corr = -0.21, $1.89 \cdot 10^{-}12$), as well as BAT (corr = -0.37, $5.85 \cdot 10^{-}38$) and GAT (corr = -0.40, $1.14 \cdot 10^{-}43$). We excluded the pancreas from the TERM analysis due to insufficient samples. Our findings suggest that age-dependent effects are evident even during the early phases of spaceflight.

Given the effects of spaceflight observed in different age groups, we explored the possibility that these variations indicate accelerated aging. We compared our findings with age-related miRNA expression patterns from the Tabula Muris Senis (TMS) non-coding RNA dataset to investigate this, allowing us to analyze age-related changes in mice exposed to spaceflight and those in mice undergoing physiological aging up to 21 months. Since our study focused on female mice, we filtered the TMS data accordingly to ensure alignment in sex and tissue type (Figure 6.56).

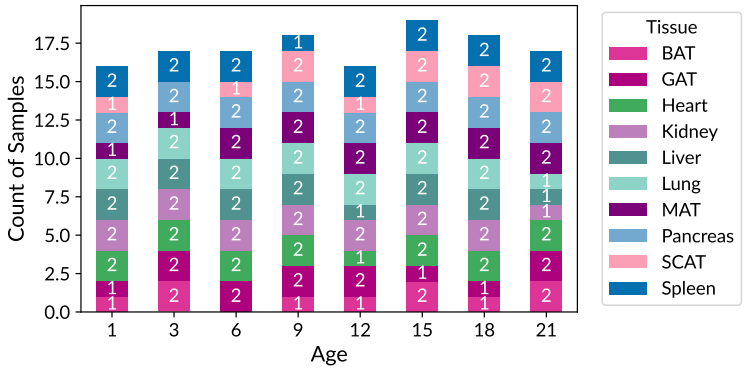


FIGURE 6.56: Number of samples per tissue and age in TMS (only female).

Notably, the M. musculus strain for our spaceflight study (BALB/cAnNTac) differed from TMS (vC57BL/6JN). As miRNA expression is tissue and age- and strain-specific, this adds noise to our comparison. We used 3-month-old mice as the baseline for TMS, compared it against progressively advancing age and considered the overlap with our 8M vs. 3M comparisons in HGC-control and Flight (Figure 6.57). In mice exposed to spaceflight, the number of overlapping miRNAs increased with age, peaking at 21 months in the GAT

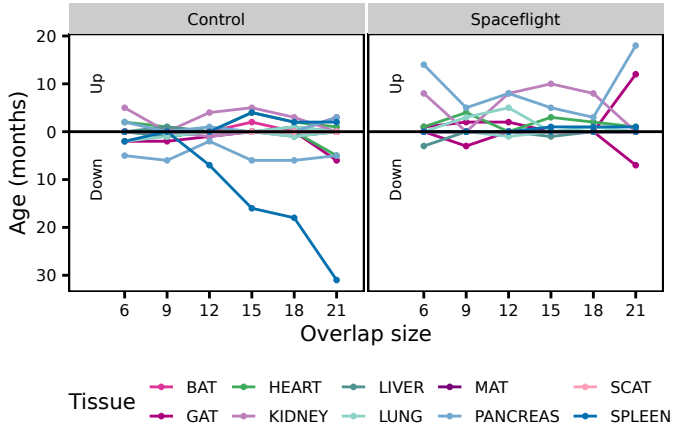


FIGURE 6.57: Overlap between deregulated miRNAs in 3M vs. 8M in FL and HGC with the deregulated miRNAs in TMS between the different ages (6,9,12,15,18, and 21 Months) and baseline (3M) with a matching direction of deregulation.

and pancreas. The comparison of TMS at 9 months with 3 months aligns closely with our control comparison of 8 months versus 3 months, which suggests we might see larger overlaps at this time point. However, we did not observe a peak for the Control group around the 9-month mark. To explore overlaps with aging-related miRNAs, we compared the miRNAs that show deregulation in at least three time points in TMS to those modified between the 3-month and 8-month periods across FL, HGC, and VGC (Figure 6.58).

This analysis identified three miRNAs (miR-92a-3p, miR-322-5p, and miR-205-5p) exhibiting consistent changes in TMS and control groups. In the Flight group, miR-92a-3p followed a trajectory consistent with aging, whereas miR-322-5p and miR-205-5p showed expression patterns that deviated from normal physiological aging. The comparison with the TMS cohort highlighted complex age-dependent differences between the spaceflight and control groups, indicating the necessity for further research using mice of the same strain and an extended age range.

Again, we performed an integrative analysis with both mRNA and miRNA data to better understand the biological mechanisms behind age-specific changes. We identified mRNAs that targeted deregulated miRNAs and displayed differential expression between 3-month-old and 8-month-old mice in both Flight and HGC conditions, aimed at clarifying the aging effects

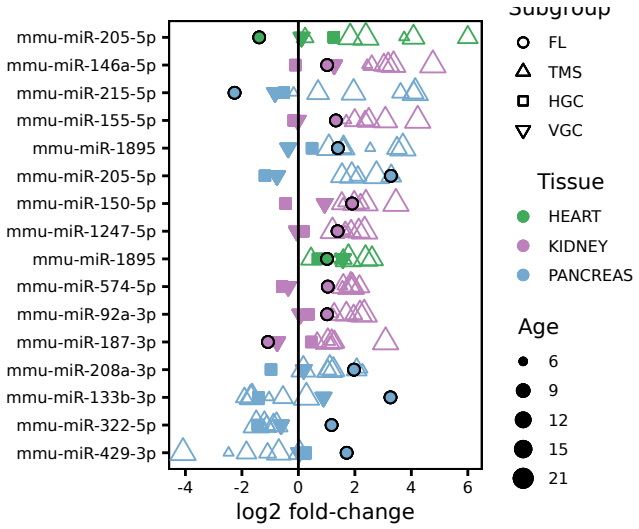


FIGURE 6.58: Log_2 fold changes in 8M vs. 3M for FL, HGC, and VGC and in the TMS in different age groups compared to the baseline (3M) for miRNAs that were deregulated in TMS in more than three age groups.

induced by spaceflight (Figure C.7). Pathway analysis using GSEA with miEAA showed that most top-enriched pathways were common between Flight and HGC (Figure C.8). These results suggest that the age-related differences observed in Flight and HGC are similar, indicating that they do not explain the negative correlations between FL and HGC in both 3-month and 8-month groups.

Baseline	**Test**	**FL 3M**		**FL 8M**	
3M HGC	3M FL	↑ and ↓	↓ and ↑		
3M FL	8M FL			↑ and ↑	↓ and ↓
8M HGC	8M FL				

TABLE 6.15: Filtering for age-dependent spaceflight effects.

Consequently, we controlled for the effects of spaceflight that similarly impacted both age groups, thereby isolating the age-dependent effects of spaceflight. We defined two groups of miRNAs: Flight 3M and Flight 8M. The Flight 3M miRNAs were deregulated in both FL vs. HGC in 3M and in 8M vs. 3M in FL, consistently showing either up- or downregulation in the

3-month FL group. Similarly, we defined Flight 8M miRNAs using the comparisons of FL vs. HGC in 8M and in 8M vs. 3M in FL (Table 6.15).

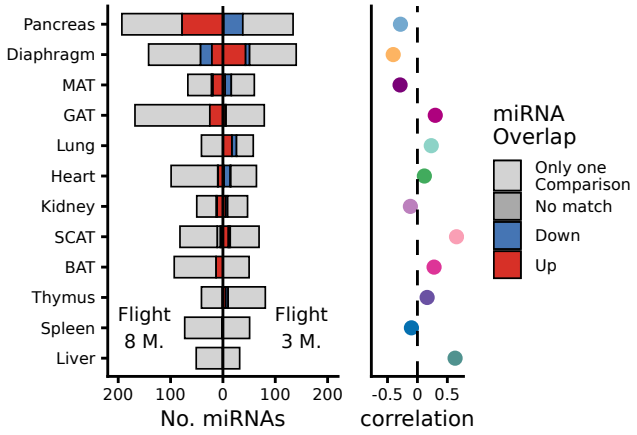


FIGURE 6.59: Number of deregulated miRNA in the comparisons for Flight 3M and Flight 8M with consistent up- or down-regulation for FL 3M or FL 8M, with different directions (No match) or with deregulation in only one comparison (Only one comparison). Correlations from Figure 6.53.

The pancreas, diaphragm, and MAT, which previously showed the weakest correlations, exhibited the highest number of age-dependent miRNAs in response to spaceflight (pancreas: 116, diaphragm: 94, and MAT: 37) (Figure 6.59).

We conducted a similar analysis on the most deregulated mRNAs and examined their overlap with miRNA targets, which we used as input for pathway analysis (ORA with clusterProfiler). The clustering of the top three pathways for each tissue (Figure 6.60) revealed only a handful of membrane-related pathways that were common across several tissues (MAT, GAT, and diaphragm). However, we found most of the age-dependent spaceflight pathways to be specific to individual tissues.

For instance, the kidney showed enrichment in "ECM structural constituent conferring tensile strength", while MAT exhibited enrichment in "regulation of cell morphogenesis involved in differentiation" in 3-month mice. MIR families group miRNAs by sequence and functional characteristics, allowing us to identify which MIR families were most influential regarding age-related effects in spaceflight.

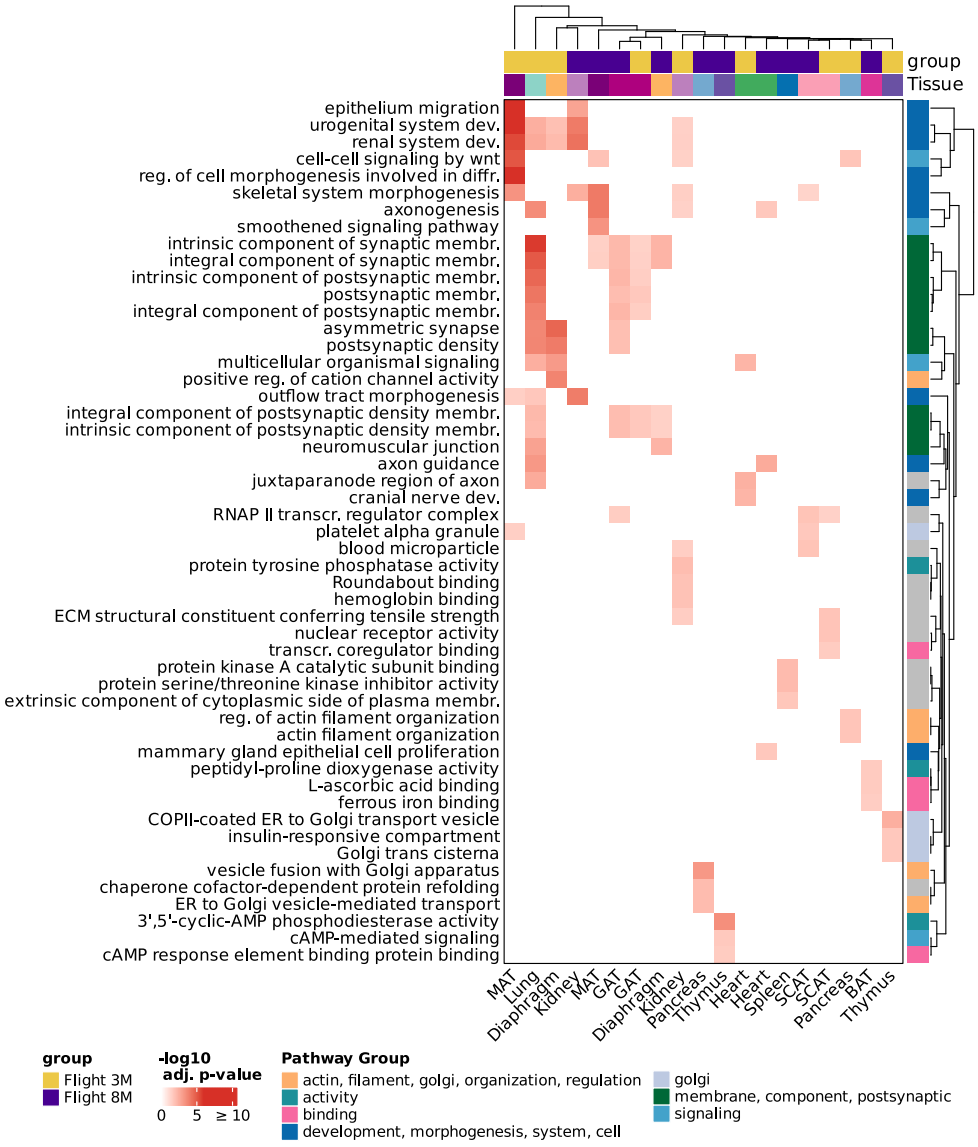


FIGURE 6.60: Pathway analysis (ORA, clusterProfiler) of the mRNAs targeted by Flight 3M and the Flight 8M miRNAs. mRNAs had to fulfill the same restrictions as the miRNA (Table 6.15). Pathways clustering used GO similarity and labeled the most frequent terms in the cluster.

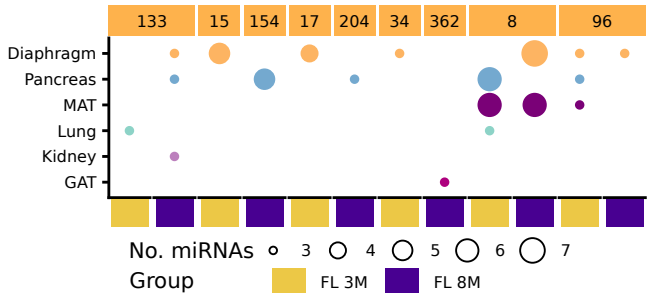


FIGURE 6.61: Number of Flight 3M and Flight 8M miRNAs from per tissue, grouped by the MIR Family (for families > 2 miRNAs in one tissue).

We pinpointed the MIR families significantly associated with age-dependent spaceflight changes by focusing on those with at least two deregulated members within a tissue. In this context, MIR-8, MIR-154, and MIR-15 emerged as the most prominent families, primarily linked to age and spaceflight in the diaphragm, pancreas, and MAT (Figure 6.61).

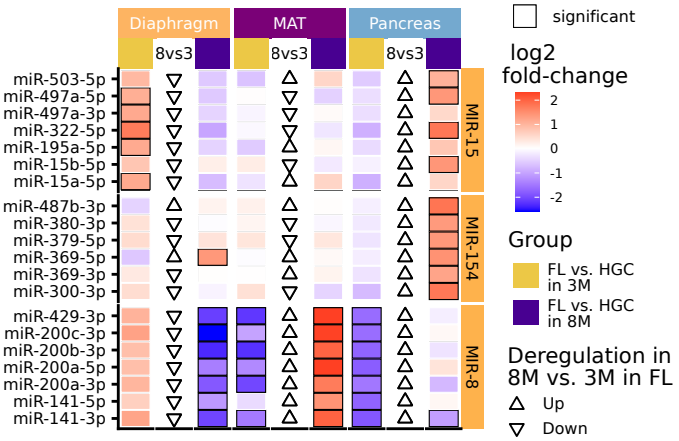


FIGURE 6.62: Log_2 fold change in FL vs. HGC in 3M and in 8M of miRNAs from the three most frequent MIR-families from Figure 6.61.

Notably, most MIR family members in each tissue show a consistent pattern of deregulation in the same direction under similar conditions 6.62. Since we did not specifically select this outcome, these findings indicate a highly coordinated regulatory mechanism. Among the 49 mRNAs targeted

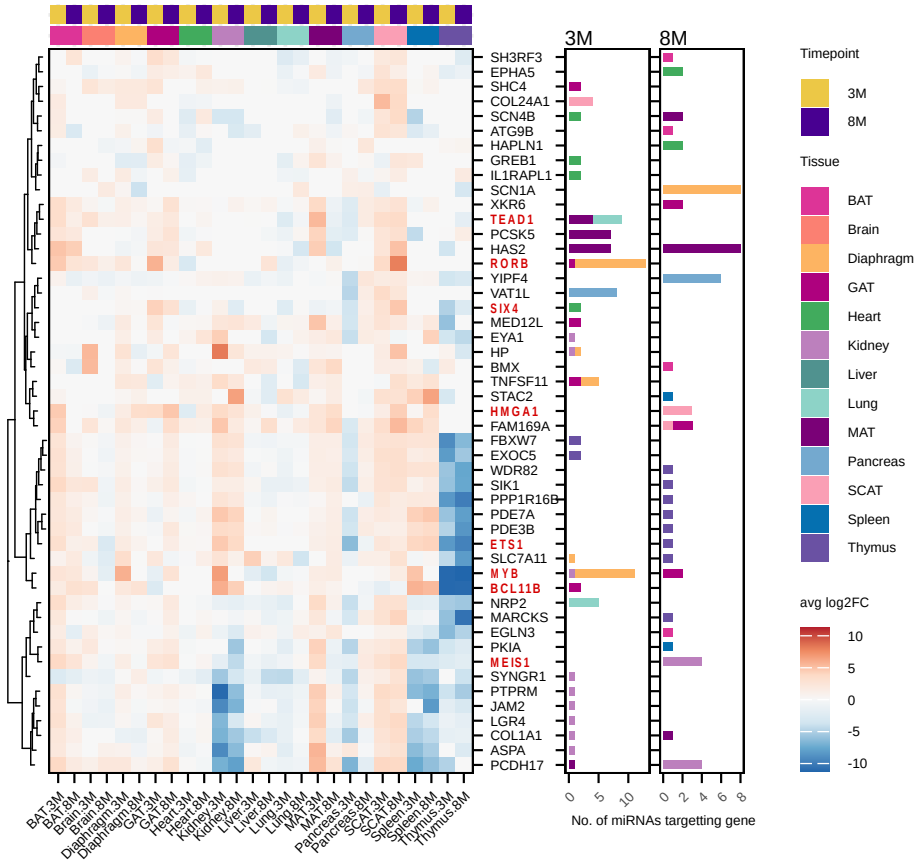


FIGURE 6.63: Log_2 fold changes for FL vs. HGC in 3M and in 8M for the most frequently targeted mRNA set that are the most heavily targeted by miRNA. Transcription factors are marked red. Barplots show the number of miRNAs targeting a particular mRNA in the corresponding tissue in Flight 3M or Flight 8M.

by miRNAs showing age-related changes in spaceflight, we identified nine transcription factors, including *Tead1*, *Rorb*, *Myb*, and *Meis1* (Figure 6.63).

The heatmap indicates that some mRNAs exhibit age-dependent patterns. We found that most miRNA targeting these mRNAs mainly came from the MAT, diaphragm, and pancreas of 3-month-old mice. For example, miRNAs from the MIR-15 and MIR-17 families in diaphragm targeted the transcription factors *Rorb* and *Myb* in 3- but not in 8-moth-old mice (Figur C.9). This suggests that miRNA regulation varies with age and results from multiple targeting mechanisms related to spaceflight. In summary, our research demonstrates that specific miRNAs, including those from the MIR-8, MIR-15, and MIR-154 families, exhibit age-dependent changes in the diaphragm, MAT, and pancreas as a response to spaceflight. Nonetheless, the overall patterns of miRNA deregulation imply a coordinated response to spaceflight stress independent of age. While there is some overlap with the signatures of natural aging, the changes triggered by spaceflight follow distinct regulatory paths. This underscores the importance of further research to clarify the balance between adaptive and degenerative processes across different tissues. These insights are vital for the context of prolonged space missions, as they enhance our understanding of age-dependent molecular responses.

6.6 DISCUSSION

This study provided a multimodal analysis of how spaceflight affects murine physiology using 686 small RNA-seq samples from 13 organs (bulk analysis) and single-cell RNA-seq samples across 28 tissues in mice from the ISS and matched ground controls. We aimed to characterize systemic vs. age-specific responses to spaceflight and the tissue adaptations that are linked to it.

We found that spaceflight causes mainly systemic and age-independent molecular effects and that age-specific effects are limited to specific cell types and tissues. The mRNA and miRNA datasets showed a substantial change in extracellular matrix (ECM) remodeling, DNA damage response, metabolic adaptation, and immune modulation. These changes largely reflect the stresses a spaceflight with microgravity, space radiation, and other environmental changes pose on the murine body.

HOUSING EFFECTS Housing factors showed the largest impact on Brain, Kidney, and MAT in our miRNA data. Even though the double-density housing has only a minimal effect on mice, factors like the elevated CO_2 levels and the temperature on board the ISS are known to affect the physiology and behavior in mice [392, 393] and impact their brain and kidney function [428, 429].

RIBOSOMAL PROTEIN GENES Our single-cell mRNA data revealed a strong downregulation of ribosomal protein genes, suggesting impaired cell growth and proliferation. These proteins are essential for RNA translation, and their reduced expression likely reflects cellular stress responses, metabolic reprogramming, or exposure to ionizing radiation [430] (mRNA). Similar ribosomal down-regulation has been observed under radiation and microgravity in Arabidopsis thaliana [431], rat muscle [432], and Drosophila melanogaster [433], while other organisms like E. coli [434] or human fibroblasts [435] have shown increased expression. Of note, sometimes under simulated microgravity.

EXTRACELLULAR MATRIX REMODELING AND TISSUE ADAPTATION At the same time, both mRNA and miRNA data consistently pointed to extracellular matrix (ECM) remodeling as a significant systemic effect. In the mRNA data, we observed spaceflight-induced deregulation of ECM genes such as *Ahnak* and *TPT1*, as well as the Collagen, Laminin, Pecam, Icam, and Thbs signaling pathways, all of which play key roles in tissue integrity and intercellular signaling. These alterations show similarity to age-related ECM degradation, suggesting shared mechanisms between space-induced [421, 436] and age-related [437] degeneration.

The miRNA data independently supported these findings. We identified tissue remodeling as a systemic miRNA-regulated process, with prominent affected GAT, SCAT, pancreas, and spleen. This remodeling likely resulted from mechanical unloading and metabolic shifts [438]. For example, we observed upregulation of miR-200b/a [439] (a member of the MIR-17 family) and miR-196a [440], both known to regulate adipocyte function and miR-196a also contributing to ECM reorganization, in the pancreas, GAT, and SCAT.

METABOLIC DYSREGULATION In addition to structural remodeling, our miRNA analysis revealed signals of metabolic dysregulation, particularly in adipose tissue and pancreas. These changes are consistent with previous findings of altered peripheral insulin sensitivity in spaceflight [441, 442]. The changes in the ECM organization might thus link these structural and metabolic pathways. Since pancreatic hormones regulate adipocyte activity and spaceflight increases pancreatic beta-cell mass [443], this points toward a complex endocrine adaptation involving tissue remodeling and metabolic shifts.

In the integrated analysis, miRNAs targeted ECM-associated mRNAs such as *ITGB8* [444] and *CRIM1* [445] in BAT, and *BSN* [446] and *TSPAN4* [447] in the liver. Systemically, members of the MIR-17 family regulated ECM and Wnt signaling in GAT, SCAT, and BAT, suggesting regulatory processes that affect both tissue structure and adipose differentiation [448, 449, 450] during a spaceflight.

CARDIOVASCULAR AND NEUROLOGICAL ADAPTATIONS Consistent with known effects of spaceflight on the cardiovascular system [398], we observed miRNA-mediated regulation of heart innervation, including changes in potassium ion homeostasis and postsynaptic membrane composition. miRNAs such as miR-133a-3p, miR-133b-3p, and miR-1a-3p (MIR-1/133 family), as well as miR-34c-5p and miR-34b-5p (MIR-34 family), were involved in these changes, suggesting a potential molecular cause for the arrhythmias reported in astronauts.

In the integrative analysis, we found synaptic membrane and neuronal function to be affected in the brain, indicated by both ECM disruption (mRNA) and deregulated miRNAs like miR-20b-5p (MIR-17 family), which is with Alzheimer's disease in humans [451]. These findings raise the possibility that spaceflight-induced metabolic stress contributes to cognitive impairment [452].

DNA DAMAGE AND STRESS RESPONSE Space radiation is known to cause DNA damage and cellular stress. In the miRNA data, this emerged as DNA metabolism and replication in multiple tissues involving miRNAs such

as miR-1a-3p and miR-206-3p (MIR-1 family), known to play a role in cancer [453].

In the integrated analysis, we observed the activation of DNA damage response pathways. This activation involved miRNAs such as miR-92b, miR-363-3p (MIR-17/92 cluster) [454], and LET-7 family members [380, 455], all previously implicated in radiation response, targeting key regulators like *PHF6*, a leukemia-associated gene in the thymus [456, 457], and *DGKE*, linked to hemolytic uremia [458].

Furthermore, the mRNA data showed disruptions in the bone marrow, indicating that spaceflight causes hematopoietic and immune remodeling at the stem cell level.

IMMUNE AND INFLAMMATORY REGULATION Immune regulation was the most prominent age-dependent effect detected in mRNA and the integrated data. Although most spaceflight effects were age-independent, immune regulation was a tissue-independent but cell type-specific change, especially in the mRNA data. We observed enrichment for inflammatory and immune pathways in epithelial, endothelial, progenitor, and immune cells.

miRNAs involved in these age-related immune effects included miR-15 (MIR-15), which targets *MYB*, a key transcription factor for NK cell maturation and muscle regeneration (miRNA) [459, 460]. Younger mice (3 months) displayed active regulation of developmental and immune pathways (e.g., *TNFSF11* (skeletal muscle development and function [461, 462]), *MYB*, *RORB* (transcription factors essential for stem cell and neuron development [459, 463])), while these interactions were absent or diminished in 8-month-old animals, indicating reduced adaptability with age.

Changes in the structure of the ECM are known to influence immune function and regulation [92], an effect that also appears in aging [93]. Observing ECM changes as a systemic effect in spaceflight suggests that along with the changes in immune regulation in endothelial, epithelial, stem/progenitor, and muscle cells, this might cause the observed immune disruptions.

COMPARISON OF AGE AND SPACEFLIGHT EFFECTS Despite the age-related changes we observed in our data, spaceflights show a much stronger influence on gene expression than age alone, both in the mRNA and the

miRNA data. While age effects were strongest in the pancreas, diaphragm, and MAT, spaceflight caused changes across nearly all tissues.

In the miRNA data, age differences also manifested in altered endocrine and immune responses. For example, *YIPF4*, involved in stress-related macroautophagy [464], was targeted in the pancreas of older mice, suggesting altered post-translational processing under combined age and space stress. Meanwhile, miR-8, miR-15, and miR-154 families showed age-specific deregulation across multiple organs.

Logistical challenges limit spaceflight research, including sample size, housing conditions, and the difficulty of determining the mice's age at launch. We must, therefore, interpret our findings in the context of controlled housing differences, which independently affect gene expression, particularly in the brain, kidney, and MAT. Despite these limitations, we prioritized robust signals observed across multiple organs, replicable across both mRNA and miRNA datasets.

Our study focused on female mice to avoid sex-based variability, but future work should also include sex-specific analyses. Aging biomarkers typically appear later in life, which limits our ability to detect aging effects in animals that are 8 months old. Nevertheless, the age range of these mice represents the usual demographics of astronauts.

By integrating small RNA and single-cell mRNA datasets, we revealed that spaceflight induces a systemic, largely age-independent deregulation of ECM, developmental, and DNA damage pathways. At the same time, immune regulation retains a degree of age sensitivity. A small set of miRNA families—MIR-17/92, MIR-1/133, MIR-34, and LET-7—emerged as key regulators across tissues and conditions.

These changes in mRNA and miRNA levels reflect systemic tissue and ECM remodeling, metabolic reprogramming, DNA damage responses, and (mainly age-dependent) immune modulation. Given the immune system's interconnected nature with tissue structure and metabolism, immune regulation emerges both as a potential regulator and a response to this remodeling of the mammalian body during spaceflight.

This dataset offers a valuable resource for understanding the biological impact of spaceflight and can inform future mission planning, countermeasure development, and studies on aging in altered gravity. Follow-up studies in simulated microgravity (e.g., endothelial cell cultures) may test whether controlling the found miRNA families can prevent spaceflight-induced structural changes.

7

DISCUSSION, OUTLOOK AND FUTURE DIRECTIONS

Understanding the immune system as a decentralized, tissue-embedded network is essential to uncover its role in systemic adaptation, disease development, and progression. We observed consistent patterns that extend beyond individual tissues or conditions across the projects presented in this thesis, including spaceflight-induced remodeling and immune involvement in neurodegeneration. These appear as systemic immune dysregulation, extracellular matrix (ECM) alterations, and metabolic adaptations in a sex- and age-dependent manner. Together, these findings highlight the importance of single-cell and small RNA sequencing for mapping transcriptomic changes into a systemic context [51].

SYSTEMS BIOLOGY VIEW ON THE IMMUNE SYSTEM

Our analyses support the idea that the immune system operates dynamically and systemically, responding to environmental signals, tissue contexts, and systemic stressors [F3, F1, F2]. Beyond their role in pathogen defense, immune cells are integral to tissue development and homeostasis [92, 22]. Specifically, transcriptomic data from spaceflight and neurodegeneration indicates that immune cells influence and are influenced by physiological shifts, including changes in the extracellular matrix, tissue metabolism, and barrier tissues [F3, F1, F2]. In neurodegeneration, PBMCs showed sex- and disease-specific transcriptional patterns, which partially align with changes found in the brain [F3]. This supports the idea that PBMCs serve as peripheral indicators of brain development and can be used for biomarker discovery [258, 332, 465]. Our sex-stratified analyses revealed variations in cell type composition and gene expression patterns, particularly in Parkinson's disease. These observations are consistent with reports of sex-specific deregulation of the immune system in neurodegeneration [262, 356, 86]. We found an age-dependent immune system dysregulation in the spaceflight studies,

although most spaceflight effects were age-independent [F1, F2]. This might be caused by changing immune adaptability with aging [466, 467]. These results highlight the necessity for a systems-level perspective on immunity, where immune responses influence and are influenced by local tissue conditions, age, and sex-based regulatory mechanisms [468].

SYSTEMIC CHANGES IN STRESS AND DISEASE

A central shared finding from both projects was the systemic deregulation of ECM-related genes. In our spaceflight data, ECM-associated genes and signaling pathways such as Collagen, Laminin, ICAM, and THBS showed consistent deregulation across multiple organs [F1, F2]. These observations are consistent with reports that microgravity and aging disrupt ECM integrity and intercellular signaling [421, 437, 438]. In neurodegeneration, we found similar ECM-related alterations in blood and brain, particularly in Alzheimer's disease [F3]. This confirms earlier studies linking ECM degradation to increased blood-brain barrier permeability and neuroinflammation [357, 469, 275]. Additionally, our single-cell mRNA data from the spaceflight project revealed the downregulation of ribosomal protein genes, suggesting reduced translational capacity and cell proliferation [F1, F2]. Such transcriptional signatures have previously been observed under microgravity or radiation exposure in other organisms, including Arabidopsis, Drosophila, and human cells [431, 432, 433, 434, 435].

We also observed miRNA-driven metabolic dysregulation, especially in adipose tissue and pancreas. These included upregulation of miR-200b/a and miR-196a, known regulators of adipocyte function, and ECM remodeling [440, 439]. Our integrated miRNA-mRNA analysis implicated MIR-17/92 and MIR-34 family members in regulating ECM and metabolic processes [450, 449, 444, 448], reinforcing the view that miRNA networks orchestrate multi-organ adaptation to systemic stress [F1, F2].

AGE-, SEX- AND SPECIES-SPECIFIC EFFECTS

One limitation we encountered is the limited number of longitudinal data. While we used RNA velocity to infer dynamic states, such models cannot substitute for the temporal resolution of longitudinal data, especially when trying to understand immune adaptations over time. In the case of disease development, like neurodegeneration, longitudinal data over the course of years can provide insights into the development and adaptations during disease progression. Considering age is thus crucial when analyzing immune cell data. For spaceflight, age comparisons showed that younger animals show more adaptive immune signatures [F1, F2]. These observations are consistent with known declines in adaptive immunity with age [466]. Even though we only used female mice in spaceflight, adaptations in spaceflight are also known to be sex-specific, a field that still requires further research. In neurodegeneration, we found sex-specific changes in both AD and PD. Female patients, for example, showed a higher number of DEGs, which aligned with prior work [F3][258, 262]. In addition to age and sex differences, a third variable must be accounted for: species differences. Even though the mouse is an essential model organism, especially for multi-organ RNA sequencing studies, environmental factors, as well as strain-specific physiology, influence the results [392, 393, 369], limiting the translatability to humans. Of note, although these analyses provide a broad overview of the mRNA and miRNA changes in the organisms and samples, they do not substitute for experimental validation. In order to explore potential treatments of the sex-specific effects in neurodegeneration, a detailed validation of the gene-expression changes in the affected cell types and clarification of the mechanistic implications is necessary. Similarly, tests in simulated microgravity or intervention experiments in real microgravity are necessary to clarify if accounting for ECM changes would mitigate some of the spaceflight effects that astronauts experience.

CHALLENGES OF SINGLE-CELL ANALYSIS

The integrative analysis of single-cell and miRNA data has enabled a systemic view of molecular changes under stress. To build on this work, unified

pipelines, including single-cell, miRNA, and spatial transcriptomics, would add spatial resolution (apart from the organ-level resolution). Similarly, sufficient longitudinal samples would allow us to capture changes in immune states over time and during disease development. Extending our understanding of the spatial and longitudinal aspects of immune cell interactions would allow us to understand the processes within the different tissues better. So far, this work integrates single-cell data with scalable computational tools and provides a framework for future research into systemic immune adaptations and their role in health and disease.

As single-cell datasets scale into millions of cells, computational and methodological challenges limit their utility. Memory constraints, long runtime, and the modularized pipelines create bottlenecks for large datasets [4].

Another major challenge is the cell type annotation. Differences in the resolution, the annotation strategy, and the reference datasets lead to inconsistent labels across studies, leading to poor reproducibility and comparability [185]. Furthermore, manual annotation is still state-of-the-art for large-scale datasets and fine-resolution annotations since most cell type annotations cannot be easily transferred between datasets or organisms. These lack reproducibility and are a time-intensive task. Improving the annotation and making it more robust to noise, batch effects, and different experimental conditions requires further research. Larger and harmonized datasets are necessary to train models for predictions of cell types reliably, but also immune states in disease [192, 470].

Technical noise further complicates analysis cross-study analysis. To address these challenges, standardization and harmonization projects such as the Human Cell Atlas (HCA) [192] and guidelines such as the FAIR data principles (Findable, Accessible, Interoperable, Reusable) are essential. Using best data sharing, annotation, and interoperability practices across tools and platforms is essential for large-scale and reproducible single-cell research.

BUILDING A HARMONIZED PBMC ATLAS

We thus created a single-cell PBMC atlas called PBMCpedia (Planned Publication[1]), collecting all openly accessible PBMC datasets sequenced on a 10X sequencer and performing uniform pre-processing of all datasets. The current version includes 4.29 million cells from 28 studies, integrating over 500 samples across 14 diseases and healthy controls. PBMCpedia provides harmonized gene expression data, differential expression and pathway analysis results, and standardized cell type annotations based on the Human Immune Health Atlas [12]. This resource improves the reproducibility of results by eliminating study-specific effects, thus promoting cross-study analysis.

CONCLUSION AND OUTLOOK

This thesis illustrates how immune regulation integrates local tissue states with systemic stress responses, using miRNA and single-cell RNA sequencing data from spaceflight and neurodegeneration. Combining single-cell and sncRNA-seq data uncovered recurring signatures of immune regulation, tissue remodeling, and metabolic adaptation across tissues, conditions, and demographic groups. We applied computational methods and developed scalable, reproducible workflows for complex, multi-modal datasets. It highlights the importance of harmonized data structures, cross-study comparability, and integrative data analysis for understanding immune behavior across biological contexts.

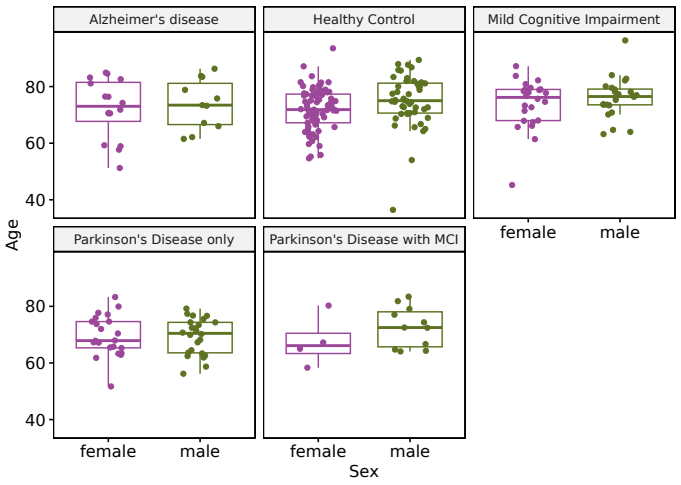
"The purpose of computation is insight, not numbers." — Richard Hamming

Based on this principle, the thesis bridges experimental biology and computational analysis, highlighting how a systems-level view can guide biological understanding in immunology.

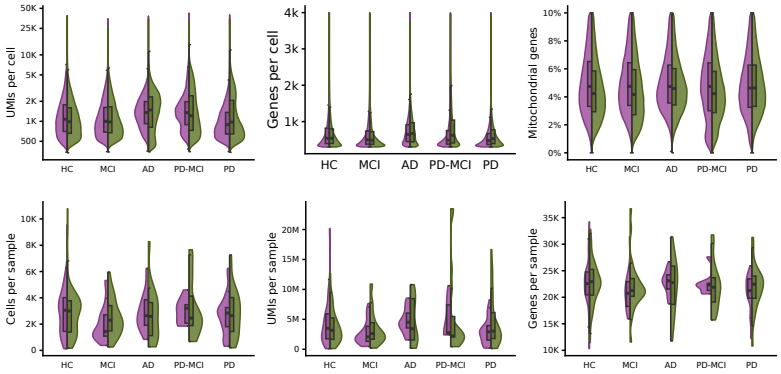
[1] Emma Hoffmann, Friederike Grandke, Lara Hombrecher, Ian Diks, Matthias Flotho, Andreas Keller. PBMCpedia: A Standardized PBMC Single-Cell RNA-seq Database with Unified Mapping and Enhanced Cell Type Annotation.

A

PBMC IN NEURODEGENERATION - SUPPLEMENTARY INFORMATION



(A) Age-distribution per disease and sex.



(B) Quality-control measurements per patient group and sex

FIGURE A.1: Sex- and age-distribution in the dataset.

		Age in females (mean, sd and number of patients)			Age in males (mean, sd and number of patients)			adj. p-value (m vs.f)		
Diagnosis	HC	$72 \pm 7 (n = 82)$			$75 \pm 9 (n = 53)$			0.318		
	AD	$71 \pm 10 (n = 21)$			$73 \pm 8 (n = 13)$			1		
	PD	$69 \pm 7 (n = 23)$			$69 \pm 6 (n = 25)$			1		
	MCI	$73 \pm 8 (n = 29)$			$76 \pm 6 (n = 26)$			1		
	PD-MCI	$68 \pm 7 (n = 6)$			$73 \pm 6 (n = 12)$			1		
Race	White (w)	$71 \pm 8 (n = 144)$			$74 \pm 7 (n = 116)$			0.101		
	Asian (a)	$72 \pm 7 (n = 12)$			$72 \pm 14 (n = 9)$			1		
	Other (o)	$73 \pm 7 (n = 5)$			$74 \pm 7 (n = 4)$			1		
	Race	w	a	o	w	a	o	w	a	o
Diagnosis	HC	$72 \pm 7 (n = 72)$	$71 \pm 6 (n = 6)$	$71 \pm 6 (n = 4)$	$76 \pm 7 (n = 49)$	$64 \pm 7 (n = 49)$	$-(n = 0)$	0.148	1	-
	AD	$70 \pm 10 (n = 18)$	$75 \pm 8 (n = 3)$	$-(n = 0)$	$73 \pm 8 (n = 13)$	$-(n = 0)$	$-(n = 0)$	1	-	-
	PD	$69 \pm 7 (n = 22)$	$67 - (n = 1)$	$-(n = 0)$	$68 \pm 6 (n = 23)$	$73 \pm 4 (n = 2)$	$-(n = 0)$	1	-	-
	MCI	$73 \pm 8 (n = 26)$	$69 \pm 11 (n = 2)$	$83 - (n = 1)$	$75 \pm 6 (n = 20)$	$83 \pm 1 (n = 2)$	$74 \pm 7 (n = 4)$	1	-	-
	PD-MCI	$68 \pm 7 (n = 6)$	$-(n = 0)$	$-(n = 0)$	$72 \pm 7 (n = 11)$	$78 - (n = 1)$	$-(n = 0)$	1	-	-

TABLE A.1: Patient demographics.

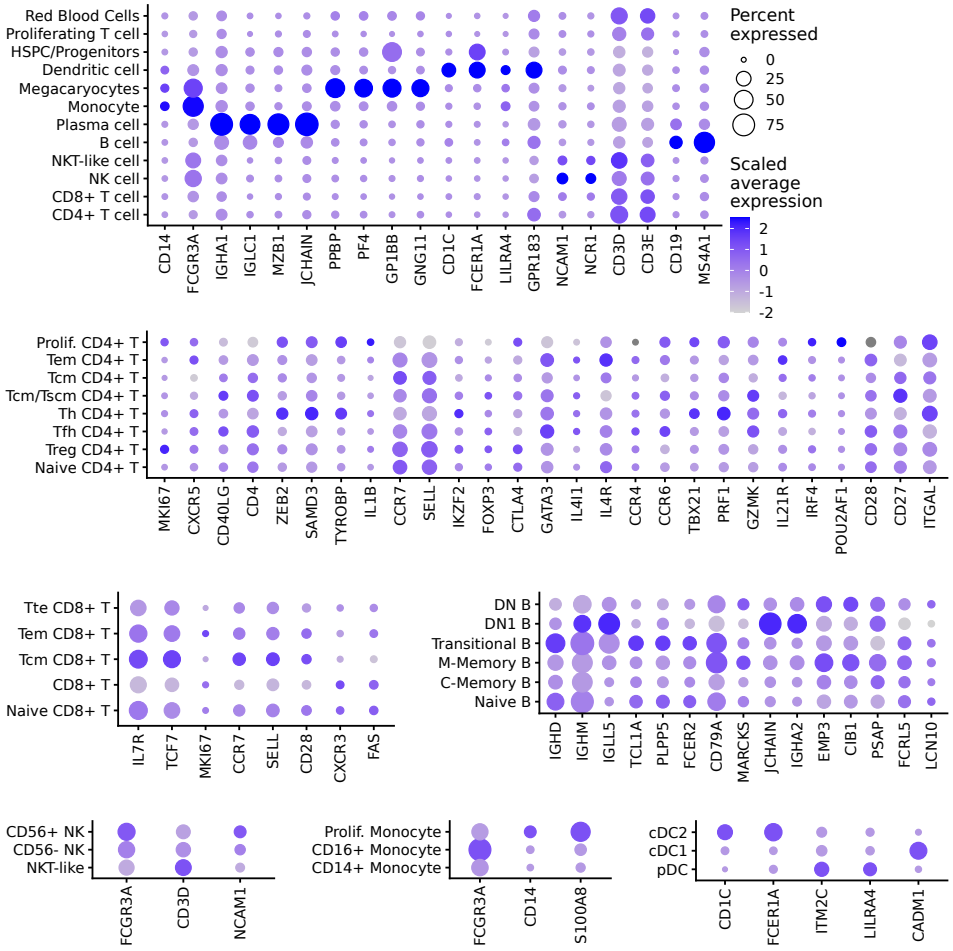
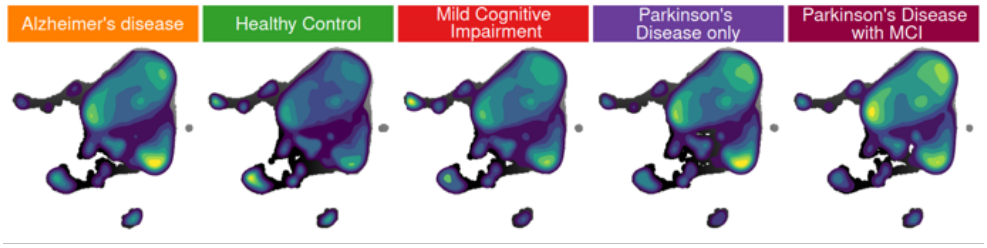
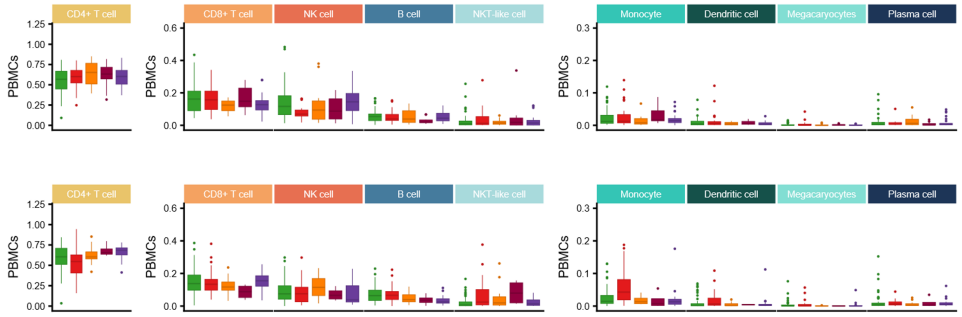


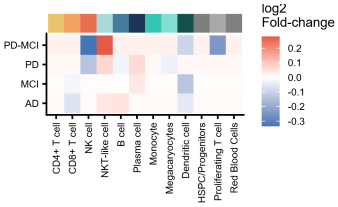
FIGURE A.2: Expression of cell type markers used for manual annotation.



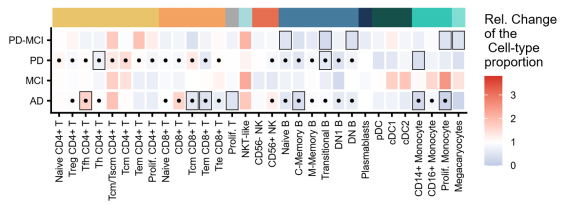
(A) Density UMAP embedding shows high (yellow) and low density (blue) areas.



(B) Cell type proportions per sample disease in males (top) and females (bottom).



(C) Fold changes using scCODA in both male and female samples.



(D) Relative change in the cell type-proportion in % between HC and disease as in Figure 5.8.

FIGURE A.3: Changes in the cell type composition.

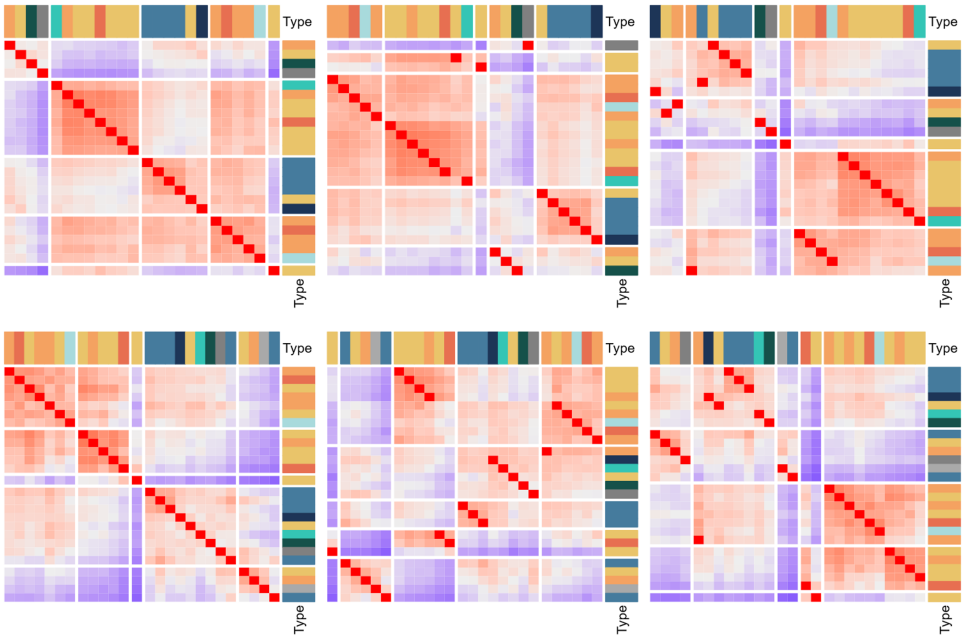
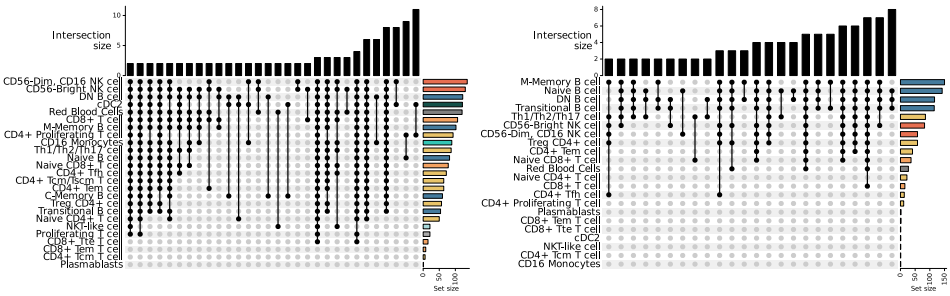
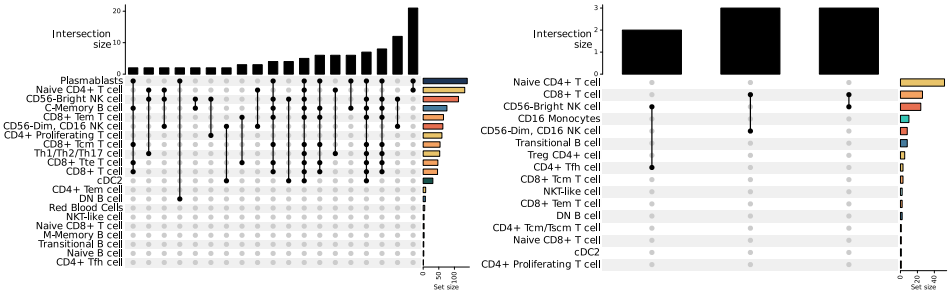


FIGURE A.4: Cosine similarity of deregulation (gene-lists ordered by fold-change) for female (top) and male (bottom) in AD vs. HC (left), PD vs. HC (middle) and MCI vs. HC (right) between different cell types.



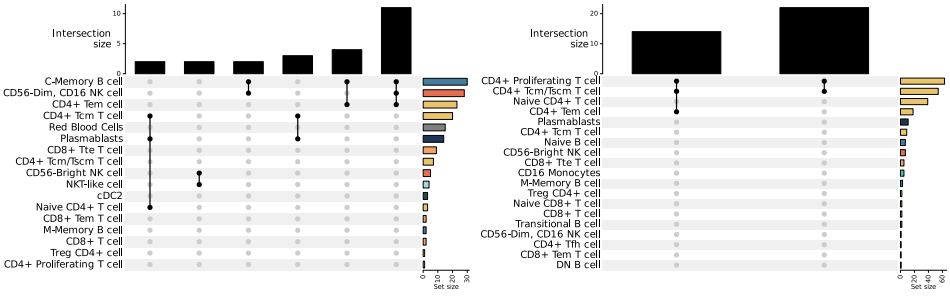
(A) AD - male

(B) AD - female



(C) PD - male

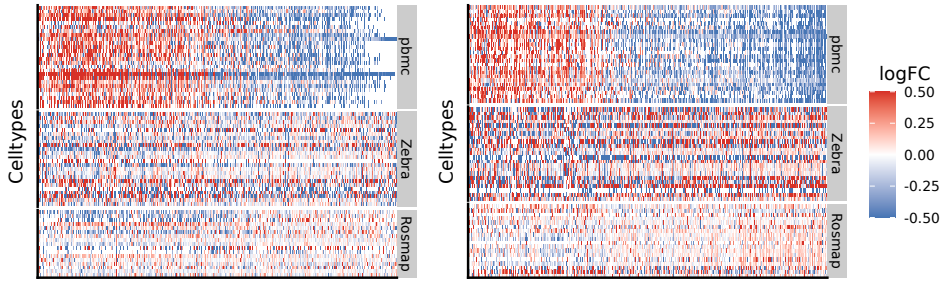
(D) PD - female



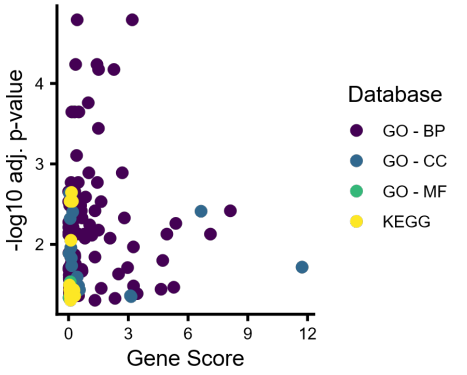
(E) MCI - male

(F) MCI - female

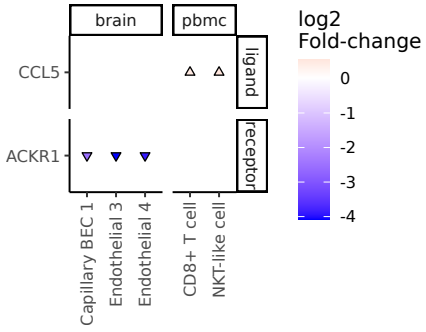
FIGURE A.5: Overlap of the enriched and depleted pathways in the different cell types.



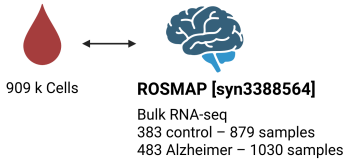
(A) Fold-changes of all DEGs in HC vs, AD in the brain (ZEBRA dataset) and PBMCs for male (left) and female (right).



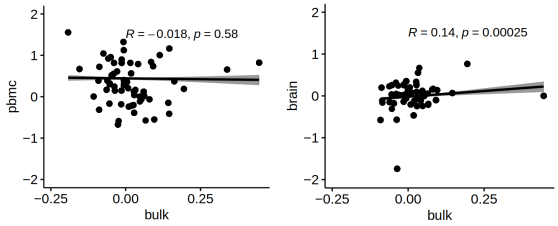
(B) Pathway analysis of the DEG overlap between PBMCs and brain cells.



(C) DEGs involved in the CCL signaling pathways in females.



(D) Bulk RNA-sequencing dataset Created with biorender.com.

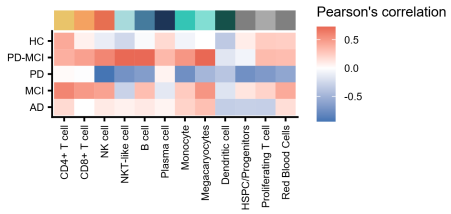


(E) Comparison of fold-changes found in single-cell data of the brain and blood with changes found in bulk data of male patients.

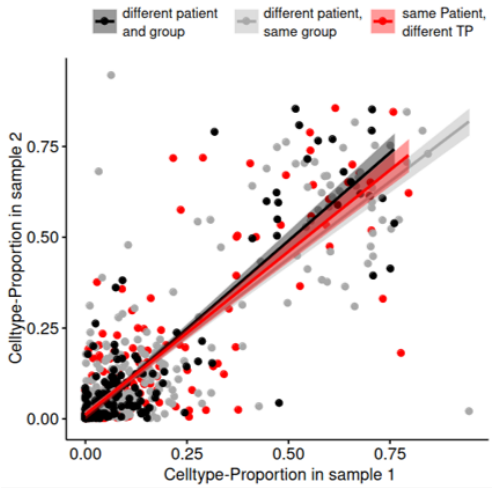
FIGURE A.6: Transcriptome patterns in brain and PBMCs.

Gene	Category	Reference
TRG-AS1	AD-general	[471]
ZNF212	-	-
GON4L	-	-
IQGAP2	BBB and Immune	[472]
CYTOR	Not reported	-
CXCR4	AD-general	[473]
CD74	Immune response	[474, 475]
CD37	Plaques	[476]
FOSB	Cognition	[477]
RBM3	-	-
PMAIP1	Mitochondrium	[478]
ITGB2-AS1	Inflammation	[479]
CYTH4	-	-
LILRB1	Microglia	[480]
IFI16	Immune response	[481]
HLA-DQB1	AD-general	[292]
RHOH	Microglia	[482]
ARL4C	-	-
HBB, HBA1, HBA2	AD-general	[483]
SLAMF7	-	-
SCML4	-	-
VIM	Astrocytes	[484]
ANXA1	Inflammation	[485]
MCUB	Mitochondrium	[486]
JUNB	-	-
FAU	-	-
RPL18	Ribosome	[487]
LYN	AD-general	[488, 489]
RRBP1	AD-general	[489]
FTH1, SAT1	Astrocytic Ferroptosis	[290]
ATP5MC3	AD-general	KEGG AD pathway
FYB1	AD-general	[490]
NBPF14	-	-
LGALS1	Microglia/Immune	[491]
CTSW	Immune response	[492, 493]
CD7	Immune response	[494]

TABLE A.2: Previously reports of the genes from Figure 5.22 in AD.



(A) Correlation of cell type proportions with the number of the visit (no significant results).



(B) Comparison of cell type proportions between patients, time-points and diagnosis.

FIGURE A.7: Cell-type composition across time points.

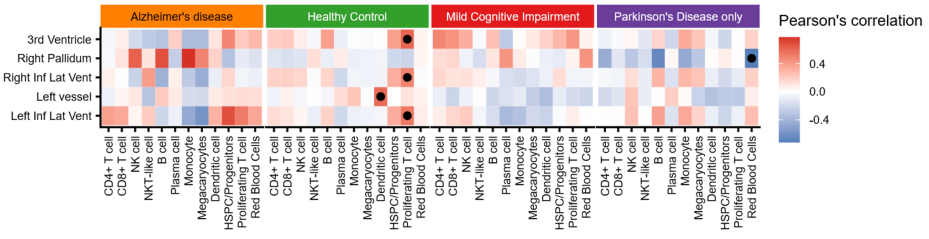


FIGURE A.8: Correlation between the CSF and brain data and the cell type proportions in the single-cell data. Significant correlations are marked with a dot.

B

SPACEFLIGHT STUDY: SINGLE-CELL - SUPPLEMENTARY INFORMATION

Step	Reagent	Catalog Nr.	Supplier
Collection	K2/EDTA-coated syringes and MiniCollect tubes		Greiner Bio-One
Red Blood Cell depletion	Histopaque-1119	RNBF0417	SigmaAldrich
Resuspension medium	2% sterile FBS in PBS + 100 U/mL DNase I	LS006344	Worthington

TABLE B.1: Tissue preparation of the Blood samples.

Tissue	Digestion	Mechanical dissociation of single cells
optic nerves and retina	Papain (15 min) → resuspended in HBSS	pipetting after digestion
corneal tissue	Dispase II (1 hour, with shaking) → 0.25% trypsin (10 mins)	pipetting after digestion
lens tissue	0.02% EDTA for 10 mins	pipetting after digestion
extraocular muscles	1mg/ml collagenase for 45 mins	pipetting every 15 minutes

TABLE B.2: Tissue digestion methods used for the Eye.

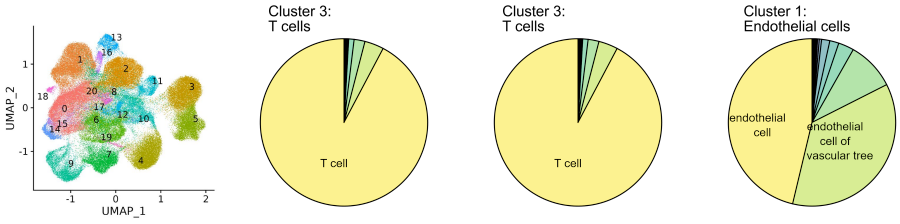


FIGURE B.1: Cell groups within cluster 1 (left), 3 (middle) and 9 (right).

Step	Reagent	Dilution ratio	Catalog Nr.	Supplier
Digestion	M199		11150067	Thermo Fisher Scientific
	collagenase II buffer	2.2mg/ml	C6885	Sigma-Aldrich
fluorochrome-conjugated antibodies	CD45	1:200	15-0451	Thermo Fisher Scientific
	TER119	1:200	15-5921	
	CD51	1:50	12-0512	
	Tie2	1:20	14-5987	
	CD31	1:200	11-0311	
	Th1.1	1:100	47-0900	
	Thy1.2	1:100	47-0902	
	6c3	1:100	17-5891	
	CD105	1:50	13-1051	
	streptavidin–PE–Cy7 conjugate	1:100	25-4317-82	

TABLE B.3: Tissue preparation of the bone samples.

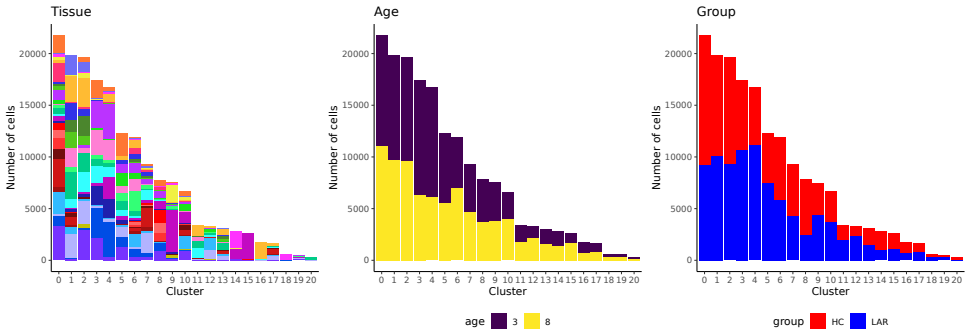


FIGURE B.2: Cell level composition of the dataset for tissue (left), age (middle) and condition (right) in the 20 clusters of the joined embedding.

Step	Kit/Instrument	Catalog Nr.	Supplier
Cell encapsulation	GemCode Single-Cell Instrument		10x Genomics
	GemCode Single-Cell 3′ Gel Bead and Library v3 Kit		10x Genomics
Enzymatic reactions (Thermal cycler)	C1000 Touch Thermal cycler with 96-Deep Well Reaction Module		Biorad
Fragment Analyzer	Fragment Analyzer		AATI
	High Sensitivity NGS Analysis Kit		Advanced Analytical
qPCR	Kapa Library Quantification kit		Illumina
Sequencing:	100 cycle run kits - (read 1: 28 bases; index 1: 8 bases; read 2: 91 bases)	20012862	Illumina
	PhiX control library		
	NovaSeq 6000 Sequencing System		Illumina

TABLE B.4: Microfluidic droplet single-cell sequencing.

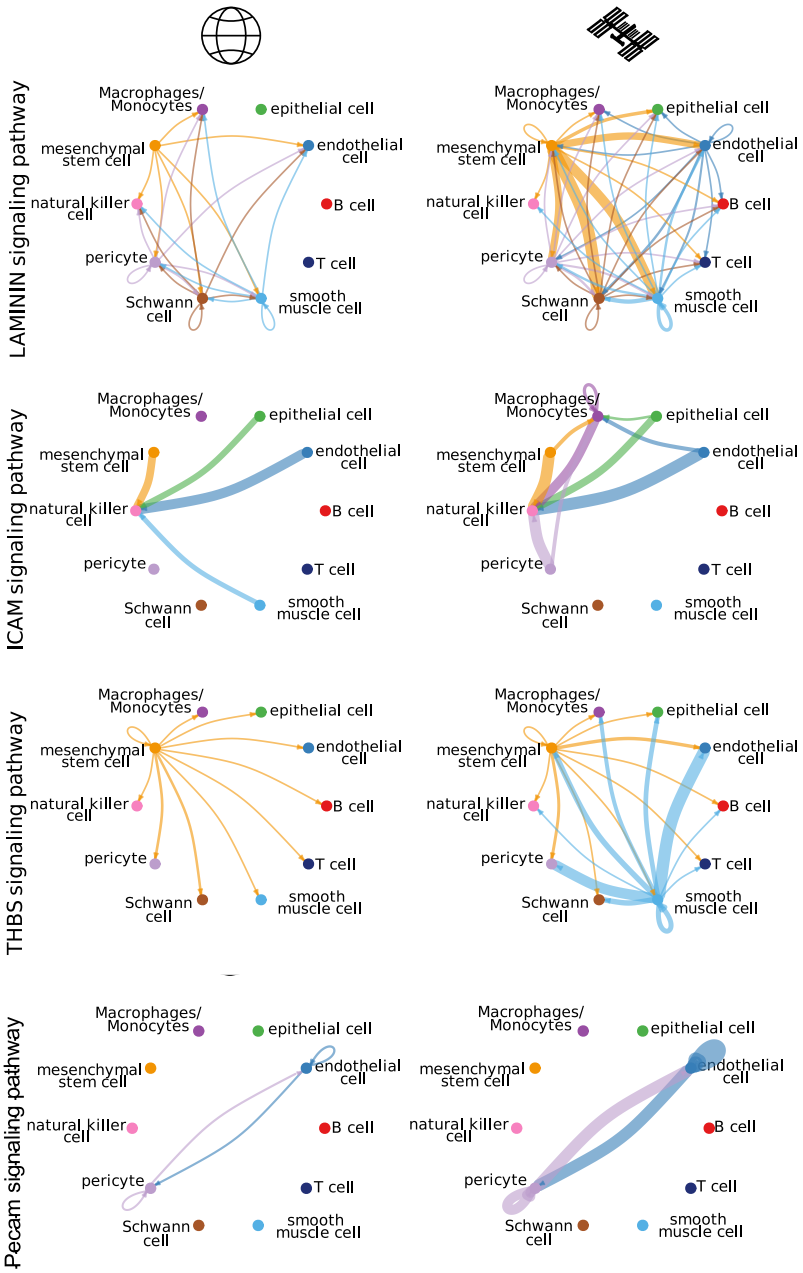


FIGURE B.3: Changes in the Laminin, Icam, Thbs and Pecam signaling pathway in SCAT after a spaceflight.

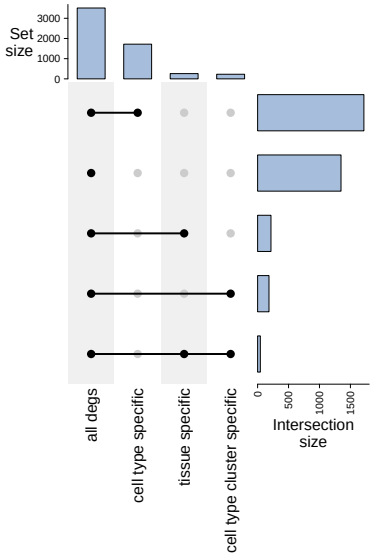


FIGURE B.4: Number of significantly deregulated genes (all genes) and genes that are cell type, cell type cluster, or tissue-specific.

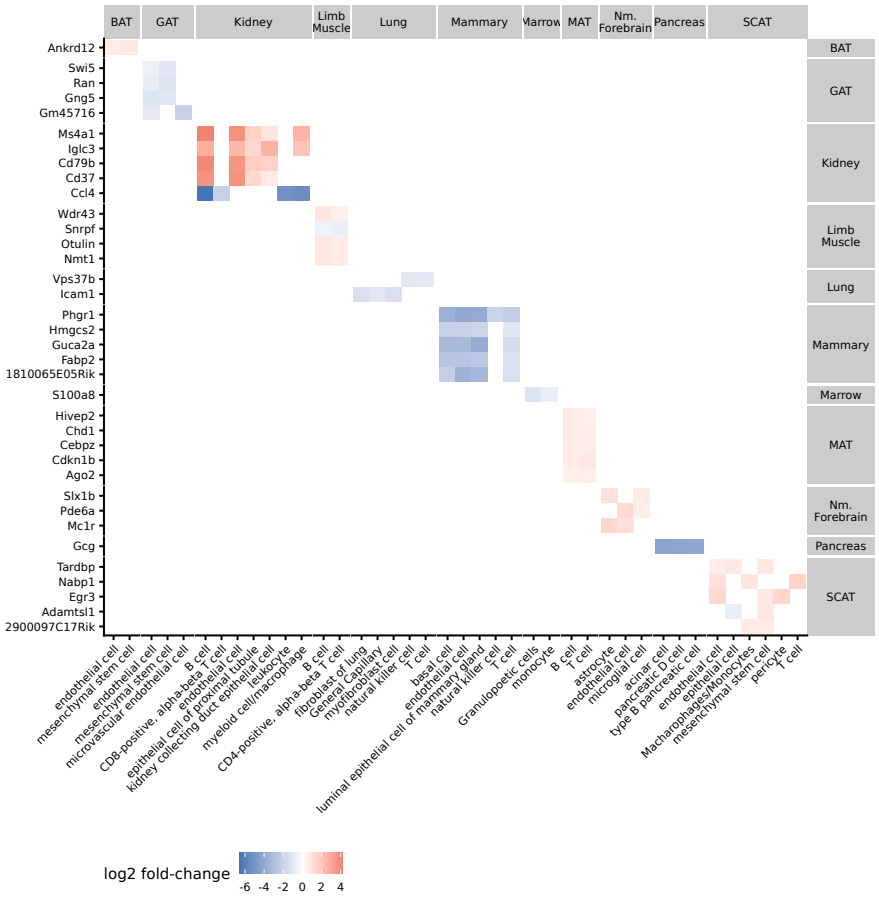


FIGURE B.5: Top 5 most frequent tissue-specific genes.

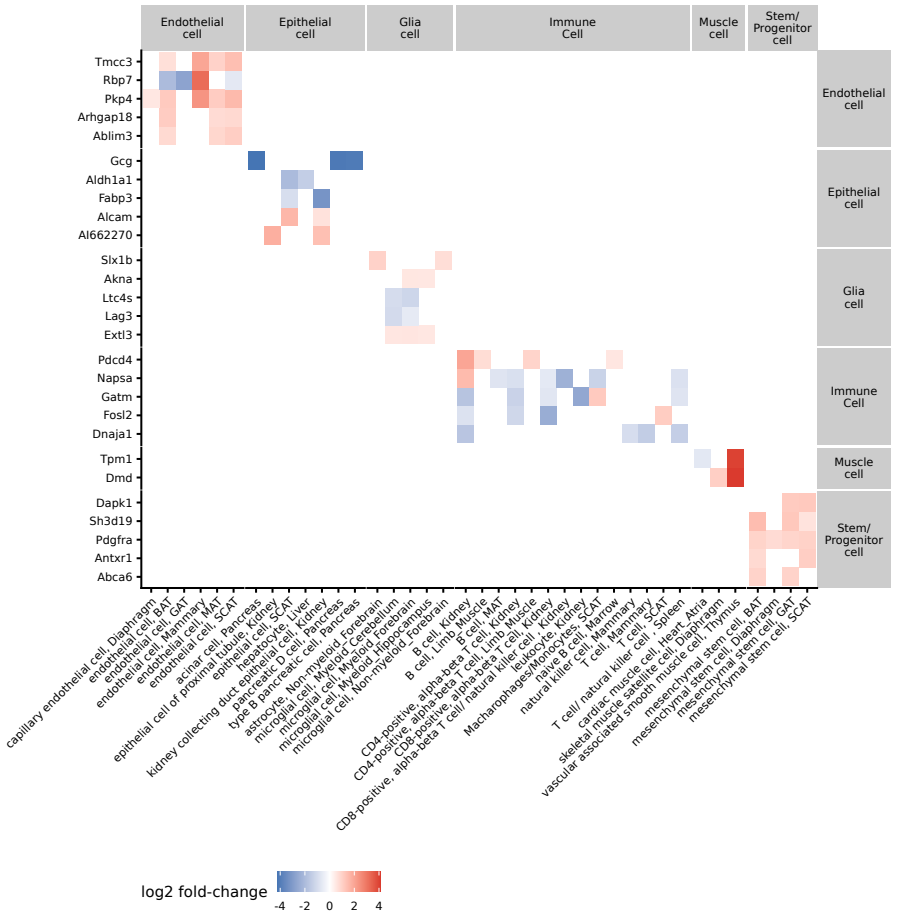


FIGURE B.6: Top 5 most frequent cell type cluster-specific genes.

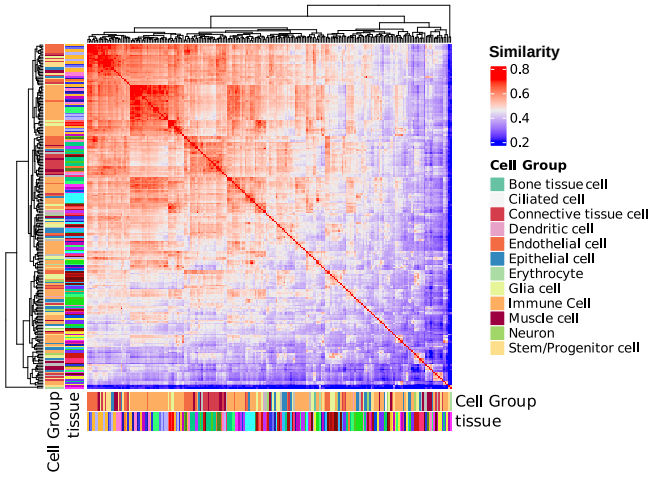


FIGURE B.7: Pairwise cosine similarity between clusters in the dataset. Cell groups were based on the Cell Ontology.

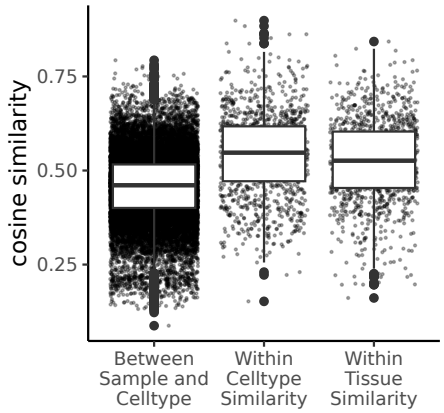


FIGURE B.8: Cosine similarities within and across cell types and tissues.

C

SPACEFLIGHT STUDY: MIRNA - SUPPLEMENTARY INFORMATION

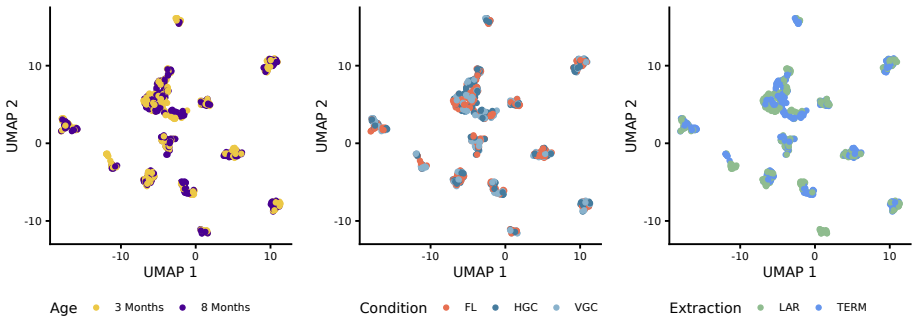


FIGURE C.1: UMAP embedding colored by Age, Condition and Extraction.

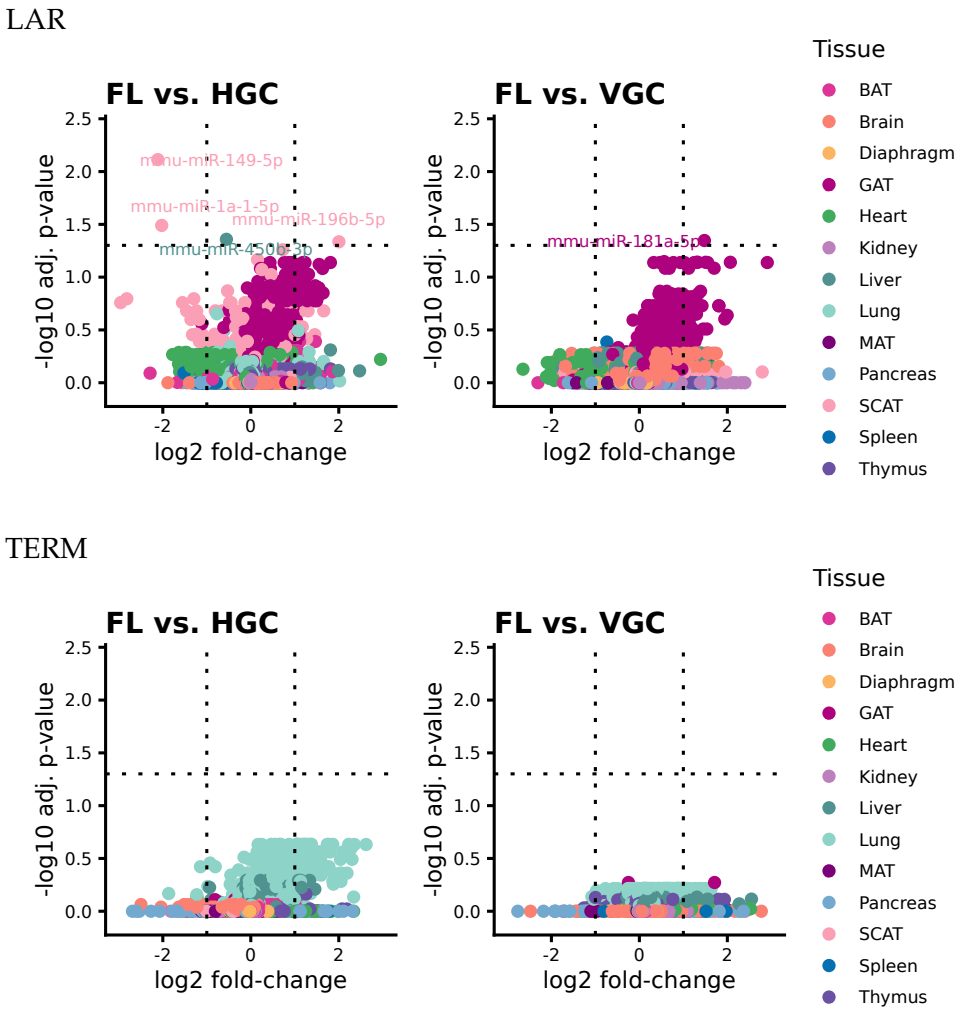


FIGURE C.2: Volcano plot of FL vs. HGC and FL vs. VGC in LAR and TERM.

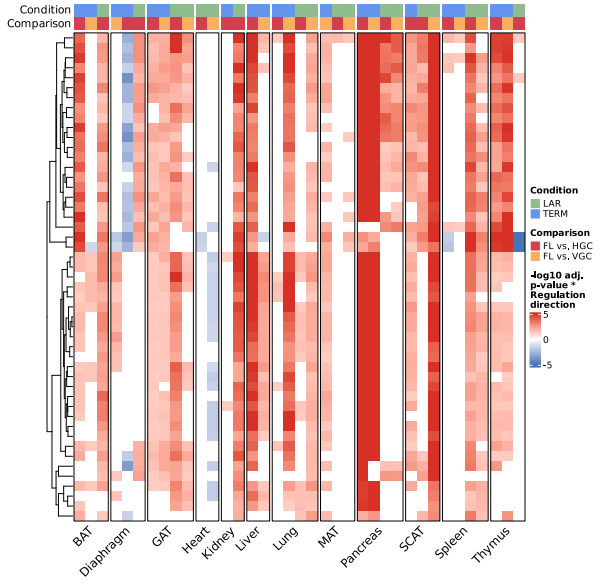


FIGURE C.3: Top 25 pathways from figure 6.40 (LAR) and 6.44 (TERM).

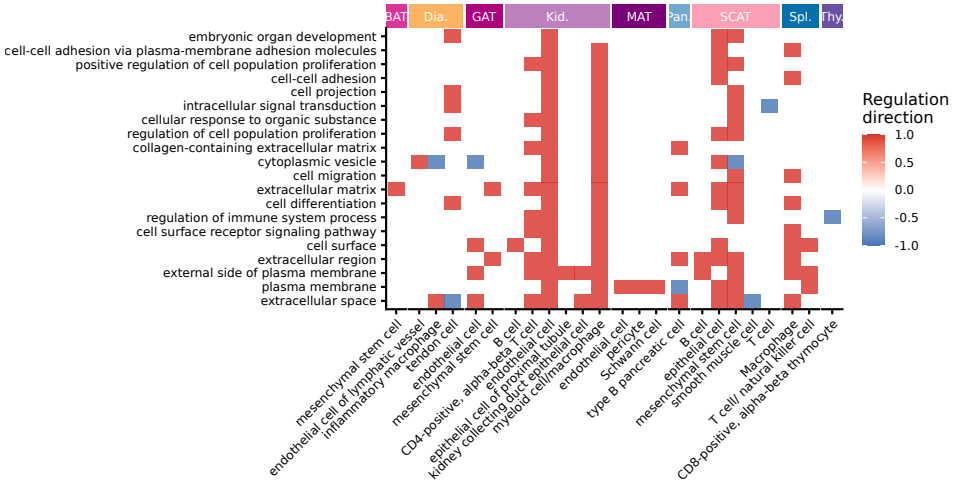


FIGURE C.4: Top 20 overlapping pathways in LAR between the miRNA and single-cell mRNA with the same direction of enrichment/depletion in $\geqslant 1$ tissue.

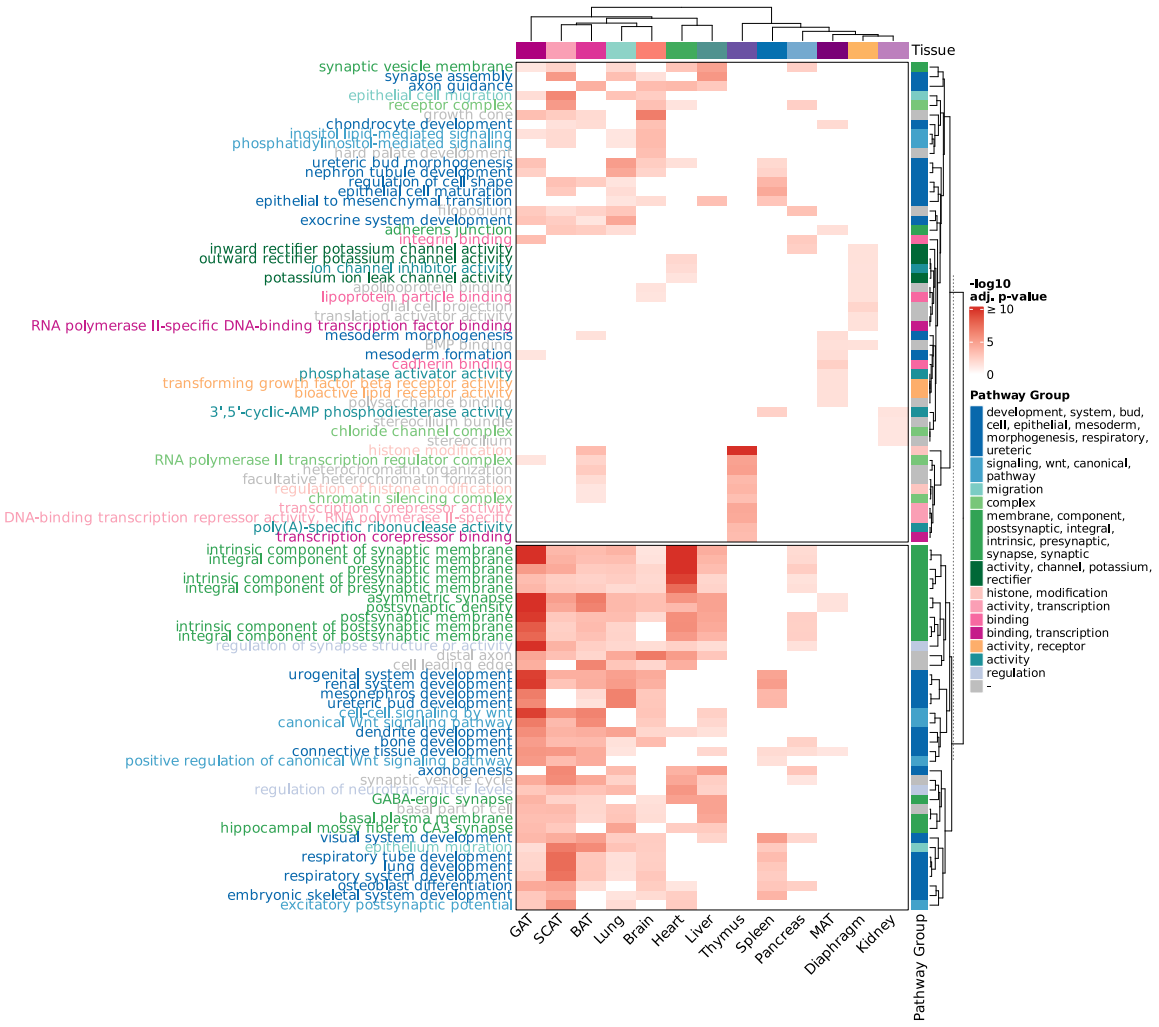


FIGURE C.5: Pathway analysis from Figure 6.46 in detail.

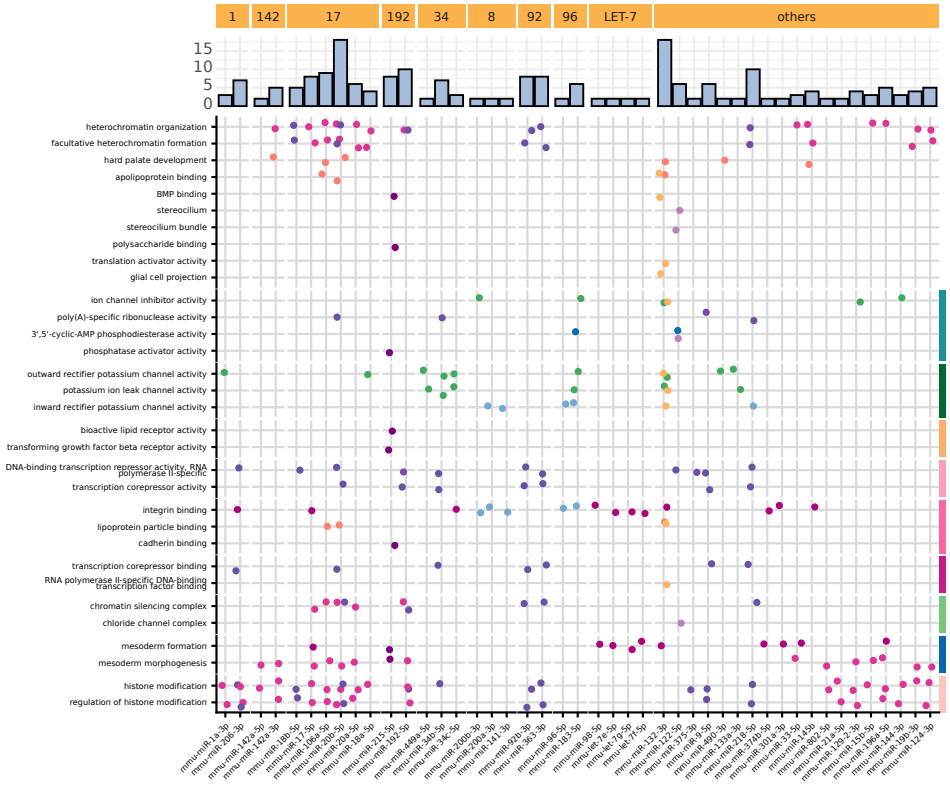


FIGURE C.6: The Cluster 1 pathways from Figure 6.46 ($\leqslant 2$ tissues) and the miRNAs that regulate their mRNAs (< 1 pathway).

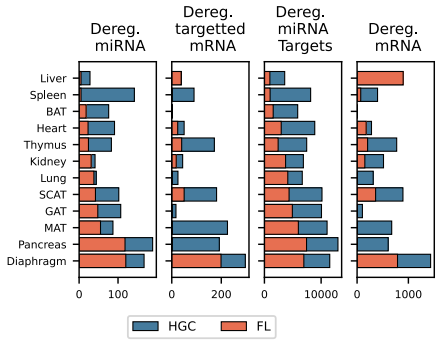


FIGURE C.7: Comparison of deregulated miRNA and targeted, deregulated mRNA for each tissue.

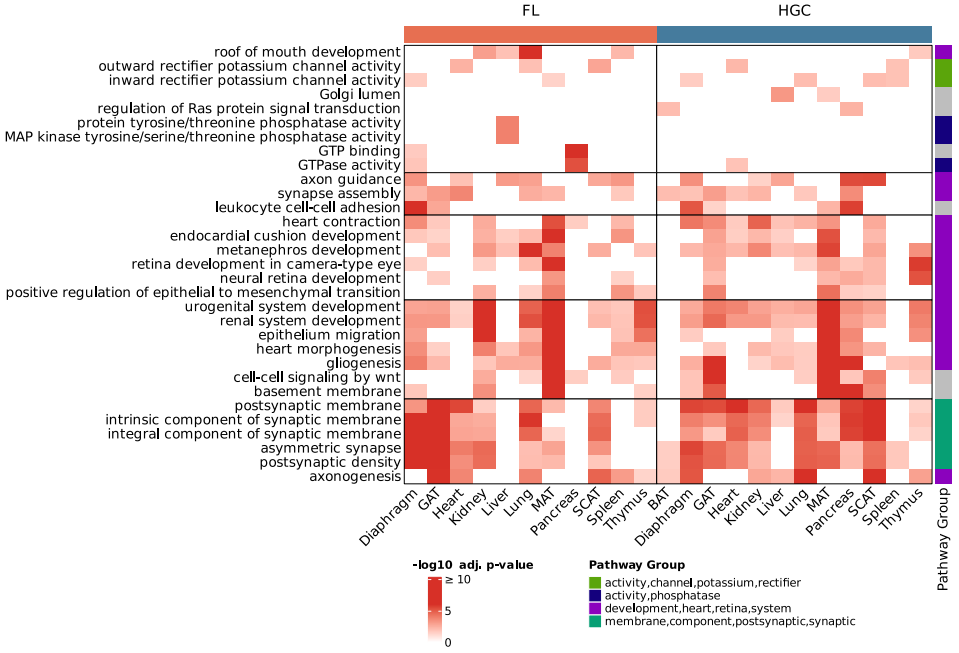


FIGURE C.8: Pathway analysis of the mRNA-miRNA overlap that is deregulated in 3M vs. 8M in HGC and show the same direction of fold-change in 3M vs. 8M in both HGC and in VGC.

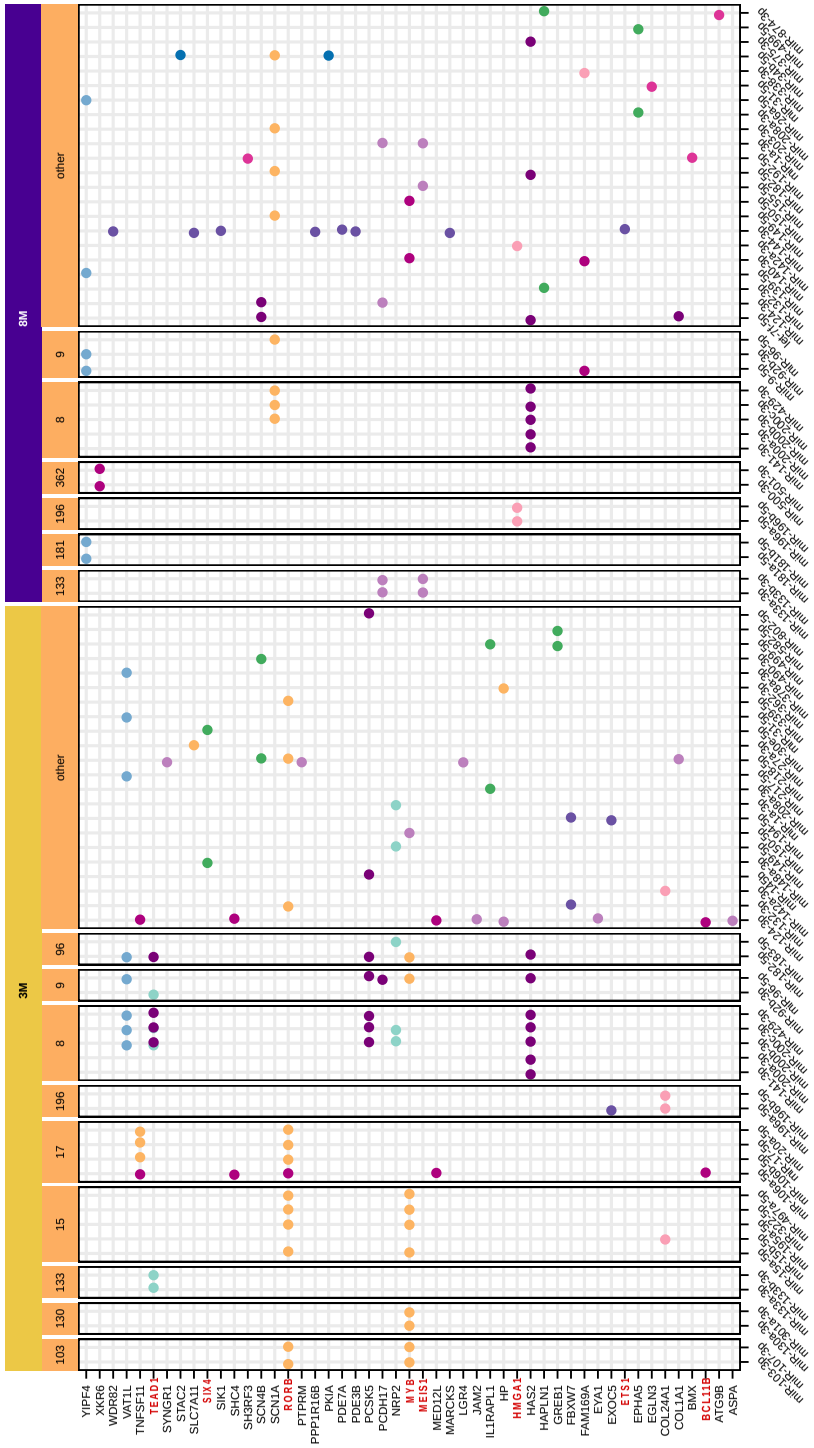


FIGURE C.9: Detailed targeting of the miRNA-mRNA pairs included in Figure 6.63 (as in 6.49).

PERSONAL CONTRIBUTIONS

[F1] **Friederike Grandke**, Nicholas Schaum, Fabian Kern, Jérémy Amand, Thomas H. Ambrosi, Teni Anbarchian, Karl Annusver, Reza Ashrafi, Kruti Calcuttawala, Shubhangi Das Barman, et al. Spaceflight induces systemic effects on extracellular matrix and immune system in different age-stages. *preprint*, 2025.

[F2] **Friederike Grandke***, Shusruto* Rishik, Viktoria Wagner, Annika Engel, Nicole Ludwig, Kruti Calcuttawala, Fabian Kern, Verena Keller, Marcin Krawczyk, Louis Stodieck, et al. Mirnas shape mouse age-independent tissue adaptation to spaceflight via ecm and developmental pathways. *preprint*, 2025.

[F3] **Friederike Grandke**, Tobias Fehlmann, Fabian Kern, David M Gate, Tobias William Wolff, Olivia Leventhal, Divya Channappa, Pascal Hirsch, Edward N Wilson, Eckart Meese, et al. A single-cell atlas to map sex-specific gene-expression changes in blood upon neurodegeneration. *Nature communications*, 16(1):1965, 2025.

[F4] Matthias Flotho, Jérémy Amand, Pascal Hirsch, **Friederike Grandke**, Tony Wyss-Coray, Andreas Keller, Fabian Kern. Zebra: a hierarchically integrated gene expression atlas of the murine and human brain at single-cell resolution. *Nucleic Acids Research*, 52(D1):D1089–D1096, 2024.

[F5] Annika Engel, Nicole Ludwig, **Friederike Grandke**, Viktoria Wagner, Fabian Kern, Tobias Fehlmann, Georges P Schmartz, Ernesto Aparicio-Puerta, Dominic Henn, Barbara Walch-Rückheim, et al. Skin treatment with non-thermal plasma modulates the immune system through mir-223-3p and its target genes. *RNA biology*, 21(1):31–44, 2024.

[F6] Shusruto Rishik, Pascal Hirsch, **Friederike Grandke**, Tobias Fehlmann, Andreas Keller. mirnatissueatlas 2025: an update to the uniformly processed and annotated human and mouse non-coding rna tissue atlas. *Nucleic Acids Research*, 53(D1):D129–D137, 2025.

[F7] **Friederike Grandke**, Tobias Fehlmann, Matthias Flotho, Tobias W. Wolff. A single-cell atlas to map sex-specific gene-expression changes in blood upon neurodegeneration, 2025. doi:10.5281/zenodo.14641642.

BIBLIOGRAPHY

[1] F. Sanger, S. Nicklen, A. R. Coulson. Dna sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12):5463–5467, 1977. doi:10.1073/pnas.74.12.5463.

[2] Fernanda Moraes, Andréa Góes. A decade of human genome project conclusion: Scientific diffusion about our genome knowledge. *Biochemistry and Molecular Biology Education*, 44(3):215–223, 2016. doi:10.1002/bmb.20952.

[3] The cost of sequencing a human genome. 11.04.2025. URL: https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost.

[4] Valentine Svensson, Roser Vento-Tormo, Sarah A Teichmann. Exponential scaling of single-cell rna-seq in the past decade. *Nature Protocols*, 13(4):599–604, 2018. arXiv:1704.01379, doi:10.1038/nprot.2017.149.

[5] Sra database growth. 09.04.2025. URL: https://www.ncbi.nlm.nih.gov/sra/docs/sragrowth/.

[6] Sra database. 07.06..2025. URL: https://www.ncbi.nlm.nih.gov/sra.

[7] Elixir. 30.05.2025. URL: https://elixir-europe.org/, doi:10.12688/f1000research.channels.151.

[8] Ncbi and nih. 30.05.2025. URL: https://www.ncbi.nlm.nih.gov/.

[9] Ddbj. 30.05.2025. URL: https://www.ddbj.nig.ac.jp/index-e.html.

[10] Ngdc. 30.05.2025. URL: https://ngdc.cncb.ac.cn/.

[11] Kimberly Siletti, Rebecca Hodge, Alejandro Mossi Albiach, Ka Wai Lee, Song-Lin Ding, Lijuan Hu, Peter Lönnerberg, Trygve Bakken, Tamara Casper, Michael Clark, et al. Transcriptomic diversity of cell types across the adult human brain. *Science*, 382(6667):eadd7046, 2023. doi:10.1126/science.add7046.

[12] Qiuyu Gong, Mehul Sharma, Emma L. Kuan, Marla C. Glass, Aishwarya Chander, Mansi Singh, Lucas T. Graybuck, Zachary J. Thomson, Christian M. LaFrance, Samir Rachid Zaim, et al. Longitudinal multi-omic immune profiling reveals age-related immune cell dynamics in healthy adults. *bioRxiv*, 2024. doi:10.1101/2024.09.10.612119.

[13] David Lähnemann, Johannes Köster, Ewa Szczurek, Davis J. McCarthy, Stephanie C. Hicks, Mark D. Robinson, Catalina A. Vallejos, Kieran R. Campbell, Niko Beerenwinkel, Ahmed Mahfouz, et al. Eleven grand challenges in single-cell data science. *Genome Biology*, 21(1):1–35, 2020. doi:10.1186/s13059-020-1926-6.

[14] Byungjin Hwang, Ji Hyun Lee, Duhee Bang. Single-cell rna sequencing technologies and bioinformatics pipelines. *Experimental & Molecular Medicine*, 50(8):1–14, 2018. doi:10.1038/s12276-018-0071-8.

[15] Lukas Heumos, Anna C. Schaar, Christopher Lance, Anastasia Litinetskaya, Felix Drost, Luke Zappia, Malte D. Lücken, Daniel C. Strobl, Juan Henao, Fabiola Curion, et al. Best practices for single-cell analysis across modalities. *Nature Reviews Genetics*, 24(8):550–572, 2023. doi:10.1038/s41576-023-00586-w.

[16] Efthymia Papalexi, Rahul Satija. Single-cell rna sequencing to explore immune cell heterogeneity. *Nature Reviews Immunology*, 18(1):35–45, 2017. doi:10.1038/nri.2017.76.

[17] Susanne Sattler. *The Role of the Immune System Beyond the Fight Against Infection*, pages 3–14. Springer International Publishing, 2017. doi:10.1007/978-3-319-57613-8_1.

[18] Kenneth Murphy, Casey Weaver. *Janeway's Immunobiology*. W.W. Norton & Company, 9th edition, 2016. doi:10.1201/9781315533247.

[19] Charles Janeway, Paul Travers, Mark Walport, Mark J Shlomchik, others. *Immunobiology: the immune system in health and disease*, volume 2. Garland Pub. New York, NY, USA, 2001.

[20] David D. Chaplin. Overview of the immune response. *Journal of Allergy and Clinical Immunology*, 125(2):S3–S23, 2010. doi:10.1016/j.jaci.2009.12.980.

[21] Shizuo Akira, Satoshi Uematsu, Osamu Takeuchi. Pathogen recognition and innate immunity. *Cell*, 124(4):783–801, 2006. doi:10.1016/j.cell.2006.02.015.

[22] Eric Vivier, David Artis, Marco Colonna, Andreas Diefenbach, James P. Di Santo, Gérard Eberl, Shigeo Koyasu, Richard M. Locksley, Andrew N.J. McKenzie, Reina E. Mebius, et al. Innate lymphoid cells: 10 years on. *Cell*, 174(5):1054–1066, 2018. doi:10.1016/j.cell.2018.07.017.

[23] Fritz Melchers. Checkpoints that control b cell development. *Journal of Clinical Investigation*, 125(6):2203–2210, 2015. doi:10.1172/jci78083.

[24] Ed Palmer. Negative selection — clearing out the bad apples from the t-cell repertoire. *Nature Reviews Immunology*, 3(5):383–391, 2003. doi:10.1038/nri1085.

[25] Ruyuan Wang, Caini Lan, Kamel Benlagha, Niels Olsen Saraiva Camara, Heather Miller, Masato Kubo, Steffen Heegaard, Pamela Lee, Lu Yang, Huamei Forsman, et al. The interaction of innate immune and adaptive immune system. *MedComm*, 5(10):e714, 2024. doi:10.1002/mco2.714.

[26] Margaret M McDaniel, Hannah E Meibers, Chandrashekhar Pasare. Innate control of adaptive immunity and adaptive instruction of innate immunity: bi-directional flow of information. *Current Opinion in Immunology*, 73:25–33, 2021. doi:10.1016/j.coi.2021.07.013.

[27] Liang Zhou, Mark MW Chong, Dan R Littman. Plasticity of cd4+ t cell lineage differentiation. *Immunity*, 30(5):646–655, 2009. doi:10.1016/j.immuni.2009.05.001.

[28] Falk Nimmerjahn, Jeffrey V Ravetch. Antibodies, fc receptors and cancer. *Current Opinion in Immunology*, 19(2):239–245, 2007. doi:10.1016/j.coi.2007.01.005.

[29] Stefano Comazzetto, Bo Shen, Sean J Morrison. Niches that regulate stem cells and hematopoiesis in adult bone marrow. *Developmental Cell*, 56(13):1848–1860, 2021. doi:10.1016/j.devcel.2021.05.018.

[30] Eliver Ghosn, Momoko Yoshimoto, Hiromitsu Nakauchi, Irving L Weissman, Leonore A Herzenberg. Hematopoietic stem cell-independent hematopoiesis and the origins of innate-like b lymphocytes. *Development*, 146(15):dev170571, 2019. doi:10.1242/dev.170571.

[31] Florent Ginhoux, Martin Guilliams. Tissue-resident macrophage ontogeny and homeostasis. *Immunity*, 44(3):439–449, 2016. doi:10.1016/j.immuni.2016.02.024.

[32] Takahiro Masuda, Roman Sankowski, Ori Staszewski, Marco Prinz. Microglia heterogeneity in the single-cell era. *Cell Reports*, 30(5):1271–1281, 2020. doi:10.1016/j.celrep.2020.01.010.

[33] Thomas A Wynn, Kevin M Vannella. Macrophages in tissue repair, regeneration, and fibrosis. *Immunity*, 44(3):450–462, 2016. doi:10.1016/j.immuni.2016.02.015.

[34] Jia Li, Chu Xiao, Chunxiang Li, Jie He. Tissue-resident immune cells: from defining characteristics to roles in diseases. *Signal Transduction and Targeted Therapy*, 10(1):12, 2025. doi:10.1038/s41392-024-02050-5.

[35] Borros Arneth. Molecular mechanisms of immune regulation: A review. *Cells*, 14(4):283, 2025. doi:10.3390/cells14040283.

[36] Shinwan Kany, Jan Tilmann Vollrath, Borna Relja. Cytokines in inflammatory disease. *International Journal of Molecular Sciences*, 20(23):6008, 2019. doi:10.3390/ijms20236008.

[37] Stephen T Smale. Transcriptional regulation in the immune system: a status report. *Trends in Immunology*, 35(5):190–194, 2014. doi:10.1016/j.it.2014.03.003.

[38] Ling Lu, Joseph Barbi, Fan Pan. The regulation of immune tolerance by foxp3. *Nature Reviews Immunology*, 17(11):703–717, 2017. doi:10.1038/nri.2017.75.

[39] Xia Li, Youyuan Ye, Kailan Peng, Zhuo Zeng, Li Chen, Yanhua Zeng. Histones: The critical players in innate immunity. *Frontiers in Immunology*, 13:1030610, 2022. doi:10.3389/fimmu.2022.1030610.

[40] Lihua Qu, Tong Yin, Yijin Zhao, Wenting Lv, Ziqi Liu, Chao Chen, Kejun Liu, Shigang Shan, Rui Zhou, Xiaoqing Li, et al. Histone demethylases in the regulation of immunity and inflammation. *Cell Death Discovery*, 9(1):188, 2023. doi:10.1038/s41420-023-01489-9.

[41] Luisa Morales-Nebreda, Fred S McLafferty, Benjamin D Singer. Dna methylation as a transcriptional regulator of the immune system. *Translational Research*, 204:1–18, 2019. doi:10.1016/j.trsl.2018.08.001.

[42] Dario A. A. Vignali, Lauren W. Collison, Creg J. Workman. How regulatory t cells work. *Nature Reviews Immunology*, 8(7):523–532, 2008. doi:10.1038/nri2343.

[43] Diego Catalán, Miguel Andrés Mansilla, Ashley Ferrier, Lilian Soto, Kristine Oleinika, Juan Carlos Aguillón, Octavio Aravena. Immunosuppressive mechanisms of regulatory b cells. *Frontiers in Immunology*, 12:611795, 2021. doi:10.3389/fimmu.2021.611795.

[44] Filippo Veglia, Emilio Sanseviero, Dmitry I Gabrilovich. Myeloid-derived suppressor cells in the era of increasing myeloid cell diversity. *Nature Reviews Immunology*, 21(8):485–498, 2021. doi:10.1038/s41577-020-00490-y.

[45] Roni Nowarski, Ruaidhrí Jackson, Richard A. Flavell. The stromal intervention: Regulation of immunity and inflammation at the epithelial-mesenchymal barrier. *Cell*, 168(3):362–375, 2017. doi:10.1016/j.cell.2016.11.040.

[46] Javier Perez-Hernandez, Valerio Chiurchiù, Sylvain Perruche, Sylvaine You. Regulation of t-cell immune responses by pro-resolving lipid mediators. *Frontiers in Immunology*, 12:768133, 2021. doi:10.3389/fimmu.2021.768133.

[47] Hao Jin, Mengtong Li, Eric Jeong, Felipe Castro-Martinez, Charles S Zuker. A body–brain circuit that regulates body inflammatory responses. *Nature*, 630(8017):695–703, 2024. doi:10.1038/s41586-024-07469-y.

[48] David A Padgett, Ronald Glaser. How stress influences the immune response. *Trends in Immunology*, 24(8):444–448, 2003. doi:10.1016/s1471-4906(03)00173-x.

[49] CHARLES J. GROSSMAN. Regulation of the immune system by sex steroids*. *Endocrine Reviews*, 5(3):435–455, 1984. doi:10.1210/edrv-5-3-435.

[50] Maya E. Kotas, Ruslan Medzhitov. Homeostasis, inflammation, and disease susceptibility. *Cell*, 160(5):816–827, 2015. doi:10.1016/j.cell.2015.02.010.

[51] Mark M Davis, Cristina M Tato, David Furman. Systems immunology: just getting started. *Nature Immunology*, 18(7):725–732, 2017. doi:10.1038/ni.3768.

[52] Stuart P Weisberg, Basak B Ural, Donna L Farber. Tissue-specific immunity for a changing world. *Cell*, 184(6):1517–1529, 2021. doi:10.1016/j.cell.2021.01.042.

[53] Baihao Zhang, Alexis Vogelzang, Sidonia Fagarasan. Secreted immune metabolites that mediate immune cell communication and function. *Trends in Immunology*, 43(12):990–1005, 2022. doi:10.1016/j.it.2022.10.006.

[54] Jeffrey A Whitsett, Theresa Alenghat. Respiratory epithelial cells orchestrate pulmonary innate immunity. *Nature Immunology*, 16(1):27–35, 2014. doi:10.1038/ni.3045.

[55] Allan M Mowat, William W Agace. Regional specialization within the intestinal immune system. *Nature Reviews Immunology*, 14(10):667–685, 2014. doi:10.1038/nri3738.

[56] Rachael A Clark. Resident memory t cells in human health and disease. *Science Translational Medicine*, 7(269):269rv1–269rv1, 2015. doi:10.1126/scitranslmed.3010641.

[57] Marco Prinz, Josef Priller. Microglia and brain macrophages in the molecular age: from origin to neuropsychiatric disease. *Nature Reviews Neuroscience*, 15(5):300–312, 2014. doi:10.1038/nrn3722.

[58] Craig N Jenne, Paul Kubes. Immune surveillance by the liver. *Nature Immunology*, 14(10):996–1006, 2013. doi:10.1038/ni.2691.

[59] Jerrold M Olefsky, Christopher K Glass. Macrophages, inflammation, and insulin resistance. *Annual Review of Physiology*, 72(1):219–246, 2010. doi:10.1146/annurev-physiol-021909-135846.

[60] Ajay Chawla, Khoa D. Nguyen, Y. P. Sharon Goh. Macrophage-mediated inflammation in metabolic disease. *Nature Reviews Immunology*, 11(11):738–749, 2011. doi:10.1038/nri3071.

[61] Ruslan Medzhitov. Origin and physiological roles of inflammation. *Nature*, 454(7203):428–435, 2008. doi:10.1038/nature07201.

[62] Alessandro Fantin, Joaquim M Vieira, Gaia Gestri, Laura Denti, Quenten Schwarz, Sergey Prykhozhij, Francesca Peri, Stephen W Wilson, Christiana Ruhrberg. Tissue macrophages act as cellular chaperones for vascular anastomosis downstream of vegf-mediated endothelial tip cell induction. *Blood*, 116(5):829–840, 2010. doi:10.1182/blood-2009-12-257832.

[63] Silvia Nucera, Daniela Biziato, Michele De Palma. The interplay between macrophages and angiogenesis in development, tissue injury and regeneration. *The International Journal of Developmental Biology*, 55(4–5):495–503, 2011. doi:10.1387/ijdb.103227sn.

[64] Paôline Laurent, Valérie Jolivel, Pauline Manicki, Lynn Chiu, Cécile Contin-Bordes, Marie-Elise Truchetet, Thomas Pradeu. Immune-mediated repair: A matter of plasticity. *Frontiers in Immunology*, 8:454, 2017. doi:10.3389/fimmu.2017.00454.

[65] Dorothy P Schafer, Emily K Lehrman, Beth Stevens. The "quad-partite" synapse: Microglia-synapse interactions in the developing and mature cns. *Glia*, 61(1):24–36, 2012. doi:10.1002/glia.22389.

[66] Gregory F. Sonnenberg, Laurel A. Monticelli, Theresa Alenghat, Thomas C. Fung, Natalie A. Hutnick, Jun Kunisawa, Naoko Shibata, Stephanie Grunberg, Rohini Sinha, Adam M. Zahm, et al. Innate lymphoid cells promote anatomical containment of lymphoid-resident commensal bacteria. *Science*, 336(6086):1321–1325, 2012. doi:10.1126/science.1222551.

[67] Srinath C. Sampath, Srihari C. Sampath, Andrew T. V. Ho, Stéphane Y. Corbel, Joshua D. Millstone, John Lamb, John Walker, Bernd Kinzel, Christian Schmedt, Helen M. Blau. Induction of muscle stem cell quiescence by the secreted niche factor oncostatin m. *Nature Communications*, 9(1):1531, 2018. doi:10.1038/s41467-018-03876-8.

[68] Raj Chovatiya, Ruslan Medzhitov. Stress, inflammation, and defense of homeostasis. *Molecular Cell*, 54(2):281–288, 2014. doi:10.1016/j.molcel.2014.03.030.

[69] Gökhan S. Hotamisligil. Foundations of immunometabolism and implications for metabolic health and disease. *Immunity*, 47(3):406–420, 2017. doi:10.1016/j.immuni.2017.08.009.

[70] Ari B Molofsky, Jesse C Nussbaum, Hong-Erh Liang, Steven J Van Dyken, Laurence E Cheng, Alexander Mohapatra, Ajay Chawla, Richard M Locksley. Innate lymphoid type 2 cells sustain visceral adipose tissue eosinophils and alternatively activated macrophages. *Journal of Experimental Medicine*, 210(3):535–549, 2013. doi:10.1084/jem.20121964.

[71] Federica Sallusto, Jens Geginat, Antonio Lanzavecchia. Central memory and effector memory t cell subsets: Function, generation, and maintenance. *Annual Review of Immunology*, 22(1):745–763, 2004. doi:10.1146/annurev.immunol.22.012703.104702.

[72] David Masopust, Jason M Schenkel. The integration of t cell migration, differentiation and function. *Nature Reviews Immunology*, 13(5):309–320, 2013. doi:10.1038/nri3442.

[73] María Yáñez-Mó, Pia R.-M. Siljander, Zoraida Andreu, Apolonija Bedina Zavec, Francesc E. Borràs, Edit I. Buzas, Krisztina Buzas, Enriqueta Casal, Francesco Cappello, Joana Carvalho, et al. Biological properties of extracellular vesicles and their physiological functions. *Journal of Extracellular Vesicles*, 4(1):27066, 2015. doi:10.3402/jev.v4.27066.

[74] Paul D Robbins, Adrian E Morelli. Regulation of immune responses by extracellular vesicles. *Nature Reviews Immunology*, 14(3):195–208, 2014. doi:10.1038/nri3622.

[75] Xinyu Yang, Di Lu, Jianyong Zhuo, Zuyuan Lin, Modan Yang, Xiao Xu. The gut-liver axis in immune remodeling: New insight into liver diseases. *International Journal of Biological Sciences*, 16(13):2357–2366, 2020. doi:10.7150/ijbs.46405.

[76] Raphaël Enaud, Renaud Prevel, Eleonora Ciarlo, Fabien Beaufils, Gregoire Wieërs, Benoit Guery, Laurence Delhaes. The gut-lung axis in health and respiratory diseases: A place for inter-organ and inter-kingdom crosstalks. *Frontiers in Cellular and Infection Microbiology*, 10:9, 2020. doi:10.3389/fcimb.2020.00009.

[77] Aliće Weiglein, Evelyn Gaffal, Anne Albrecht. Probing the skin–brain axis: New vistas using mouse models. *International Journal of Molecular Sciences*, 23(13):7484, 2022. doi:10.3390/ijms23137484.

[78] Kevin Man, Axel Kallies, Ajithkumar Vasanthakumar. Resident and migratory adipose immune cells control systemic metabolism and thermogenesis. *Cellular & Molecular Immunology*, 19(3):421–431, 2021. doi:10.1038/s41423-021-00804-7.

[79] Aurélien Trompette, Eva S Gollwitzer, Koshika Yadava, Anke K Sichelstiel, Norbert Sprenger, Catherine Ngom-Bru, Carine Blanchard, Tobias Junt, Laurent P Nicod, Nicola L Harris, et al. Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. *Nature Medicine*, 20(2):159–166, 2014. doi:10.1038/nm.3444.

[80] Markus G Manz, Steffen Boettcher. Emergency granulopoiesis. *Nature Reviews Immunology*, 14(5):302–314, 2014. doi:10.1038/nri3660.

[81] Forum Raval. The bidirectional relationship between metabolism and immune responses. *Discoveries*, 1(1):e6, 2013. doi:10.15190/d.2013.6.

[82] A Katharina Simon, Georg A Hollander, Andrew McMichael. Evolution of the immune system in humans from infancy to old age. *Proceedings of the Royal Society B: Biological Sciences*, 282(1821):20143085, 2015. doi:10.1098/rspb.2014.3085.

[83] Jun Guo, Xiuqing Huang, Lin Dou, Mingjing Yan, Tao Shen, Weiqing Tang, Jian Li. Aging and aging-related diseases: from molecular mechanisms to interventions and treatments. *Signal Transduction and Targeted Therapy*, 7(1):391, 2022. doi:10.1038/s41392-022-01251-0.

[84] Shannon Marchal, Alexander Choukér, Jürgen Bereiter-Hahn, Armin Kraus, Daniela Grimm, Marcus Krüger. Challenges for the human immune system after leaving earth. *npj Microgravity*, 10(1):106, 2024. doi:10.1038/s41526-024-00446-9.

[85] Nitya Jain. The early life education of the immune system: Moms, microbes and (missed) opportunities. *Gut Microbes*, 12(1):1824564, 2020. doi:10.1080/19490976.2020.1824564.

[86] Sabra L Klein, Katie L Flanagan. Sex differences in immune responses. *Nature Reviews Immunology*, 16(10):626–638, 2016. doi:10.1038/nri.2016.90.

[87] Joseph P Hoffmann, Jennifer A Liu, Kumba Seddu, Sabra L Klein. Sex hormone signaling and regulation of immune function. *Immunity*, 56(11):2472–2491, 2023. doi:10.1016/j.immuni.2023.10.008.

[88] Zaoqu Liu, Qimeng Liang, Yuqing Ren, Chunguang Guo, Xiaoyong Ge, Libo Wang, Quan Cheng, Peng Luo, Yi Zhang, Xinwei Han. Immunosenescence: molecular mechanisms and diseases. *Signal Transduction and Targeted Therapy*, 8(1):200, 2023. doi:10.1038/s41392-023-01451-2.

[89] Cornelia M. Weyand, Jörg J. Goronzy. Aging of the immune system. mechanisms and therapeutic targets. *Annals of the American Thoracic Society*, 13(Supplement 5):S422–S428, 2016. doi:10.1513/annalsats.201602-095aw.

[90] Miruna G. Gaspar, Carmen Núñez-Carro, Margarita Blanco-Blanco, Francisco J. Blanco, María C. de Andrés. Inflammaging contributes to osteoarthritis development and human microbiota variations and vice versa: A systematic review. *Osteoarthritis and Cartilage*, 33(2):218–230, 2025. doi:10.1016/j.joca.2024.11.005.

[91] Daniel Baylis, David B Bartlett, Harnish P Patel, Helen C Roberts. Understanding how we age: insights into inflammaging. *Longevity & Healthspan*, 2(1):1–8, 2013. doi:10.1186/2046-2395-2-8.

[92] Tara E Sutherland, Douglas P Dyer, Judith E Allen. The extracellular matrix and the immune system: A mutually dependent relationship. *Science*, 379(6633):eabp8964, 2023. doi:10.1126/science.abp8964.

[93] Carlos López-Otín, Maria A. Blasco, Linda Partridge, Manuel Serrano, Guido Kroemer. Hallmarks of aging: An expanding universe. *Cell*, 186(2):243–278, 2023. doi:10.1016/j.cell.2022.11.001.

[94] Rakesh Kumar Sahoo, Tanisha Gupta, Smily, Vinay Kumar, Sarita Rani, Umesh Gupta. *Aetiology and pathophysiology of neurodegenerative disorders*, pages 1–16. Elsevier, 2022. doi:10.1016/b978-0-323-85544-0.00006-x.

[95] David M Wilson, Mark R Cookson, Ludo Van Den Bosch, Henrik Zetterberg, David M Holtzman, Ilse Dewachter. Hallmarks of neurodegenerative diseases. *Cell*, 186(4):693–714, 2023. doi:10.1016/j.cell.2022.12.032.

[96] David S. Knopman, Helene Amieva, Ronald C. Petersen, Gäel Chételat, David M. Holtzman, Bradley T. Hyman, Ralph A. Nixon, David T. Jones. Alzheimer disease. *Nature Reviews Disease Primers*, 7(1):33, 2021. doi:10.1038/s41572-021-00269-y.

[97] Nóra Baligács, Giulia Albertini, Sarah C. Borrie, Lutgarde Serneels, Clare Pridans, Sriram Balusu, Bart De Strooper. Homeostatic microglia initially seed and activated microglia later reshape amyloid plaques in alzheimer's disease. *Nature Communications*, 15(1):1–14, 2024. doi:10.1038/s41467-024-54779-w.

[98] Christopher R Beam, Cody Kaneshiro, Jung Yun Jang, Chandra A Reynolds, Nancy L Pedersen, Margaret Gatz. Differences between women and men in incidence rates of dementia and alzheimer's disease. *Journal of Alzheimer's Disease*, 64(4):1077–1083, 2018. doi:10.3233/jad-180141.

[99] Chih-Ching Liu, Chung-Yi Li, Yu Sun, Susan C Hu. Gender and age differences and the trend in the incidence and prevalence of dementia and alzheimer's disease in taiwan: A 7-year national population-based study. *BioMed Research International*, 2019(1):1–12, 2019. doi:10.1155/2019/5378540.

[100] Werner Poewe, Klaus Seppi, Caroline M Tanner, Glenda M Halliday, Patrik Brundin, Jens Volkmann, Anette-Eleonore Schrag, Anthony E Lang. Parkinson disease. *Nature Reviews Disease Primers*, 3(1):1–21, March 2017. doi:10.1038/nrdp.2017.13.

[101] Sigurlaug Sveinbjornsdottir. The clinical symptoms of parkinson's disease. *Journal of Neurochemistry*, 139(S1):318–324, 2016. doi:10.1111/jnc.13691.

[102] Shofiul Azam, Md. Ezazul Haque, Rengasamy Balakrishnan, In-Su Kim, Dong-Kug Choi. The ageing brain: Molecular and cellular basis of neurodegeneration. *Frontiers in Cell and Developmental Biology*, 9:683459, 2021. doi:10.3389/fcell.2021.683459.

[103] Yujun Hou, Xiuli Dan, Mansi Babbar, Yong Wei, Steen G Hasselbalch, Deborah L Croteau, Vilhelm A Bohr. Ageing as a risk factor for neurodegenerative disease. *Nature Reviews Neurology*, 15(10):565–581, 2019. doi:10.1038/s41582-019-0244-7.

[104] Hannah D Mason, Dorian B McGavern. How the immune system shapes neurodegenerative diseases. *Trends in Neurosciences*, 45(10):733–748, 2022. doi:10.1016/j.tins.2022.08.001.

[105] Citlalli Netzahualcoyotzi, Juan Jair Santillán-Cigales, Laura Virginia Adalid-Peralta, Iván Velasco. Infiltration of immune cells to the brain and its relation to the pathogenesis of alzheimer's and parkinson's diseases. *Journal of Neurochemistry*, 168(9):2316–2334, 2024. doi:10.1111/jnc.16106.

[106] Callum N Watson, Antonio Belli, Valentina Di Pietro. Small non-coding rnas: New class of biomarkers and potential therapeutic targets in neurodegenerative disease. *Frontiers in Genetics*, 10:364, 2019. doi:10.3389/fgene.2019.00364.

[107] Anton I Petrov, Simon J E Kay, Ioanna Kalvari, Kevin L Howe, Kristian A Gray, Elspeth A Bruford, Paul J Kersey, Guy Cochrane, Robert D Finn, Alex Bateman, et al. Rnacentral: a comprehensive database of non-coding rna sequences. *Nucleic Acids Research*, 45(D1):D128–D134, 2016. doi:10.1093/nar/gkw1008.

[108] Vincent Boivin, Laurence Faucher-Giguère, Michelle Scott, Sherif Abou-Elela. The cellular landscape of mid-size noncoding rna. *WIREs RNA*, 10(4):e1530, 2019. doi:10.1002/wrna.1530.

[109] Danyella Barbosa Dogini, Vinícius D'Avila Bittencourt Pascoal, Simoni Helena Avansini, André Schwambach Vieira, Tiago Campos Pereira, Iscia Lopes-Cendes. The new world of rnas. *Genetics and Molecular Biology*, 37(1 suppl 1):285–293, 2014. doi:10.1590/s1415-47572014000200014.

[110] Donny D Licatalosi, Robert B Darnell. Rna processing and its regulation: global insights into biological networks. *Nature Reviews Genetics*, 11(1):75–87, 2010. doi:10.1038/nrg2673.

[111] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, Peter Walter. From dna to rna. In *Molecular Biology of the Cell. 4th edition*. Garland Science, 2002.

[112] Theresa Phillips. Regulation of transcription and gene expression in eukaryotes. *Nature Education*, 1(1):199, 2008.

[113] Jie Zhu, Ke Zhang, Yuan Chen, Xinyu Ge, Junqing Wu, Peng Xu, Jie Yao. Progress of single-cell rna sequencing combined with spatial transcriptomics in tumour microenvironment and treatment of pancreatic cancer. *Journal of Translational Medicine*, 22(1):563, 2024. doi:10.1186/s12967-024-05307-3.

[114] Malte D Luecken, Fabian J Theis. Current best practices in single-cell rna-seq analysis: a tutorial. *Molecular Systems Biology*, 15(6):e8746, 2019. doi:10.15252/msb.20188746.

[115] 10X Genomics. Cell ranger's gene expression algorithm, 2025. URL: https://www.10xgenomics.com/support/software/cell-ranger/latest/algorithms-overview/cr-gex-algorithm.

[116] Tobias Fehlmann, Fabian Kern, Omar Laham, Christina Backes, Jeffrey Solomon, Pascal Hirsch, Carsten Volz, Rolf Müller, Andreas Keller. mirmaster 2.0: multi-species noncoding rna sequencing analyses at scale. *Nucleic Acids Research*, 49(W1):W397–W408, 2021. doi:10.1093/nar/gkab268.

[117] Matthew D Young, Sam Behjati. Soupx removes ambient rna contamination from droplet-based single-cell rna sequencing data. *GigaScience*, 9(12):giaa151, 2020. doi:10.1093/gigascience/giaa151.

[118] Shiyi Yang, Sean E Corbett, Yusuke Koga, Zhe Wang, W Evan Johnson, Masanao Yajima, Joshua D Campbell. Decontamination of ambient rna in single-cell rna-seq with decontx. *Genome Biology*, 21(1):1–15, 2020. doi:10.1186/s13059-020-1950-6.

[119] E. Madissoon, A. Wilbrey-Clark, R. J. Miragaia, K. Saeb-Parsy, K. T. Mahbubani, N. Georgakopoulos, P. Harding, K. Polanski, N. Huang, K. Nowicki-Osuch, et al. scrna-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. *Genome Biology*, 21(1):1–16, 2019. doi:10.1186/s13059-019-1906-x.

[120] Elisa Floriddia. The impact of ambient rna. *Nature Neuroscience*, 25(12):1583–1583, 2022. doi:10.1038/s41593-022-01232-0.

[121] Marijn Berg, Ilya Petoukhov, Inge van den Ende, Kerstin B. Meyer, Victor Guryev, Judith M. Vonk, Orestes Carpaij, Martin Banchero, Rudi W. Hendriks, Maarten van den Berge, et al. Fastcar: fast correction for ambient rna to facilitate differential gene expression analysis in single-cell rna-sequencing datasets. *BMC Genomics*, 24(1):722, 2023. doi:10.1186/s12864-023-09822-3.

[122] Christopher S McGinnis, Lyndsay M Murrow, Zev J Gartner. Doubletfinder: Doublet detection in single-cell rna sequencing data using artificial nearest neighbors. *Cell Systems*, 8(4):329–337.e4, 2019. doi:10.1016/j.cels.2019.03.003.

[123] Pierre-Luc Germain, Aaron Lun, Carlos Garcia Meixide, Will Macnair, Mark D. Robinson. Doublet identification in single-cell sequencing data using scdblfinder. *F1000Research*, 10:979, 2022. URL: https://doi.org:10, doi:10.12688/f1000research.73600.2.

[124] Oliver Stegle, Sarah A Teichmann, John C Marioni. Computational and analytical challenges in single-cell transcriptomics. *Nature Reviews Genetics*, 16(3):133–145, 2015. doi:10.1038/nrg3833.

[125] Tomislav Ilicic, Jong Kyoung Kim, Aleksandra A Kolodziejczyk, Frederik Otzen Bagger, Davis James McCarthy, John C Marioni, Sarah A Teichmann. Classification of low quality cells from single-cell rna-seq data. *Genome Biology*, 17(1):1–15, 2016. doi:10.1186/s13059-016-0888-1.

[126] Nan Miles Xi, Jingyi Jessica Li. Benchmarking computational doublet-detection methods for single-cell rna sequencing data. *Cell Systems*, 12(2):176–194.e6, 2021. doi:10.1016/j.cels.2020.11.008.

[127] 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*, 37(7):963–967, 2020. doi:10.1093/bioinformatics/btaa751.

[128] Tomàs Montserrat-Ayuso, Anna Esteve-Codina. High content of nuclei-free low-quality cells in reference single-cell atlases: a call for more stringent quality control using nuclear fraction. *BMC Genomics*, 25(1):1124, 2024. doi:10.1186/s12864-024-11015-5.

[129] Georgi K Marinov, Brian A Williams, Ken McCue, Gary P Schroth, Jason Gertz, Richard M Myers, Barbara J Wold. From single-cell to cell-pool transcriptomes: Stochasticity in gene expression and rna splicing. *Genome Research*, 24(3):496–510, 2014. doi:10.1101/gr.161034.113.

[130] Philip Brennecke, Simon Anders, Jong Kyoung Kim, Aleksandra A Kołodziejczyk, Xiuwei Zhang, Valentina Proserpio, Bianka Baying, Vladimir Benes, Sarah A Teichmann, John C Marioni, et al. Accounting for technical noise in single-cell rna-seq experiments. *Nature Methods*, 10(11):1093–1095, 2013. doi:10.1038/nmeth.2645.

[131] Jonathan M. Raser, Erin K. O'Shea. Noise in gene expression: Origins, consequences, and control. *Science*, 309(5743):2010–2013, 2005. doi:10.1126/science.1105891.

[132] Benjamin B Kaufmann, Alexander van Oudenaarden. Stochastic gene expression: from single molecules to the proteome. *Current Opinion in Genetics & Development*, 17(2):107–112, 2007. doi:10.1016/j.gde.2007.02.007.

[133] Laleh Haghverdi, Aaron T L Lun, Michael D Morgan, John C Marioni. Batch effects in single-cell rna-sequencing data are corrected by matching mutual nearest neighbors. *Nature Biotechnology*, 36(5):421–427, 2018. doi:10.1038/nbt.4091.

[134] Hoa Thi Nhu Tran, Kok Siong Ang, Marion Chevrier, Xiaomeng Zhang, Nicole Yee Shin Lee, Michelle Goh, Jinmiao Chen. A benchmark of batch-effect correction methods for single-cell rna sequencing data. *Genome Biology*, 21(1):1–32, 2020. doi:10.1186/s13059-019-1850-9.

[135] Stephanie C Hicks, F William Townes, Mingxiang Teng, Rafael A Irizarry. Missing data and technical variability in single-cell rna-sequencing experiments. *Biostatistics*, 19(4):562–578, 2017. doi:10.1093/biostatistics/kxx053.

[136] Ilya Korsunsky, Nghia Millard, Jean Fan, Kamil Slowikowski, Fan Zhang, Kevin Wei, Yuriy Baglaenko, Michael Brenner, Po-ru Loh, Soumya Raychaudhuri. Fast, sensitive and accurate integration of single-cell data with harmony. *Nature Methods*, 16(12):1289–1296, 2019. doi:10.1038/s41592-019-0619-0.

[137] Andrew Butler, Paul Hoffman, Peter Smibert, Efthymia Papalexi, Rahul Satija. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology*, 36(5):411–420, 2018. doi:10.1038/nbt.4096.

[138] Yuqing Zhang, Giovanni Parmigiani, W Evan Johnson. Combat-seq: batch effect adjustment for rna-seq count data. *NAR Genomics and Bioinformatics*, 2(3):lqaa078, 2020. doi:10.1093/nargab/lqaa078.

[139] Peter V Kharchenko, Lev Silberstein, David T Scadden. Bayesian approach to single-cell differential expression analysis. *Nature Methods*, 11(7):740–742, 2014. doi:10.1038/nmeth.2967.

[140] Peng Qiu. Embracing the dropouts in single-cell rna-seq analysis. *Nature Communications*, 11(1):1169, 2020. doi:10.1038/s41467-020-14976-9.

[141] Tae Hyun Kim, Xiang Zhou, Mengjie Chen. Demystifying "drop-outs" in single-cell umi data. *Genome Biology*, 21(1):196, 2020. doi:10.1186/s13059-020-02096-y.

[142] Dominic Grün, Lennart Kester, Alexander van Oudenaarden. Validation of noise models for single-cell transcriptomics. *Nature Methods*, 11(6):637–640, 2014. doi:10.1038/nmeth.2930.

[143] Vladimir Yu Kiselev, Tallulah S Andrews, Martin Hemberg. Challenges in unsupervised clustering of single-cell rna-seq data. *Nature Reviews Genetics*, 20(5):273–282, 2019. doi:10.1038/s41576-018-0088-9.

[144] RICHARD BELLMAN. *Dynamic Programming*. Rand Corporation research study. Princeton University Press, 2021. URL: https://books.google.de/books?id=rZW4ugAACAAJ, doi:10.2307/j.ctv1nxcw0f.

[145] Naomi Altman, Martin Krzywinski. The curse(s) of dimensionality. *Nature Methods*, 15(6):399–400, 2018. doi:10.1038/s41592-018-0019-x.

[146] Yusuke Imoto, Tomonori Nakamura, Emerson G Escolar, Michio Yoshiwaki, Yoji Kojima, Yukihiro Yabuta, Yoshitaka Katou, Takuya Yamamoto, Yasuaki Hiraoka, Mitinori Saitou. Resolution of the curse of dimensionality in single-cell rna sequencing data analysis. *Life Science Alliance*, 5(12):e202201591, 2022. doi:10.26508/lsa.202201591.

[147] F William Townes, Stephanie C Hicks, Martin J Aryee, Rafael A Irizarry. Feature selection and dimension reduction for single-cell rna-seq based on a multinomial model. *Genome Biology*, 20(1):1–16, 2019. doi:10.1186/s13059-019-1861-6.

[148] Pengyi Yang, Hao Huang, Chunlei Liu. Feature selection revisited in the single-cell era. *Genome Biology*, 22(1):1–17, 2021. arXiv:2110.14329, doi:10.1186/s13059-021-02544-3.

[149] Tim Stuart, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William M Mauck, Yuhan Hao, Marlon Stoeckius, Peter Smibert, Rahul Satija. Comprehensive integration of single-cell data. *Cell*, 177(7):1888–1902.e21, 2019. doi:10.1016/j.cell.2019.05.031.

[150] Svante Wold, Kim Esbensen, Paul Geladi. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1-3):37–52, 1987. doi:10.1016/0169-7439(87)80084-9.

[151] Laurens Van der Maaten, Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(11), 2008.

[152] Etienne Becht, Leland McInnes, John Healy, Charles-Antoine Dutertre, Immanuel W H Kwok, Lai Guan Ng, Florent Ginhoux, Evan W Newell. Dimensionality reduction for visualizing single-cell data using umap. *Nature Biotechnology*, 37(1):38–44, 2018. doi:10.1038/nbt.4314.

[153] Ruizhi Xiang, Wencan Wang, Lei Yang, Shiyuan Wang, Chaohan Xu, Xiaowen Chen. A comparison for dimensionality reduction methods of single-cell rna-seq data. *Frontiers in Genetics*, 12:646936, 2021. doi:10.3389/fgene.2021.646936.

[154] Zoe A Clarke, Tallulah S Andrews, Jawairia Atif, Delaram Pouyabahar, Brendan T Innes, Sonya A MacParland, Gary D Bader. Tutorial: guidelines for annotating single-cell transcriptomic maps using automated and manual methods. *Nature Protocols*, 16(6):2749–2764, 2021. doi:10.1038/s41596-021-00534-0.

[155] Leland McInnes, John Healy, James Melville. Umap: Uniform manifold approximation and projection for dimension reduction, 2018. doi:10.48550/ARXIV.1802.03426.

[156] Dmitry Kobak, Philipp Berens. The art of using t-sne for single-cell transcriptomics. *Nature Communications*, 10(1):5416, 2019. doi:10.1038/s41467-019-13056-x.

[157] Shixiong Zhang, Xiangtao Li, Jiecong Lin, Qiuzhen Lin, Ka-Chun Wong. Review of single-cell rna-seq data clustering for cell-type identification and characterization. *RNA*, 29(5):517–530, 2023. arXiv:2001.01006, doi:10.1261/rna.078965.121.

[158] Dominic Grün, Anna Lyubimova, Lennart Kester, Kay Wiebrands, Onur Basak, Nobuo Sasaki, Hans Clevers, Alexander van Oudenaarden. Single-cell messenger rna sequencing reveals rare intestinal cell types. *Nature*, 525(7568):251–255, 2015. doi:10.1038/nature14966.

[159] Lu Yang, Jiancheng Liu, Qiang Lu, Arthur D Riggs, Xiwei Wu. Saic: an iterative clustering approach for analysis of single cell rna-seq data. *BMC Genomics*, 18(S6):9–17, 2017. doi:10.1186/s12864-017-4019-5.

[160] Justina žurauskienė, Christopher Yau. pcareduce: hierarchical clustering of single cell transcriptional profiles. *BMC Bioinformatics*, 17(1):1–11, 2016. doi:10.1186/s12859-016-0984-y.

[161] Vladimir Yu Kiselev, Kristina Kirschner, Michael T Schaub, Tallulah Andrews, Andrew Yiu, Tamir Chandra, Kedar N Natarajan, Wolf Reik, Mauricio Barahona, Anthony R Green, et al. Sc3: consensus clustering of single-cell rna-seq data. *Nature Methods*, 14(5):483–486, 2017. doi:10.1038/nmeth.4236.

[162] Peijie Lin, Michael Troup, Joshua W. K. Ho. Cidr: Ultrafast and accurate clustering through imputation for single-cell rna-seq data. *Genome Biology*, 18(1):1–11, 2017. doi:10.1186/s13059-017-1188-0.

[163] Amit Zeisel, Ana B. Muñoz-Manchado, Simone Codeluppi, Peter Lönnerberg, Gioele La Manno, Anna Juréus, Sueli Marques, Hermany Munguba, Liqun He, Christer Betsholtz, et al. Cell types in the mouse cortex and hippocampus revealed by single-cell rna-seq. *Science*, 347(6226):1138–1142, 2015. doi:10.1126/science.aaa1934.

[164] Minzhe Guo, Hui Wang, S Steven Potter, Jeffrey A Whitsett, Yan Xu. Sincera: A pipeline for single-cell rna-seq profiling analysis. *PLOS Computational Biology*, 11(11):e1004575, 2015. doi:10.1371/journal.pcbi.1004575.

[165] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008. doi:10.1088/1742-5468/2008/10/p10008.

[166] V. A. Traag, L. Waltman, N. J. van Eck. From louvain to leiden: guaranteeing well-connected communities. *Scientific Reports*, 9(1):1–12, 2019. arXiv:1810.08473, doi:10.1038/s41598-019-41695-z.

[167] Xiaojie Qiu, Qi Mao, Ying Tang, Li Wang, Raghav Chawla, Hannah A Pliner, Cole Trapnell. Reversed graph embedding resolves complex single-cell trajectories. *Nature Methods*, 14(10):979–982, 2017. doi:10.1038/nmeth.4402.

[168] Lan Jiang, Huidong Chen, Luca Pinello, Guo-Cheng Yuan. Giniclust: detecting rare cell types from single-cell gene expression data with gini index. *Genome Biology*, 17(1):1–13, 2016. doi:10.1186/s13059-016-1010-4.

[169] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, others. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231, 1996.

[170] Xiangjie Li, Kui Wang, Yafei Lyu, Huize Pan, Jingxiao Zhang, Dwight Stambolian, Katalin Susztak, Muredach P Reilly, Gang Hu, Mingyao Li. Deep learning enables accurate clustering with batch effect removal in single-cell rna-seq analysis. *Nature Communications*, 11(1):2338, 2020. doi:10.1038/s41467-020-15851-3.

[171] Liang Chen, Weinan Wang, Yuyao Zhai, Minghua Deng. Deep soft k-means clustering with self-training for single-cell rna sequence data. *NAR Genomics and Bioinformatics*, 2(2):lqaa039, 2020. doi:10.1093/nargab/lqaa039.

[172] Christopher Heje Grønbech, Maximillian Fornitz Vording, Pascal N Timshel, Casper Kaae Sønderby, Tune H Pers, Ole Winther. scvae: variational auto-encoders for single-cell gene expression data. *Bioinformatics*, 36(16):4415–4422, 2020. doi:10.1093/bioinformatics/btaa293.

[173] M. E. J. Newman, M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113, 2004. arXiv:cond-mat/0308217, doi:10.1103/physreve.69.026113.

[174] Detlev Arendt. The evolution of cell types in animals: emerging principles from molecular studies. *Nature Reviews Genetics*, 9(11):868–882, 2008. doi:10.1038/nrg2416.

[175] Hongkui Zeng. What is a cell type and how to define it? *Cell*, 185(15):2739–2755, 2022. doi:10.1016/j.cell.2022.06.031.

[176] David Osumi-Sutherland, Chuan Xu, Maria Keays, Adam P Levine, Peter V Kharchenko, Aviv Regev, Ed Lein, Sarah A Teichmann. Cell type ontologies of the human cell atlas. *Nature Cell Biology*, 23(11):1129–1135, 2021. doi:10.1038/s41556-021-00787-7.

[177] Alexander Stewart, Joseph Chi-Fung Ng, Gillian Wallis, Vasiliki Tsioligka, Franca Fraternali, Deborah K Dunn-Walters. Single-cell transcriptomic analyses define distinct peripheral b cell subsets and discrete development pathways. *Frontiers in Immunology*, 12:602539, 2021. doi:10.3389/fimmu.2021.602539.

[178] Ignacio Sanz, Chungwen Wei, Scott A Jenks, Kevin S Cashman, Christopher Tipton, Matthew C Woodruff, Jennifer Hom, F Eun-Hyung Lee. Challenges and opportunities for consistent classification of human b cell and plasma cell populations. *Frontiers in Immunology*, 10:2458, 2019. doi:10.3389/fimmu.2019.02458.

[179] Kevin Mulder, Amit Ashok Patel, Wan Ting Kong, Cécile Piot, Evelyn Halitzki, Garett Dunsmore, Shabnam Khalilnezhad, Sergio Erdal Irac, Agathe Dubuisson, Marion Chevrier, et al. Cross-tissue single-cell landscape of human monocytes and macrophages in health and disease. *Immunity*, 54(8):1883–1900.e5, 2021. doi:10.1016/j.immuni.2021.07.007.

[180] Charles-Antoine Dutertre, Etienne Becht, Sergio Erdal Irac, Ahad Khalilnezhad, Vipin Narang, Shabnam Khalilnezhad, Pei Y. Ng, Lucas L. van den Hoogen, Jing Yao Leong, Bernett Lee, et al. Single-cell analysis of human mononuclear phagocytes reveals subset-defining markers and identifies circulating inflammatory dendritic cells. *Immunity*, 51(3):573–589.e8, 2019. doi:10.1016/j.immuni.2019.08.008.

[181] Alexandra-Chloé Villani, Rahul Satija, Gary Reynolds, Siranush Sarkizova, Karthik Shekhar, James Fletcher, Morgane Griesbeck, Andrew Butler, Shiwei Zheng, Suzan Lazo, et al. Single-cell rna-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science*, 356(6335):eaah4573, 2017. doi:10.1126/science.aah4573.

[182] Yolanda D Mahnke, Tess M Brodie, Federica Sallusto, Mario Roederer, Enrico Lugli. The who's who of t-cell differentiation: Human memory t-cell subsets. *European Journal of Immunology*, 43(11):2797–2809, 2013. doi:10.1002/eji.201343751.

[183] Nadia Caccamo, Simone A. Joosten, Tom H. M. Ottenhoff, Francesco Dieli. Atypical human effector/memory cd4+ t cells with a naive-like phenotype. *Frontiers in Immunology*, 9:2832, 2018. doi:10.3389/fimmu.2018.02832.

[184] Kerry A Mullan, Nicky de Vrij, Sebastiaan Valkiers, Pieter Meysman. Current annotation strategies for t cell phenotyping of single-cell rna-seq data. *Frontiers in Immunology*, 14:1306169, 2023. doi:10.3389/fimmu.2023.1306169.

[185] Tamim Abdelaal, Lieke Michielsen, Davy Cats, Dylan Hoogduin, Hailiang Mei, Marcel J. T. Reinders, Ahmed Mahfouz. A comparison of automatic cell identification methods for single-cell rna sequencing data. *Genome Biology*, 20(1):1–19, 2019. doi:10.1186/s13059-019-1795-z.

[186] Oscar Franzén, Li-Ming Gan, Johan L M Björkegren. Panglaodb: a web server for exploration of mouse and human single-cell rna sequencing data. *Database*, 2019:baz046, 2019. doi:10.1093/database/baz046.

[187] Congxue Hu, Tengyue Li, Yingqi Xu, Xinxin Zhang, Feng Li, Jing Bai, Jing Chen, Wenqi Jiang, Kaiyue Yang, Qi Ou, et al. Cellmarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scrna-seq data. *Nucleic Acids Research*, 51(D1):D870–D876, 2022. doi:10.1093/nar/gkac947.

[188] Sara Aibar, Carmen Bravo González-Blas, Thomas Moerman, Vân Anh Huynh-Thu, Hana Imrichova, Gert Hulselmans, Florian Rambow, Jean-Christophe Marine, Pierre Geurts, Jan Aerts, et al. Scenic: single-cell regulatory network inference and clustering. *Nature Methods*, 14(11):1083–1086, 2017. doi:10.1038/nmeth.4463.

[189] Ze Zhang, Danni Luo, Xue Zhong, Jin Huk Choi, Yuanqing Ma, Stacy Wang, Elena Mahrt, Wei Guo, Eric W Stawiski, Zora Modrusan, et al. Scina: A semi-supervised subtyping algorithm of single cells and bulk samples. *Genes*, 10(7):531, 2019. doi:10.3390/genes10070531.

[190] R. Edgar. Gene expression omnibus: Ncbi gene expression and hybridization array data repository. *Nucleic Acids Research*, 30(1):207–210, 2002. doi:10.1093/nar/30.1.207.

[191] Irene Papatheodorou, Pablo Moreno, Jonathan Manning, Alfonso Muñoz-Pomer Fuentes, Nancy George, Silvie Fexova, Nuno A Fonseca, Anja Füllgrabe, Matthew Green, Ni Huang, et al. Expression atlas update: from tissues to single cells. *Nucleic Acids Research*, 48(D1):D77–D83, 2019. doi:10.1093/nar/gkz947.

[192] Aviv Regev, Sarah A. Teichmann, Eric S. Lander, Ido Amit, Christophe Benoist, Ewan Birney, Bernd Bodenmiller, Peter Campbell, Piero Carninci, Menna Clatworthy, et al. The human cell atlas. *elife*, 6:e27041, 2017. doi:10.1101/121202.

[193] Dvir Aran, Agnieszka P. Looney, Leqian Liu, Esther Wu, Valerie Fong, Austin Hsu, Suzanna Chak, Ram P. Naikawadi, Paul J. Wolters, Adam R. Abate, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nature Immunology*, 20(2):163–172, 2019. doi:10.1038/s41590-018-0276-y.

[194] Vladimir Yu Kiselev, Andrew Yiu, Martin Hemberg. scmap: projection of single-cell rna-seq data across data sets. *Nature Methods*, 15(5):359–362, 2018. doi:10.1038/nmeth.4644.

[195] Rahul Satija, Jeffrey A Farrell, David Gennert, Alexander F Schier, Aviv Regev. Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology*, 33(5):495–502, 2015. doi:10.1038/nbt.3192.

[196] Jacob C Kimmel, David R Kelley. Semisupervised adversarial neural networks for single-cell classification. *Genome Research*, 31(10):1781–1793, 2021. doi:10.1101/gr.268581.120.

[197] Feiyang Ma, Matteo Pellegrini. Actinn: automated identification of cell types in single cell rna sequencing. *Bioinformatics*, 36(2):533–538, 2019. doi:10.1093/bioinformatics/btz592.

[198] C. Domínguez Conde, C. Xu, L. B. Jarvis, D. B. Rainbow, S. B. Wells, T. Gomes, S. K. Howlett, O. Suchanek, K. Polanski, H. W. King, et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science*, 376(6594):eabl5197, 2022. doi:10.1126/science.abl5197.

[199] Jordan W. Squair, Matthieu Gautier, Claudia Kathe, Mark A. Anderson, Nicholas D. James, Thomas H. Hutson, Rémi Hudelle, Taha Qaiser, Kaya J. E. Matson, Quentin Barraud, et al. Confronting false discoveries in single-cell differential expression. *Nature Communications*, 12(1):5692, 2021. doi:10.1038/s41467-021-25960-2.

[200] Jianying Li, Pierre R Bushel, Tzu-Ming Chu, Russell D Wolfinger. Principal variance components analysis: Estimating batch effects in microarray gene expression data, 2009. doi:10.1002/9780470685983.ch12.

[201] Arfa Anjum, Seema Jaggi, Eldho Varghese, Shwetank Lall, Arpan Bhowmik, Anil Rai. Identification of differentially expressed genes in rna-seq data of arabidopsis thaliana: A compound distribution approach. *Journal of Computational Biology*, 23(4):239–247, 2016. doi:10.1089/cmb.2015.0205.

[202] Michael I Love, Wolfgang Huber, Simon Anders. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biology*, 15(12):1–17, 2014. doi:10.1186/s13059-014-0550-8.

[203] Mark D Robinson, Davis J McCarthy, Gordon K Smyth. edger: a bioconductor package for differential expression analysis of digital gene expression data. *bioinformatics*, 26(1):139–140, 2010.

[204] Matthew E. Ritchie, Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, Gordon K. Smyth. limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Research*, 43(7):e47–e47, 2015. doi:10.1093/nar/gkv007.

[205] Greg Finak, Andrew McDavid, Masanao Yajima, Jingyuan Deng, Vivian Gersuk, Alex K. Shalek, Chloe K. Slichter, Hannah W. Miller, M. Juliana McElrath, Martin Prlic, et al. Mast: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell rna sequencing data. *Genome Biology*, 16(1):1–13, 2015. doi:10.1186/s13059-015-0844-5.

[206] Ken Kelley, Kristopher J Preacher. On effect size. *Psychological Methods*, 17(2):137–152, 2012. doi:10.1037/a0028086.

[207] Dirk Enzmann. Notes on effect size measures for the difference of means from two independent groups: The case of cohen's d and hedges' g, 2015. doi:10.13140/2.1.1578.2725.

[208] Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Routledge, 2013. doi:10.4324/9780203771587.

[209] Matthew S Thiese, Brenden Ronna, Ulrike Ott. P value interpretations and considerations. *Journal of Thoracic Disease*, 8(9):E928–E931, 2016. doi:10.21037/jtd.2016.08.16.

[210] Patrick Schober, Christa Boer, Lothar A Schwarte. Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia*, 126(5):1763–1768, 2018. doi:10.1213/ane.0000000000002864.

[211] Michael P Cary, Gary D Bader, Chris Sander. Pathway information for systems biology. *FEBS Letters*, 579(8):1815–1820, 2005. doi:10.1016/j.febslet.2005.02.005.

[212] M. Kanehisa. Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1):27–30, 2000. doi:10.1093/nar/28.1.27.

[213] Gene Ontology Consortium. The gene ontology (go) database and informatics resource. *Nucleic Acids Research*, 32(90001):258D–261, 2004. doi:10.1093/nar/gkh036.

[214] D. Croft, G. O'Kelly, G. Wu, R. Haw, M. Gillespie, L. Matthews, M. Caudy, P. Garapati, G. Gopinath, B. Jassal, et al. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Research*, 39(Database):D691–D697, 2010. doi:10.1093/nar/gkq1018.

[215] Purvesh Khatri, Marina Sirota, Atul J Butte. Ten years of pathway analysis: Current approaches and outstanding challenges. *PLoS Computational Biology*, 8(2):e1002375, 2012. doi:10.1371/journal.pcbi.1002375.

[216] Miguel A. García-Campos, Jesús Espinal-Enríquez, Enrique Hernández-Lemus. Pathway analysis: State of the art. *Frontiers in Physiology*, 6:383, 2015. doi:10.3389/fphys.2015.00383.

[217] Aravind Subramanian, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, Scott L. Pomeroy, Todd R. Golub, Eric S. Lander, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43):15545–15550, 2005. doi:10.1073/pnas.0506580102.

[218] Suoqin Jin, Christian F Guerrero-Juarez, Lihua Zhang, Ivan Chang, Raul Ramos, Chen-Hsiang Kuan, Peggy Myung, Maksim V Plikus, Qing Nie. Inference and analysis of cell-cell communication using cellchat. *Nature Communications*, 12(1):1088, 2021. doi:10.1038/s41467-021-21246-9.

[219] Mirjana Efremova, Miquel Vento-Tormo, Sarah A Teichmann, Roser Vento-Tormo. Cellphonedb: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nature Protocols*, 15(4):1484–1506, 2020. doi:10.1038/s41596-020-0292-x.

[220] Daniel Dimitrov, Philipp Sven Lars Schäfer, Elias Farr, Pablo Rodriguez-Mier, Sebastian Lobentanzer, Pau Badia-i Mompel, Aurelien Dugourd, Jovan Tanevski, Ricardo Omar Ramirez Flores, Julio Saez-Rodriguez. Liana+ provides an all-in-one framework for cell–cell communication inference. *Nature Cell Biology*, 26(9):1613–1622, 2024. doi:10.1038/s41556-024-01469-w.

[221] Orsolya Liska, Balázs Bohár, András Hidas, Tamás Korcsmáros, Balázs Papp, Dávid Fazekas, Eszter Ari. Tflink: an integrated gateway to access transcription factor–target gene interactions for multiple species. *Database*, 2022:baac083, 2022. doi:10.1093/database/baac083.

[222] Ieva Rauluseviciute, Rafael Riudavets-Puig, Romain Blanc-Mathieu, Jaime A Castro-Mondragon, Katalin Ferenc, Vipin Kumar, Roza Berhanu Lemma, Jérémy Lucas, Jeanne Chèneby, Damir Baranasic, et al. Jaspar 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Research*, 52(D1):D174–D182, 2023. doi:10.1093/nar/gkad1059.

[223] Vikram Agarwal, George W Bell, Jin-Wu Nam, David P Bartel. Predicting effective microrna target sites in mammalian mrnas. *eLife*, 4:e05005, 2015. doi:10.7554/elife.05005.

[224] Yuhao Chen, Xiaowei Wang. mirdb: an online database for prediction of functional microrna targets. *Nucleic Acids Research*, 48(D1):D127–D131, 2019. doi:10.1093/nar/gkz757.

[225] Fabian Kern, Ernesto Aparicio-Puerta, Yongping Li, Tobias Fehlmann, Tim Kehl, Viktoria Wagner, Kamalika Ray, Nicole Ludwig, Hans-Peter Lenhof, Eckart Meese, et al. mirtargetlink 2.0—interactive mirna target gene and target pathway networks. *Nucleic Acids Research*, 49(W1):W409–W416, 2021. doi:10.1093/nar/gkab297.

[226] Haidong Yi, Alec Plotkin, Natalie Stanley. Benchmarking differential abundance methods for finding condition-specific prototypical cells in multi-sample single-cell datasets. *Genome Biology*, 25(1):9, 2024. doi:10.1186/s13059-023-03143-0.

[227] Emma Dann, Neil C Henderson, Sarah A Teichmann, Michael D Morgan, John C Marioni. Differential abundance testing on single-cell data using k-nearest neighbor graphs. *Nature Biotechnology*, 40(2):245–253, 2021. doi:10.1038/s41587-021-01033-z.

[228] M. Büttner, J. Ostner, C. L. Müller, F. J. Theis, B. Schubert. sccoda is a bayesian model for compositional single-cell data analysis. *Nature Communications*, 12(1):6876, 2021. doi:10.1038/s41467-021-27150-6.

[229] Jun Zhao, Ariel Jaffe, Henry Li, Ofir Lindenbaum, Esen Sefik, Ruaidhrí Jackson, Xiuyuan Cheng, Richard A. Flavell, Yuval Kluger. Detection of differentially abundant cell subpopulations in scrna-seq data. *Proceedings of the National Academy of Sciences*, 118(22):e2100293118, 2021. doi:10.1073/pnas.2100293118.

[230] Daniel B. Burkhardt, Jay S. Stanley, Alexander Tong, Ana Luisa Perdigoto, Scott A. Gigante, Kevan C. Herold, Guy Wolf, Antonio J. Giraldez, David van Dijk, Smita Krishnaswamy. Quantifying the effect of experimental perturbations at single-cell resolution. *Nature Biotechnology*, 39(5):619–629, 2021. doi:10.1038/s41587-020-00803-5.

[231] Yakir A Reshef, Laurie Rumker, Joyce B Kang, Aparna Nathan, Ilya Korsunsky, Samira Asgari, Megan B Murray, D Branch Moody, Soumya Raychaudhuri. Co-varying neighborhood analysis identifies cell populations associated with phenotypes of interest from single-cell transcriptomics. *Nature Biotechnology*, 40(3):355–363, 2021. doi:10.1038/s41587-021-01066-4.

[232] Cole Trapnell, Davide Cacchiarelli, Xiaojie Qiu. Monocle: Cell counting, differential expression, and trajectory analysis for single-cell rna-seq experiments. *Bioconductor*, 2017.

[233] Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, Maria E. Kastriti, Peter Lönnerberg, Alessandro Furlan, et al. Rna velocity of single cells. *Nature*, 560(7719):494–498, 2018. doi:10.1038/s41586-018-0414-6.

[234] Volker Bergen, Marius Lange, Stefan Peidli, F Alexander Wolf, Fabian J Theis. Generalizing rna velocity to transient cell states through dynamical modeling. *Nature Biotechnology*, 38(12):1408–1414, August 2020. URL: http://dx.doi.org/10.1038/s41587-020-0591-3, doi:10.1038/s41587-020-0591-3.

[235] F. Alexander Wolf, Philipp Angerer, Fabian J. Theis. Scanpy: large-scale single-cell gene expression data analysis. *Genome Biology*, 19(1), 2018. doi:10.1186/s13059-017-1382-0.

[236] Felix Mölder, Kim Philipp Jablonski, Brice Letcher, Michael B. Hall, Christopher H. Tomkins-Tinch, Vanessa Sochat, Jan Forster, Soohyun Lee, Sven O. Twardziok, Alexander Kanitz, et al. Sustainable data analysis with snakemake. *F1000Research*, 10:33, 2021. doi:10.12688/f1000research.29032.2.

[237] Anaconda documentation. anaconda inc, 2020.

[238] Philip Scheltens, Bart De Strooper, Miia Kivipelto, Henne Holstege, Gael Chételat, Charlotte E Teunissen, Jeffrey Cummings, Wiesje M van der Flier. Alzheimer's disease. *The Lancet*, 397(10284):1577–1590, 2021. URL: https://doi.org:10, doi:10.1016/s0140-6736(20)32205-4.

[239] Bastiaan R Bloem, Michael S Okun, Christine Klein. Parkinson's disease. *The Lancet*, 397(10291):2284–2303, 2021. URL: https://doi.org:10, doi:10.1016/s0140-6736(21)00218-x.

[240] Li Gan, Mark R Cookson, Leonard Petrucelli, Albert R La Spada. Converging pathways in neurodegeneration, from genetics to mechanisms. *Nature Neuroscience*, 21(10):1300–1309, 2018. doi:10.1038/s41593-018-0237-7.

[241] Frank Jessen, Rebecca E Amariglio, Rachel F Buckley, Wiesje M van der Flier, Ying Han, José Luis Molinuevo, Laura Rabin, Dorene M Rentz, Octavio Rodriguez-Gomez, Andrew J Saykin, et al. The characterisation of subjective cognitive decline. *The Lancet Neurology*, 19(3):271–278, 2020. URL: https://doi.org:10, doi:10.1016/s1474-4422(19)30368-0.

[242] Serge Gauthier, Barry Reisberg, Michael Zaudig, Ronald C Petersen, Karen Ritchie, Karl Broich, Sylvie Belleville, Henry Brodaty, David Bennett, Howard Chertkow, et al. Mild cognitive impairment. *The Lancet*, 367(9518):1262–1270, 2006. URL: https://doi.org:10, doi:10.1016/s0140-6736(06)68542-5.

[243] Cecilia S Lindestam Arlehamn, Francesca Garretti, David Sulzer, Alessandro Sette. Roles for the adaptive immune system in parkinson's and alzheimer's diseases. *Current Opinion in Immunology*, 59:115–120, 2019. doi:10.1016/j.coi.2019.07.004.

[244] Richard M Ransohoff. How neuroinflammation contributes to neurodegeneration. *Science*, 353(6301):777–783, 2016. doi:10.1126/science.aag2590.

[245] Andrew C. Yang, Ryan T. Vest, Fabian Kern, Davis P. Lee, Maayan Agam, Christina A. Maat, Patricia M. Losada, Michelle B. Chen, Nicholas Schaum, Nathalie Khoury, et al. A human brain vascular atlas reveals diverse mediators of alzheimer's risk. *Nature*, 603(7903):885–892, 2022. doi:10.1038/s41586-021-04369-3.

[246] Francisco J. Garcia, Na Sun, Hyeseung Lee, Brianna Godlewski, Hansruedi Mathys, Kyriaki Galani, Blake Zhou, Xueqiao Jiang, Ayesha P. Ng, Julio Mantero, et al. Single-cell dissection of the human brain vasculature. *Nature*, 603(7903):893–899, 2022. doi:10.1038/s41586-022-04521-7.

[247] Harald Hampel, Sid E. O'Bryant, José L. Molinuevo, Henrik Zetterberg, Colin L. Masters, Simone Lista, Steven J. Kiddle, Richard Batrla, Kaj Blennow. Blood-based biomarkers for alzheimer disease: mapping the road to the clinic. *Nature Reviews Neurology*, 14(11):639–652, 2018. doi:10.1038/s41582-018-0079-7.

[248] David Gate, Naresha Saligrama, Olivia Leventhal, Andrew C. Yang, Michael S. Unger, Jinte Middeldorp, Kelly Chen, Benoit Lehallier, Divya Channappa, Mark B. De Los Santos, et al. Clonally expanded cd8 t cells patrol the cerebrospinal fluid in alzheimer's disease. *Nature*, 577(7790):399–404, 2020. doi:10.1038/s41586-019-1895-7.

[249] David Gate, Emma Tapp, Olivia Leventhal, Marian Shahid, Tim J. Nonninger, Andrew C. Yang, Katharina Strempfl, Michael S. Unger, Tobias Fehlmann, Hamilton Oh, et al. Cd4 + t cells contribute to neurodegeneration in lewy body dementia. *Science*, 374(6569):868–874, 2021. doi:10.1126/science.abf7266.

[250] William A. Banks, May J. Reed, Aric F. Logsdon, Elizabeth M. Rhea, Michelle A. Erickson. Healthy aging and the blood–brain barrier. *Nature Aging*, 1(3):243–254, 2021. doi:10.1038/s43587-021-00043-5.

[251] Melanie D. Sweeney, Kassandra Kisler, Axel Montagne, Arthur W. Toga, Berislav V. Zlokovic. The role of brain vasculature in neurodegenerative disorders. *Nature Neuroscience*, 21(10):1318–1331, 2018. URL: https://doi.org:10, doi:10.1038/s41593-018-0234-x.

[252] Melanie D. Sweeney, Abhay P. Sagare, Berislav V. Zlokovic. Blood–brain barrier breakdown in alzheimer disease and other neurodegenerative disorders. *Nature Reviews Neurology*, 14(3):133–150, 2018. doi:10.1038/nrneurol.2017.188.

[253] Andrew C. Yang, Marc Y. Stevens, Michelle B. Chen, Davis P. Lee, Daniel Stähli, David Gate, Kévin Contrepois, Winnie Chen, Tal Iram, Lichao Zhang, et al. Physiological blood–brain transport is impaired with age by a shift in transcytosis. *Nature*, 583(7816):425–430, 2020. URL: https://doi.org:10, doi:10.1038/s41586-020-2453-z.

[254] Jay Amin, Delphine Boche, Zoe Clough, Jessica Teeling, Anthony Williams, Yifang Gao, Lindsey Chudley, Laurie Lau, Florence Smith, Scott Harris, et al. Peripheral immunophenotype in dementia with lewy bodies and alzheimer's disease: an observational clinical study. *Journal of Neurology, Neurosurgery & Psychiatry*, 91(11):1219–1226, 9 2020. doi:10.1136/jnnp-2020-323603.

[255] Wei Cao, Hui Zheng. Peripheral immune system in aging and alzheimer's disease. *Molecular Neurodegeneration*, 13(1):51, 2018. doi:10.1186/s13024-018-0284-2.

[256] Caitlín Ní Chasaide, Marina A Lynch. The role of the immune system in driving neuroinflammation. *Brain and Neuroscience Advances*, 4:239821281990108, 2020. doi:10.1177/2398212819901082.

[257] Marco Prinz, Josef Priller. The role of peripheral immune cells in the cns in steady state and disease. *Nature Neuroscience*, 20(2):136–144, 2017. doi:10.1038/nn.4475.

[258] Hansruedi Mathys, Jose Davila-Velderrain, Zhuyu Peng, Fan Gao, Shahin Mohammadi, Jennie Z. Young, Madhvi Menon, Liang He, Fatema Abdurrob, Xueqiao Jiang, et al. Single-cell transcriptomic analysis of alzheimer's disease. *Nature*, 570(7761):332–337, 2019. doi:10.1038/s41586-019-1195-2.

[259] Marta Sochocka, Michał Ochnik, Maciej Sobczyński, Beata Orzechowska, Jerzy Leszek. Sex differences in innate immune response of peripheral blood leukocytes of alzheimer's disease patients. *Archivum Immunologiae et Therapiae Experimentalis*, 70(1):16, 6 2022. doi:10.1007/s00005-022-00653-w.

[260] Wencan Ji, Ke An, Canjun Wang, Shaohua Wang. Bioinformatics analysis of diagnostic biomarkers for alzheimer's disease in peripheral blood based on sex differences

and support vector machine algorithm. *Hereditas*, 159(1):38, 10 2022. doi:10.1186/s41065-022-00252-x.

[261] Isabelle Coales, Stergios Tsartsalis, Nurun Fancy, Maria Weinert, Daniel Clode, David Owen, Paul M. Matthews. Alzheimer's disease-related transcriptional sex differences in myeloid cells. *Journal of Neuroinflammation*, 19(1):247, 10 2022. doi:10.1186/s12974-022-02604-w.

[262] Samantha M. Carlisle, Hongwei Qin, R. Curtis Hendrickson, Jordana E. Muwanguzi, Elliot J. Lefkowitz, Richard E. Kennedy, Zhaoqi Yan, Talene A. Yacoubian, Etty N. Benveniste, Andrew B. West, et al. Sex-based differences in the activation of peripheral blood monocytes in early parkinson disease. *npj Parkinson's Disease*, 7(1):36, 2021. URL: https://doi.org:10, doi:10.1038/s41531-021-00180-z.

[263] Liu-Lin Xiong, Lu-Lu Xue, Ruo-Lan Du, Rui-Ze Niu, Li Chen, Jie Chen, Qiao Hu, Ya-Xin Tan, Hui-Fang Shang, Jia Liu, et al. Single-cell rna sequencing reveals b cell–related molecular biomarkers for alzheimer's disease. *Experimental & Molecular Medicine*, 53(12):1888–1901, 2021. doi:10.1038/s12276-021-00714-8.

[264] G.-R HU, R. S. Walls, H. Creasey, E. McCusker, G. A. Broe. Peripheral blood lymphocyte subset distribution and function in patients with alzheimer's disease and other dementias. *Australian and New Zealand Journal of Medicine*, 25(3):212–217, 1995. doi:10.1111/j.1445-5994.1995.tb01525.x.

[265] Tuula Pirttilä, Satu Mattinen, Harry Frey. The decrease of cd8-positive lymphocytes in alzheimer's disease. *Journal of the Neurological Sciences*, 107(2):160–165, 1992. doi:10.1016/0022-510x(92)90284-r.

[266] Kaneez Shad, Yashar Aghazadeh, Sagheer Ahmad, Bodo Kress. Peripheral markers of alzheimer's disease: Surveillance of white blood cells. *Synapse*, 67(8):541–543, 3 2013. doi:10.1002/syn.21651.

[267] Xinyi Wang, Guangqiang Sun, Teng Feng, Jing Zhang, Xun Huang, Tao Wang, Zuoquan Xie, Xingkun Chu, Jun Yang, Huan Wang, et al. Sodium oligomannate therapeutically remodels gut microbiota and suppresses gut bacterial amino acids-shaped neuroinflammation to inhibit alzheimer's disease progression. *Cell Research*, 29(10):787–803, 9 2019. doi:10.1038/s41422-019-0216-x.

[268] Xiao-Hang Qian, Xiao-Li Liu, Sheng-Di Chen, Hui-Dong Tang. Integrating peripheral blood and brain transcriptomics to identify immunological features associated with alzheimer's disease in mild cognitive impairment patients. *Frontiers in Immunology*, 13:986346, 9 2022. doi:10.3389/fimmu.2022.986346.

[269] Rui Li, Thomas Tropea, Laura Baratta, Leah Zuroff, Maria Diaz-Ortiz, Bo Zhang, Koji Shinoda, Ayman Rezk, Roy Alcalay, Alice Chen-Plotkin, et al. Abnormal b-cell and tfh-cell profiles in patients with parkinson disease. *Neurology Neuroimmunology & Neuroinflammation*, 9(2), 2022. doi:10.1212/nxi.0000000000001125.

[270] Xueping Chen, Weihua Feng, Ruwei Ou, Jiao Liu, Jing Yang, Jiajia Fu, Bei Cao, Yongping Chen, Qianqian Wei, Huifang Shang. Evidence for peripheral immune activation in parkinson's disease. *Frontiers in Aging Neuroscience*, 13:617370, 4 2021. doi:10.3389/fnagi.2021.617370.

[271] Natalia Pessoa Rocha, Frankcinéia Assis, Paula Luciana Scalzo, Érica Leandro Marciano Vieira, Izabela Guimarães Barbosa, Mariana Soares de Souza, Paulo Pereira Christo, Helton José Reis, Antonio Lucio Teixeira. Reduced activated t lymphocytes (cd4+cd25+) and plasma levels of cytokines in parkinson's disease. *Molecular Neurobiology*, 55(2):1488–1497, 2017. doi:10.1007/s12035-017-0404-y.

[272] S Garfias, B Tamaya Domínguez, A Toledo Rojas, M Arroyo, U Rodríguez, C Boll, A Sosa, E Sciutto, L Adalid-Peralta, Y Martinez López, et al. Peripheral blood lymphocyte phenotypes in alzheimer and parkinson's diseases. *Neurología (English Edition)*, 37(2):110–121, 2022. doi:10.1016/j.nrleng.2018.10.022.

[273] Natasa Kustrimovic, Cristoforo Comi, Luca Magistrelli, Emanuela Rasini, Massimiliano Legnaro, Raffaella Bombelli, Iva Aleksic, Fabio Blandini, Brigida Minafra, Giulio Riboldazzi, et al. Parkinson's disease patients have a complex phenotypic and functional th1 bias: cross-sectional studies of cd4+ th1/th2/t17 and treg in drug-naïve and drug-treated patients. *Journal of Neuroinflammation*, 15(1):205, 7 2018. doi:10.1186/s12974-018-1248-8.

[274] Sejal Patel, Derek Howard, Alana Man, Deborah Schwartz, Joelle Jee, Daniel Felsky, Zdenka Pausova, Tomas Paus, Leon French. Donor specific transcriptomic analysis of alzheimer's disease associated hypometabolism highlights a unique donor, microglia, and ribosomal proteins. *Ribosomal Proteins and Microglia. eNeuro*, 7(6), 2020. doi:10.1101/2019.12.23.887364.

[275] Daniel Felsky, Sanjeev Sariya, Ismael Santa-Maria, Leon French, Julie Schneider, David Bennett, Richard Mayeux, Philip De Jager, Giuseppe Tosto. The caribbean-hispanic alzheimer's brain transcriptome reveals ancestry-specific disease mechanisms. *Neurobiol Dis*, 176:105938, 2023. doi:10.1101/2020.05.28.122234.

[276] Yanxin Zhao, Wei Tan, Wenhua Sheng, Xiaohong Li. Identification of biomarkers associated with alzheimer's disease by bioinformatics analysis. *American Journal of Alzheimer's Disease & Other Dementias®*, 31(2):163–168, 2016. doi:10.1177/1533317515588181.

[277] Shirin Moradifard, Moslem Hoseinbeyki, Shahla Ganji, Zarrin Minuchehr. Analysis of microrna and gene expression profiles in alzheimer's disease: A meta-analysis approach. *Scientific Reports*, 8(1):4767, 3 2018. doi:10.1038/s41598-018-20959-0.

[278] Annika Sommer, Franz Marxreiter, Florian Krach, Tanja Fadler, Janina Grosch, Michele Maroni, Daniela Graef, Esther Eberhardt, Markus J. Riemenschneider, Gene W. Yeo, et al. Th17 lymphocytes induce neuronal cell death in a human ipsc-based model of parkinson's disease. *Cell Stem Cell*, 23(1):123–131.e6, 2018. doi:10.1016/j.stem.2018.06.015.

[279] M. Mogi, A. Togari, T. Kondo, Y. Mizuno, O. Komure, S. Kuno, H. Ichinose, T. Nagatsu. Caspase activities and tumor necrosis factor receptor r1 (p55) level are elevated in the substantia nigra from parkinsonian brain. *Journal of Neural Transmission*, 107(3):335–341, 2000. doi:10.1007/s007020050028.

[280] Ibrahim Abdulwahid Arif, Haseeb Ahmad Khan. Environmental toxins and parkinson's disease: Putative roles of impaired electron transport chain and oxidative stress. *Toxicology and Industrial Health*, 26(2):121–128, 2010. doi:10.1177/0748233710362382.

[281] Charmaine Lang, Kieran R. Campbell, Brent J. Ryan, Phillippa Carling, Moustafa Attar, Jane Vowles, Olga V. Perestenko, Rory Bowden, Fahd Baig, Meike Kasten, et al. Single-cell sequencing of ipsc-dopamine neurons reconstructs disease progression and identifies hdac4 as a regulator of parkinson cell phenotypes. *Cell Stem Cell*, 24(1):93–106.e6, 2019. doi:10.1016/j.stem.2018.10.023.

[282] Merina Varghese, Mritunjay Pandey, Ananda Samanta, Prasanta Kumar Gangopadhyay, Kochupurackal P. Mohanakumar. Reduced nadh coenzyme q dehydrogenase activity in platelets of parkinson's disease, but not parkinson plus patients, from an indian population. *Journal of the Neurological Sciences*, 279(1–2):39–42, 2009. doi:10.1016/j.jns.2008.12.021.

[283] Shaletha Holmes, Meharvan Singh, Chang Su, Rebecca L. Cunningham. Effects of oxidative stress and testosterone on pro-inflammatory signaling in a female rat dopaminergic neuronal cell line. *Endocrinology*, 157(7):2824–2835, 2016. doi:10.1210/en.2015-1738.

[284] Mina Maki, Noriyuki Matsukawa, Hiroyuki Yuasa, Yasushi Otsuka, Takayuki Yamamoto, Hiroyasu Akatsu, Takashi Okamoto, Ryuzo Ueda, Kosei Ojika. Decreased expression of hippocampal cholinergic neurostimulating peptide precursor protein mrna in the hippocampus in alzheimer disease. *Journal of Neuropathology & Experimental Neurology*, 61(2):176–185, 2002. doi:10.1093/jnen/61.2.176.

[285] Yasukuni Tsugu, Kosei Ojika, Noriyuki Matsukawa, Tamaki Iwase, Yasushi Otsuka, Eiichi Katada, Shigehisa Mitake. High levels of hippocampal cholinergic neurostimulating peptide (hdnp> in the csf of some patients with alzhemer's disease. *European Journal of Neurology*, 5(6):561–569, 1998. doi:10.1046/j.1468-1331.1998.560561.x.

[286] Johanna Nilsson, Johan Gobom, Simon Sjödin, Gunnar Brinkmalm, Nicholas J. Ashton, Johan Svensson, Per Johansson, Erik Portelius, Henrik Zetterberg, Kaj Blennow, et al. Cerebrospinal fluid biomarker panel for synaptic dysfunction in alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 13(1), 2021. doi:10.1002/dad2.12179.

[287] Stephen J. Kish, Iscia Lopes-Cendes, Mark Guttman, Yoshiaki Furukawa, Massimo Pandolfo, Guy A. Rouleau, Brian M. Ross, Martha Nance, Lawrence Schut, Lee Ang, et al. Brain glyceraldehyde-3-phosphate dehydrogenase activity in human trinucleotide repeat disorders. *Archives of Neurology*, 55(10):1299, 1998. doi:10.1001/archneur.55.10.1299.

[288] D. Allan Butterfield, Sarita S. Hardas, Miranda L. Bader Lange. Oxidatively modified glyceraldehyde-3-phosphate dehydrogenase (gapdh) and alzheimer's disease: Many pathways to neurodegeneration. *Journal of Alzheimer's Disease*, 20(2):369–393, 2010. doi:10.3233/jad-2010-1375.

[289] Chen Wei Tsai, Chia Fan Tsai, Kuan Hung Lin, Wei Jung Chen, Muh Shi Lin, Cho Chen Hsieh, Chai Ching Lin. An investigation of the correlation between the s-glutathionylated gapdh levels in blood and alzheimer's disease progression. *PLOS ONE*, 15(5):e0233289, 2020. doi:10.1371/journal.pone.0233289.

[290] Yini Dang, Qing He, Siyu Yang, Huaiqing Sun, Yin Liu, Wanting Li, Yi Tang, Yu Zheng, Ting Wu. Fth1- and sat1-induced astrocytic ferroptosis is involved in alzheimer's disease: Evidence from single-cell transcriptomic analysis. *Pharmaceuticals*, 15(10):1177, 2022. doi:10.3390/ph15101177.

[291] Srinu Tumpara, Matthias Ballmaier, Sabine Wrenger, Mandy König, Matthias Lehmann, Ralf Lichtinghagen, Beatriz Martinez-Delgado, Elena Korenbaum, David DeLuca, Nils Jedicke, et al. Polymerization of misfolded z alpha-1 antitrypsin protein lowers cx3cr1 expression in human pbmcs. *eLife*, 10, 2021. URL: https://doi.org:10, doi:10.7554/elife.64881.

[292] Daichi Shigemizu, Taiki Mori, Shintaro Akiyama, Sayuri Higaki, Hiroshi Watanabe, Takashi Sakurai, Shumpei Niida, Kouichi Ozaki. Identification of potential blood biomarkers for early diagnosis of alzheimer's disease through rna sequencing analysis. *Alzheimer's Research & Therapy*, 12(1):87, 2020. URL: https://doi.org:10, doi:10.1186/s13195-020-00654-x.

[293] Johannes C. M. Schlachetzki, Iryna Prots, Jenhan Tao, Hyun B. Chun, Kaoru Saijo, David Gosselin, Beate Winner, Christopher K. Glass, Jürgen Winkler. A monocyte gene expression signature in the early clinical course of parkinson's disease. *Scientific Reports*, 8(1):10757, 2018. doi:10.1038/s41598-018-28986-7.

[294] Sara Konstantin Nissen, Kristine Farmen, Mikkel Carstensen, Claudia Schulte, David Goldeck, Kathrin Brockmann, Marina Romero-Ramos. Changes in cd163+, cd11b+, and ccr2+ peripheral monocytes relate to parkinson's disease and cognition. *Brain, Behavior, and Immunity*, 101:182–193, 2022. doi:10.1016/j.bbi.2022.01.005.

[295] Fabian Kern, Tobias Fehlmann, Ivo Violich, Eric Alsop, Elizabeth Hutchins, Mustafa Kahraman, Nadja L. Grammes, Pedro Guimarães, Christina Backes, Kathleen L. Poston, et al. Deep sequencing of sncrnas reveals hallmarks and regulatory modules of the transcriptome during parkinson's disease progression. *Nature Aging*, 1(3):309–322, 2021. doi:10.1038/s43587-021-00042-6.

[296] Marta Milà-Alomà, Nicholas J. Ashton, Mahnaz Shekari, Gemma Salvadó, Paula Ortiz-Romero, Laia Montoliu-Gaya, Andrea L. Benedet, Thomas K. Karikari, Juan Lantero-Rodriguez, Eugeen Vanmechelen, et al. Plasma p-tau231 and p-tau217 as state markers of amyloid-β pathology in preclinical alzheimer's disease. *Nature Medicine*, 28:1797–1801, 2022. URL: https://doi.org:10, doi:10.1038/s41591-022-01925-w.

[297] Suzanne E. Schindler, Thomas K. Karikari. Comorbidities confound alzheimer's blood tests. *Nature Medicine*, 28(7):1349–1351, 2022. doi:10.1038/s41591-022-01875-3.

[298] Sebastian Palmqvist, Pontus Tideman, Nicholas Cullen, Henrik Zetterberg, Kaj Blennow, Jeffery L. Dage, Erik Stomrud, Shorena Janelidze, Niklas Mattsson-Carlgren, Oskar Hansson. Prediction of future alzheimer's disease dementia using plasma phospho-tau combined with other accessible measures. *Nature Medicine*, 27(6):1034–1042, 2021. URL: https://doi.org:10, doi:10.1038/s41591-021-01348-z.

[299] Randall J. Bateman, Nicolas R. Barthélemy, Kanta Horie. Another step forward in blood-based diagnostics for alzheimer's disease. *Nature Medicine*, 26(3):314–316, 2020. doi:10.1038/s41591-020-0797-4.

[300] Henrik Zetterberg, Jonathan M. Schott. Biomarkers for alzheimer's disease beyond amyloid and tau. *Nature Medicine*, 25(2):201–203, 2019. URL: https://doi.org:10, doi:10.1038/s41591-019-0348-z.

[301] Alexandra Grubman, Gabriel Chew, John F. Ouyang, Guizhi Sun, Xin Yi Choo, Catriona McLean, Rebecca K. Simmons, Sam Buckberry, Dulce B. Vargas-Landin, Daniel Poppe, et al. A single-cell atlas of entorhinal cortex from individuals with alzheimer's disease reveals cell-type-specific gene expression regulation. *Nature Neuroscience*, 22(12):2087–2097, 2019. doi:10.1038/s41593-019-0539-4.

[302] Thanaphong Phongpreecha, Rosemary Fernandez, Dunja Mrdjen, Anthony Culos, Chandresh R. Gajera, Adam M. Wawro, Natalie Stanley, Brice Gaudilliere, Kathleen L. Poston, Nima Aghaeepour, et al. Single-cell peripheral immunoprofiling of alzheimer's and parkinson's diseases. *Science Advances*, 6(48), 2020. doi:10.1126/sciadv.abd5575.

[303] Theodore S. Kapellos, Lorenzo Bonaguro, Ioanna Gemünd, Nico Reusch, Adem Saglam, Emily R. Hinkley, Joachim L. Schultze. Human monocyte subsets and phenotypes in major chronic inflammatory diseases. *Frontiers in Immunology*, 10:2035, 2019. doi:10.3389/fimmu.2019.02035.

[304] Vahid Gazestani, Tushar Kamath, Naeem M. Nadaf, Antonios Dougalis, S.J. Burris, Brendan Rooney, Antti Junkkari, Charles Vanderburg, Anssi Pelkonen, Mireia Gomez-Budia, et al. Early alzheimer's disease pathology in human cortex involves transient cell states. *Cell*, 186(20):4438–4453.e23, 2023. doi:10.1016/j.cell.2023.08.005.

[305] Na Sun, Matheus B. Victor, Yongjin P. Park, Xushen Xiong, Aine Ni Scannail, Noelle Leary, Shaniah Prosper, Soujanya Viswanathan, Xochitl Luna, Carles A. Boix, et al. Human microglial state dynamics in alzheimer's disease progression. *Cell*, 186(20):4386–4403.e29, 2023. doi:10.1016/j.cell.2023.08.037.

[306] Hansruedi Mathys, Zhuyu Peng, Carles A. Boix, Matheus B. Victor, Noelle Leary, Sudhagar Babu, Ghada Abdelhady, Xueqiao Jiang, Ayesha P. Ng, Kimia Ghafari, et al. Single-cell atlas reveals correlates of high cognitive function, dementia, and resilience to alzheimer's disease pathology. *Cell*, 186(20):4365–4385.e27, 2023. doi:10.1016/j.cell.2023.08.039.

[307] Mariano I. Gabitto, Kyle J. Travaglini, Victoria M. Rachleff, Eitan S. Kaplan, Brian Long, Jeanelle Ariza, Yi Ding, Joseph T. Mahoney, Nick Dee, Jeff Goldy, et al. Integrated multimodal cell atlas of alzheimer's disease. *Nature Neuroscience*, 27(12):2366–2383, 2024. URL: https://doi.org:10, doi:10.1038/s41593-024-01774-5.

[308] Andrew C. Yang, Fabian Kern, Patricia M. Losada, Maayan R. Agam, Christina A. Maat, Georges P. Schmartz, Tobias Fehlmann, Julian A. Stein, Nicholas Schaum, Davis P. Lee, et al. Dysregulation of brain and choroid plexus cell types in severe covid-19. *Nature*, 595(7868):565–571, 2021. doi:10.1038/s41586-021-03710-0.

[309] Xianwen Ren, Wen Wen, Xiaoying Fan, Wenhong Hou, Bin Su, Pengfei Cai, Jiesheng Li, Yang Liu, Fei Tang, Fan Zhang, et al. Covid-19 immune features revealed by a large-scale single-cell transcriptome atlas. *Cell*, 184(7):1895–1913.e19, 2021. doi:10.1016/j.cell.2021.01.053.

[310] Emily Stephenson, Gary Reynolds, Rachel A. Botting, Fernando J. Calero-Nieto, Michael D. Morgan, Zewen Kelvin Tuong, Karsten Bach, Waradon Sungnak, Kaylee B. Worlock, Masahiro Yoshida, et al. Single-cell multi-omics analysis of the immune response in covid-19. *Nature Medicine*, 27(5):904–916, 2021. doi:10.1038/s41591-021-01329-2.

[311] Masashi Fujita, Zongmei Gao, Lu Zeng, Cristin McCabe, Charles C. White, Bernard Ng, Gilad Sahar Green, Orit Rozenblatt-Rosen, Devan Phillips, Liat Amir-Zilberstein, et al. Cell subtype-specific effects of genetic variation in the alzheimer's disease brain. *Nature Genetics*, 56(4):605–614, 2024. URL: https://doi.org:10, doi:10.1038/s41588-024-01685-y.

[312] Daniel W. Sirkis, Caroline Warly Solsberg, Taylor P. Johnson, Luke W. Bonham, Alexis P. Oddi, Ethan G. Geier, Bruce L. Miller, Gil D. Rabinovici, Jennifer S. Yokoyama. Expansion of highly interferon-responsive t cells in early-onset alzheimer's disease. *Alzheimer's & Dementia*, 20(7):5062–5070, 2024. doi:10.1002/alz.13892.

[313] Pingping Wang, Meng Luo, Wenyang Zhou, Xiyun Jin, Zhaochun Xu, Shi Yan, Yiqun Li, Chang Xu, Rui Cheng, Yan Huang, et al. Global characterization of peripheral b cells in parkinson's disease by single-cell rna and bcr sequencing. *Frontiers in Immunology*, 13, 2022. doi:10.3389/fimmu.2022.814239.

[314] Pingping Wang, Lifen Yao, Meng Luo, Wenyang Zhou, Xiyun Jin, Zhaochun Xu, Shi Yan, Yiqun Li, Chang Xu, Rui Cheng, et al. Single-cell transcriptome and tcr profiling reveal activated and expanded t cell populations in parkinson's disease. *Cell Discovery*, 7(1), 2021. doi:10.1038/s41421-021-00280-3.

[315] Shi Yan, Yao Si, Wenyang Zhou, Rui Cheng, Pingping Wang, Di Wang, Wencai Ding, Wanying Shi, Qinghua Jiang, Fan Yang, et al. Single-cell transcriptomics reveals the interaction between peripheral cd4+ ctls and mesencephalic endothelial cells mediated by ifng in parkinson's disease. *Computers in Biology and Medicine*, 158:106801, 2023. doi:10.1016/j.compbiomed.2023.106801.

[316] Hui Xu, Jianping Jia. Single-cell rna sequencing of peripheral blood reveals immune cell signatures in alzheimer's disease. *Frontiers in Immunology*, 12, 2021. doi:10.3389/fimmu.2021.645666.

[317] Yanjun Lu, Ke Li, Yu Hu, Xiong Wang. Expression of immune related genes and possible regulatory mechanisms in alzheimer's disease. *Frontiers in Immunology*, 12, 2021. doi:10.3389/fimmu.2021.768966.

[318] Xueyi Zhang, Lissette Gomez, Jennifer E. Below, Adam C. Naj, Eden R. Martin, Brian W. Kunkle, William S. Bush. An x chromosome transcriptome wide association study implicates armcx6 in alzheimer's disease. *Journal of Alzheimer's Disease*, 98(3):1053–1067, 2024. doi:10.3233/jad-231075.

[319] Sulev Koks, Abigail L Pfaff, Vivien J Bubb, John P Quinn. Longitudinal intronic rna-seq analysis of parkinson's disease patients reveals disease-specific nascent transcription. *Experimental Biology and Medicine*, 247(11):945–957, 2022. doi:10.1177/15353702221081027.

[320] David W. Craig, Elizabeth Hutchins, Ivo Violich, Eric Alsop, J. Raphael Gibbs, Shawn Levy, Madison Robison, Nripesh Prasad, Tatiana Foroud, Karen L. Crawford, et al. Rna sequencing of whole blood reveals early alterations in immune cells and gene expression in parkinson's disease. *Nature Aging*, 1(8):734–747, 2021. doi:10.1038/s43587-021-00088-6.

[321] Giulietta Maria Riboldi, Ricardo A. Vialle, Elisa Navarro, Evan Udine, Katia de Paiva Lopes, Jack Humphrey, Amanda Allan, Madison Parks, Brooklyn Henderson, Kelly Astudillo, et al. Transcriptome deregulation of peripheral monocytes and whole blood in gba-related parkinson's disease. *Molecular Neurodegeneration*, 17(1), 2022. doi:10.1186/s13024-022-00554-8.

[322] Liting Song, Yucheng T. Yang, Qihao Guo, Xing-Ming Zhao. Cellular transcriptional alterations of peripheral blood in alzheimer's disease. *BMC Medicine*, 20(1), 2022. doi:10.1186/s12916-022-02472-4.

[323] Huan Zhong, Xiaopu Zhou, Hyebin Uhm, Yuanbing Jiang, Han Cao, Yu Chen, Tiffany T. W. Mak, Ronnie Ming Nok Lo, Bonnie Wing Yan Wong, Elaine Yee Ling Cheng, et al. Using blood transcriptome analysis for alzheimer's disease diagnosis and patient stratification. *Alzheimer's & Dementia*, 20(4):2469–2484, 2024. doi:10.1002/alz.13691.

[324] Anthony J. Griswold, Sathesh K. Sivasankaran, Derek Van Booven, Olivia K. Gardner, Farid Rajabli, Patrice L. Whitehead, Kara L. Hamilton-Nelson, Larry D. Adams, Aja M. Scott, Natalia K. Hofmann, et al. Immune and inflammatory pathways implicated by whole blood transcriptomic analysis in a diverse ancestry alzheimer's disease cohort. *Journal of Alzheimer's Disease*, 76(3):1047–1060, 2020. doi:10.3233/jad-190855.

[325] Olivia K Gardner, Lily Wang, Derek Van Booven, Patrice L Whitehead, Kara L Hamilton-Nelson, Larry D Adams, Takiyah D Starks, Natalia K Hofmann, Jeffery M Vance, Michael L Cuccaro, et al. Rna editing alterations in a multi-ethnic alzheimer disease cohort converge on immune and endocytic molecular pathways. *Human Molecular Genetics*, 28(18):3053–3061, 2019. doi:10.1093/hmg/ddz110.

[326] Kosar Hooshmand, Glenda M. Halliday, Sandy S. Pineda, Greg T. Sutherland, Boris Guennewig. Overlap between central and peripheral transcriptomes in parkinson's disease but not alzheimer's disease. *International Journal of Molecular Sciences*, 23(9):5200, 2022. doi:10.3390/ijms23095200.

[327] Liting Song, Jingqi Chen, Chun-Yi Zac Lo, Qihao Guo, Jianfeng Feng, Xing-Ming Zhao. Impaired type i interferon signaling activity implicated in the peripheral blood transcriptome of preclinical alzheimer's disease. *eBioMedicine*, 82:104175, 2022. doi:10.1016/j.ebiom.2022.104175.

[328] Rebecca Panitch, Junming Hu, Weiming Xia, David A. Bennett, Thor D. Stein, Lindsay A. Farrer, Gyungah R. Jun. Blood and brain transcriptome analysis reveals apoe genotype-mediated and immune-related pathways involved in alzheimer disease. *Alzheimer's Research & Therapy*, 14(1), 2022. doi:10.1186/s13195-022-00975-z.

[329] Sian M. J. Hemmings, Patricia Swart, Jacqueline S. Womersely, Ellen S. Ovenden, Leigh L. van den Heuvel, Nathaniel W. McGregor, Stuart Meier, Soraya Bardien, Shameemah Abrahams, Gerard Tromp, et al. Rna-seq analysis of gene expression profiles in posttraumatic stress disorder, parkinson's disease and schizophrenia identifies roles for common and distinct biological pathways. *Discover Mental Health*, 2(1), 2022. doi:10.1007/s44192-022-00009-y.

[330] Dallin Dressman, Thomas Buttrick, Maria Cimpean, David Bennett, Vilas Menon, Elizabeth M Bradshaw, Badri Vardarajan, Wassim Elyaman. Genotype–phenotype correlation of t-cell subtypes reveals senescent and cytotoxic genes in alzheimer's disease. *Human Molecular Genetics*, 31(19):3355–3366, 2022. doi:10.1093/hmg/ddac126.

[331] Jon Infante, Carlos Prieto, María Sierra, Pascual Sánchez-Juan, Isabel González-Aramburu, Coro Sánchez-Quintana, José Berciano, Onofre Combarros, Jesús Sainz. Identification of candidate genes for parkinson's disease through blood transcriptome analysis in lrrk2-g2019s carriers, idiopathic cases, and controls. *Neurobiology of Aging*, 36(2):1105–1109, 2015. doi:10.1016/j.neurobiolaging.2014.10.039.

[332] Rekha Dhanwani, João Rodrigues Lima-Junior, Ashu Sethi, John Pham, Gregory Williams, April Frazier, Yaqian Xu, Amy W. Amara, David G. Standaert, Jennifer G. Goldman, et al. Transcriptional analysis of peripheral memory t cells reveals parkinson's disease-specific gene signatures. *npj Parkinson's Disease*, 8(1):30, 2022. doi:10.1038/s41531-022-00282-2.

[333] Jon Infante, Carlos Prieto, María Sierra, Pascual Sánchez-Juan, Isabel González-Aramburu, Coro Sánchez-Quintana, José Berciano, Onofre Combarros, Jesús Sainz. Comparative blood transcriptome analysis in idiopathic and lrrk2 g2019s–associated parkinson's disease. *Neurobiology of Aging*, 38:214.e1–214.e5, 2016. doi:10.1016/j.neurobiolaging.2015.10.026.

[334] Sarah Jane Annesley, Claire Yvonne Allan, Oana Sanislav, Andrew Evans, Paul Robert Fisher. Dysregulated gene expression in lymphoblasts from parkinson's disease. *Proteomes*, 10(2):20, 2022. doi:10.3390/proteomes10020020.

[335] Maria Garofalo, Cecilia Pandini, Matteo Bordoni, Orietta Pansarasa, Federica Rey, Alfredo Costa, Brigida Minafra, Luca Diamanti, Susanna Zucca, Stephana Carelli, et al. Alzheimer's, parkinson's disease and amyotrophic lateral sclerosis gene expression patterns divergence reveals different grade of rna metabolism involvement. *International Journal of Molecular Sciences*, 21(24):9500, 2020. doi:10.3390/ijms21249500.

[336] Adrienne R. Henderson, Qi Wang, Bessie Meechoovet, Ashley L. Siniard, Marcus Naymik, Matthew De Both, Matthew J. Huentelman, Richard J. Caselli, Erika Driver-Dunckley, Travis Dunckley. Dna methylation and expression profiles of whole blood in parkinson's disease. *Frontiers in Genetics*, 12, 2021. doi:10.3389/fgene.2021.640266.

[337] Lille Kurvits, Freddy Lättekivi, Ene Reimann, Liis Kadastik-Eerme, Kristjan M Kasterpalu, Sulev Kõks, Pille Taba, Anu Planken. Transcriptomic profiles in parkinson's disease. *Experimental Biology and Medicine*, 246(5):584–595, 2020. doi:10.1177/1535370220967325.

[338] Yong Hu, Kunshan Zhang, Tianyu Zhang, Junbang Wang, Fei Chen, Wenting Qin, Weifang Tong, Qiang Guan, Yijing He, Chunya Gu, et al. Exercise reverses dysregulation of t-cell-related function in blood leukocytes of patients with parkinson's disease. *Frontiers in Neurology*, 10, 2020. doi:10.3389/fneur.2019.01389.

[339] Kelly Cardona, Javier Medina, Mary Orrego-Cardozo, Francia Restrepo de Mejía, Xabier Elcoroaristizabal, Carlos Andrés Naranjo Galvis. Inflammatory gene expression profiling in peripheral blood from patients with alzheimer's disease reveals key pathways and hub genes with potential diagnostic utility: a preliminary study. *PeerJ*, 9:e12016, 2021. doi:10.7717/peerj.12016.

[340] Li Wang, Chunjiang Yu, Ye Tao, Xiumei Yang, Qiao Jiang, Haiyu Yu, Jiejun Zhang. Transcriptome analysis reveals potential marker genes for diagnosis of alzheimer's disease and vascular dementia. *Frontiers in Genetics*, 13, 2022. doi:10.3389/fgene.2022.1038585.

[341] Chuanyu Liu, Tao Wu, Fei Fan, Ya Liu, Liang Wu, Michael Junkin, Zhifeng Wang, Yeya Yu, Weimao Wang, Wenbo Wei, et al. A portable and cost-effective microfluidic system

for massively parallel single-cell transcriptome profiling. *BioRxiv*, page 818450, 2019. doi:10.1101/818450.

[342] Alexander Dobin, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, Thomas R. Gingeras. Star: ultrafast universal rna-seq aligner. *Bioinformatics*, 29(1):15–21, 2012. doi:10.1093/bioinformatics/bts635.

[343] Artem Tarasov, Albert J. Vilella, Edwin Cuppen, Isaac J. Nijman, Pjotr Prins. Sambamba: fast processing of ngs alignment formats. *Bioinformatics*, 31(12):2032–2034, 2015. doi:10.1093/bioinformatics/btv098.

[344] Adam Frankish, Mark Diekhans, Irwin Jungreis, Julien Lagarde, Jane E Loveland, Jonathan M Mudge, Cristina Sisu, James C Wright, Joel Armstrong, If Barnes, et al. Gencode 2021. *Nucleic Acids Research*, 49(D1):D916–D923, 2020. doi:10.1093/nar/gkaa1087.

[345] Quan Shi, Shiping Liu, Karsten Kristiansen, Longqi Liu. The fastq+ format and pisa. *Bioinformatics*, 38(19):4639–4642, 2022. doi:10.1093/bioinformatics/btac562.

[346] Matthias Flotho, Jérémy Amand, Pascal Hirsch, Friederike Grandke, Tony Wyss-Coray, Andreas Keller, Fabian Kern. Zebra: a hierarchically integrated gene expression atlas of the murine and human brain at single-cell resolution. *Nucleic Acids Research*, 52(D1):D1089–D1096, 2023. doi:10.1093/nar/gkad990.

[347] Nico Gerstner, Tim Kehl, Kerstin Lenhof, Anne Müller, Carolin Mayer, Lea Eckhart, Nadja Liddy Grammes, Caroline Diener, Martin Hart, Oliver Hahn, et al. Genetrail 3: advanced high-throughput enrichment analysis. *Nucleic Acids Research*, 48(W1):W515–W520, 2020. doi:10.1093/nar/gkaa306.

[348] Zhaohao Huang, Binyao Chen, Xiuxing Liu, He Li, Lihui Xie, Yuehan Gao, Runping Duan, Zhaohuai Li, Jian Zhang, Yingfeng Zheng, et al. Effects of sex and aging on the immune cell landscape as assessed by single-cell transcriptomic analysis. *Proceedings of the National Academy of Sciences*, 118(33):e2023216118, 2021. doi:10.1073/pnas.2023216118.

[349] Hugo Lövheim, Jonathan Gilthorpe, Anders Johansson, Sture Eriksson, Göran Hallmans, Fredrik Elgh. Herpes simplex infection and the risk of alzheimer's disease: A nested case-control study. *Alzheimer's & Dementia*, 11(6):587–592, 2015. doi:10.1016/j.jalz.2014.07.157.

[350] Morgane Linard, Luc Letenneur, Isabelle Garrigue, Angélique Doize, Jean-François Dartigues, Catherine Helmer. Interaction between apoe4 and herpes simplex virus type 1 in alzheimer's disease. *Alzheimer's & Dementia*, 16(1):200–208, 2020. doi:10.1002/alz.12008.

[351] Suoqin Jin, Maksim V. Plikus, Qing Nie. Cellchat for systematic analysis of cell–cell communication from single-cell transcriptomics. *Nature Protocols*, 20(1):180–219, 2024. doi:10.1038/s41596-024-01045-4.

[352] Hecheng Wang, Yu Zong, Lei Zhu, Weiyi Wang, Yanshuo Han. Chemokines in patients with alzheimer's disease: A meta-analysis. *Frontiers in Aging Neuroscience*, 15, 2023. doi:10.3389/fnagi.2023.1047810.

[353] Min-Seok Kim, Kwangmin Cho, Mi-Hyang Cho, Na-Young Kim, Kyunggon Kim, Dong-Hou Kim, Seung-Yong Yoon. Neuronal mhc-i complex is destabilized by amyloid-β and its implications in alzheimer's disease. *Cell & Bioscience*, 13(1):181, 2023. doi:10.1186/s13578-023-01132-1.

[354] Yuyan Zhu, Huayan Hou, Kavon Rezai-Zadeh, Brian Giunta, Amanda Ruscin, Carmelina Gemma, JingJi Jin, Natasa Dragicevic, Patrick Bradshaw, Suhail Rasool, et al. Cd45 deficiency drives amyloid-β peptide oligomers and neuronal loss in alzheimer's disease mice. *The Journal of Neuroscience*, 31(4):1355–1365, 2011. doi:10.1523/jneurosci.3268-10.2011.

[355] Yukie Kashima, Keiya Kaneko, Patrick Reteng, Nina Yoshitake, Lucky Ronald Runtuwene, Satoi Nagasawa, Masaya Onishi, Masahide Seki, Ayako Suzuki, Sumio Sugano, et al. Intensive single-cell analysis reveals immune-cell diversity among healthy individuals. *Life Science Alliance*, 5(7):e202201398, 2022. URL: https://doi.org:10, doi:10.26508/lsa.202201398.

[356] Léon-Charles Tranchevent, Rashi Halder, Enrico Glaab. Systems level analysis of sex-dependent gene expression changes in parkinson's disease. *npj Parkinson's Disease*, 9(1), 2023. doi:10.1038/s41531-023-00446-8.

[357] Sejal Patel, Derek Howard, Alana Man, Deborah Schwartz, Joelle Jee, Daniel Felsky, Zdenka Pausova, Tomas Paus, Leon French. Donor-specific transcriptomic analysis of alzheimer's disease-associated hypometabolism highlights a unique donor, ribosomal proteins and microglia. *eneuro*, 7(6):ENEURO.0255–20.2020, 2020. doi:10.1523/eneuro.0255-20.2020.

[358] Dani Flinkman, Ye Hong, Jelena Gnjatovic, Prasannakumar Deshpande, Zsuzsanna Ortutay, Sirkku Peltonen, Valtteri Kaasinen, Peter James, Eleanor Coffey. Regulators of proteostasis are translationally repressed in fibroblasts from patients with sporadic and lrrk2-g2019s parkinson's disease. *npj Parkinson's Disease*, 9(1):20, 2023. URL: https://doi.org:10, doi:10.1038/s41531-023-00460-w.

[359] Sarah M. Hernandez, Elena B. Tikhonova, Kristen R. Baca, Fanpeng Zhao, Xiongwei Zhu, Andrey L. Karamyshev. Unexpected implication of srp and ago2 in parkinson's disease: Involvement in alpha-synuclein biogenesis. *Cells*, 10(10):2792, 2021. URL: https://doi.org:10, doi:10.3390/cells10102792.

[360] Daniel Felsky, Ismael Santa-Maria, Mehmet Ilyas Cosacak, Leon French, Julie A. Schneider, David A. Bennett, Philip L. De Jager, Caghan Kizil, Giuseppe Tosto. The caribbean-hispanic alzheimer's disease brain transcriptome reveals ancestry-specific disease mechanisms. *Neurobiology of Disease*, 176:105938, 2023. doi:10.1016/j.nbd.2022.105938.

[361] Vladimir Ilievski, Paulina K. Zuchowska, Stefan J. Green, Peter T. Toth, Michael E. Ragozzino, Khuong Le, Haider W. Aljewari, Neil M. O'Brien-Simpson, Eric C. Reynolds, Keiko Watanabe. Chronic oral application of a periodontal pathogen results in brain inflammation, neurodegeneration and amyloid beta production in wild type mice. *PLOS ONE*, 13(10):e0204941, 2018. doi:10.1371/journal.pone.0204941.

[362] Stephen S. Dominy, Casey Lynch, Florian Ermini, Malgorzata Benedyk, Agata Marczyk, Andrei Konradi, Mai Nguyen, Ursula Haditsch, Debasish Raha, Christina Griffin, et al. Porphyromonas gingivalis in alzheimer's disease brains: Evidence for disease causation and treatment with small-molecule inhibitors. *Science Advances*, 5(1), 2019. doi:10.1126/sciadv.aau3333.

[363] H.W. Jones. The recent large reduction in space launch cost. In *48th International Conference on Environmental Systems*. NASA Ames Research Center, 2018.

[364] Raghavan Gopalakrishnan, Kerim O. Genc, Andrea J. Rice, Stuart M. C. Lee, Harlan J. Evans, Christian C. Maender, Hakan Ilaslan, Peter R. Cavanagh. Muscle volume, strength, endurance, and exercise loads during 6-month missions in space. *Aviation, Space, and Environmental Medicine*, 81(2):91–104, 2010. doi:10.3357/asem.2583.2010.

[365] Merja A. Perhonen, Fatima Franco, Lynda D. Lane, Jay C. Buckey, C. Gunnar Blomqvist, Joseph E. Zerwekh, Ronald M. Peshock, Paul T. Weatherall, Benjamin D. Levine. Cardiac atrophy after bed rest and spaceflight. *Journal of Applied Physiology*, 91(2):645–653, 2001. doi:10.1152/jappl.2001.91.2.645.

[366] Scott Trappe, David Costill, Philip Gallagher, Andrew Creer, Jim R. Peters, Harlan Evans, Danny A. Riley, Robert H. Fitts. Exercise in space: human skeletal muscle after

6 months aboard the international space station. *Journal of Applied Physiology*, 106(4):1159–1168, 2009. doi:10.1152/japplphysiol.91578.2008.

[367] Victor A. Convertino. *Exercise and adaptation to microgravity environments*. Handbook of physiology ; sect. 4. Published for the American Physiological Society by Oxford University Press, 1996.

[368] Ginger Flores, Danny Harris, Rachel McCauley, Shane Canerday, Lindsey Ingram, Nicole Herrmann. Deep space habitation: Establishing a sustainable human presence on the moon and beyond. In *2021 IEEE Aerospace Conference (50100)*, pages 1–7. IEEE, 2021. doi:10.1109/aero50100.2021.9438260.

[369] Francine E. Garrett-Bakelman, Manjula Darshi, Stefan J. Green, Ruben C. Gur, Ling Lin, Brandon R. Macias, Miles J. McKenna, Cem Meydan, Tejaswini Mishra, Jad Nasrini, et al. The nasa twins study: A multidimensional analysis of a year-long human spaceflight. *Science*, 364(6436), 2019. doi:10.1126/science.aau8650.

[370] M.R. Barratt, S.L. Pool. Principles of clinical medicine for space flight, 2008.

[371] Afshin Beheshti, Jack Miller, Yared Kidane, Daniel Berrios, Samrawit G Gebre, Sylvain V Costes. Nasa genelab project: Bridging space radiation omics with ground studies. *Radiation Research*, 189(6):553–559, 2018. doi:10.1667/rr15062.1.

[372] M. Muratani. *Cell-free RNA analysis of plasma samples collected from six astronauts in JAXA Cell-Free Epigenome (CFE) Study*. NASA Open Science Data Repository, 2022.

[373] Nailil Husna, Tatsuya Aiba, Shin-Ichiro Fujita, Yoshika Saito, Dai Shiba, Takashi Kudo, Satoru Takahashi, Satoshi Furukawa, Masafumi Muratani. Release of cd36-associated cell-free mitochondrial dna and rna as a hallmark of space environment response. *Nature Communications*, 15(1):4814, 2024. doi:10.1038/s41467-023-41995-z.

[374] Michael A. Schmidt, Caleb M. Schmidt, Robert M. Hubbard, Christopher E. Mason. Why personalized medicine is the frontier of medicine and performance for humans in space. *New Space*, 8(2):63–76, 2020. doi:10.1089/space.2019.0037.

[375] Christopher W. Jones, Eliah G. Overbey, Jerome Lacombe, Adrian J. Ecker, Cem Meydan, Krista Ryon, Braden Tierney, Namita Damle, Matthew MacKay, Evan E. Afshin, et al. Molecular and physiological changes in the spacex inspiration4 civilian crew. *Nature*, 632(8027):1155–1164, 2024. doi:10.1038/s41586-024-07648-x.

[376] Eliah G. Overbey, Krista Ryon, JangKeun Kim, Braden T. Tierney, Remi Klotz, Veronica Ortiz, Sean Mullane, Julian C. Schmidt, Matthew MacKay, Namita Damle, et al. Collection of biospecimens from the inspiration4 mission establishes the standards for the space omics and medical atlas (soma). *Nature Communications*, 15(1):4964, 2024. doi:10.1038/s41467-024-48806-z.

[377] Eliah G. Overbey, JangKeun Kim, Braden T. Tierney, Jiwoon Park, Nadia Houerbi, Alexander G. Lucaci, Sebastian Garcia Medina, Namita Damle, Deena Najjar, Kirill Grigorev, et al. The space omics and medical atlas (soma) and international astronaut biobank. *Nature*, 632(8027):1145–1154, 2024. doi:10.1038/s41586-024-07639-y.

[378] Samrawit G Gebre, Ryan T Scott, Amanda M Saravia-Butler, Danielle K Lopez, Lauren M Sanders, Sylvain V Costes. Nasa open science data repository: open science for life in space. *Nucleic Acids Research*, 53(D1):D1697–D1710, 2024. doi:10.1093/nar/gkae1116.

[379] Afshin Beheshti, Shayoni Ray, Homer Fogle, Daniel Berrios, Sylvain V. Costes. A microrna signature and tgf-β1 response were identified as the key master regulators for spaceflight response. *PLOS ONE*, 13(7):e0199621, 2018. doi:10.1371/journal.pone.0199621.

[380] Sherina Malkani, Christopher R. Chin, Egle Cekanaviciute, Marie Mortreux, Hazeem Okinula, Marcel Tarbier, Ann-Sofie Schreurs, Yasaman Shirazi-Fard, Candice G.T. Tahimic, Deyra N. Rodriguez, et al. Circulating mirna spaceflight signature reveals targets for countermeasure development. *Cell Reports*, 33(10):108448, 2020. doi:10.1016/j.celrep.2020.108448.

[381] A. Yu Ratushnyy, L. B Buravkova. Microgravity effects and aging physiology: similar changes or common mechanisms? *Biochemistry (Mosc*, 88(11):2138–2155, 2023. doi:10.31857/s032097252311009x.

[382] Holly E. Kinser, Zachary Pincus. Micrornas as modulators of longevity and the aging process. *Human Genetics*, 139(3):291–308, 2019. doi:10.1007/s00439-019-02046-0.

[383] Viktoria Wagner, Fabian Kern, Oliver Hahn, Nicholas Schaum, Nicole Ludwig, Tobias Fehlmann, Annika Engel, Dominic Henn, Shusruto Rishik, Alina Isakova, et al. Characterizing expression changes in noncoding rnas during aging and heterochronic parabiosis across mouse tissues. *Nature Biotechnology*, 42(1):109–118, 2023. doi:10.1038/s41587-023-01751-6.

[384] Aristeidis G. Telonis, Rogan Magee, Phillipe Loher, Inna Chervoneva, Eric Londin, Isidore Rigoutsos. Knowledge about the presence or absence of mirna isoforms (isomirs) can successfully discriminate amongst 32 tcga cancer types. *Nucleic Acids Research*, 45(6):2973–2985, 2017. doi:10.1093/nar/gkx082.

[385] A. K. Aissiou, S. Jha, K. Dhunnoo, Z. Ma, D. X. Li, R. Ravin, M. Kunze, K. Wong, A. B. Adesida. Transcriptomic response of bioengineered human cartilage to parabolic flight microgravity is sex-dependent. *npj Microgravity*, 9(1), 2023. doi:10.1038/s41526-023-00255-6.

[386] Hyun Hwang, Antonio Rampoldi, Parvin Forghani, Dong Li, Jordan Fite, Gene Boland, Kevin Maher, Chunhui Xu. Space microgravity increases expression of genes associated with proliferation and differentiation in human cardiac spheres. *NPJ Microgravity*, 9(1):88, 2023. doi:10.1038/s41526-023-00336-6.

[387] Francesca Ferranti, Marta Del Bianco, Claudia Pacelli. Advantages and limitations of current microgravity platforms for space biology research. *Applied Sciences*, 11(1):68, 2020. doi:10.3390/app11010068.

[388] N. Schaum. Single-cell transcriptomics of 20 mouse organs creates a tabula muris. *Nature*, 562(7727):367–372, 2018. doi:10.1038/s41586-018-0590-4.

[389] Nicole Almanzar, Jane Antony, Ankit S. Baghel, Isaac Bakerman, Ishita Bansal, Ben A. Barres, Philip A. Beachy, Daniela Berdnik, Biter Bilen, Douglas Brownfield, et al. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature*, 583(7817):590–595, 2020. doi:10.1038/s41586-020-2496-1.

[390] Nicholas Schaum, Benoit Lehallier, Oliver Hahn, Robert Palovics, Shayan Hosseinzadeh, Song E Lee, Rene Sit, Davis P Lee, Patricia Moran Losada, Macy E Zardeneta, et al. Ageing hallmarks exhibit organ-specific temporal signatures. *Yearbook of Paediatric Endocrinology*, 583:596–602, 2021. doi:10.1530/ey.18.14.11.

[391] Róbert Pálovics, Andreas Keller, Nicholas Schaum, Weilun Tan, Tobias Fehlmann, Michael Borja, Fabian Kern, Liana Bonanno, Kruti Calcuttawala, James Webber, et al. Molecular hallmarks of heterochronic parabiosis at single-cell resolution. *Nature*, 603(7900):309–314, 2022. doi:10.1038/s41586-022-04461-2.

[392] Judith L. Morgan, Karen L. Svenson, Jeffrey P. Lake, Weidong Zhang, Timothy M. Stearns, Michael A. Marion, Luanne L. Peters, Beverly Paigen, Leah Rae Donahue. Effects of housing density in five inbred strains of mice. *PLoS ONE*, 9(3):e90012, 2014. doi:10.1371/journal.pone.0090012.

[393] Linda A. Toth. The influence of the cage environment on rodent physiology and behavior: Implications for reproducibility of pre-clinical rodent research. *Experimental Neurology*, 270:72–77, 2015. doi:10.1016/j.expneurol.2015.04.010.

[394] Gian C. Demontis, Marco M. Germani, Enrico G. Caiani, Ivana Barravecchia, Claudio Passino, Debora Angeloni. Human pathophysiological adaptations to the space environment. *Frontiers in Physiology*, 8:547, 2017. doi:10.3389/fphys.2017.00547.

[395] Satoshi Iwase, Naoki Nishimura, Kunihiko Tanaka, Tadaaki Mano. *Effects of Microgravity on Human Physiology*, page 2. IntechOpen, IntechOpen, Rijeka, 2020. doi:10.5772/intechopen.90700.

[396] Daniel Stratis, Guy Trudel, Lynda Rocheleau, Martin Pelchat, Odette Laneuville. The transcriptome response of astronaut leukocytes to long missions aboard the international space station reveals immune modulation. *Frontiers in Immunology*, 14:1171103, 2023. doi:10.3389/fimmu.2023.1171103.

[397] Miriam Capri, Maria Conte, Erika Ciurca, Chiara Pirazzini, Paolo Garagnani, Aurelia Santoro, Federica Longo, Stefano Salvioli, Patrick Lau, Ralf Moeller, et al. Long-term human spaceflight and inflammaging: Does it promote aging? *Ageing Research Reviews*, 87:101909, 2023. doi:10.1016/j.arr.2023.101909.

[398] Nicholas A Vernice, Cem Meydan, Ebrahim Afshinnekoo, Christopher E Mason. Long-term spaceflight and the cardiovascular system. *Precision Clinical Medicine*, 3(4):284–291, 2020. doi:10.1093/pcmedi/pbaa022.

[399] Marta Murgia, Jörn Rittweger, Carlo Reggiani, Roberto Bottinelli, Matthias Mann, Stefano Schiaffino, Marco V. Narici. Spaceflight on the iss changed the skeletal muscle proteome of two astronauts. *npj Microgravity*, 10(1):60 10 1038 41526–024–00406–3, 2024. doi:10.1038/s41526-024-00406-3.

[400] Chayakrit Krittanawong, Nitin Kumar Singh, Richard A. Scheuring, Emmanuel Urquieta, Eric M. Bershad, Timothy R. Macaulay, Scott Kaplin, Carly Dunn, Stephen F. Kry, Thais Russomano, et al. Human health during space travel: State-of-the-art review. *Cells*, 12(1):40, 2022. doi:10.3390/cells12010040.

[401] Ronni Baran, Shannon Marchal, Sebastian Garcia Campos, Emil Rehnberg, Kevin Tabury, Bjorn Baselet, Markus Wehland, Daniela Grimm, Sarah Baatout. The cardiovascular system in space: Focus on in vivo and in vitro studies. *Biomedicines*, 10(1):59, 2021. doi:10.3390/biomedicines10010059.

[402] Charles K.F. Chan, Eun Young Seo, James Y. Chen, David Lo, Adrian McArdle, Rahul Sinha, Ruth Tevlin, Jun Seita, Justin Vincent-Tompkins, Taylor Wearda, et al. Identification and specification of the mouse skeletal stem cell. *Cell*, 160(1–2):285–298, 2015. doi:10.1016/j.cell.2014.12.002.

[403] Gunsagar S Gulati, Matthew P Murphy, Owen Marecic, Michael Lopez, Rachel E Brewer, Lauren S Koepke, Anoop Manjunath, Ryan C Ransom, Ankit Salhotra, Irving L Weissman, et al. Isolation and functional assessment of mouse skeletal stem cell lineage. *Nature Protocols*, 13(6):1294–1309, 2018. doi:10.1038/nprot.2018.041.

[404] Yuhan Hao, Stephanie Hao, Erica Andersen-Nissen, William M. Mauck, Shiwei Zheng, Andrew Butler, Maddie J. Lee, Aaron J. Wilk, Charlotte Darby, Michael Zager, et al. Integrated analysis of multimodal single-cell data. *Cell*, 184(13):3573–3587.e29, 2021. doi:10.1016/j.cell.2021.04.048.

[405] Guangchuang Yu, Li-Gen Wang, Yanyan Han, Qing-Yu He. clusterprofiler: an r package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology*, 16(5):284–287, 2012. doi:10.1089/omi.2011.0118.

[406] Ana Kozomara, Maria Birgaoanu, Sam Griffiths-Jones. mirbase: from microrna sequences to function. *Nucleic Acids Research*, 47(D1):D155–D162, 2018. doi:10.1093/nar/gky1141.

[407] Alexander W Clarke, Eirik Høye, Anju Angelina Hembrom, Vanessa Molin Paynter, Jakob Vinther, Łukasz Wyrożemski, Inna Biryukova, Alessandro Formaggioni, Vladimir Ovchinnikov, Holger Herlyn, et al. Mirgenedb 3.0: improved taxonomic sampling, uniform nomenclature of novel conserved microrna families and updated covariance models. *Nucleic Acids Research*, 53(D1):D116–D128, 2024. doi:10.1093/nar/gkae1094.

[408] Helena L. Crowell, Charlotte Soneson, Pierre-Luc Germain, Daniela Calini, Ludovic Collin, Catarina Raposo, Dheeraj Malhotra, Mark D. Robinson. muscat detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data. *Nature Communications*, 11(1):6077 10 1038 41467–020–19894–4, 2020. doi:10.1038/s41467-020-19894-4.

[409] Ernesto Aparicio-Puerta, Pascal Hirsch, Georges P Schmartz, Fabian Kern, Tobias Fehlmann, Andreas Keller. mieaa 2023: updates, new functional microrna sets and improved enrichment visualizations. *Nucleic Acids Research*, 51(W1):W319–W325, 2023. doi:10.1093/nar/gkad392.

[410] Hsi-Yuan Huang, Yang-Chi-Dung Lin, Shidong Cui, Yixian Huang, Yun Tang, Jiatong Xu, Jiayang Bao, Yulin Li, Jia Wen, Huali Zuo, et al. mirtarbase update 2022: an informative resource for experimentally validated mirna–target interactions. *Nucleic Acids Research*, 50(D1):D222–D230, 2021. doi:10.1093/nar/gkab1079.

[411] Shidong Cui, Sicong Yu, Hsi-Yuan Huang, Yang-Chi-Dung Lin, Yixian Huang, Bojian Zhang, Jihan Xiao, Huali Zuo, Jiayi Wang, Zhuoran Li, et al. mirtarbase 2025: updates to the collection of experimentally validated microrna–target interactions. *Nucleic Acids Research*, 53(D1):D147–D156, 2024. doi:10.1093/nar/gkae1072.

[412] Carsten Sticht, Carolina De La Torre, Alisha Parveen, Norbert Gretz. mirwalk: An online resource for prediction of microrna binding sites. *PLOS ONE*, 13(10):e0206239, 2018. doi:10.1371/journal.pone.0206239.

[413] Shuangbin Xu, Erqiang Hu, Yantong Cai, Zijing Xie, Xiao Luo, Li Zhan, Wenli Tang, Qianwen Wang, Bingdong Liu, Rui Wang, et al. Using clusterprofiler to characterize multiomics data. *Nature Protocols*, 19(11):3292–3320, 2024. doi:10.1038/s41596-024-01020-z.

[414] Daniel Greene, Sylvia Richardson, Ernest Turro. ontologyx: a suite of r packages for working with ontological data. *Bioinformatics*, 33(7):1104–1106, 2016. doi:10.1093/bioinformatics/btw763.

[415] Zuguang Gu, Daniel Hübschmann. Simplifyenrichment: A bioconductor package for clustering and visualizing functional enrichment results. *Genomics, Proteomics & Bioinformatics*, 21(1):190–202, 2022. doi:10.1016/j.gpb.2022.04.008.

[416] Sean E. McGeary, Kathy S. Lin, Charlie Y. Shi, Thy M. Pham, Namita Bisaria, Gina M. Kelley, David P. Bartel. The biochemical basis of microrna targeting efficacy. *Science*, 366(6472), 2019. doi:10.1126/science.aav1741.

[417] Hassaan Maan, Lin Zhang, Chengxin Yu, Michael J Geuenich, Kieran R Campbell, Bo Wang. Characterizing the impacts of dataset imbalance on single-cell data integration. *Nature Biotechnology*, 42(12):1899–1908, 2024. doi:10.1038/s41587-023-02097-9.

[418] Shusen Zhang, Zhigang Cai, Hui Li. Ahnaks roles in physiology and malignant tumors. *Frontiers in Oncology*, 13:1258951, 2023. doi:10.3389/fonc.2023.1258951.

[419] Gelan Miao, Yulian Yang, Xuelian Yang, Dexiu Chen, Li Liu, Xianying Lei. The multifaceted potential of tpt1 as biomarker and therapeutic target. *Heliyon*, 10(19):e38819, 2024. doi:10.1016/j.heliyon.2024.e38819.

[420] Gayatri Arun, Disha Aggarwal, David L. Spector. Malat1 long non-coding rna: Functional implications. *Non-Coding RNA*, 6(2):22, 2020. doi:10.3390/ncrna6020022.

[421] Elena Andreeva, Diana Matveeva, Olga Zhidkova, Ivan Zhivodernikov, Oleg Kotov, Ludmila Buravkova. Real and simulated microgravity: Focus on mammalian extracellular matrix. *Life*, 12(9):1343, 2022. doi:10.3390/life12091343.

[422] Nathan J. White, Andrew Wenthe. Managing hemostasis in space. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 43(11):2079–2087, 2023. doi:10.1161/atvbaha.123.318783.

[423] Marissé Masís Solano, Remy Dumas, Mark R Lesk, Santiago Costantino. Ocular biomechanical responses to long-duration spaceflight. *IEEE Open Journal of Engineering in Medicine and Biology*, 6:127–132, 2025. doi:10.1109/ojemb.2024.3453049.

[424] F. Strollo, S. Gentile, A. M. V. Pipicelli, A. Mambro, M. Monici, P. Magni. Space flight-promoted insulin resistance as a possible disruptor of wound healing. *Frontiers in Bioengineering and Biotechnology*, 10:868999, 2022. doi:10.3389/fbioe.2022.868999.

[425] Yizhou Liu, Xiaojian Cao, Qiuzhi Zhou, Chunchu Deng, Yujie Yang, Danxia Huang, Hongmei Luo, Song Zhang, Yajie Li, Jia Xu, et al. Mechanisms and countermeasures for muscle atrophy in microgravity. *Cells*, 13(24):2120, 2024. doi:10.3390/cells13242120.

[426] Janice M. Fritsch-Yelle, Urs A. Leuenberger, Dominick S. D'Aunno, Alfred C. Rossum, Troy E. Brown, Margie L. Wood, Mark E. Josephson, Ary L. Goldberger. An episode of ventricular tachycardia during long-duration spaceflight. *The American Journal of Cardiology*, 81(11):1391–1392, 1998. doi:10.1016/s0002-9149(98)00179-9.

[427] Yu-Ping Wang, Kuo-Bin Li. Correlation of expression profiles between microrna and mrna targets using nci-60 data. *BMC Genomics*, 10(1):218, 2009. doi:10.1186/1471-2164-10-218.

[428] Adam E. Ziemann, Jason E. Allen, Nader S. Dahdaleh, Iuliia I. Drebot, Matthew W. Coryell, Amanda M. Wunsch, Cynthia M. Lynch, Frank M. Faraci, Matthew A. Howard, Michael J. Welsh, et al. The amygdala is a chemosensor that detects carbon dioxide and acidosis to elicit fear behavior. *Cell*, 139(5):1012–1021, 2009. URL: https://doi.org/10.1016/j.cell.2009.10.029., doi:10.1016/j.cell.2009.10.029.

[429] D. Elkon, R.E. Fechner, M.J. Homzie, D.G. Baker, W.C. Constable. Response of mouse kidney to hyperthermia: pathology and temperature–dependence of injury. *Arch Pathol Lab Med*, 104:153–158, 1980.

[430] K. Weber, J. Kiefer. Inhibition of ribosomal rna synthesis in yeast by ionizing radiations. *International Journal of Radiation Biology and Related Studies in Physics, Chemistry and Medicine*, 46(6):691–702, 1984. doi:10.1080/09553008414551931.

[431] Khaled Y. Kamal, Raúl Herranz, Jack J.W.A. van Loon, F. Javier Medina. Cell cycle acceleration and changes in essential nuclear functions induced by simulated microgravity in a synchronized arabidopsis cell culture. *Plant, Cell & Environment*, 42(2):480–494, 2018. doi:10.1111/pce.13422.

[432] Sergey V. Rozhkov, Kristina A. Sharlo, Timur M. Mirzoev, Boris S. Shenkman. Temporal changes in the markers of ribosome biogenesis in rat soleus muscle under simulated microgravity. *Acta Astronautica*, 186:252–258, 2021. doi:10.1016/j.actaastro.2021.05.036.

[433] Alberto Benguría, Enrique Grande, Emilio De Juan, Cristina Ugalde, Jaime Miquel, Rafael Garesse, Roberto Marco. Microgravity effects on drosophila melanogaster behavior and aging. implications of the iml-2 experiment. *Journal of Biotechnology*, 47(2–3):191–201, 1996. doi:10.1016/0168-1656(96)01407-1.

[434] Jie Huangfu, Hye Su Kim, Ke Xu, Xiaoyu Ning, Lei Qin, Jun Li, Chun Li. Omics analysis reveals the mechanism of enhanced recombinant protein production under simulated microgravity. *Frontiers in Bioengineering and Biotechnology*, 8:30, 2020. doi:10.3389/fbioe.2020.00030.

[435] Yongqing Liu, Eugenia Wang. Transcriptional analysis of normal human fibroblast responses to microgravity stress. *Genomics, Proteomics & Bioinformatics*, 6(1):29–41, 2008. doi:10.1016/s1672-0229(08)60018-2.

[436] Otto J. Juhl, Evan G. Buettmann, Michael A. Friedman, Rachel C. DeNapoli, Gabriel A. Hoppock, Henry J. Donahue. Update on the effects of microgravity on the musculoskeletal system. *npj Microgravity*, 7(1):28, 2021. doi:10.1038/s41526-021-00158-4.

[437] Rahel Schnellmann, Sharon Gerecht. Reconstructing the ageing extracellular matrix. *Nature Reviews Bioengineering*, 1(7):458–459, 2023. doi:10.1038/s44222-023-00049-1.

[438] Trent Davis, Kevin Tabury, Shouan Zhu, Debora Angeloni, Sarah Baatout, Alexandra Benchoua, Juergen Bereiter-Hahn, Daniele Bottai, Judith-Irina Buchheim, Marco Calvaruso, et al. How are cell and tissue structure and function influenced by gravity and what are the gravity perception mechanisms? *npj Microgravity*, 10(1):10 1038 41526–024–00357–9, 2024. doi:10.1038/s41526-024-00357-9.

[439] Cong Tao, Hongyan Ren, Pan Xu, Jia Cheng, Sujuan Huang, Rong Zhou, Yulian Mu, Shulin Yang, Desheng Qi, Yanfang Wang, et al. Adipocyte mir-200b/a/429 ablation in mice leads to high-fat-diet-induced obesity. *Oncotarget*, 7(42):67796–67807, 2016. doi:10.18632/oncotarget.12080.

[440] Catriona Hilton, Matt J. Neville, Laura B.L. Wittemans, Marijana Todorcevic, Katherine E. Pinnick, Sara L. Pulit, Jian'an Luan, Agné Kulyté, Ingrid Dahlman, Nicholas J. Wareham, et al. Microrna-196a links human body fat distribution to adipose tissue extracellular matrix composition. *eBioMedicine*, 44:467–475, 2019. doi:10.1016/j.ebiom.2019.05.047.

[441] Richard L. Hughson, Andrew D. Robertson, Philippe Arbeille, J. Kevin Shoemaker, James W. E. Rush, Katelyn S. Fraser, Danielle K. Greaves. Increased postflight carotid artery stiffness and inflight insulin resistance resulting from 6-mo spaceflight in male and female astronauts. *American Journal of Physiology-Heart and Circulatory Physiology*, 310(5):H628–H638, 2016. doi:10.1152/ajpheart.00802.2015.

[442] Brian W Tobin, Peter N Uchakin, Sandra K Leeper-Woodford. Insulin secretion and sensitivity in space flight. *Nutrition*, 18(10):842–848, 2002. doi:10.1016/s0899-9007(02)00940-1.

[443] A.E. Proshchina, Y.S. Krivova, S.C. Saveliev. Pancreas of c57 black mice after long-term space flight (bion-m1 space mission). *Life Sciences in Space Research*, 7:22–26, 2015. URL: https://doi.org/10.1016/j.lssr.2015.09.001., doi:10.1016/j.lssr.2015.09.001.

[444] Vijay Kumar, Upendra Kumar Soni, Vineet Kumar Maurya, Kiran Singh, Rajesh Kumar Jha. Integrin beta8 (itgb8) activates vav-rac1 signaling via fak in the acquisition of endometrial epithelial cell receptivity for blastocyst implantation. *Scientific Reports*, 7(1):10 1038 41598–017–01764–7, 2017. doi:10.1038/s41598-017-01764-7.

[445] Virgilio G. Ponferrada, Jieqing Fan, Jefferson E. Vallance, Shengyong Hu, Aygun Mamedova, Scott A. Rankin, Matthew Kofron, Aaron M. Zorn, Rashmi S. Hegde, Richard A. Lang. Crim1 complexes with ß-catenin and cadherins, stabilizes cell-cell junctions and is critical for neural morphogenesis. *PLoS ONE*, 7(3):e32635, 2012. doi:10.1371/journal.pone.0032635.

[446] Yajie Zhao, Maria Chukanova, Katherine A. Kentistou, Zammy Fairhurst-Hunter, Anna Maria Siegert, Raina Y. Jia, Georgina K. C. Dowsett, Eugene J. Gardner, Katherine Lawler, Felix R. Day, et al. Protein-truncating variants in bsn are associated with severe

adult-onset obesity, type 2 diabetes and fatty liver disease. *Nature Genetics*, 56(4):579–584, 2024. doi:10.1038/s41588-024-01694-x.

[447] Raviv Dharan, Yuwei Huang, Sudheer Kumar Cheppali, Shahar Goren, Petr Shendrik, Weisi Wang, Jiamei Qiao, Michael M. Kozlov, Li Yu, Raya Sorkin. Tetraspanin 4 stabilizes membrane swellings and facilitates their maturation into migrasomes. *Nature Communications*, 14(1), 2023. doi:10.1038/s41467-023-36596-9.

[448] Qiang Wang, Yan Chun Li, Jinhua Wang, Juan Kong, Yuchen Qi, Richard J. Quigg, Xinmin Li. mir-17-92 cluster accelerates adipocyte differentiation by negatively regulating tumor-suppressor rb2/p130. *Proceedings of the National Academy of Sciences*, 105(8):2889–2894, 2008. doi:10.1073/pnas.0800178105.

[449] Lili Tian, Zhuolun Song, Weijuan Shao, William W Du, Lisa R Zhao, Kejing Zeng, Burton B Yang, Tianru Jin. Curcumin represses mouse 3t3-l1 cell adipogenic differentiation via inhibiting mir-17-5p and stimulating the wnt signalling pathway effector tcf7l2. *Cell Death & Disease*, 8(1):e2559–e2559, 2017. doi:10.1038/cddis.2016.455.

[450] Zinger Yang Loureiro, Shannon Joyce, Tiffany DeSouza, Javier Solivan-Rivera, Anand Desai, Pantos Skritakis, Qin Yang, Rachel Ziegler, Denise Zhong, Tammy T. Nguyen, et al. Wnt signaling preserves progenitor cell multipotency during adipose tissue development. *Nature Metabolism*, 5(6):1014–1028, 2023. doi:10.1038/s42255-023-00813-y.

[451] Ruizhi Wang, Nipun Chopra, Kwangsik Nho, Bryan Maloney, Alexander G. Obukhov, Peter T. Nelson, Scott E. Counts, Debomoy K. Lahiri. Human microrna (mir-20b-5p) modulates alzheimer's disease pathways and neuronal function, and a specific polymorphism close to the mir20b gene influences alzheimer's biomarkers. *Molecular Psychiatry*, 27(2):1256–1273, 2022. doi:10.1038/s41380-021-01351-3.

[452] Meysam Amidfar, Gholamreza Askari, Yong-Ku Kim. Association of metabolic dysfunction with cognitive decline and alzheimer's disease: A review of metabolomic evidence. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 128:110848, 2024. doi:10.1016/j.pnpbp.2023.110848.

[453] Amin Safa, Zahra Bahroudi, Hamed Shoorei, Jamal Majidpoor, Atefe Abak, Mohammad Taheri, Soudeh Ghafouri-Fard. mir-1: A comprehensive review of its role in normal development and diverse disorders. *Biomedicine & Pharmacotherapy*, 132:110903, 2020. doi:10.1016/j.biopha.2020.110903.

[454] Kunyu Liao, Pengda Chen, Mengdi Zhang, Jiazhen Wang, Teri Hatzihristidis, Xiaoxi Lin, Liang Yang, Nan Yao, Chenfeng Liu, Yazhen Hong, et al. Critical roles of the mir-17~92 family in thymocyte development, leukemogenesis, and autoimmunity. *Cell Reports*, 43(6):114261, 2024. doi:10.1016/j.celrep.2024.114261.

[455] J. Tyson McDonald, JangKeun Kim, Lily Farmerie, Meghan L. Johnson, Nidia S. Trovao, Shehbeel Arif, Keith Siew, Sergey Tsoy, Yaron Bram, Jiwoon Park, et al. Space radiation damage rescued by inhibition of key spaceflight associated mirnas. *Nature Communications*, 15(1):4825 10 1038 41467–024–48920–, 2024. doi:10.1038/s41467-024-48920-y.

[456] Pieter Van Vlierberghe, Teresa Palomero, Hossein Khiabanian, Joni Van der Meulen, Mireia Castillo, Nadine Van Roy, Barbara De Moerloose, Jan Philippé, Sara González-García, María L Toribio, et al. Phf6 mutations in t-cell acute lymphoblastic leukemia. *Nature Genetics*, 42(4):338–342, 2010. doi:10.1038/ng.542.

[457] P Van Vlierberghe, J Patel, O Abdel-Wahab, C Lobry, C V Hedvat, M Balbin, C Nicolas, A R Payer, H F Fernandez, M S Tallman, et al. Phf6 mutations in adult acute myeloid leukemia. *Leukemia*, 25(1):130–134, 2010. doi:10.1038/leu.2010.247.

[458] Mathieu Lemaire, Véronique Frémeaux-Bacchi, Franz Schaefer, Murim Choi, Wai Ho Tang, Moglie Le Quintrec, Fadi Fakhouri, Sophie Taque, François Nobili, Frank Martinez,

et al. Recessive mutations in dgke cause atypical hemolytic-uremic syndrome. *Nature Genetics*, 45(5):531–536, 2013. doi:10.1038/ng.2590.

[459] R. J. Watson, P. J. Dyson, J. McMahon. Multiple c-myb transcript cap sites are variously utilized in cells of mouse haemopoietic origin. *The EMBO Journal*, 6(6):1643–1651, 1987. URL: https://doi.org/10.1002/j.1460-2075.1987.tb02413.x., doi:10.1002/j.1460-2075.1987.tb02413.x.

[460] Ryan P Sullivan, Jeffrey W Leong, Stephanie E Schneider, Aaron R Ireland, Melissa M Berrien-Elliott, Anvita Singh, Timothy Schappe, Brea A Jewell, Veronika Sexl, Todd A Fehniger. Microrna-15/16 antagonizes myb to control nk cell maturation. *The Journal of Immunology*, 195(6):2806–2817, 2015. doi:10.4049/jimmunol.1500949.

[461] Sébastien S. Dufresne, Nicolas A. Dumont, Antoine Boulanger-Piette, Val A. Fajardo, Daniel Gamu, Sandrine-Aurélie Kake-Guena, Rares Ovidiu David, Patrice Bouchard, Éliane Lavergne, Josef M. Penninger, et al. Muscle rank is a key regulator of ca2+ storage, serca activity, and function of fast-twitch skeletal muscles. *American Journal of Physiology-Cell Physiology*, 310(8):C663–C672, 2016. doi:10.1152/ajpcell.00285.2015.

[462] Guixing Ma, Siyuan Cheng, Yingying Han, Wanze Tang, Wei Pang, Litong Chen, Zhen Ding, Huiling Cao. The p53-mir17 family-rankl axis bridges liver-bone communication. *Molecular Therapy*, 33(2):631–648, 2025. doi:10.1016/j.ymthe.2024.12.046.

[463] Koji Oishi, Michihiko Aramaki, Kazunori Nakajima. Mutually repressive interaction between brn1/2 and rorb contributes to the establishment of neocortical layer 2/3 and layer 4. *Proceedings of the National Academy of Sciences*, 113(12):3371–3376, 2016. doi:10.1073/pnas.1515949113.

[464] Kelsey L. Hickey, Sharan Swarup, Ian R. Smith, Julia C. Paoli, Enya Miguel Whelan, Joao A. Paulo, J. Wade Harper. Proteome census upon nutrient stress reveals golgiphagy membrane receptors. *Nature*, 623(7985):167–174, 2023. doi:10.1038/s41586-023-06657-6.

[465] Daniel W. Sirkis, Caroline Warly Solsberg, Taylor P. Johnson, Luke W. Bonham, Virginia E. Sturm, Suzee E. Lee, Katherine P. Rankin, Howard J. Rosen, Adam L. Boxer, William W. Seeley, et al. Single-cell rna-seq reveals alterations in peripheral cx3cr1 and nonclassical monocytes in familial tauopathy. *Genome Medicine*, 15(1), July 2023. doi:10.1186/s13073-023-01205-3.

[466] Anna Aiello, Farzin Farzaneh, Giuseppina Candore, Calogero Caruso, Sergio Davinelli, Caterina Maria Gambino, Mattia Emanuela Ligotti, Nahid Zareian, Giulia Accardi. Immunosenescence and its hallmarks: How to oppose aging strategically? a review of potential options for therapeutic intervention. *Frontiers in Immunology*, 10, September 2019. doi:10.3389/fimmu.2019.02247.

[467] Janko Nikolich-Žugich. The twilight of immunity: emerging concepts in aging of the immune system. *Nature Immunology*, 19(1):10–19, December 2017. doi:10.1038/s41590-017-0006-x.

[468] Graham Pawelec. Immunosenescence and cancer. *Biogerontology*, 18(4):717–721, February 2017. doi:10.1007/s10522-017-9682-z.

[469] Alyssa Soles, Adem Selimovic, Kaelin Sbrocco, Ferris Ghannoum, Katherine Hamel, Emmanuel Labrada Moncada, Stephen Gilliat, Marija Cvetanovic. Extracellular matrix regulation in physiology and in brain disease. *International Journal of Molecular Sciences*, 24(8):7049, April 2023. doi:10.3390/ijms24087049.

[470] Mengjie Chen, Qi Zhan, Zepeng Mu, Lili Wang, Zhaohui Zheng, Jinlin Miao, Ping Zhu, Yang I. Li. Alignment of single-cell rna-seq samples without overcorrection using kernel density matching. *Genome Research*, 31(4):698–712, March 2021. doi:10.1101/gr.261115.120.

[471] Verena Haage, Philip L. De Jager. Neuroimmune contributions to alzheimer's disease: a focus on human data. *Molecular Psychiatry*, 27(8):3164–3181, 2022. doi:10.1038/s41380-022-01637-0.

[472] Ketaki A. Katdare, Andrew Kjar, Natasha M. O'Brown, Emma H. Neal, Alexander G. Sorets, Alena Shostak, Wilber Romero-Fernandez, Alexander J. Kwiatkowski, Kate Mlouk, Hyosung Kim, et al. Iqgap2 regulates blood-brain barrier immune dynamics. *iScience*, 28(3):111994, 2025. doi:10.1016/j.isci.2025.111994.

[473] Rong Wang, Hongyan Li. A focus on cxcr4 in alzheimer's disease. *Brain Circulation*, 3(4):199, 2017. doi:10.4103/bc.bc_13_17.

[474] Kathryn J Bryan, Xiongwei Zhu, Peggy L Harris, George Perry, Rudy J Castellani, Mark A Smith, Gemma Casadesus. Expression of cd74 is increased in neurofibrillary tangles in alzheimer's disease. *Molecular Neurodegeneration*, 3(1), 2008. doi:10.1186/1750-1326-3-13.

[475] Huiting Su, Ning Na, Xiaodong Zhang, Yong Zhao. The biological function and significance of cd74 in immune diseases. *Inflammation Research*, 66(3):209–216, 2016. doi:10.1007/s00011-016-0995-1.

[476] Marie-Victoire Guillot-Sestier, Terrence Town. Innate immunity in alzheimer's disease: A complex affair. *CNS & Neurological Disorders - Drug Targets*, 12(5):593–607, 2013. doi:10.2174/18715273113120500008.

[477] Brian F. Corbett, Jason C. You, Xiaohong Zhang, Mark S. Pyfer, Umberto Tosi, Daniel M. Iascone, Iraklis Petrof, Anupam Hazra, Chia-Hsuan Fu, Gabriel S. Stephens, et al. δfosb regulates gene expression and cognitive dysfunction in a mouse model of alzheimer's disease. *Cell Reports*, 20(2):344–355, 2017. doi:10.1016/j.celrep.2017.06.040.

[478] Yingchun Ling, Lingmin Hu, Jie Chen, Mingyong Zhao, Xinyang Dai. The mechanism of mitochondrial metabolic gene pmaip1 involved in alzheimer's disease process based on bioinformatics analysis and experimental validation. *Clinics*, 79:100373, 2024. doi:10.1016/j.clinsp.2024.100373.

[479] Wenjia Liu, Sophia Chen, Xin Rao, Yisong Yang, Xiaodong Chen, Liyang Yu. The inflammatory gene pycard of the entorhinal cortex as an early diagnostic target for alzheimer's disease. *Biomedicines*, 11(1):194, 2023. doi:10.3390/biomedicines11010194.

[480] Jinchao Hou, Yun Chen, Zhangying Cai, Gyu Seong Heo, Carla M. Yuede, Zuoxu Wang, Kent Lin, Fareeha Saadi, Tihana Trsan, Aivi T. Nguyen, et al. Antibody-mediated targeting of human microglial leukocyte ig-like receptor b4 attenuates amyloid pathology in a mouse model. *Science Translational Medicine*, 16(741), 2024. doi:10.1126/scitranslmed.adj9052.

[481] Ethan R. Roy, Baiping Wang, Ying-wooi Wan, Gabriel Chiu, Allysa Cole, Zhuoran Yin, Nicholas E. Propson, Yin Xu, Joanna L. Jankowsky, Zhandong Liu, et al. Type i interferon response drives neuroinflammation and synapse loss in alzheimer disease. *Journal of Clinical Investigation*, 130(4):1912–1930, 2020. doi:10.1172/jci133737.

[482] Céline Bellenguez, Fahri Küçükali, Iris E. Jansen, Luca Kleineidam, Sonia Moreno-Grau, Najaf Amin, Adam C. Naj, Rafael Campos-Martin, Benjamin Grenier-Boley, Victor Andrade, et al. New insights into the genetic etiology of alzheimer's disease and related dementias. *Nature Genetics*, 54(4):412–436, 2022. doi:10.1038/s41588-022-01024-z.

[483] Meric A. Altinoz, Sinan Guloksuz, Rainald Schmidt-Kastner, Gunter Kenis, Bahri Ince, Bart P.F. Rutten. Involvement of hemoglobins in the pathophysiology of alzheimer's disease. *Experimental Gerontology*, 126:110680, 2019. doi:10.1016/j.exger.2019.110680.

[484] Willem Kamphuis, Lieneke Kooijman, Marie Orre, Oscar Stassen, Milos Pekny, Elly M. Hol. Gfap and vimentin deficiency alters gene expression in astrocytes and microglia in

wild-type mice and changes the transcriptional response of reactive glia in mouse model for alzheimer's disease. *Glia*, 63(6):1036–1056, 2015. doi:10.1002/glia.22800.

[485] Miriam Ries, Rodrigo Loiola, Urvi N. Shah, Steve M. Gentleman, Egle Solito, Magdalena Sastre. The anti-inflammatory annexin a1 induces the clearance and degradation of the amyloid-β peptide. *Journal of Neuroinflammation*, 13(1), 2016. doi:10.1186/s12974-016-0692-6.

[486] Paloma Garcia-Casas, Michela Rossini, Riccardo Filadi, Paola Pizzo. Mitochondrial ca2+ signaling and alzheimer's disease: Too much or too little? *Cell Calcium*, 113:102757, 2023. doi:10.1016/j.ceca.2023.102757.

[487] Masayoshi Suzuki, Kenta Tezuka, Takumi Handa, Risa Sato, Hina Takeuchi, Masaki Takao, Mitsutoshi Tano, Yasuo Uchida. Upregulation of ribosome complexes at the blood-brain barrier in alzheimer's disease patients. *Journal of Cerebral Blood Flow & Metabolism*, 42(11):2134–2150, 2022. doi:10.1177/0271678x221111602.

[488] Youngdae Gwon, Seo-Hyun Kim, Hyun Tae Kim, Tae-In Kam, Jisu Park, Bitna Lim, Hyunju Cha, Ho-Jin Chang, Yong Rae Hong, Yong-Keun Jung. Amelioration of amyloid β-fcγriib neurotoxicity and tau pathologies by targeting lyn. *The FASEB Journal*, 33(3):4300–4313, 2018. doi:10.1096/fj.201800926r.

[489] Annerieke Sierksma, Ashley Lu, Renzo Mancuso, Nicola Fattorelli, Nicola Thrupp, Evgenia Salta, Jesus Zoco, David Blum, Luc Buée, Bart De Strooper, et al. Novel alzheimer risk genes determine the microglia response to amyloid-β but not to tau pathology. *EMBO Molecular Medicine*, 12(3), 2020. doi:10.15252/emmm.201910606.

[490] Lynn Pulliam, Bing Sun, Erin McCafferty, Steven A. Soper, Malgorzata A. Witek, Mengjia Hu, Judith M. Ford, Sarah Song, Dimitrios Kapogiannis, Marshall J. Glesby, et al. Microfluidic isolation of neuronal-enriched extracellular vesicles shows distinct and common neurological proteins in long covid, hiv infection and alzheimer's disease. *International Journal of Molecular Sciences*, 25(7):3830, 2024. doi:10.3390/ijms25073830.

[491] Yi Lu, Carolina Saibro-Girardi, Nicholas Francis Fitz, Mikayla Ranae McGuire, Mary Ann Ostach, A.N.M. Mamun-Or-Rashid, Iliya Lefterov, Radosveta Koldamova. Multi-transcriptomics reveals brain cellular responses to peripheral infection in alzheimer's disease model mice. *Cell Reports*, 42(7):112785, 2023. doi:10.1016/j.celrep.2023.112785.

[492] Caiyun Qi, Fang Liu, Wenjun Zhang, Yali Han, Nan Zhang, Qiang Liu, Handong Li. Alzheimer's disease alters the transcriptomic profile of natural killer cells at single-cell resolution. *Frontiers in Immunology*, 13, 2022. doi:10.3389/fimmu.2022.1004885.

[493] Yujun Hou, Xixia Chu, Jae-Hyeon Park, Qing Zhu, Mansoor Hussain, Zhiquan Li, Helena Borland Madsen, Beimeng Yang, Yong Wei, Yue Wang, et al. Urolithin a improves alzheimer's disease cognition and restores mitophagy and lysosomal functions. *Alzheimer's & Dementia*, 20(6):4212–4233, 2024. doi:10.1002/alz.13847.

[494] Zhuohang Liu, Hang Li, Shuyi Pan. Discovery and validation of key biomarkers based on immune infiltrates in alzheimer's disease. *Frontiers in Genetics*, 12, 2021. doi:10.3389/fgene.2021.658323.

ACKNOWLEDGEMENTS

This acknowledgment is written in English and German to reflect the international and local support I have experienced throughout this journey.

This thesis marks the end of a long journey — almost 11 years at Saarland University and 4 years of doctoral research at the Department of Clinical Bioinformatics. To everyone who accompanied me along the way: thank you.

Diese Arbeit ist der Abschluss einer langen Reise — fast 11 Jahre an der Universität des Saarlandes und vier Jahre Promotion an der Klinischen Bioinformatik. Ich danke allen, die mich auf diesem Weg begleitet haben.

I would like to express my sincere gratitude to Prof. Andreas Keller, who has supported me since my Bachelor's thesis and gave me the opportunity to contribute to such inspiring projects. I also thank the reviewers of this thesis for their time and effort. During my journey, I have worked with many amazing researchers, especially in Homburg, Stanford, and Barcelona. Thank you for the inspiring collaborations and interesting projects.

Ein besonderer Dank gilt meinen Kolleginnen und Kollegen am Lehrstuhl für Klinische Bioinformatik. Sabine und Anna danke ich dafür, dass sie mir bei allen bürokratischen Problemen immer mit Rat und Tat zur Seite standen. Tobias Fehlmann danke ich für die Unterstützung zu Beginn meiner Arbeit, und Fabian Kern dafür, dass er die ganze Zeit ein offenes Ohr und hilfreiche Ratschläge hatte. Vielen Dank auch an Georges und Pascal für die vielen guten Gespräche beim Mittagessen und die großartige Zusammenarbeit. An meine (zum Teil ehemaligen) Studenten und HiWis — Tobias, Emma und Lara— es war mir eine Ehre mit euch zusammenzuarbeiten. Ich kann es kaum erwarten zu sehen was ihr in der Zukunft alles erreichen werdet.

Over time, many colleagues have become close friends. Annika, Matthias, and Alejandra: thank you for being such incredible companions on this journey. I also want to especially thank Shusruto (aka Rick) for countless hours

of discussion and unwavering support. Thank you for always being willing to "add your Senf" and for believing in me even when I doubted myself.

Ich danke besonders auch meinen Freund:innen und meiner Familie für ihre bedingungslose Unterstützung — und die gelegentliche Erinnerung daran, dass es auch ein Leben außerhalb der Universität gibt.

CURRICULUM VITAE

Aus datenschutzrechtlichen Gründen wird der Lebenslauf in der elektronischen Fassung der Dissertation nicht veröffentlicht.