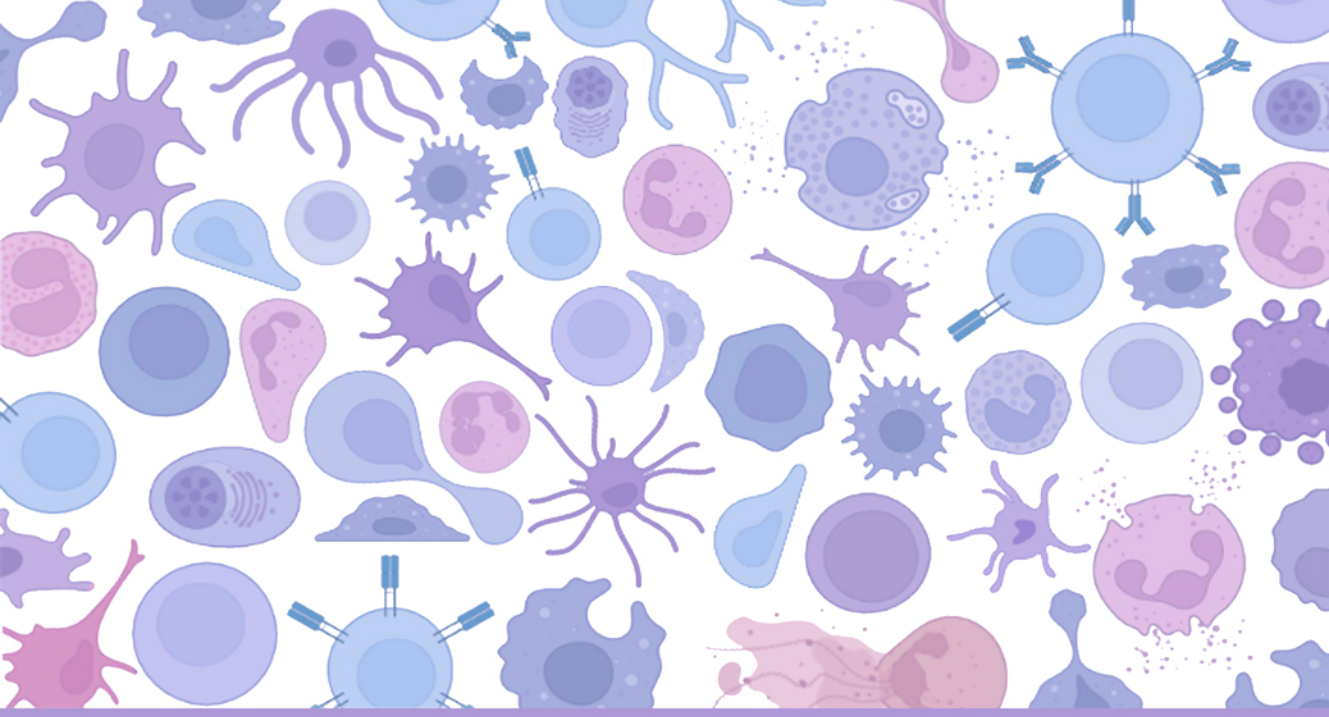


MULTI-OMIC DATA INTEGRATION STUDY OF IMMUNE SYSTEM ALTERATIONS IN THE DEVELOPMENT OF MINIMAL HEPATIC ENCEPHALOPATHY IN PATIENTS WITH LIVER CIRRHOSIS

Mª Teresa Rubio Martinez-Abarca

PhD Thesis

2022

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Supervisors

Dr. Ana Conesa Cegarra
Dr. Vicente Felipo Orts
Dr. Sonia Tarazona Campos

Multi-omic data integration study of immune system alterations in the development of minimal hepatic encephalopathy in patients with liver cirrhosis

María Teresa Rubio Martínez-Abarca

Supervisors:
Dr. Ana Conesa
Dr. Vicente Felipo
Dr. Sonia Tarazona

A doctoral thesis submitted to

Department of Biotechnology

April 2022

Graphic design: David Martínez Álvarez

Abstract

A high proportion of patients with liver cirrhosis develop minimal hepatic encephalopathy (MHE), a neuropsychiatric syndrome that produces attention deficits, cognitive impairment, and motor incoordination. MHE reduces both the quality of life and the life span of patients and results in high economic costs for healthcare systems given that it has now become a major socioeconomic problem. Although the etiology of MHE is not completely understood, it has been demonstrated that the appearance of mild cognitive and coordination impairments in patients with MHE is due to the synergistic action of both hyperammonemia and peripheral inflammation arising from liver failure.

The main objective of this work was to understand the immunological alterations associated with the peripheral inflammation that trigger MHE in patients with cirrhosis. These changes can be monitored through the signaling cascades of different immune system cell types. In this work, in a preliminary study, changes in gene expression (transcriptomics), plasma metabolites (metabolomics), and a panel of extracellular cytokines were analyzed in blood samples from patients with cirrhosis with and without MHE.

Transcriptomic analysis supported the hypothesis that alternations in the Th1/Th2 and Th17 lymphocyte cell populations are the major drivers of MHE. Cluster analysis of serum molecules highlighted 6 groups of chemically similar compounds, suggesting that functional modules operate during the induction of MHE. We also developed a multi-omic integration analysis pipeline to detect covariation between intra- and extracellular components that could contribute to the induction of cognitive impairment. Results of this integrative analysis suggested a relationship between cytokines CCL20, CX3CL1, CXCL13, IL-15, IL-22, and IL-6 and altered chemotaxis, as well as a link between long-chain unsaturated phospholipids and increased fatty acid transport and prostaglandin production.

A shift in peripheral inflammation in patients with MHE, mainly orchestrated by $CD4^+$ T cells, had been proposed in previous studies as a critical factor that triggers cognitive impairment. Thus, the second part of this thesis focused on understanding the pathways and mechanisms by which alterations in $CD4^+$ lymphocytes may contribute to peripheral inflammation in MHE. Thus, the expression levels of genes, transcription

factors, and miRNAs were analyzed in this lymphocyte subtype by high throughput sequencing (RNA-seq and miRNA-seq).

Separate analysis of each dataset showed mRNA and miRNA expression differences and altered biological pathways in $CD4^+$ lymphocytes when comparing patients with cirrhosis with and without MHE. We found alterations in 167 mRNAs and 20 pathways in patients with MHE, including toll-like receptors, IL-17 signaling, histidine, and tryptophan metabolism pathways. In addition, 13 miRNAs and 7 transcription factors presented alterations in patients with MHE. We used public databases to determine the target genes of these regulatory molecules and found key genes involved in the immunological shift triggering MHE. For instance, increased miR-494-39, miR-656-3p, and miR-130b-3p expression may modulate TNFAIP3 (A20) and ZFP36 (TTP) to increase levels of pro-inflammatory cytokines such as IL-17 and TNFα.

The last part of this thesis comprised a case study of the T-cell receptor (TCR) repertoire profiles of control patients and patients with cirrhosis with and without MHE obtained from the bulk RNA-seq dataset previously generated from isolated $CD4^+$ T cells. Given that RNA-seq experiments contain the TCR genes in a fraction of the data, an opportunity for receptor repertoire analysis without the need to generate additional data was created, thereby reducing the sample number required and the associated economic cost.

After read alignment to the VDJ genes performed with the MiXCR tool, we successfully recovered 498–1,114 distinct TCR beta chains per patient. In addition, profiling of the $CD4^+$ TCR repertoire could be used to help us understand the immune status of patients with MHE. Results showed fewer public clones (clonal convergence), higher diversity (clonal expansion), and elevated sequence architecture similarity within repertoires, independently of the immune status of the 3 groups of patients. Additionally, we detected significant overrepresentation of celiac disease and inflammatory bowel disease related TCRs in MHE patient repertoires. To the best of our knowledge, this is one of the few studies to have shown a step-by-step pipeline for the analysis of immune repertoires using whole transcriptome RNA-seq reads as source data.

In conclusion, our work identified potentially relevant molecular mechanisms of the changes in the immune system associated with the onset of MHE in patients with cirrhosis. Nonetheless, future work with a large sample cohort will be required to validate these results in terms of biomarker determination and the development of new, more effective treatments for MHE.

Resumen

Una alta proporción de pacientes con cirrosis hepática desarrollan encefalopatía hepática mínima (EHM), un síndrome neuropsiquiátrico que produce déficits de atención, deterioro cognitivo e incoordinación motora. La EHM reduce la calidad y la duración de la vida de los pacientes y produce elevados costos económicos para los sistemas sanitarios, siendo un importante problema socioeconómico actual. Aunque la etiología de la EHM no se conoce por completo, se ha demostrado que la aparición de alteraciones leves en la cognición y coordinación de los pacientes con EHM se debe a la acción sinérgica tanto de la hiperamonemia como de la inflamación periférica, derivadas de la insuficiencia hepática.

El objetivo principal de este trabajo fue conocer las alteraciones inmunológicas asociadas a la inflamación periférica que desencadenan deterioro cognitivo en los pacientes cirróticos. Estos cambios pueden ser monitorizados como cascadas de señalización a lo largo de los tipos celulares del sistema inmune. Como estudio preliminar, se analizaron los cambios en la expresión génica (transcriptómica), los metabolitos de plasma (metabolómica) y un panel de citoquinas extracelulares en muestras de sangre de pacientes cirróticos con y sin EHM. Los resultados del análisis transcriptómico apoyaron la hipótesis de alternancias en las poblaciones celulares de linfocitos Th1/Th2 y Th17 como principales impulsores de la EHM. El análisis clúster de las moléculas del suero dio como resultado 6 grupos de compuestos químicamente similares, sugiriendo módulos funcionales que operan durante la inducción de la EHM. También se ha realizado un análisis de integración multiómica para detectar las relaciones entre los componentes intra y extracelulares que podrían contribuir a la inducción del deterioro cognitivo. Los resultados de este análisis integrativo sugirieron una relación entre las citocinas CCL20, CX3CL1, CXCL13, IL-15, IL-22 e IL-6 con la alteración de la quimiotaxis, así como un vínculo entre los fosfolípidos insaturados de cadena larga y el aumento del transporte de ácidos grasos y la producción de prostaglandinas.

Estudios previos sugieren que un cambio en la inflamación periférica, orquestado principalmente por las células T $CD4^+$, es un factor crítico que desencadena el deterioro cognitivo en EHM. Así, la segunda parte de la tesis se centró en la comprensión de las rutas genéticas y los mecanismos por los que las alteraciones en los linfocitos $CD4^+$ pueden contribuir a la inflamación periférica en EHM. Se analizaron los niveles de expresión de genes, factores de transcripción y miARNs en este subtipo de linfocitos mediante secuenciación de alto rendimiento (RNA-seq y

miRNA-seq). El análisis individual de cada grupo de datos mostró las diferencias de expresión de ARNm y miARN, así como las vías biológicas alteradas en los linfocitos $CD4^+$ comparando pacientes cirróticos con y sin EHM. Encontramos alteraciones en 167 ARNm y 20 rutas biológicas en los pacientes con EHM, incluyendo los receptores tipo Toll, la señalización de la IL-17 y las vías del metabolismo de la histidina y el triptófano. Además, 13 miRNAs y 7 factores de transcripción presentaron alteraciones en los pacientes con EHM. Después utilizamos bases de datos para determinar sus genes diana, los cuales resultaron ser codificantes para proteínas clave implicadas en el cambio inmunológico que desencadena la EHM. Por ejemplo, la modulación por el aumento de miR-494-39, miR-656-3p y miR-130b-3p de la expresión de TNFAIP3 (proteína A20) y ZFP36 (proteína TTP) podría estar aumentando los niveles de citoquinas proinflamatorias como IL-17 y TNFα.

La última parte de la tesis comprende un supuesto práctico en el que se estudia el repertorio de receptores de células T (TCR) de pacientes control, y de pacientes cirróticos con y sin EHM, a partir del conjunto de datos de RNA-seq procedentes de células T $CD4^+$ aisladas previamente. Dado que los experimentos de RNA-seq contienen genes del TCR en una fracción de los datos, se crea una oportunidad para el análisis del repertorio sin necesidad de generar datos adicionales, lo que reduce la cantidad y los costes de las muestras. Tras el alineamiento de las lecturas contra la base de datos de los genes VDJ realizada con la herramienta MiXCR, recuperamos con éxito entre 498-1114 cadenas TCR beta distintas por paciente en nuestros datos. Además, estudiar los perfiles del repertorio TCR puede ayudar a comprender el estado inmunológico de los pacientes con EHM. Los resultados mostraron un bajo número de clones públicos (convergencia clonal), una alta diversidad (expansión clonal) y una elevada similitud en la arquitectura de la secuencia dentro de los repertorios, independientemente del estado inmunitario de los 3 grupos de pacientes. Además, detectamos una sobrerrepresentación significativa de los TCRs relacionados con la enfermedad celíaca y la enfermedad inflamatoria intestinal en los repertorios de los pacientes con EHM. Cabe destacar que este es uno de los pocos estudios que muestran el análisis paso a paso de los repertorios TCR utilizando lecturas del transcriptoma completo como datos de partida.

En conclusión, nuestro trabajo identifica potenciales mecanismos moleculares de los cambios en el sistema inmune asociados a la aparición de EHM en pacientes cirróticos. Sin embargo, se requieren trabajos futuros para validar estos resultados utilizando una mayor cohorte de pacientes que permita la determinación de biomarcadores y el desarrollo de nuevos tratamientos más eficaces para los pacientes con EHM.

Resum

Una alta proporció de pacients amb cirrosi hepàtica desenvolupen encefalopatia hepàtica mínima (EHM), una síndrome neuropsiquiàtric que produeix dèficits d'atenció, deteriorament cognitiu i incoordinació motora. La EHM redueix la qualitat i la durada de la vida dels pacients i produeix elevats costos econòmics per als sistemes sanitaris, sent un important problema socioeconòmic actual. Encara que l'etiologia de la EHM no es coneix del tot, s'ha demostrat que l'aparició d'alteracions lleus en la cognició i coordinació dels pacients amb EHM es deu a l'acció sinèrgica tant de la hiperamonemia com de la inflamació perifèrica, derivades de la insuficiència hepàtica.

L'objectiu principal d'aquest treball va ser conèixer les alteracions immunològiques associades a la inflamació perifèrica que desencadenen deteriorament cognitiu en els pacients cirròtics. Aquests canvis poden ser monitoritzats com cascades de senyalització al llarg dels tipus cel·lulars del sistema immune. Com a estudi preliminar, es van analitzar els canvis en l'expressió gènica (transcriptòmica), els metabòlits de plasma (metabolòmica) i un conjunt de citocines extracel·lulars en mostres de sang de pacients cirròtics amb i sense EHM. Els resultats de l'anàlisi transcriptòmica van recolzar la hipòtesi d'alternances en les poblacions cel·lulars de limfòcits Th1/Th2 i Th17 com a principals impulsors de la EHM. L'anàlisi clúster de les molècules del sèrum va donar com a resultat 6 grups de compostos químicament similars, suggerint mòduls funcionals que operen durant la inducció de la EHM. També s'ha realitzat una anàlisi d'integració multiòmica per detectar les relacions entre els components intra i extracel·lulars que podrien contribuir a la inducció del deteriorament cognitiu. Els resultats d'aquesta anàlisi d'integració van suggerir una relació entre les citocines CCL20, CX3CL1, CXCL13, IL-15, IL-22 i IL-6 amb l'alteració de la quimiotaxis, així com un vincle entre els fosfolípids insaturats de cadena llarga i l'augment del transport d'àcids grassos i la producció de prostaglandines.

Estudis previs suggereixen que un canvi en la inflamació perifèrica, orquestrat principalment per les cèl·lules T $CD4^+$, és un factor crític que desencadena el deteriorament cognitiu en EHM. Així, la segona part de la tesi es va centrar en la comprensió de les rutes genètiques i els mecanismes pels quals les alteracions en els limfòcits $CD4^+$ poden contribuir a la inflamació perifèrica en EHM. Es van analitzar els nivells d'expressió de gens, factors de transcripció i miARNs en aquest subtipus

de limfòcits mitjançant seqüenciació d'alt rendiment (RNA-seq i miRNA-seq). L'anàlisi individual de cada grup de dades va mostrar les diferències d'expressió d'ARNm i miARN, així com les vies biològiques alterades en els limfòcits $CD4^+$ comparant pacients cirròtics amb i sense EHM. Trobàrem alteracions en 167 ARNm i 20 rutes biològiques en els pacients amb EHM, incloent els receptors tipus Toll, la senyalització de la IL-17 i les vies del metabolisme de la histidina i el triptòfan. A més, 13 miRNAs i 7 factors de transcripció van presentar alteracions en els pacients amb EHM. Després utilitzàrem bases de dades per determinar els seus gens diana, els quals van resultar ser codificants per proteïnes clau implicades en el canvi immunològic que desencadena la EHM. Per exemple, la modulació per l'augment de miR-494-39, miR-656-3p i miR-130b-3p de l'expressió de TNFAIP3 (proteïna A20) i ZFP36 (proteïna TTP) augmentaria els nivells de citocines proinflamatòries com IL-17 i TNFα.

L'última part de la tesi comprèn un cas pràctic en el qual s'estudia el repertori de receptors de cèl·lules T (TCR) de pacients control, i de pacients cirròtics amb i sense EHM, a partir del conjunt de dades de RNA-seq procedents de cèl·lules T $CD4^+$ aïllades prèviament. Atès que els experiments de RNA-seq contenen gens del TCR en una fracció de les dades, es crea una oportunitat per a l'anàlisi del repertori sense necessitat de generar dades addicionals, el qual redueix la quantitat i els costos de les mostres. Després de l'alineament de les lectures amb la base de dades dels gens VDJ realitzada per l'eina MiXCR, recuperàrem amb èxit entre 498-1114 cadenes TCR beta diferents per pacient en les nostres dades. A més, estudiar els perfils del repertori TCR pot ajudar a comprendre l'estat immunològic dels pacients amb EHM. Els resultats van mostrar un baix nombre de clons públics (convergència clonal), una alta diversitat (expansió clonal) i una elevada similitud en l'arquitectura de la seqüència dins dels repertoris, independentment de l'estat immunitari dels 3 grups de pacients. A més, detectàrem una sobrerepresentació significativa dels TCRs relacionats amb la malaltia celíaca i la malaltia inflamatòria intestinal en els repertoris dels pacients amb EHM. Cal destacar que aquest és un dels pocs estudis que mostren l'anàlisi pas a pas dels repertoris TCR utilitzant lectures del transcriptoma complet com a dades de partida.

En conclusió, el nostre treball identifica els mecanismes moleculars dels canvis en el sistema immune associats a l'aparició d'EHM en pacients cirròtics. No obstant això, es requereixen treballs futurs per validar aquests resultats utilitzant una major cohort de pacients que permeta la determinació de biomarcadors i el desenvolupament de nous tractaments més eficaços per als pacients amb EHM.

A mi abuelo Juan Pedro
y a mi hermana Helena.

Agradecimientos

Aquí se cierra una etapa de crecimiento y superación de la cual solo tengo palabras de agradecimiento para todas las personas que me han acompañado durante el camino. Gracias a mi familia, por apoyarme en los buenos momentos y aguantarme en los malos; ellos siempre se alegran por mis pequeños logros a pesar de no pertenecer al mundo científico, lo que les convierte en las personas más especiales para mí. Gracias a David, por su capacidad para alegrarme todos los días, eres el pilar fundamental que consiguió revivir la esperanza en esos momentos en los que quise tirar la toalla. Gracias a mis amigos, por soportar mis largas charlas, interesarse por mis avances e incluso ser buenos confidentes en muchas ocasiones. Gracias también a mis compañeros y amigos del Conesalab (Ángeles, Pedro, Manu, Salva, Fran, Carlos, Cristina, Lorena, Rocío, Tian, Leandro y Tatyana, a todos nuestros estudiantes, Héctor, Alberto, Víctor, Marina, Anastasiya, Sergio, Paco, Zulema, Antonio y Blanca), por tantos buenos momentos amenizando el laboratorio con chistes o teniendo intensos debates para arreglar este mundo. Guardo con especial cariño todos los cursos que hemos hecho juntos, los viajes a congresos, los almuerzos de los viernes y nuestras quedadas de juegos de mesa, de ruta hamburguesera o cualquier otro plan de los nuestros. Y gracias a mis compañeras del Laboratorio de Neurobiología (Annie, María, Paula, Yaiza, Paola, Andrea, Marta, Mar, Mari Carmen, Anna, Fernando, Franc, Juanjo, Carla, Raquel y Lucas), que me recibieron como a una más, aunque fuera una "friki" de la bioinformática. Por último, especial gratitud a mis directores de tesis, Vicente, Ana y Sonia, de los cuales me llevo tres visiones complementarias de la ciencia, el pensamiento crítico, la paciencia y la constancia del trabajo.

Muchas gracias.

Contents

Glossary

a: acyl
aa: diacyl
ae: acyl-alkyl
AIRR: adaptive immune receptor repertoire
alpha-AAA: alpha-aminoadipic acid
AMPA: α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid
APO: apolipoprotein
ARE: AU rich-element
ATAC-seq: assay for transposase-accessible chromatin using sequencing
BCC: Bayesian consensus clustering
BCR: B cell receptor
C: constant gene
C18:2: octadecenoylcarnitine
CAD: collisionally activated dissociation
CAMs: cell adhesion molecules
CCL: C-C motif chemokine ligand
CCR: C-C motif chemokine receptor
CD: cluster of differentiation
CDR1–3: complementary determining regions 1–3
CDR3β: CDR3 region of the T-cell receptor β chain
CE: collision energy
ChIP-seq: chromatin immunoprecipitation sequencing
CLRs: C-type lectin receptors
CMV: cytomegalovirus
CPM: counts per million
CQN: conditional quantile normalization
CXCL: C-X-C motif chemokine ligand
CXCR: C-X-C motif chemokine receptor
CXP: collision cell exit potential
D: diversity gene
DEA: differential expression analysis
DENV: dengue virus
DNase-seq: DNase I hypersensitive sites sequencing
DP: declustering potential
e: alkyl
EBV: Epstein-Barr virus
EDTA: ethylenediaminetetraacetic acid
EP: entrance potential

ESI: electrospray ionization
FACS: fluorescent activated cell sorter
FC: fold-change
FDR: false discovery rate
FEA: functional enrichment analysis
FIA: flow injection analysis
FSMKL: feature selection multiple kernel learning
FWER: family-wise error rate
GC/MS: gas chromatography-mass spectrometry
GC: guanine and cytosine
GO: gene ontology
GS1–2: ion source gas 1–2
GSEA: gene set enrichment analysis
HCV: hepatitis C virus
HE: hepatic encephalopathy
HGNC: HUGO gene nomenclature committee
HIV: human immunodeficiency virus
HLA: human leukocyte antigen
HNE: hydroxynonenal
HPV: human papilloma virus
HSV2: herpes simplex virus 2
HTLV-1: human T-cell lymphotropic virus-1
HUGO: human genome organization
HuR: human antigen R
IBD: inflammatory bowel disease
IFNG: interferon gamma
IG: immunoglobulin
IL: interleukin
IMGT: immunogenetics information system
iNMF: integrative nonnegative matrix factorization
IQR: interquartile range
J: joining gene
JAK: Janus kinase-signal transducer
JIVE: joint and individual variation explained
KEGG: Kyoto encyclopedia of genes and genomes
L: ligand
LC/MS: liquid chromatography-mass spectrometry
LC: liquid chromatography
lncRNA: long non-coding RNA
logFC: $\log_2$-fold-change
LOR: log odds ratio
LPP: locality preserving projections
LPS: lipopolysaccharide

LRAcluster: low rank approximation clustering
lysoPC: lysophosphatidylcholine
m/z: mass divided by charge
M2: alternatively activated macrophages and monocytes
MAPK: mitogen-activated protein kinase
MAS5: microarray analysis suite 5
MCIA: multiple co-inertia analysis
MCPyV: Merkel cell polyomavirus
MD: mean-difference
MDA: malondialdehyde
MDI: multiple dataset integration
Met: methionine
Methyl-seq: methylation sequencing
MFA: multiple factor analysis
MHC: major histocompatibility complex
MHE: minimal hepatic encephalopathy
MIP: macrophage inflammatory protein
miRNAs: microRNAs
miRNA-seq: miRNA sequencing
MOFA: multi-omics factor analysis
MRM: multiple reaction monitoring
mRNA: messenger RNA
MS/MS: tandem mass spectrometry
MS: mass spectrometry
NA: non-available data
NASH: non-alcoholic steatohepatitis
NC: negative control
NEMO: neighborhood-based multi-omics clustering
NetICS: network-based integration of multi-omic data
NF-κB: nuclear factor kappa-B
NGS: next-generation sequencing
NLRs: nucleotide oligomerization domain (NOD)-like receptors
NMDA: N-methyl-D-aspartate
NMR: nuclear magnetic resonance
OH: hydroxyl group
ORA: over-representation analysis
PAM: partitioning around medoids
PARADIGM: Pathway Recognition Algorithm using Data Integration on Genomic Models
PBMCs: peripheral blood mononuclear cells
PC: positive control
PC1–2: principal component 1–2
PCA: principal component analysis

PCR: polymerase chain reaction
PCs: phosphatidylcholines
PFA: pattern fusion analysis
PHES: psychometric hepatic encephalopathy score
PINSPlus: perturbation clustering for data integration and disease subtyping
PLA2: phospholipase
PLS: partial least squares
PMA: penalized multivariate analysis
PRESS: predictive residual error sum of squares
PRRs: pattern-recognition receptors
PSDF: patient-specific data fusion
Q2: goodness of prediction
R: receptor gene
R2: goodness of fit
RACE: rapid amplification of 5'-complementary DNA ends
RISC: RNA induced silencing complex
RLRs: retinoic acid-inducible gene (RIG)-I-like receptors
RMA: robust multichip average
rMKL: regularized multiple kernel learning
RNA-seq: RNA sequencing
ROS: reactive oxygen species
RPKM: reads per kilobase per million
RSS: residual sum of squares
SARS-CoV-2: severe acute respiratory syndrome coronavirus
S-E: Shannon evenness
SHM: somatic hypermutation
SM: sphingomyelin
sMBPLS: sparse multi-block partial least squares
SNF: similarity network fusion
SR: species richness
STAT: signal transducer activator of transcription
T1D: type 1 diabetes
Tc: T cytotoxic
TCR: T-cell receptor
Tfh: T follicular helper cell
TFs: transcription factors
TGFβ: transforming growth factor beta
Th: T helper cell
TLRs: toll-like receptors
TMM: trimmed mean of M values
TNF: tumor necrosis factor alpha
TRA: T-cell receptor α chain

TRB: T-cell receptor β chain
TRC: T-cell receptor γ chain
TRD: T-cell receptor δ chain
Treg: T regulatory cells
TSH: thyroid-stimulating hormone
TSS: total sum of squares
T-SVD: thresholding singular value decomposition
TTP: tristetraprolin or ZFP36
UMI: unique molecular identifier
UTR: untranslated region
V: variable gene
Val: valine
VSN: variance stabilizing normalization
YFV: yellow fever virus

Nomenclature

Gene names follow the Human Genome Organization (HUGO) gene nomenclature (symbols contain only uppercase Latin letters and Arabic numerals); they were written in italic font to distinguish them from cases where the encoded protein had the same name (e.g., in the case of interleukin 17, the *IL-17* gene and IL-17 protein). Additionally, the mentioned software and R/Bioconductor packages were written in bold font.

Chapter 1

Introduction

1.1. Minimal hepatic encephalopathy

Minimal hepatic encephalopathy (MHE) is an alteration in cerebral function resulting from previous liver failure. Approximately 1.5 billion people worldwide present chronic liver disease[1], mainly cirrhosis. Around 5.5 million people in the USA and a similar number in Europe are affected by the disease, with most of these patients developing cognitive impairment[2]. According to the latest updated guidelines, hepatic encephalopathy (HE) is classified as type A, which is caused by acute liver disease; type B, secondary to portosystemic bypass or shunting, without liver failure; or type C, resulting from chronic liver disease, especially cirrhosis[3].

The latter group can be subdivided into covert HE (also known as MHE) and overt HE. Patients with cirrhosis with MHE or overt HE suffer progressive deterioration of their neurological function which reduces both their quality of life and overall life span. Characteristic symptoms associated with MHE are attention deficits, mild cognitive impairment, psychomotor slowing, and impaired visuo-motor coordination that can be unveiled and evaluated with psychometric tests[4,5]. Therefore, the ability of these patients to perform daily life tasks is reduced and their risk of falls, fractures, hospitalizations, and traffic accidents increases[6,7], making this an important health, social, and economic problem.

MHE can progress to a more advanced stage which involves severe cognitive impairment and, in the worst-case scenario, coma and death. Even though MHE is a potentially reversible alteration, most patients currently remain undiagnosed and untreated due the lack of simple diagnostic procedures. The gold standard test used to reveal these less-

evident symptoms and diagnose MHE was established in the consensus reached in 2002 as a battery of 5 psychometric tests denominated the Psychometric Hepatic Encephalopathy Score (PHES)[8]. However, recent findings have shown problems with the sensitivity of the PHES, and have suggested the use of other more sensitive tests such as the oral Symbol Digit Modalities Test, d2 Attention Test, and Bimanual and Visuo-motor Coordination Tests, and moreover, have proven that neurological alterations emerge differently among patients[9].

1.1.1. Factors contributing to minimal hepatic encephalopathy

Although the etiology of MHE is not completely understood, it is believed that multiple underlying mechanisms induce the functional impairment of the central nervous system observed in this syndorme[10,11]. The appearance of mild cognitive and coordination impairments in patients with MHE is due to the synergistic action of both hyperammonemia and peripheral inflammation arising from liver failure. The proposed general process starts with (1) chronic hyperammonemia and peripheral inflammation which are transmitted to the brain where they induce neuroinflammation; (2) thereafter, the neuroinflammation disrupts neurotransmission and neuronal connectivity which leads to (3), impairment of cognitive and motor function[12] (Figure 1.1).

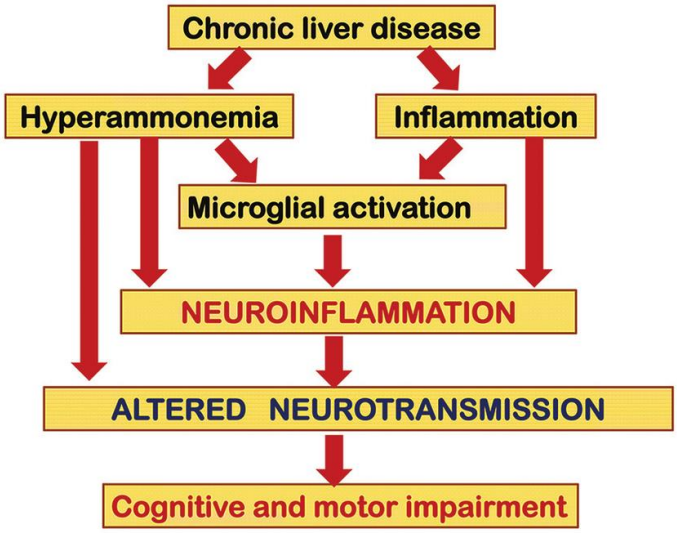


Figure 1.1 Factors contributing to cognitive impairment in minimal hepatic encephalopathy (Figure from Montoliu et al.[13]). Synergistic action of both hyperammonemia and peripheral inflammation, arising from liver failure, produces neurological alterations that triggers cognitive and motor impairment.

Immunological changes in the blood of patients with cirrhosis and MHE have also been uncovered[14], together with the appearance of neuroinflammation (microglial and astrocytic activation and Purkinje cell loss) in post-mortem cerebellum samples[15]. Studies using MHE animal models have demonstrated that peripheral inflammation induces neuroinflammation which in turn, alters neurotransmission, leading to neurological alterations. For instance, neuroinflammation and NMDA/AMPA receptor membrane expression alterations that impair spatial learning and memory have been detected in the hippocampus of rats with MHE. Additionally, the cerebellar neuroinflammation detected in this model was related with alterations in extracellular GABA and its transporters, thereby impairing motor coordination and learning. Treatments that reduce peripheral inflammation, neuroinflammation, and

GABAergic tone, or that increase extracellular cGMP can reverse these alterations and improve mild cognitive and coordination impairments[12].

1.1.1.1. Hyperammonemia

Regarding the induction of MHE as the first stage in the spectrum of HE, one of the main questions remaining is which factors associated with liver failure induce the neuropsychological alterations. The principal factors contributing to cognitive impairment in MHE are hyperammonemia together with systemic inflammation. The liver normally efficiently removes gut-derived ammonia produced by the digestion of dietary protein in order to maintain arterial ammonia levels at low concentrations. Both acute and chronic liver failure results in a reduced hepatic capacity for ammonia removal, producing elevated concentrations of ammonia (hyperammonemia) which are toxic both to neurons and astrocytes in the brain.

The only organ capable of processing a complete urea cycle and ammonia removal is the liver, although other tissues, including brain and skeletal muscle, express some of the constituent enzymes. Chronic liver failure (mainly cirrhosis) results in spontaneous portal-systemic shunting of portal blood, a moderate increase in arterial ammonia concentrations, and excessive ammonia concentrations in the brain. Chronic liver failure also results in the accumulation of additional toxins, including manganese, mercaptans, and short-chain fatty acids that may have deleterious effects on brain functions[16].

Hyperammonemia is one of the principal factors in the induction of mild cognitive impairment, as shown by the development of MHE in ammonia-fed model animals[17], and even in situations without liver cirrhosis[18]. In

addition, hyperammonemia impairs performance in psychometric tests in patients with cirrhosis during the inflammatory state but not after its resolution, thereby indicating that hyperammonemia exacerbates the neurological alterations induced by inflammation[19].

1.1.1.2. Alterations during peripheral inflammation

This subsection focuses on alterations in systemic inflammation related to MHE. A more general explanation of the fundamental concepts of the human immune system and the inflammatory response can be found in section 1.2 of this chapter.

In addition to hyperammonemia, systemic inflammation also results from cirrhosis and diverse studies have shown that inflammatory markers are higher in patients with MHE[20]. For instance, Montoliu et al.[21] showed a strong correlation between the high serum levels of interleukin 6 (IL-6) and IL-18 and the presence and grade of MHE, suggesting that the concentrations of these 2 interleukins may be useful to reveal cognitive impairment. Furthermore, a synergistic relationship between hyperammonemia and inflammation was found in the induction of MHE. Felipo et al.[18] recruited patients with liver or dermatological diseases to study different grades of hyperammonemia and inflammation and concluded that the combination of moderate levels of both factors together, but not alone, induced mild cognitive impairment. Indeed, this ammonia–inflammation synergy was able to induce cognitive impairment even in the case of diseases without liver cirrhosis, like those affecting patients with non-alcoholic steatohepatitis (NASH) or keloids[18].

It has been proposed that qualitative changes in the innate and adaptive immune systems of these patients contribute to triggering the appearance

of MHE (Figure 1.2). Patients with MHE exhibit selective T and B lymphocyte activation, as indicated by increased CD69 marker expression[14]. The same study also showed a pro-inflammatory state with a higher number of $CD14^{+}CD16^{+}$ monocytes, autoreactive $CD4^{+}CD28^{-}$ T lymphocytes, and enhanced serum levels of pro-inflammatory cytokines (IL-6, IL-21, IL-17, IL-18, tumor necrosis factor [TNFα], IL-1β, IL-15, and IL-22) and chemokines (CCL20, CXCL13, and CX3CL1).

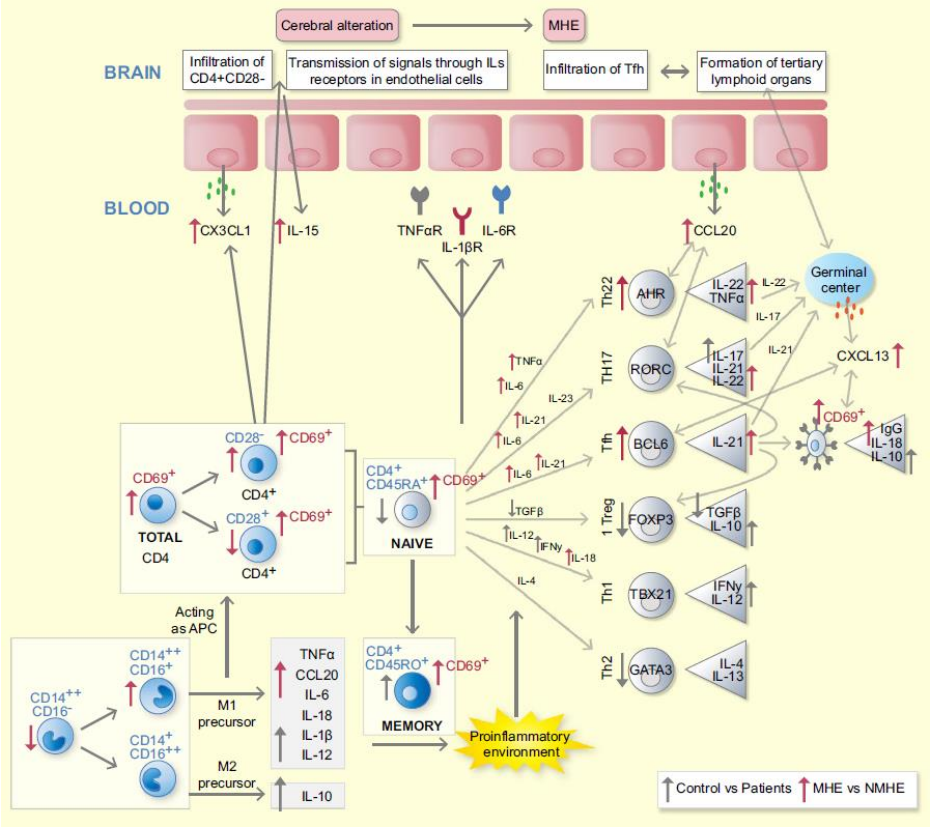


Figure 1.2 Changes in the innate and adaptive immune system of the patients with MHE (Figure from Cabrera-Pastor et al.[12]). Immunophenotype study of controls, and patients with cirrhosis with and without MHE, by measuring subtypes of monocytes and $CD4^{+}$ lymphocytes, a panel of cytokines of the cellular environment, and transcription factors that act as marker genes.

Furthermore, $CD4^+$ T lymphocytes have been shown to play a vital role in MHE. $CD4^+$ T lymphocytes can differentiate into various subtypes that are normally characterized by patterns of cytokine secretion and/or selective transcription factors acting as marker genes. Of note, Mangas-Losada et al.[14] isolated $CD4^+$ T cells from patients with and without MHE and reported higher levels of IL-17, IL-21, IL-22, and TNFα together with upregulated levels of the AHR and BCL6 transcription factors in MHE, which are respectively Th22 (T helper) and Tfh (T follicular helper) lymphocyte markers. Thus, these authors concluded that a shift in the immune system towards increased activation of Th17, Th22, and Tfh $CD4^+$ lymphocytes is associated with the appearance of MHE.

Despite this knowledge, few studies have addressed the precise pathways that trigger the systemic immune responses leading to neurological disorders, nor have they simultaneously analyzed all the contributing elements (chemokines, metabolites, regulatory components, etc.). In particular regard to MHE, as mentioned above, $CD4^+$ T lymphocytes are key players in the shift in peripheral inflammation, but the signals that trigger the appearance of neurological alterations remain poorly understood.

1.1.1.3. The interplay between neuroinflammation and systemic inflammation

Neuroinflammation is the immune response of the nervous system and is mainly caused by activation of microglia, the resident innate immune cells in the brain. Analysis of this phenomenon in MHE animal models has shown that neuroinflammation can alter neurotransmission, thereby disturbing neuronal communication and leading to cognitive and motor

impairment[12]. These alterations can occur in different brain areas including the hippocampus[17,22–24] and cerebellum[25–27].

Diverse studies in rat models have demonstrated that peripheral inflammation plays a leading role in the induction of neuroinflammation in hepatic encephalopathy, both in the hippocampus[28] and the cerebellum[29]. Three main possible mechanisms have been proposed as the origin of the cerebral function alterations caused by the immune system shift in patients with cirrhosis with MHE[14] (Figure 1.2), as follows:

- Infiltration of peripheral blood cells into the brain. The increased chemoattractants (IL-15, CX3CL1, and CXCL13) present in plasma samples from patients with MHE may promote the recruitment of autoreactive $CD4^+CD28^-$ T cells, as shown for other diseases such as multiple sclerosis[30]. Additionally, higher levels of CCL20 promotes T and B lymphocyte infiltration[31] and could contribute to neuroinflammation in patients with MHE.
- Signal transmission through endothelial cell receptors at the blood brain barrier. The upregulated circulating cytokines (TNFα, IL-1β, and IL-6) present in the blood of patients with MHE could activate their receptors on endothelial cells, triggering the release of inflammatory factors into the brain.
- The formation of tertiary lymphoid organs. This phenomenon (which also occurs, for example, in multiple sclerosis[32]) happens when effector cells infiltrate target tissues and organize into B cell follicles that also contain T-cell areas. Factors that may contribute to this formation in patients with MHE are the increased differentiation of $CD4^+$ lymphocytes into Th22 and Tfh cells, higher B lymphocyte activation (CD69), and elevated levels of IL-22 and CXCL13[33].

As discussed above, cognitive impairment may be present in patients with liver disease with certain ammonia and inflammation levels, even before the presentation of cirrhosis. Indeed, Balzano et al.[15] analyzed post-mortem cerebellum samples from patients with progressive stages of liver disease and found that neuroinflammation (i.e., activation of microglia and astrocytes) was present in patients before liver cirrhosis. The same study also showed the infiltration of $CD4^+$ T cells (mainly Th17, Tfh, and autoreactive $CD4^+CD28^-$ lymphocytes) during the early stages of liver failure.

These results reinforce the idea of the infiltration of lymphocytes and monocytes into the brain as a possible mechanism involved in the appearance of MHE in patients with cirrhosis, as proposed by Mangas-Losada et al.[14]. Furthermore, these data support the idea that infiltrated lymphocytes contribute to the induction of neuroinflammation and cognitive/motor impairment in patients with MHE. Moreover, this neuro-immune axis has been previously described in other neurological human diseases including multiple sclerosis, Alzheimer's disease, and psychiatric disorders in which pro-inflammatory biomarkers are thought to be responsible for leukocyte migration into damaged tissues[34,35].

1.1.1.4. Inflammation in cases of liver cirrhosis

As discussed in this section, regardless of their level of cognitive impairment, patients with cirrhosis present chronic inflammation of the liver. The general inflammatory process of the immune system is described at molecular level in section 1.2.3.

Liver hepatocytes and macrophages (Kupffer cells) recognize toxic agents entering the portal vein circulation. When endotoxin levels

increase, these liver cells release inflammatory cytokines and attractant chemokines to recruit neutrophils and other inflammation-mediating cells. In the context of liver damage, Kupffer cells act as source of inflammatory intermediaries such as pro-inflammatory cytokines (TNFα, IL-1β, IL-6, or IL-12), superoxide ions, nitric oxide, eicosanoids, chemokines, and proteolytic enzymes, which generate a cytotoxic environment[36]. Additional chemokines, including those in the macrophage inflammatory protein (MIP) family, attract peripheral system cells with phagocytic (neutrophils) or proinflammatory/cytotoxic (monocytes and lymphocytes) functions towards the inflamed tissue to help control the agents causing the liver damage[37].

Lymphopenia, an abnormally low level of lymphocytes in the blood, is one of the most characteristic attributes of patients with cirrhosis. In particular, Th cell lymphopenia in patients with cirrhosis has been the focus of many studies and is usually attributed to splenic sequestration[38]. Different studies, both in humans and animal models, have shown a reduction in naive Th cells in liver diseases of diverse etiologies[39].

For example, Lairo et al.[38] highlighted the different factors involved in immune deficiency in patients with cirrhotic liver by focusing exclusively on $CD4^+$ cells. Firstly, inefficient thymopoiesis results in reduced production of naive Th lymphocytes. Secondly, increased splenic sequestration leads to a decrease in the population of circulating naive Th lymphocytes. Finally, bacterial translocation causes strong memory lymphocyte activation and increased cell death by apoptosis, both in the naive and memory cell pools.

Furthermore, patients with cirrhosis present immune cell alterations at the systemic level. Some of the abnormalities in the main circulating

populations of immune cells include neutrophil–phagocyte dysfunction, decreased natural killer cell activity, increased proinflammatory cytokine expression, and intensified induction of TNFα from monocytes[38,40]. Of note, some of these cytokines have been proposed as therapeutic targets for the treatment of patients with liver disease[41] and the effect that cytokines have directly on the liver has also been studied. For instance, IL-17 promotes hepatic fibrogenesis by activating hepatic stellate cells which facilitates the development of liver cancer by recruiting myeloid suppressor cells[42]. In turn, IL-22 protects against the development of fibrosis and steatohepatitis[43], although it participates in the development of hepatocellular carcinoma[44]. Controversy still exists regarding IL-6 because it plays a critical role during the acute phase of liver damage[45] but IL-6 signaling plays a protective role during the progression of fibrosis[46].

Cirrhosis contributes to peripheral inflammation by compromising the homeostatic role of the liver in the systemic immune response. Local damage to the reticuloendothelial system reduces hepatic synthesis of proteins involved in the pattern recognition of phagocytic cells, impairing their systemic bactericidal capacity. Furthermore, necrotic liver cells release damage-associated molecular signals that persistently activate immune circulatory cells. As cirrhosis progresses, the pathogen-associated molecular patterns produced by impairment of intestinal permeability further activate the immune system and aggravate systemic inflammation[47].

1.1.1.5. Microbiota

Over the past decade, microbiota has been added as a new factor contributing to MHE and is now viewed as an important part of the impaired gut–liver–brain axis in cirrhosis. Besides ammonia, other toxic compounds derived from bacterial metabolism are not catabolized due to impaired liver function and are thereby transported to the brain where they exert neurotoxic effects[48]. Some microbial metabolites like endotoxins (lipopolysaccharides [LPS] from bacterial membranes) or bacterial DNA itself, are also able to activate immune cells and therefore may contribute to the inflammatory pathogenesis of MHE. In fact, Jain et al.[49] proved the correlation between serum levels of endotoxins and the grade of encephalopathy in patients with MHE.

With the aim of looking for new diagnostic targets, the work of Bajaj et al.[50] determined specific stool and salivary microbial signatures for individual cognitive testing strategies in patients with MHE. Recent studies have found that the fecal microbiota in patients with MHE contains higher levels of *Veillonellaceae* and *Streptococcus salivarius* linked to cognition and ammonia[51,52]. Although it has been shown that gut microbial composition and function in cirrhosis can impact cognition, the mechanistic pathways linking the gut, liver, and brain have not yet been investigated in depth.

Indeed, different potential mechanisms that may regulate microbiota–brain communication have been proposed, including activation of the afferent sensory neurons of the vagus nerve, neuro-immune and neuro-endocrine pathways, microbial metabolites such as short-chain fatty acids, microbial derived neurotransmitters, and tryptophan–kynurenine pathway modulation. Notably, the gut microbiota can regulate tryptophan availability for kynurenine pathway metabolism, which then exerts effects

in the periphery and in the central nervous system functionality[53]. In addition, short-chain fatty acids such as acetate, propionate, and butyrate are the main products of microbiota metabolism after fiber or starch anaerobic fermentation[54]. They mainly influence systemic inflammation by inducing T regulatory (Treg) cell differentiation and by regulating the secretion of interleukins[55].

1.1.2. Treatments that reduce peripheral inflammation

Unfortunately, only clinical HE has approved treatment, with mainstay therapy being aimed at reducing blood ammonia levels. As discussed in the previous section, inflammation plays a key role in the pathogenesis of HE and therefore, inflammation is an important therapeutic target for HE. Current HE therapies with clinical evidence for their efficacy involve non-absorbable disaccharides (lactulose and lactitol), that reduce intestinal ammonia production and absorption[56].

Other treatments like antibiotics and probiotics inhibit pathogenic bacterial activity in the intestinal tract, which ultimately reduces hyperammonemia and inflammation. Rifaximin is a systematically studied antibiotic for the treatment of HE which has minimal systemic absorption, a subtle adverse effect profile, and a low risk for bacterial resistance due its gut-selective nature. The efficacy and safety of this medication makes rifaximin an important therapeutic tool for the management of HE because it alleviates endotoxemia and improves cognition[57].

A few experimental drugs like infliximab, a commercial anti-TNFα drug with anti-inflammatory properties which prevents systemic inflammation, are available for the treatment of MHE. Infliximab has been shown to normalize the serum levels of the pro-inflammatory prostaglandin E2, as

well as IL-17, IL-6, and IL-10 in rat models with MHE[28]. However, a recent meta-analysis of data from 25 trials by Dhiman et al.[58] found that rifaximin and lactulose were the most effective agents for the reversion of MHE. Nonetheless, while future personalized treatments may consider the stage of HE alongside the disease source and patient clinical history, MHE treatment still remains a huge unmet need that requires attention.

1.2. The human immune system and inflammatory responses

Given that patients with MHE undergo a shift in peripheral inflammation in which $CD4^+$ T cells are key players, this section briefly introduces the inflammatory response process, immune system cell types, cell communication pathways, and T-cell receptor characteristics.

The immune system is the set of molecules, cells, tissues, and organs that protect organisms against extraneous agents. Immunity respects the 'self' (body cells, food, symbiotic microorganisms, etc.) while also eliminating the 'non-self' (pathogens and toxins, etc.). Danger signals are usually produced by pathogens, however, some self-components can also act as danger signals as in the case of burns, radiation, anoxia, or infection-associated tissue damage. In addition to providing protection against invasive organisms, the immune system also destroys modified self-cells/molecules that may compromise the health of the organism.

On the one hand, the first line of defense is 'innate immunity', which comprises barriers (skin or mucous membranes), small molecules (the complement system), and cells (macrophages and dendritic cells). The natural or nonspecific immune system is present from host birth and reacts immediately to remove pathogens within hours. On the other hand,

the immune system can generate an extremely specific response known as ‘adaptive immunity’. In this case, specific proteins (antibodies) and cells (lymphocytes) achieve an extensive range of antigen recognition. The adaptive immune response can take days to fully develop but it is highly specialized and saves immunological memory for future reinfections.

1.2.1. Immune system cells

The cells of the immune system are mainly produced in the bone marrow from pluripotent hematopoietic stem cells where they go through erythroid, myeloid, or lymphoid series differentiation until mature blood cells are produced. Erythroid progenitors produce erythrocytes and platelets; myeloid progenitors produce the innate immune cell types (neutrophils, basophils, eosinophils, mast cells, macrophages, and dendritic cells); and lymphoid progenitors generate the adaptive immune cell types (B cells and T cells) and natural killer cells[59].

Neutrophils, whose primary function is phagocytosis, are the first cells to arrive at infection sites. Macrophages are mobilized at a later stage and engage in phagocytosis as well as antigen presentation to other cells. Eosinophils release antimicrobial proteins and inflammation mediators, mainly to eliminate large microbial agents. Basophils and mast cells produce inflammatory mediators that participate in acute inflammation and are specialized in the liberation of histamine in allergic responses.

Neutrophils, basophils, and eosinophils all have an abundant content of cytoplasmic granules and are therefore commonly known as granulocytes. The early arrival of granulocytes during an infection induces acute inflammation and dilation of blood vessels to allow for the rapid influx of other immune cell types. Finally, dendritic cells can capture,

process, and present antigens to resident T cells in lymphoid organs by using specific receptors referred to as the major histocompatibility complex (MHC).

As part of the adaptive immune response, T cells recognize specific antigens using the T-cell receptor (TCR) and can differentiate into 2 groups. On the one hand are T cytotoxic (Tc) cells which express CD8 on the surface and interact with MHC type I (MHC-I) and on the other are Th cells that express CD4 and interact with MHC type II (MHC-II). After antigen recognition, $CD4^+$ T cells start to mature into other cell subtypes with new effector functions.

In contrast, B cells can recognize antigens by themselves due to the nature of the B cell receptor (BCR) which is composed of immunoglobulin molecules, also known as antibodies. Additionally, B cells also can also act as antigen presenting cells to T cells. Finally, natural killer cells are a special cell type derived from adaptive progenitors but which exert innate functions. They scan neighboring cells for signs of infection and kill eukaryotic cells by natural cytotoxicity or antibody-dependent cellular cytotoxicity.

1.2.1.1. $CD4^+$ T cells

The various functions of each immune system cell subtype are determined by different signaling routes. In the case of $CD4^+$ T lymphocytes, the cell differentiation pathways are determined by specific transcription factors driven by cytokines (detailed in the next section). However, additional processes such as cell–cell interaction pathways, signaling pathways, and metabolic pathways, etc. are at play in $CD4^+$ T cells.

Interestingly, metabolism and nutrient availability influence T-cell activation and function. For instance, activated T cells undergo metabolic reprogramming which promotes biomass induction (increased production of lipids, proteins, nucleic acids, and carbohydrates). This means that different T-cell stages have distinct metabolic profiles. For instance, memory T cells but not effector T cells undergo fatty acid oxidation[60].

The metabolism of amino acids like tryptophan can also modulate the immune system. Dendritic cells and macrophages catabolize tryptophan to produce proapoptotic intermediates known as kynurenines. Kynurenines can translocate the transcription factor AHR to the nucleus where it performs 2 functions: firstly, it blocks STAT signaling resulting in Th17 differentiation downregulation and secondly, it promotes IL-10, IL-22, and Treg cell expression[61]. In addition, the expression of functional histamine receptors has also been noted on $CD4^+$ cells from patients with inflammatory skin diseases and furthermore, stimulation with histamine increased the production of IL-17 in Th17 cells[62].

1.2.2. Communication between cells during immune responses

Communication between cells of the immune system is driven by cytokines: small, secreted proteins that coordinate the immune response. Different cell subtypes release cytokines into the extracellular space and they then travel to the target cell where they are recognized by membrane receptors. Paracrine cytokine secretion produces an effect in neighboring cells while endocrine action occurs in distant cells. Cytokines may even generate a local or autocrine response in the cells that produce them.

Functional classification of cytokines divides them into 3 main groups: innate immune cytokines (mainly secreted by macrophages), adaptive immune cytokines (produced by Th cells), and growth factors that regulate hematopoiesis. A single cytokine can perform various functions (pleiotropy) and many cytokines may produce the same effect (redundancy). Additionally, synergy is produced when 2 or more cytokines potentiate their effect, while antagony describes the case when the effect is neutralized. Additionally, cytokines may be either pro-inflammatory (such as IL-1β, IL-6, and TNFα), anti-inflammatory (including IL-4, IL-10, IL-11, and IL-13), or under specific circumstances, some can be either anti-inflammatory or pro-inflammatory (e.g., IFN-α, IL-6, or TGFβ)[63].

The $CD4^+$ T-cell differentiation process is defined by cytokine patterns which promote the expression of essential transcription factors that characterize each cell subtype. For instance, IFN-γ and IL-12 stimulate the transcription factor T-bet that promotes Th1 differentiation, while IL-4 activates GATA-3 in the case of Th2 cells, and TGFβ and IL-6 stimulate RORgT in Th17 production.

In turn, each Th subtype releases specific cytokine patterns to perform their various functions. Th1 normally secretes IFN-γ and TNFβ[64] to coordinate the immune response by activating macrophages, natural killer cells, $CD8^+$, and B cells. Th2 liberates IL-4, IL-5, IL-9, and IL-25 which regulate the action of eosinophils, basophils, and mastocytes. Th17 produces IL-17, IL-6, and IL-22 which together influence the production of inflammatory chemokines and growth factors in an aggressive response at the beginning of adaptive immunity. Polarization to Th1 or Th2 cells tends to reduce the exacerbated Th17 response, but in chronic inflammatory processes, especially autoimmunity, sustained Th17 activity remains.

Finally, chemokines (or chemotactic cytokines), a large family of structurally similar molecules that stimulate leukocyte movement and regulate their migration from blood to tissues, are a special type of cytokine. The nomenclature and classification of chemokines is based on their 2 internal disulfide bonds. The CC subfamily has 2 adjacent cysteines while in the CXC branch there is 1 amino acid between cysteines. Receptors and ligands are denominated with an additional R or L respectively (e.g., CCR2, CCL2, CXCR2, and CXCL2)[65].

1.2.3. Inflammatory response

Inflammation is one of the first defense mechanisms that the immune system initiates to minimize damage after injury or infection. An acute inflammatory response includes local alterations in vascular permeability, the accumulation of recruited leukocytes, and produces inflammatory factor secretion which is all restored by homeostasis after the lesion is resolved. However, when acute inflammation is not mitigated, it may become chronic, leading to disease. Examples of chronic inflammatory diseases include rheumatoid arthritis, systemic lupus erythematosus, Graves' disease, psoriasis, and multiple sclerosis[66].

Independently of the stimulus origin, a general mechanism exists which orchestrates the inflammatory process in all these diseases. Local immune and nonimmune cells with pattern-recognition receptors (PRRs) are primarily responsible for the recognition of exogenous or endogenous damage signals. The 4 main families of PRRs are toll-like receptors (TLRs), nucleotide oligomerization domain (NOD)-like receptors (NLRs), C-type lectin receptors (CLRs), and retinoic acid-inducible gene (RIG)-I-like receptors (RLRs).

Antigen–ligand binding activates a series of intracellular signaling pathways: nuclear factor kappa-B (NF-κB), mitogen-activated protein kinase (MAPK), Janus kinase (JAK)-signal transducer, and activator of transcription (STAT) pathways, which promote the production and release of inflammatory molecules. These molecules, such as pro-inflammatory cytokines (e.g., IL-1β, IL-6, and TNFα) or oxidative stress products including reactive oxygen species (ROS) and malondialdehyde (MDA), are used as biomarkers of inflammatory diseases. The last step in the inflammatory response is recruitment of inflammatory cells from the general circulation to sites of damage via chemokine gradients[67].

1.2.4. T-cell receptor

The 2 main phases of the immunological response are recognition and activation. Each cell of the innate immunity system recognizes various foreign substances through identical receptors. However, in adaptive immunity, every cell expresses a unique receptor that detects 1 single antigen. Thus, the potential antigen receptor repertoire in each individual is estimated to comprise around 2×10^{12} different B- and T-cell immune receptor sequences[68].

Adaptive immune cells acquire this huge receptor diversity during their development through a combinatorial rearrangement process at the DNA level. Briefly, multiple segments of the V (variable), D (diversity), and J (joining) genes are available at the germline level and a random V(D)J combination is spliced into the C (constant) regions in the final receptor transcript. An additional process of somatic hypermutation (SHM) occurs in B cells that increases the receptor diversity and antigen specificity.

TCR antigen recognition is dependent on the interactions with antigen-MHC molecules. Most T cells express a heterodimeric receptor formed by *α* and *β* chains while only 1–5% of T cells have *γ* and *δ* chains[69] (Figure 1.3). *α*/*γ* chains comprise a combination of 1 V and 1 J gene (VJ) while *βIδ* chains also contain 1 D gene between the V and J (VDJ). The TCR recognizes MHC molecules by the complementary determining regions 1–3 (CDR1–3) located at the distal region of the receptor.

CDR1 and CDR2 are encoded by V genes while CDR3 is encoded in the junction between the V-J or D-J genes. CDR3 is the receptor region which directly contacts the antigen and each T-cell 'clonotype' has a unique CDR3 sequence. Of note, the concept of clonotype is defined in different ways in the literature: it may indicate either a complete antigen receptor (e.g., *α/β* TCR), only 1 of the 2 chains (e.g., TRA or TRB), 1 domain (e.g., VJ or VDJ), or the CDR3 sequence of a domain[68].

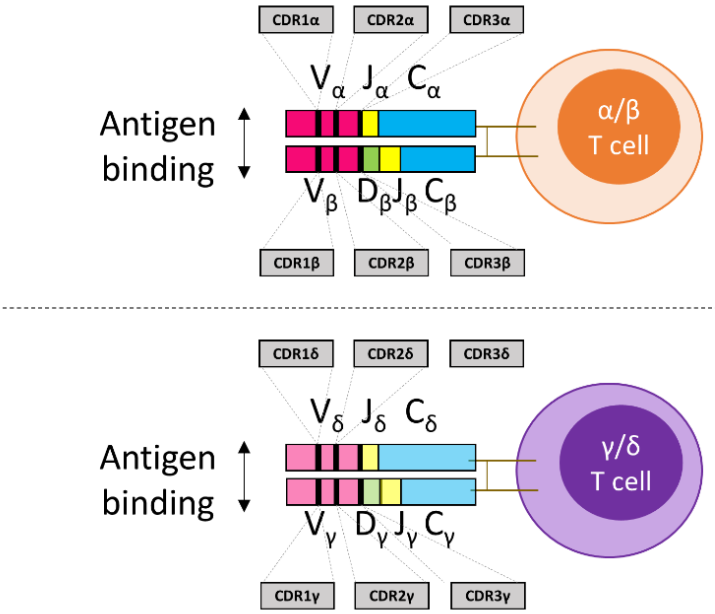


Figure 1.3 T-cell receptor gene structure. Scheme of the regions that comprise the T-cell receptor in both α/β and γ/δ T cells after variable (V), diversity (D), joining (J), and constant (C) gene rearrangement. The distal region of the receptor recognizes antigens through the complementary determining regions 1–3 (CDR1–3).

The international reference that has standardized immunogenetic concepts and nomenclature is the IMGT ('ImMunoGeneTics') information system[70]. IMGT follows a hierarchical classification for the BCR and TCR genes based on the concepts 'group', 'subgroup', 'gene', and 'allele'. The term group allows the classification of a set of genes that belong to the same multigene family; there are 14 groups for the TCR: TRAV, TRAJ, TRAC, TRBV, TRBD, TRBJ, TRBC, TRDV, TRDD, TRDJ, TRDC, TRGV, TRGJ, and TRGC.

The subgroup classification identifies genes from the same species that share at least 75% of their identity at the nucleotide level. Finally, the allele classification highlights variants with a few mutations (different nucleotides) compared to the reference sequence. Subgroups, genes, and alleles are always associated with a species name. For instance, *Homo sapiens* TRAV8–6*00 is the allele *00 of the TRAV8–6 gene that belongs to the TRAV8 subgroup and the TRAV group.

T-cell progenitors from the bone marrow travel to the thymus, where TCR rearrangement takes place and $CD4^+CD8^+$ double positive thymocytes are generated. These cells are then submitted to a stringent process of positive selection via the human leukocyte antigen (HLA) present in thymic epithelial cells and negative selection of self-reactive clones through interactions with thymic dendritic cells. After selection, thymocytes emerge to the periphery as $CD4^+$ or $CD8^+$ single positive naive T cells. When the TCR recognizes its specific antigen, T cells start a clonal expansion and differentiation into effector T cells. After infection

resolution, effector T cells are eliminated by apoptosis, although a fraction of them are retained to protect against future reinfections (memory T cells).

In humans, naive T cells can persist 5–10 years and are characterized by their high diversity and low clonality levels compared to memory cells. In contrast, memory cells represent a major circulating population in the blood that may be classified as 'central memory' with a high proliferative capacity, 'effector-memory' which produces effector cytokines, and 'stem-cell memory', a rare subset with no effector function. Importantly, T-cell dynamics change with age as a result of thymic output decline, with a general decrease in naive cells and an increase in memory cells across all the T-cell compartments[71].

1.3. Omics and systems biology

Considerable knowledge about the MHE has been gathered through clinical and biomedical studies of target molecules[14,21]. However, the effective design of diagnostic tools and treatments for MHE requires the application of comprehensive approaches to evaluate which underlying molecular mechanisms of peripheral inflammation trigger the appearance of neurological alterations. These types of studies are now possible thanks to the generalization of high-throughput multi-omic technologies for the study of human disease.

This section includes an overview of high-throughput multi-omic methods, as well as of the bioinformatic analysis of these multi-dimensional data. Here we discuss data processing, statistical analysis, data integration, and biological interpretation of diverse types of omic datasets. Lastly, the

final part of this section considers computational immunology, a branch of bioinformatics tailored to the study of the adaptive immune system and used here to characterize T-cell receptors in patients with MHE.

1.3.1. Overview of multi-omic data

The 20th century advent of high-throughput molecular techniques that measure thousands of biomolecules of the same type at the same time in a single experiment[72] led to the development of the '–omics' sciences. When attached to a word describing a particular type of molecule, the '–omics' suffix indicates 'all constituents considered collectively'. For instance, transcriptomics is the study of the full set of RNA transcripts for a particular sample; likewise, metabolomics considers metabolites and proteomics analyses proteins. Multi-omics is the combination of omic data with the aim of analyzing the interrelationships among biomolecules and has the potential to considerably contribute to our understanding of the complexity of biological systems.

These multivariate biotechnologies, which allow almost everything contained in living cells to be tracked, can be applied to the study of the changes that biological, cellular, and molecular systems undergo over time and/or under different perturbations. However, systems biologists face huge challenges in terms of the amount of data generated by these methods. Hence, the creation of comprehensive molecular models, methods for extracting useful knowledge and making predictions from such data, and approaches to best visualize and integrate highly dimensional datasets is still required[73].

On the one hand, high-throughput techniques can be classified according to the type of source material and analytical platform they use. For

instance, next-generation sequencing (NGS) platforms measure DNA or RNA and produce RNA-seq (RNA quantification), miRNA-seq (microRNA [miRNA] quantification), ChIP-seq (DNA binding sites for transcription factors and other proteins using chromatin immunoprecipitation followed by sequencing), DNase-seq (DNA regulatory region locations), ATAC-seq (assay for transposase-accessible chromatin using sequencing), or Methyl-seq (patterns of DNA methylation sequencing as epigenetic marks), among other datasets. On the other hand, mass spectrometry (MS) or nuclear magnetic resonance (NMR) platforms quantify proteins and metabolites to obtain datasets such as proteomics, metabolomics, lipidomics, glycomics, and glycoproteomics, etc.

In this subsection, we describe the study of transcriptomics using microarrays, RNA-seq, and miRNA-seq, and metabolomics using MS because these methods were used to carry out the work described in this thesis. Figure 1.4 summarizes the analysis steps required for these 4 data types, which are detailed in more detail in the following subsections.

Figure 1.4 Summary of 'single-omic' data generation, processing, and analysis. Each column shows all the steps required for the analysis of distinct types of omic datasets; shared steps are aligned on the same row.

1.3.1.1. High-throughput RNA quantification

The first RNA profiles were measured using microarray experiments in 1999[74,75]; however, they were rapidly replaced by NGS technologies (RNA-seq), which have been evolving since the finalization of the human genome project in 2004[76]. The main difference between microarrays and RNA-seq is that the former profiles a list of predetermined genes while the later captures the whole transcriptome.

Microarray technology (e.g., Affymetrix or Agilent, etc.) is based on fluorescent or radioactive hybridization. Briefly, the RNA extracted from a tissue sample is labeled with fluorophores or radioactive nucleotides and

is then hybridized to immobilized cDNA probes, with each one representing a different gene. The labeled RNA binds to its complementary sequence and emits radioactivity or a fluorescence signal which allows estimation of the amount of RNA present for each transcript type in the sample[77]. Microarray analysis includes image processing to extract measurements, data normalization to avoid technological variation, and statistical analysis to determine which genes are differentially expressed between groups of samples (Figure 1.4).

However, microarray technology is outdated and NGS technologies have been gaining ground over the last 2 decades thanks to considerable improvements in their chemistry, quality, yield, and costs. Short-read sequencing approaches are provided by different companies but the most extensive kits, and the ones applied in this current work to produce the RNA-seq dataset, are sold by Illumina. The Illumina sequencing platform can simultaneously collect information from many millions of hybridization reactions, thus sequencing many millions of DNA or RNA molecules in parallel.

RNA-seq datasets produced by NGS are useful for studying the RNA profiles contained in biological samples but if one is interested in quantifying non-coding small RNAs, the sequencing protocol must incorporate an additional step of enrichment by size after RNA extraction (Figure 1.4) in order to specifically target small RNAs. Typical miRNA library construction entails an initial miRNA gel purification step from total RNA to increase the efficiency of 3’ and 5’ ligation adapters, followed by cDNA library construction to integrate barcodes into polymerase chain reaction (PCR) primers, and final validation of the miRNA library purification[78].

1.3.1.2. Mass spectrometry platforms

MS has the potential to identify and quantify numerous molecules with diverse physico-chemical properties such as proteins and metabolites. Thanks to its potential for large-scale phenotyping, MS has become the most popular platform in metabolomics, and so it is increasingly being incorporated into human health studies. For targeted metabolite analysis, the compounds to be measured must already be known and stable isotope-labeled standards are used for analyte identification. In contrast, global or untargeted metabolomics registers all ions within a certain mass range, including ions belonging to structurally novel metabolites[79].

A typical metabolomic workflow includes sample collection and preparation, derivatization (soft chemical structure modification) if needed, instrumental measurement, raw data pre-processing, metabolite identification, statistical analysis, and interpretation[80]. Liquid-liquid or solid-phase sample extraction and dilution are also frequently used in metabolomics. Blood-derived metabolites are either studied in plasma (obtained by centrifugation) or serum (obtained by coagulation procedures), although plasma is the recommended biological material for metabolite identification[81].

MS platforms include a separation technique, an ion source, and a mass analyzer (Figure 1.4). Liquid/gas chromatography columns or capillary electrophoresis is coupled with mass spectrometers to separate molecules by their retention time. For example, gas chromatography-mass spectrometry (GC/MS) frequently couples electron ionization to analyze volatile and thermally stable compounds such as fatty acids, amino acids, and organic acids. However, liquid chromatography-mass

spectrometry (LC/MS) with atmospheric pressure ionization is the most used platform in untargeted metabolomics[82].

Additional chromatographic techniques, complementary ionization approaches and multiple analytical platforms have been proposed to increase the metabolome coverage using these techniques. Finally, a subset of samples or reference material can be spiked with reference compounds as a quality control to evaluate signal performance and additionally, to correct data using the normalization methods[83].

1.3.2. Analysis steps and integration of multi-omic data

1.3.2.1. Single-omic analysis

Here, we denote 'single-omic analysis' as the processing and subsequent statistical analysis of 1 independent omic dataset. The general steps in single-omic analysis are common to all the approaches: (1) pre-processing with the appropriate methods to correct possible technical/platform-dependent noise, (2) normalization to make the samples comparable, and (3) statistical analysis and biological interpretation (Figure 1.4). Moreover, different high-throughput techniques produce diverse types of data with specific properties which require normalization with dedicated algorithms. A brief description of the steps required to independently analyze microarrays, RNA-seq, miRNA-seq, and metabolomic datasets are provided in this subsection.

Microarray raw data are scanned images which can be processed with specific image analysis software (**ArrayVision**, **ImaGene**, **GenePix**, **QuantArray**, or **SPOT**) to quantify the amount of hybridization at the probe level. Generally, the higher the intensity observed for the probe, the

higher the expression of the gene with which it is associated. The main data processing steps include background correction, normalization, filtering of low intensity/expression genes, and replicate-spot averaging (Figure 1.4). These operations can be performed sequentially, or alternatively, methods are also available for performing them in combination. For instance, Microarray Analysis Suite 5 (MAS5) and Robust Multichip Average (RMA) are 2 common methods for Affymetrix microarray analysis. MAS5 and RMA are implemented in the **affy** R/Bioconductor package[84] and perform sequential pre-processing steps: MAS5 corrects each microarray independently while RMA works simultaneously with all the microarrays in the experiment.

In the case of Agilent microarrays, the feature extraction software can estimate foreground and background signals for each spot using the mean/median of the foreground and background pixels. Agilent Feature Extraction output files contain probe annotation columns and intensity columns that can be read using the **limma** R/Bioconductor package[85]. This package also allows subsequent background correction, control probe removal, and filtering of probes below a certain background level. After all these data pre-processing steps, we obtain a gene expression table (genes as rows and samples as columns) containing continuous data without technical noise.

In the case of RNA-seq and miRNA-seq, the raw data consist of 1 file per sample which contains read sequences of nucleotides in fasta or fastq (fasta + quality) format. The general scheme for the analysis with this type of data is (1) filtering of low-quality short reads and adapters; (2) the remaining sequences are mapped to a reference genome or transcriptome; and (3), the expression levels of each biological feature are estimated (Figure 1.4). The resulting gene expression quantification has

limited accuracy at low expression levels and may present technical biases such as expression level dependance on gene length or guanine and cytosine nucleotide gene content (GC content), PCR artifacts, uneven transcript read coverage, off-target transcript contamination, and differences in transcript distribution.

These potential limitations can be detected and corrected by combining different strategies. For instance, the R/Bioconductor package **NOISeq**[86] includes exploratory plots for the early detection of these errors as well as methods to appropriately reduce the noise at each step. One important aspect of RNA-seq data analysis is to confirm that the sequencing depth (total number of sequencing reads per sample) is adequate to accurately quantify gene expression. The expression level estimates of genes detected with a low number of reads are generally less reliable than those that accumulate a high number of counts. A widespread practice is to remove genes with total counts for all the samples below a certain cutoff threshold[87]. However, this runs the risk of excluding genes with relatively high expression in 1 of the conditions. An alternative method implemented in **NOISeq** calculates the average expression per condition measured in counts per million reads (CPM) and provides different strategies to filter out genes with low expression levels in all the conditions.

The use of CPM data instead of raw counts facilitates the application of comparable thresholds between samples with different total read numbers. Regarding the sequencing biases that can alter the expression levels of RNA-seq data, gene length bias results from longer transcripts rather than shorter transcripts accumulating more short read counts for the same number of molecules, which leads to a more accurate estimation of expression levels for longer genes[88]. This bias can be corrected by dividing the CPM value by the length of each transcript. Finally, in the case

of the Illumina RNA-seq platform, expression levels might be underestimated when gene GC content is very low or very high, which is known as 'GC contact bias'[89].

Normalization is usually the subsequent step in RNA-seq analysis (Figure 1.4). Two main groups of normalization procedures can be considered. While within-sample normalization methods are focused on the mitigation of gene biases (gene length or GC content), between-sample normalization reduces sample differences either because of the sequencing depth (the more reads for a given sample, the higher the gene expression estimation for that sample) or RNA composition (genes with extremely high expression in 1 sample could artificially increase the biological variability among samples, possibly leading to the detection of false expression changes between conditions).

A standard normalization method is RPKM (reads per kilobase per million), which divides the number of reads in each transcript by its length in kilobases and the sequencing depth of the sample in millions. Two commonly applied between-sample normalization methods are quantile normalization and TMM (trimmed mean of *M* values), which both match the gene count distribution across samples and assign them the same quartiles or median, respectively. Finally, different normalization methods have been developed to deal with potential sequencing bias. These are based on regression models of the count data on a gene feature (GC content, gene length, or both) and then subtract the fit from the counts to remove the dependence. These data pre-processing steps result in a gene expression matrix (genes as rows and samples as columns) that is free of technical noise and containing comparable samples.

For metabolomics, the first analysis step is the conversion of the raw data file into a suitable format (NetCDF, mzML, or mzXML) for pre-processing tools. The software most widely used by the metabolomic community to pre-process data from separation techniques is **XCMS**[90]. Data pre-processing typically involves (1) data filtration to remove the contamination found in all samples, (2) peak picking to extract features, i.e., m/z (mass divided by charge) values and the retention time, (3) peak alignment to correct possible retention time shifts in LC/MS from 1 sample with respect to another, and (4) addition of the NA (non-available data) when no signal is detected in some samples[82].

The annotation of metabolite species and ultimately, the characterization of their structures, is crucial for accurate biological interpretation. This is a very laborious and time-consuming procedure which includes database searches based on m/z values to find chemical formulas, different species (adducts, fragments, and isotopes), grouping within the same metabolite, annotation confirmation by retention time, and comparison with reference MS/MS datasets.

Downstream statistical analyses such as differential expression analysis (DEA) identify features with statistically significant different average expression levels between experimental conditions (e.g., diseased vs. healthy patients, treatment vs. control, mutant vs. wildtype, etc.). Although MS data platforms measure feature abundances rather than expression, for convenience, DEA notation is also applied to the differential analysis of metabolite levels. For microarray or metabolomic data, where the estimated expression/abundance level is a continuous variable, most of the DEA methods assume a normal data distribution. In contrast, RNA-seq expression levels have a discrete nature and DEA approaches use

discrete probability distributions such as the negative binomial to model read-count data.

The **limma** R/Bioconductor package[85] is based on linear regression models for DEA that are equally applicable to microarrays, RNA-seq, or mass spectrometry data. Unlike microarray or MS data, RNA-seq datasets require a previous data distribution transformation (from discrete to normally distributed data using $\log_2$ or voom transformation[91]) to fit the **limma** linear models. Other R/Bioconductor packages such as **edgeR**[92] or **DESeq2**[93] are suitable for the DEA of count data without the normalization steps described above because they apply generalized linear models based on the negative binomial distribution and possible biases or correction factors are included in the model.

All the **limma**, **edgeR**, and **DESeq2** packages use Empirical Bayes methods (an approach that borrows information across genes) to produce robust variance estimations in situations with a low number of samples[94]. Additionally, non-parametric methods like **NOISeq**[86] have also been proposed which have the advantage of not requiring such strong data distributional assumptions for unfulfilled cases. The output obtained from differential expression analysis is a table of genes/metabolites ordered by statistical *p*-value and fold-change (i.e., mean expression ratios).

Finally, the high number of statistical tests (1 per gene or metabolite) required in the analysis of omic datasets tends to create problems in the control of type-I errors, meaning that multiple testing corrections to adjust *p*-values and reduce the number of false positives must be applied. A procedure traditionally used in bioinformatics to address these problems is the restrictive Bonferroni adjustment[95] for family-wise error rate (FWER) reduction, i.e., the number of rejected true null hypothesis. In addition, the

more permissive Benjamini and Hochberg[96] adjustment is often used to control the false discovery rate (FDR), i.e., the proportion of true null hypotheses with respect to the number of rejected null hypotheses.

Once the DEA has identified the genes and metabolites that change their levels between experimental conditions, the next challenge is identification of the corresponding biological pathways represented by these molecular changes. Functional enrichment analysis (FEA) is the approach used to address this question. FEA identifies the functional label associated with the differentially expressed genes/metabolites. This is interpreted as the functionality of the functional label being altered in the experimental condition, thereby providing a functional readout of the high-throughput assay. Gene set enrichment analysis (GSEA) and over-representation analysis (ORA) have been widely used for FEA using different functional annotation databases such as Gene Ontology (GO)[97] or the Kyoto Encyclopedia of Genes and Genomes (KEGG)[98].

On the one hand, the goal of GSEA is to determine whether a predefined gene set (usually with an associated functional label) is proportionally more frequent (i.e., enriched) at the top or bottom of a list of genes ranked based on expression differences between 2 experimental classes. Enrichment *p*-values calculated in the original **GSEA** software[99] use Kolmogorov–Smirnov-like statistics, while other parameterizations are applied in alternative software such as **PAGE**[100] or **FatiScan**[101]; the R package **mdgsa** incorporates sophisticated multidimensional logistic models[102].

On the other hand, ORA determines the differential proportion of genes annotated to a given functional label between the groups of differentially and non-differentially expressed genes. The differential proportion can be

calculated using common and well-known statistical methods, including Fisher's exact test, the Chi-squared test, or the binomial proportions test[103]. **PaintOmics**[104] is a good example of a user-friendly web tool used to visualize and interpret FEA. **PaintOmics** maps genes and metabolites into biological pathways and represents their expression values within the pathway maps and can also combine different omic data types into the same FEA. FEA is valuable for detecting the variation in gene function that contributes to human diseases and which typically results from moderate variation in the activity of multiple members of a pathway.

1.3.2.2. Multi-omic analysis

The results of single-omic analysis can highlight the molecules under study that may be involved in different conditions, as well as their functions. However, multi-omic data integration provides an opportunity to study the relationship between several types of molecules and to propose complex biological models that span multiple molecular layers. The biological questions that integrated approaches address respond to 3 main goals:

- Molecular regulatory mechanisms: gaining knowledge about the signaling pathways affected in each disease is essential for improved diagnosis and the development of new disease interventions.
- Sample clustering: disease subtyping and classification based on multi-omic profiles. Suitable interventions are different for patients belonging to each subtype of a disease and consequently, sample classification is required to understand the disease etiology.

- Prediction: searching for biomarkers for various applications including disease diagnosis and prognostic predictions. Because biomarker validation is time-consuming, *in silico* prioritization of candidate biomolecules is an effective alternative.

While the different approaches for multi-omic data integration focus on 1 or several of the previous goals, classification of the multiple existing integrative tools is a non-trivial task. In this work we adopted the classification proposed by Subramanian et al.[105] which is based on the integrative approach general analysis methodology of similarity, correlation, network, Bayesian, multivariate, and fusion (Figure 1.5). Some of these approaches belong to more than 1 group because they use different approaches in combination to reach their final goal.

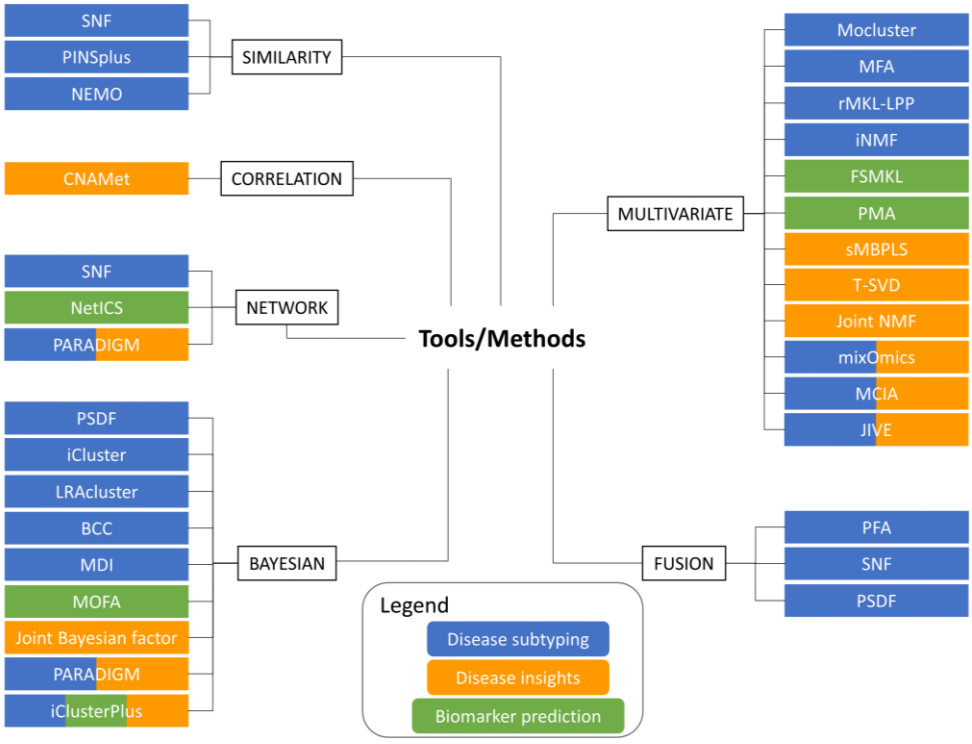


Figure 1.5 Classification of multi-omic integration approaches (figure modified from Subramanian et al.[105]). The classification is based on the tools/methods that integrative approaches are based on, as well as their analysis objective. Abbreviations: SNF: similarity network fusion, PINSPlus: perturbation clustering for data integration and disease subtyping, NEMO: neighborhood-based multi-omic clustering, NetICS: network-based integration of multi-omic data, PARADIGM: Pathway Recognition Algorithm using Data Integration on Genomic Models, PSDF: patient-specific data fusion, LRAcluster: low rank approximation clustering, BCC: Bayesian consensus clustering, MDI: multiple dataset integration, MOFA: multi-omics factor analysis, MFA: multiple factor analysis, rMKL: regularized multiple kernel learning, LPP: locality preserving projections, iNMF: integrative nonnegative matrix factorization, FSMKL: feature selection multiple kernel learning, PMA: penalized multivariate analysis, sMBPLS: sparse multi-block partial least squares, T-SVD: thresholding singular value decomposition, MCIA: multiple co-inertia analysis, JIVE: joint and individual variation explained, PFA: pattern fusion analysis.

Similarity methods generate a graph representing patients as nodes and their similarity as edges. One graph is created for each omic dataset and then a similarity matrix is calculated by merging edges across data types to identify subgroups of patients that are 'connected' through multiple omic types[106,107]. Correlation-based approaches reveal synergistic effects between omic features by linking their values (e.g., gene expression, copy number variation, and DNA methylation). Network-based methods consider already known (protein–protein interactions) or predicted relationships (correlation analysis) between biological variables[108]. Thus, metrics based on mathematical graph theory (degree, centrality, connectivity, etc.) can be calculated to identify features or patient groups that are central in the network and that therefore might be drivers or important components of the disease.

Bayesian approaches have been used for both disease subtyping (e.g. **iCluster**[109]), disease insight discovery (e.g. **PARADIGM**[110]), and/or biomarker prediction (e.g. **MOFA**[111]). Fusion-based approaches analyze the contribution by each data type to a common feature space, which allows the inference of relationships among them[112]. Multivariate methods

are dimension reduction techniques that can perform variable selection, identify patterns across multi-omic datasets, and capture co-relationships among them[105]. Here we will focus on discussing multivariate approaches in more detail because they were the main methods used in this work.

1.3.2.2.1. Multivariate approaches to omic data analysis

Integrative analysis of multi-omic data is a challenging task because a much greater number of biological variables are involved compared to the number of biological samples. There are also problems related to the noisy nature of the data which has missing values and is affected by batch effects and outliers, among other problems. Moreover, biological systems are complex and consist of non-linear relationships that are difficult to model with the data and mathematical methods currently available. Although simple approaches such as correlation analysis cannot provide a full explanation of the system behavior, they are often effective thanks to their simplicity and interpretability[73].

Multivariate methods based on dimension reduction and analysis of the latent space are frequently used for the analysis of multi-omic data thanks to their inherent ability to deal with many correlated variables while remaining robust even with a reasonable proportion of missing values. The number of components (or latent variables) and the sparsity level for variable selection must be optimized for these dimension-reduction approaches. Examples of multivariate methods frequently used in the multi-omic field are principal component analysis (PCA), partial least squares (PLS) in its multiple variations (i.e., sparse PLS, PLS discriminant analysis, multi-block PLS, etc.), and multi-way multidimensional data reduction models.

PCA decomposes a data matrix into 2 small matrices (a score matrix for observations and a loading matrix for variables) that capture the essential data patterns from the original table. Multiple applications have been described for PCA such as the identification of class membership (i.e., similar molecules or patient groups), detection of outliers, and variable selection for biomarker identification[113]. In contrast, PLS can relate 2 data matrices (e.g., transcriptomics and metabolomics) via a linear multivariate model, thereby overcoming the limitations of traditional regression to process data with many noisy, collinear, and even incomplete variables[114]. PLS can also simultaneously model several response variables to answer the biological question of which biomolecules might be related to each other (e.g., which transcription factors are related to which gene(s), or which metabolite changes are similar to gene expression changes, etc.).

The wide range of PLS versions can be used to address specific biological questions: sparse PLS[115] improves model interpretability because it allows selection of subsets of the most positive or negative correlated variables across the samples; PLS discriminant analysis[116] can classify and discriminate samples, and predict the class if new samples are added; and multi-block PLS[117] was developed to estimate co-variation between more than 2 data matrices, thereby broadening the spectrum of molecules that can be integrated.

Multi-way approaches can accommodate the complex structure of repeated measurements from different assays, where different treatments are applied to the same biological samples and at different time points[118]. Moreover, multi-way data can be arranged in a 3-way structure (a cube) instead of a matrix to accommodate the multiple dimensions of the datasets[119].

1.3.3. Computational immunology

High-throughput sequencing has been applied to the study of adaptive immune response in a wide range of diseases. Indeed, research in autoimmunity, lymphocyte biology, vaccine profiling, infections, allergies, and aging are some examples of its recent applications. However, these data require bioinformatics pipelines to analyze, interpret, and visualize immune repertoires. The bioinformatic and statistical analysis of high-throughput sequencing data from the immune repertoire typically follows a set of steps like the following: pre-processing of raw sequencing reads, V(D)J gene identification, clonal assignment, and repertoire analysis (i.e., diversity, architecture study, and immune repertoire specificity). Additional steps such as phylogenetic (lineage tree) construction and somatic hypermutation modelling are useful for studying B cells because of the characteristics of their cell receptors.

Depending on sequencing depth, read length, paired/single-end reads, and inclusion of unique molecular identifiers (UMIs), the pre-processing analysis is affected at various stages. Pre-processing steps include, when required, quality control and read annotation of fasta/fastq files, clustering by UMIs to reduce PCR amplification biases, and paired-end read assembly (which is expected to overlap in most experimental designs).

Identification of V(D)J regions involves reconstructing segments and boundaries that were rearranged to generate the BCR/TCR receptor. The most common alignment algorithms use known VDJ genes from a database as a reference. The main annotation platforms and software are **IMGT, IgBlast, iHMMune-align, MiGEC,** and **MiXCR** and an extensive comparison between them can be found in Greiff et al.[120]. Clustering of CDR3 sequences with a high homology at the nucleotide level and

identical V(D)J gene usage is a common approach to define clonotypes (i.e., lymphocytes with the same BCR or TCR). The set of unique cell clonotypes in an individual is referred to as its immune 'repertoire'.

Frequent steps in immune repertoire analysis include clone diversity, clone convergence, and sequence architecture analysis. Furthermore, ecology-derived diversity measures such as species richness and the Shannon and Simpson indices are traditionally used to reflect different aspects of diversity. Species richness is defined as the number of different clonotypes in a repertoire. The Shannon and Simpson indices include measurement of the diversity within a sample where the former is strongly influenced by species richness and rare species, while the later gives more weight to evenness and common species. In order to capture more information, use of the Hill family of diversity indices[121] is recommended to avoid contradicting qualitative outcomes when using single diversity indices[122]. Useful software for calculating diversity profiles includes **Vegan**[123], **tcR**[124], and **VDJtools**[125]. In addition, tools like **Change-O**[126] compare differences between diversity profiles in the presence of differently sized repertoires.

Identical immune receptor sequences shared by 2 or more individuals are denominated as 'public' clones and the convergence of the immune repertoires of 2 samples is quantified by their clonotype overlap. Overlap calculation measurements like the Morisita-Horn index[127] or Jaccard index[128] integrate the clonal frequency of compared clones to the measurement of clonal overlap.

High sequence similarity among immune receptor genes may occur when they recognize the same antigen. Network analysis allows interrogation of sequence similarity, defining each receptor sequence as a node and

adding edges when sequences satisfy a defined similarity condition or distance (e.g., if 2 sequences differ in only 3 amino acids, the Levenshtein distance = 3). **imNet**[129] is a pipeline that computes distance matrices between receptor sequences to make large-scale repertoire networks. This tool incorporates the PageRank measurement, a local network parameter (like 'degree', 'closeness', or 'betweenness') that measures the importance of the similarity between 2 clonotypes within the network.

Many tools are currently available for studying immune repertoires in computational immunology. However, there are still no consensus protocols or best practice guidelines in the field. Thus, to help reach the goal of defining a standard methodology for immune repertoire analysis, the Adaptive Immune Receptor Repertoire (AIRR) consortium was created and gathers expert scientists in the field. Additionally, very few attempts have been made to link immune receptor and transcriptomic data, which may provide a deep understanding of how BCR and TCR are regulated at the genetic level. Transcriptomic data has the advantage of capturing multiple aspects of immune repertoire complexity—from clonotypes to the expression of regulatory molecules—which could be of immense help when developing vaccines and immunodiagnostic techniques.

Chapter 2

Hypothesis, aims, and contributions

2.1. Hypothesis and objectives

Previous studies demonstrated that sustained peripheral inflammation induces neuroinflammation that alters neurotransmission, thereby producing cognitive and functional impairment both in patients and in animal models of MHE. However, the molecular mechanisms underlying these processes are not yet fully understood. The overall objective of this work was to understand the biology of the immunological changes associated with the appearance of MHE in patients with cirrhosis.

The hypotheses upon which this thesis is based are:

- The occurrence of MHE in patients with cirrhosis is associated with specific changes in peripheral inflammation and in their immunophenotype.
- These immune system changes result in a shift in peripheral inflammation that triggers the appearance of MHE in patients with cirrhosis; changes in $CD4^+$ T-lymphocyte differentiation and activation play a key role in this shift.
- Immunological changes are also associated with specific metabolic changes in immune cells and can be detected in the peripheral blood of patients with MHE.

2.2. Aims

To reach the general goal set out for this thesis, we defined the following specific aims:

- Characterize changes in gene expression, metabolites, and cytokines in blood samples from patients with cirrhosis and MHE versus patients without MHE. In particular:
    - Identify the biological pathways altered by the disease.

- Understand the joint contribution of gene expression and extracellular components associated with the onset of MHE.
- Identify potential predictors or biomarkers for the early diagnosis of MHE by using extracellular components (cytokines and metabolites) and the expression levels of different genes in the blood of these patients.

- Study of the specific molecular alterations in $CD4^+$ T cells to:
  - Unravel gene and signaling pathway alterations that contribute to the appearance of MHE.
  - Identify miRNAs and transcription factors involved in the modulation of differentially expressed messenger RNAs (mRNAs) in MHE as well as their underlying mechanisms of action.
  - Pinpoint mechanisms by which the altered levels of miRNAs and transcription factors may contribute to the immune system shift that triggers MHE.
  - Perform a detailed analysis of $CD4^+$ T-cell receptor repertoires in controls and in patients with cirrhosis with or without MHE.

2.3. Contributions

The development of this thesis started in 2017 at the Centro de Investigación Príncipe Felipe as a synergistic project between translational, clinical, and computational research. I awarded an

IMP/IMFAHE (International Mentor Program/International Mentoring Foundation for the Advancement of Higher Education) engineering fellowship a few months prior, which allowed me to complete an international internship in the laboratory of Dr. Claudia Angelini (Naples, Italy) at the beginning of my thesis work. This stay taught me the basics of different statistical methods for multi-omic data integration, which was later extended through a bachelor's course in the Statistics Department at the Polytechnic University of Valencia.

This multi-omic integration training was applied to the analysis of multi-omic datasets in our first study (Chapter 3). The preliminary results were presented at the VI Student Symposium organized by the International Society for Computational Biology Student Council in 2018 (Granada, Spain), where I won the award for the best talk. A more extended version of the same work was also selected for oral presentation at the ComBioPreVal 2019 conference (Valencia, Spain) and as poster in the international ISCB/ECCB (International Society for Computational Biology/European Conferences on Computational Biology) 2019 conference (Basel, Switzerland).

Finally, this work was published at the beginning of 2021 in the Scientific Reports journal, where we described a multi-omic (microarray, metabolomics, and cytokine panel) integration analysis to unravel the molecular mechanisms of the immune system in whole blood samples of patients with and without MHE (Chapter 3). Study of the etiology of MHE was continued with a transcriptomic analysis of $CD4^+$ T lymphocytes from patients with and without MHE (Chapter 4) which introduced me to the analysis of RNA-seq and miRNA-seq datasets. This study resulted in a paper that is currently submitted to the Translational Research journal.

During the last year of my PhD, I became extremely interested in the use of next-generation sequencing technologies to study antibody/B-cell and T-cell receptor repertoires and their applications in patients with MHE. An additional online international internship in the laboratory of Dr. Victor Greiff at the University of Oslo (Norway) introduced me to immune receptor analysis using computational methods. Taking advantage of our $CD4^+$ lymphocyte dataset, we developed a Nextflow pipeline for T-cell receptor analysis from RNA-seq data (Chapter 5) that resulted in a third scientific article (published in the ImmunoInformatics journal, 2022). Additionally, this crossover between bioinformatic fields resulted in the development of an R package in collaboration with Dr. Greiff's lab to integrate both immune receptor and transcriptomic datasets.

Working under the supervision of 3 thesis directors from 2 different laboratories (neurobiology and biostatistics/bioinformatics) has greatly enriched my knowledge. This experience has also allowed me to participate in different research projects in both groups which has further contributed to boosting my training.

Several collaborations also emerged during my PhD training, including (1) the multi-omic study in type 1 diabetes (TEDDY project) in which I compared the multi-way latent variable approach used for patient classification in this study (NPLS-DA) with a sparse regression model (logistic regression with variable selection) to contrast the performance of both models; (2) a multi-omic study of obesity in germ-free mice fed with a high fat diet, in which I supervised and helped a collaborator's PhD student; (3) a metabolomic study in patients with MHE to understand the effects of the antibiotic rifaximin, in which I analyzed the dataset; (4) analysis of the exosome cargo from patients with MHE, in which I analyzed miRNA-seq and proteomic data; and (5) the effect of rifaximin

on the microbiome of MHE model rats, in which I examined 16S rRNA gene sequencing data for species analysis.

Another job I really enjoyed throughout my PhD work and which showed me my own intellectual maturity was the opportunity to be a reviewer for the Scientific Data journal. In this capacity I evaluated the correct technical and biological validation of a large multi-omic dataset measured in patients with rheumatoid arthritis and checked the adequate deposition of these datasets into appropriate repositories.

In addition, the knowledge of multi-omic analysis I acquired during my thesis work gave me the opportunity to co-supervise a bachelor's degree thesis from the Biotechnology Program at the Polytechnic University of Valencia. This multi-omic study was undertaken in collaboration with the La Fe University Hospital to determine the differences in breast milk exosomes between preterm and to-term births.

Finally, in the last year of my thesis, I also had the opportunity to demonstrate the state of the art in multi-omic analyses to other scientists by teaching in different courses including "Multi-omic Integrative Analysis of Gene Expression" (Centro de Investigación Príncipe Felipe, Valencia) in 2018 and "Transcriptómica en biomedicina: bases de datos y análisis de expresión" [Transcriptomics in biomedicine: databases and expression analysis] (Universidad Católica de Valencia San Vicente Mártir, Valencia) in 2019. Lastly, I have also been a teacher of the "Integrative Analysis of Multiple Omics" master's degree course in Omic Data Analysis and Systems Biology at the University of Sevilla in 2022.

2.3.1. Journal papers

- Teresa Rubio, Vicente Felipo, Sonia Tarazona, Roberta Pastorelli, Desamparados Escudero-García, Joan Tosca, Amparo Urios, Ana Conesa, Carmina Montoliu. *Multi-omic analysis unveils biological pathways in peripheral immune system associated to minimal hepatic encephalopathy appearance in cirrhotic patients*. Scientific Reports, 11, 2021. DOI: 10.1038/s41598-020-80941-7

- Teresa Rubio, Ana Conesa, Sonia Tarazona, Cristina Marti, Paula Izuierdo-Altarejos, María-Pilar Ballester, Juan José Gallego, Desamparados Escudero-García, Amparo Urios, Carmina Montoliu, Vicente Felipo. *Gene expression and signaling pathways in lymphocytes of cirrhotic patients with cognitive impairment. Role of miRNAs*. Submitted to Translational Research, 2021.

- Leandro Balzano-Nogueira, Ricardo Ramirez, Tatyana Zamkovaya, Jordan Dailey, Alexandria N. Ardissone, Srikar Chamala, Joan Serrano-Quilez, Teresa Rubio, Michael J. Haller, Patrick Concannon, Mark A. Atkinson, Desmond A. Schatz, Eric W. Triplett, Ana Conesa. *Integrative analyses of TEDDY Omics data reveal lipid metabolism abnormalities, increased intracellular ROS and heightened inflammation prior to autoimmunity for type 1 diabetes*. Genome Biology, 22, 2021. DOI: 10.1186/s13059-021-02262-w.

- Teresa Rubio, Maria Chernigovskaya, Susanna Marquez, Cristina Marti, Paula Izquierdo-Altarejos, Amparo Urios, Carmina Montoliu, Vicente Felipo, Ana Conesa, Victor Greiff, Sonia Tarazona. *A*

Nextflow pipeline for T-cell receptor repertoire reconstruction and analysis from RNA sequencing data. Immunoinformatics, 6, 2022. DOI: 10.1016/j.immuno.2022.100012

- Cédric R. Weber, Teresa Rubio, Longlong Wang, Wei Zhang, Philippe A. Robert, Rahmad Akbar, Igor Snapkov, Jinghua Wu, Marieke L. Kuijjer, Sonia Tarazona, Ana Conesa, Geir K. Sandve, Xiao Liu, Sai T. Reddy, Victor Greiff. 2022. *Reference-based comparison of adaptive immune receptor repertoires*. bioRxiv. DOI: 10.1101/2022.01.23.476436

2.3.2. Conferences

- XIV Symposium on Bioinformatics (JBI 2018). Granada, Spain. November 2018. Teresa Rubio, Vicente Felipo, Carmina Montoliu, Sonia Tarazona, Ana Conesa. “Integrative multi-omics analysis to explain immune alterations in minimal hepatic encephalopathy patients” (Oral presentation by Teresa Rubio).

- III Congreso Nacional de Jóvenes Investigadores en Biomedicina [3rd National Conference of Young Researchers in Biomedicine]. Valencia, Spain. April 2019. Teresa Rubio, Vicente Felipo, Sonia Tarazona, Carmina Montoliu, Ana Conesa. “Multi-omics data integration uncovers candidate molecular biomarkers for minimal hepatic encephalopathy” (Oral presentation by Teresa Rubio).

- 27th Conference on Intelligent Systems for Molecular Biology and the 18th European Conference on Computational Biology (ISMB/ECCB 2019). Basel, Switzerland. July 2019. Teresa Rubio, Vicente Felipo, Sonia Tarazona, Carmina Montoliu, Ana Conesa.

"Integrative multi-omics analysis to explain immune alterations in minimal hepatic encephalopathy patients" (Poster).

- EASL - THE INTERNATIONAL LIVER CONGRESS™ 2021. Online. June 2021. Teresa Rubio, María-Pilar Ballester-Ferré, Juan José Gallego Roig, Alessandra Fiorillo, Franc Casanova, Sonia Tarazona, Carla Giménez-Garzó, Amparo Urios, Paloma Lluch, Desamparados Escudero-García, Juan Tosca, Cristina Montón, José Ballester, María Pilar Ríos, Lucía Durbán, Ana Conesa, Vicente Felipo, Carmina Montoliu. "Multi-omic analysis unveils biological pathways in peripheral immune system associated to minimal hepatic encephalopathy appearance in cirrhotic patients" (Poster).

Chapter 3

Multi-omic analysis of changes in the peripheral immune system associated with the appearance of minimal hepatic encephalopathy in patients with cirrhosis

This chapter was adapted from the paper '*Multi-omic analysis unveils biological pathways in peripheral immune system associated to minimal hepatic encephalopathy appearance in patients with cirrhosis'* by Teresa Rubio, Vicente Felipo, Sonia Tarazona, Roberta Pastorelli, Desamparados Escudero-García, Joan Tosca, Amparo Urios, Ana Conesa, Carmina Montoliu. Scientific Reports, 11, 2021. DOI: 10.1038/s41598-020-80941-7

3.1. Introduction

The precise pathways that trigger a systemic immune response leading to a neurological disorder are still poorly understood. Next to the role of pro-inflammatory chemokines, other signaling, metabolic, and regulatory components are also likely to contribute. Multi-omic approaches help us to investigate the set of molecular and cellular events associated with these kinds of diseases. In fact, multi-omic technologies such as genomics, epigenomics, transcriptomics, proteomics, and metabolomics are increasingly being used to profile the multi-layered components of living cells and pathological processes[130–132].

The rationale behind this strategy is that disease usually impacts several types of biomolecules and expanding the molecular space under study will increase the likelihood of identifying relevant biomarkers. In the context of neurological pathologies, multi-omics has allowed the modeling of the highly sophisticated brain metabolic networks and their contribution to human health[133,134], as well as the identification of exclusive and common features for these disease types[135].

The revolution in omic data generation holds great promise in life sciences fields. Although multi-omic datasets are easy to produce in terms of economic cost and time, the effort required for data curation and integration is much greater and requires specialized data analysts. The generation of biological knowledge from this large amount of data is probably the biggest challenge in this context. Traditional functional enrichment methods in bioinformatics give an overview of the biological pathways altered by the disease, allowing scientists to jointly look at the biological functions of all the altered genes.

Useful web-based tools such as **Reactome**[136] or **PaintOmics**[104] have been developed for the visualization of functional enrichment results represented as biological pathways, without the need for advanced bioinformatic skills. Other tools like **STRING**[137] collect protein information from numerous sources (i.e., experimental data, computational prediction methods, and public text collections) to construct protein–protein interaction networks that allow biological interpretation at the protein level. All these approaches generally focus on single-omic datasets of each data type which can be linked to genes (e.g., proteins to mRNA, miRNA to target genes, etc.), although **PaintOmics** is one of the first to combine multiple omics into the same functional enrichment analysis. In contrast, the output from very sophisticated integration models such as machine or deep learning methods are often exceedingly difficult to interpret from a biological viewpoint.

In this chapter we deployed a bioinformatics analysis pipeline that combined univariate, multivariate, and enrichment methods to find intracellular and extracellular compounds that shared variation patterns linked to the MHE phenotype. Specifically, we combined blood transcriptomics, serum metabolomics, and cytokines to identify pathways and biomarkers associated with MHE in patients with cirrhosis.

Using this approach, we discovered a relationship between extracellular CCL20, CX3CL1, CXCL13, IL-15, IL-22, and IL-6 with alterations in chemotactic receptors and ligands in patients with MHE. In addition, this methodology suggested a link between long-chain unsaturated phospholipids and increased fatty acid transport and prostaglandin production, which may jointly contribute to the onset of mild cognitive impairment. These results illustrate the power of integrative statistical

analysis of multi-omic data in modelling disease processes and connecting phenotypic changes across molecular layers.

3.2. Methods

3.2.1. Overview of the analysis strategy

To understand the joint contribution of gene expression and extracellular metabolite changes to the induction of MHE, we developed an extensive pipeline that combined univariate and multivariate regression analysis for feature selection while also considering the biology of the system (Figure 3.1). Briefly, first we identified genes and associated pathways that significantly changed between cirrhotic individuals with and without MHE. In parallel, metabolites and cytokines with significant serum level changes were identified and grouped by their variation pattern across individuals. These groups represented metabolic modules that changed in a coordinated fashion, suggesting common underlying mechanisms.

Next, we asked if a specific gene expression signature was associated with each of these modules. Therefore, we applied the PLS method to the multi-omic data, taking the selected genes as explanatory variables and each module of extracellular compounds as a response variable. This analysis selected the genes with the strongest associations with each metabolic group which were then further investigated by pathway enrichment and biological interaction modeling (Figure 3.1).

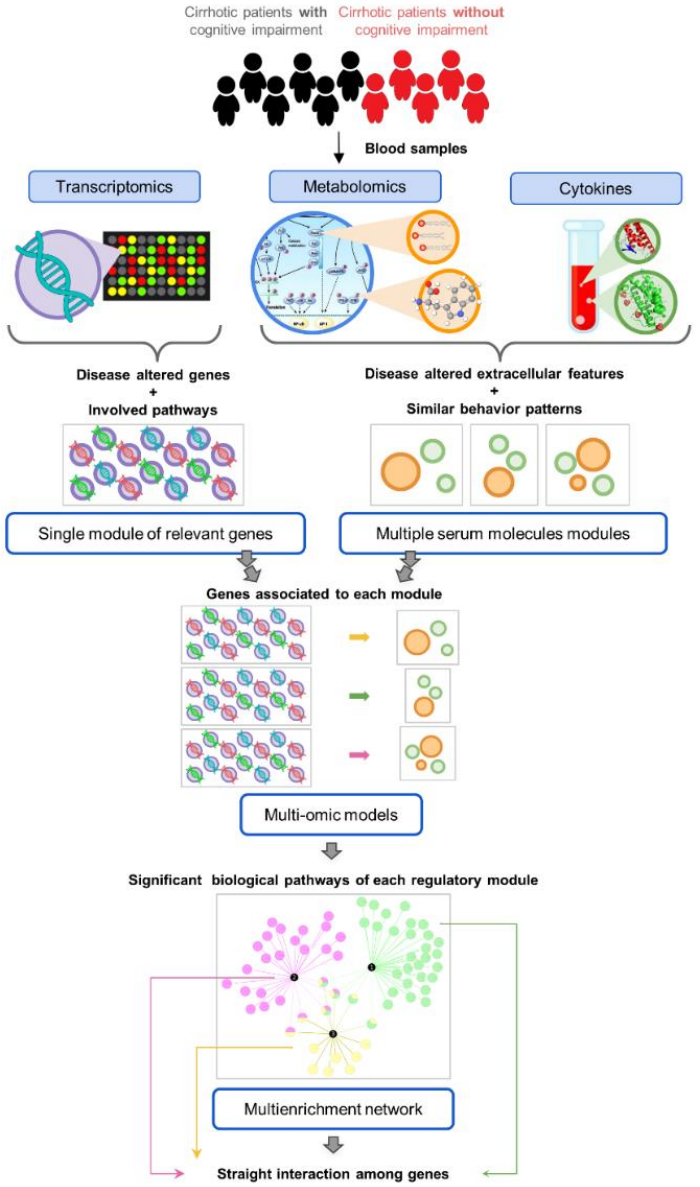


Figure 3.1 Overview of the analysis strategy. Transcriptomics, metabolomics, and cytokines were measured in blood samples from 11 patients with cirrhosis (6 with and 5 without minimal hepatic encephalopathy). Step 1: genes and pathways altered in minimal hepatic encephalopathy. Step 2: altered metabolites and cytokines grouped by correlation. Step 3: integration of extracellular and intracellular datasets. Step 4: enrichment and network analysis of iterative models.

3.2.2. Patients and sample collection

Eleven patients with liver cirrhosis were recruited from the outpatient clinics at the Hospital Clínico Universitario de Valencia, Spain. The diagnosis of cirrhosis was based on clinical, biochemical, and ultrasonographic data. The exclusion criteria were overt hepatic encephalopathy, recent (< 6 months) alcohol consumption, infection, recent (< 6 weeks) antibiotic use or gastrointestinal bleeding, recent (< 6 weeks) use of drugs affecting cognitive function, presence of a hepatocellular carcinoma, or a diagnosis with a neurological or psychiatric disorder.

Patients included in the study did not show fever or any clinical or biological signs of recent infection and none of them had hypothyroidism or altered thyroid stimulating hormone (TSH) levels. All the participants were included in the study after signing their written informed consent to participation. The study protocols were approved by the Scientific and Ethics Committees at the Hospital Clínico Universitario de Valencia. The procedures followed were in accordance with the ethical guidelines set out in the Declaration of Helsinki.

Diagnosis of minimal hepatic encephalopathy. MHE was diagnosed using the PHES which comprises 5 psychometric tests[8,138]. The scores were adjusted for patient age and education levels using Spanish normality tables (www.redeh.org). Patients were classified as having MHE when the PHES score was ≤ −4 points. From the 11 recruited patients with cirrhosis, 5 did not have MHE and 6 had been diagnosed with MHE. All the patients were males and the mean age was 70.7 ± 4.0 in the group

with MHE and 63.8 ± 1.2 in the group without MHE, with no significant differences between them (*t*-test; *p*-value = 0.157).

Sample collection. The transcriptomic dataset was measured in blood samples collected in PAXgene® blood RNA tubes (BD Biosciences) that had been frozen at −20 °C for 24 h and kept at −80 °C for subsequent analysis. For the metabolomic measurements, plasma samples were obtained from blood collected in BD P100 tubes (BD Biosciences) containing EDTA and protein stabilizers. Plasma samples were obtained after 2 serial centrifugations, the first at 5 °C, 2500 *g*, for 20 min, and the second at 10 °C, 2500 *g*, for 10 min after prior transfer of the plasma to a new tube. The plasma samples were distributed into several aliquots and stored at −80 °C. Cytokines were measured in the serum fraction after blood sample collection in tubes without EDTA.

3.2.3. Transcriptomic profiling of plasma samples

RNA extracted from peripheral blood cells was quantified with a NanoDrop ND1000 spectrophotometer (NanoDrop Technologies, Wilmington, Delaware USA) and RNA quality was confirmed with an RNA 6000 Nano Bioanalyzer assay (Agilent Technologies, Palo Alto, California USA). Using the One-Color Low Input Quick Amp Labelling Kit (Agilent p/n 5190-2305) according to the manufacturer's instructions; 200 ng of total RNA were used to produce Cyanine 3-CTP-labeled cRNA. Following the 'One-Color Microarray-Based Gene Expression Analysis' protocol Version 6.7 (Agilent p/n G4140-90040), 600 ng of labeled cRNA was hybridized with SurePrint G3 Human Gene Expression Microarrays v3 8X60K (Agilent p/n G4858A-072363). The microarrays were scanned in an Agilent Microarray Scanner (Agilent G2565C) as described in the guidelines from the

manufacturer and the images were analyzed with **Agilent Feature Extraction Software** 11.5.1.1 using the default parameters (protocol GE1_1105_Oct12, grid template 072363_D_F_20150612, and QC Metric Set GE1_QCMT_Oct12).

3.2.4. Metabolomic profiling of serum samples

Targeted metabolomic analysis was performed using the Absolute-IDQ™ P180 kit (BIOCRATES Life Sciences AG, Innsbruck, Austria). The metabolite extracts were processed following the manufacturer's instructions and were analyzed on a triple-quadrupole mass spectrometer (AB SCIEX triple-quad 5500) operating in the multiple reaction monitoring (MRM) mode. This assay is based on phenylisothiocyanate-derivatization in the presence of internal standards for the analysis of resolved amino acids and biogenic amines and was analyzed by liquid chromatography (LC) tandem mass spectrometry (MS/MS) using scheduled MRMs.

Separation was achieved on an Agilent Zorbax Eclipse XDB C18 column (3 × 100 mm, 3.5 µm) with mobile phases of 0.2% formic acid in water (A) and 0.2% formic acid in acetonitrile (B). The gradient program was as follows: 0 min, 100% A; 0.5 min, 100% A; 5.5 min, 5% A; 6.5 min, 5% A; 7.0 min, 100% A; and 9.5 min 100% A. The column was maintained at 50 °C, with a 0.5 ml/min flow rate and 10 µl volume injection.

Sample analysis in the mass spectrometer was performed in positive electrospray ionization (ESI) mode. The declustering potential (DP) and collision energy (CE) were specified by the kit. The MRM MS/MS detector conditions were set as follows: temperature, 500 °C; ion spray voltage, 5500 V; curtain gas, 20 psi; collisionally activated dissociation (CAD), medium; entrance potential (EP) 10 V; collision cell exit potential (CXP),

15 V; ion source gas 1 (GS1), 40 psi; and ion source gas 2 (GS2), 50 psi. The retention times for every metabolite were previously adjusted to the experiment.

To analyze acylcarnitines, glycerophospholipids, and hexoses, subsequent flow injection analysis tandem mass spectrometry (FIA-MS/MS) was performed. Red PEEK tubing (1/16 × 0.005) was used for the separation with Biocrates solvent I diluted in 290 ml methanol as the mobile phase ('isocratic elution mode'). The gradient program was as follows: 0 min, 0.03 ml/min; 1.6 min, 0.03 ml/min; 2.4 min, 0.20 ml/min; 2.8 min, 0.20 ml/min; 3.0 min, and 0.03 ml/min. The positive ESI mode was used in the spectrometry, although some parameters were different for lipids and sugars. The same parameters were used for lipids in the LC-MS/MS, except for the temperature (200 °C). For sugars, the following parameters were applied: ion spray voltage, −4500 V (negative); curtain gas, 20 psi; CAD, medium; EP, −10 V; CXP, −15 V; GS1, 40 psi; GS2, 50 psi; DP, −55 V; and CE, −12 V.

The lipid concentrations were automatically calculated in µM using **MetIDQTM** software (Biocrates Life Science AG, Innsbruck, Austria) and the rest of metabolites were quantified from standard curves with **Analyst** software from SCIEX. Isotope-labeled internal standards were integrated into the platform for absolute metabolite quantification. The measurements were performed in a 96-well format. The metabolite limit of detection was set to 3 times the value of the 'zero samples' and the average coefficient of variation of the metabolites among the biological replicates was 30%.

Based on the 5 internal quality controls, technical variation was below 15%. A metabolite was excluded from further analyses if its concentration measurement data did not meet all of the following criteria: (1) fewer than 20% missing values (non-detectable peak) for each quantified metabolite in each experimental group; (2) 50% of all the measured sample concentrations for the metabolite had to exceed the limit of detection[139]. The provided metabolomic data matrix included 143 metabolites without missing values: 8 acylcarnitines, 21 amino acids, 17 biogenic amines, the sum of all the hexoses, 15 sphingomyelins (SMs), and 81 glycerophospholipids including lysophosphatidylcholines (lysoPC) and phosphatidylcholines (PCs).

According to platform annotation, lipid side-chain composition was abbreviated as 'C x:y', where 'x' denotes the number of carbon atoms and 'y' denotes the number of double bonds present in the fatty acid residues. The presence of a hydroxyl group in SMs was indicated with an −OH. Glycerophospholipids were differentiated according to the presence of ester and ether bonds in the glycerol moiety. Double letters (aa = diacyl or ae = acyl-alkyl) indicate that 2 glycerol positions were bound to a fatty acid residue, while a single letter (a = acyl or e = alkyl) indicated a bond with only 1 fatty acid residue.

3.2.5. Analysis of cytokines in serum samples

The concentrations of IL-4, IL-6, IL-13, IL-17, IL-18, IL-22, and TGFβ (Affymetrix eBioscience, Vienna, Austria) and IL-10, IL-12, IL-15, CXCL13, CCL20, CX3CL1, and TNFα (R&D Systems, Minneapolis, MN, USA) were measured by ELISA according to the manufacturer's instructions. High sensitivity kits were required to assess IL-17, IL-4, and TNFα.

3.2.6. Data pre-processing plots

Mean-Difference (MD) plots. MD plots are XY scatter plots that compare differences versus the means of 2 quantitative measurements[140]. We used the plotMD() function in the **limma** R/Bioconductor package[85] for microarray data, which represents the log-expression of 1 sample as the x-axis and the expression log-ratio of this sample against the rest of samples as the y-axis.

Principal component analysis plots. PCA[141] is a multivariate dimension-reduction approach applied to 1 dataset (e.g., transcriptomics or metabolomics). This technique calculates a set of new variables (latent variables or components) which are linear combinations of the original variables and capture most of the variability in the data. A small number of components—in comparison with the high number of original variables—are often sufficient to describe the dataset. PCA score plots show projections of the samples in the new latent space and these plots are useful for inspecting the presence of outlier samples, possible batch effects, or sample groups.

3.2.7. Single-omic analysis

Transcriptomic data analysis. The variance stabilizing normalization (VSN) method implemented in the **vsn** R/Bioconductor package[142], was applied to stabilize the variance of the microarray intensity data. This method also considers possible variations between samples/arrays and sets the measurements on a common scale before they are compared. Additionally, a long-standing problem in the analysis of microarray gene expression experiments is the mean-variance dependency, that is, the

standard deviation increases linearly with the mean at high intensities, meaning that the obtained values usually cannot be compared directly but rather, only after appropriate transformation (called 'normalization'). Logarithmic transformation alleviates this problem but does not solve it entirely. However, VSN achieves a constant variance independently of the spot intensity. This transformation is an arsinh function as follows:

$$arsinh(x) = log(x + \sqrt{x^2 + 1}), \quad [1]$$

where x represents the fluorescence intensity values.

After data normalization, weighted linear modeling was performed with the **limma** R/Bioconductor package[85] to find genes that were differentially expressed between the 2 groups: patients with cirrhosis with and without cognitive impairment. Genes in the group with MHE were considered to be significantly altered when the FDR < 0.05. Enrichment analysis was calculated using Fisher's exact test implemented in **PaintOmics**[104] to determine whether genes from predefined biological pathways were more present in the list resulting from the differential expression test than would be expected by chance. **PaintOmics** includes the KEGG database[98] for enrichment analysis annotations and subsequent gene expression representations.

Metabolomic data analysis. As in the case of transcriptomics, the **vsn** normalization[142] and **limma**[85] procedures were applied to metabolomic data to find the altered metabolites between the 2 groups of patients. Metabolites in the MHE group were considered to be significantly altered when the FDR < 0.05. Although **vsn** and **limma** were primarily developed for microarray data, they can also be applied to mass spectrometry-based data because the assumption of normality holds. Moreover, the

effectiveness of using these methods for metabolomic datasets has already been validated in several studies[143–145].

Cytokine data analysis. Significant differences in cytokines between patients with and without MHE were tested using the non-parametric Wilcoxon test after homoscedasticity was checked. Multiple testing correction was applied using the Benjamini–Hochberg procedure.

3.2.8. Omic power analysis

The statistical power of each omic data type was evaluated with the **MultiPower** tool[146] as a quality control for differential expression results. For gene expression and metabolomic data, we found power values exceeding 0.75 when aiming to detect features with relatively large changes and low variability (a Cohen's *d* of at least 1.5). For cytokines, a power of at least 0.75 was obtained for a higher proportion of features because they presented bigger effect sizes compared to the other omics. Thus, we concluded that, despite the small sample size available, we had enough power to validate our results and proceed with the analysis.

3.2.9. Obtaining modules of coordinated metabolites and cytokines

Clustering analysis was performed to classify similar metabolites and cytokines into groups called 'clusters'. The endpoint is a set of clusters, where each one is distinct from the others and the objects within every cluster are broadly similar. A comparative analysis of several methods, including hierarchical clustering, K-means, and partitioning around medoids (PAM)[147] was performed. The agglomerative hierarchical clustering algorithm starts by treating each observation as a separate

cluster and iteratively executes the identification of the 2 closest clusters, merging together the 2 most similar clusters.

Both the K-means and PAM algorithms aim to partition the dataset into groups and attempt to minimize the distance between points labeled for each cluster and a point designated as the center of that cluster. The difference between them is that PAM chooses actual data points as the centers (medoids), thereby allowing better interpretability and robustness, while the center in K-means is the average between the points in the cluster. These 3 clustering methods were evaluated using a Silhouette analysis[148], which measures clustering quality to determine how well each observation lies within its cluster.

We finally decided to apply PAM clustering analysis to our data because it provided the best Silhouette results. We used the absolute Spearman correlation coefficient value (*r*) between all the significant serum features, including metabolites and cytokines. Specifically, the distance measure was $\sqrt{2(1-r^2)}$. Every cluster was a group of features with similar variation patterns across patients and clusters were denoted as 'modules'.

3.2.10. Integration of multi-omic datasets

We used PLS regression[114] to link changes in gene expression with the variation in the serum levels of metabolites and cytokines. PLS is a multivariate dimension-reduction approach that integrates 2 data matrices with the same number (n) of observations: X, is the explanatory or independent variable with p variables (e.g., transcriptomics), and Y, is the response or dependent variable with q variables (e.g., metabolomics). Like PCA, PLS aims to determine new variables or components (k) that

are linear combinations of predictive variables. The formal problem that must be optimized in PLS is finding the latent variables or components in X and Y so as to maximize the covariance between them.

X and Y were modeled using regression models:

$$X = TP^T + E_X \quad [2]$$
$$Y = UQ^T + E_Y \quad [3]$$

where T and U are score matrices, P and Q are loading matrices, and E_X and E_Y are residual matrices of X and Y, respectively. In addition, T and U are linear combinations of X and Y and contain information about the observations (as well as their similarities or dissimilarities with respect to the given problem). There is also an internal relationship between T and U:

$$U = TD + H \quad [4]$$

where $D_{k\times k}$ is a diagonal matrix, whose main diagonal includes the elements d_1, d_2, …, d_k and $H_{n\times k}$ is the residual matrix. Therefore:

$$Y = TDQ^T + HQ^T + E_Y = TC^T + E_Y^* \quad [5]$$

where $C_{q\times k}$ is the Y-weighting matrix.

Let $W_{n\times k}$ be the X-weighting matrix that satisfies orthogonality $W^TW = 1$. The w and c weights give us information on how the variables combine to form the quantitative relationship between X and Y, thus providing an interpretation of the t and u projections. Therefore, these weights are essential for understanding which X variables are important (numerically large values of w) and which ones provide the same information (similar w values).

In this study, the same set of genes were used as explanatory variables (X) in all the PLS models created. These genes were selected because they had an FDR < 0.05 in the **limma** analysis (i.e., 'significant' genes). In this way, we compiled a gene expression matrix of explanatory variables that were relevant to the disease. Next, the serum levels of the metabolites and cytokines included in each module were taken as the response Y matrix, thereby creating module-specific PLS models.

Two parameters were calculated to assess the performance of the PLS models: R^2 (multiple correlation coefficient; degree of Y variance explained by X) and Q^2 (predictive power). R^2 was calculated as 1 – the residual sum of squares (RSS) and the total sum of squares (TSS) thus:

$$R^2 = 1 - RSS/TSS \quad [6]$$

$$RSS = \sum (y - \hat{y})^2 \quad [7]$$

$$TSS = \sum (y - \bar{y})^2 \quad [8]$$

While Q^2 was calculated as 1 − predictive residual error sum of squares (PRESS) / TSS:

$$Q^2 = 1 - PRESS/TSS \quad [9]$$

$$PRESS = \sum (y - \hat{y})^2 \quad [10]$$

The calculations for R^2 and Q^2 were almost identical, with the only difference being that RSS was calculated from the data upon which the algorithm was trained and PRESS was calculated by cross-validation procedures. Because the number of samples was small in this study, leave-one-out cross-validation was used to compute Q^2. Gene loadings of the resulting PLS models are a measure of the relative importance of the gene variables to the model. Enrichment analyses using the **mdgsa**

R/Bioconductor package[102] were run for the list of genes ranked by the loading values of the first component of the PLS models. The Biological Process branch of the GO[97] database was chosen to annotate the enrichment analyses.

3.2.11. Data and code availability

The transcriptomic dataset used in this study is available in the GEO database repository, GSE149741, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE149741. The R scripts are publicly available from the Github repository: https://github.com/ConesaLab/MHE_Multi-Omics_Integration.

3.3. Results and discussion

3.3.1. Data pre-processing

An initial exploratory analysis was required to choose the appropriate pre-processing methods for correcting possible technical/platform-dependent noise. Different plots were used to inspect the data distributions and to detect the presence of outlier samples, possible batch effects, or sample grouping (Figures 3.2 to 3.4). For the Agilent microarray dataset, after reading the raw microarray data files which include the hybridization quantification at the probe level, as well as local background intensities to be subtracted, 60,901 probes were detected—1,767 positive control (PC) and 308 negative control (NC) spots.

As an example, the MD plots (see methods) for patient PC55 (with cirrhosis but without MHE) were also generated to show the normalization

effect using PC and NC probes (Figure 3.2A–C). The sample distributions were comparable after normalization (Figure 3.2D), although separation of the patient groups in the PCA score plot was unclear (Figure 3.2E).

The metabolite concentration levels in the serum samples of the same patients with cirrhosis varied from serotonin (0.11 micromolar) to sugars (81,569.91 micromolar; Figure 3.3A). The exceptionally large value in the latter was because the metabolomic platform used in this work quantifies all the hexoses (including glucose) as a single variable. Metabolomic analysis showed different distributions across samples corrected by applying the vsn normalization method (see Methods) to make the samples comparable (Figure 3.3A and C). After normalization, the PCA of this dataset revealed almost perfect separation between the groups of patients with and without MHE, except for PC57 in PC1 (with 32.65% explained variability).

The cytokines measured in the serum samples also showed different concentration levels (Figure 3.4A), from IL-17 with a mean of 0.75 pg/ml to TGFβ which presented 4 orders of magnitude more concentration than the rest of the variables across the samples (an average of 14,124 pg/ml). The PCA score plot of cytokine data (Figure 3.4B) showed that the first principal component (PC1) perfectly separated patients with cirrhosis with and without MHE.

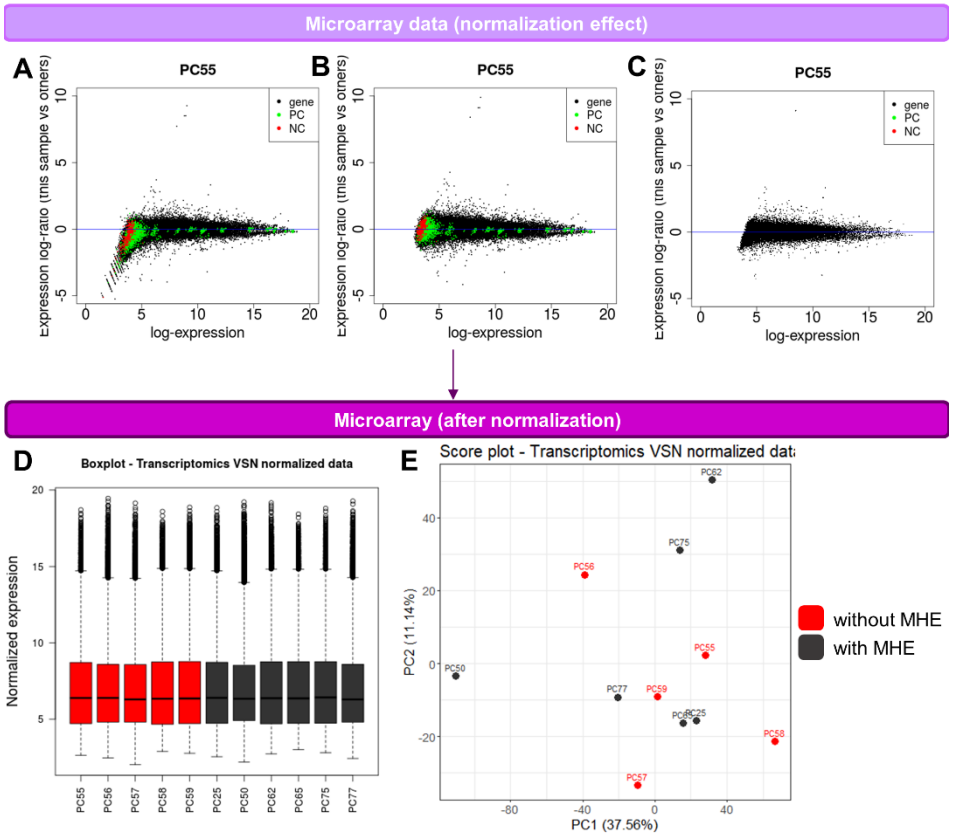


Figure 3.2 Microarray exploratory analysis. Median-Difference (MD) plots of microarray raw data (A), normalized data (B), and normalized data without background control probes (C), graphically representing the normalization effect. The red dots are the negative controls (NC), green dots are the positive controls (PC), and black dots represent genes. A boxplot (D) and PCA score plot (E) showed the data distribution and samples after normalization, respectively. Sample groups are colored in red (without MHE) or dark grey (with MHE).

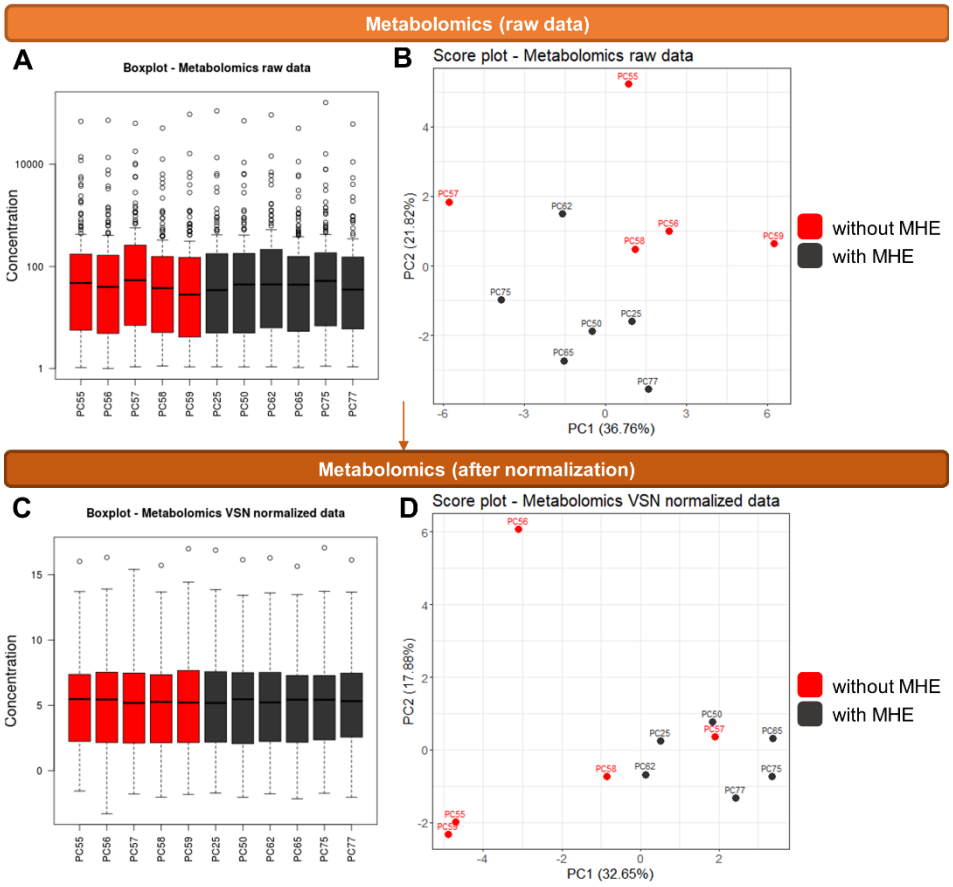


Figure 3.3 Metabolomic exploratory analysis. Boxplot and PCA score plots of the metabolomic dataset from patients with cirrhosis without (red) and with (dark grey) MHE before (A–B) and after (C–D) normalization. For visualization purposes, the boxplot x-axis was represented on a logarithmic scale and data for the PCA analysis were log-transformed before normalization. In PCA plot axes, the percentage between the parenthesis indicates the amount of explained variability. Abbreviations: PC1 and PC2, principal components 1 and 2, respectively.

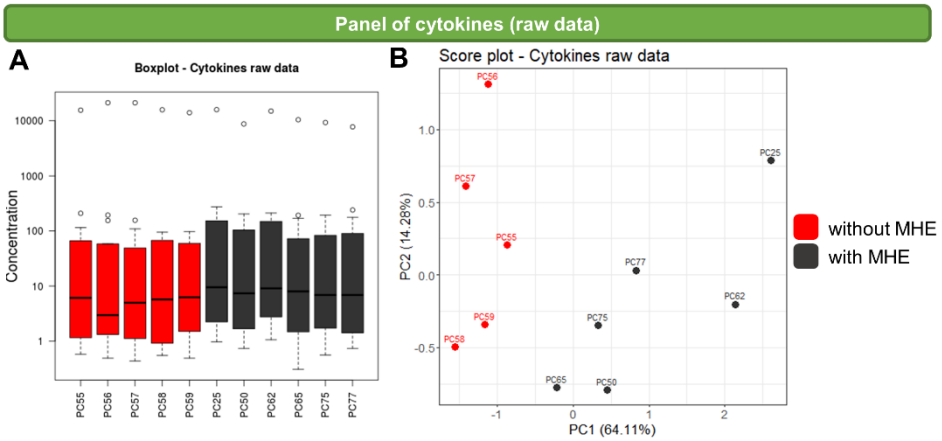


Figure 3.4 Cytokine exploratory analysis. (A) a boxplot and (B) PCA score plots of cytokine dataset from patients with cirrhosis without (red) and with (dark grey) MHE. For visualization purposes, the boxplot x-axis was represented on logarithmic scale and the data for the PCA analysis were previously log-transformed. The percentage shown in parentheses on the PCA plot axes indicates the amount of explained variability. Abbreviations: PC1 and PC2, principal components 1 and 2, respectively.

3.3.2. Transcriptomic analysis confirms changes in the immunophenotype in patients with minimal hepatic encephalopathy and unveils new altered pathways

Gene expression analysis identified 847 differentially expressed genes with an FDR < 0.05 (Figure 3.5) from among the 24,460 measured variables. The top 10 significantly altered genes were *MS4A4A*, *APOB*, *ACVRL1*, *AREG*, *FAM65B*, *PSPH*, *PDK4*, *SEMA3F*, *CLEC6A,* and *PLIN1*, with some of them showing important roles within the immune system. For instance, MS4A4A (membrane-spanning 4-domains, subfamily A, member 4) has immunotherapeutic potential in blood cancer because of its exclusive expression in 'alternatively activated' (M2) macrophages and monocytes, but not in granulocytes or lymphocytes[149].

Overexpressed AREG (amphiregulin) levels have been associated with inflammatory conditions such as asthma and rheumatoid arthritis[150]. SEMA3F (semaphorin 3F) signaling regulates neutrophil retention in inflamed tissues, with consequences for neutrophil clearance and the resolution of inflammation. The decrease in FAM65B (family with sequence similarity 65 member B, also known as PL48 or RIPOR2) expression is required for T-lymphocyte proliferation upon T-cell receptor engagement[151]. The other genes are involved in different lipid, serine, or glucose metabolism processes in immune system cells.

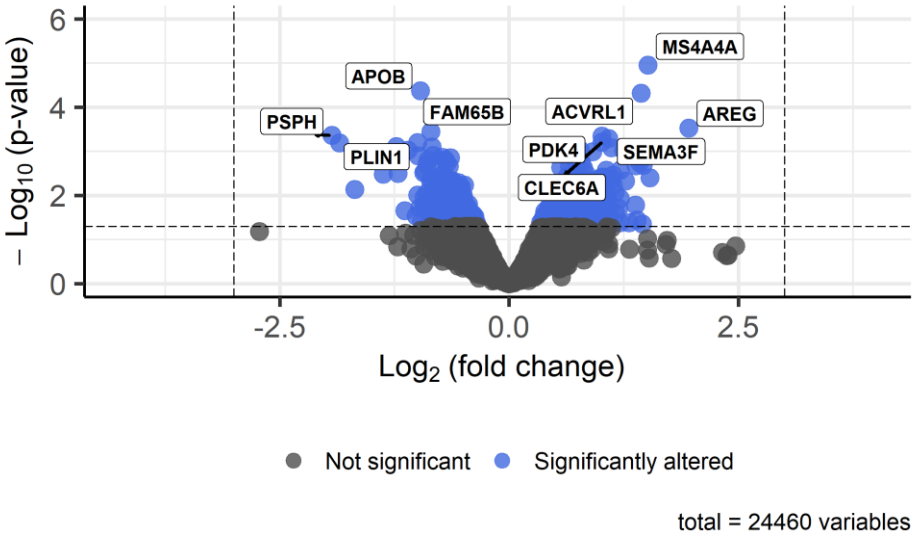


Figure 3.5 Results of the differential expression analysis comparing patients with and without MHE. The volcano plot shows the statistical significance (logarithmically transformed *p*-values) versus the ratio between groups of patients (logarithmically transformed fold-changes). The blue dots represent significantly altered genes (FDR < 0.05) while genes with no significant changes are colored in grey. The top 10 differentially expressed genes are labeled using the nomenclature of the HUGO Gene Nomenclature Committee (HGNC).

To jointly look at the biological functions of all the altered genes, we also applied a pathway enrichment analysis that resulted in 11 significant

pathways with *p*-values < 0.05 (Table 3.1) when comparing patients with MHE to those without MHE. Three additional pathways with *p*-values < 0.1 ('primary bile acid biosynthesis', 'intestinal immune network for IgA production', and 'Th17 cell differentiation') were also included in the discussion because of their strong relevance to MHE.

Table 3.1 Pathway enrichment analysis comparing patients with and without MHE.

KEGG pathway	Unique genes	*p*-value
Cytokine–cytokine receptor interaction	196	0.000001
Hematopoietic cell lineage	83	0.000139
Peroxisome proliferator-activated receptor (PPAR) signaling pathway	49	0.000259
Fat digestion and absorption	31	0.004812
Cholesterol metabolism	35	0.007909
Antigen processing and presentation	68	0.008833
Jak–STAT signaling pathway	116	0.018337
Oxidative phosphorylation	119	0.020249
Steroid biosynthesis	16	0.021786
Th1 and Th2 cell differentiation	84	0.024095
Ovarian steroidogenesis	32	0.024764
Histidine metabolism	22	0.037213
Cell adhesion molecules (CAMs)	116	0.050105
Primary bile acid biosynthesis	10	0.051761
Intestinal immune network for IgA production	40	0.053478
Th17 cell differentiation	99	0.058921

Ranked list of KEGG pathways ordered by their *p*-values, according to the enriched analysis completed with **PaintOmics**. The 'unique genes' column indicates the number of genes in our study that overlapped with each pathway.

Pathway enrichment analysis was carried out with **PaintOmics**[104] to create a network of interconnected significant pathways (Figure 3.6). This analysis revealed a network of immunity-related pathways consisting of 'Th1 and Th2 cell differentiation' and 'Th17 cell differentiation' linked with 'cell adhesion molecules (CAMs)' and 'Jak–STAT signaling pathways', in

turn connected with ‘cytokine–cytokine receptor interaction’. The Jak–STAT pathway plays critical roles in the immune system, especially cytokine receptors and the polarization of T helper cells[152]. In addition, ‘antigen processing and presentation’ was joined with ‘intestinal immune network for IgA production’, while ‘hematopoietic cell lineage’ remained a single node. A second group of processes in this network was related to lipid metabolism and included ‘fat digestion and absorption’, ‘cholesterol metabolism’, ‘steroid biosynthesis’, ‘primary bile acid biosynthesis’, and the ‘peroxisome proliferator-activated receptor (PPAR) signaling pathway’.

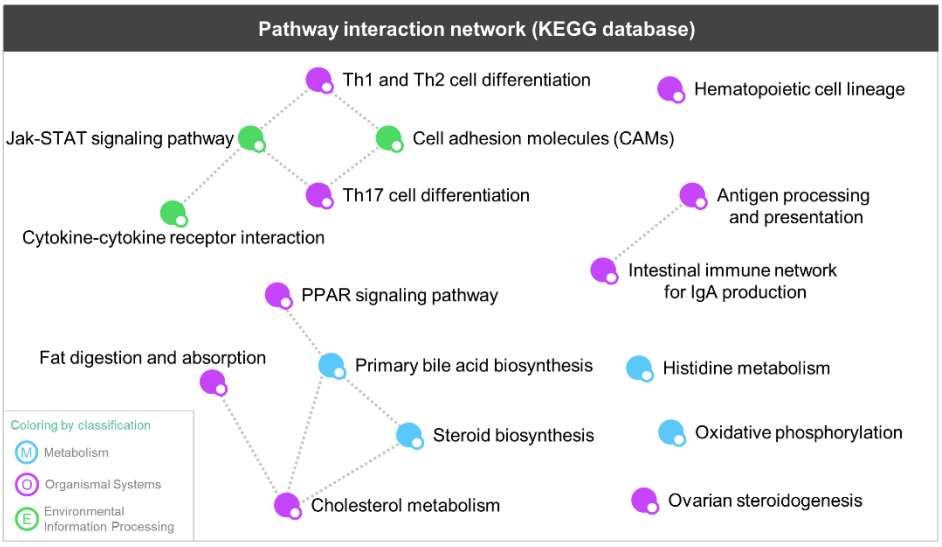


Figure 3.6 Network of interconnected significant KEGG pathways. Nodes represent pathways with a *p*-value lower than 0.05 and are colored according to their KEGG categories. The edges between 2 nodes indicate that both biological processes were closely related.

The advantage of using **PaintOmics** was that this web-based tool allowed the visualization of gene expression levels onto KEGG[98] pathways. Indeed, inspection of the ‘hematopoietic cell lineage’ pathway revealed

significant differences in genes related to B- and T-cell differentiation (Figure 3.7).

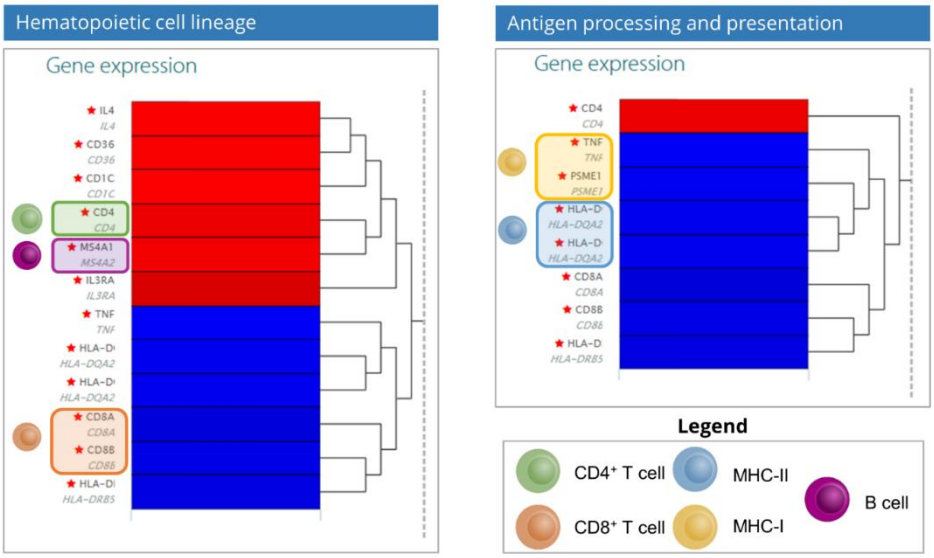


Figure 3.7 Heatmap of relevant genes involved in 2 significant pathways in MHE. Red and blue boxes represent significant gene expression level up- and downregulation, respectively, in patients with MHE versus without MHE. Gene markers of specific immune cell subtypes are highlighted in distinct colors according to the legend palette.

In particular, the mature B-cell marker *MS4A1* (also known as CD20) was upregulated in patients with MHE (Figure 3.7). This gene is a relevant target for autoimmune disease therapies[153] and has also been related to T-cell-dependent immune responses[154]. Taken together, these results suggested new altered immune subtypes in patients with MHE such as upregulated B-cells and downregulated $CD8^+$ T cells.

It has been proposed that the appearance of MHE is associated with a shift in peripheral inflammation to an autoimmune-like profile[14]. Thus, the upregulation of B cells may contribute to potentiating this shift. In addition, B cells may contribute to autoimmune diseases through different functions such as autoantibody secretion, autoantigen presentation, secretion of

inflammatory cytokines, modulation of antigen processing, and the presentation or generation of ectopic germinal centers[155]. It is possible that the upregulation of B cells in MHE may induce some of these functions, thereby contributing to triggering the appearance of MHE.

T cells are divided into different subsets like $CD8^+$ Tc cells and $CD4^+$ Th cells, according to their markers and functions. Of note, our results showed significant upregulation of *CD4* and significant downregulation of *CD8A* and *CD8B* in patients with MHE (Figure 3.7). $CD8^+$ T-cell deficiency and an increased $CD4^+/CD8^+$ ratio are features of many human chronic autoimmune diseases[156]; for instance, $CD8^+$ T-cell deficiency is an early and persistent feature of patients with multiple sclerosis[157]. In fact, the reduced total number of $CD8^+$ T cells in the blood of these patients has been attributed to $CD8^+$ T-cell sequestration in the brain[158]. Therefore, it is likely that early $CD8^+$ T-cell downregulation could also contribute to triggering the autoimmune shift towards peripheral inflammation and promotion of T-cell infiltration into the brain in patients with MHE.

Our data also indicated a general decrease in the 'antigen processing and presentation' pathway (Figure 3.7). Two key upstream genes in the MHC class I subpathway, tumor necrosis factor (*TNF*) and proteasome activator subunit (*PSME1*), were also significantly downregulated. *TNF* can activate *PSME1*, which in turn is implicated in immunoproteasome assembly and is required for efficient antigen processing[159]. Moreover, the MHC class II antigen presentation genes *HLA-DQA2* and *HLA-DRB5* were also significantly decreased. Together, these results describe an immune response signature in patients with MHE in which antigen processing and presentation were downregulated; the opposite regulation

patterns were found in T-cell populations, with Tc cells being downregulated and Th cells being upregulated.

More details regarding the changes in Th lymphocytes were found when analyzing the 'Th1 and Th2 cell differentiation' pathway (Figure 3.8). We found significant differences between Th1 and Th2 subtypes in cytokine receptors and products (left and right part of the pathway, respectively). For instance, the Th1-related genes *IFNGR, IL12R*, *IFNG,* and *IL12* were increased while Th2 showed decreased *IL2R* and *IL4R*, significantly increased *IL4*, and unaltered *IL13* expression.

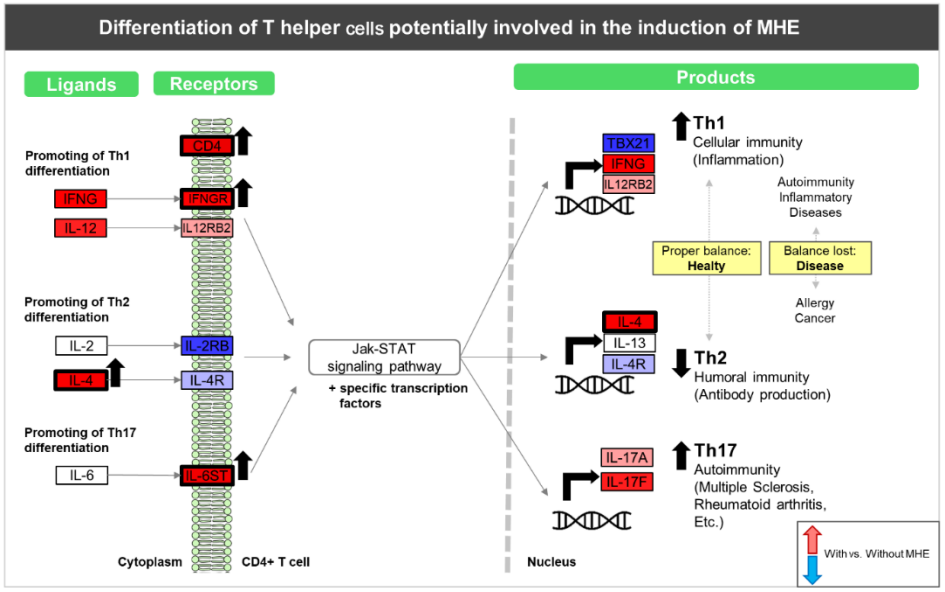


Figure 3.8 Pathway scheme that summarizes Th1, Th2, and Th17 differentiation pathways. Rectangles represent up- (red) or downregulation (blue) of genes in patients with versus without MHE. Black arrows and thicker black borders mark statistically significant genes (FDR < 0.05).

Th17 lymphocytes are the third most important Th subset for the regulation of immunity[160]. We found that the 'Th17 cell differentiation' pathway showed significant upregulation of the *IL6ST* receptor gene and

an increase in *IL17A* and *IL17F* products, suggesting the promotion of Th17 cell types in patients with MHE. These new results corroborate previous observations in which the activity of Th1 and Th17 cells tends to be more elevated in patients with MHE while the activity of Th2 cells tends to decrease[14].

We also found significant alterations in the expression of chemokine receptors (*CCR2* and *CXCR3*) and ligands (*CCL5, CXCL5, PF4 [CXCL4], PF4V1 [CXCL4V1],* and *XCL1*) in patients with MHE, which were downregulated in all cases except for *CCR2*. Chemokine receptor regulation on T cells is a complex mechanism that is dependent on both T-cell activation and the cell differentiation state[161]. For instance, Th1 pro-inflammatory cells preferentially express CCR5, CXCR3, and CXCR6, while the expression of CCR3, CCR4, and CCR8 is associated with Th2 anti-inflammatory cells[162–165]. Chemokine receptors have been proposed to play critical roles in neurological diseases like Alzheimer's[166,167] and multiple sclerosis[34]. In the latter case, CCR2 was found to be a critical modulator of the aberrant migration of peripheral T cells towards the site of inflammation. Thus, the changes in chemokine receptor expression we observed in patients with MHE reinforce the hypothesis of a readjustment of the population of immune cell types in these patients.

The results reported here also confirmed the inflammatory signature previously detected in MHE individuals and suggested cytokine signaling pathways as putative mechanisms for the decrease in Th2 cells and increase in Th1 and Th17 cells in patients with MHE. In a previous study, using isolated and cultured $CD4^+$ lymphocytes, we identified higher levels of IL-17, IL-21, IL-22, and TNFα, (a signature of factors released by Th17 and Th22 cells) in the culture medium[14]. This indeed suggests that

activation of these Th cell subtypes had increased and further validates the results of our analysis approach. Th17 cells are one of the major pathogenic Th cell populations underlying the development of many autoimmune diseases, including multiple sclerosis, rheumatoid arthritis, and psoriasis[168]. Therefore, the increase in Th17 would also likely contribute to promoting an autoimmune profile in patients with MHE. In turn, this shift towards autoimmune-like inflammation could trigger mild cognitive impairment.

Beyond immunological processes, lipid metabolism pathways also appeared to be significantly enriched in patients with MHE. Differentially expressed genes in this set of pathways were the upregulated *CD36, ABCA1* (both transporters), *PLPP3, NCEH1* (both enzymes), and the downregulated apolipoproteins *APOA1* and *APOB*. These changes pointed towards increased fatty acid digestion and absorption in the immune cells of patients with MHE. Altered lipid metabolism in immune system cells has also been previously reported in other diseases associated with cognitive impairment such as Alzheimer's disease[169] and mild cognitive impairment[170].

3.3.3. A coordinated metabolic and cytokinetic signature is present in patients with minimal hepatic encephalopathy

We identified 29 metabolites with significant differences (FDR < 0.05) between patients with and without MHE (Table 3.2). These increased metabolites included a methionine (Met), 3 SMs, 5 PCs, and an octadecenoylcarnitine (C18:2). Metabolites with reduced levels included 17 PCs and 2 lysoPCs, spermine, alpha-aminoadipic acid (alpha-AAA), and valine (Val). The most altered molecules in patients with MHE were

PCs and lysoPCs, which were mainly downregulated (negative logFC values). Of note, these phospholipids were previously found to be decreased in peripheral blood samples[171] and postmortem brain samples[172] from patients with Alzheimer's disease.

Table 3.2 Metabolites with significant alterations between patients with and without MHE.

ID Biocrates	logFC[a]	FDR[b]	ID Biocrates	logFC[a]	FDR[b]
PC aa C36:4	−0.7436	0.0005	**PC aa C38:5**	−0.6788	0.0009
PC aa C38:4	−0.7106	0.0006	**PC aa C38:3**	−0.5105	0.0165
PC aa C38:5	−0.6788	0.0009	**alpha-AAA**	−0.4547	0.0167
PC aa C40:5	−0.6272	0.0015	**PC aa C40:4**	−0.4155	0.0228
PC aa C34:4	−0.8185	0.0038	**PC aa C36:6**	−0.6148	0.0230
lysoPC a C20:4	−0.5669	0.0041	**lysoPC a C20:3**	−0.6249	0.0245
PC aa C40:6	−0.7801	0.0046	**PC aa C34:3**	−0.5725	0.0250
PC aa C38:6	−0.8598	0.0051	**Valine**	−0.4825	0.0296
PC ae C38:0	−0.5767	0.0055	**SM C22:3**	1.1125	0.0372
Methionine	0.6004	0.0063	**PC aa C32:3**	−0.3489	0.0379
PC aa C36:5	−0.8773	0.0067	**PC aa C42:1**	0.3616	0.0386
SM (OH) C14:1	0.5611	0.0075	**SM C16:0**	0.3501	0.0399
PC ae C44:6	0.6000	0.0137	**PC ae C30:0**	0.4426	0.0409
Spermine	−0.3971	0.0139	**PC ae C42:5**	0.5168	0.0424
PC aa C36:4	−0.7436	0.0005	**PC ae C44:5**	0.5536	0.0473
PC aa C38:4	−0.7106	0.0006	**C18:2**	0.4845	0.0477

Lipids are described with the notation 'Cx:y', where x denotes the number of carbons in the side chain and y denotes the number of double bonds. Abbreviations: PC, phosphatidylcholine; lysoPC, lysophosphatidylcholine; SM, sphingomyelin. [a]logFC: $\log_2$-fold-change; [b]*p*-value of the *t*-test calculated and adjusted by FDR using the **limma** R package.

Six cytokines (IL-15, CXCL13, CCL20, CX3CL1, IL-6, and IL-22) were identified with significantly different levels (FDR < 0.05) in patients with or without MHE, all of which presented increased levels in patients with MHE (Figure 3.9). Most cytokines showed the same direction of change at the mRNA and protein levels in patients with MHE (Table 3.3). Nonetheless,

given that the cytokine proteins were measured in plasma while mRNA was obtained from blood cells, we did not expect a perfect correlation between these 2 levels. Firstly, mRNAs for cytokines are subjected to post-transcriptional regulation by RNA binding proteins[173,174] or miRNAs[175]. In addition, plasma cytokines can be derived from sources other than blood cells. For instance, IL-15, CX3CL1, IL-17, and TNFα can come from the liver[176–178]. Furthermore, peripheral lymphocyte subtypes Th17 and Th22 can infiltrate the liver, resulting in IL-17 and TNFα production that may contribute to the upregulation of these cytokines in the plasma of patients with MHE[178].

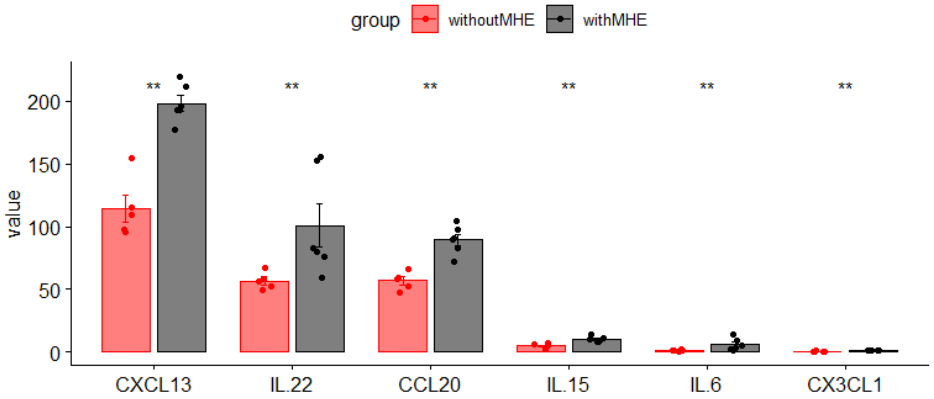


Figure 3.9 Serum cytokines with significant differences between patients with and without MHE. Differences between patients without MHE (red) and with MHE (dark grey) were tested using the Wilcoxon test with a threshold of an FDR < 0.05.

Table 3.3 Cytokine direction of change in patients with MHE at the mRNA and protein levels.

	Cytokines dataset			Microarray dataset			Coincidence
	FDR	logFC	Regulation	FDR	logFC	Regulation	
CCL20	0.02	0.63	UP	NA	NA	NA	?
CX3CL1	0.02	0.47	UP	0.23	−0.35	DOWN	NO

CXCL13	0.02	0.83	UP	NA	NA	NA	?
IL-10	0.13	0.35	UP	0.13	0.43	UP	YES
IL-12	0.4	0.18	UP	0.12	0.3	UP	YES
IL-13	0.77	−0.16	DOWN	NA	NA	NA	?
IL-15	0.02	1.06	UP	0.36	−0.2	DOWN	NO
IL-17	0.35	0.43	UP	0.93	−0.02	DOWN	NO
IL-18	0.28	0.31	UP	0.37	0.21	UP	YES
IL-22	0.02	0.53	UP	NA	NA	NA	?
IL-4	0.92	0	UP	0.01	0.66	UP	YES
IL-6	0.02	1.68	UP	0.89	0.04	UP	YES
TGFβ	0.09	−0.68	DOWN	0.27	−0.25	DOWN	YES
TNFα	0.12	0.75	UP	0.02	−0.45	DOWN	NO

The table shows the significance (FDR) and log fold-change (logFC) between patients with MHE and without MHE in both cytokine and transcriptomic datasets extracted from their respective statistical tests. Some cytokines were not represented in the Agilent microarray (shown as NA). Abbreviations: FDR, false discovery rate; NA, non-available data.

We then evaluated the co-abundance patterns of these differentially expressed plasma compounds across patients with cirrhosis. The data clustered into 6 distinct groups with highly correlated elements (Figure 3.10a) and had abundance profiles that distinguished between the groups with and without MHE (Figure 3.10b). Several correlation groups comprised chemically similar compounds, suggesting a functional significance for this co-variation pattern. For example, 1 relevant cluster was composed of the cytokines CCL20, CX3CL1, CXCL13, IL-15, IL-22, and IL-6 (module 1, shown in green in Figure 3.10), which have already been proposed as candidate modulators of lymphocyte infiltration into the brain and may contribute to cognitive impairment in MHE[14].

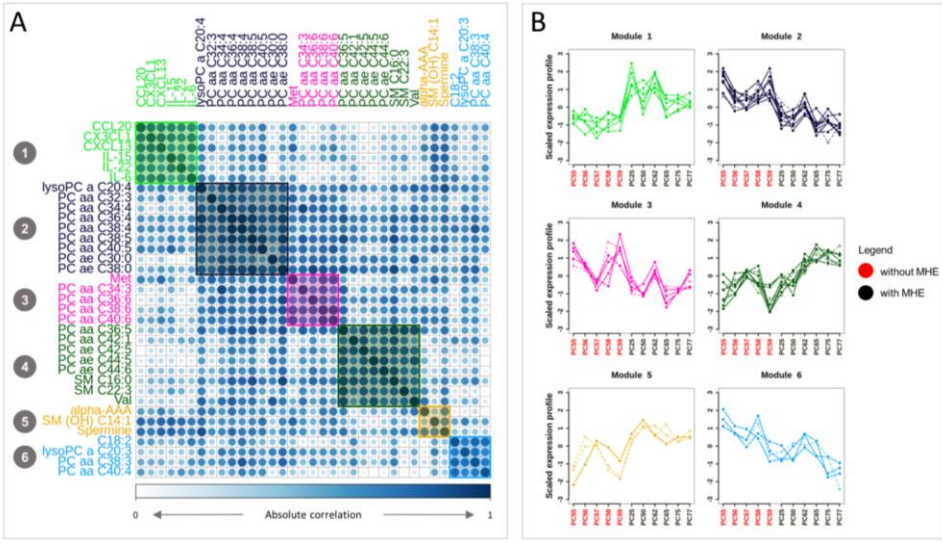


Figure 3.10 Metabolic and cytokine signatures present in patients with MHE. (A) Absolute correlation plot and clustering analysis of extracellular molecules. (B) Profiles of scaled compound levels across patients indicate the ability to distinguish between the groups with (red) and without (dark grey) MHE. Dashed lines represent compounds with negative correlations with the average block profile, which were set to positive for representation purposes.

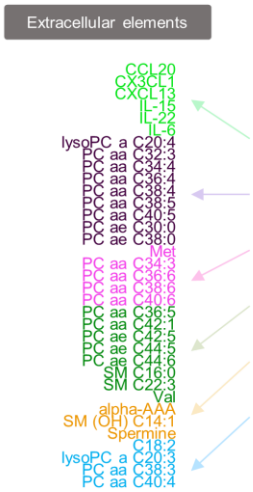
In addition, the phospholipids almost perfectly clustered according to their molecular masses (total number of carbon atoms) and saturation levels (total number of double bonds) in 3 main functional modules: (a) short and softly unsaturated lipids with ≤ 40C and ≤ 5 double bonds (purple module); (b) short and highly unsaturated lipids with ≤ 40C and > 5 double bonds (pink module); (c) long and highly unsaturated lipids with > 40C and ≥ 5 double bonds (dark green module). Interestingly, the relationship between phospholipids and inflammation has been extensively reported in the context of autoimmune diseases[179,180]. For instance, metabolomic measurements using the same technological platform as we used previously identified a panel of lipids that predict mild cognitive impairment or Alzheimer's disease[181] and 3 of them, PC aa C36:6, PC aa C38:6, and PC aa C40:6, were also significantly downregulated in patients with MHE.

During inflammation, or under conditions of oxidative stress, the polyunsaturated fatty acid side chains of the phospholipids in lipoproteins or cellular membranes can be oxidatively modified, generating structurally diverse oxidized phospholipid species and reactive aldehydes (e.g., MDA and hydroxynonenal [HNE])[182], each of which may exert both pro-inflammatory and anti-inflammatory effects. In previous studies we showed that patients with MHE had signs of increased oxidative stress in their blood samples compared to patients with cirrhosis without MHE. Furthermore, MDA, an indicator of oxidative damage to lipids, was also increased in patients with MHE and its presence correlated with psychometric test results[183]. Here we describe altered lipid metabolism in the context of mild cognitive impairment in MHE for the first time. The mechanisms by which these changes in lipids contribute to MHE could be similar to those involved in mild cognitive impairment[181] and Alzheimer's disease[184].

3.3.4. Multi-omic integration analysis highlights altered genes and associated pathways related with metabolites and cytokines in patients with minimal hepatic encephalopathy

The previous analyses identified transcriptional signatures in whole blood cells and a metabolic signature in serum that was associated with MHE. We then asked whether these signatures might be related by identifying gene expression signals that correlated with the extracellular compound changes linked to cognitive impairment. To answer this question, we established PLS multivariable regression models in which differentially expressed genes were taken as explanatory variables and each of the 6 metabolic modules were used as joint response variables to obtain a model per module. The accuracy and predictive ability of these models

were determined by their goodness of fit (R^2) and goodness of prediction (Q^2), respectively. All 6 PLS models had high (> 0.9) R^2 scores, indicating that they had a good explanatory values, and most of the models also had good (> 0.5) Q^2 predictive values (Figure 3.11).



INTEGRATIVE MODEL PARAMETERS

Functional module	Number of components	R^2	Q^2
1	4	0.953	0.436
2	4	0.92	0.599
3	4	0.936	0.403
4	4	0.948	0.315
5	3	0.947	0.597
6	4	0.913	0.337

Figure 3.11 Multi-omic integration analysis using partial least squares integrative models. R^2: goodness of fit. Q^2: goodness of prediction.

One interesting property of the PLS method was the direct interpretability of the loading values as a measure of the relative importance of the variables in the model. To further understand the biological significance of the models, gene set enrichment tests were run for genes ranked by the loading vectors of the first component, which provided the best separation between patients with and without MHE in each of the 6 models (Figure 3.12).

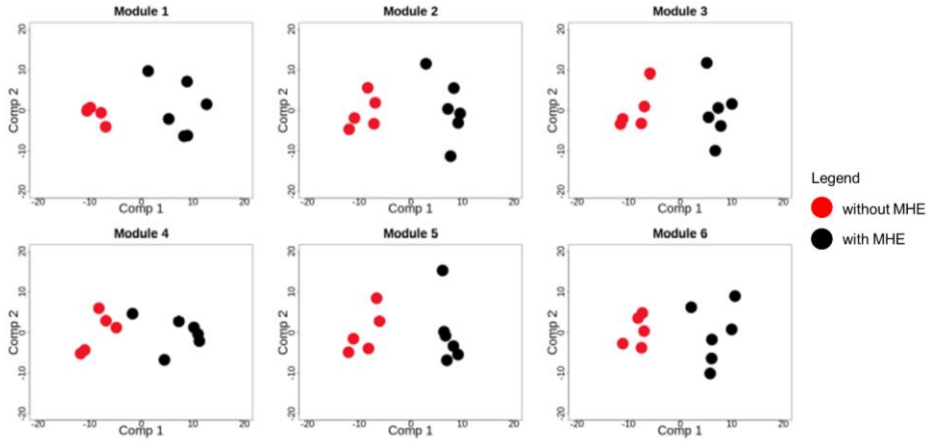


Figure 3.12 Score plots of the 6 partial least squares models. First component (Comp 1) achieved the best separation between patients with and without MHE in each of the 6 models. Red represents patients without MHE and dark grey represents patients with MHE.

The results were summarized in a multi-enrichment graph (Figure 3.13), where every node is a selected GO term from the enrichment analysis of the PLS loadings and the edges link them to the compound modules in which they were found to be significant. This representation locates module-specific pathways at the periphery of the network, while shared pathways are grouped towards the center. Overall, frequently shared pathways included ‘cell adhesion’, ‘immune response’, ‘proteolysis’, ‘ion transport’ and ‘adaptive immune response’, while module-specific pathways involved distinct signaling and metabolic pathways. Details of the significant genes within each enriched pathway are provided in Table 3.4.

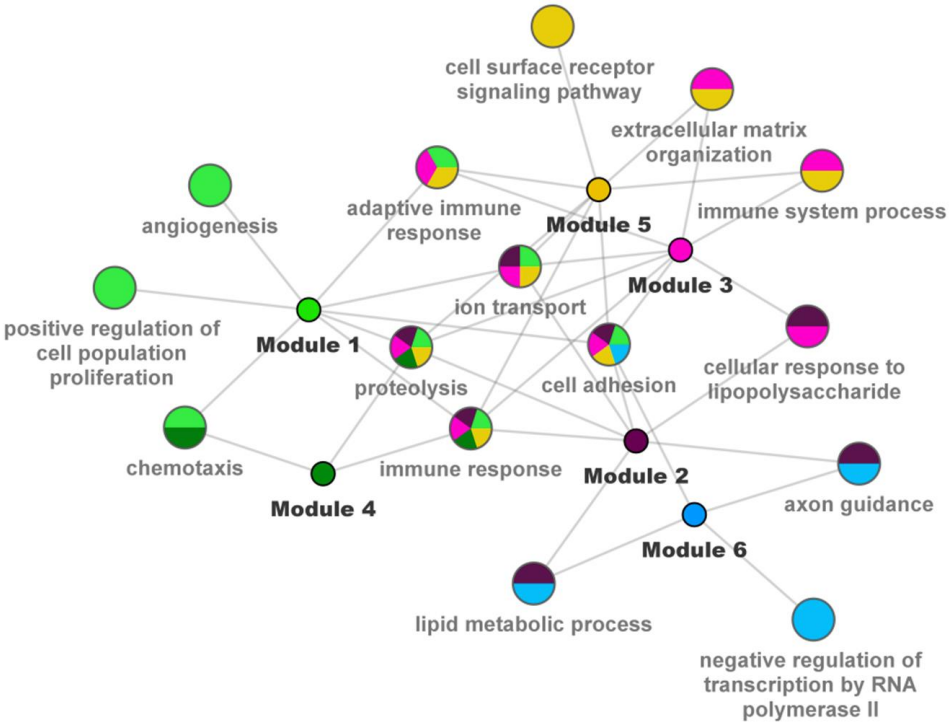


Figure 3.13 Multi-enrichment network. Every node is a significant GO biological process from the different enrichment analysis based on the partial least squares models. Edges denote links to the metabolite/cytokine modules in which these pathways were found to be significant.

Table 3.4 Altered genes in patients with MHE within the biological processes obtained from the multi-enrichment analysis.

	GO term	*p*-value	LOR	Number	Genes
Module 1: CCL20, CX3CL1, CXCL13, IL-15, IL-22, and IL-6					
1	proteolysis	0.003	0.012	13	CPA3, CASP10, HP, CFB, SPCS3, MMP14, ST14, BACE1, USP14, HPR, ADAM30, CPNE1, PCSK5
2	cell adhesion	0.010	0.013	20	VCAN, CYP1B1, SPP1, DCHS2, OLFM4, COL8A2, TNR, THBS1, COL13A1, CD4, SIGLEC12, ICAM5, CSTA, EPHA1, CD36, CNTN5, COL5A1, PCDH12, CXCR3, CLSTN2
3	immune response	0.013	−0.014	22	PF4V1, CMKLR1, SMAD6, IL4, CD8B, THBS1, CD274, CCL5, MS4A2, CD8A, CXCL5, CD4, TNF, XCL1, HLA-DRB5, LY86, PF4, IFITM2, CD36, CXCR3, HLA-DQA2, CCR2
4	adaptive immune response	0.019	−0.011	16	CLEC6A, CLEC10A, CD8B, CD274, MCOLN2, CD8A, CD4, ZNF683, LAG3, MICB, HLA-DRB5, HAVCR2, ANXA1, TNFRSF17, CD1C, HLA-DQA2
5	angiogenesis	0.023	0.010	13	ACVRL1, ANG, APOLD1, MEIS1, CYP1B1, COL8A2, EREG, FGFR2, MMP14, EPHA1, JAG1, FLT4, CXCR3
6	chemotaxis	0.027	−0.009	12	PF4V1, CMKLR1, CCL5, PDGFRB, CXCL5, HMGB2, MSMP, CMTM5, XCL1, PF4, CXCR3, CCR2
7	positive regulation of cell population proliferation	0.036	0.010	15	AREG, SAPCD2, AKR1C3, PRAMEF8, IL4, THBS1, EREG, PRAMEF12, PDGFB, PDGFRB, FGFR2, CXCL5, EPHA1, FLT4, IL6ST
8	ion transport	0.043	−0.008	12	SLC12A7, CATSPER1, COMMD3, MCOLN2, HFE, SCN4A, CACNG2, CACNB2, TRPM6, SLC13A4, SLC4A4, EFHB
Module 2: lysoPC a C20:4, PC aa C32:3, PC aa C34:4, PC aa C36:4, PC aa C38:4, PC aa C38:5, PC aa C40:5, PC ae C30:0, and PC ae C38:0					
	GO term	***p*-value**	**LOR**	**Number**	**Genes**
1	lipid metabolic process	0.008	−0.016	27	APOB, PLIN1, PLA2G4A, CYP1B1, DHRS9, RAB7A, SLC27A2, SC5D, TM7SF2, DHCR24, APOA1, ACOX2, NCEH1, ACSBG1, PLA1A, CPNE7, CYP2J2, MGST2, CD36, LRP10, CYB5R2, CPT1A, CPNE1, ABCA1, CERS4, ARV1, NR1H2
2	immune response	0.025	−0.012	22	PF4V1, CMKLR1, SMAD6, IL4, CD8B, THBS1, CD274, CCL5, MS4A2, CD8A, CXCL5, CD4, TNF,

					XCL1, HLA-DRB5, LY86, PF4, IFITM2, CD36, CXCR3, HLA-DQA2, CCR2
3	proteolysis	0.025	0.009	13	CPA3, CASP10, HP, CFB, SPCS3, MMP14, ST14, BACE1, USP14, HPR, ADAM30, CPNE1, PCSK5
4	cellular response to lipopolysaccharide	0.027	−0.010	15	PF4V1, CD274, NLRP7, CXCL5, HMGB2, TNF, HAVCR2, PF4, CD36, ASS1, ABCA1, NUGGC, STAR, NR1H2, LY96
5	cell adhesion	0.029	0.011	20	VCAN, CYP1B1, SPP1, DCHS2, OLFM4, COL8A2, TNR, THBS1, COL13A1, CD4, SIGLEC12, ICAM5, CSTA, EPHA1, CD36, CNTN5, COL5A1, PCDH12, CXCR3, CLSTN2
6	ion transport	0.047	−0.008	12	SLC12A7, CATSPER1, COMMD3, MCOLN2, HFE, SCN4A, CACNG2, CACNB2, TRPM6, SLC13A4, SLC4A4, EFHB
7	axon guidance	0.049	0.007	10	SEMA3F, PALLD, CYFIP1, TNR, GFRA2, EPHA1, SEMA4F, NGFR, KIF5B, BMP7
Module 3: Met, PC aa C34:3, PC aa C36:6, PC aa C38:6, and PC aa C40:6					
	GO term	***p*-value**	**LOR**	**Number**	**Genes**
1	immune response	0.012	−0.014	22	PF4V1, CMKLR1, SMAD6, IL4, CD8B, THBS1, CD274, CCL5, MS4A2, CD8A, CXCL5, CD4, TNF, XCL1, HLA-DRB5, LY86, PF4, IFITM2, CD36, CXCR3, HLA-DQA2, CCR2
2	cell adhesion	0.014	0.013	20	VCAN, CYP1B1, SPP1, DCHS2, OLFM4, COL8A2, TNR, THBS1, COL13A1, CD4, SIGLEC12, ICAM5, CSTA, EPHA1, CD36, CNTN5, COL5A1, PCDH12, CXCR3, CLSTN2
3	proteolysis	0.017	0.010	13	CPA3, CASP10, HP, CFB, SPCS3, MMP14, ST14, BACE1, USP14, HPR, ADAM30, CPNE1, PCSK5
4	extracellular matrix organization	0.021	0.010	13	VCAN, SPP1, COL8A2, TNR, THBS1, COL13A1, PDGFB, TNF, ICAM5, MMP14, SPARC, COL5A1, GFOD2
5	adaptive immune response	0.023	−0.011	16	CLEC6A, CLEC10A, CD8B, CD274, MCOLN2, CD8A, CD4, ZNF683, LAG3, MICB, HLA-DRB5, HAVCR2, ANXA1, TNFRSF17, CD1C, HLA-DQA2
6	ion transport	0.023	−0.009	12	SLC12A7, CATSPER1, COMMD3, MCOLN2, HFE, SCN4A, CACNG2, CACNB2, TRPM6, SLC13A4, SLC4A4, EFHB
7	cellular response to lipopolysaccharide	0.039	−0.009	15	PF4V1, CD274, NLRP7, CXCL5, HMGB2, TNF, HAVCR2, PF4, CD36, ASS1, ABCA1, NUGGC, STAR, NR1H2, LY96

8	immune system process	0.047	−0.013	32	CLEC6A, CLEC10A, HP, CD8B, VSIG4, CFB, CD274, MPEG1, MCOLN2, CD8A, KLRG1, HMGB2, CD4, PGLYRP1, FFAR2, IFIT3, ZNF683, LAG3, MICB, HLA-DRB5, HAVCR2, NLRC4, APOBEC3H, ANXA1, LY86, TNFRSF17, USP14, IFITM2, CD1C, APOBEC3G, HLA-DQA2, LY96

Module 4: PC aa C36:5, PC aa C42:1, PC ae C42:5, PC ae C44:5, PC ae C44:6, SM C16:0, and SM C22:3, Val

	GO term	*p*-value	LOR	Number	Genes
1	proteolysis	0.006	0.012	13	CPA3, CASP10, HP, CFB, SPCS3, MMP14, ST14, BACE1, USP14, HPR, ADAM30, CPNE1, PCSK5
2	immune response	0.008	−0.014	22	PF4V1, CMKLR1, SMAD6, IL4, CD8B, THBS1, CD274, CCL5, MS4A2, CD8A, CXCL5, CD4, TNF, XCL1, HLA-DRB5, LY86, PF4, IFITM2, CD36, CXCR3, HLA-DQA2, CCR2
3	chemotaxis	0.010	−0.011	12	PF4V1, CMKLR1, CCL5, PDGFRB, CXCL5, HMGB2, MSMP, CMTM5, XCL1, PF4, CXCR3, CCR2

Module 5: alpha-AAA, SM (OH) C14:1, and spermine

	GO term	*p*-value	LOR	Number	Genes
1	immune response	0.005	−0.015	22	PF4V1, CMKLR1, SMAD6, IL4, CD8B, THBS1, CD274, CCL5, MS4A2, CD8A, CXCL5, CD4, TNF, XCL1, HLA-DRB5, LY86, PF4, IFITM2, CD36, CXCR3, HLA-DQA2, CCR2
2	adaptive immune response	0.005	−0.013	16	CLEC6A, CLEC10A, CD8B, CD274, MCOLN2, CD8A, CD4, ZNF683, LAG3, MICB, HLA-DRB5, HAVCR2, ANXA1, TNFRSF17, CD1C, HLA-DQA2
3	cell adhesion	0.005	0.014	20	VCAN, CYP1B1, SPP1, DCHS2, OLFM4, COL8A2, TNR, THBS1, COL13A1, CD4, SIGLEC12, ICAM5, CSTA, EPHA1, CD36, CNTN5, COL5A1, PCDH12, CXCR3, CLSTN2
4	immune system process	0.006	−0.018	32	CLEC6A, CLEC10A, HP, CD8B, VSIG4, CFB, CD274, MPEG1, MCOLN2, CD8A, KLRG1, HMGB2, CD4, PGLYRP1, FFAR2, IFIT3, ZNF683, LAG3, MICB, HLA-DRB5, HAVCR2, NLRC4, APOBEC3H, ANXA1, LY86, TNFRSF17, USP14, IFITM2, CD1C, APOBEC3G, HLA-DQA2, LY96
5	proteolysis	0.009	0.011	13	CPA3, CASP10, HP, CFB, SPCS3, MMP14, ST14, BACE1, USP14, HPR, ADAM30, CPNE1, PCSK5
6	ion transport	0.014	−0.010	12	SLC12A7, CATSPER1, COMMD3, MCOLN2, HFE, SCN4A, CACNG2, CACNB2, TRPM6, SLC13A4, SLC4A4, EFHB
7	extracellular matrix	0.019	0.010	13	VCAN, SPP1, COL8A2, TNR, THBS1, COL13A1, PDGFB, TNF, ICAM5, MMP14, SPARC,

	organization				COL5A1, GFOD2
8	cell surface receptor signaling pathway	0.039	−0.010	16	CASP10, MERTK, CD274, MS4A2, CD8A, KLRG1, ASGR2, CLEC1A, CD4, LAG3, EPHA1, ANXA1, CD36, GALR2, CXCR3, LY96
Module 6: C18:2, lysoPC a C20:3, PC aa C38:3, and PC aa C40:4					
	GO term	***p*-value**	**LOR**	**Number**	**Genes**
1	lipid metabolic process	0.011	−0.015	27	APOB, PLIN1, PLA2G4A, CYP1B1, DHRS9, RAB7A, SLC27A2, SC5D, TM7SF2, DHCR24, APOA1, ACOX2, NCEH1, ACSBG1, PLA1A, CPNE7, CYP2J2, MGST2, CD36, LRP10, CYB5R2, CPT1A, CPNE1, ABCA1, CERS4, ARV1, NR1H2
2	axon guidance	0.029	0.008	10	SEMA3F, PALLD, CYFIP1, TNR, GFRA2, EPHA1, SEMA4F, NGFR, KIF5B, BMP7
3	negative regulation of transcription by RNA polymerase II	0.034	−0.010	15	LEFTY1, MSC, FGFR2, TNF, ZNF157, GATA2, DUSP26, NFATC4, SOX15, HBZ, CD36, RARB, NR1H2, TBX15, ZHX1
4	cell adhesion	0.045	0.010	20	VCAN, CYP1B1, SPP1, DCHS2, OLFM4, COL8A2, TNR, THBS1, COL13A1, CD4, SIGLEC12, ICAM5, CSTA, EPHA1, CD36, CNTN5, COL5A1, PCDH12, CXCR3, CLSTN2

The table shows the summary of the 6 final enrichment analyses (1 per module of cytokines/metabolites). The gene ontology (GO) biological processes are sorted by their *p*-values according to gene set enrichment analysis performed with **mdgsa** implemented in the R/Bioconductor package. The log odds ratio (LOR) indicates whether the biological process was mainly enriched in downregulated genes in patients with MHE (negative values) or upregulated genes (positive values). The last 2 columns summarize the number and HGNC (HUGO Gene Nomenclature Committee) symbols of altered genes in patients with MHE that enriched each pathway.

Interestingly, the model for module 1, which comprised the previously highlighted cytokines CCL20, CX3CL1, CXCL13, IL-15, IL-22, and IL-6, showed enrichment for 3 immune-related pathways including 'chemotaxis', 'adaptive immune response', and 'immune response'. This indicated a connection between the cellular immune response discussed above and the serum accumulation of inflammatory markers (Figure 3.13). Further inspection of the 'chemotaxis' GO term (Figure 3.14) revealed that this extracellular module of cytokines was related with the previously discussed chemokine receptors and ligands (*CCR2*, *CXCR3*, *CMKLR1*, *CCL5*, *CXCL5*, *PF4, PF4V1,* and *MSMP*).

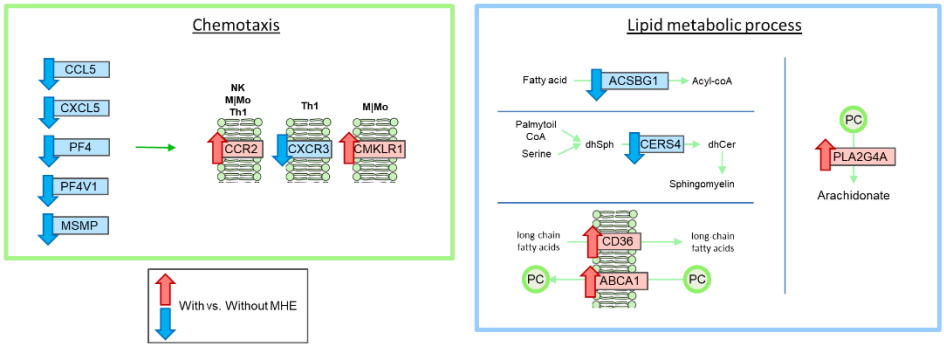


Figure 3.14 Biological interpretation of the multi-omic integrative analysis. The altered GO terms 'chemotaxis' and 'lipid metabolic process' represented at the gene (rectangles) and metabolite (circles) levels. Arrows indicate significant up- (red) or downregulation (blue) in patients with versus without MHE.

Many of these chemokine receptors are altered in patients with multiple sclerosis, who show elevated levels of CCR2, CCR5, and CXCR3[185]. Similarly, mouse models of inflammatory liver injury showed increased numbers of circulating CCR2-expressing monocytes that are attracted to the brain by activated microglia[186], while *in vitro* models of transmigration revealed that T cells from patients with multiple sclerosis exhibit an increased attraction to CCL3 and CCL5[187]. Moreover, the ligand *MSMP* acts as ligand for CCR2 and exhibits a chemotactic activity for monocytes

and lymphocytes[188], while the receptor *CMKLR1* plays a key role in directing plasmacytoid dendritic cell trafficking[189]. In a similar way, these results suggest that chemokines play an important role in patients with MHE in relation to trafficking of peripheral blood cells into the brain, as already shown in MHE patients[14].

Another interesting model was obtained for module 2, comprising PCs with ≤ 40 carbons and 0–5 double bonds in the fatty acid chain, that appeared to be correlated to genes in the 'lipid metabolic process', 'cellular response to lipopolysaccharide', and 'immune response' pathways, among others (Figure 3.13). The lipid metabolic process pathway was also shared by module 6 (which was composed of similar phospholipids in terms of length and saturation levels) and included annotated genes from different pathways. In particular, we identified the downregulation of *ACSBG1* and *CERS4* (Figure 3.14).

ACSBG1 catalyzes the conversion of very long-chain fatty acids into their active form (acyl-CoA) both for the biosynthesis and degradation of cellular lipids[190], while *CERS4* catalyzes the formation of sphingomyelins with high selectivity towards long and very-long chain fatty acids[191]. This result suggests that impairment of long-chain fatty acid activation and SM production may be ongoing in patients with MHE. Additionally, we found 2 significantly upregulated lipid transporters within this pathway: *ABCA1* (involved in transport of PCs from the cell into environment[192]) and *CD36* (responsible for internalization of long-chain fatty acids into cells[193]), pointing towards increased fatty acid transport in patients with MHE.

CD36 is also a scavenger receptor that cooperates with toll-like receptors in the internalization of oxidized phospholipid and activating the inflammasome in macrophages[194]. Moreover, the *PLA2G4A* gene was

also upregulated in patients with MHE. *PLA2G4A* selectively hydrolyzes PCs into arachidonic acid which is eventually formed into prostaglandins and is also implicated in the initiation of the inflammatory response[195].

Lipidomic screening in patients with Alzheimer's disease[171] identified 3 PCs that were significantly diminished. These authors hypothesized that altered phospholipase (PLA2) activity could lead to an increased PC metabolism and thereby, a subsequent decrease in their plasma levels. This forms part of the Lands cycle for the synthesis and degradation of PCs, in which plasma PCs are synthesized in the liver and secreted as components of lipoprotein particles. PCs are then hydrolyzed by PLA2 to produce lysoPC, which is rapidly cleared from circulation by transporters to the liver for the synthesis of PCs, closing the cycle[196].

In agreement with the above, our study showed an increase in *PLA2G4A* in the immune system cells from patients with MHE which may contribute to the decrease of PCs we observed. Of note, a differential phospholipid pattern was detected in the peripheral blood of patients with cirrhosis with or without MHE, as well as in patients with Alzheimer's[171] or Parkinson's disease[197].

In summary, our integrated model was able to link the extracellular information (metabolites/cytokines) with gene expression in human blood samples from patients with MHE. On the one hand, the results pointed towards the chemotactic function associated with a group of altered cytokines. On the other hand, 2 groups of PCs related to different aspects of lipid metabolic processes were identified that may contribute to the regulation of peripheral inflammation in patients with MHE.

3.4. Conclusion

This work identified immune system pathways that are potentially involved in the induction of MHE by integrating multiple molecular assays from blood samples. Firstly, we detected biological pathways that corroborated the observed decrease in Th2 cells and the increase in Th1 and Th17 cells in patients with MHE, while other pathways suggested B-cell upregulation and Tc downregulation. Lipid metabolism was described for the first time in the context of mild cognitive impairment in MHE. Secondly, specific sets of serum metabolites and cytokines were identified that pointed towards chemotactic function or structural patterns of lipids as mediators of MHE induction.

Finally, integrative modelling suggested a link between the cytokines CCL20, CX3CL1, CXCL13, IL-15, IL-22, and IL-6 and the alteration of chemotactic receptors (*CCR2*, *CXCR3*, and *CMKLR1*) and ligands (*CCL5*, *CXCL5*, *PF4*, *PF4V1,* and *MSMP*) in patients with MHE. This revealed a link between the production and reception of inflammatory signals operating in these patients. The PLS analysis also revealed that long-chain unsaturated PCs, increased fatty acid transport, and prostaglandin production are all strongly associated in samples from patients with MHE. This suggested a possible pathway for the dysregulation of structural phospholipids during this mild cognitive decline. Nonetheless, future work with a large cohort will be required to validate these results and to establish the utility of these molecular mechanisms as biomarkers of MHE.

Chapter 4

Identification of altered signaling pathways in CD4$^+$ lymphocytes isolated from patients with minimal hepatic encephalopathy

This chapter was adapted from the paper '*Gene expression and signaling pathways in lymphocytes of cirrhotic patients with cognitive impairment. Role of miRNAs'* by Teresa Rubio, Ana Conesa, Sonia Tarazona, Cristina Marti, Paula Izuierdo-Altarejos, María-Pilar Ballester, Juan José Gallego, Desamparados Escudero-García, Amparo Urios, Carmina Montoliu, Vicente Felipo. Submitted to Translational Research, 2021.

4.1. Introduction

Peripheral inflammation and hyperammonemia act synergistically to induce MHE[18]. All patients with cirrhosis show peripheral inflammation. However, it has been reported that the appearance of MHE is associated with a shift in the type of peripheral inflammation, with increased activation of all subtypes of $CD4^+$ T lymphocytes; increased $CD4^+CD28^-$ T lymphocytes, and intensified differentiation of $CD4^+$ T lymphocytes into Tfh and Th22 cells[14]. This aforementioned study suggested that changes in $CD4^+$ T lymphocytes are key in the shift in peripheral inflammation in patients with cirrhosis, thereby triggering the appearance of MHE.

However, the precise biological pathways that are altered in $CD4^+$ T lymphocytes to trigger this immune system shift and the emergence of neurological alterations are still poorly understood. Therefore, the aim of this current work was to shed light on the differences in gene expression in $CD4^+$ T cells in patients with cirrhosis with and without MHE as well as the possible regulatory mechanisms behind these differences. Here we studied 2 layers of transcriptional regulation: modulation of mRNA transcription by transcription factors (TFs) and modulation of mRNA stability or translation by miRNAs.

miRNAs are small non-coding RNA molecules that induce target mRNA silencing. miRNA biogenesis is a multi-step process in which the primary hairpin structure in the nucleus is reduced to a 21–25 nt double-stranded RNA in the cytoplasm. This miRNA–miRNA* complex is composed of 1 miRNA which acts as the functional arm and another complementary miRNA* arm which is normally degraded. The complex is loaded into an RNA induced silencing complex (RISC) which, in cooperation with

additional proteins, recognizes the target mRNA. Consequently, the target mRNA molecules are silenced through mRNA strand cleavage, destabilization by shortening of its poly(A) tail, or by reducing the efficiency of the mRNA translation into proteins by ribosomes[198].

Altered miRNA expression has been observed in several pathological situations including immunological diseases[199], liver diseases[200], and in different neurological diseases such as multiple sclerosis or Alzheimer's and Parkinson's diseases[201–203]. Common pathophysiological processes have been reported for several neurological diseases, leading to the identification of shared miRNA dysregulation[204]. A subset of these miRNAs (termed 'NeurimmiRs') may play a role in regulating aspects of both the neuronal and immune responses during neurological disease, perhaps making them new therapeutic targets[205].

Research into miRNAs is attracting considerable clinical interest because of the potential role of these molecules as diagnostic and prognostic biomarkers[206,207] or as therapeutic targets[208]. Indeed, NGS methods and the development of bioinformatics tools for the analysis of the resulting datasets[209] is now allowing the detailed characterization of both miRNA expression profiles (miRNA-seq) and their target genes (RNA-seq)[210].

The aim of the work described in this chapter was to identify differences in gene expression in $CD4^+$ T lymphocytes between patients with cirrhosis with and without MHE. In addition, the possible mechanisms by which these differences in mRNA levels appear and the signaling pathways affected were also examined. Here we isolated $CD4^+$ T lymphocytes from patients with cirrhosis with and without MHE and performed expression analysis to identify: (1) gene expression changes and associated biological pathways altered in $CD4^+$ T cells by analyzing mRNA levels

using RNA-seq; (2) changes in miRNA levels associated with MHE by employing miRNA-seq; (3) miRNAs or transcription factors which could be modulating the levels of key mRNAs; and (4) signaling pathways affected by the alterations in miRNAs and mRNAs that might play a relevant role in the shift in peripheral inflammation associated with the appearance of MHE.

4.2. Methods

4.2.1. Patients

Patients with liver cirrhosis were recruited from the outpatient clinics at the Hospital Clínico Universitario de Valencia and Hospital Arnau de Vilanova (Valencia, Spain). The diagnosis of cirrhosis was based on clinical, biochemical, and ultrasonographic data. Inclusion criteria were the presence of chronic liver cirrhosis and patient age of 18 years or more. The exclusion criteria were a diagnosis of overt HE or a history of overt HE, recent (< 6 months) alcohol intake, infections, antibiotic use, gastrointestinal bleeding (< 6 weeks), use of drugs affecting cognitive function (< 6 weeks), and the presence of hepatocellular carcinoma or neurological or psychiatric disorders. The study protocols were approved by the Scientific and Research Ethics Committees at the Hospital Clínico Universitario de Valencia and Hospital Arnau de Vilanova in Valencia, Spain and were in accordance with the ethical guidelines set out in the Helsinki Declaration.

4.2.2. Assessment of cognitive function using psychometric tests

Patients were evaluated using the PHES, a battery of 5 psychometric tests which is used as the gold standard for MHE diagnosis[8]. The PHES score was calculated, adjusting for age and educational level, using Spanish normality tables (www.redeh.org). Patients were classified as having MHE when the score was ≤ −4 points[4]. After performing the PHES, 6 patients without MHE and 6 patients with MHE were selected for inclusion in the study. All the patients were males except 1 female in the group of patients without MHE (PC130). The mean age in the group with MHE was 63 ± 4 years and in the group without MHE it was 60 ± 2 years, with no significant differences between the groups (*t*-test; *p*-value = 0.59).

4.2.3. Isolation of $CD4^+$ T lymphocytes

Blood samples were collected in BD Vacutainer® tubes (Becton, Dickinson and Company, Franklin Lakes, NJ, USA) containing EDTA. Peripheral blood mononuclear cells (PBMCs) were prepared from whole blood by centrifugation over a density gradient medium (Lymphoprep™, Palex Medical, SA), according to the manufacturer's instructions. We then purified $CD4^+$ T cells from 5×10^6 PBMCs by immunomagnetic negative selection using an EasySep™ Human $CD4^+$ T-cell Isolation Kit (Stemcell Technologies Inc.).

4.2.4. RNA extraction and sequencing

The total RNA, including small RNAs, was isolated from $CD4^+$ lymphocytes using an miRNeasy Micro Kit (Qiagen) according to the manufacturer's instructions. Sample extractions were sent to the

Genomics Unit at the Centre for Genomic Regulation (Barcelona, Spain) which used a HiSeq2500 sequencer with HiSeq Sequencing v4 Chemistry. Ultra-low input RNA sequencing was selected for RNA-seq with poly A selection, a reverse stranded-specific library, with paired-end short-reads of 125 bp; single-end reads of 50 bp were used for small RNA sequencing. The **HiSeq Control Software** 2.2.58 was employed by the sequencing service for both datasets.

4.2.5. RNA-seq analysis

Reads were trimmed with **Trimmomatic**[211] software (version 0.38) when the mean quality score was below 20 in a sliding window of 20 bp and was removed if the resulting length was less than 80 pb. Read alignment was performed with **STAR**[212] software (version 2.6.1c) with the default parameters and using the Gencode v29 human genome[213] as a reference. Next, alignment sequencing quality was checked with **Qualimap2**[214] software (version 2.2.1). Subsequently, the reads were counted to obtain the gene expression matrix with the htseq-count function in **HTSeq**[215] software (version 0.9.1), applying the intersection-strict mode. Conditional quantile normalization (CQN), implemented in the **cqn** R/Bioconductor package, was applied to remove technical variability in the RNA-seq data. The **cqn** algorithm combines robust generalized regression to remove systematic biases such as GC-content and quantile normalization to correct for global RNA-seq data distortions[216]. Finally, differential expression analysis was performed in the **edgeR** R/Bioconductor package[92] and pathway enrichment analysis was carried out using the **Paintomics**[104] web tool.

4.2.6. miRNA-seq analysis

Adapter sequences (AGATCGGAAGAGCACACGTCT) were removed with **cutadapt**[217] software (version 1.9.1), which was also used to trim low-quality ($q < 30$) ends from reads and to retain reads with a minimum length of 15 bp. Read alignment was performed with the same software and reference genome as for RNA-seq, but with specific parameters for small RNA-seq as follows: outFilterMatchNmin 16 (≥ 16 bp matched to the genome), outFilterMismatchNoverLmax 0.05 (number of mismatches ≤ 5% of the mapped length), outFilterScoreMinOverLread 0, outFilterMatchNminOverLread 0 (both options turned off), and alignIntronMax 1 (splicing switched off). **Qualimap2**[214] (version 2.2.1) was used to check the alignment with default parameters. The expression matrix was generated with **featureCounts**[218] using miRBase annotation[219] (release 22.1) and was normalized with **cqn**[216]. Finally, differential expression analysis was performed in the **edgeR** R/Bioconductor package[92].

4.2.7. Omic power analysis

The statistical power of each omic data type was evaluated with the **MultiPower** tool[146] as a quality control for the differential expression results. For RNA-seq and miRNA-seq data, the power values exceeded 0.7 when aiming to detect features, with relatively large changes and low variability (a Cohen's *d* of at least 1.3). Thus, we concluded that, despite the small sample size available, we had sufficient power to validate our results and proceed with the analysis.

4.2.8. Biological integration analysis

To biologically relate miRNA and mRNA, we obtained target genes for differentially expressed miRNAs from the functional annotation database included in **tappAS** software[220] that contains curated miRNA annotations and target predictions from at least 5 different tools. Only target genes for which we had obtained statistically significant differences were considered in this analysis. Additionally, we filtered out any pairs with the same fold-change tendency because opposite expression profiles between regulator and target genes were expected in the miRNA context.

To collect TF–mRNA relationships, we first extracted differentially expressed TFs from the RNA-seq analysis. The target genes of these differentially expressed TFs were then obtained from the **tftargets** R package available at github[221], selecting the ENCODE database[222] as the source of the TF–target gene annotations. Only differentially expressed target genes were included and both up- and downregulatory relationships between TFs and target genes were considered.

4.2.9. Data availability

The RNA-seq dataset used in this study is available in the GEO database repository, GSE184200, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE184200. The miRNA-seq dataset used in this study is available in the GEO database repository, GSE184201, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE184201.

4.3. Results and discussion

4.3.1. Data pre-processing

RNA-seq data contained > 50 million 150 bp reads which we processed using different software packages (as described in the Methods section). First, we performed read quality control and found that read trimming was required to remove the low quality read 3'-ends. Second, reads were aligned (mapped) against a reference genome to identify which ones belonged to each gene. After quality control of the mapping step, the genes were quantified counting the number of reads assigned to every gene. By repeating this process for each sample, we obtained an expression matrix with 58,676 genes and 12 samples. Hereinafter, when we mention 'raw data', we refer to this expression matrix.

Our exploratory analysis of RNA-seq raw data showed a high proportion of genes with low expression levels (Figure 4.1A) in all the quantified genes. Genes containing low count values across most of samples were removed and 12,352 genes were retained for subsequent analysis. After the low-count filtering, boxplots highlighted a slight variable data distribution between the samples (Figure 4.1C). PCA score plot of this dataset (Figure 4.1D) revealed an almost perfect separation between the patients with and without MHE by PC1 (35.97% explained variability) except for PC132 and PC134.

After applying CQN normalization, the data distribution was transformed (Figure 4.1E) to allow sample comparison in subsequent analyses but with very few alterations to the groupings between patients without and with MHE, as shown in the score plot (Figure 4.1C). Additionally, CQN normalization also corrected the GC bias frequently produced when using RNA-seq platforms, although in our case this effect was minimal.

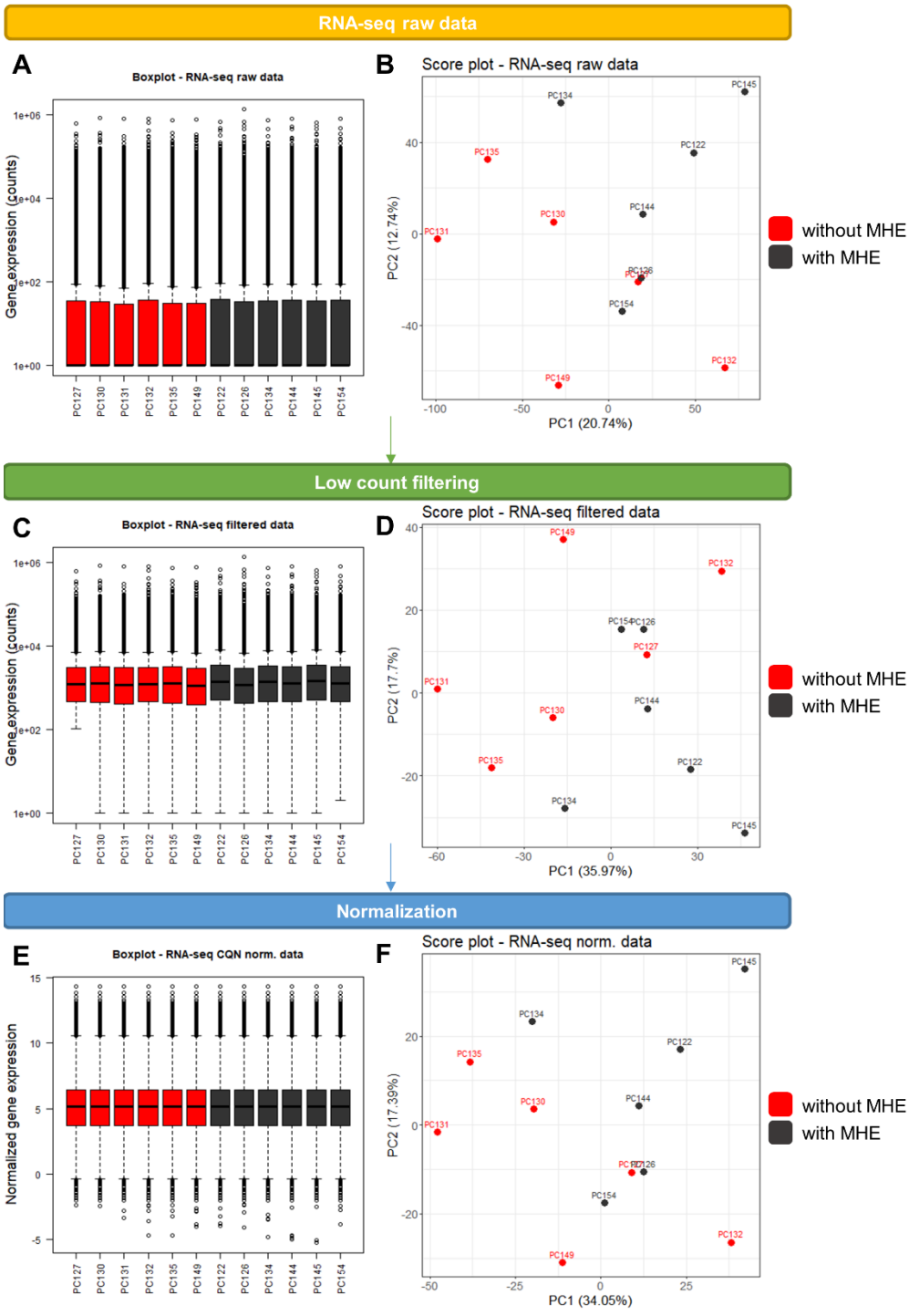

RNA-seq raw data
A
Boxplot - RNA-seq raw data
Gene_expression (counts)
1e+06
1e+04
1e+02
1e+00
PC127 PC130 PC131 PC132 PC135 PC149 PC122 PC126 PC134 PC144 PC145 PC154
B
Score plot - RNA-seq raw data
PC2 (12.74%)
PC1 (20.74%)
without MHE
with MHE
Low count filtering
C
Boxplot - RNA-seq filtered data
D
Score plot - RNA-seq filtered data
PC2 (17.7%)
PC1 (35.97%)
Normalization
E
Boxplot - RNA-seq CQN norm. data
Normalized gene expression
F
Score plot - RNA-seq norm. data
PC2 (17.39%)
PC1 (34.05%)

Figure 4.1 RNA-seq exploratory analysis. Boxplots and PCA score plots of RNA-seq raw data (A, B), after low-count filtering (C, D), and after normalization (E, F), from patients with cirrhosis without MHE (red) and with MHE (dark grey). For visualization purposes, the data were log-transformed in panels A, B, C and D. The percentages shown in parenthesis for the PCA plot axes indicate the amount of explained variability. Abbreviations: PC1 and PC2, principal components 1 and 2, respectively.

miRNA-seq data also contained > 50 million reads but with lengths of 50 bp. Illumina adapter content was required for sequencing but had to be removed for quantification, as described in the Methods section. After adapter removal and 3'-end trimming, the quality of the reads was optimal for alignment against the reference genome. miRNA quantification was performed for each sample to obtain an expression matrix with 2,652 miRNAs and 12 samples. Boxplots showed a high content in low-count miRNAs (Figure 4.2A) and PCA score plots separated well the 2 groups of patients (Figure 4.2B), except for the same 2 patients highlighted in the RNA-seq dataset.

miRNAs with low counts were filtered out to retain 395 miRNAs for subsequent analysis, with very variable data distribution between the samples, as shown in the boxplots (Figure 4.2C). The PCA score plot showed that the group separation remained after the low-count filtering (Figure 4.2D). As in the case of RNA-seq, the application of CQN normalization transformed the data distribution to make the samples comparable (Figure 4.2E) while leaving the grouping between patients without and with MHE almost unchanged, as shown in the corresponding score plot (Figure 4.2F). Additionally, as previously mentioned, CQN normalization also corrected the GC bias in the miRNA-seq dataset with a minimal effect.

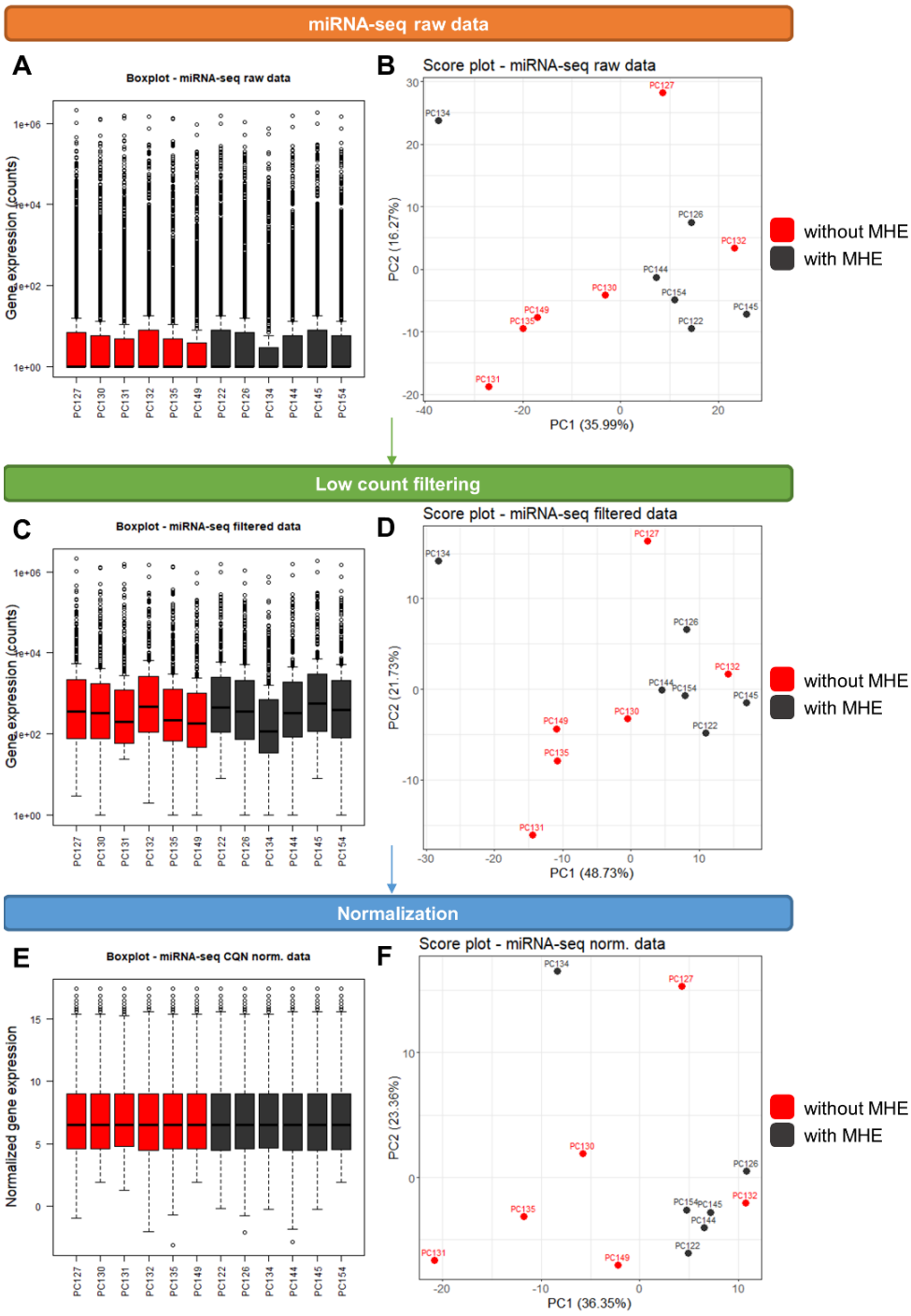

miRNA-seq raw data
A
Boxplot - miRNA-seq raw data
Gene expression (counts)
1e+06
1e+04
1e+02
1e+00
PC127
PC130
PC131
PC132
PC135
PC149
PC122
PC126
PC134
PC144
PC145
PC154
B
Score plot - miRNA-seq raw data
PC2 (16.27%)
PC1 (35.99%)
without MHE
with MHE
Low count filtering
C
Boxplot - miRNA-seq filtered data
D
Score plot - miRNA-seq filtered data
PC2 (21.73%)
PC1 (48.73%)
without MHE
with MHE
Normalization
E
Boxplot - miRNA-seq CQN norm. data
Normalized gene expression
F
Score plot - miRNA-seq norm. data
PC2 (23.36%)
PC1 (36.35%)
without MHE
with MHE

Figure 4.2 miRNA-seq exploratory analysis. Boxplot and PCA score plots of the miRNA-seq raw data (A, B), after low-count filtering (C, D), and after normalization (E, F), from patients with cirrhosis without MHE (red) and with MHE (dark grey). For visualization purposes, the data were log-transformed in panels A, B, C and D. The percentages shown in parenthesis for the PCA plot axes indicate the amount of explained variability. Abbreviations: PC1 and PC2, principal components 1 and 2, respectively.

4.3.2. CD4$^+$ lymphocytes from patients with minimal hepatic encephalopathy showed alterations in 167 mRNA and 20 signaling pathways compared to patients without minimal hepatic encephalopathy

Differential gene expression analysis comparing patients with cirrhosis with and without MHE revealed 167 genes with significant expression changes in CD4$^+$ lymphocytes (p-value < 0.05; 2-fold change; Figure 4.3). Some biological insights could be extracted from this information by looking at the 10 genes with the highest fold-changes: *ADGRG7, IGHV2-70D, RGS16, HES1, HLA-DRB5, IGLV4-69, PDK4, NLRC4, FPR2,* and *HSPA6*.

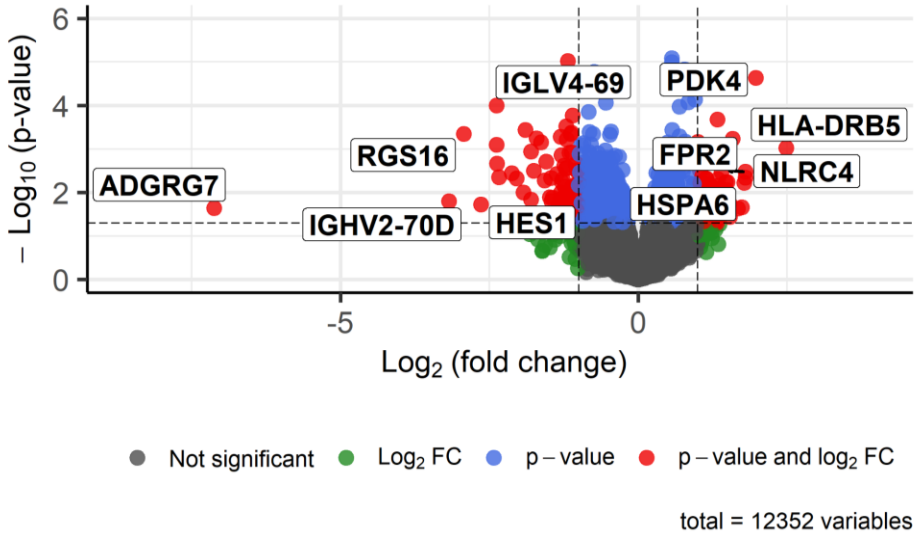


Figure 4.3 Differential gene expression analysis of the RNA-seq dataset comparing patients with cirrhosis with and without MHE. The volcano plot shows the statistical significance (logarithmically transformed p-values) versus

the ratio between groups of patients (logarithmically transformed fold-changes). Red dots represent significantly altered genes (p-value < 0.05 and $|\log_2$ fold-change$|$ > 1); genes out of our significance criteria were colored in blue, green, or grey. The top 10 differentially expressed genes were labeled using the nomenclature from the HUGO Gene Nomenclature Committee (HGNC).

Of these, the decreased *IGHV2-70D* (immunoglobulin heavy variable 2-70D) and *IGLV4-69* (immunoglobulin lambda variable 4-69) and the increased *HLA-DRB5* (major histocompatibility complex, class II, DR beta 5) genes were involved in immune system function. The highly downregulated *ADGRG7* (adhesion G protein-coupled receptor G7) has been previously linked to lymphoma[223] and its deletion produces body weight loss and increased intestinal contraction frequency[224], however, its physiological role and regulation have not yet been elucidated.

RGS16 (regulator of G protein signaling 16) can regulate T-lymphocyte activation and migration in response to inflammatory stimuli[225]. *HES1* (hes family BHLH transcription factor 1) plays a crucial role in controlling and regulating T-lymphocyte progenitors[226]. While *NLRC4* (NOD-like receptor C4) and *FPR2* (formyl peptide receptor 2) are key components in inflammation and antibacterial host defense[227,228]. The remaining genes, *PDK4* (pyruvate dehydrogenase kinase 4) and *HSPA6* (heat shock protein family A member 6) are a kinase that regulates glucose and fatty acid metabolism and a chaperone implicated in a wide variety of cellular processes, respectively.

Enrichment analysis to jointly look at the biological functions of all the differentially expressed genes identified 20 altered pathways (p-value < 0.05; Figure 4.4) which were mainly related to the immune system. In particular, the ‘IL-17 signaling pathway’ and ‘toll-like receptor signaling pathway’ processes were directly linked to $CD4^+$ T-cell activity. Other affected pathways were ‘histidine metabolism’ and ‘tryptophan

metabolism', which also modulate the immune system to produce intermediates that act as modulators of peripheral inflammation[61,62].

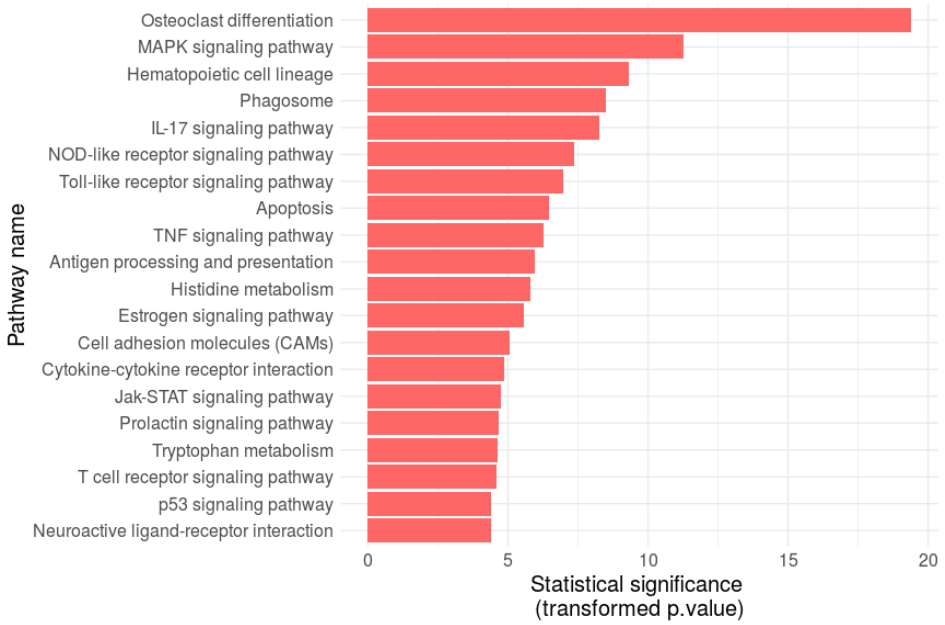
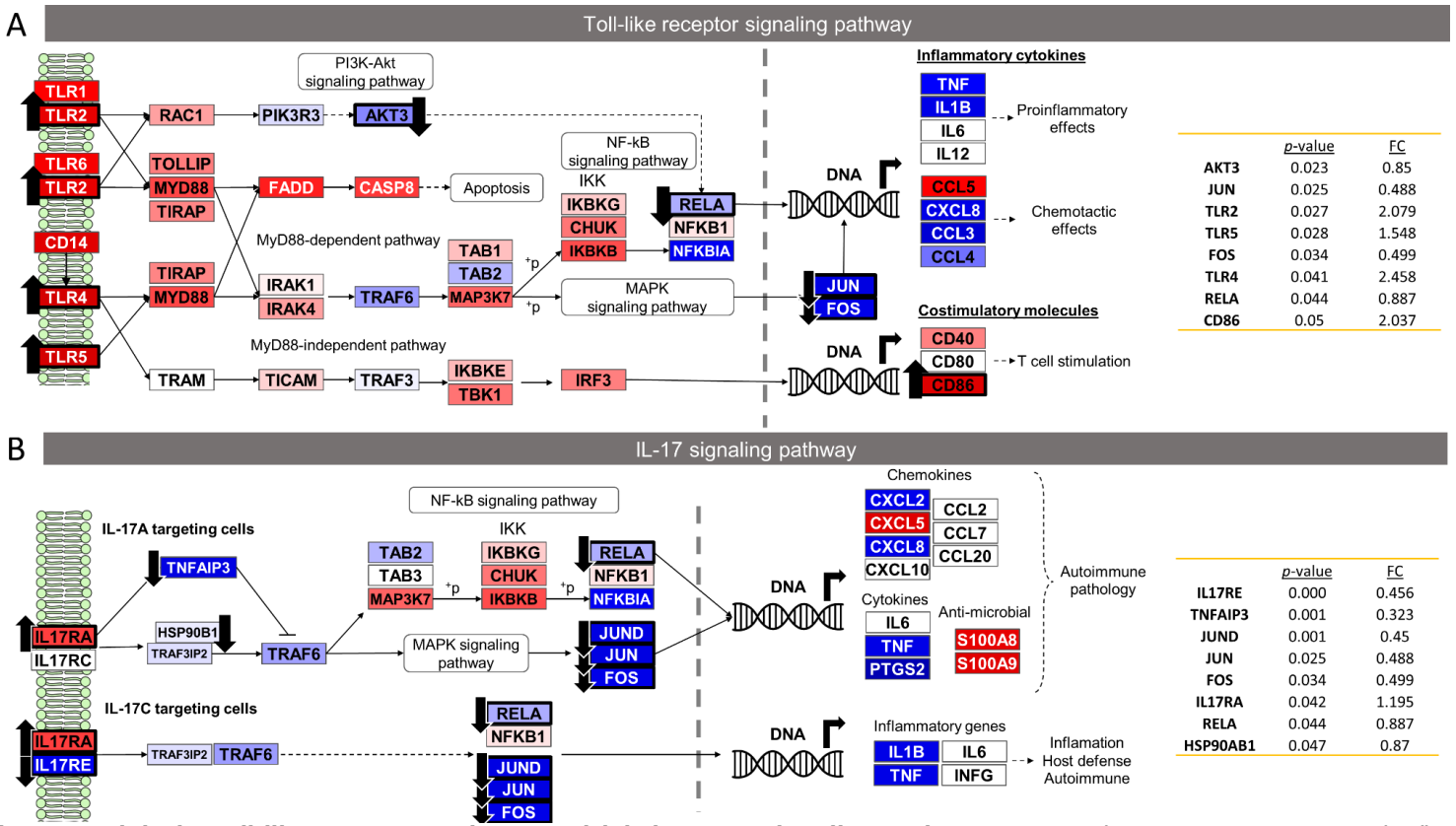


Figure 4.4 RNA-seq enrichment analysis identified 20 signaling pathways that were altered in CD4$^+$ lymphocytes from patients with MHE. The graph shows the KEGG pathways that were significantly (*p*-value < 0.05) altered in CD4$^+$ lymphocytes from patients with MHE, as identified by transcriptomic enrichment analysis. The *p*-values were mathematically transformed in this representation by applying the formula $1 - log10(p)$ to improve the interpretability of the data (the higher value, the higher the statistical significance).

Regarding the altered genes included in the 'toll-like receptor signaling pathway', TLR2 and TLR4 mRNAs were significantly increased and TLR1, TLR5, and TLR6 mRNAs also tended to increase (Figure 4.5). Toll-like receptors 2 and 4 (TLR-2 and TLR-4), which were upregulated, could play a role in the shift in peripheral inflammation in patients with cirrhosis towards an 'autoimmune-like' type of inflammation, which would then trigger MHE[14,229]. Indeed, TLR2 and TLR4 have already been proposed as prognostic markers for autoimmune pathologies including Graves' disease[230].



	p-value	FC
AKT3	0.023	0.85
JUN	0.025	0.488
TLR2	0.027	2.079
TLR5	0.028	1.548
FOS	0.034	0.499
TLR4	0.041	2.458
RELA	0.044	0.887
CD86	0.05	2.037

	p-value	FC
IL17RE	0.000	0.456
TNFAIP3	0.001	0.323
JUND	0.001	0.45
JUN	0.025	0.488
FOS	0.034	0.499
IL17RA	0.042	1.195
RELA	0.044	0.887
HSP90AB1	0.047	0.87

Figure 4.5 (A) The toll-like receptor pathway and (B) the IL-17 signaling pathway. Rectangles represent gene up- (red) or downregulation (blue) in patients with MHE versus those without MHE. The black arrows and thicker black borders mark genes with statistically significant changes (*p*-value < 0.05 and 2-fold increases/decreases). Tables of *p*-values and fold-changes (FCs) are included to show genes with *p*-value < 0.05 which exceeded our significance cutoff but were still relevant to MHE.

The expression of these 2 TLRs is also increased in multiple sclerosis, where resident cells and leukocytes that have infiltrated into the central nervous system exhibit high levels of expression of TLR-2 and its ligands[231,232]. TLR-2 may contribute to the development of multiple sclerosis by reinforcing Th1/Th17 cell-related responses and/or by downregulating regulatory T cells[233]. Furthermore, depletion of TLR-2 in murine multiple sclerosis models alleviated disease susceptibility and reduced $CD4^+$ cell infiltration into the central nervous system[234]. In addition, TLR-4 regulates Th17 differentiation in multiple sclerosis and has been proposed as a therapeutic target for Th17-mediated neurodegeneration in neuroinflammatory diseases[235].

Taken together, these data suggest that the increase in TLR2 and TLR4 in $CD4^+$ T lymphocytes in patients with MHE could contribute to the shift towards auto-immune-type peripheral inflammation, the associated activation of Th17 $CD4^+$ lymphocytes[14], and promotion of the infiltration of peripheral Th17 and Tfh lymphocytes into the brain which would then trigger cognitive impairment[15].

Following ligand binding, activation of toll-like receptors initiates a series of signaling cascades (e.g., MAPK and PI3K-Akt) that lead to the translocation of NF-kB, FOS, and JUN transcription factors into the nucleus where they promote the transcription of genes for pro-inflammatory cytokines such as TNFα, IL-1β, CXCL8, CCL5, CCL3L, and CCL4[236,237]. The levels of mRNAs for these cytokines tended to decrease in $CD4^+$ T lymphocytes from patients with MHE, except for *CCL5,* whose expression increased. Another significantly upregulated mRNA coded for CD86, a protein with costimulatory properties on effector memory T lymphocytes[238].

The 'IL-17 signaling pathway' gene expression was also altered in patients with MHE. The interleukin-17 superfamily consists of 6 ligands (from IL-17A to IL-17F), although IL-17A is the principal ligand and is mainly produced by Th17 lymphocytes[239]. This increase would also contribute to the enhanced response and activation of Th17 lymphocytes reported in patients with MHE and to the infiltration of these cells into the brain[14,15]. Although *IL17RA* receptor gene expression was increased in patients with MHE, intermediates like *TRAF6* tended to decrease and *NF-kB*, *FOS,* and *JUN* were downregulated. This was associated with a tendency towards the decrease in mRNAs coding for cytokines such as *CXCL2*, *CXCL8*, *TNF*, and *PTGS2* and an increase in *CXCL5*, *S100A8,* and *S100A9* (Figure 4.5). In the case of IL-17C, the subunit receptor IL17RE and subsequent intermediates (JUND, FOS, and JUN) were significantly downregulated and this was associated with a tendency towards decreased expression of the mRNAs for IL-1β and TNFα.

Taken together, this analysis identified the alteration of 2 signaling pathways that were relevant for $CD4^+$ lymphocyte function in patients with MHE. It is noteworthy that the initial elements of both for the toll-like receptor and IL-17 signaling pathways (the mRNAs for TLR2, TLR4, or IL17RA) were upregulated (Figure 4.5), while the intermediate products of the pathways were downregulated (RELA, JUN, and FOS) and the final products tended to be downregulated. This occurred for the mRNAs for TNFα, IL-1β, CXCL8, CCL3, and CCL4 in the TLR pathway and TNFα, IL-1β, CXCL2, CXCL8, and PTGS2 in the IL-17 signaling pathway. The role of miRNAs in this paradoxical regulation is discussed below.

We also identified 2 metabolic pathways that were altered in $CD4^+$ lymphocytes from patients with MHE that may contribute to alterations in the immune response. Firstly, the 'histidine metabolism' pathway showed

increased expression of mRNAs for enzymes involved in the first reactions of histidine catabolism (Figure 4.6). L-Histidine is transformed into histamine via increased histidine decarboxylase (*HDC*), with upregulated histamine N-methyltransferase (*HNMT*) later adding a methyl group to histamine, while aldehyde dehydrogenase (*ALDH2*, which was also upregulated) transforms it into imidazole-4-acetate. In addition, L-Histidine is also metabolized into urocanate by upregulated histidine ammonia-lyase (*HAL*). Histamine favors polarization of the immune response towards a Th1 profile[240] and induces IL-17 production in Th17 cells in patients with atopic dermatitis or psoriasis[62]. The data reported here suggest that upregulation of histamine production in $CD4^+$ T cells in patients with MHE could contribute to their increased Th1 and Th17 responses[14,229].

.

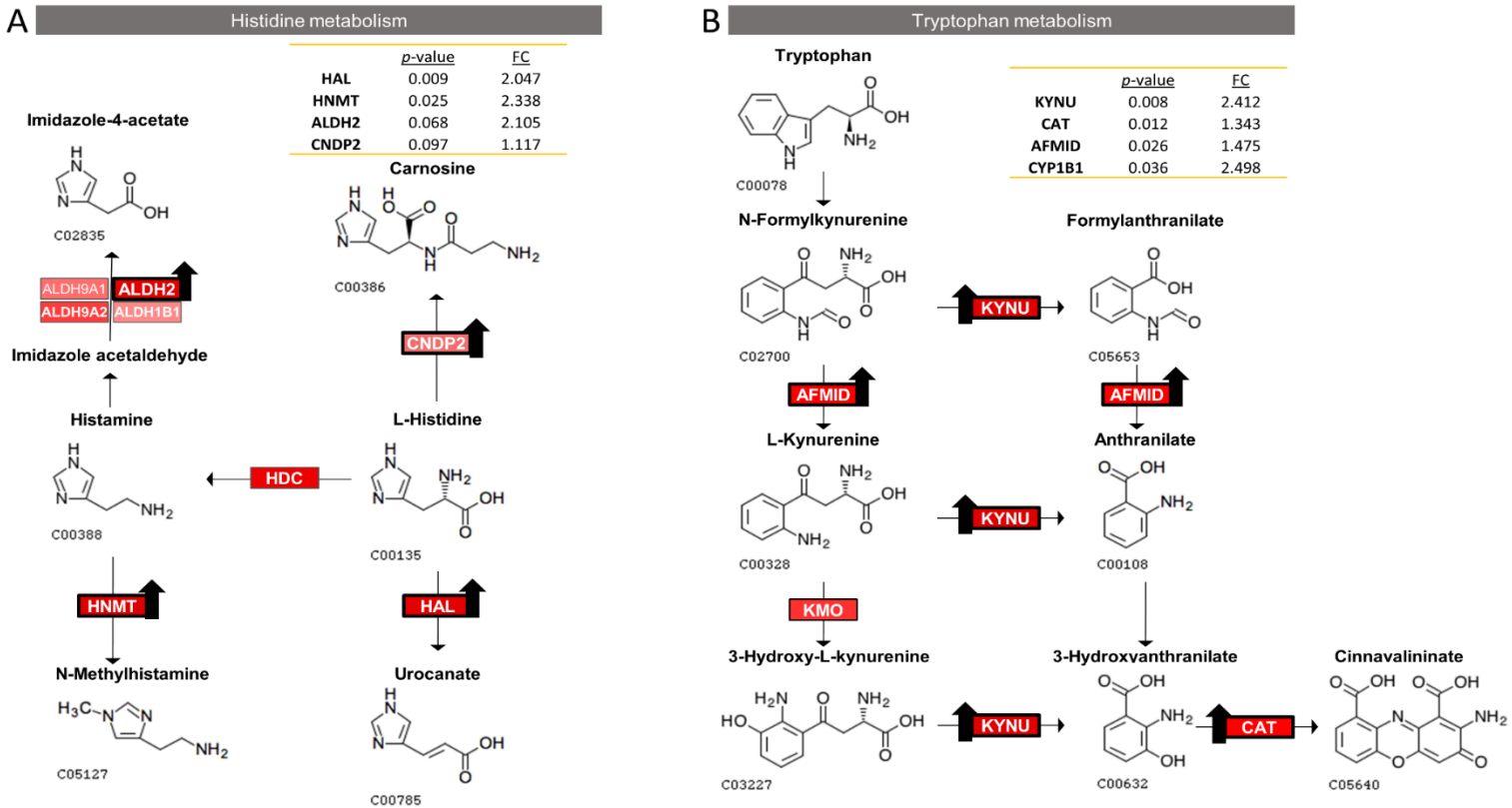


Figure 4.6 (A) Histidine metabolism pathway and (B) the tryptophan metabolism pathway. Rectangles represent gene up- (red) in patients with MHE versus those without MHE. The black arrows and thicker black borders mark genes with statistically significant changes (*p*-value < 0.05 and 2-fold increases). The *p*-value and fold-change (FC) tables were included to show genes with *p*-value < 0.05 which exceeded our significance cutoff but were still relevant to MHE.

Secondly, the ‘tryptophan metabolism’ pathway was also altered in patients with MHE (Figure 4.6), with increased levels of the mRNAs for proteins that produce L-kynurenine via arylformamidase (*AFMID*) and 3-hydroxy-L-kynurenine via kynurenine 3-monooxygenase (*KMO*). Both molecules are transformed into anthranilate and 3-hydroxy-anthranilate, respectively, by the upregulated kynureninase (*KYNU*). Tryptophan metabolism dysregulation, which shifts catabolic routes towards oxidative breakdown along the kynurenine axis is inextricably linked with immune activation[241].

Mezrich et al.[242] showed that kynurenine activates AHR, a Th22-inducing transcription factor. Kynurenines induce AHR translocation to the nucleus and promote IL-10 and IL-22 expression[61]. It is likely that enhanced production of kynurenine in $CD4^+$ T lymphocytes in patients with MHE contributes to the increased differentiation to Th22 cells and production of IL-22 reported by Mangas-Losada et al.[14], as well as the reduced cognitive performance in patients with MHE. Indeed, an association between increased levels of kynurenine and poorer cognitive performance has been reported in several studies[243–245]. Thus, the use of kynurenine pathway enzyme inhibition has been proposed as a potential therapeutic strategy for treating neurological diseases[246]. These data suggest that the upregulation of enzymes producing kynurenine and its derivatives in $CD4^+$ T lymphocytes from patients with MHE could contribute to the immune shift and appearance of cognitive impairment in these patients.

4.3.3. CD4$^+$ lymphocytes showed altered levels for 13 miRNAs and 7 transcription factors comparing patients with versus without minimal hepatic encephalopathy

Differential expression analysis of miRNA-seq detected 13 miRNAs whose levels were significantly different between patients with cirrhosis with or without MHE (*p*-value < 0.05; Table 4.1). Most of these miRNAs were upregulated in patients with MHE while only 3 (miR-1246, miR-4999-5p, and miR-4521) were downregulated. Additionally, 7 transcription factors showed significant differences in their gene expression levels. In this case, only 1 transcription factor was upregulated (*GATA2*) while the rest were downregulated.

Table 4.1 miRNAs and transcription factors with altered levels in CD4$^+$ lymphocytes from patients with MHE.

micro-RNAs			
	log CPM	**fold change**	***p*-value**
miR-1246	1.729	0.501	0.015
miR-127-5p	5.460	2.641	0.019
miR-4999-5p	1.996	0.498	0.020
miR-130b-3p	5.250	1.896	0.022
miR-656-3p	5.640	2.512	0.023
miR-4645-3p	3.330	1.602	0.024
miR-494-3p	8.866	2.291	0.034
miR-4433b-5p	9.020	2.213	0.035
miR-539-3p	6.191	2.244	0.044
miR-130a-3p	6.226	2.285	0.047
miR-33a-3p	2.530	1.989	0.049
miR-335-5p	8.380	1.862	0.049
miR-4521	6.150	0.461	0.049

Transcription Factors			
	log CPM	**fold change**	***p*-value**
JUND	8.688	0.450	0.001
BHLHE40	7.411	0.387	0.004
FOS	10.486	0.499	0.034
JUNB	10.680	0.474	0.008
JUN	9.683	0.488	0.025
GATA2	3.445	2.131	0.037
MAFF	4.011	0.416	0.005

A list of miRNAs and transcription factors whose levels were significantly (*p*-value < 0.05) altered in $CD4^+$ lymphocytes from patients with MHE. miRNA annotation was derived from the miRBase database. The level quantification is shown as the logarithmic value of counts per million (CPM). The relative differences between the groups are presented as the fold-change scores.

4.3.4. Identification of miRNAs and transcription factors that modulate the differential expression of key genes involved in regulating immune responses in minimal hepatic encephalopathy

A main goal of this study was to unveil the regulatory mechanisms involved in gene expression changes in $CD4^+$ T cells from patients with MHE. To identify these mechanisms, we searched public databases for evidence of binding between the detected miRNAs and TFs and the genes that were differentially expressed in patients with MHE (as described in the Methods section). A total of 16 miRNA–mRNA associations involving the altered genes were found (Figure 4.7) in which 4 miRNAs (miR-130a/b-3p, miR-656-3p, miR-335-5p, and miR-494-3p) participated. According to the annotation database, these 4 miRNAs modulated the levels of 14 mRNAs and 3 transcription factors (Figure 4.7).

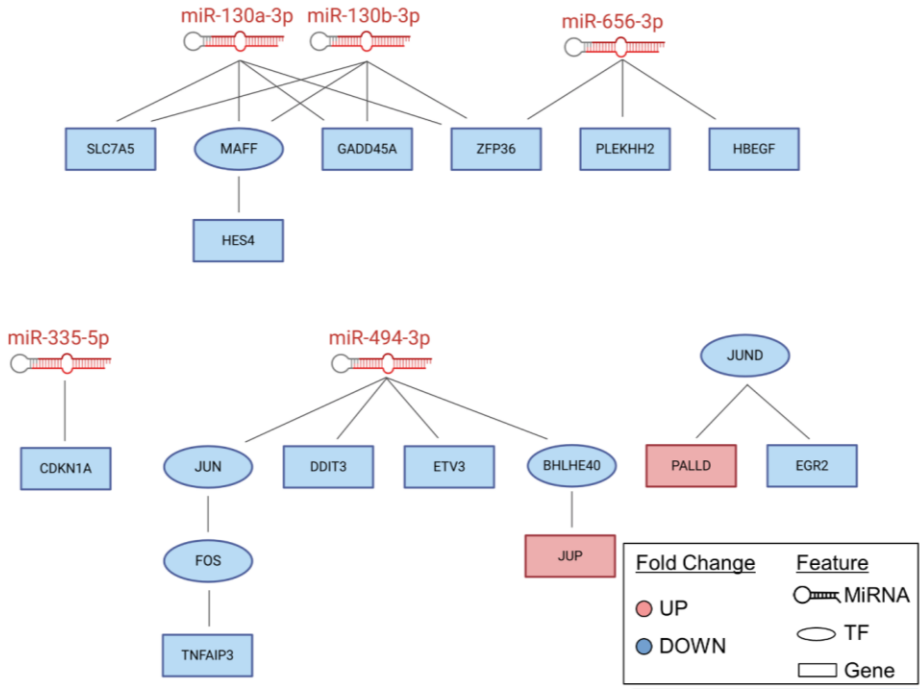


Figure 4.7 miRNAs and transcription factors modulating the differential expression of key genes involved in the regulation of immune responses in patients with MHE. The colors indicate upregulation (red) or downregulation (blue) in $CD4^+$ lymphocytes from patients with cirrhosis with MHE versus those without MHE. Shapes were used to distinguish the distinct features with hairpins representing miRNA, ovals showing transcription factors (TF), and rectangles denoting genes.

In all the cases, the regulatory miRNAs were upregulated while target mRNAs were downregulated in patients with MHE. In the case of transcription factors, the regulatory analysis indicated that previous studies have described *JUN*, *FOS*, *MAFF,* and *BHLHE40* (which were downregulated in this work), as binding to the promoter regions of 5 of the MHE-altered genes as well as 1 transcription factor (Table 4.2).

Table 4.2 Associations between miRNAs or transcription factors altered in CD4$^+$ T lymphocytes from patients with MHE, and their target genes.

miRNAs – mRNAs					
Regulator (miRNA)			**Target gene (mRNA)**		
Name	Fold Change	*p*-value	Name	Fold Change	*p*-value
miR-130a-3p	2.285	0.047	SLC7A5	0.482	0.002
miR-130a-3p	2.285	0.047	GADD45A	0.411	0.006
miR-130a-3p	2.285	0.047	ZFP36	0.362	0.005
miR-130a-3p	2.285	0.047	MAFF	0.416	0.005
miR-130b-3p	1.896	0.022	SLC7A5	0.482	0.002
miR-130b-3p	1.896	0.022	GADD45A	0.411	0.006
miR-130b-3p	1.896	0.022	ZFP36	0.362	0.005
miR-130b-3p	1.896	0.022	MAFF	0.416	0.005
miR-335-5p	1.863	0.049	CDKN1A	0.327	0.031
miR-494-3p	2.292	0.034	ETV3	0.455	0.017
miR-494-3p	2.292	0.034	BHLHE40	0.387	0.004
miR-494-3p	2.292	0.034	DDIT3	0.491	0.003
miR-494-3p	2.292	0.034	JUN	0.488	0.025
miR-656-3p	2.513	0.023	HBEGF	0.263	0.010
miR-656-3p	2.513	0.023	ZFP36	0.362	0.005
miR-656-3p	2.513	0.023	PLEKHH2	0.41	0.001
Transcription factors (TF) – mRNAs					
Regulator (TF)			**Target gene (mRNA)**		
Name	Fold Change	*p*-value	Name	Fold Change	*p*-value
BHLHE40	0.387	0.004	JUP	2.295	0.008
FOS	0.499	0.034	TNFAIP3	0.323	0.001
JUN	0.488	0.025	FOS	0.499	0.034
JUND	0.450	0.001	EGR2	0.351	0.044
JUND	0.450	0.001	PALLD	2.119	0.010
MAFF	0.416	0.005	HES4	0.407	0.010

The table is divided in 2 sub-tables: the first summarizes all the predicted miRNA–mRNA interactions and the second shows the predicted transcription factors (TF)–mRNA associations. In both cases, the left part (salmon) shows the regulators (miRNA or TFs) and the right (yellow) shows the target genes (mRNAs). Both fold-change and *p*-values are given for each feature to compare the changes and statistical significance in patients with cirrhosis with MHE versus those without MHE.

This analysis identified increased levels of miR-130a-3p and miR-130b-3p as potential inducers of the downregulation of *GADD45A* in $CD4^+$ lymphocytes from patients with MHE. GADD45A is a negative regulator of T-cell proliferation that suppresses the TCR signaling pathway mediated by p38 (Figure 4.8). Inactivation of GADD45A in knockout mice induces a lupus-like autoimmune response[247]. In mammalian cells, p38 constitutes an alternative TCR-signaling pathway in T cells, bypassing the upstream MAPK cascade[248]. p38 plays a role in the regulation of Th1 differentiation and so p38 inhibitors are now being used to treat Th1-mediated inflammatory diseases such as rheumatoid arthritis[249]. These data suggest that miR-130a-3p and miR-130b-3p mediated downregulation of *GADD45A* in $CD4^+$ T cells from patients with MHE may be one contributor to Th1 cell polarization. Moreover, this molecule may be an attractive therapeutic target because of its disruption of the TCR activation pathway is selectively restricted to activated T cells.

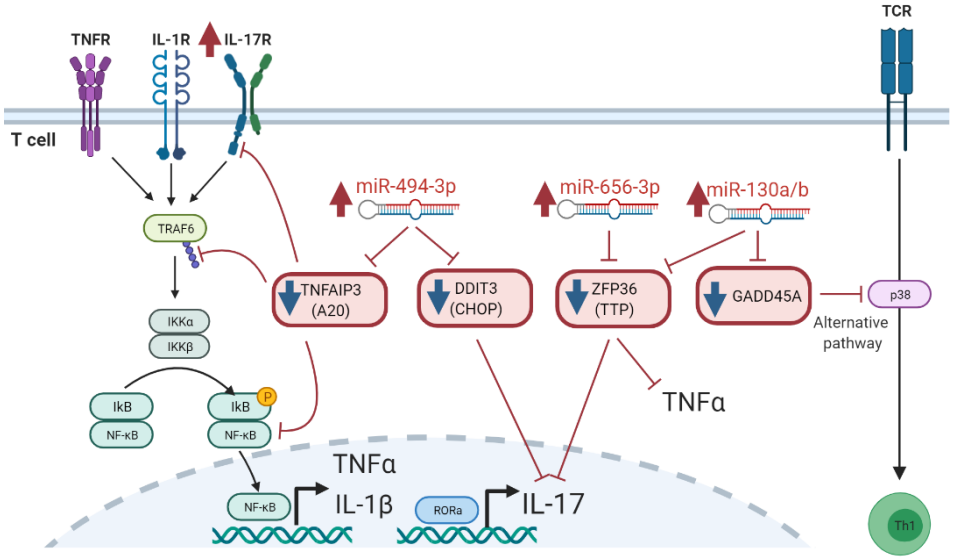


Figure 4.8 Potential mechanisms by which increased levels of miR-494-3p, miR-656-3p, and miR-130a/b may contribute to altering $CD4^+$ cell pathways and could be promoting a pro-inflammatory shift in peripheral inflammation

in patients with MHE. TNFAIP3 (A20) negatively regulates the production of pro-inflammatory cytokines (TNFα and IL-1β) by inhibiting the action of the NF-kB pathway at various levels (IL-17R, TRAF6, or NF-kB). DDIT3 (CHOP) and ZFP36 (TTP) act as post-transcriptional inhibitors of IL-17 and/or TNFα expression. GADD45A inhibits the T-cell receptor signaling pathway via p38, which then suppresses T-cell proliferation. Created with BioRender.com.

It has been proposed that activation of Th17 lymphocytes and increased IL-17 levels play a key role in triggering MHE[14]. Here we identified 2 miRNAs (miR-656-3p and miR-494-3p), which seem to contribute to this enhanced IL-17 response. Increased levels of miR-656-3p would downregulate *ZFP36*; the *ZFP36* gene encodes the tristetraprolin (TTP) protein which binds directly to AU rich-elements (AREs) in different mRNAs, including those for the cytokines TNFα, IL-10, IL-2, and IL-23. Interestingly, the stability of a large class of mammalian mRNAs is regulated by AREs located in the mRNA 3'-untranslated region (UTR).

mRNAs with AREs are inherently labile although, as a response to different cellular cues, they can either become stabilized, allowing gene expression, or destabilized, resulting in silencing[250]. Previous reports showed that TTP reduces IL-17 expression in human T cells by directly binding to a region of ARE motifs in the mRNA 3' UTR[251]. The downregulation of TTP by miR-656-3p in $CD4^+$ lymphocytes from patients with MHE would contribute to increased translation of *IL-17* mRNA and IL-17 protein levels.

Furthermore, DDIT3 (also known as CHOP or GADD153) is a suppressor of Th17 cytokine production, thereby inhibiting mRNA-translation. Chang et al.[252] showed that overexpression of CHOP reduces IL-17, IL-17F, and IL-22 levels. In $CD4^+$ lymphocytes from patients with MHE, increased levels of miR-494-3p would downregulate *DDIT3*, contributing to the reported increased production of IL-17 and IL-22 by isolated $CD4^+$ T cells from these patients[14].

We also identified miR-494-3p as an indirect regulator of *TNFAIP3* (Figure 4.8), via the AP-1 complex (FOS and JUN transcription factors). *TNFAIP3* is a gene that encodes the protein A20, whose dysfunction has been linked with excessive inflammation and autoimmunity[253]. *TNFAIP3* is also decreased in patients with multiple sclerosis and Parkinson's disease, contributing to the persistence of inflammation in these disorders[254]. Clinical studies of multiple sclerosis have linked reduced *TNFAIP3* levels (mainly in monocytes and to a lesser extent in $CD4^+$ T cells[255]) with increased disease progression[256].

A20 is expressed in practically all cell types and acts as a negative regulator in multiple inflammatory processes including TNFR, IL-1R, TLR, and NOD signals[257]. A20 negatively regulates inflammation by inhibiting the action of the NF-kB transcription factor[258]. In fact, the reduced levels of A20 in patients with MHE may enhance NF-kB-mediated pro-inflammatory responses (Figure 4.8). A20 is also able to interact directly with the IL-17RA receptor to reduce IL-17 responses[259,260].

In other neurological disorders, *TNFAIP3* gene expression is decreased in $CD4^+$ T cells and serum, and is associated with the upregulation of NF-κB expression[261]. Similar mechanisms may occur in $CD4^+$ T lymphocytes from patients with MHE: the increased levels of miR-494-3p would downregulate A20 and thus, increase the production of pro-inflammatory cytokines, as reported by Mangas-Losada et al.[14], in serum and $CD4^+$ T cells isolated from patients with MHE.

The increased levels of miR-494-3p in $CD4^+$ T lymphocytes from patients with MHE may explain some of the paradoxical effects on the toll-like receptors and IL-17 receptor pathways mentioned above. In these cases, the initial pathway elements (*TLR2*, *TLR4*, and *IL17RA*) are upregulated

while the final products of the pathways (the mRNAs for TNFα, IL-1β, CXCL8, CCL3, and CCL4 for the TLR pathway and TNFα, IL-1β, CXCL2, CXCL8, and PTGS2 for the IL-17 pathway) tend to be downregulated. As shown in Figure 4.5, the AP-1 complex (formed by the FOS and JUN transcription factors) is an essential element in the pathway and is significantly downregulated in both pathways. Therefore, upregulation of miR-494-3p would explain the downregulation of the *FOS*/*JUN* mRNAs in $CD4^+$ lymphocytes from patients with MHE (Figure 4.7 and Table 4.2).

Moreover, miR-494-3p may also contribute to reducing the expression of mRNAs for NF-kB-dependent pro-inflammatory cytokines in patients with MHE (Figure 4.9). Zhang et al.[262] showed that LPS-induced overexpression of miR-494-3p in RAW264.7 cells downregulated IL-1β and TNFα mRNA expression; miR-494-3p also specifically binds to the 3'-UTR of PTEN mRNA, reducing its transcription. PTEN is a phosphatase that dephosphorylates phosphatidylinositol (3,4,5)-triphosphate (PIP3), resulting in the inactivation of Akt1. Thus, the reduction of PTEN by miR-494-3p enhances LPS-induced Akt1 activation, which in turn suppresses nuclear translocation of the p65 subunit of NF-kB. This reduces the mRNA expression of NF-kB-dependent pro-inflammatory cytokines such as TNFα and IL-1β[262]. A similar process could occur in the $CD4^+$ T lymphocytes of patients with MHE and so an increase in miR-494-3p would likely reduce TNFα, IL-1β, and IL-17 mRNA levels by a similar mechanism.

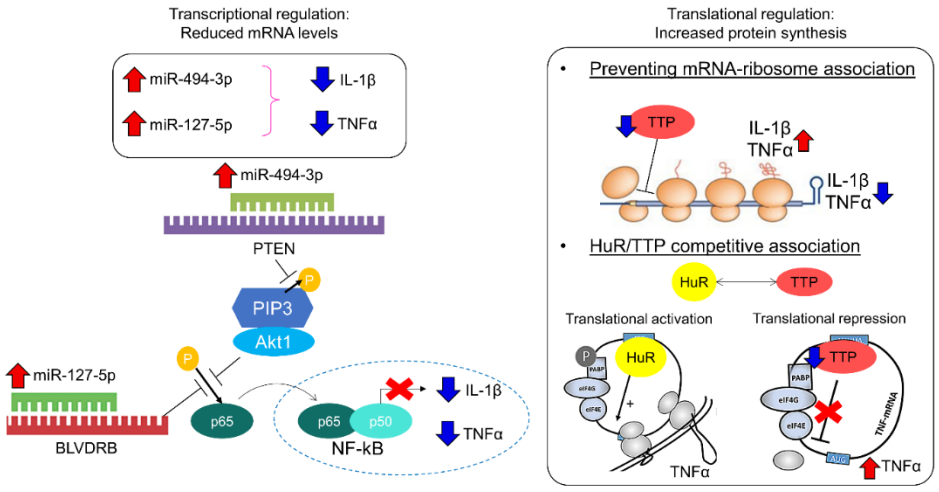


Figure 4.9 Summary of the molecular mechanisms that could explain the paradoxical effects found in our functional enrichment analysis. Transcriptional regulation mechanisms such as mRNA silencing by upregulation of miRNA activity can reduce the levels of mRNA for proinflammatory cytokines (IL-1β and TNFα). Translational regulation mechanisms can explain the increased protein levels of proinflammatory cytokines while messenger levels remain decreased.

miR-127-5p is another miRNA which may downregulate the TNFα and IL-1β mRNAs by reducing NF-kB activity in $CD4^+$ lymphocytes from patients with MHE (Figure 4.9). Huan et al.[263] showed that miR-127-5p binds to the 3'-UTR of biliverdin reductase B (BLVDRB) mRNA, thereby reducing its transcription. The reduction of BVDRB inhibits phosphorylation of the p65 subunit of NF-kB, reducing its translocation to the nucleus and the expression of NF-kB-dependent mRNAs including TNFα, IL-1β, and IL-17. The increase in miR-127-5p in the $CD4^+$ T cells of individuals with MHE may therefore also contribute to reducing the levels of TNFα and IL-1β mRNAs by reducing NF-kB activity.

These data support the idea that the increase in specific miRNAs such as miR-494-3p and miR-127-5p could contribute to the downregulation of mRNAs for the final steps of the toll-like receptor and IL-17 signaling pathways in $CD4^+$ T cells from patients with MHE. This would be mediated

by a reduction in NF-kB- and/or AP1-dependent mRNA transcription.

Paradoxically, we found that the mRNAs for most of the inflammatory cytokines were downregulated (e.g., for TNFα and IL-1β) while in a previous study, we found upregulated protein levels of the same cytokines in cultures of isolated $CD4^+$ T cells from patients with MHE[14]. This suggests that the synthesis of the proteins for TNFα or IL-1β is increased in patients with MHE despite their reduced mRNA levels. However, this is not unusual and indeed does occur under certain conditions for distinct types of proteins (Figure 4.9).

The increase in miR-656-3p and the associated reduction in *ZFP36* (encoding tristetraprolin protein, or TTP) in $CD4^+$ T lymphocytes from patients with MHE (Figure 4.8) may also contribute to this discrepancy between the mRNA and protein levels for some cytokines, including TNFα. ZFP36 RNA-binding proteins are prominent inflammatory regulators linked to autoimmunity and modulate the levels of cytokines such as TNFα, IL-17, and IFNγ, which are all increased in T cells from knockout mice lacking the *ZFP36* gene[251,264].

In fact, Tiedje et al.[265] proposed that the translational control of TNFα mRNA is achieved by the competitive association of human antigen R protein (HuR) and TTP, with the latter acting as a translation repressor and the former acting as translation activator. A reduction in TTP (*ZFP36*) levels therefore leads to enhanced TNFα mRNA translation without a need for increased transcription of its mRNA, thus allowing faster adaptation to environmental factors and inflammatory events.

Indeed, Moore et al.[264] showed that TTP (ZFP36) represses mRNA target translation by preventing the association of mRNAs with ribosomes and that a reduction in *ZFP36* levels in Zfp36 knockout cells led to increased

ribosome binding to TNFα mRNA, resulting in increased protein levels for TNFα despite the trend towards reduced levels of its mRNA. A similar process would explain why TNFα protein levels are also increased in $CD4^+$ T lymphocytes from patients with MHE in spite of the tendency towards a reduction in its mRNA.

The increased levels of miR-656-3p would reduce *ZFP36* (TTP) levels, thus reducing its repression of TNFα mRNA translation, allowing increased ribosome binding, mRNA translation, and increased TNFα protein levels. *ZFP36* (TTP) levels are also modulated by miR-130b-3p (Figure 4.8) whose expression is likewise increased in $CD4^+$ T lymphocytes from patients with MHE. Thus, this mechanism could also induce a similar increase in binding of ribosomes, mRNA translation, and increased TNFα protein levels.

4.4. Conclusion

In summary, in this work we analyzed gene expression changes in $CD4^+$ T lymphocytes from patients with MHE, as well as the possible involvement of miRNAs and transcription factors in the regulation of these altered mRNAs. We identified 167 genes and 20 associated signaling pathways whose expression was significantly altered in $CD4^+$ T lymphocytes from patients with MHE. Four altered pathways which may contribute to the immune system shift associated with the appearance of MHE are the toll-like receptor and IL-17 signaling pathways (which are directly involved in modulating immune system responses) and the histidine and tryptophan metabolism pathways (which also modulate immune system function). The increase of mRNAs for TLR-2, TLR-4, and IL17RA receptors in $CD4^+$ lymphocytes from patients with MHE further

supports the idea that Th1/Th17 cell related responses play a key role in inducing MHE.

We also identified 13 altered miRNAs and performed an mRNA–miRNA integration analysis to discover 4 miRNAs that may modulate mRNAs for key proteins involved in the immunological shift that triggers MHE in patients with cirrhosis. Modulation of the expression of proteins such as A20 (TNFAIP3) and TTP (ZFP36) in $CD4^+$ T lymphocytes by increased levels of miR-494-39, miR-656-3p, and miR-130b-3p would contribute to the changes in signal transduction pathways, thereby resulting in an increase in pro-inflammatory cytokines such as IL-17 and TNFα.

These results describe several pathways and mechanisms by which alterations in $CD4^+$ lymphocytes may contribute to the shift in peripheral inflammation that triggers MHE in patients with cirrhosis. Thus, our work could pave the way towards the development of new diagnostic and therapeutic approaches to MHE.

Chapter 5

A Nextflow pipeline for T-cell receptor repertoire reconstruction and analysis from RNA sequencing data

This chapter was adapted from the paper '*A Nextflow pipeline for T-cell receptor repertoire reconstruction and analysis from RNA sequencing data'* by Teresa Rubio, María Chernigovskaya, Susanna Marquez, Cristina Marti, Paula Izquierdo-Altarejos, Amparo Urios, Carmina Montoliu, Vicente Felipo, Ana Conesa, Victor Greiff, Sonia Tarazona; Immunoinformatics, 6, 2022. DOI: 10.1016/j.immuno.2022.100012

5.1. Introduction

TCRs can recognize an immense variety of processed antigens. In fact, the potential diversity of unique TCRs is approximately 10^8–10^{10} in each human individual[266–268]. A T-cell clonotype is a set of cells that share the same TCR, while the set of unique T-cell clonotypes in an individual is called the TCR repertoire. The TCR is a two-chain protein and most human T cells comprise α/β chains (TRA and TRB) with a small proportion which comprise the γ/δ chains (TRC and TRD). The receptors are synthesized during VDJ recombination, which involves rearrangement of the *V*, *D*, *J*, and *C* genes as described in Chapter 1.

The immense diversity in the TCR repertoire of each individual results from the many possible combinations of these genes, particularly at the N-terminal end of the receptor which recognizes and binds antigens[68]. Three CDRs are important for binding antigens, with CDR3β (the CDR3 region of the TRB chain) being the preferential target of many TCR repertoire studies because of its high diversity and primary importance for antigen binding[269].

High-throughput sequencing is a powerful tool for the analysis of these highly diverse immune repertoires and has contributed to lymphocyte biology research, antibody engineering, and the production of vaccines[270,271]. To capture the TCRs of α/β T cells, most high-throughput sequencing immune repertoire studies apply specific library preparation methods that target the receptor transcripts. These sequencing protocols include multiplex PCR with guided primers for the amplification of TCR transcripts at the CDR3 region; target enrichment to capture the sequences of interest using complementary RNA baits; and the more

standard 5'RACE (rapid amplification of 5'-complementary DNA ends), which can retrieve the complete 5'-end of mRNAs[271].

However, immune repertoires can also be efficiently extracted from transcriptomic RNA-seq data[272–274] given that TCR transcripts are part of the bulk-sequenced transcriptome. Using RNA-seq for immune receptor analysis reduces economic costs and the required sample quantities because both gene expression and immune receptor transcripts are measured in the same experiment. Conversely, less sensitivity for TCR detection can be expected when using RNA-seq data because only the most abundant and expanded clones will be recovered.

Different computational methods such as **MiXCR**[272], **CATT**[275], or **TRUST4**[274] can reconstruct TCR receptors from RNA-seq data. All these repertoire reconstruction algorithms have been evaluated and compared with each other in terms of their sensitivity and specificity for TCR extraction. For instance, **MiXCR** software is able to extract high-frequency clonotypes better than the first version of **TRUST** at every tested read-sequencing length, and most of the MiXCR-reported clonotypes were confirmed by control TCR-seq data[272]. However, the number of clonotypes reported by **MiXCR**, **CATT**, **TRUST3**, and **TRUST4** from *in silico* RNA-seq data was very similar when using sequences equal to or longer than 150 bp, except for **CATT** which produced twice the number of mismatches than perfect matches[274].

Despite the wide availability of TCR repertoire reconstruction tools, a detailed step-by-step pipeline for the processing of immune repertoire data from standard bulk RNA-seq data is not readily available, representing a missed opportunity for complementing gene expression

studies of immunological conditions with the analysis of the immune repertoire.

5.2. Methods

5.2.1. Overview of the analysis strategy

Here, we present an end-to-end pipeline for the analysis of TCR repertoire profiles from bulk RNA-seq, implemented in Nextflow (https://www.nextflow.io/) for easy distribution and robust utilization. Nextflow is a workflow management system that provides native support to run pipelines in multiple compute environments and with containerization systems. Nextflow pipelines have become very popular mainly due to the simplicity to run an analysis while transparently managing common issues of shell scripting (e.g., required dependencies, computational resources, code failure tracking, cumbersome transfer between collaborators).

We applied our Nextflow pipeline to a dataset of $CD4^+$ T cells isolated from controls (healthy individuals) and patients with cirrhosis with and without MHE (Figure 5.1). As previously discussed, specific alterations in the immunophenotype of patients with cirrhosis with MHE pointed towards $CD4^+$ cells as key players in the immune shift that triggers the appearance of MHE[14]. Therefore, studying $CD4^+$ T-cell repertoires might help us to understand the immune status of patients with MHE.

The RNA-seq based pipeline for TCR repertoire analysis starts by performing quality control on the raw sequencing reads, followed by bias detection and pre-processing. Read alignment to the reference VDJ genes (IMGT database[276]) is subsequently carried out to assemble and

define T-cell clonotypes by using **MiXCR**[277] software (Figure 5.1). Secondary analysis of the TCR repertoire includes clonal convergence to identify shared clones; clonal expansion to measure diversity; a comparison of the clonal architecture and repertoire similarity to evaluate sequence likeness; and antigen specificity analysis using public databases (McPAS-TCR[278] and VDJdb[279]).

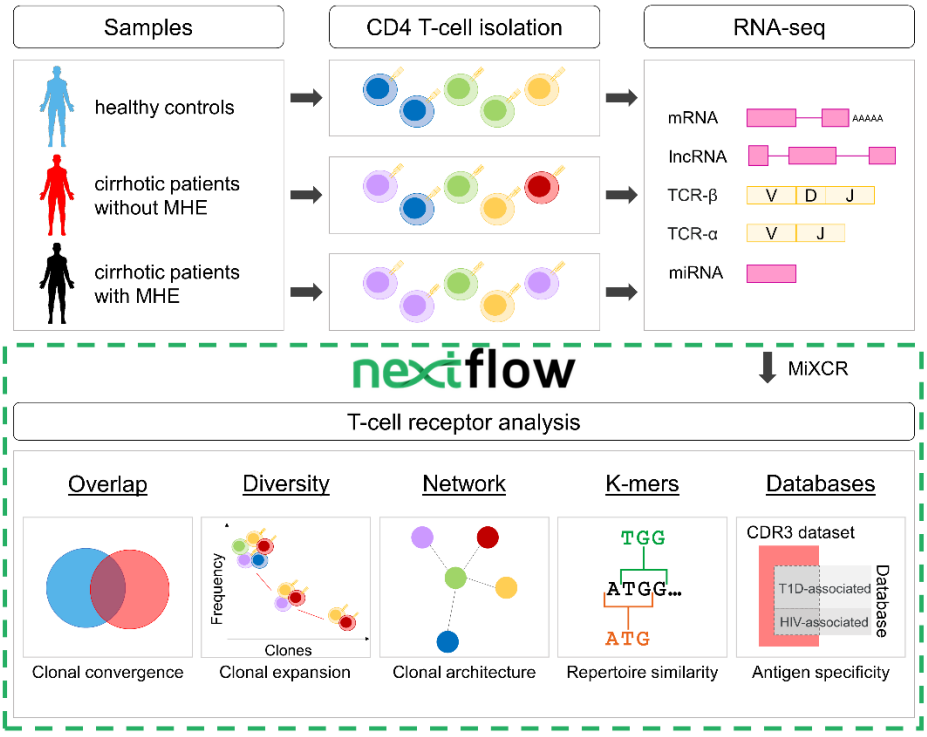


Figure 5.1 Overview of the CD4$^+$ T-cell repertoire analysis based on RNA sequencing data in patients with MHE. The experimental steps comprised sample collection, cell isolation, and RNA extraction and sequencing. The computational analysis included repertoire quantification and repertoire properties screening. Abbreviations: mRNA, messenger RNA; lncRNA, long non-coding RNA; TCR, T-cell receptor; miRNA, microRNA; T1D, type 1 diabetes, and HIV, human immunodeficiency virus.

5.2.2. Sample collection, RNA sequencing, and read pre-processing

The RNA-seq dataset used in this study was previously described in Chapter 4 (in the Methods section). Briefly, RNA-seq data (125 bp paired-end reads, with a sequencing depth of 50 million reads for each sample, were selected for short-read sequencing) were obtained from patients with cirrhosis with and without MHE, as well as from healthy controls. The reads were trimmed with **Trimmomatic**[211] software (version 0.38) when the average quality score was below 20 (in a sliding window of 20 bp) and were removed if the resulting length was less than 80 bp. These parameters were selected as optimal after a comparative analysis of different sliding window values (from 4–20bp).

5.2.3. Repertoire reconstruction

MiXCR[277] (version 3.0.13) was used to align and assemble the TCR repertoire from RNA-seq data with the 'analyze shotgun' command, applying the default parameters. Clones (i.e., the CDR3 amino acid sequence of the TRB chain) were included in the analysis if they had a minimal read abundance of 2 counts and their CDR3βs were a minimum of 4 amino acids in length, as previously described[280].

5.2.4. Hill-based evenness profiles

The Hill numbers are diversity indexes defined as follows:

$$D_{\alpha} = \left(\sum_{i=1}^{n} {f_i}^{\alpha}\right)^{\frac{1}{1-\alpha}}, \qquad [1]$$

where n is the number of clones within each repertoire, f_i is the frequency distribution (proportional abundance of clones), and α is a scale parameter in (0,1) and (1, +∞). The diversity profile (D_α) was previously defined[122] as:

$$D_\alpha = SR \times E_\alpha, \quad [2]$$

where E_α is the evenness and SR is the species richness, or the number of unique clones in each repertoire dataset.

Here, we calculated evenness profiles (E_α), defined as D_α/SR according to equation 2. We used different values of α, ranging from 0 to 10 with a step size of 0.2, to obtain the diversity profiles (D_α). Diversity is not defined for the case α = 1, however L'Hospital's rule states that, as α tends towards 1, the diversity tends towards the exponential of the Shannon entropy:

$$D_{\alpha=1} = exp(-\sum_{i=1}^{n} f_i \ln(f_i)) . \quad [3]$$

Finally, all the pairwise Pearson correlation coefficients of the evenness profiles for the samples were calculated. The significance of the Pearson correlation was tested, and *p*-values were adjusted for multiple testing using Holm's method, implemented in the *rcorr.adjust()* function of the **RcmdrMisc**[281] R package. Hierarchical clustering was performed based on Euclidean distances for correlations and heatmaps were generated for visualization purposes.

5.2.5. Shannon Evenness

Shannon evenness (S-E) is defined as the Shannon entropy divided by the species richness (SR). S-E is 1 if all clones in a repertoire have the same frequency (an 'even' repertoire), or it converges to 0 if very few clones dominate in the repertoire (a 'polarized' repertoire).

5.2.6. Jaccard similarity

Pairwise clonal convergence between 2 repertoires (A and B) was quantified using the Jaccard similarity coefficient, defined as the size of the intersection of A and B divided by the size of the union of A and B:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \quad [4]$$

Values range between 0 and 1, where 1 means complete overlap of repertoire A and B, and 0 indicates no overlapping receptor sequences between repertoires A and B.

5.2.7. k-mer-based CDR3β analysis

Overlapping k-mers of length 3 (k = 3) were extracted from the CDR3β amino acid sequences[280] in each TCR repertoire and were condensed into a k-mer frequency distribution matrix using the **immunarch** R package[282]. The correlation significance test and hierarchical clustering based on correlation distance and heatmap visualization were calculated as described above in section 5.2.4.

5.2.8. T-cell receptor sequence similarity architecture

The TRB repertoire networks were generated as previously described[129,280], where nodes represent amino acid CDR3β sequences and edges were drawn between sequences differing by 1 amino acid (Levenshtein distance = 1). The degree (number of linked nodes) distributions of each repertoire were calculated using the degree function from the **igraph**[283] R package.

5.2.9. Generation of the graphics

Statistical analyses were performed and graphics were produced using the programming environment R[284] (version 4.0.5). The matrix of public clones generated in the repertoire overlap analysis was created with the **immunarch** R package[282] using the *repOverlap()* and *vis()* functions. All the heatmaps were created using the *aheatmap()* function in the **NMF** R package[285]. Mean quality sequencing plots for paired-end reads were obtained from **FastQC**[286] reports, and bar graphs summarizing the **MiXCR** output were drawn using the **ggplot2** R package[287]. The *ggboxplot()* and *ggscatter()* functions in the **ggpubr** R package[288] were used to statistics and correlations of clones, respectively.

5.2.10. Antigen/Disease-specific T-cell receptor databases

McPAS-TCR is a curated database of TCR sequences linked to the associated antigen target or pathology based on the published academic literature[278]. The database was downloaded and filtered according to the pathological category, maintaining human TRB sequences associated with autoimmunity and pathogens. In turn, VDJdb is a curated repository of antigen-specific TCR sequences that utilizes experimental information from recently published TCR specificity assays[279]. At the time of download, the McPAS-TCR and VDJdb databases had last been updated on 6 March 2021 and 2 February 2021, respectively.

5.2.11. Fisher's exact test analysis

The overrepresentation of clones associated with diseases or antigens (McPAS-TCR and VDJdb) in our CDR3β sequences was evaluated by a

one-tailed Fisher's exact test applied to each group of patients (control, with MHE, and without MHE), using the disease categories included in the McPAS-TCR and VDJdb databases. The TCRs present in the samples but absent in the McPAS-TCR and VDJdb databases were excluded from the analysis as described previously[289]. Fisher's exact test was used to test the overrepresentation of specific disease-associated receptors in the database within the measured receptors of the sample. The obtained *p*-values were adjusted for multiple testing using Benjamini-Hochberg FDR correction, considering both the number of diseases and the number of sample groups tested.

5.2.12. Nextflow pipeline

Nextflow v21.10.6 was used to implement the pipeline. In addition, the DSL2 syntax extension was enabled at the beginning of the workflow script to allow the definition of module libraries and simplify the writing of the data analysis pipeline.

5.2.13. Data and code availability

The transcriptomic dataset used in this study is available in the GEO database repository, GSE184200, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE184200. The code and complete documentation of the Nextflow pipeline are publicly available from the Github repository: https://github.com/ConesaLab/TCR_nextflow.

5.3. Results and discussion

5.3.1. Step-by-step analysis overview

The step-by-step pipeline for processing immune repertoire data from whole transcriptome RNA-seq reads is summarized in Figure 5.2. The pipeline consists of 4 main blocks representing both the experimental procedure and the computational analysis: (1) experimental design, (2) transcriptome profiling, (3) AIRR pre-processing, and (4) AIRR analysis. The experimental procedures include sample collection, immune cell isolation, RNA extraction, sequencing design (e.g., library strand specificity and paired- or single-end reads), and RNA sequencing (RNA-seq). The computational analysis starts with 'fastq' files in which the sequencing quality (step 1) needs to be verified.

The **MiXCR** software[277] was used to assemble TCR repertoires from sequencing reads after quality control. The **MiXCR** assembly algorithm reduces the introduction of false-positive clones, which might appear either by alignment of reads to non-target molecules or by overlapping between 2 sequences from different clones in the reconstruction of partially covered CDR3 regions. The Nextflow pipeline starts with the **MiXCR** repertoire extraction step (Figure 5.2) and was performed using the MiXCR 'analyze shotgun' command, which consists of the following workflow: sequence alignment against reference *V*, *D*, *J*, and *C* genes (step 2), followed by clustering of identical sequences into clonotypes (default, clustering by CDR3β; step 3), correction of PCR/sequencing errors (step 4), and output of a tab-delimited file containing the quantification as a clonotype matrix (step 5).

Additional AIRR analysis steps are needed to study different immune receptor features: overlap for clonal convergence (step 6a), diversity indexes for clonal expansion (step 6b), network analysis for clonal sequence architecture (step 6c), k-mer distribution for repertoire sequence similarity (step 6d), and public database screening for antigen specificity (step 6e). Steps 2–6 were implemented in the Nextflow pipeline as parallel processes that receive MiXCR files as input and provide ready-to-publish plots and tables as well as a final report summarising all results for better user interpretability.

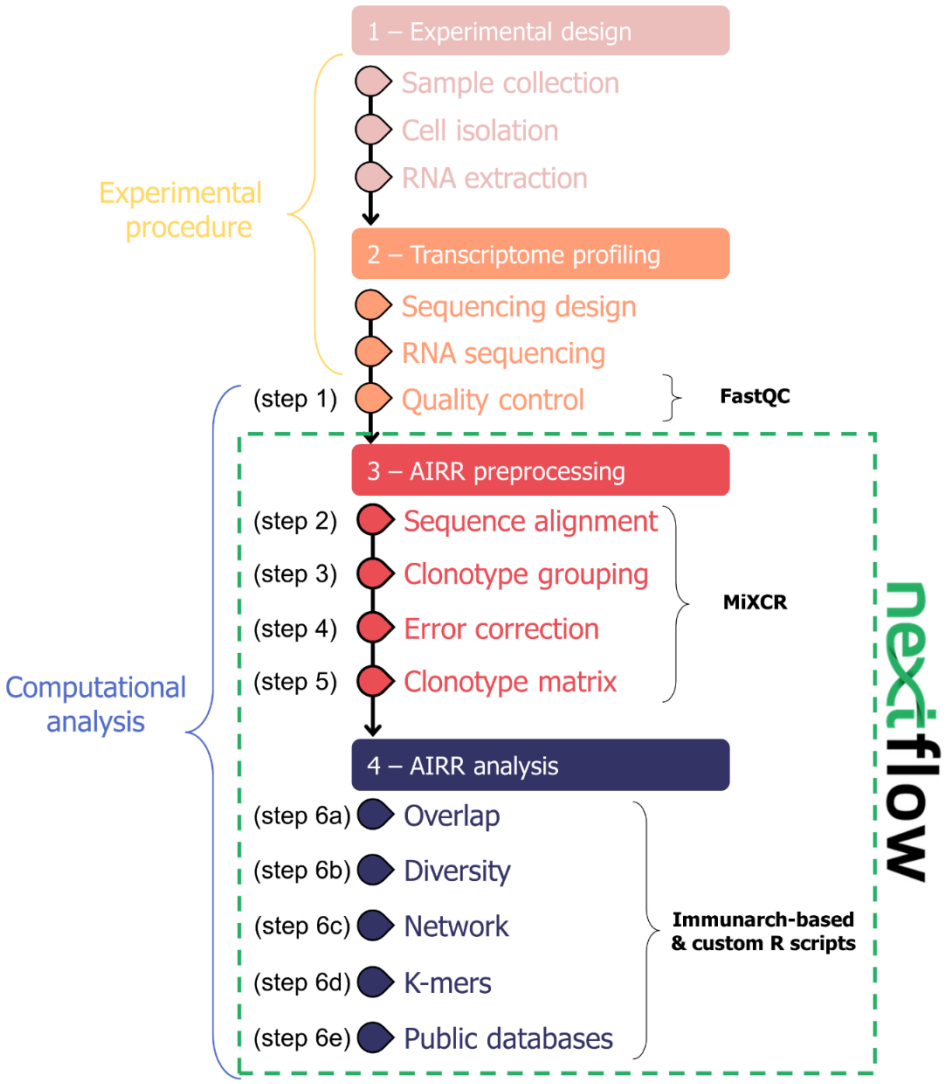


Figure 5.2 Overview of the pipeline for processing immune repertoire data from whole transcriptome RNA-seq data. Experimental steps comprise sample collection, cell isolation, RNA extraction, and sequencing. The computational pipeline analysis implemented in Nextflow included repertoire quantification, and repertoire properties screening. AIRR: Adaptive Immune Receptor Repertoire.

5.3.2. T-cell receptor sequences can be recovered from RNA-seq data (steps 1–5)

CD4$^+$ T cells were isolated and sequenced by bulk paired-end RNA-seq from a total of 20 patients (8 controls, 6 patients with cirrhosis without MHE, and 6 patients with cirrhosis with MHE). Sequencing read pre-processing included trimming and filtering (see the Methods section), which resulted in good quality scores (mean q > 30) for all the samples (Figure 5.3).

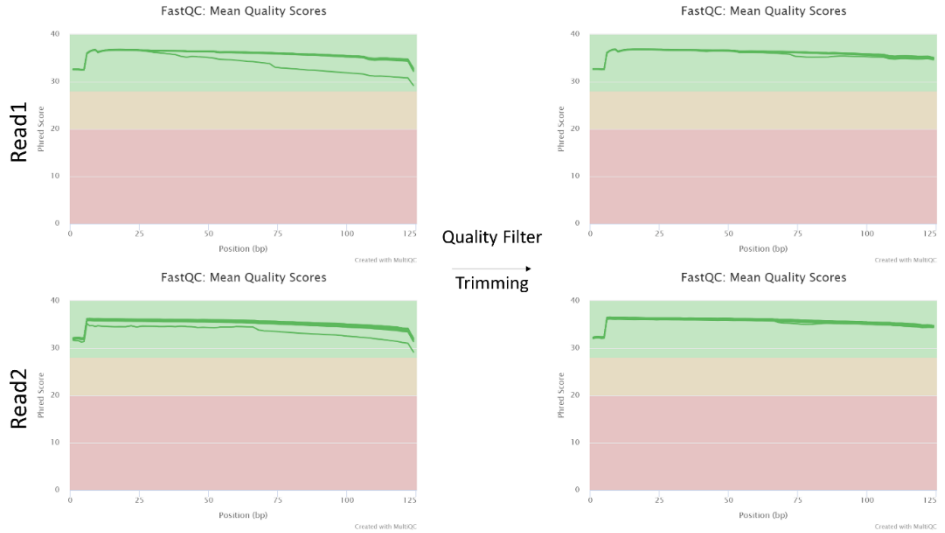


Figure 5.3 Sequencing read quality control of RNA-seq data measured in CD4$^+$ T cells isolated from patients with MHE. The sequencing-read quality of all the samples measured using the Phred score for both read1 and read2 in the paired-end dataset before and after read trimming is shown. The green area indicates excellent read sequencing quality (q > 28), the orange area indicates good quality (28 > q > 20), and the red area indicates poor quality (q < 20).

Read alignment against the reference VDJ genes (IMGT database[276]) showed a range of successfully aligned reads between 0.05–0.1%. The majority of reads matched TCR regions (24.1–67.6% of successfully aligned reads), although some reads aligned with immunoglobulin (IG) chains (27.1–75.0% of successfully aligned reads), thereby indicating

slight contamination during T-cell isolation (Figure 5.4). A high proportion (52.5–87.7%) of the recovered clones (i.e., CDR3 amino acid sequences), matched TRA and TRB chains. Bulk RNA-seq data cannot determine the pairing of specific α/β receptor chains within the population of T cells because this can only be achieved by sequencing single T cells. Therefore, we decided to focus on the TRB chain for our subsequent analyses. TRB is more appropriate than TRA to identify T-cell clones because around 7–30% of T cells can have 2 different alpha chains expressed on the same clone[290] while only 1% of T cells can have 2 different beta chains on the same clone[291].

Figure 5.4 Read alignment and clonotype assembly quality control of RNA-seq data measured in CD4$^+$ T cells isolated from patients with MHE. Summary of lymphocyte chain quantification reported after MiXCR read alignment (upper graph) and clone assembly (lower graph). Colors indicate the number of aligned reads/assembled clonotypes for the different TCR (TRA, TRB, TRD, TRG) or BCR (IGH, IGK, IGL) chains in the 3 patient groups (blue: controls; red: patients with cirrhosis without MHE; dark grey: patients with cirrhosis with MHE).

We have compared our 3 groups of patients using the Kruskal-Wallis test for various repertoire statistics – number of isolated cells, RNA quantity, number of reads obtained, number of recovered clones, i.e., the CDR3β amino acid sequence, and their Shannon evenness (Figure 5.5). The clone recovery yield ranged from 498 to 1,114 distinct CDR3 amino acid sequences per patient for the TRB. None of the mentioned measurements showed any significant differences between the groups of patients (Kruskal–Wallis test, *p*-value > 0.05) except for the number of clones, which was significantly increased in patients with cirrhosis without MHE versus the controls (post hoc Wilcoxon test, *p*-value = 0.024).

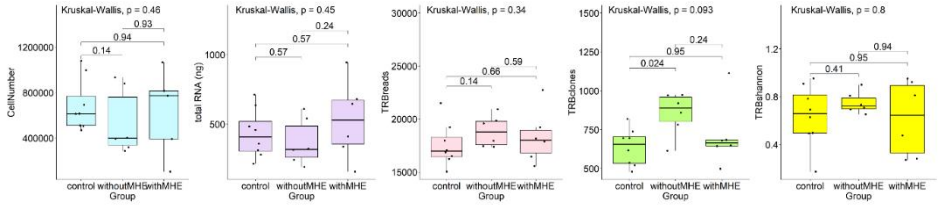


Figure 5.5 T-cell receptor beta chain (TRB) repertoire statistics. Total CD4$^+$ T-cell number, total RNA, number of reads, number of TRB clones, and Shannon evenness statistics (global *p*-values from the Kruskal–Wallis test and pairwise *p*-values from Wilcoxon post hoc tests).

To determine whether the sequencing depth sufficiently covered the clonal repertoire of the samples, we calculated pairwise correlations between the cell number, number of clones, and the Shannon evenness across all the samples (Figure 5.6). When the sequencing depth is saturating with respect to clone detection, the number of clones solely

depends on the sample type and not on the number of reads. We found a positive correlation (Pearson coefficient = 0.77, *p*-value = 6.5×10^{-5}), that may indicate insufficient sequencing depth. Nevertheless, the number of distinct CDR3 sequences assembled was of similar magnitude as reported in other studies of TCR reconstruction from bulk RNA-seq data: 367–936 TRB clonotypes extracted from the central nervous system and 1,684–2,977 TRB clonotypes extracted from the spleen using paired-end data from isolated $CD4^+$ T cells[272].

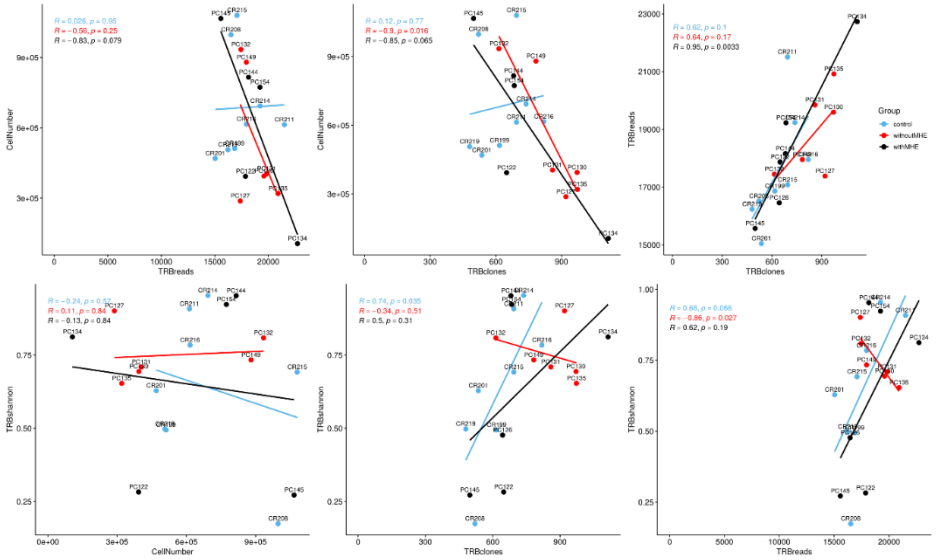


Figure 5.6 T-cell receptor beta chain (TRB) repertoire sequencing correlation plots. Pairwise Pearson correlation plots between various repertoire statistics: the number of isolated cells (CellNumber), number of TRB reads obtained (TRBreads), number of recovered TRB clones (TRBclones), and their Shannon evenness (TRBshannon). Colors indicate the 3 patient groups (blue: controls; red: patients with cirrhosis without MHE; dark grey: patients with cirrhosis with MHE).

5.3.3. T-cell receptor sequence profiling in patients with minimal hepatic encephalopathy

5.3.3.1. Low clonal convergence among patient samples (step 6a)

Repertoire overlap analysis is the most common approach to uncover clonotypes shared between given individuals, which are also denominated as 'public' clones[292–294]. Using the Jaccard similarity measure (see the Methods section), we found a low clonal convergence (0.00027 ± 0.00041 [average ± standard deviation] Jaccard similarity) between healthy controls and patients with cirrhosis with or without MHE (Figure 5.7A).

5.3.3.2. High clonal expansion in all the samples, independently of the immune status (step 6b)

The expansion of individual T-cell clones that bind their matching antigen was analyzed using Hill-based evenness profiles, a diversity measurement derived from ecology (see the Methods section). Unlike single diversity indices, which can produce different clonal expansion results, diversity profiles capture the entire immune repertoire and more sensitively reflect immunological statuses[122]. The $CD4^+$ T cells in this work positively correlated (Pearson coefficient from 0.59 to 1, *p*-value from 0 to 5.32×10^{-6}) in diversity profiles of T-cell clones, regardless of the cognitive impairment or cirrhosis condition of the studied patients (Figure 5.7B).

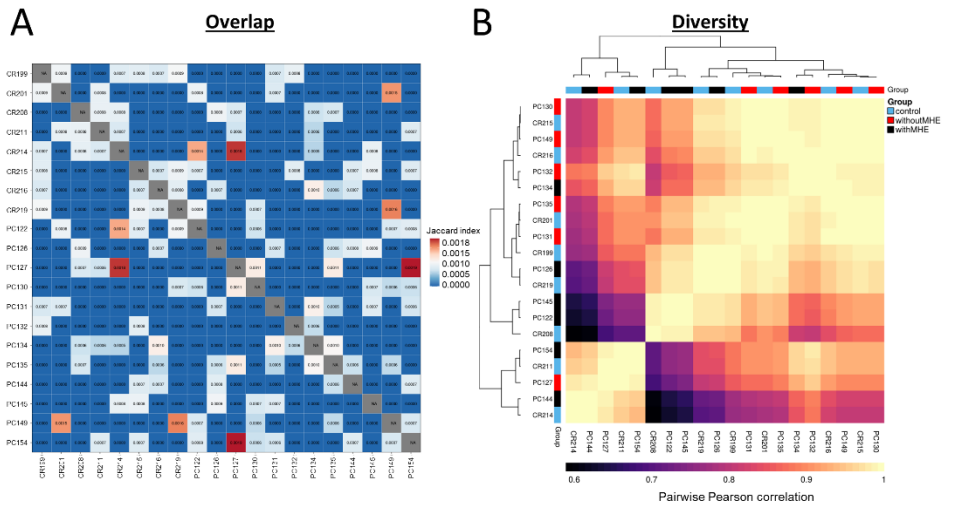


Figure 5.7 Overlap and diversity analysis of TCR repertoires showed specific within-sample profiles which were independent of immune status. (A) Overlap analysis of CDR3 sequences to uncover clonal convergence or shared clonotypes between individuals. Jaccard similarity coefficient matrix indicated a strong (red) or weak (blue) overlap between 2 sample sets. (B) Diversity profile analysis to measure clonal expansion. The heatmap contains high (light yellow) or low (dark purple) pairwise Pearson correlation coefficients of the evenness profiles calculated for the control group (blue), patients with cirrhosis without MHE (red), or patients with cirrhosis with MHE (dark grey).

5.3.3.3. Increased within-repertoire similarity based on repertoire architecture was unconstrained by immune status (steps 6c–d)

The adaptive immune response was determined by immune receptor sequences: the higher their dissimilarity, the wider the range of antigens they can recognize. The all-to-all sequence similarity within a repertoire represents the repertoire architecture, which was measured in our patients using both k-mer and network analysis.

First, the number of overlapped 3-mers (k-mer length = 3 amino acids) were calculated for each patient, as previously described[280]. A large positive correlation, ranging from 0.96 to 1 with a *p*-value = 0 was obtained

between the k-mer vectors (Figure 5.8A). This result might indicate that patients share similar sequence architecture patterns independently of their immune status.

To complete the repertoire architecture analysis, a sequence similarity network was constructed using the CDR3β amino acid sequences as nodes and adding an edge when the sequences were deferred by 1 amino acid (Levenshtein distance = 1). The number of similar clones (network degree) was then calculated and represented as a heatmap (Figure 5.8B). A total of 96.8% of the clones had degree = 0 (single nodes) in all the samples, and the maximum degree obtained was 5 in 1 cirrhotic patient without MHE (PC149). This substantial proportion of clones with clone degree 0 was also found in $CD4^+$ cells from patients with multiple sclerosis, while $CD8^+$ cells presented a more homogeneous degree distribution[280]. Moreover, most patients with cirrhosis without MHE showed a significantly higher (Wilcoxon test, *p*-value = 0.013) number of single nodes (Figure 5.8C), which may be related to underlying differences in the number of TRB clones between the 3 different groups (Figure 5.5). Noteworthy, sample PC134 (a patient with cirrhosis without MHE) presented 1,064 single nodes, the highest number compared with the rest of the patients (472–935 single nodes), which could be explained by the highest number of clones (1,114 clones) in this sample, and we considered it as an outlier in this analysis (Figure 5.8C).

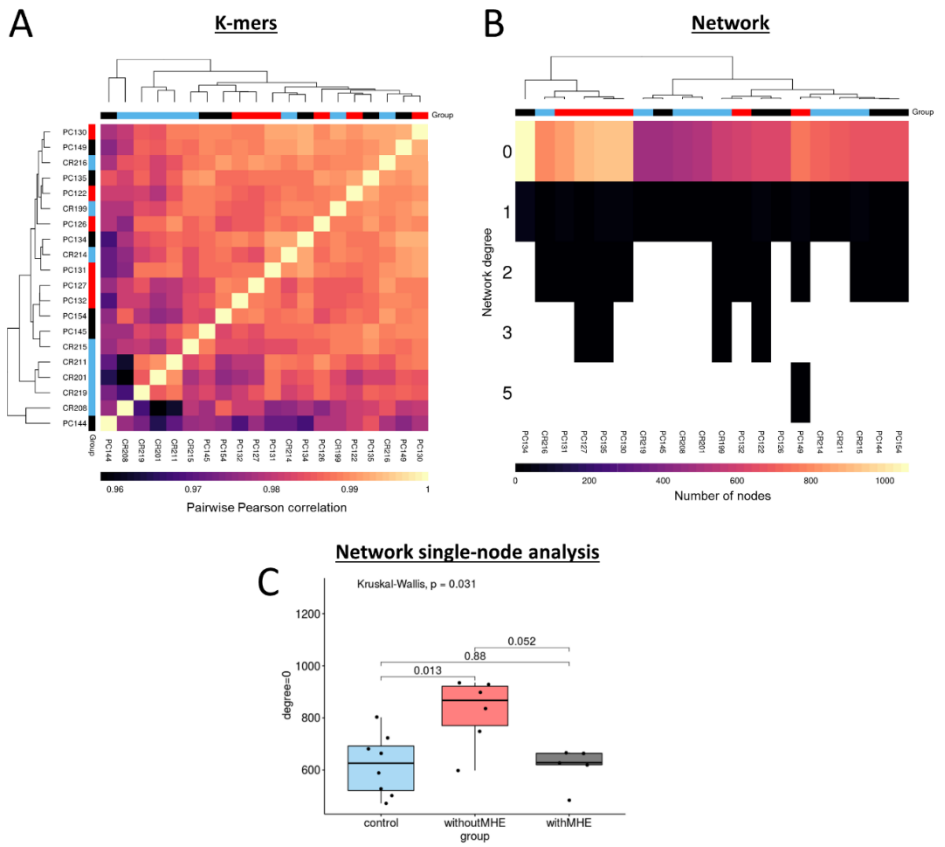


Figure 5.8 Repertoire architecture analysis of TCR repertoires showed specific within-sample profiles which were independent of immune status. (A) K-mer analysis to examine the repertoire similarity. The heatmap represents the pairwise Pearson correlation coefficients of the 3-mer (k-mer length = 3 amino acids) frequency distribution matrix from the highest (light yellow) to the lowest (dark purple) results. (B) Network analysis to study the clonal architecture. The colors, from light yellow to dark purple, represent the number of nodes with degree 0–5 for each repertoire. (C) Single network node (degree = 0) distribution between patient groups to test if the number of single nodes in patients with cirrhosis without MHE (red) were significantly higher than the controls (blue) or patients with cirrhosis with MHE (dark grey).

5.3.4. T-cell receptors with known disease associations were overrepresented in minimal hepatic encephalopathy CDR3 repertoires (step 6e)

To evaluate if the clones in our patients were significantly associated with the described diseases or antigens, we assessed the overlap (see the Methods section) between our CDR3β sequences and 2 different TCR databases: VDJdb and McPAS-TCR. The former contained a total of 41,169 human TRB sequences with Cytomegalovirus being the species epitope with the highest number of sequences, comprising nearly a half of them. The latter database contained a total of 30,052 TRB sequences within the autoimmune and pathology categories, with over half of them belonging to *Mycobacterium tuberculosis*.

The sequence intersection between the 2 databases was low, ranging from 2 to 1,032 in the common pathologies/pathogens: Influenza A, Cytomegalovirus, Epstein-Barr Virus, Human Immunodeficiency Virus, Yellow fever virus, Human T-cell Lymphotropic Virus, Hepatitis C virus, *Mycobacterium tuberculosis*, Herpes Simplex Virus 2, and COVID-19, sorted by decreasing order of the shared sequences (Table 5.1).

The CDR3β sequences were grouped by patient type (control, patients with cirrhosis without MHE, or patients with cirrhosis with MHE) to test the overlap with the McPAS-TCR database (Table 5.2).

Table 5.1 Common pathologies/pathogens and the number of shared TRB sequences between the McPAS[278] and VDJbd[279] databases.

McPAS vs. VDJdb database intersection	
Common pathologies/pathogens	Shared TRB sequences
Influenza A	1,032
Cytomegalovirus (CMV)	765
Epstein-Barr Virus (EBV)	640
Human immunodeficiency virus (HIV)	577
Yellow fever virus (YFV)	177
Human T-cell lymphotropic virus (HTLV)	131
Hepatitis C virus (HCV)	52
Mycobacterium tuberculosis	22
Herpes simplex virus 2 (HSV2)	16
COVID-19	2

Table 5.2 *P*-values for the overrepresentation CDR3 sequence analysis for all the tested diseases collected in the McPAS-TCR database[278].

McPAS-TCR database			
Disease association	control	without MHE	with MHE
Alzheimer's disease	1	1	1
Celiac disease	0.011*	0.214	4.836×10^{-4}**
Churg Strauss syndrome	1	1	1
COVID-19	1	1	1
Cytomegalovirus (CMV)	1	1	0.723
Diabetes Type 1	1	1	0.510
Diabetes type 2	1	1	1
Epstein Barr virus (EBV)	1	0.510	1
Hepatitis C virus	1	1	1
Hepatitis E virus infection (cHEV)	1	1	1
Herpes simplex virus 2 (HSV2)	1	1	1
HTLV-1	1	0.877	0.823
Human herpes virus 1	1	1	1

Human immunodeficiency virus (HIV)	1	1	1
Human papilloma virus	1	1	1
IgG4-related disease	1	1	1
Inflammatory bowel disease (IBD)	1	3.082×10^{-5}**	5.470×10^{-4}**
Influenza	0.510	0.510	0.066
Mycobacterium tuberculosis	1	1	1
Multiple sclerosis	1	1	1
Myasthenia gravis	1	1	1
Narcolepsy	0.723	1	1
Parkinson disease	1	1	1
Psoriatic arthritis	0.064	1	1
Rheumatoid arthritis	1	1	1
Sarcoidosis	1	1	1
Ulcerative colitis	1	0.066	1
Yellow fever virus	1	0.723	1

P-values were adjusted for multiple testing using Benjamini-Hochberg FDR correction considering both number of diseases and number of sample groups. HTLV1, human T-cell lymphotropic virus-1; IgG4, immunoglobulin G4; **p*-value < 0.05; ***p*-value < 0.01

We found that inflammatory bowel disease (IBD) was significantly overrepresented both in patients with cirrhosis without MHE (*p*-value = 3.082×10^{-5}) and in those with MHE (*p*-value = 5.470×10^{-4}). Celiac disease was also significant in the controls (*p*-value = 1.127×10^{-2}) and in patients with cirrhosis without MHE (*p*-value = 4.836×10^{-4}). There was no significant overlap (*p*-value > 0.05) between our dataset and the VDJdb associated CDR3βs (Table 5.3). While the overlap studies of bulk sequencing data with antigen-specific data are interesting, results remain challenging to interpret, as it is unclear why specific antigens are enriched. The polyreactivity of the TCR repertoire might be a reason. Specifically, each TCR maps to several antigen specificities, or in other words, there is no one-to-one TCR-antigen map, only a many-to-many. A similar

association of bulk sequencing data with seemingly unrelated antigens have been observed in recent publications[280,289].

Table 5.3 *P*-values for the overrepresentation CDR3 sequence analysis for all the pathogens collected in the VDJdb database[279].

VDJdb database			
Antigen specificity	control	without MHE	with MHE
CMV	1	0.373	0.677
DENV1	1	1	1
DENV2	1	0.959	1
DENV3/4	1	1	1
EBV	1	0.524	0.677
HCV	1	1	1
HIV-1	1	1	1
Homo Sapiens	0.816	1	0.816
HPV	1	1	1
HSV2	1	1	1
HTLV1	1	1	1
Influenza A	1	1	0.816
MCPyV	1	1	1
Mycobacterium tuberculosis	1	0.816	1
Saccharomyces cerevisiae	1	1	1
SARS-CoV-2	1	1	1
Streptomyces kanamyceticus	1	1	1
synthetic	1	1	1
Triticum aestivum	1	1	1
YFV	0.816	1	1

The *p*-values were adjusted for multiple testing using Benjamini–Hochberg FDR correction considering both the number of diseases and the number of sample groups. CMV, cytomegalovirus; DENV, Dengue virus; EBV, Epstein Barr virus; HCV, hepatitis C virus; HIV, human immunodeficiency virus; HPV, human papilloma virus; HSV2, herpes simplex virus 2; HTLV1, human T-cell lymphotropic virus-1; MCPyV, Merkel cell polyomavirus; SARS-CoV-2, severe acute respiratory syndrome coronavirus; YFV, yellow fever virus.

5.4. Conclusion

We present a step-by-step computational pipeline for the processing of immune repertoire data from whole transcriptome RNA-seq reads that leverages the presence of immunological receptor sequences (TCRs) extracted from RNA-seq transcriptomic datasets. To the best of our knowledge, this is the first pipeline that includes both TCR extraction from RNA-seq data as well as a complete immune repertoire data analysis.

Different repertoire features can be calculated to interpret the immune repertoire variation. Repertoire overlap analysis is the most common approach to uncover shared clonotypes between individuals, also known as 'public' clones[294]. Diversity measurement helps understand the expansion of individual T-cell clones[122]. Repertoire architecture is represented by receptor sequence likeness, which determines the adaptive immune response and can be quantified both by k-mers and network analysis[280]. Finally, evaluating the overrepresentation of immune receptors with known pathological associations in patient immune repertoires can guide the assessment of cross-reactivity with other immunological conditions[289].

We have presented here a case study where TCR repertoire profiles were recovered from bulk RNA-seq data from $CD4^+$ T cells isolated from control patients, and patients with cirrhosis, either without or with MHE. A total of 498–1,114 distinct clones (i.e., CDR3β amino acid sequences) per individual were reconstructed using **MiXCR**. The authors of this tool have shown a similar magnitude yield using 100 bp paired-end RNA-seq data using ileocecal lymph node metastasis samples (around 3,000 recovered TRBs), small intestine resection samples (around 150 TRBs), or isolated $CD4^+$ T cells from the spleen (1,684–2,977 TRBs) or central nervous

system (367–936 TRBs)[272]. In this work, the range of clones we obtained from isolated CD4$^+$ T cells of human blood samples was halfway between those previously obtained from spleen and central nervous system CD4$^+$ cells, thereby indicating that the extraction of clonotypes from our RNA-seq dataset was adequate.

Based on our clonotype data analysis, we found that the immune repertoires of our 3 groups of patients were very similar. The high similarity could have either a biological or technological origin. Three main different reasons can be suggested. (1) The fact that repertoire changes to immune perturbations are more subtle than previously thought. This is in line with recent results on larger cohorts. Specifically, our findings suggest that for autoimmune diseases, the immune signal is very weak if not isolated by cell type for example[295]. (2) Another reason could be the low sample number. However, we have previously shown that even large sample numbers may not lead to separation between patient classes unless specific machine learning algorithms and feature encoding are used[296]. (3) It may also be that the proposed features (repertoire overlap, repertoire diversity, network analysis, k-mer, and comparison with existing databases) do not capture the full biological heterogeneity of TCR repertoires. However, we have recently shown that these features cover a large part of immune repertoire diversity[295]. Our pipeline can be applied to any bulk RNA-seq dataset obtained from a sample containing T cells, thanks to the Nextflow implementation. We believe that is a useful resource to study the immune repertoire similarity landscape across different biological scenarios (e.g., health, disease, autoimmunity, infection, or vaccination).

Taking advantage of the generation of a vast number of RNA-seq datasets that have become available for different immune cell populations over the

last few decades, our Nextflow pipeline can be applied to study TCR repertoires to understand patient immune statuses in the contexts of multiple diseases. The current version of this pipeline is tailored to the study of T-cell subtypes ($CD8^+$ and $CD4^+$ subpopulations) and it can be easily adapted to the study of B cells. Additionally, it only supports 2 input species for the moment (*Homo sapiens* and *Mus musculus*), but they can be expanded as soon as the information on antigen/disease-specific TCR databases increases.

However, one limitation is that only the most abundant and expanded clones can be recovered using bulk RNA-seq data for immune repertoire quantification and specific α/β receptor chain pairings cannot be determined. Nonetheless, since bulk RNA-seq datasets combine dual biological information measured in 1 single experiment, the Nextflow pipeline will allow for parallel analysis of immune repertoires and gene expression. In the previous chapter, we performed the gene expression analysis of the RNA-seq dataset used here, in which the integration of both RNA-seq and miRNA-seq datasets from $CD4^+$ T-cells of our MHE patient cohort were analyzed. Formal integration of TCR and gene expression analysis results will require the development of adequate mathematical methods that are able to deal with the different structure of both datasets[295].

Chapter 6

General discussion and conclusions

6.1. General discussion

This thesis focused on understanding the immunological changes associated with the appearance of MHE in patients with cirrhosis. The analysis of human blood samples is minimally invasive for patients and it allows researchers to collect information from the cells and molecules circulating in each patient at a given time. These samples can then be subjected to a variety of high-throughput molecular profiling techniques that can measure gene activity, metabolite concentrations, and circulating protein levels. However, the biological interpretation of such data is not trivial because neither the origin nor the destination of each molecule is known. On the one hand, the immune system contains unique genes that can serve as markers to distinguish between different cell types (e.g., $CD4^+$ and $CD8^+$, among others) and can be used to infer the presence and activity of different blood cell types. On the other hand, alternative strategies such as fluorescent activated cell sorting or single cell methods are required to understand the ongoing processes of each cellular subpopulation.

Here we addressed the study of peripheral inflammation in patients with cirrhosis with and without MHE, both in a general and in a cell-type specific manner. Chapter 3 was a preliminary study using whole blood which included a mix of several cell types. This had the advantage of providing a global overview of the immune system in these groups but the disadvantage of being unable to differentiate between cell populations except for specific marker genes. In this work we characterized changes in the gene expression, metabolites, and cytokines associated with the onset of MHE. Moreover, we presented a novel analysis pipeline for the integration of multiple molecular assays performed on blood samples,

showing the joint contribution of several types of molecules to a disease phenotype.

It should be noted that establishing a case-effect among identified biomarkers was difficult because our approach searched for co-variation patterns, which do not always imply causation. However, we did link extra- and intracellular components of the immune system that were altered in patients with MHE and proposed a model for their participation in biological processes that could contribute to the onset of MHE. Additionally, as described in Chapters 4 and 5, the isolation of $CD4^+$ lymphocytes allowed the study of the specific molecular alterations in this cell subtype without the confounding effects from other immune system cell types. In Chapter 4, we identified mechanisms in $CD4^+$ cells by which the altered levels of miRNAs and transcription factors may contribute to the immune system shift that triggers MHE, whereas in Chapter 5 we presented an analytical pipeline implemented in Nextflow for the characterization of T-cell receptors in our patients.

These 3 studies share the limitation of operating with a low number of source samples because of our limited access to patients at only 2 hospitals. Thus, the creation of a consortium for the study of MHE that includes the participation of multiple clinical centers will be required to recruit a large cohort of patients. More sophisticated statistical approaches such as deep or machine learning could also be applied to patient classification and biomarker prediction in studies with larger cohort sizes. The methods we applied in this thesis are suitable for use both in the fields of multi-omics and computational immunology when the analysis of a large number of samples is impossible.

6.2. Conclusions

- This study integrated multiple molecular assays from blood samples to identify immune system pathways that are potentially involved in the induction of MHE.
- In whole blood samples we detected biological pathways that corroborate the decrease in Th2 and increase in Th1 and Th17 observed in patients with MHE, while other activated pathways suggested B-cell upregulation and Tc-cell downregulation.
- Lipid metabolism was described here for the first time in the context of mild cognitive impairment in MHE.
- Specific sets of serum metabolites and cytokines were identified that pointed towards a chemotactic function or structural lipid patterns which act as mediators of MHE induction.
- Our integrative modelling suggested a link between the CCL20, CX3CL1, CXCL13, IL-15, IL-22, and IL-6 cytokines and the alteration of chemotactic receptors (CCR2, CXCR3, and CMKLR1) and ligands (CCL5, CXCL5, PF4, PF4V1, and MSMP) in MHE. This suggested a link between the production and reception of inflammatory signals operating in these patients.
- PLS analysis indicated that long-chain unsaturated PCs, increased fatty acid transport, and prostaglandin production were strongly linked in patients with MHE, implying a possible pathway for the dysregulation of structural phospholipids during mild cognitive decline.

- Our results illustrate the power of the integrative statistical analysis of multi-omics in modelling disease processes and connecting phenotypic changes across molecular layers.
- Altered pathways in $CD4^+$ T cells that may contribute to triggering the appearance of MHE are:
    - The toll-like receptors and IL-17 signaling pathways, which are directly involved in modulating immune system responses.
    - The histidine and tryptophan metabolism pathways, which also modulate immune system function.
- mRNA–miRNA integration analysis highlighted 4 miRNAs that may modulate the expression of transcripts coding for key proteins involved in the onset of MHE.
- Upregulation of miR-494-3p was linked with the downregulation of FOS/JUN, which may explain the paradoxical observation whereby early elements of the pathways were upregulated while the final products tended to be downregulated.
- Discrepancies in the transcript and protein levels of the most inflammatory cytokines (e.g., TNFα and IL-1β) could perhaps be explained by the upregulation of miR-494-3p or miR-127-5p.
- Modulation of proteins such as A20 (*TNFAIP3*) and TTP (*ZFP36*) by increased levels of miR-494-3p, miR-656-3p, and miR-130b-3p in $CD4^+$ T lymphocytes would contribute to changes in signal transduction pathways, resulting in increased pro-inflammatory cytokines such as IL-17 and TNFα.

- The immune TCR repertoires of controls, patients with cirrhosis with and without MHE were very similar, which might indicate that repertoire changes leading to immune perturbations are more subtle than previously thought.

- We have developed a publicly available Nextflow pipeline that include both TCR extraction from RNA-seq data and a complete step-by-step immune repertoire data analysis, applicable in the context of multiple diseases.

- This work could help researchers to develop new diagnostic and therapeutic approaches for patients with MHE.

6.3. Future perspectives

The synergetic role between peripheral inflammation and hyperammonemia is the main factor contributing to the cerebral alterations present in MHE. Thus, our neurobiology laboratory is currently engaged in different lines of work designed to study the molecular mechanisms occurring in the immune system and brain, both in rat models and patients. This synergic project between our neurobiology and bioinformatics laboratories is contributing to the study of this neuropsychiatric syndrome by employing multi-omic approaches and developing bioinformatic pipelines applicable to other research areas or disease studies.

Some future lines of research in the field of peripheral inflammation from the multi-omic standpoint includes:

1. Analysis of more immune system cell subtypes. Given that $CD4^+$ lymphocytes play a key role in MHE, the isolation of different subtypes of these cells in these patients would be useful to

understand the alterations in more specific pathways in each T helper cell subtype (e.g., Th1, Th2, Th17, etc.) without confounding effects. Together with $CD4^+$ cells, the activation of subtypes of monocytes was also higher in patients with MHE and so the isolation of these populations could help researchers understand their molecular changes between different groups of patients. In addition, although the immunophenotype study using flow cytometry showed no significant changes for B lymphocytes and $CD8^+$ lymphocytes, a study of the transcriptomic profile of these subtypes after their isolation could reveal intracellular changes (such as the few we observed in Chapter 3, even with the handicap of analyzing multiple subtypes together). Finally, the role of the innate immunity in MHE still remains unknown even though populations such as neutrophils are the most abundant in human blood and other cells like natural killer cells or dendritic cells are essential elements for the immune response. Thus, the same single-omic analysis methodologies used here, such as differential expression and functional enrichment analysis, could also be applied to different immune system populations in future work.

2. The study of intercellular communication in the immune system. Application of the pipelines we developed here to different isolated cell subtypes (e.g., monocytes, B cells, $CD8^+$, Th1, Th2, and Th17 cells, etc.) could help the discovery of potential associations between extracellular elements (e.g., metabolites and cytokines) and a wide range of genes, separated by each cell population. Gene expression regulation can also be studied inside cells, as we did here using miRNA-seq and RNA-seq, but additional datasets such as ChIP-seq, ATAC-seq, or even Methyl-seq would be useful for measuring DNA binding sites and chromatin accessibility for

transcription factors as well as epigenetic marks. Additionally, more sophisticated models such as multi-block PLS or PLS path modeling, would allow the simultaneous multivariate integration of all this information to find possible interactions not previously described in databases and which could serve as good indicators of biological relationships.

3. A large-scale study of immune system repertoires. Taking advantage of the generation of multiple RNA-seq datasets for different immune cell populations, our pipeline could be applied to study TCR repertoires in other diseases. The current version of this pipeline is useful for studying T-cell subtypes ($CD8^+$ and $CD4^+$ subpopulations) but in the future it could easily be modified to study B cells.

4. Discovering the antigen(s) that trigger alterations in the immune response of patients with MHE. Our work has unraveled some immune system alterations that occur in patients with MHE. We also identified possible mechanisms that may be responsible for the alteration of cerebral functions because of the immune system shift present in patients with cirrhosis with MHE. However, it is not yet clear what mechanism triggers of all these alterations. Thus, antigen discovery for adaptive immunity is a branch of computational immunology that could be helpful in solving this problem.

References

1. Moon, A. M., Singal, A. G., Tapper, E. B. (2020). Contemporary Epidemiology of Chronic Liver Disease and Cirrhosis. *Clinical Gastroenterology and Hepatology*, *18*(12), 2650-2666. https://doi.org/10.1016/j.cgh.2019.07.060
2. Felipo, V. (2013). Hepatic encephalopathy: Effects of liver failure on brain function. *Nature Reviews Neuroscience*, *14*(12), 851-858. https://doi.org/10.1038/nrn3587
3. Vilstrup, H., Amodio, P., Bajaj, J., Cordoba, J., Ferenci, P., Mullen, K. D., Weissenborn, K., Wong, P. (2014). Hepatic encephalopathy in chronic liver disease: 2014 Practice Guideline by the American Association for the Study Of Liver Diseases and the European Association for the Study of the Liver: Vilstrup et al. *Hepatology*, *60*(2), 715-735. https://doi.org/10.1002/hep.27210
4. Weissenborn, K., Ennen, J. C., Schomerus, H., Rückert, N., Hecker, H. (2001). Neuropsychological characterization of hepatic encephalopathy. *Journal of Hepatology*, *34*(5), 768-773. https://doi.org/10.1016/S0168-8278(01)00026-5
5. Weissenborn, K., Giewekemeyer, K., Heidenreich, S., Bokemeyer, M., Berding, G., Ahl, B. (2005). Attention, Memory, and Cognitive Function in Hepatic Encephalopathy. *Metabolic Brain Disease*, *20*(4), 359-367. https://doi.org/10.1007/s11011-005-7919-z
6. Román, E., Córdoba, J., Torrens, M., Torras, X., Villanueva, C., Vargas, V., Guarner, C., Soriano, G. (2011). Minimal Hepatic Encephalopathy Is Associated With Falls: *American Journal of Gastroenterology*, *106*(3), 476-482. https://doi.org/10.1038/ajg.2010.413
7. Formentin, C., De Rui, M., Zoncapè, M., Ceccato, S., Zarantonello, L., Senzolo, M., Burra, P., Angeli, P., Amodio, P., Montagnese, S. (2019). The psychomotor vigilance task: Role in the diagnosis of hepatic encephalopathy and relationship with driving ability. *Journal of Hepatology*, *70*(4), 648-657. https://doi.org/10.1016/j.jhep.2018.12.031
8. Ferenci, P., Lockwood, A., Mullen, K., Tarter, R., Weissenborn, K., Blei, A. T. (2002). Hepatic encephalopathy-Definition, nomenclature, diagnosis, and quantification: Final report of the Working Party at the 11th World Congresses of Gastroenterology, Vienna, 1998. *Hepatology*, *35*(3), 716-721. https://doi.org/10.1053/jhep.2002.31250
9. Giménez-Garzó, C., Garcés, J. J., Urios, A., Mangas-Losada, A., García-García, R., González-López, O., Giner-Durán, R., Escudero-García, D., Serra, M. A., Soria, E., Felipo, V., Montoliu, C. (2017). The PHES battery does not detect all cirrhotic patients with early neurological deficits, which are different in different patients. *PLOS ONE*, *12*(2), e0171211. https://doi.org/10.1371/journal.pone.0171211
10. Luo, M. (2015). Inflammation: A novel target of current therapies for hepatic encephalopathy in liver cirrhosis. *World Journal of Gastroenterology*, *21*(41), 11815. https://doi.org/10.3748/wjg.v21.i41.11815

11. McPhail, M. J. W., Bajaj, J. S., Thomas, H. C., Taylor-Robinson, S. D. (2010). Pathogenesis and diagnosis of hepatic encephalopathy. *Expert Review of Gastroenterology Hepatology*, *4*(3), 365-378. https://doi.org/10.1586/egh.10.32
12. Cabrera-Pastor, A., Llansola, M., Montoliu, C., Malaguarnera, M., Balzano, T., Taoro-Gonzalez, L., García-García, R., Mangas-Losada, A., Izquierdo-Altarejos, P., Arenas, Y. M., Leone, P., Felipo, V. (2019). Peripheral inflammation induces neuroinflammation that alters neurotransmission and cognitive and motor function in hepatic encephalopathy: Underlying mechanisms and therapeutic implications. *Acta Physiologica*, *226*(2), e13270. https://doi.org/10.1111/apha.13270
13. Montoliu, C., Llansola, M., Felipo, V. (2015). Neuroinflammation and neurological alterations in chronic liver diseases. *Neuroimmunology and Neuroinflammation*, *2*(3), 138. https://doi.org/10.4103/2347-8659.160845
14. Mangas-Losada, A., García-García, R., Urios, A., Escudero-García, D., Tosca, J., Giner-Durán, R., Serra, M. A., Montoliu, C., Felipo, V. (2017). Minimal hepatic encephalopathy is associated with expansion and activation of CD4+CD28-, Th22 and Tfh and B lymphocytes. *Scientific Reports*, *7*(1), 6683. https://doi.org/10.1038/s41598-017-05938-1
15. Balzano, T., Forteza, J., Molina, P., Giner, J., Monzó, A., Sancho-Jiménez, J., Urios, A., Montoliu, C., Felipo, V. (2018). The Cerebellum of Patients with Steatohepatitis Shows Lymphocyte Infiltration, Microglial Activation and Loss of Purkinje and Granular Neurons. *Scientific Reports*, *8*(1), 3004. https://doi.org/10.1038/s41598-018-21399-6
16. Felipo, V., Butterworth, R. F. (2002). Neurobiology of ammonia. *Progress in Neurobiology*, *67*(4), 259-279. https://doi.org/10.1016/s0301-0082(02)00019-9
17. Hernández-Rabaza, V., Cabrera-Pastor, A., Taoro-González, L., Malaguarnera, M., Agustí, A., Llansola, M., Felipo, V. (2016). Hyperammonemia induces glial activation, neuroinflammation and alters neurotransmitter receptors in hippocampus, impairing spatial learning: Reversal by sulforaphane. *Journal of Neuroinflammation*, *13*(1), 41. https://doi.org/10.1186/s12974-016-0505-y
18. Felipo, V., Urios, A., Montesinos, E., Molina, I., Garcia-Torres, M. L., Civera, M., Olmo, J. A. D., Ortega, J., Martinez-Valls, J., Serra, M. A., Cassinello, N., Wassel, A., Jordá, E., Montoliu, C. (2012). Contribution of hyperammonemia and inflammatory factors to cognitive impairment in minimal hepatic encephalopathy. *Metabolic brain disease*, *27*(1), 51-58. https://doi.org/10.1007/s11011-011-9269-3
19. Shawcross, D. L., Davies, N. A., Williams, R., Jalan, R. (2004). Systemic inflammatory response exacerbates the neuropsychological effects of induced hyperammonemia in cirrhosis. *Journal of Hepatology*, *40*(2), 247-254. https://doi.org/10.1016/j.jhep.2003.10.016
20. Shawcross, D. L., Wright, G., Olde Damink, S. W. M., Jalan, R. (2007). Role of ammonia and inflammation in minimal hepatic encephalopathy. *Metabolic Brain Disease*, *22*(1), 125-138. https://doi.org/10.1007/s11011-006-9042-1

21. Montoliu, C., Piedrafita, B., Serra, M. A., del Olmo, J. A., Urios, A., Rodrigo, J. M., Felipo, V. (2009). IL-6 and IL-18 in blood may discriminate cirrhotic patients with and without minimal hepatic encephalopathy. *Journal of Clinical Gastroenterology*, *43*(3), 272-279. https://doi.org/10.1097/MCG.0b013e31815e7f58
22. Hernandez-Rabaza, V., Agusti, A., Cabrera-Pastor, A., Fustero, S., Delgado, O., Taoro-Gonzalez, L., Montoliu, C., Llansola, M., Felipo, V. (2015). Sildenafil reduces neuroinflammation and restores spatial learning in rats with hepatic encephalopathy: Underlying mechanisms. *Journal of Neuroinflammation*, *12*(1), 195. https://doi.org/10.1186/s12974-015-0420-7
23. Cabrera-Pastor, A., Hernandez-Rabaza, V., Taoro-Gonzalez, L., Balzano, T., Llansola, M., Felipo, V. (2016). In vivo administration of extracellular cGMP normalizes TNF-α and membrane expression of AMPA receptors in hippocampus and spatial reference memory but not IL-1β, NMDA receptors in membrane and working memory in hyperammonemic rats. *Brain, Behavior, and Immunity*, *57*, 360-370. https://doi.org/10.1016/j.bbi.2016.05.011
24. Taoro-Gonzalez, L., Arenas, Y. M., Cabrera-Pastor, A., Felipo, V. (2018). Hyperammonemia alters membrane expression of GluA1 and GluA2 subunits of AMPA receptors in hippocampus by enhancing activation of the IL-1 receptor: Underlying mechanisms. *Journal of Neuroinflammation*, *15*(1), 36. https://doi.org/10.1186/s12974-018-1082-z
25. Hernandez-Rabaza, V., Cabrera-Pastor, A., Taoro-Gonzalez, L., Gonzalez-Usano, A., Agusti, A., Balzano, T., Llansola, M., Felipo, V. (2016). Neuroinflammation increases GABAergic tone and impairs cognitive and motor function in hyperammonemia by increasing GAT-3 membrane expression. Reversal by sulforaphane by promoting M2 polarization of microglia. *Journal of Neuroinflammation*, *13*(1), 83. https://doi.org/10.1186/s12974-016-0549-z
26. Gonzalez-Usano, A., Cauli, O., Agusti, A., Felipo, V. (2014). Pregnenolone Sulfate Restores the Glutamate-Nitric-Oxide-cGMP Pathway and Extracellular GABA in Cerebellum and Learning and Motor Coordination in Hyperammonemic Rats. *ACS Chemical Neuroscience*, *5*(2), 100-105. https://doi.org/10.1021/cn400168y
27. Cabrera-Pastor, A., Malaguarnera, M., Taoro-Gonzalez, L., Llansola, M., Felipo, V. (2016). Extracellular cGMP Modulates Learning Biphasically by Modulating Glycine Receptors, CaMKII and Glutamate-Nitric Oxide-cGMP Pathway. *Scientific Reports*, *6*(1), 33124. https://doi.org/10.1038/srep33124
28. Dadsetan, S., Balzano, T., Forteza, J., Cabrera-Pastor, A., Taoro-Gonzalez, L., Hernandez-Rabaza, V., Gil-Perotín, S., Cubas-Núñez, L., García-Verdugo, J. M., Agusti, A., Llansola, M., Felipo, V. (2016). Reducing Peripheral Inflammation with Infliximab Reduces Neuroinflammation and Improves Cognition in Rats with Hepatic Encephalopathy. *Frontiers in Molecular Neuroscience*, *9*. https://doi.org/10.3389/fnmol.2016.00106
29. Dadsetan, S., Balzano, T., Forteza, J., Agusti, A., Cabrera-Pastor, A., Taoro-

Gonzalez, L., Hernandez-Rabaza, V., Gomez-Gimenez, B., ElMlili, N., Llansola, M., Felipo, V. (2016). Infliximab reduces peripheral inflammation, neuroinflammation, and extracellular GABA in the cerebellum and improves learning and motor coordination in rats with hepatic encephalopathy. *Journal of Neuroinflammation*, *13*(1), 245. https://doi.org/10.1186/s12974-016-0710-8

30. Broux, B., Mizee, M. R., Vanheusden, M., van der Pol, S., van Horssen, J., Van Wijmeersch, B., Somers, V., de Vries, H. E., Stinissen, P., Hellings, N. (2015). IL-15 amplifies the pathogenic properties of CD4+CD28- T cells in multiple sclerosis. *The Journal of Immunology*, *194*(5), 2099-2109. https://doi.org/10.4049/jimmunol.1401547
31. Meares, G. P., Ma, X., Qin, H., Benveniste, E. N. (2012). Regulation of CCL20 expression in astrocytes by IL-6 and IL-17. *Glia*, *60*(5), 771-781. https://doi.org/10.1002/glia.22307
32. Grogan, J. L., Ouyang, W. (2012). A role for Th17 cells in the regulation of tertiary lymphoid follicles. *European Journal of Immunology*, *42*(9), 2255-2262. https://doi.org/10.1002/eji.201242656
33. Barone, F., Nayar, S., Campos, J., Cloake, T., Withers, D. R., Toellner, K. M., Zhang, Y., Fouser, L., Fisher, B., Bowman, S., Rangel-Moreno, J., Garcia-Hernandez, M. de la L., Randall, T. D., Lucchesi, D., Bombardieri, M., Pitzalis, C., Luther, S. A., Buckley, C. D. (2015). IL-22 regulates lymphoid chemokine production and assembly of tertiary lymphoid organs. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(35), 11024-11029. https://doi.org/10.1073/pnas.1503315112
34. Fantuzzi, L., Tagliamonte, M., Gauzzi, M. C., Lopalco, L. (2019). Dual CCR5/CCR2 targeting: Opportunities for the cure of complex disorders. *Cellular and Molecular Life Sciences*, *76*(24), 4869-4886. https://doi.org/10.1007/s00018-019-03255-6
35. Stuart, M. J., Baune, B. T. (2014). Chemokines and chemokine receptors in mood disorders, schizophrenia, and cognitive impairment: A systematic review of biomarker studies. *Neuroscience and Biobehavioral Reviews*, *42*, 93-115. https://doi.org/10.1016/j.neubiorev.2014.02.001
36. Neuman, M. G., Maor, Y., Nanau, R. M., Melzer, E., Mell, H., Opris, M., Cohen, L., Malnick, S. (2015). Alcoholic Liver Disease: Role of Cytokines. *Biomolecules*, *5*(3), 2023-2034. https://doi.org/10.3390/biom5032023
37. Kolios, G., Valatas, V., Kouroumalis, E. (2006). Role of Kupffer cells in the pathogenesis of liver disease. *World Journal of Gastroenterology*, *12*(46), 7413-7420. https://doi.org/10.3748/wjg.v12.i46.7413
38. Lario, M., Muñoz, L., Ubeda, M., Borrero, M. J., Martínez, J., Monserrat, J., Díaz, D., Alvarez-Mon, M., Albillos, A. (2013). Defective thymopoiesis and poor peripheral homeostatic replenishment of T-helper cells cause T-cell lymphopenia in cirrhosis. *Journal of Hepatology*, *59*(4), 723-730. https://doi.org/10.1016/j.jhep.2013.05.042
39. Yonkers, N. L., Sieg, S., Rodriguez, B., Anthony, D. D. (2011). Reduced naive CD4 T cell numbers and impaired induction of CD27 in response to T cell receptor stimulation reflect a state of immune activation in chronic

hepatitis C virus infection. *The Journal of Infectious Diseases*, *203*(5), 635-645. https://doi.org/10.1093/infdis/jiq101
40. Girón-González, J. A., Martínez-Sierra, C., Rodriguez-Ramos, C., Macías, M. A., Rendón, P., Díaz, F., Fernández-Gutiérrez, C., Martín-Herrera, L. (2004). Implication of inflammation-related cytokines in the natural history of liver cirrhosis. *Liver International: Official Journal of the International Association for the Study of the Liver*, *24*(5), 437-445. https://doi.org/10.1111/j.1478-3231.2004.0951.x
41. Hammerich, L., Tacke, F. (2014). Interleukins in chronic liver disease: Lessons learned from experimental mouse models. *Clinical and Experimental Gastroenterology*, *7*, 297-306. https://doi.org/10.2147/CEG.S43737
42. Lan, R. Y. Z., Salunga, T. L., Tsuneyama, K., Lian, Z. X., Yang, G. X., Hsu, W., Moritoki, Y., Ansari, A. A., Kemper, C., Price, J., Atkinson, J. P., Coppel, R. L., Gershwin, M. E. (2009). Hepatic IL-17 responses in human and murine primary biliary cirrhosis. *Journal of Autoimmunity*, *32*(1), 43-51. https://doi.org/10.1016/j.jaut.2008.11.001
43. Yang, L., Zhang, Y., Wang, L., Fan, F., Zhu, L., Li, Z., Ruan, X., Huang, H., Wang, Z., Huang, Z., Huang, Y., Yan, X., Chen, Y. (2010). Amelioration of high fat diet induced liver lipogenesis and hepatic steatosis by interleukin-22. *Journal of Hepatology*, *53*(2), 339-347. https://doi.org/10.1016/j.jhep.2010.03.004
44. Jiang, R., Tan, Z., Deng, L., Chen, Y., Xia, Y., Gao, Y., Wang, X., Sun, B. (2011). Interleukin-22 promotes human hepatocellular carcinoma by activation of STAT3. *Hepatology*, *54*(3), 900-909. https://doi.org/10.1002/hep.24486
45. Yeh, S. H., Chen, P. J. (2010). Gender disparity of hepatocellular carcinoma: The roles of sex hormones. *Oncology*, *78*(Suppl 1), 172-179. https://doi.org/10.1159/000315247
46. Streetz, K. L., Tacke, F., Leifeld, L., Wüstefeld, T., Graw, A., Klein, C., Kamino, K., Spengler, U., Kreipe, H., Kubicka, S., Müller, W., Manns, M. P., Trautwein, C. (2003). Interleukin 6/gp130-dependent pathways are protective during chronic liver diseases. *Hepatology*, *38*(1), 218-229. https://doi.org/10.1053/jhep.2003.50268
47. Albillos, A., Lario, M., Álvarez-Mon, M. (2014). Cirrhosis-associated immune dysfunction: Distinctive features and clinical relevance. *Journal of Hepatology*, *61*(6), 1385-1396. https://doi.org/10.1016/j.jhep.2014.08.010
48. Williams, R. (2007). Review article: Bacterial flora and pathogenesis in hepatic encephalopathy. *Alimentary Pharmacology Therapeutics*, *25*, 17-22. https://doi.org/10.1111/j.1746-6342.2006.03217.x
49. Jain, L., Sharma, B. C., Sharma, P., Srivastava, S., Agrawal, A., Sarin, S. K. (2012). Serum endotoxin and inflammatory mediators in patients with cirrhosis and hepatic encephalopathy. *Digestive and Liver Disease*, *44*(12), 1027-1031. https://doi.org/10.1016/j.dld.2012.07.002
50. Bajaj, J. S., Fagan, A., White, M. B., Wade, J. B., Hylemon, P. B., Heuman, D. M., Fuchs, M., John, B. V., Acharya, C., Sikaroodi, M., Gillevet, P. M. (2019). Specific Gut and Salivary Microbiota Patterns Are Linked With

Different Cognitive Testing Strategies in Minimal Hepatic Encephalopathy. *The American Journal of Gastroenterology*, *114*(7), 1080-1090. https://doi.org/10.14309/ajg.0000000000000102
51. Bajaj, J. S., Hylemon, P. B., Ridlon, J. M., Heuman, D. M., Daita, K., White, M. B., Monteith, P., Noble, N. A., Sikaroodi, M., Gillevet, P. M. (2012). Colonic mucosal microbiome differs from stool microbiome in cirrhosis and hepatic encephalopathy and is linked to cognition and inflammation. *American Journal of Physiology. Gastrointestinal and Liver Physiology*, *303*(6), G675-685. https://doi.org/10.1152/ajpgi.00152.2012
52. Zhang, Z., Zhai, H., Geng, J., Yu, R., Ren, H., Fan, H., Shi, P. (2013). Large-scale survey of gut microbiota associated with MHE Via 16S rRNA-based pyrosequencing. *The American Journal of Gastroenterology*, *108*(10), 1601-1611. https://doi.org/10.1038/ajg.2013.221
53. Kennedy, P. J., Cryan, J. F., Dinan, T. G., Clarke, G. (2017). Kynurenine pathway metabolism and the microbiota-gut-brain axis. *Neuropharmacology*, *112*(Pt B), 399-412. https://doi.org/10.1016/j.neuropharm.2016.07.002
54. Dalile, B., Van Oudenhove, L., Vervliet, B., Verbeke, K. (2019). The role of short-chain fatty acids in microbiota-gut-brain communication. *Nature Reviews. Gastroenterology Hepatology*, *16*(8), 461-478. https://doi.org/10.1038/s41575-019-0157-3
55. Smith, P. M., Howitt, M. R., Panikov, N., Michaud, M., Gallini, C. A., Bohlooly-Y, M., Glickman, J. N., Garrett, W. S. (2013). The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. *Science*, *341*(6145), 569-573. https://doi.org/10.1126/science.1241165
56. Mukherjee, S., John, S. (2021). Lactulose. *StatPearls [Internet]*. Available from: http://www.ncbi.nlm.nih.gov/books/NBK536930/
57. Reinert, J. P., Burnham, K. (2020). Non-Lactulose Medication Therapies for the Management of Hepatic Encephalopathy: A Literature Review. *Journal of Pharmacy Practice*, 089719002095302. https://doi.org/10.1177/0897190020953024
58. Dhiman, R. K., Thumburu, K. K., Verma, N., Chopra, M., Rathi, S., Dutta, U., Singal, A. K., Taneja, S., Duseja, A., Singh, M. (2020). Comparative Efficacy of Treatment Options for Minimal Hepatic Encephalopathy: A Systematic Review and Network Meta-Analysis. *Clinical Gastroenterology and Hepatology*, *18*(4), 800-812.e25. https://doi.org/10.1016/j.cgh.2019.08.047
59. McComb, S., Thiriot, A., Akache, B., Krishnan, L., Stark, F. (2019). Introduction to the Immune System. *Immunoproteomics, 2024*, 1-24. https://doi.org/10.1007/978-1-4939-9597-4_1
60. Maciolek, J. A., Pasternak, J. A., Wilson, H. L. (2014). Metabolism of activated T lymphocytes. *Current Opinion in Immunology*, *27*, 60-74. https://doi.org/10.1016/j.coi.2014.01.006
61. de la Fuente, H., Cibrián, D., Sánchez-Madrid, F. (2012). Immunoregulatory molecules are master regulators of inflammation during the immune response. *FEBS Letters*, *586*(18), 2897-2905. https://doi.org/10.1016/j.febslet.2012.07.032
62. Mommert, S., Gschwandtner, M., Koether, B., Gutzmer, R., Werfel, T. (2012).

Human memory Th17 cells express a functional histamine H4 receptor. *The American Journal of Pathology*, *180*(1), 177-185. https://doi.org/10.1016/j.ajpath.2011.09.028
63. Zhang, J. M., An, J. (2007). Cytokines, inflammation, and pain. *International Anesthesiology Clinics*, *45*(2), 27-37. https://doi.org/10.1097/AIA.0b013e318034194e
64. Borish, L. C., Steinke, J. W. (2003). 2. Cytokines and chemokines. *The Journal of Allergy and Clinical Immunology*, *111*(Suppl 2), S460-475. https://doi.org/10.1067/mai.2003.108
65. Hughes, C. E., Nibbs, R. J. B. (2018). A guide to chemokines and their receptors. *The FEBS Journal*, *285*(16), 2944-2971. https://doi.org/10.1111/febs.14466
66. Straub, R. H., Schradin, C. (2016). Chronic inflammatory systemic diseases: An evolutionary trade-off between acutely beneficial but chronically harmful programs. *Evolution, Medicine, and Public Health*, *2016*(1), 37-51. https://doi.org/10.1093/emph/eow001
67. Chen, L., Deng, H., Cui, H., Fang, J., Zuo, Z., Deng, J., Li, Y., Wang, X., Zhao, L. (2018). Inflammatory responses and inflammation-associated diseases in organs. *Oncotarget*, *9*(6), 7204-7218. https://doi.org/10.18632/oncotarget.23208
68. Lefranc, M. P. (2014). Immunoglobulin and T Cell Receptor Genes: IMGT(®) and the Birth and Rise of Immunoinformatics. *Frontiers in Immunology*, *5*, 22. https://doi.org/10.3389/fimmu.2014.00022
69. Lo Presti, E., Dieli, F., Meraviglia, S. (2014). Tumor-Infiltrating γδ T Lymphocytes: Pathogenic Role, Clinical Significance, and Differential Programing in the Tumor Microenvironment. *Frontiers in Immunology*, *5*, 607. https://doi.org/10.3389/fimmu.2014.00607
70. Lefranc, M. P., Giudicelli, V., Ginestoux, C., Jabado-Michaloud, J., Folch, G., Bellahcene, F., Wu, Y., Gemrot, E., Brochet, X., Lane, J., Regnier, L., Ehrenmann, F., Lefranc, G., Duroux, P. (2009). IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Research*, *37*(Database issue), D1006-1012. https://doi.org/10.1093/nar/gkn838
71. Kumar, B. V., Connors, T. J., Farber, D. L. (2018). Human T Cell Development, Localization, and Function throughout Life. *Immunity*, *48*(2), 202-213. https://doi.org/10.1016/j.immuni.2018.01.007
72. Reuter, J. A., Spacek, D. V., Snyder, M. P. (2015). High-throughput sequencing technologies. *Molecular Cell*, *58*(4), 586-597. https://doi.org/10.1016/j.molcel.2015.05.004
73. Ma'ayan, A. (2017). Complex systems biology. *Journal of the Royal Society, Interface*, *14*(134), 20170391. https://doi.org/10.1098/rsif.2017.0391
74. Lee, C. K., Klopp, R. G., Weindruch, R., Prolla, T. A. (1999). Gene expression profile of aging and its retardation by caloric restriction. *Science*, *285*(5432), 1390-1393. https://doi.org/10.1126/science.285.5432.1390
75. Perou, C. M., Jeffrey, S. S., van de Rijn, M., Rees, C. A., Eisen, M. B., Ross, D. T., Pergamenschikov, A., Williams, C. F., Zhu, S. X., Lee, J. C., Lashkari, D., Shalon, D., Brown, P. O., Botstein, D. (1999). Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proceedings of the National Academy of Sciences of the United*

States of America, *96*(16), 9212-9217. https://doi.org/10.1073/pnas.96.16.9212
76. International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, *431*(7011), 931-945. https://doi.org/10.1038/nature03001
77. Page, G. P., Zakharkin, S. O., Kim, K., Mehta, T., Chen, L., Zhang, K. (2007). Microarray analysis. *Methods in Molecular Biology*, *404*, 409-430. https://doi.org/10.1007/978-1-59745-530-5_20
78. Liu, P. (2018). MicroRNA Expression Analysis: Next-Generation Sequencing. *Methods in Molecular Biology*, *1783*, 171-183. https://doi.org/10.1007/978-1-4939-7834-2_8
79. Vinayavekhin, N., Saghatelian, A. (2010). Untargeted metabolomics. *Current Protocols in Molecular Biology*, *Chapter 30*, Unit 30 11-24. https://doi.org/10.1002/0471142727.mb3001s90
80. Kim, Y. M., Heyman, H. M. (2018). Mass Spectrometry-Based Metabolomics. *Methods in Molecular Biology*, *1775*, 107-118. https://doi.org/10.1007/978-1-4939-7804-5_10
81. Liu, X., Hoene, M., Wang, X., Yin, P., Häring, H. U., Xu, G., Lehmann, R. (2018). Serum or plasma, what is the difference? Investigations to facilitate the sample material selection decision making process for metabolomics studies and beyond. *Analytica Chimica Acta*, *1037*, 293-300. https://doi.org/10.1016/j.aca.2018.03.009
82. Alves, S., Paris, A., Rathahao-Paris, E. (2020). Mass spectrometry-based metabolomics for an in-depth questioning of human health. *Advances in Clinical Chemistry*, *99*, 147-191. https://doi.org/10.1016/bs.acc.2020.02.009
83. Rathahao-Paris, E., Alves, S., Boussaid, N., Picard-Hagen, N., Gayrard, V., Toutain, P. L., Tabet, J. C., Rutledge, D. N., Paris, A. (2019). Evaluation and validation of an analytical approach for high-throughput metabolomic fingerprinting using direct introduction-high-resolution mass spectrometry: Applicability to classification of urine of scrapie-infected ewes. *European Journal of Mass Spectrometry*, *25*(2), 251-258. https://doi.org/10.1177/1469066718806450
84. Gautier, L., Cope, L., Bolstad, B. M., Irizarry, R. A. (2004). affy—Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, *20*(3), 307-315. https://doi.org/10.1093/bioinformatics/btg405
85. Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47. https://doi.org/10.1093/nar/gkv007
86. Tarazona, S., Furió-Tarí, P., Turrà, D., Pietro, A. D., Nueda, M. J., Ferrer, A., Conesa, A. (2015). Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Research*, *43*(21), e140. https://doi.org/10.1093/nar/gkv711
87. Soneson, C., Delorenzi, M. (2013). A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*, *14*, 91. https://doi.org/10.1186/1471-2105-14-91
88. Oshlack, A., Wakefield, M. J. (2009). Transcript length bias in RNA-seq data

confounds systems biology. *Biology Direct*, *4*, 14. https://doi.org/10.1186/1745-6150-4-14
89. Risso, D., Schwartz, K., Sherlock, G., Dudoit, S. (2011). GC-content normalization for RNA-Seq data. *BMC Bioinformatics*, *12*, 480. https://doi.org/10.1186/1471-2105-12-480
90. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., Siuzdak, G. (2006). XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry*, *78*(3), 779-787. https://doi.org/10.1021/ac051437y
91. Law, C. W., Chen, Y., Shi, W., Smyth, G. K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, *15*(2), R29. https://doi.org/10.1186/gb-2014-15-2-r29
92. Robinson, M. D., McCarthy, D. J., Smyth, G. K. (2010). EdgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139-140. https://doi.org/10.1093/bioinformatics/btp616
93. Love, M. I., Huber, W., Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8
94. Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, *3*(1), Article3. https://doi.org/10.2202/1544-6115.1027
95. Bland, J. M., Altman, D. G. (1995). Multiple significance tests: The Bonferroni method. *BMJ*, *310*(6973), 170. https://doi.org/10.1136/bmj.310.6973.170
96. Benjamini, Y., Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289-300.
97. The Gene Ontology Consortium. (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research*, *47*(D1), D330-D338. https://doi.org/10.1093/nar/gky1055
98. Kanehisa, M., Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, *28*(1), 27-30. https://doi.org/10.1093/nar/28.1.27
99. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43), 15545-15550. https://doi.org/10.1073/pnas.0506580102
100. Kim, S. Y., Volsky, D. J. (2005). PAGE: Parametric analysis of gene set enrichment. *BMC Bioinformatics*, *6*, 144. https://doi.org/10.1186/1471-2105-6-144
101. Al-Shahrour, F., Arbiza, L., Dopazo, H., Huerta-Cepas, J., Mínguez, P., Montaner, D., Dopazo, J. (2007). From genes to functional classes in the study of biological systems. *BMC Bioinformatics*, *8*, 114. https://doi.org/10.1186/1471-2105-8-114
102. Montaner, D., Dopazo, J. (2010). Multidimensional Gene Set Analysis of

Genomic Data. *PLOS ONE*, *5*(4), e10348. https://doi.org/10.1371/journal.pone.0010348
103. Huang, D. W., Sherman, B. T., Lempicki, R. A. (2009). Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, *37*(1), 1-13. https://doi.org/10.1093/nar/gkn923
104. Hernández-de-Diego, R., Tarazona, S., Martínez-Mira, C., Balzano-Nogueira, L., Furió-Tarí, P., Pappas, G. J., Conesa, A. (2018). PaintOmics 3: A web resource for the pathway analysis and visualization of multi-omics data. *Nucleic Acids Research*, *46*(W1), W503-W509. https://doi.org/10.1093/nar/gky466
105. Subramanian, I., Verma, S., Kumar, S., Jere, A., Anamika, K. (2020). Multi-omics Data Integration, Interpretation, and Its Application. *Bioinformatics and Biology Insights*, *14*, 1177932219899051. https://doi.org/10.1177/1177932219899051
106. Nguyen, H., Shrestha, S., Draghici, S., Nguyen, T. (2019). PINSPlus: A tool for tumor subtype discovery in integrated genomic data. *Bioinformatics*, *35*(16), 2843-2846. https://doi.org/10.1093/bioinformatics/bty1049
107. Rappoport, N., Shamir, R. (2019). NEMO: Cancer subtyping by integration of partial multi-omic data. *Bioinformatics*, *35*(18), 3348-3356. https://doi.org/10.1093/bioinformatics/btz058
108. Barabási, A. L., Gulbahce, N., Loscalzo, J. (2011). Network medicine: A network-based approach to human disease. *Nature Reviews. Genetics*, *12*(1), 56-68. https://doi.org/10.1038/nrg2918
109. Shen, R., Olshen, A. B., Ladanyi, M. (2009). Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics*, *25*(22), 2906-2912. https://doi.org/10.1093/bioinformatics/btp543
110. Sedgewick, A. J., Benz, S. C., Rabizadeh, S., Soon-Shiong, P., Vaske, C. J. (2013). Learning subgroup-specific regulatory interactions and regulator independence with PARADIGM. *Bioinformatics*, *29*(13), i62-i70. https://doi.org/10.1093/bioinformatics/btt229
111. Argelaguet, R., Velten, B., Arnol, D., Dietrich, S., Zenz, T., Marioni, J. C., Buettner, F., Huber, W., Stegle, O. (2018). Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. *Molecular Systems Biology*, *14*(6), e8124. https://doi.org/10.15252/msb.20178124
112. Shi, Q., Zhang, C., Peng, M., Yu, X., Zeng, T., Liu, J., Chen, L. (2017). Pattern fusion analysis by adaptive alignment of multiple heterogeneous omics data. *Bioinformatics*, *33*(17), 2706-2714. https://doi.org/10.1093/bioinformatics/btx176
113. Wold, S., Esbensen, K., Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, *2*(1), 37-52. https://doi.org/10.1016/0169-7439(87)80084-9
114. Wold, S., Sjöström, M., Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*(2), 109-130. https://doi.org/10.1016/S0169-7439(01)00155-1
115. Lê Cao, K. A., Martin, P. G., Robert-Granié, C., Besse, P. (2009). Sparse

canonical methods for biological data integration: Application to a cross-platform study. *BMC Bioinformatics*, *10*(1), 34. https://doi.org/10.1186/1471-2105-10-34
116. Barker, M., Rayens, W. (2003). Partial least squares for discrimination. *Journal of chemometrics*, *17*(3), 166-173. https://doi.org/10.1002/cem.785
117. Westerhuis, J. A., Coenegracht, P. M. J. (1997). Multivariate modelling of the pharmaceutical two-step process of wet granulation and tableting with multiblock partial least squares. *Journal of Chemometrics*, *11*(5). https://doi.org/10.1002/(sici)1099-128x(199709/10)11:5<379::aid-cem482>3.0.co;2-8
118. Liquet, B., Lê Cao, K. A., Hocini, H., Thiébaut, R. (2012). A novel approach for biomarker selection and the integration of repeated measures experiments from two assays. *BMC Bioinformatics*, *13*, 325. https://doi.org/10.1186/1471-2105-13-325
119. Bro, R. (1997). PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, *38*(2), 149-171. https://doi.org/10.1016/S0169-7439(97)00032-4
120. Greiff, V., Miho, E., Menzel, U., Reddy, S. T. (2015). Bioinformatic and Statistical Analysis of Adaptive Immune Repertoires. *Trends in Immunology*, *36*(11), 738-749. https://doi.org/10.1016/j.it.2015.09.006
121. Tuomisto, H. (2010). A consistent terminology for quantifying species diversity? Yes, it does exist. *Oecologia*, *164*(4), 853-860. https://doi.org/10.1007/s00442-010-1812-0
122. Greiff, V., Bhat, P., Cook, S. C., Menzel, U., Kang, W., Reddy, S. T. (2015). A bioinformatic framework for immune repertoire diversity profiling enables detection of immunological status. *Genome Medicine*, *7*(1), 49. https://doi.org/10.1186/s13073-015-0169-8
123. Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H. (2020). *vegan: Community Ecology Package*. Available from: https://CRAN.R-project.org/package=vegan
124. Nazarov, V. I., Pogorelyy, M. V., Komech, E. A., Zvyagin, I. V., Bolotin, D. A., Shugay, M., Chudakov, D. M., Lebedev, Y. B., Mamedov, I. Z. (2015). tcR: An R package for T cell receptor repertoire advanced data analysis. *BMC Bioinformatics*, *16*, 175. https://doi.org/10.1186/s12859-015-0613-1
125. Shugay, M., Bagaev, D. V., Turchaninova, M. A., Bolotin, D. A., Britanova, O. V., Putintseva, E. V., Pogorelyy, M. V., Nazarov, V. I., Zvyagin, I. V., Kirgizova, V. I., Kirgizov, K. I., Skorobogatova, E. V., Chudakov, D. M. (2015). VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. *PLoS Computational Biology*, *11*(11), e1004503. https://doi.org/10.1371/journal.pcbi.1004503
126. Gupta, N. T., Vander Heiden, J. A., Uduman, M., Gadala-Maria, D., Yaari, G., Kleinstein, S. H. (2015). Change-O: A toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics*, *31*(20), 3356-3358. https://doi.org/10.1093/bioinformatics/btv359
127. Morisita, M. (1959). Measuring of the dispersion of individuals and analysis

of the distributional patterns. *Memories of the Faculty of Science, Kyushu University. Series E: Biology*, *2*, 215-235.
128. Jaccard, P. (1912). The Distribution of the Flora of the Alpine Zone. *New Phytologist*, *11*(2), 37-50. https://doi.org/10.1111/j.1469-8137.1912.tb05611.x
129. Miho, E., Roškar, R., Greiff, V., Reddy, S. T. (2019). Large-scale network analysis reveals the sequence space architecture of antibody repertoires. *Nature Communications*, *10*(1), 1321. https://doi.org/10.1038/s41467-019-09278-8
130. Gomez-Cabrero, D., Tarazona, S., Ferreirós-Vidal, I., Ramirez, R. N., Company, C., Schmidt, A., Reijmers, T., von Saint Paul, V., Marabita, F., Rodríguez-Ubreva, J., Garcia-Gomez, A., Carroll, T., Cooper, L., Liang, Z., Dharmalingam, G., van der Kloet, F., Harms, A. C., Balzano-Nogueira, L., Lagani, V., Tsamardinos, I., Lappe, M., Maier, D., Westerhuis, J. A., Hankemeier, T., Imhof, A., Ballestar, E., Mortazavi, A., Merkenschlager, M., Tegner, J. Conesa, A. (2019). STATegra, a comprehensive multi-omics dataset of B-cell differentiation in mouse. *Scientific data*, *6*(1), 256. https://doi.org/10.1038/s41597-019-0202-7
131. Mens, M. M. J., Maas, S. C. E., Klap, J., Weverling, G. J., Klatser, P., Brakenhoff, J. P. J., van Meurs, J. B. J., Uitterlinden, A. G., Ikram, M. A., Kavousi, M., Ghanbari, M. (2020). Multi-Omics Analysis Reveals MicroRNAs Associated With Cardiometabolic Traits. *Frontiers in genetics*, *11*, 110. https://doi.org/10.3389/fgene.2020.00110
132. Yu, X., Lai, S., Chen, H., Chen, M. (2020). Protein-protein interaction network with machine learning models and multi-omics data reveals potential neurodegenerative disease-related proteins. *Human Molecular Genetics, 29*(8), 1378-1387. https://doi.org/10.1093/hmg/ddaa065
133. Martín-Jiménez, C. A., Salazar-Barreto, D., Barreto, G. E., González, J. (2017). Genome-Scale Reconstruction of the Human Astrocyte Metabolic Network. *Frontiers in aging neuroscience*, *9*, 23. https://doi.org/10.3389/fnagi.2017.00023
134. Sertbas, M., Ulgen, K. O. (2018). Unlocking Human Brain Metabolism by Genome-Scale and Multiomics Metabolic Models: Relevance for Neurology Research, Health, and Disease. *Omics : a journal of integrative biology*, *22*(7), 455-467. https://doi.org/10.1089/omi.2018.0088
135. Sertbaş, M., Ulgen, K., Cakır, T. (2014). Systematic analysis of transcription-level effects of neurodegenerative diseases on human brain metabolism by a newly reconstructed brain-specific metabolic network. *FEBS open bio*, *4*, 542-553. https://doi.org/10.1016/j.fob.2014.05.006
136. Fabregat, A., Sidiropoulos, K., Viteri, G., Forner, O., Marin-Garcia, P., Arnau, V., D'Eustachio, P., Stein, L., Hermjakob, H. (2017). Reactome pathway analysis: A high-performance in-memory approach. *BMC Bioinformatics*, *18*(1), 142. https://doi.org/10.1186/s12859-017-1559-2
137. Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., von Mering, C. (2019). STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in

genome-wide experimental datasets. *Nucleic Acids Research*, *47*(D1), D607-D613. https://doi.org/10.1093/nar/gky1131
138. De Rui, M., Montagnese, S., Amodio, P. (2016). Recent developments in the diagnosis and treatment of covert/minimal hepatic encephalopathy. *Expert Review of Gastroenterology Hepatology*, *10*(4), 443-450. https://doi.org/10.1586/17474124.2016.1141675
139. Ferrario, M., Cambiaghi, A., Brunelli, L., Giordano, S., Caironi, P., Guatteri, L., Raimondi, F., Gattinoni, L., Latini, R., Masson, S., Ristagno, G., Pastorelli, R. (2016). Mortality prediction in patients with severe septic shock: A pilot study using a target metabolomics approach. *Scientific Reports*, *6*, 20391. https://doi.org/10.1038/srep20391
140. Bland, J. M., Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, *1*(8476), 307-310.
141. Bro, R., Smilde, A. K. (2014). Principal component analysis. *Analytical Methods*, *6*(9), 2812-2831. https://doi.org/10.1039/C3AY41907J
142. Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A., Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, *18*(Suppl 1), S96-104. https://doi.org/10.1093/bioinformatics/18.suppl_1.s96
143. Li, B., Tang, J., Yang, Q., Cui, X., Li, S., Chen, S., Cao, Q., Xue, W., Chen, N., Zhu, F. (2016). Performance Evaluation and Online Realization of Data-driven Normalization Methods Used in LC/MS based Untargeted Metabolomics Analysis. *Scientific Reports*, *6*(1), 38881. https://doi.org/10.1038/srep38881
144. Purohit, P. V., Rocke, D. M., Viant, M. R., Woodruff, D. L. (2004). Discrimination Models Using Variance-Stabilizing Transformation of Metabolomic NMR Data. *OMICS: A Journal of Integrative Biology*, *8*(2), 118-130. https://doi.org/10.1089/1536231041388348
145. Shin, H., Medriano, C. A., Park, B., Park, Y. H., Lee, K. Y. (2018). Screening and identification of neuroprotective compounds from Scrophularia buergeriana using cell extraction coupled with LC–MS. *Journal of Pharmaceutical and Biomedical Analysis*, *148*, 355-360. https://doi.org/10.1016/j.jpba.2017.10.018
146. Tarazona, S., Balzano-Nogueira, L., Gómez-Cabrero, D., Schmidt, A., Imhof, A., Hankemeier, T., Tegnér, J., Westerhuis, J. A., Conesa, A. (2020). Harmonization of quality metrics and power calculation in multi-omic studies. *Nature Communications*, *11*(1), 3092. https://doi.org/10.1038/s41467-020-16937-8
147. Reynolds, A. P., Richards, G., de la Iglesia, B., Rayward-Smith, V. J. (2006). Clustering rules: A comparison of partitioning and hierarchical clustering algorithms. *Journal of Mathematical Modelling and Algorithms*, *5*(4), 475-504. https://doi.org/10.1007/s10852-005-9022-1
148. Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53-65. https://doi.org/10.1016/0377-0427(87)90125-7
149. Sanyal, R., Polyak, M. J., Zuccolo, J., Puri, M., Deng, L., Roberts, L., Zuba, A., Storek, J., Luider, J. M., Sundberg, E. M., Mansoor, A., Baigorri, E.,

Chu, M. P., Belch, A. R., Pilarski, L. M., Deans, J. P. (2017). MS4A4A: A novel cell surface marker for M2 macrophages and plasma cells. *Immunology and Cell Biology*, *95*(7), 611-619. https://doi.org/10.1038/icb.2017.18

150. Berasain, C., Avila, M. A. (2014). Amphiregulin. *Seminars in Cell Developmental Biology*, *28*, 31-41. https://doi.org/10.1016/j.semcdb.2014.01.005
151. Froehlich, J., Versapuech, M., Megrelis, L., Largeteau, Q., Meunier, S., Tanchot, C., Bismuth, G., Delon, J., Mangeney, M. (2016). FAM65B controls the proliferation of transformed and primary T cells. *Oncotarget*, *7*(39), 63215-63225. https://doi.org/10.18632/oncotarget.11438
152. Seif, F., Khoshmirsafa, M., Aazami, H., Mohsenzadegan, M., Sedighi, G., Bahar, M. (2017). The role of JAK-STAT signaling pathway and its regulators in the fate of T helper cells. *Cell Communication and Signaling*, *15*(1), 23. https://doi.org/10.1186/s12964-017-0177-y
153. Glennie, M. J., French, R. R., Cragg, M. S., Taylor, R. P. (2007). Mechanisms of killing by anti-CD20 monoclonal antibodies. *Molecular Immunology*, *44*(16), 3823-3837. https://doi.org/10.1016/j.molimm.2007.06.151
154. Morsy, D. E. D., Sanyal, R., Zaiss, A. K., Deo, R., Muruve, D. A., Deans, J. P. (2013). Reduced T-dependent humoral immunity in CD20-deficient mice. *The Journal of Immunology*, *191*(6), 3112-3118. https://doi.org/10.4049/jimmunol.1202098
155. Hampe, C. S. (2012). B Cells in Autoimmune Diseases. *Scientifica*, *2012*, 1-18. https://doi.org/10.6064/2012/215308
156. Pender, M. P. (2012). CD8+ T-Cell Deficiency, Epstein-Barr Virus Infection, Vitamin D Deficiency, and Steps to Autoimmunity: A Unifying Hypothesis. *Autoimmune Diseases*, *2012*, 1-16. https://doi.org/10.1155/2012/189096
157. Pender, M. P., Csurhes, P. A., Pfluger, C. M., Burrows, S. R. (2014). Deficiency of CD8 $^{+}$ effector memory T cells is an early and persistent feature of multiple sclerosis. *Multiple Sclerosis Journal*, *20*(14), 1825-1832. https://doi.org/10.1177/1352458514536252
158. Kreuzfelder, E., Shen, G., Bittorf, M., Scheiermann, N., Thraenhart, O., Seidel, D., Grosse-Wilde, H. (1992). Enumeration of T, B and Natural Killer Peripheral Blood Cells of Patients with Multiple Sclerosis and Controls. *European Neurology*, *32*(4), 190-194. https://doi.org/10.1159/000116820
159. McCusker, D., Wilson, M., Trowsdale, J. (1999). Organization of the genes encoding the human proteasome activators PA28alpha and beta. *Immunogenetics*, *49*(5), 438-445. https://doi.org/10.1007/s002510050517
160. Steinman, L. (2007). A brief history of T(H)17, the first major revision in the T(H)1/T(H)2 hypothesis of T cell-mediated tissue damage. *Nature Medicine*, *13*(2), 139-145. https://doi.org/10.1038/nm1551
161. Langenkamp, A., Nagata, K., Murphy, K., Wu, L., Lanzavecchia, A., Sallusto, F. (2003). Kinetics and expression patterns of chemokine receptors in human CD4+ T lymphocytes primed by myeloid or plasmacytoid dendritic cells. *European Journal of Immunology*, *33*(2), 474-482.

https://doi.org/10.1002/immu.200310023
162. Kim, C. H., Kunkel, E. J., Boisvert, J., Johnston, B., Campbell, J. J., Genovese, M. C., Greenberg, H. B., Butcher, E. C. (2001). Bonzo/CXCR6 expression defines type 1-polarized T-cell subsets with extralymphoid tissue homing potential. *The Journal of Clinical Investigation*, *107*(5), 595-601. https://doi.org/10.1172/JCI11902
163. Sallusto, F., Lenig, D., Mackay, C. R., Lanzavecchia, A. (1998). Flexible programs of chemokine receptor expression on human polarized T helper 1 and 2 lymphocytes. *The Journal of Experimental Medicine*, *187*(6), 875-883. https://doi.org/10.1084/jem.187.6.875
164. Watanabe, S., Yamada, Y., Murakami, H. (2019). Expression of Th1/Th2 cell-related chemokine receptors on CD4+ lymphocytes under physiological conditions. *International Journal of Laboratory Hematology*. https://doi.org/10.1111/ijlh.13141
165. Zingoni, A., Soto, H., Hedrick, J. A., Stoppacciaro, A., Storlazzi, C. T., Sinigaglia, F., D'Ambrosio, D., O'Garra, A., Robinson, D., Rocchi, M., Santoni, A., Zlotnik, A., Napolitano, M. (1998). The chemokine receptor CCR8 is preferentially expressed in Th2 but not Th1 cells. *The Journal of Immunology*, *161*(2), 547-551.
166. Lee, Y. K., Kwak, D. H., Oh, K. W., Nam, S. Y., Lee, B. J., Yun, Y. W., Kim, Y. B., Han, S. B., Hong, J. T. (2009). CCR5 deficiency induces astrocyte activation, Abeta deposit and impaired memory function. *Neurobiology of Learning and Memory*, *92*(3), 356-363. https://doi.org/10.1016/j.nlm.2009.04.003
167. Reale, M., Iarlori, C., Feliciani, C., Gambi, D. (2008). Peripheral chemokine receptors, their ligands, cytokines and Alzheimer's disease. *Journal of Alzheimer's Disease*, *14*(2), 147-159. https://doi.org/10.3233/jad-2008-14203
168. Zambrano-Zaragoza, J. F., Romo-Martínez, E. J., Durán-Avelar, M. J., García-Magallanes, N., Vibanco-Pérez, N. (2014). Th17 Cells in Autoimmune and Infectious Diseases. *International Journal of Inflammation*, *2014*, 1-12. https://doi.org/10.1155/2014/651503
169. Fiala, M., Kooij, G., Wagner, K., Hammock, B., Pellegrini, M. (2017). Modulation of innate immunity of patients with Alzheimer's disease by omega-3 fatty acids. *The FASEB Journal*, *31*(8), 3229-3239. https://doi.org/10.1096/fj.201700065R
170. Olivera-Perez, H. M., Lam, L., Dang, J., Jiang, W., Rodriguez, F., Rigali, E., Weitzman, S., Porter, V., Rubbi, L., Morselli, M., Pellegrini, M., Fiala, M. (2017). Omega-3 fatty acids increase the unfolded protein response and improve amyloid-β phagocytosis by macrophages of patients with mild cognitive impairment. *The FASEB Journal*, *31*(10), 4359-4369. https://doi.org/10.1096/fj.201700290R
171. Whiley, L., Sen, A., Heaton, J., Proitsi, P., García-Gómez, D., Leung, R., Smith, N., Thambisetty, M., Kloszewska, I., Mecocci, P., Soininen, H., Tsolaki, M., Vellas, B., Lovestone, S., Legido-Quigley, C. (2014). Evidence of altered phosphatidylcholine metabolism in Alzheimer's disease. *Neurobiology of Aging*, *35*(2), 271-278. https://doi.org/10.1016/j.neurobiolaging.2013.08.001

172. Grimm, M. O. W., Grösgen, S., Riemenschneider, M., Tanila, H., Grimm, H. S., Hartmann, T. (2011). From brain to food: Analysis of phosphatidylcholins, lyso-phosphatidylcholins and phosphatidylcholin-plasmalogens derivates in Alzheimer's disease human post mortem brains and mice model via mass spectrometry. *Journal of Chromatography. A*, *1218*(42), 7713-7722. https://doi.org/10.1016/j.chroma.2011.07.073
173. Mino, T., Takeuchi, O. (2018). Post-transcriptional regulation of immune responses by RNA binding proteins. *Proceedings of the Japan Academy, Series B*, *94*(6), 248-258. https://doi.org/10.2183/pjab.94.017
174. Vlasova-St. Louis, I., Bohjanen, P. R. (2017). Post-transcriptional regulation of cytokine and growth factor signaling in cancer. *Cytokine Growth Factor Reviews*, *33*, 83-93. https://doi.org/10.1016/j.cytogfr.2016.11.004
175. Salvi, V., Gianello, V., Tiberio, L., Sozzani, S., Bosisio, D. (2019). Cytokine Targeting by miRNAs in Autoimmune Diseases. *Frontiers in Immunology*, *10*, 15. https://doi.org/10.3389/fimmu.2019.00015
176. Aoyama, T., Inokuchi, S., Brenner, D. A., Seki, E. (2010). CX3CL1-CX3CR1 interaction prevents carbon tetrachloride-induced liver inflammation and fibrosis in mice. *Hepatology*, *52*(4), 1390-1400. https://doi.org/10.1002/hep.23795
177. Golden-Mason, L., Kelly, A. M., Doherty, D. G., Traynor, O., Mcentee, G., Kelly, J., Hegarty, J. E., O'Farrelly, C. (2004). Hepatic interleuklin 15 (IL-15) expression: Implications for local NK/NKT cell homeostasis and development. *Clinical and Experimental Immunology*, *138*(1), 94-101. https://doi.org/10.1111/j.1365-2249.2004.02586.x
178. Rolla, S., Alchera, E., Imarisio, C., Bardina, V., Valente, G., Cappello, P., Mombello, C., Follenzi, A., Novelli, F., Carini, R. (2016). The balance between IL-17 and IL-22 produced by liver-infiltrating T-helper cells critically controls NASH development in mice. *Clinical Science*, *130*(3), 193-203. https://doi.org/10.1042/CS20150405
179. Paapstel, K., Kals, J., Eha, J., Tootsi, K., Ottas, A., Piir, A., Jakobson, M., Lieberg, J., Zilmer, M. (2018). Inverse relations of serum phosphatidylcholines and lysophosphatidylcholines with vascular damage and heart rate in patients with atherosclerosis. *Nutrition, Metabolism, and Cardiovascular Diseases*, *28*(1), 44-52. https://doi.org/10.1016/j.numecd.2017.07.011
180. Weismann, D., Hartvigsen, K., Lauer, N., Bennett, K. L., Scholl, H. P. N., Charbel Issa, P., Cano, M., Brandstätter, H., Tsimikas, S., Skerka, C., Superti-Furga, G., Handa, J. T., Zipfel, P. F., Witztum, J. L., Binder, C. J. (2011). Complement factor H binds malondialdehyde epitopes and protects from oxidative stress. *Nature*, *478*(7367), 76-81. https://doi.org/10.1038/nature10449
181. Mapstone, M., Cheema, A. K., Fiandaca, M. S., Zhong, X., Mhyre, T. R., MacArthur, L. H., Hall, W. J., Fisher, S. G., Peterson, D. R., Haley, J. M., Nazar, M. D., Rich, S. A., Berlau, D. J., Peltz, C. B., Tan, M. T., Kawas, C. H., Federoff, H. J. (2014). Plasma phospholipids identify antecedent memory impairment in older adults. *Nature Medicine*, *20*(4), 415-418. https://doi.org/10.1038/nm.3466

182. Freigang, S. (2016). The regulation of inflammation by oxidized phospholipids. *European Journal of Immunology*, *46*(8), 1818-1825. https://doi.org/10.1002/eji.201545676
183. Gimenez-Garzó, C., Urios, A., Agustí, A., González-López, O., Escudero-García, D., Escudero-Sanchis, A., Serra, M. A., Giner-Durán, R., Montoliu, C., Felipo, V. (2015). Is cognitive impairment in cirrhotic patients due to increased peroxynitrite and oxidative stress? *Antioxidants Redox Signaling*, *22*(10), 871-877. https://doi.org/10.1089/ars.2014.6240
184. Klavins, K., Koal, T., Dallmann, G., Marksteiner, J., Kemmler, G., Humpel, C. (2015). The ratio of phosphatidylcholines to lysophosphatidylcholines in plasma differentiates healthy controls from patients with Alzheimer's disease and mild cognitive impairment. *Alzheimer's Dementia*, *1*(3), 295-302. https://doi.org/10.1016/j.dadm.2015.05.003
185. Mahad, D. J., Lawry, J., Howell, S. J. L., Woodroofe, M. N. (2003). Longitudinal study of chemokine receptor expression on peripheral lymphocytes in multiple sclerosis: CXCR3 upregulation is associated with relapse. *Multiple Sclerosis*, *9*(2), 189-198. https://doi.org/10.1191/1352458503ms899oa
186. D'Mello, C., Le, T., Swain, M. G. (2009). Cerebral microglia recruit monocytes into the brain in response to tumor necrosis factoralpha signaling during peripheral organ inflammation. *The Journal of Neuroscience*, *29*(7), 2089-2102. https://doi.org/10.1523/JNEUROSCI.3567-08.2009
187. Zang, Y. C., Samanta, A. K., Halder, J. B., Hong, J., Tejada-Simon, M. V., Rivera, V. M., Zhang, J. Z. (2000). Aberrant T cell migration toward RANTES and MIP-1 alpha in patients with multiple sclerosis. Overexpression of chemokine receptor CCR5. *Brain: A Journal of Neurology*, *123 (9)*, 1874-1882. https://doi.org/10.1093/brain/123.9.1874
188. Pei, X., Sun, Q., Zhang, Y., Wang, P., Peng, X., Guo, C., Xu, E., Zheng, Y., Mo, X., Ma, J., Chen, D., Zhang, Y., Zhang, Y., Song, Q., Guo, S., Shi, T., Zhang, Z., Ma, D., Wang, Y. (2014). PC3-secreted microprotein is a novel chemoattractant protein and functions as a high-affinity ligand for CC chemokine receptor 2. *The Journal of Immunology*, *192*(4), 1878-1886. https://doi.org/10.4049/jimmunol.1300758
189. Vermi, W., Riboldi, E., Wittamer, V., Gentili, F., Luini, W., Marrelli, S., Vecchi, A., Franssen, J. D., Communi, D., Massardi, L., Sironi, M., Mantovani, A., Parmentier, M., Facchetti, F., Sozzani, S. (2005). Role of ChemR23 in directing the migration of myeloid and plasmacytoid dendritic cells to lymphoid organs and inflamed skin. *The Journal of Experimental Medicine*, *201*(4), 509-515. https://doi.org/10.1084/jem.20041310
190. Steinberg, S. J., Morgenthaler, J., Heinzer, A. K., Smith, K. D., Watkins, P. A. (2000). Very long-chain acyl-CoA synthetases. Human «bubblegum» represents a new family of proteins capable of activating very long-chain fatty acids. *The Journal of Biological Chemistry*, *275*(45), 35162-35169. https://doi.org/10.1074/jbc.M006403200
191. Tidhar, R., Zelnik, I. D., Volpert, G., Ben-Dor, S., Kelly, S., Merrill, A. H., Futerman, A. H. (2018). Eleven residues determine the acyl chain

specificity of ceramide synthases. *The Journal of Biological Chemistry*, *293*(25), 9912-9921. https://doi.org/10.1074/jbc.RA118.001936
192. Quazi, F., Molday, R. S. (2013). Differential phospholipid substrates and directional transport by ATP-binding cassette proteins ABCA1, ABCA7, and ABCA4 and disease-causing mutants. *The Journal of Biological Chemistry*, *288*(48), 34414-34426. https://doi.org/10.1074/jbc.M113.508812
193. Smith, J., Su, X., El-Maghrabi, R., Stahl, P. D., Abumrad, N. A. (2008). Opposite regulation of CD36 ubiquitination by fatty acids and insulin: Effects on fatty acid uptake. *The Journal of Biological Chemistry*, *283*(20), 13578-13585. https://doi.org/10.1074/jbc.M800008200
194. Sheedy, F. J., Grebe, A., Rayner, K. J., Kalantari, P., Ramkhelawon, B., Carpenter, S. B., Becker, C. E., Ediriweera, H. N., Mullick, A. E., Golenbock, D. T., Stuart, L. M., Latz, E., Fitzgerald, K. A., Moore, K. J. (2013). CD36 coordinates NLRP3 inflammasome activation by facilitating intracellular nucleation of soluble ligands into particulate ligands in sterile inflammation. *Nature Immunology*, *14*(8), 812-820. https://doi.org/10.1038/ni.2639
195. Murakami, M., Nakatani, Y., Atsumi, G. I., Inoue, K., Kudo, I. (2017). Regulatory functions of phospholipase A2. *Critical Reviews in Immunology*, *37*(2-6), 121-179. https://doi.org/10.1615/CritRevImmunol.v37.i2-6.20
196. Law, S. H., Chan, M. L., Marathe, G. K., Parveen, F., Chen, C. H., Ke, L. Y. (2019). An Updated Review of Lysophosphatidylcholine Metabolism in Human Diseases. *International Journal of Molecular Sciences*, *20*(5), 1149. https://doi.org/10.3390/ijms20051149
197. Miletić Vukajlović, J., Drakulić, D., Pejić, S., Ilić, T. V., Stefanović, A., Petković, M., Schiller, J. (2020). Increased plasma phosphatidylcholine/lysophosphatidylcholine ratios in patients with Parkinson's disease. *Rapid Communications in Mass Spectrometry*, *34*(4), Article 4. https://doi.org/10.1002/rcm.8595
198. Vishnoi, A., Rani, S. (2017). MiRNA Biogenesis and Regulation of Diseases: An Overview. *MicroRNA Profiling, 1509*, 1-10. https://doi.org/10.1007/978-1-4939-6524-3_1
199. Luo, Y., Wang, H. (2020). Effects of Non-Coding RNA on Regulatory T Cells and Implications for Treatment of Immunological Diseases. *Frontiers in Immunology*, *11*, 612060. https://doi.org/10.3389/fimmu.2020.612060
200. Jin, B. X., Zhang, Y. H., Jin, W. J., Sun, X. Y., Qiao, G. F., Wei, Y. Y., Sun, L. B., Zhang, W. H., Li, N. (2015). MicroRNA panels as disease biomarkers distinguishing hepatitis B virus infection caused hepatitis and liver cirrhosis. *Scientific Reports*, *5*, 15026. https://doi.org/10.1038/srep15026
201. Caggiu, E., Paulus, K., Mameli, G., Arru, G., Sechi, G. P., Sechi, L. A. (2018). Differential expression of miRNA 155 and miRNA 146a in Parkinson's disease patients. *ENeurologicalSci*, *13*, 1-4. https://doi.org/10.1016/j.ensci.2018.09.002
202. Jain, G., Stuendl, A., Rao, P., Berulava, T., Pena Centeno, T., Kaurani, L., Burkhardt, S., Delalle, I., Kornhuber, J., Hüll, M., Maier, W., Peters, O.,

Esselmann, H., Schulte, C., Deuschle, C., Synofzik, M., Wiltfang, J., Mollenhauer, B., Maetzler, W., Schneider, A. Fischer, A. (2019). A combined miRNA-piRNA signature to detect Alzheimer's disease. *Translational Psychiatry*, *9*(1), 250. https://doi.org/10.1038/s41398-019-0579-2

203. McCoy, C. E. (2017). MiR-155 Dysregulation and Therapeutic Intervention in Multiple Sclerosis. *Regulation of Inflammatory Signaling in Health and Disease, 1024*, 111-131. https://doi.org/10.1007/978-981-10-5987-2_5
204. Cherone, J. M., Jorgji, V., Burge, C. B. (2019). Cotargeting among microRNAs in the brain. *Genome Research*, *29*(11), 1791-1804. https://doi.org/10.1101/gr.249201.119
205. Soreq, H., Wolf, Y. (2011). NeurimmiRs: MicroRNAs in the neuroimmune interface. *Trends in Molecular Medicine*, *17*(10), 548-555. https://doi.org/10.1016/j.molmed.2011.06.009
206. López-Riera, M., Conde, I., Quintas, G., Pedrola, L., Zaragoza, Á., Perez-Rojas, J., Salcedo, M., Benlloch, S., Castell, J. V., Jover, R. (2018). Non-invasive prediction of NAFLD severity: A comprehensive, independent validation of previously postulated serum microRNA biomarkers. *Scientific Reports*, *8*(1), 10606. https://doi.org/10.1038/s41598-018-28854-4
207. Markovic, J., Sharma, A. D., Balakrishnan, A. (2020). MicroRNA-221: A Fine Tuner and Potential Biomarker of Chronic Liver Injury. *Cells*, *9*(8), E1767. https://doi.org/10.3390/cells9081767
208. Wang, L., Zhang, L. (2020). MicroRNAs in amyotrophic lateral sclerosis: From pathogenetic involvement to diagnostic biomarker and therapeutic agent development. *Neurological Sciences: Official Journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*, *41*(12), 3569-3577. https://doi.org/10.1007/s10072-020-04773-z
209. Chen, L., Heikkinen, L., Wang, C., Yang, Y., Sun, H., Wong, G. (2019). Trends in the development of miRNA bioinformatics tools. *Briefings in Bioinformatics*, *20*(5), 1836-1852. https://doi.org/10.1093/bib/bby054
210. Hu, Y., Lan, W., Miller, D. (2017). Next-Generation Sequencing for MicroRNA Expression Profile. En J. Huang, G. M. Borchert, D. Dou, J. Huan, W. Lan, M. Tan, B. Wu (Eds.), *Bioinformatics in MicroRNA Research* (Vol. 1617, pp. 169-177). Springer New York. https://doi.org/10.1007/978-1-4939-7046-9_12
211. Bolger, A. M., Lohse, M., Usadel, B. (2014). A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120. https://doi.org/10.1093/bioinformatics/btu170
212. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15-21. https://doi.org/10.1093/bioinformatics/bts635
213. Frankish, A., Diekhans, M., Ferreira, A. M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J. M., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Carbonell Sala, S., Chrast, J., Cunningham, F., Di Domenico, T., Donaldson, S., Fiddes, I. T., García-Girón, C.,

Gonzalez, J. M., Grego, T., Hardy, M., Hourlier, T., Hunt, T., Izuogu, O. G., Lagarde, J., J Martin, F. J., Martínez, L., Mohanan, S., Muir, P., Navarro, F. C. P., Parker, A., Pei, B., Pozo, F., Ruffier, M., Schmitt, B. M., Stapleton, E., Suner, M. M., Sycheva, I., Uszczynska-Ratajczak, B., Xu, J., Yates, A., Zerbino, D., Zhang, Y., Aken, B., Choudhary, J. S., Gerstein, M., Guigó, R., Hubbard, T. J. P., Kellis, M., Paten, B., Reymond, A., Tress, M. L., Flicek, P. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Research*, *47*(D1), D766-D773. https://doi.org/10.1093/nar/gky955

214. Okonechnikov, K., Conesa, A., García-Alcalde, F. (2015). Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, *32(2)*, 292-294. https://doi.org/10.1093/bioinformatics/btv566
215. Anders, S., Pyl, P. T., Huber, W. (2015). HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166-169. https://doi.org/10.1093/bioinformatics/btu638
216. Hansen, K. D., Irizarry, R. A., Wu, Z. (2012). Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics*, *13*(2), 204-216. https://doi.org/10.1093/biostatistics/kxr054
217. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*(1), 10. https://doi.org/10.14806/ej.17.1.200
218. Liao, Y., Smyth, G. K., Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923-930. https://doi.org/10.1093/bioinformatics/btt656
219. Kozomara, A., Birgaoanu, M., Griffiths-Jones, S. (2019). miRBase: From microRNA sequences to function. *Nucleic Acids Research*, *47*(D1), D155-D162. https://doi.org/10.1093/nar/gky1141
220. de la Fuente, L., Arzalluz-Luque, Á., Tardáguila, M., Del Risco, H., Martí, C., Tarazona, S., Salguero, P., Scott, R., Lerma, A., Alastrue-Agudo, A., Bonilla, P., Newman, J. R. B., Kosugi, S., McIntyre, L. M., Moreno-Manzano, V., Conesa, A. (2020). tappAS: A comprehensive computational framework for the analysis of the functional impact of differential splicing. *Genome Biology*, *21*(1), 119. https://doi.org/10.1186/s13059-020-02028-w
221. Slowikowski, K. (2015). *Tftargets* [R]. Available from: https://github.com/slowkow/tftargets
222. ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57-74. https://doi.org/10.1038/nature11247
223. López-Nieva, P., Fernández-Navarro, P., Graña-Castro, O., Andrés-León, E., Santos, J., Villa-Morales, M., Cobos-Fernández, M. Á., González-Sánchez, L., Malumbres, M., Salazar-Roa, M., Fernández-Piqueras, J. (2019). Detection of novel fusion-transcripts by RNA-Seq in T-cell lymphoblastic lymphoma. *Scientific Reports*, *9*(1), 5179. https://doi.org/10.1038/s41598-019-41675-3
224. Ni, Y. Y., Chen, Y., Lu, S. Y., Sun, B. Y., Wang, F., Wu, X. L., Dang, S. Y.,

Zhang, G. H., Zhang, H. X., Kuang, Y., Fei, J., Gu, M. M., Rong, W. F., Wang, Z. G. (2014). Deletion of Gpr128 results in weight loss and increased intestinal contraction frequency. *World Journal of Gastroenterology*, *20*(2), 498-508. https://doi.org/10.3748/wjg.v20.i2.498

225. Lippert, E., Yowe, D. L., Gonzalo, J. A., Justice, J. P., Webster, J. M., Fedyk, E. R., Hodge, M., Miller, C., Gutierrez-Ramos, J. C., Borrego, F., Keane-Myers, A., Druey, K. M. (2003). Role of regulator of G protein signaling 16 in inflammation-induced T lymphocyte migration and activation. *The Journal of Immunology*, *171*(3), 1542-1555. https://doi.org/10.4049/jimmunol.171.3.1542
226. Rani, A., Greenlaw, R., Smith, R. A., Galustian, C. (2016). HES1 in immunity and cancer. *Cytokine Growth Factor Reviews*, *30*, 113-117. https://doi.org/10.1016/j.cytogfr.2016.03.010
227. Damiano, J. S., Oliveira, V., Welsh, K., Reed, J. C. (2004). Heterotypic interactions among NACHT domains: Implications for regulation of innate immune responses. *The Biochemical Journal*, *381*(Pt 1), 213-219. https://doi.org/10.1042/BJ20031506
228. Ye, R. D., Cavanagh, S. L., Quehenberger, O., Prossnitz, E. R., Cochrane, C. G. (1992). Isolation of a cDNA that encodes a novel granulocyte N-formyl peptide receptor. *Biochemical and Biophysical Research Communications*, *184*(2), 582-589. https://doi.org/10.1016/0006-291x(92)90629-y
229. Rubio, T., Felipo, V., Tarazona, S., Pastorelli, R., Escudero-García, D., Tosca, J., Urios, A., Conesa, A., Montoliu, C. (2021). Multi-omic analysis unveils biological pathways in peripheral immune system associated to minimal hepatic encephalopathy appearance in cirrhotic patients. *Scientific Reports*, *11*(1), 1907. https://doi.org/10.1038/s41598-020-80941-7
230. Polak, A., Grywalska, E., Klatka, J., Roliński, J., Matyjaszek-Matuszek, B., Klatka, M. (2019). Toll-Like Receptors-2 and -4 in Graves' Disease-Key Players or Bystanders? *International Journal of Molecular Sciences*, *20*(19), Article 19. https://doi.org/10.3390/ijms20194732
231. Hasheminia, S. J., Zarkesh-Esfahani, S. H., Tolouei, S., Shaygannejad, V., Shirzad, H., Hashemzadeh-Chaleshtory, M. (2014). Toll like receptor 2 and 4 expression in peripheral blood mononuclear cells of multiple sclerosis patients. *Iranian journal of immunology : IJI*, *11*(2), 74-83.
232. van Noort, J. M., Bsibsi, M. (2009). Toll-like receptors in the CNS: implications for neurodegeneration and repair. *Progress in Brain Research*, *175*, 139-148. https://doi.org/10.1016/S0079-6123(09)17509-X
233. Jafarzadeh, A., Nemati, M., Khorramdelazad, H., Mirshafiey, A. (2019). The Toll-like Receptor 2 (TLR2)-related Immunopathological Responses in the Multiple Sclerosis and Experimental Autoimmune Encephalomyelitis. *Iranian journal of allergy, asthma, and immunology*, *18*(3), 230-250. https://doi.org/10.18502/ijaai.v18i3.1117
234. Miranda-Hernandez, S., Gerlach, N., Fletcher, J. M., Biros, E., Mack, M., Körner, H., Baxter, A. G. (2011). Role for MyD88, TLR2 and TLR9 but not TLR1, TLR4 or TLR6 in experimental autoimmune

encephalomyelitis. *The Journal of Immunology*, *187*(2), 791-804. https://doi.org/10.4049/jimmunol.1001992
235. Qu, X., Han, J., Zhang, Y., Wang, X., Fan, H., Hua, F., Yao, R. (2019). TLR4-RelA-miR-30a signal pathway regulates Th17 differentiation during experimental autoimmune encephalomyelitis development. *Journal of Neuroinflammation*, *16*(1), 183. https://doi.org/10.1186/s12974-019-1579-0
236. Fitzgerald, K. A., Kagan, J. C. (2020). Toll-like Receptors and the Control of Immunity. *Cell*, *180*(6), 1044-1066. https://doi.org/10.1016/j.cell.2020.02.041
237. Vidya, M. K., Kumar, V. G., Sejian, V., Bagath, M., Krishnan, G., Bhatta, R. (2018). Toll-like receptors: Significance, ligands, signaling pathways, and functions in mammals. *International Reviews of Immunology*, *37*(1), 20-36. https://doi.org/10.1080/08830185.2017.1380200
238. Jeannin, P., Herbault, N., Delneste, Y., Magistrelli, G., Lecoanet-Henchoz, S., Caron, G., Aubry, J. P., Bonnefoy, J. Y. (1999). Human effector memory T cells express CD86: A functional role in naive T cell priming. *The Journal of Immunology*, *162*(4), 2044-2048.
239. Chen, K., Kolls, J. K. (2017). Interluekin-17A (IL17A). *Gene*, *614*, 8-14. https://doi.org/10.1016/j.gene.2017.01.016
240. Branco, A. C. C. C., Yoshikawa, F. S. Y., Pietrobon, A. J., Sato, M. N. (2018). Role of Histamine in Modulating the Immune Response and Inflammation. *Mediators of Inflammation*, *2018*, 9524075. https://doi.org/10.1155/2018/9524075
241. Strasser, B., Becker, K., Fuchs, D., Gostner, J. M. (2017). Kynurenine pathway metabolism and immune activation: Peripheral measurements in psychiatric and co-morbid conditions. *Neuropharmacology*, *112*(Pt B), 286-296. https://doi.org/10.1016/j.neuropharm.2016.02.030
242. Mezrich, J. D., Fechner, J. H., Zhang, X., Johnson, B. P., Burlingham, W. J., Bradfield, C. A. (2010). An interaction between kynurenine and the aryl hydrocarbon receptor can generate regulatory T cells. *The Journal of Immunology*, *185*(6), 3190-3198. https://doi.org/10.4049/jimmunol.0903670
243. Küster, O. C., Laptinskaya, D., Fissler, P., Schnack, C., Zügel, M., Nold, V., Thurm, F., Pleiner, S., Karabatsiakis, A., von Einem, B., Weydt, P., Liesener, A., Borta, A., Woll, A., Hengerer, B., Kolassa, I. T., von Arnim, C. A. F. (2017). Novel Blood-Based Biomarkers of Cognition, Stress, and Physical or Cognitive Training in Older Adults at Risk of Dementia: Preliminary Evidence for a Role of BDNF, Irisin, and the Kynurenine Pathway. *Journal of Alzheimer's Disease: JAD*, *59*(3), 1097-1111. https://doi.org/10.3233/JAD-170447
244. Solvang, S. E. H., Nordrehaug, J. E., Tell, G. S., Nygård, O., McCann, A., Ueland, P. M., Midttun, Ø., Meyer, K., Vedeler, C. A., Aarsland, D., Refsum, H., Smith, A. D., Giil, L. M. (2019). The kynurenine pathway and cognitive performance in community-dwelling older adults. The Hordaland Health Study. *Brain, Behavior, and Immunity*, *75*, 155-162. https://doi.org/10.1016/j.bbi.2018.10.003
245. Talarowska, M., Galecki, P. (2016). Cognition and Emotions in Recurrent

Depressive Disorders—The Role of Inflammation and the Kynurenine Pathway. *Current Pharmaceutical Design*, *22*(8), 955-962. https://doi.org/10.2174/1381612822666151230110738
246. Fujigaki, H., Yamamoto, Y., Saito, K. (2017). L-Tryptophan-kynurenine pathway enzymes are therapeutic target for neuropsychiatric diseases: Focus on cell type differences. *Neuropharmacology*, *112*(Pt B), 264-274. https://doi.org/10.1016/j.neuropharm.2016.01.011
247. Salvador, J. M., Hollander, M. C., Nguyen, A. T., Kopp, J. B., Barisoni, L., Moore, J. K., Ashwell, J. D., Fornace, A. J. (2002). Mice lacking the p53-effector gene Gadd45a develop a lupus-like syndrome. *Immunity*, *16*(4), 499-508. https://doi.org/10.1016/s1074-7613(02)00302-3
248. Heiko, J., Salvador, J. (2009). The Role of Gadd45a in Suppression of Autoimmunity. *Molecular and Cellular Pharmacology*, *1*. https://doi.org/10.4255/mcpharmacol.09.35
249. López-Santalla, M., Salvador-Bernáldez, M., González-Alvaro, I., Castañeda, S., Ortiz, A. M., García-García, M. I., Kremer, L., Roncal, F., Mulero, J., Martínez-A, C., Salvador, J. M. (2011). Tyr^{323}-dependent p38 activation is associated with rheumatoid arthritis and correlates with disease activity. *Arthritis and Rheumatism*, *63*(7), 1833-1842. https://doi.org/10.1002/art.30375
250. Damgaard, C. K., Lykke-Andersen, J. (2013). Regulation of ARE-mRNA Stability by Cellular Signaling: Implications for Human Cancer. *Cancer Treatment and Research*, *158*, 153-180. https://doi.org/10.1007/978-3-642-31659-3_7
251. Lee, H. H., Yoon, N. A., Vo, M. T., Kim, C. W., Woo, J. M., Cha, H. J., Cho, Y. W., Lee, B. J., Cho, W. J., Park, J. W. (2012). Tristetraprolin down-regulates IL-17 through mRNA destabilization. *FEBS Letters*, *586*(1), 41-46. https://doi.org/10.1016/j.febslet.2011.11.021
252. Chang, S. H., Chung, Y., Dong, C. (2010). Vitamin D suppresses Th17 cytokine production by inducing C/EBP homologous protein (CHOP) expression. *The Journal of Biological Chemistry*, *285*(50), 38751-38755. https://doi.org/10.1074/jbc.C110.185777
253. Abbasi, A., Forsberg, K., Bischof, F. (2015). The role of the ubiquitin-editing enzyme A20 in diseases of the central nervous system and other pathological processes. *Frontiers in Molecular Neuroscience*, *8*, 21. https://doi.org/10.3389/fnmol.2015.00021
254. Perga, S., Martire, S., Montarolo, F., Navone, N. D., Calvo, A., Fuda, G., Marchet, A., Leotta, D., Chiò, A., Bertolotto, A. (2017). A20 in Multiple Sclerosis and Parkinson's Disease: Clue to a Common Dysregulation of Anti-Inflammatory Pathways? *Neurotoxicity Research*, *32*(1), 1-7. https://doi.org/10.1007/s12640-017-9724-y
255. Navone, N. D., Perga, S., Martire, S., Berchialla, P., Malucchi, S., Bertolotto, A. (2014). Monocytes and CD4+ T cells contribution to the under-expression of NR4A2 and TNFAIP3 genes in patients with multiple sclerosis. *Journal of Neuroimmunology*, *272*(1-2), 99-102. https://doi.org/10.1016/j.jneuroim.2014.04.017
256. Gilli, F., Navone, N. D., Perga, S., Marnetto, F., Caldano, M., Capobianco, M., Pulizzi, A., Malucchi, S., Bertolotto, A. (2011). Loss of braking signals

during inflammation: A factor affecting the development and disease course of multiple sclerosis. *Archives of Neurology*, *68*(7), 879-888. https://doi.org/10.1001/archneurol.2011.32
257. Ma, A., Malynn, B. A. (2012). A20: Linking a complex regulator of ubiquitylation to immunity and human disease. *Nature Reviews Immunology*, *12*(11), 774-785. https://doi.org/10.1038/nri3313
258. Shembade, N., Ma, A., Harhaj, E. W. (2010). Inhibition of NF-kappaB signaling by A20 through disruption of ubiquitin enzyme complexes. *Science*, *327*(5969), 1135-1139. https://doi.org/10.1126/science.1182364
259. Garg, A. V., Ahmed, M., Vallejo, A. N., Ma, A., Gaffen, S. L. (2013). The deubiquitinase A20 mediates feedback inhibition of interleukin-17 receptor signaling. *Science Signaling*, *6*(278), ra44. https://doi.org/10.1126/scisignal.2003699
260. Urbano, P. C. M., Aguirre-Gamboa, R., Ashikov, A., van Heeswijk, B., Krippner-Heidenreich, A., Tijssen, H., Li, Y., Azevedo, V. F., Smits, L. J. T., Hoentjen, F., Joosten, I., Koenen, H. J. P. M. (2018). TNF-α-induced protein 3 (TNFAIP3)/A20 acts as a master switch in TNF-α blockade-driven IL-17A expression. *The Journal of Allergy and Clinical Immunology*, *142*(2), 517-529. https://doi.org/10.1016/j.jaci.2017.11.024
261. Saxena, S., Lokhande, H., Gombolay, G., Raheja, R., Rooney, T., Chitnis, T. (2020). Identification of TNFAIP3 as relapse biomarker and potential therapeutic target for MOG antibody associated diseases. *Scientific Reports*, *10*(1), 12405. https://doi.org/10.1038/s41598-020-69182-w
262. Zhang, S., He, K., Zhou, W., Cao, J., Jin, Z. (2019). MiR-494-3p regulates lipopolysaccharide-induced inflammatory responses in RAW264.7 cells by targeting PTEN. *Molecular Medicine Reports*, *19*(5), 4288-4296. https://doi.org/10.3892/mmr.2019.10083
263. Huan, L., Bao, C., Chen, D., Li, Y., Lian, J., Ding, J., Huang, S., Liang, L., He, X. (2016). MicroRNA-127-5p targets the biliverdin reductase B/nuclear factor-κB pathway to suppress cell growth in hepatocellular carcinoma cells. *Cancer Science*, *107*(3), 258-266. https://doi.org/10.1111/cas.12869
264. Moore, M. J., Blachere, N. E., Fak, J. J., Park, C. Y., Sawicka, K., Parveen, S., Zucker-Scharff, I., Moltedo, B., Rudensky, A. Y., Darnell, R. B. (2018). ZFP36 RNA-binding proteins restrain T cell activation and anti-viral immunity. *ELife*, *7*, e33057. https://doi.org/10.7554/eLife.33057
265. Tiedje, C., Ronkina, N., Tehrani, M., Dhamija, S., Laass, K., Holtmann, H., Kotlyarov, A., Gaestel, M. (2012). The p38/MK2-driven exchange between tristetraprolin and HuR regulates AU-rich element-dependent translation. *PLoS Genetics*, *8*(9), e1002977. https://doi.org/10.1371/journal.pgen.1002977
266. Elhanati, Y., Murugan, A., Callan, C. G., Mora, T., Walczak, A. M. (2014). Quantifying selection in immune receptor repertoires. *Proceedings of the National Academy of Sciences*, *111*(27), 9875-9880. https://doi.org/10.1073/pnas.1409572111
267. Murugan, A., Mora, T., Walczak, A. M., Callan, C. G. (2012). Statistical inference of the generation probability of T-cell receptors from sequence

repertoires. *Proceedings of the National Academy of Sciences*, *109*(40), 16161-16166. https://doi.org/10.1073/pnas.1212755109
268. Robins, H. S., Campregher, P. V., Srivastava, S. K., Wacher, A., Turtle, C. J., Kahsai, O., Riddell, S. R., Warren, E. H., Carlson, C. S. (2009). Comprehensive assessment of T-cell receptor β-chain diversity in αβ T cells. *Blood*, *114*(19), 4099-4107. https://doi.org/10.1182/blood-2009-04-217604
269. Davis, M. M., Bjorkman, P. J. (1988). T-cell antigen receptor genes and T-cell recognition. *Nature*, *334*(6181), 395-402. https://doi.org/10.1038/334395a0
270. Brown, A. J., Snapkov, I., Akbar, R., Pavlović, M., Miho, E., Sandve, G. K., Greiff, V. (2019). Augmenting adaptive immunity: Progress and challenges in the quantitative engineering and analysis of adaptive immune receptor repertoires. *Molecular Systems Design Engineering*, *4*(4), 701-736. https://doi.org/10.1039/C9ME00071B
271. Rosati, E., Dowds, C. M., Liaskou, E., Henriksen, E. K. K., Karlsen, T. H., Franke, A. (2017). Overview of methodologies for T-cell receptor repertoire analysis. *BMC Biotechnology*, *17*(1), 61. https://doi.org/10.1186/s12896-017-0379-9
272. Bolotin, D. A., Poslavsky, S., Davydov, A. N., Frenkel, F. E., Fanchi, L., Zolotareva, O. I., Hemmers, S., Putintseva, E. V., Obraztsova, A. S., Shugay, M., Ataullakhanov, R. I., Rudensky, A. Y., Schumacher, T. N., Chudakov, D. M. (2017). Antigen receptor repertoire profiling from RNA-seq data. *Nature Biotechnology*, *35*(10), 908-911. https://doi.org/10.1038/nbt.3979
273. Farmanbar, A., Kneller, R., Firouzi, S. (2019). RNA sequencing identifies clonal structure of T-cell repertoires in patients with adult T-cell leukemia/lymphoma. *Npj Genomic Medicine*, *4*(1), 1-9. https://doi.org/10.1038/s41525-019-0084-9
274. Song, L., Cohen, D., Ouyang, Z., Cao, Y., Hu, X., Liu, X. S. (2021). TRUST4: Immune repertoire reconstruction from bulk and single-cell RNA-seq data. *Nature Methods*. https://doi.org/10.1038/s41592-021-01142-2
275. Chen, S. Y., Liu, C. J., Zhang, Q., Guo, A. Y. (2020). An ultra-sensitive T-cell receptor detection method for TCR-Seq and RNA-Seq data. *Bioinformatics*, *36*(15), 4255-4262. https://doi.org/10.1093/bioinformatics/btaa432
276. Giudicelli, V., Chaume, D., Lefranc, M. P. (2005). IMGT/GENE-DB: A comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Research*, *33*(Suppl 1), D256-D261. https://doi.org/10.1093/nar/gki010
277. Bolotin, D. A., Poslavsky, S., Mitrophanov, I., Shugay, M., Mamedov, I. Z., Putintseva, E. V., Chudakov, D. M. (2015). MiXCR: software for comprehensive adaptive immunity profiling. *Nature Methods*, *12*(5), 380-381. https://doi.org/10.1038/nmeth.3364
278. Tickotsky, N., Sagiv, T., Prilusky, J., Shifrut, E., Friedman, N. (2017). McPAS-TCR: A manually curated catalogue of pathology-associated T cell receptor sequences. *Bioinformatics*, *33*(18), 2924-2929. https://doi.org/10.1093/bioinformatics/btx286

279. Bagaev, D. V., Vroomans, R. M. A., Samir, J., Stervbo, U., Rius, C., Dolton, G., Greenshields-Watson, A., Attaf, M., Egorov, E. S., Zvyagin, I. V., Babel, N., Cole, D. K., Godkin, A. J., Sewell, A. K., Kesmir, C., Chudakov, D. M., Luciani, F., Shugay, M. (2020). VDJdb in 2019: Database extension, new analysis infrastructure and a T-cell receptor motif compendium. *Nucleic Acids Research*, *48*(D1), D1057-D1062. https://doi.org/10.1093/nar/gkz874
280. Amoriello, R., Greiff, V., Aldinucci, A., Bonechi, E., Carnasciali, A., Peruzzi, B., Repice, A. M., Mariottini, A., Saccardi, R., Mazzanti, B., Massacesi, L., Ballerini, C. (2020). The TCR Repertoire Reconstitution in Multiple Sclerosis: Comparing One-Shot and Continuous Immunosuppressive Therapies. *Frontiers in Immunology*, *11*, 559. https://doi.org/10.3389/fimmu.2020.00559
281. Fox, J., Muenchen, R., Putler, D. (2020). *RcmdrMisc: R Commander Miscellaneous Functions*. Available from: https://CRAN.R-project.org/package=RcmdrMisc
282. ImmunoMind Team. (2019). *immunarch: An R Package for Painless Analysis of Large-Scale Immune Repertoire Data*. Available from: https://CRAN.R-project.org/package=immunarch
283. Csardi, G., Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695. https://igraph.org.
284. R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Available from: https://www.R-project.org/
285. Gaujoux, R., Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics*, *11*, 367. https://doi.org/10.1186/1471-2105-11-367
286. Andrews, S. (2010). *FASTQC. A quality control tool for high throughput sequence data*. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
287. Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* Available from: https://ggplot2.tidyverse.org
288. Kassambara, A. (2020). *Ggpubr: «ggplot2» Based Publication Ready Plots.* Available from: https://CRAN.R-project.org/package=ggpubr
289. Amoriello, R., Chernigovskaya, M., Greiff, V., Carnasciali, A., Massacesi, L., Barilaro, A., Repice, A. M., Biagioli, T., Aldinucci, A., Muraro, P. A., Laplaud, D. A., Lossius, A., Ballerini, C. (2021). TCR repertoire diversity in Multiple Sclerosis: High-dimensional bioinformatics analysis of sequences from brain, cerebrospinal fluid and peripheral blood. *EBioMedicine*, *68*, 103429. https://doi.org/10.1016/j.ebiom.2021.103429
290. Rybakin, V., Westernberg, L., Fu, G., Kim, H. O., Ampudia, J., Sauer, K., Gascoigne, N. R. J. (2014). Allelic exclusion of TCR α-chains upon severe restriction of Vα repertoire. *PloS One*, *9*(12), e114320. https://doi.org/10.1371/journal.pone.0114320
291. Steinel, N. C., Brady, B. L., Carpenter, A. C., Yang-Iott, K. S., Bassing, C. H. (2010). Posttranscriptional silencing of VbetaDJbetaCbeta genes contributes to TCRbeta allelic exclusion in mammalian lymphocytes. *The*

Journal of Immunology, *185*(2), 1055-1062. https://doi.org/10.4049/jimmunol.0903099

292. Elhanati, Y., Sethna, Z., Callan, C. G., Mora, T., Walczak, A. M. (2018). Predicting the spectrum of TCR repertoire sharing with a data-driven model of recombination. *Immunological Reviews*, *284*(1), 167-179. https://doi.org/10.1111/imr.12665
293. Greiff, V., Menzel, U., Miho, E., Weber, C., Riedel, R., Cook, S., Valai, A., Lopes, T., Radbruch, A., Winkler, T. H., Reddy, S. T. (2017). Systems Analysis Reveals High Genetic and Antigen-Driven Predetermination of Antibody Repertoires throughout B Cell Development. *Cell Reports*, *19*(7), 1467-1478. https://doi.org/10.1016/j.celrep.2017.04.054
294. Greiff, V., Weber, C. R., Palme, J., Bodenhofer, U., Miho, E., Menzel, U., Reddy, S. T. (2017). Learning the High-Dimensional Immunogenomic Features That Predict Public and Private Antibody Repertoires. *The Journal of Immunology*, *199*(8), 2985-2997. https://doi.org/10.4049/jimmunol.1700594
295. Weber, C. R., Rubio, T., Wang, L., Zhang, W., Robert, P. A., Akbar, R., Snapkov, I., Wu, J., Kuijjer, M. L., Tarazona, S., Conesa, A., Sandve, G. K., Liu, X., Reddy, S. T., Greiff, V. (2022). Reference-based comparison of adaptive immune receptor repertoires. *bioRxiv*. https://doi.org/10.1101/2022.01.23.476436
296. Pavlović, M., Scheffer, L., Motwani, K., Kanduri, C., Kompova, R., Vazov, N., Waagan, K., Bernal, F. L. M., Costa, A. A., Corrie, B., Akbar, R., Al Hajj, G. S., Balaban, G., Brusko, T. M., Chernigovskaya, M., Christley, S., Cowell, L. G., Frank, R., Grytten, I., Gundersen, S., Haff, I. H., Hovig, E., Hsieh, P. H., Klambauer, G., Kuijjer, M. L., Lund-Andersen, C., Martini, A., Minotto, T., Pensar, J., Rand, K., Riccardi, E., Robert, P. A., Rocha, A., Slabodkin, A., Snapkov, I., Sollid, L. M., Titov, D., Weber, C. R., Widrich, M., Yaari, G., Greiff, V., Sandve, G. K. (2021). The immuneML ecosystem for machine learning analysis of adaptive immune receptor repertoires. *Nature Machine Intelligence*, *3*(11), 936-944. https://doi.org/10.1038/s42256-021-00413-z