Édi Pedro Correia

Complement Gene expression in the fetal brain after maternal immune activation in a rodent model

UNIVERSIDADE DO ALGARVE

Faculdade de Medicina e Ciências Biomédicas

2024

Édi Pedro Correia

Complement Gene expression in the fetal brain after maternal immune activation in a rodent model

Master's degree in Biomedical Sciences – Disease Mechanisms

This work was done under the supervision of:

Carolin Hoffmann phD (supervisor), Centre of Biomedical Research, University of Algarve

UNIVERSIDADE DO ALGARVE

Faculdade de Medicina e Ciências Biomédicas

2024

Complement Gene expression in the fetal brain after maternal immune activation in a rodent model

Declaração de autoria de trabalho.

Declaro ser o autor deste trabalho, que é original e inédito. Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída.

_____

Édi Pedro Correia

Dedicatória e Agradecimentos

Dedication

Dedication

To my family and friends, whose unwavering support and love have been the foundation of my journey.

To my wife Laura, for your patience and understanding during this challenging time and to my parents Zélia and Francisco, for their sacrifices and encouragement.

And especially to my daughter Sofia. This work is for you.

Acknowledgments

Acknowledgments

I would like to express my sincere gratitude to all those who contributed to the completion of this project.

First and foremost, I would like to thank God who has given me more than everything that I needed in my life through His unconditional love.

I would like to thank my supervisor, Carolin Hoffman, for their invaluable guidance, encouragement, and support throughout my research. Your expertise and constructive feedback have been instrumental in shaping this work.

Lastly, I would like to express my heartfelt appreciation to my family and friends for their love, patience, and understanding. Your encouragement kept me motivated during the most challenging times.

Thank you all for being a part of this journey.

Abstract

The developmental stage consists of a wide variety of fine-tuned biological processes. Disruptions in these mechanisms may induce alterations in the fetus, compromising its viability or causing malformations and/or deficits. Pathogens such as bacteria and viruses that cross the placenta have been well established as some of these disruptors with some of these being able to directly harm neurodevelopment. However, the immune system cells themselves have also been described as having a key role in the development of the brain being relevant in processes such as the cortical plate formation and neurogenesis. It is also well known that the mother's health habits and even diseases are likely to be mirrored on the fetus' development, since it relies on the mother's organism for survival. Recent studies have dwelled on the maternal immune system and its effects on development being more specifically focused on its activation by inflammatory signals. Maternal Immune Activation (MIA) during development has been described as having a causative relation to neurodevelopmental disorders, such as schizophrenia, autism specter disorder (ASD), anxiety and depression. Experimental models of MIA can reproduce disease biomolecular alterations and behavioral phenotypes in rodent models. Alongside these findings, the complement system – a group of proteins that play a role in the innate immune response – has also been discovered as having a key role in neurodevelopment, mainly in neuron progenitor cells (NPC) proliferation, migration and differentiation. With both MIA and the complement system having roles in neurodevelopment, this project sets out to evaluate the complement gene expression in fetal brain after MIA in the rodent model using bioinformatics to assess differential gene expression, co-expressed genes and the biological processes in which these are involved. Differential gene expression results of this project proved to be consistent with literature while gene ontology enrichment analysis of co-expressed genes resulted in significant neurodevelopmental biological processes.

Keywords:

Neurodevelopment, Maternal Immune Activation, Complement system, Bioinformatics

Resumo

O estágio de desenvolvimento embrionário e fetal envolve uma vasta e complexa gama de processos biológicos finamente ajustados e interdependentes. Durante esse período crítico, o desenvolvimento ocorre por meio de uma série de mecanismos coordenados que garantem a formação correta de tecidos e órgãos, bem como a organização estrutural e funcional de todos os sistemas do organismo. A interrupção ou disrupção desses mecanismos, mesmo que de forma sutil, pode resultar em alterações significativas no feto, comprometendo sua viabilidade e potencial de crescimento. Isso pode causar desde defeitos leves até malformações graves, além de deficiências funcionais, que podem afetar a saúde física e mental do indivíduo ao longo da vida. Fatores externos, como a exposição a agentes teratogénicos, bem como fatores internos, como mutações genéticas ou alterações no ambiente intrauterino, têm sido amplamente estudados como possíveis causas dessas disfunções.

Patógenos, como bactérias e vírus que conseguem atravessar a barreira placentária e invadir o ambiente intrauterino, foram bem estabelecidos na literatura científica como disruptores potenciais. Esses agentes podem causar prejuízos ao desenvolvimento ao interferir diretamente nos processos celulares do feto. Alguns desses patógenos demonstraram a capacidade de prejudicar o neurodesenvolvimento ao afetar a formação e a organização das células neurais e dos circuitos cerebrais. No entanto, não são apenas os agentes externos que desempenham esse papel; as próprias células do sistema imunológico do organismo materno também têm sido descritas como desempenhando um papel essencial no desenvolvimento do cérebro fetal. Essas células atuam de maneira crítica na regulação de processos como a formação da placa cortical e a neurogénese, que são fundamentais para o desenvolvimento adequado das estruturas cerebrais.

Sabe-se que os hábitos de saúde da mãe durante a gestação, como alimentação, consumo de substâncias nocivas (álcool, drogas, tabaco) e até mesmo o estado emocional e psicológico, têm um impacto profundo no desenvolvimento do feto. Doenças maternas, como diabetes, hipertensão, infeções e outras condições inflamatórias, são frequentemente refletidas no desenvolvimento fetal, já que o feto depende inteiramente do organismo materno para garantir sua nutrição, oxigenação e proteção. Assim, o estado de saúde geral da mãe é um fator determinante para o ambiente intrauterino. Recentemente, estudos têm explorado com mais detalhes como o sistema imunológico materno influencia o desenvolvimento do feto, com um

foco específico em sua ativação por sinais inflamatórios. Essa ativação é conhecida como Ativação Imunológica Materna (MIA, na sigla em inglês).

A ativação imunológica materna tem sido amplamente investigada como um fator potencial para o desenvolvimento de distúrbios neuropsiquiátricos e de neuro desenvolvimento, tanto em humanos quanto em modelos animais. Em modelos de roedores, foi demonstrado que a ativação imunológica materna está relacionada a uma maior incidência de condições como esquizofrenia, transtorno do espectro autista, transtornos de ansiedade e depressão. Esses resultados sugerem que o ambiente imunológico durante a gestação pode influenciar profundamente a maneira como os circuitos neurais se formam e se organizam, resultando em alterações comportamentais e cognitivas que persistem até a vida adulta. Essas descobertas sublinham a importância de entender como fatores imunológicos podem interferir no desenvolvimento do cérebro, já que as consequências podem ser amplas e duradouras.

Além dessas observações, o sistema complemento, que é uma rede complexa de proteínas que desempenham um papel fundamental na resposta imune inata, também tem sido destacado como um regulador chave do neuro desenvolvimento. O sistema complemento, originalmente identificado como um componente de defesa contra patógenos, foi descoberto como desempenhando papéis inesperados na proliferação, migração e diferenciação de células progenitoras neurais. Essas células progenitoras são responsáveis pela geração de novos neurônios durante o desenvolvimento fetal e pela formação das estruturas iniciais do cérebro. Quando há alterações na atividade do sistema complemento, esses processos podem ser afetados, levando a mudanças na organização das camadas neurais e na conectividade sináptica.

Com ambos os fatores – a ativação imunológica materna e o sistema complemento – desempenhando papéis críticos no neuro desenvolvimento, há um crescente interesse em entender como esses sistemas interagem e afetam a formação do cérebro em diferentes estágios do desenvolvimento. Este projeto de pesquisa, portanto, propõe investigar a expressão gênica de genes relacionados ao sistema complemento no cérebro fetal após a ativação imunológica materna, utilizando um modelo de roedores para esse estudo. A abordagem utilizada envolverá o uso de ferramentas de bioinformática para avaliar a expressão diferencial de genes, identificar genes co expressos e investigar os processos biológicos nos quais esses genes estão envolvidos.

Os resultados obtidos com a análise da expressão diferencial de genes neste projeto foram consistentes com a literatura existente, indicando que a ativação imunológica materna afeta

significativamente a expressão de genes do sistema complemento no cérebro fetal. Além disso, a análise de enriquecimento de ontologia gênica, que examina as funções biológicas associadas aos genes coexpressos, revelou processos biológicos no foro do neuro desenvolvimento significativos, como a regulação da neurogénese, a diferenciação neuronal e a organização do córtex cerebral. Estes resultados reforçam a hipótese de que o sistema complemento desempenha um papel ativo no desenvolvimento do cérebro, não apenas como um mediador da resposta imune, mas também como um regulador dos processos de desenvolvimento celular e estrutural.

Dessa forma, o presente estudo contribui para um entendimento mais abrangente de como fatores imunológicos maternos, como a ativação imunológica materna e o sistema complemento, influenciam o desenvolvimento cerebral. A compreensão dessas interações poderá abrir novos caminhos para a identificação de biomarcadores precoces de distúrbios neuropsiquiátricos e para o desenvolvimento de estratégias terapêuticas que visem minimizar os efeitos negativos da disfunção imunológica materna no desenvolvimento do cérebro fetal.

Palavras-chave:

Neuro desenvolvimento, Ativação Imunológica Materna

Table of Contents

List of Figures

List of Abbreviations, Acronyms & Symbols

.cel: Affymetrix cell file format

Adn: Adenosine Deaminase

ADHD: Attention Deficit Hyperactive Disorder

ASD: Autism Spectrum Disorder

C1qa: Complement Component 1, q subcomponent, a

C1qb: Complement Component 1, q subcomponent, b

C1qc: Complement Component 1, q subcomponent

C3: Complement Component 3

C3a: Complement Component 3a

C3ar1: Complement Component 3a Receptor 1

C4a: Complement Component 4a

C4bpa: Complement Component 4b Binding Protein

C5: Complement Component 5

C6: Complement Component 6

C7: Complement Component 7

C8b: Complement Component 8, beta

C8g: Complement Component 8, gamma

C9: Complement Component 9

Clu: Clusterin

CNS: Central Nervous System

CFD: Complement Factor D

CFI: Complement Factor I

Cfh: Complement Factor H

CRP: C reactive protein

CXCL8: interleukin-8

DAMPs: Damage-associated molecular patterns

DEG: Differentially Expressed Genes

DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition

Fncb: Fibronectin Type III Domain Containing

GAD-1: glutamate decarboxylase-1

GD: Gestational day

GEO: Gene Expression Omnibus

Gtf3c2: General Transcription Factor IIIC Subunit 2

GO: Gene Ontology

HGNC: HUGO Gene Nomenclature Committee

IL: Interleukin

IFN-γ: Interferon-gamma

LPS: Lipopolysaccharide

Lcn12: Lactoferrin

MA-plot: M-Array Plot

MIA: Maternal Immune Activation

NCBI: National Center for Biotechnology Information

NDDs: Neurodevelopmental Disorders

NFκB: Nuclear factor kappa B

NPC: Neuroprogenitor Cells

PAMPs: Pathogen-associated molecular patterns

PCA: Principal Component Analysis

Poly(I:C): Polyinosinic:plycytidylic acid

PPI: Prepulse Inhibition

PVALB: parvalbumin

P-value: Probability Value

RMA: Robust Multi-Array Average

Rps7: Ribosomal Protein S7

SCZ: Schizophrenia

soft: A file format for GEO data

SST: somatostatin

SSTR2: somatostatin receptor 2

Tbxa2r: Thromboxane A2 Receptor

TLR: Toll-like receptor

TNF: tumour necrosis factor

Z-scores: Standard scores indicating how many standard deviations an element is from the mean

1. Introduction

1.1. The role of the immune system in neurodevelopment

The first weeks of gestation are crucial to the development of the neurons and subsequently, the brain. Relying on a myriad of mechanisms and signaling pathways, the formation of the neuronal network of the body's most complex and fascinating organ is a fragile yet fine-tuned process. One of the most commonly associated mechanisms that act on the central nervous system (CNS) development are the morphogens, which determine cell differentiation through their temporal and spatial gradient-like expression (Pagliaro et al., 2023). This peculiar expression pattern constitutes in precise regions which define distinct functional domains due to the differentiation signals of the morphogen gradient (Sansom & Livesey, 2009). In the cortex alone, a subarea of the brain, there are hundreds of different neuron subtypes. This neuronal diversity is key to the functional development of the brain, changing in cell morphology, connectivity, electrophysiology and more (Tiberi et al., 2012). All these different factors need to be in unison to create a complex vast network that is capable of the brain's most convoluted tasks such as information processing, motor function, memory, speech and much more.

Like the brain's neuronal diversity, the immune system also has the ability of having such specifications. This being for the identification of several different antigens, or in the release of specific antibodies from plasmocytes to combat already known pathogens. The immune system is even capable of memorizing previous pathogens with the aid of the memory B-cells. Additionally, a common link between these two systems has been discovered, showing that the immune system also has a role in the development of the CNS. Cells such as microglia, the main immune system cells in the brain, have a key role in the formation of the cortical plate and neurogenesis when these invade the early brain tissue (Morimoto & Nakajima, 2019). Other immune cells also infiltrate the brain such as the T and B-cells in the late stage of development and reside mostly on the pial surface and ventricle. These cells have also been studied as having a distinct role in synaptic pruning, oligodendrogenesis and myelination.

These findings have established an interest for the since now, hidden role of the immune system on the neurodevelopment, their influence on malformations, short-term and long-term effects on the functioning of the brain and even neuropsychiatric disorders.

It is well known that pathogens that disrupt the brain development mechanisms may result in life-time effects and may even result in the non-viability of the fetus. Certain viruses and bacteria such as zika, rubella and toxoplasma that cross the placenta and the blood brain barrier, are well-known and studied regarding their harm on the fetus' brain. Maternal Immune Activation (MIA), which consists of the immune response from the maternal body facing infection has been a topic of interest in several epidemiological studies that have associated this phenomenon with the onset of neurodevelopmental disorders. These being autism spectrum disorder (ASD), schizophrenia (SCZ), epilepsy, cerebral palsy, anxiety and depression (O'Connor & Ciesla, 2022).

1.2. Neurodevelopmental disorders

Neurodevelopmental disorders or NDDs are defined as "a group of conditions with onset in the developmental period" with a wide range of symptoms which impact personal, social, academic, or occupational functioning. With this term's recent entrance on the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5), an interest in these disorders has erupted, when in the past, these disorders were often associated with the subject's intellectual capacities provoking the creation of a stigma against individuals affected by them. However, with NDDs being classified as legitimate mental disorders and a growing body of research, there has been a change of perspective in understanding. Modern insights emphasize the complex interplay of genetic, neurological and environmental factors in play that aggravate these conditions (Cainelli & Bisiacchi, 2023).

Although, these factors have been implicated in the development of NDDs, they are not sufficient on their own to induce the onset of these conditions. Some NDDs such as, schizophrenia have been associated with a two-hit hypothesis regarding the aggravation of the disease (Bayer et al., 1999). This comes from the fact that SCZ typically manifests in early to late adulthood during the 20s, indicating that events occurring during adolescence may play a role in the onset of the disorder – this is mainly associated with psychiatric trauma. In summary, the two-hit hypothesis combines a predisposition (first hit) with an exacerbation that leads to the emergence of the disorder (second hit). In the case of SCZ, innate immune dysfunction and neuroinflammation have been observed as possible key factors in inducing the first hit of this disorder (Feigenson et al., 2014).

These predisposing factors for NDDs have also been traced back to epigenetic sources (Ciptasari & van Bokhoven, 2020), immune dysfunction through cytokines and microglia leading to neuroinflammation (Hughes et al., 2023) and previously mentioned, MIA. These external factors lead to brain structure changes, causing an abnormal neural network impairing the basal neuronal functions, inciting symptoms of anxiety, inattentiveness, restlessness, inability to focus and other neuropsychiatric ailments.

1.3. Maternal Immune Activation in humans

Research on the effects of MIA in humans has been short-sighted due to the ethical barriers in accessing and analyzing sensitive samples like fetal brain tissue, maternal blood and placenta. To overcome these challenges, alternative methods of study are implemented such as population-based case-control and prospective investigations. These studies examine MIA patients, their offspring's symptoms, neurodevelopmental disorders and associated risk factors. Risk factors include epigenetic influences such as environmental factors, socioeconomic status, lifestyle (e.g. sleep and diet), as well as disease states like obesity, pre-eclampsia, diabetes and depression.

The suggested path of action for the cause of Maternal Immune Activation is the activation of the cytokine signaling pathway by pathogen-associated molecular patterns (PAMPs) and damage-associated molecular patterns (DAMPs). These molecular patterns cause alterations in the levels of inflammatory signals such as cytokines, chemokines, C reactive protein (CRP) and complement system proteins. These inflammatory signals are disseminated by the circulatory system and, during pregnancy, are transmitted to the fetus altering its immune system. Human umbilical vein endothelial cells isolated from newborns with mothers living with inactive Systemic Lupus Erythematosus – a chronic disease that alters the normal function of the immune system due to the production of autoantibodies – showed an upregulation in pro-inflammatory receptors and secretion of cytokines such as interleukin-6 (IL-6) and interleukin-8 (CXCL8) (Rodriguez et al., 2008). Other mechanisms proposed for the transmission of the effects of MIA to the fetus causing abnormal CNS development are the maternal innate/adaptive immune system and complement pathways (Knuesel et al., 2014).

While cytokines are often associated with inducing inflammation, some also have anti-inflammatory roles that help maintain immune system homeostasis. However, when this balance is disrupted, complications can arise, potentially impacting other systems. During fetal

brain development, cytokines and their receptors are constitutively expressed, highlighting their role in modulating and regulating normal brain development. In the adult central nervous system, cytokines are expressed in glial and neuronal cell types, further underscoring their importance in maintaining neurological function (Meyer et al., 2009). Furthermore, cytokines, chemokines, complement proteins, and major histocompatibility complex proteins are also present at synapses, where changes in their levels can disrupt synaptic function and hinder the refinement of neural connectivity (Estes & McAllister, 2015) (Estes & McAllister, 2016).

Epidemiological studies have consistently supported preclinical findings, linking changes in inflammatory biomarker levels in maternal and offspring tissues with the diagnosis of ASD and Attention Deficit Hyperactive Disorder (ADHD) in offspring. Neuroanatomical changes, including alterations in the amygdala and brain connectivity, were observed in offspring of mothers with elevated IL-6 levels in their serum during pregnancy (Graham et al., 2018). Postnatal studies on individuals diagnosed with neurodevelopmental disorders (NDDs) show increased expression of cytokines like IL-6, tumour necrosis factor (TNF), and interferon-gamma (IFN-γ) in the central nervous system and cerebrospinal fluid (Xu et al., 2015). However, conflicting results have also been reported, such as studies linking increased midgestational and neonatal IL-4 levels to a higher incidence of ASD (Goines et al., 2011) (Krakowiak et al., 2017), while another study found the opposite, with reduced levels of IL-4 (Abdallah et al., 2012).

Given these conflicting findings, it is crucial to conduct experimental studies using well-controlled Maternal Immune Activation (MIA) models to disentangle the complex relationship between prenatal inflammation and neurodevelopmental outcomes. In an effort to reduce unpredictability and inconsistencies, methodological elements that lead to incoherent study outcomes have been identified. These include immunogen type, dosage, timing, caging techniques, and rodent strains (Harvey & Boksa, 2012). Researchers can directly evaluate the effects of manipulating particular immune pathways on fetal brain development and child behavior by using MIA models. These models aid in elucidating causality, identifying important mediators of neurodevelopmental risk, and investigating prospective therapeutic approaches by offering a controlled setting for the replication and testing of epidemiological relationships. Our knowledge of how maternal inflammation affects the neurodevelopment of kids is still lacking in the absence of experimental validation, which restricts the creation of focused preventative and therapeutic approaches.

1.4. Animal models of Maternal Immune Activation

A number of MIA models, some especially made for neurodevelopment, have been created to further examine the effects of the immune system. Understanding the processes of maternal immune activation and how it affects the neurodevelopment of the offspring has been greatly aided by the use of animal models, such as mice, rats and non-human primates. Regarding MIA induction, the introduction of viruses has been especially helpful in simulating the impact of maternal infections on the development of the fetal brain.

Rodent models have long been one of the most commonly used animal models in science due to their genetic resemblance to humans and ease of control over experimental circumstances. While the use of non-human primates, such as macaques, frequently yields more similar results due to the species' evolutionary proximity to humans, it is more expensive and has ethical implications. When using a rodent model for MIA, pregnant mice or rats are usually exposed to immune stimulants, like bacterial endotoxins or synthetic viral mimics, in order to elicit an immunological response. Sub-lethal dosages of live viruses were initially used, such as the influenza virus (Fatemi et al., 2004), but because handling these pathogens has biosafety issues, efforts were made to find other means to stimulate the immune system (Massarali et al., 2022).

Other MIA models, although having the same objective of inducing MIA, can be designed for different finalities. For example, to assess and identify individual and specific factors responsible for MIA-induced neurodevelopmental changes, the cytokine-specific model is indicated. This consists of injecting pro-inflammatory cytokines such as IL-6 or IFN-γ (Smith et al., 2007) or drugs that act as cytokine receptor agonists (Sheng & Tobet, 2024), into pregnant animals to study their isolated effects on fetal brain development. Similarly to this model, the cytokine-transfer model comprises the transfer of endogenous maternal cytokines to the developing fetus, either through placental pathways or artificial means.

To simulate real-world infection and the recruitment of the immune system, exogenous polymers have been found to be useful. Polyinosinic:plycytidylic acid [Poly(I:C)], a synthetic analog of double-stranded RNA, mimics viral infections by activating toll-like receptor 3 (TLR-3) when injected in pregnant mice. This leads to an innate immune response by the activation of the TLR cascade culminating in the activation of the transcription factor nuclear factor kappa B (NFκB) and the transcription of pro-inflammatory cytokines. On the other hand,

the lipopolysaccharide (LPS) model, a constituent of the outer membrane present in Gram-negative bacteria, resembles a bacterial infection. This polymer connects to and activates toll-like receptor 4 (TLR-4) signaling cascade inducing an acute and fast inflammatory response accompanied by elevated cytokine levels such as tumor necrosis factor alpha (TNF-a) and IL-1B (Bucknor et al., 2022).

1.5. Maternal Immune Activation effects in the nervous system

As different MIA induction protocols emanate different results, the timing of gestational period in which MIA is actuated also alters the results of the effects in brain development. Despite both mice and rats being rodents, embryonic development varies through species and therefore, the brain developmental process. This means that certain brain development processes occur in different gestational days across rats, mice and humans. Therefore, special attention related to the brain development phases has to be taken into consideration when studying MIA interference with neurodevelopment. For example, to study

Mouse GD	Rat GD	Human GD	Brain developmental process
7	8	17-23	Neural plate and neural groove formation.
8	9	23-32	Neural tube and neural crest formation.
9	10	28-42	MICROGLIAL INVASION. Formation of three primary brain vesicles (forebrain, midbrain, hindbrain).
10	11	35-52	NEUROGENESIS Retinal ganglion cells; cerebral hemispheres form; posterior commissure develops; initiation of the olfactory bulbs; formation of five secondary brain vesicles.
11	12	43-61	Cranial sensory nuceli develops; external and internal capsules develop (white matter tracts); Corticol layer IV starts development; hippocampal neurons begin to develop.
12	13	50-68	Optic nerve begins development; amygdala begins development; cerebellar primordium forms; substantia nigra begins development; entorhinal cortex begins development.
13	14	55-74	Choriod plexus forms; corticol layer V begins to develop; glial precursors generated.
14	15	60-84	Fornix develops; CA1, CA2 and denate gyrus (hippocampus) develop; Anterior commissure forms; corticol layers II and III start to develop.
15	16	67-91	Cerebellar enlargement; nucleus accumbens development; cortical layer V development ends; hippocampal commissure forms.
16	17	72-100	Corpus callosum forms.
17	18	79-107	Corticol layer IV ends; denate gyrus becomes visible.
18	19	84-117	ASTROGENESIS. Corticol layer II and III finish development; cerebellar lobulisation and lamination.
	20		
	21		

the effects of MIA in neurogenesis, the immune system must be active during the tenth gestational day in mice and during the eleventh to twelfth gestational day in rats (Figure 1).

Figure 1 - Brain development chronology marked by gestation days (GD) in the mouse, rat and human – adapted from Woods et al., 2021

The timing of MIA activation has been proven to alter neurodevelopment trajectory. A study where groups of mice were injected with Poly(I:C) in GD 9 and 17, to simulate early and late gestation MIA respectively and then submitted to MRI scans, showed alterations in brain development regarding brain volume in different regions which are implicated in neuropsychiatric disorders and NDDs. In early gestational MIA mice, these brain regions were decreased in volume at post-natal day 21 which then increased relative to controls between days 38 and 60, normalizing at day 90. This was observed in the hippocampus, cingulate cortex, nucleus accumbens and striatum. In late gestational MIA mice, these resulted in a flatter developmental trajectory in the nucleus accumbens, auditory cortex, reticular nucleus, subiculum and hypothalamus. Adversely, the amygdala volume decreased in later adulthood. With these findings, early MIA exposure is predicted to have a largest effect in neuroanatomy (Guma et al., 2021). In another study, mice were injected with Poly(I:C) at GD 10 and 19 and subsequently tested for alterations on inhibitory markers previously observed in schizophrenia patients. The evaluated inhibitory markers were glutamate decarboxylase-1 (GAD-1), parvalbumin (PVALB) and somatostatin (SST) on different subareas of the brain like the cortex, striatum and hippocampus. These markers were chosen due to the role of inhibitory interneurons in the regulation of motor, cognitive and limbic function. These functions are observed as compromised in people with schizophrenia, being postulated to be a contributor to the disease. In this study, these inhibitory markers were reduced in MIA in a subregion-specific manner whereas, the somatostatin receptor 2 (SSTR2) was reduced in a timing-specific manner mainly in late-gestational MIA in the striatum (Rahman et al., 2020).

Nevertheless, studies conducted with these models have produced offspring with neuropsychiatric symptoms, regarding their behavior and social skills, commonly associated with ASD and SCZ, somewhat confirming previous epidemiological studies already mentioned. Matter of fact, when evaluating the neurite density in the brains of these offsprings, a decrease in neurite outgrowth was observed, consistent with post-mortem brains of human patients suffering from SCZ and/or ASD (Glantz & Lewis, 2000). Behavior tests conducted usually consist in the acoustic startle response, delayed non-match to position testing and drug-

induced locomotor activity assessments. On the acoustic startle response test, the mice startle reflex is monitored after a loud noise is played. Prepulse inhibition (PPI) is a phenomenon observed when a softer sound is played before (prepulse) the startle inducing loud noise (pulse) inducing an alert state in the mice which inhibits the startle reaction on the loud pulse. On SCZ patients, the ability to filter out sounds and distractions is compromised leading to a decrease in the PPI. In another study, other neuroanatomical changes were identified such as corticogenesis deficits and decreased hippocampal volume in mice neonates where the pregnant mothers were infected with an adapted human influenza virus. These neural abnormalities accompanied with decreased γ-aminobutyric acid (GABA) marker expression and changes in the behavior, predominantly on the decrease of the PPI, were linked with schizophrenia (Kneeland & Fatemi, 2013).

For the delayed non-match position test, the working memory is assessed. The test starts with two levers being presented to the rodent, one activated and another deactivated. Upon pressing the deactivated lever, a reward is dispensed. When the rodent collects the reward, a short delay is established where the levers are reset to the deactivated position, this delay induces the recruitment of working memory. For the rodent to receive another reward, it must press the lever that was not previously pressed, hence the non-match position (Figure 2). A study that used this test with MIA induced rats by Poly I:C injection at gestational day (GD) 10 and 19, recorded a significant reduction in performance in GD19 MIA rats when comparing to GD19 Control rats. Curiously, GD10 MIA rats showed no significant differences when compared to GD10 Controls, suggesting that working memory deficits specific to MIA are predominantly associated with late but not early gestation (Meehan et al., 2017).

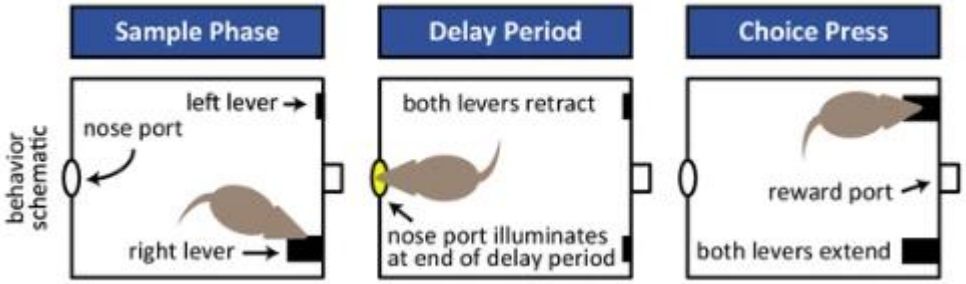


Figure 2 – Diagram describing the delayed non-match to position test (DNMPT), considered an appropriate behavior test for schizophrenia-like symptoms in animal models, evoking the working memory to collect a reward – adapted from Akhlaghpour et al., 2016.

To better understand the effects of MIA on neurogenesis, investigators used human induced pluripotent stem cells (iPSC or hiPSC), which can easily be differentiated into neuroprogenitor cells (NPC), mimicking the embryonic cells that lead to the formation of the fetus' brain. The use of iPSC as a model of embryonic development is highly indicated due to the pluripotent nature of these cells. Studies with NPC derived from hiPSC, which have been exposed to inflammatory cytokines resulted in increased neurite outgrowth, increased branching and decreased dendritic spine density (Kathuria et al., 2022).

Additionally, structural and anatomical abnormalities of the brain in MIA offspring were observed. MIA in rodents and nonhuman primates (NHP) resulted in enlarged ventricles alongside reductions in volume of various brain regions (da Silveira et al., 2017). Other studies found no changes in total brain volume however, relative volume of several regions was identified as increased or decreased unilaterally (Richetto et al., 2017). A constant change in the increase or decrease of brain volume has not yet been described although, these changes in volume have been correlated with the MIA response to influenza in NHP (Short et al., 2010). Histologically, neuron density seems to be affected in cases of MIA depending on time of gestation (Bergdolt & Dunaevsky, 2019). In early stages of gestation, MIA resulted in the total neuron density decrease in the dentate gyrus and subiculum in the offspring of the MIA porcine model (Antonson et al., 2018). Contrary to this, late gestation induced MIA showed a slight increase in neuron density in the corpus callosum (Duchatel et al., 2016). Despite these inconsistent results, one can affirm that MIA induces changes in brain structure and formation while also impacting neuron density and viability in the developmental stage.

Due to these studies, MIA is thought to be a possible risk factor for neuropsychiatric disease, due to the elevated presence of cytokines such as interleukin-6 (IL-6) and interferon gamma (IFN-γ) which caused morphological and transcriptomic alterations in the neuroprogenitor cells (Warre-Cornish et al., 2020). This leads to interest in studying the different strands of the innate immune response system.

1.6. The Complement System and its role in innate immune response

The complement system consists of a group of plasma and membrane-associated serum proteins that act on the presence of foreign organisms like bacteria, parasites, viruses, foreign objects and even in the presence of damaged tissue. This system acts on the general and immediate response to the threats presented to the organism, making itself part of the innate immune response, rather than the specific and specialized adaptive immune system which contain the lymphocytes T and B-cells and other leukocytes.

Complement consists of a group of proteins which work together to activate a strong inflammatory and cytolytic response by recruiting the anaphylatoxins, the pro-inflammatory signaling proteins, the opsins who surround the surface of the foreign body/pathogen and the membrane attack complex (MAC) proteins which form pores in the membrane of the intruder and proceed to its lysis. There are three pathways which all lead to the aforementioned inflammation, opsonization and lysis: the classical pathway, the lectin pathway and the alternative pathway.

The classical pathway works by identifying a connection between an antibody and an antigen of the pathogen. When the protein C1q binds to an antibody attached to an antigen it associates with the complement proteins C1r and C1s which lead to a cleaving cascade of more complement proteins. The lectin pathway continues this trend differing in the first step of the identification of the pathogen. While the classical pathway relies on the presence of pre-established connection between antigen and antibody, the lectin pathway identifies pathogens by using its namesake, mannose-binding lectin (MBL) receptors. Different to the previous two methods, the alternative pathway initiates when the complement protein C3, one of the most prevalent in the plasma, is spontaneously hydrolyzed by a pathogen. This hydrolyzed C3 protein ($C3H_2O$) then binds to factor B and through a series of cleavages, produces the protein C3b which then creates a positive feedback loop, amplifying the immune response. All these three pathways then converge for the generation of the proteins C3 and C5 convertases which cause inflammation, lysis, and opsonization (Dunkelberger & Song, 2010).

Finally, the complement pathway converges into the creation of the membrane attack complex (MAC) a pore forming structure that provokes damage into the lipidic bilayer of the plasma membrane of several pathogens, mainly Gram-negative bacteria, enveloped viruses and parasites. This complex is characterized as an oligomer formed by C5α, C5β, C5a, C6. C7 and several copies of the protein C9. This structure can either be directly inserted into the

pathogen's membrane via endocytosis or gradually formed by the MAC complement proteins. The fully assembled MAC creates a pore allowing the efflux of K^+ and the influx of ATP and Ca^{2+}, this leads to the secretion of inflammatory cytokines due to the activation of inflammasomes and the activation of apoptosis and necroptosis signaling pathways (Xie et al., 2020).

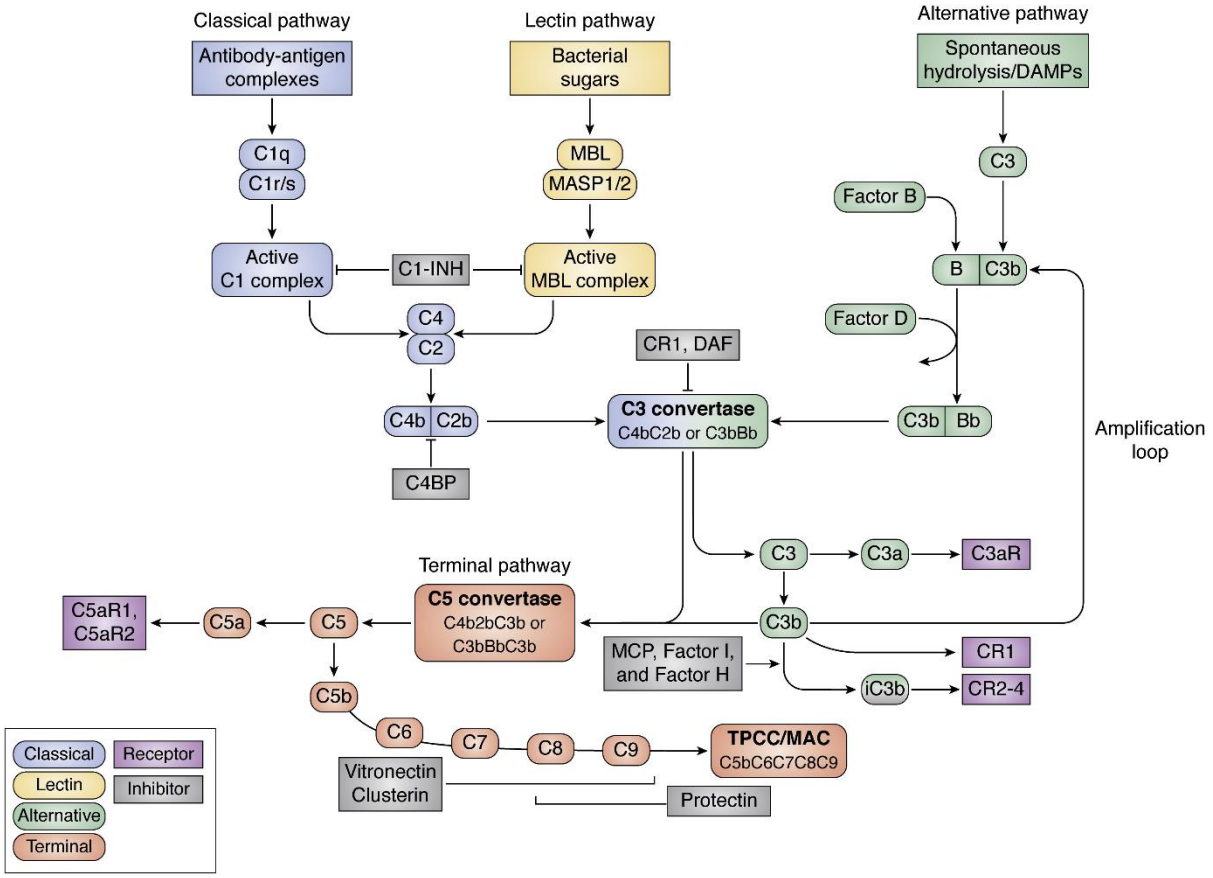


Figure 3 – Complement system pathways depicting classical, lectin and alternative pathway converging into the terminal pathway and the creation of the MAC - adapted from Warwick et al., 2021

1.7. Intracellular complement signaling

Although being described as a liver-derived system, complement activation and effector functions present a more compartmentalized role. While being discovered and originally perpetrated as an extracellular system, evidence regarding the intracellular role of the complement system have emerged across a broad range of cell populations and tissues. This intracellular complement system, coined as the 'complosome', has been originally found in T cell leukocytes contributing to cell homeostasis and metabolic pathway regulation. Ongoing

work on the complosome continues to reveal clues that complement activity might have novel and non-canonical functions, as the complosome participates in basic cellular processes, such as metabolism, autophagy and gene expression. Interestingly, up and downregulation of the complosome components have been also associated with the tissue occupying capacity of the immune cells. As previously established, and in similarity to cytokines, immune cells have had a beneficial role in the brain's development and normal function. In addition, the metabolic state of these cells and their capacity to produce interferon cytokines contribute to the brain's activities determining behavior. Therefore, when observing the immune system's role in the incidence of neurodevelopmental and neuropsychiatric disorders, the intracellular complement system, or complosome, must be taken into account.

In cells where the complosome is identified, the origins of these complement proteins are often observed as self-produced by encoding the same genes that give rise to the liver-derived circulating extracellular complement system. The main complosome proteins, C3 and C5, can also be activated intracellularly as these can be cleaved into C3a, C3b and C5a by specific proteases and/or by intracellular C3 and C5 convertases beneath the plasma membrane and on the surface of the subcellular compartments. These compartments being the endoplasmatic reticulum, outer mitochondrial membrane and the nucleus. These core complement proteins are also identified in immune cells such as antigen presenting cells (APCs) and T cells where, when secreted, can "kickstart" the inflammation process by activating the extracellular complement system. Additionally, although being generally derived from cell-intrinsic expression, intracellular C3 can be collected from the extracellular space via opsonized pathogens, C3(H2O) uptake and C3 re-ingestion that had been previously secreted. When cleaved, the C3a and C5a proteins bind to mitochondrial, lysosomal and/or endosomal C3aR and C5aR1 where signaling pathways are triggered similar to the classical pathway in the extracellular complement (Kunz & Kemper, 2021).

The main identified function of the complosome is the cell metabolism control, where the lysosomes and mitochondria play significant roles in this process. This is observed in T cell populations denominated $CD4^+$, commonly associated with T helper cells, and $CD8^+$, the cytotoxic T cell counterpart, referring to the type of the surface receptor CD. In these cells, the lysosome contains stores of C3 which are continuously cleaved into C3a and C3b by the protease cathepsin L (CTSL). The presence of intracellular C3a engages the lysosomal C3a receptor (C3aR) inducing the low-level activation of mammalian target of rapamycin (mTOR), supporting peripheral T cell survival. This is confirmed by previous studies with mice deficient

in C3aR or C5aR on APCs and/or T cells which resulted in the decreased recruitment of T helper type 1 (Th1) cells and subsequent deficient immunity. In addition, upon T cell receptor (TCR) activation, intracellular C3a and C3b rapidly shuttle to the cell surface where they bind to surface expressed C3aR and C3b receptor CD46 – an ubiquitously expressed transmembrane receptor. While this receptor is expressed on all nucleated human cells, mice and rats lack CD46 on somatic cells and an exact functional and complement derived homolog to human CD46 has not been discovered in rodents. Therefore, signaling events and functions controlled and induced by CD46 are human-specific. CD46 also has four distinct isoforms which carries either one of two distinct isoforms termed as CYT-1 and CYT-2. The autocrine CD46 activation mediated by TCR activation induces the rapid upregulation of the $CD46^{CYT1}$ isoform which has its intracellular tails processed by the γ-secretase complex. These intracellular tails translocate to the nucleus acting as transcription regulators activating the expression of genes encoding the glucose transporters such as SLC2A1 gene for the glucose transporter 1 (GLUT1) protein, which is a key element for the nutrient influx required for Th1 cell activation and IFN-γ production. Furthermore, in the mitochondrial membrane, C3aR and C5aR1 stimulation controls Ca^{2+} influx and the direction of the electron transport chain which results in the production of reactive oxygen species (ROS) versus ATP and the cellular induction of glycolysis and oxidative phosphorylation (OXPHOS). These trigger the assembly of the NLR family pyrin domain 3 protein (NLRP3) inflammasome, which controls the processing and secretion of IL-1β, a cytokine which determines the duration of Th1 responses in an autocrine and paracrine fashion while sustaining the secretion of IFN-γ (West & Kemper, 2023).

Conversely, resting and circulating T cells present the $CD46^{CYT-2}$ isoform which has been described as preventing T cell activation. This is due to its capability of binding and sequestering the Notch-1 ligand, Jagged-1 (Le Friec et al., 2012). By preventing the interaction between Jagged-1 and its receptor Notch-1, the signaling pathways that would lead to Th1 differentiation remain inactive resulting in T cells remaining quiescent. Although, when these T cells are activated due to TCR actuation, metalloproteinases induce CD46 shedding, irreversibly removing the receptor by proteolysis. This leads to the release of Jagged-1 and the removal of the "brake" on the Notch-1 signaling pathway which leads to Th1 differentiation. To cease Th1 response, the anti-inflammatory and immunosuppressive cytokine IL-10 has emerged as an important mediator. This has been observed in expanded human Th1 cells, where the IL-2 receptor (IL-2R) signals together with CD46 induce co-expression of IL-10. However, the definite mechanisms downstream of the IL-2R and CD46 that lead to this

cytokine's production are not fully described yet. Nevertheless, they have been traced to the reversion of CD46 isoforms to a predominant $CD46^{CYT-2}$ domain. As described earlier, this leads to the downregulation of genes encoding nutrient channels including SLC2A1, which encodes GLUT-1, and SLC7A5, which encodes large neutral amino acid transporter 1 (LAT1) leading to a limitation in nutrient influx. This reversion also reduces mTOR complex 1 (mTORC1) activity which provokes the return of the T cell to a metabolically OXPHOS-dependent state and subsequently reversing the cell to a quiescent state (Zheng et al., 2007). This is due to the CD46 intracellular domain CYT-1 capability of acting as transcription regulators when cleaved. Curiously, this protein lacks a DNA-binding site, thus being described as a transcription regulator and not a transcription factor, with the exact CD46-controlled gene activation mechanisms not being currently defined. In addition to the negative control of Th1 cells provided by CD46, T-cell intrinsic stores of complement protein C5 also play a role in this process. The already cleaved C5a complement protein is processed to C5a-desArg, a des-Arginized form of C5a. C5a-desArg then engages with C5aR2, an inhibitory receptor for C5a, in an autocrine/paracrine fashion. This interaction leads to C5aR2 signals which reduce ROS production, NLRP3 inflammasome activation and subsequently, IL1-β secretion shutting down Th1 responses in the periphery.

The complosome also presents itself to be an integral part of cytotoxic T cells responses. In human $CD8^+$ T cells, aside from strong oxidation and phosphorylation induction, CD46 activation is also needed for the raised glycolysis levels and fatty acid synthesis for ideal IFN-γ secretion (Arbore et al., 2018). Additionally, although $CD8^+$ T cells express NLRP3, the inflammasome is not required for normal IFN-γ secretion or cytotoxicity, indicating important but distinct roles of complosome-inflammasome interactions in $CD4^+$ and $CD8^+$ T cells. Furthermore, intracellular complement components can negatively regulate cytotoxic T cell responses, this being the complement-derived pattern recognition protein C1q, which originates from the extracellular environment, acting on human and mouse $CD8^+$ T cell mitochondria to inhibit glycolysis in the setting of infection.

In addition, a study regarding upstream regulators for C3 and C5 gene transcription has led to the observation that different tissues present different patterns of complement gene activity, suggesting an influence by the environment on how these genes are expressed. Alongside this analysis, researchers found that in both $CD4^+$ and $CD8^+$ T cells, the induction of C3 gene expression was dependent on the integrin leukocyte adhesion factor 1 (LFA-1) signaling pathway. This LFA-1 integrin is composed of two distinct integrin protein chains

aLb2 or CD11a/CD18 encoded by the genes ITGAL and ITGB2 respectively. LFA-1 is expressed on immune cells and plays a key role in immune cell migration and activation by binding to intercellular adhesion molecule 1 (ICAM-1) located on endothelial cells during diapedesis - the process where the immune cell travels through the blood vessel walls. This LFA-1 and ICAM-1 interaction also occurs in APC-T cell communication during priming. During this interaction, intracellular C3 stores increase in anticipation of antigen encounter, this process is denominated as ‘C3 licensing’ which remains stable for about 12 hours. In the events of TCR or TLR activation, intracellular C3 is processed into C3a and C3b initiating processes regarding elevated nutrient influx and metabolic reprogramming (Walling & Kim, 2018).

Figure 4 – Described mechanisms of the intracellular complement system, the complosome, in $CD4^+$ and $CD8^+$ T cells, acting on Th1 recruitment and changes in T cell metabolism adapted from (Kunz & Kemper, 2021)

Not only T cells present intracellular complement system regarding immune system cells. Other immune cells like B cells, monocytes, macrophages and neutrophils also reveal complement system derived proteins with different mechanisms of action in each type of cell. With metabolism control being the main similarity between the different cell types, some differ from the norm such as B cells which utilize the complosome for gene transcription purposes and to perform epigenetic alterations. This is evident in a study which revealed that internalized C3 and C3a in human B cells translocate to the nucleus and bind to histones. This activity is hypothesized to contribute to DNA packing as a C3 inhibited histone-DNA interaction, with further studies needed to consolidate this assumption(Kremlitzka et al., 2019). Another study demonstrated the importance of extracellular C3 for human B cell memory formation but did not evaluate if this complement protein was ingested to the intracellular space and its subsequent intracellular mechanisms. However, on the surface, B cell anaphaylatoxin receptors are surging as important contributors alongside C3aR to normal germinal center formation. Despite liver-derived C3 being the main source of C3aR activation in this process, a less noticeable but indicative effect of intracellular C3 is observed. Although, the source of this intracellular C3 is still to be known, as it is not clear if the B cell autonomously produces this complement protein or if it is absorbed from the external environment (i.e., in an autocrine or paracrine fashion).

The role of the immune system in the brain, specifically in neurons, has been thoroughly observed, as the complement system contributes to synaptic pruning (Yilmaz et al., 2021). This process is described as the removal of unwanted neural connections during the formation of the neuronal network. However, the source of the complement system in the brain is currently hypothesized to be systemic with the main thesis being the occurrence of a defect in the blood-brain barrier or by the expression of complement genes in other brain cells like astrocytes (Jacob & Alexander, 2014). The existence of intracellular active components of the complement system and their function in the brain is still to be explored with the latest strides in this knowledge being made in the discovery of C1q in the neurons of the rhesus macaque species, still with no functional mechanisms being explored (Datta et al., 2020). Nevertheless, in mice C1q present in cortical neurons is capable of activating mitochondria and the subsequent production of ROS with changes in mitochondrial respiration and augmented

neuronal activity, due to these metabolic alterations. These mechanisms are thought to be activated in response to hypoxic brain injury with the cells prioritizing the anaerobic ATP production pathways in the events of depleted oxygen volumes present in the blood (Ten et al., 2010).

With these and other discoveries relying on the intra- and extracellular complement system in the immune system and other surrounding tissue and systems, the further study of this exchange can lead to unearthing groundbreaking insights into how the complement system influences brain development, neural pruning, synaptic plasticity and the onset or progression of neurodevelopmental disorders, potentially unlocking novel therapeutic avenues for treatment of such conditions.

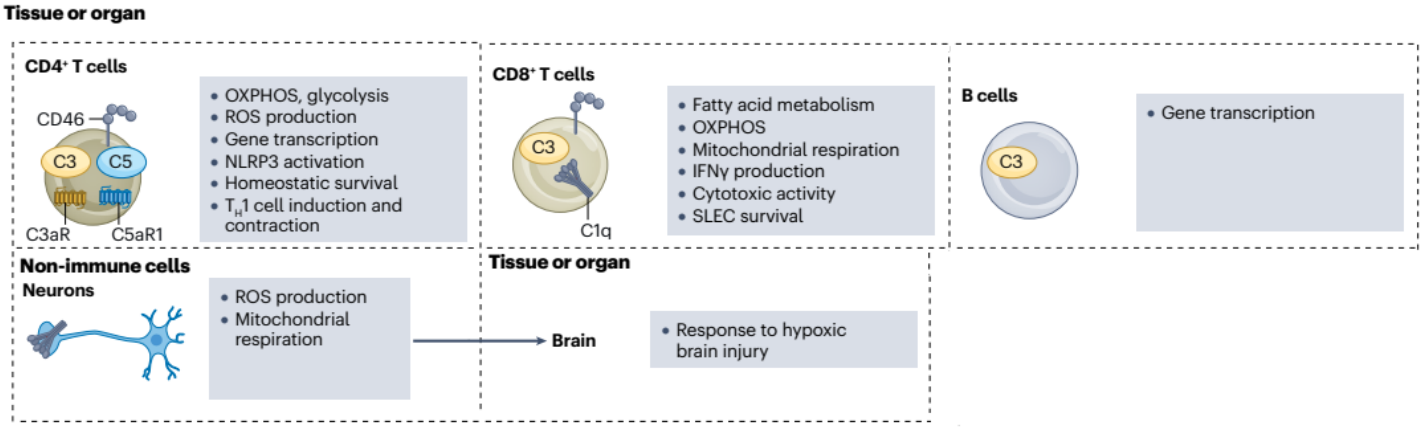


Figure 5 – Key components of the intracellular complement system, the complosome, and their functions across different immune cells types, $CD4^+$, $CD8^+$ T cells and B cells as well as in the neurons and brain tissue adapted from (West & Kemper, 2023)

1.8. Complement System in neurodevelopment

The pre- and post-natal neurodevelopmental stages rely on a series of well-regulated proteins and genes, and any disruption in these pathways may lead to severe consequences. To prevent any complications, the microglial cells work as a protection network. It has also been found that astrocytes, and even the neurons themselves, produce complement system components to serve as an innate immune response system to provide a better defense against pathogens and other harmful signals, detecting aberrant proteins and cellular debris (Haga et

al., 1996). Although, when this system develops faults, it can lead to dysregulation and acceleration of potential harmful pathways leading to neurodegenerative disorders (Chen et al., 2022).

Mainly observed in the pre-natal stage, neurogenesis consists in the differentiation, proliferation of neurons and formation of neural connections. Mouse embryos with deficiencies for complement components C3 and Masp2, or treated with a C3aR antagonist, show signs of an increased concentration of NPC in the ventricular and subventricular zones of the brain, proposing an inhibition of NPC proliferation in the early stages of development. Interestingly, C5aR deficiency during neurogenesis leads to the inhibition of the NPC proliferation in the ventricular zone although, the C5ar knockout mice show increased NPC proliferation in the same zone (Coulthard et al., 2018). Early *in vitro* studies with adult mouse NPC, which are normally situated in the neurogenic niches, show that C3a stimulates neuronal differentiation without any alterations in their survival and proliferation, these same studies also show that the C5a-C5aR1 signaling pathway does not have any role in the adult neurogenic niches (Shinjyo et al., 2009).

Alongside its effects on neurogenesis, the complement system has also been described as a mediator in the neuronal migration during the development of the CNS. More specifically, the knockout or knockdown of the proteins MASP1, MASP2 and C3 result in the irregular cortical thickness and deficient neuronal migration; however, cleaved C3 products seem to reverse this effect (Gorelik, Sapir, Haffner-Krausz, et al., 2017). A different study regarding the SERPING I and the C1INH complement inhibitory proteins, show that these assist in the radial migration and neural proliferation in the pre-natal stage (Gorelik, Sapir, Woodruff, et al., 2017). These proteins are all present in the mannose-binding lectin complement cascade pathway. These findings strongly suggest that the complement proteins may have a spatiotemporal expression pattern during neurogenesis and the neuronal migration steps of the neurodevelopment (Magdalon et al., 2020).

In the MIA mouse model, induced by LPS injection, an increase in C5a levels was detected in the brain and amniotic fluid in premature fetuses. Further inspection of the brain in preterm fetuses, revealed cortical abnormalities, showing increased cell death and low expression of neuronal markers. The neurotoxic properties of C5a were later confirmed *in vitro* by treating fetal cortical neurons with this protein, resulting in inhibited neurite growth and increased cell death (Pedroni et al., 2014). Phenotypically, offspring of malaria-infected mice

during pregnancy, showed difficulties in learning and in memory recall. These characteristics were also accompanied with depressive-like behavior. Disrupting the C5a-C5aR signaling pathway in dams, rescued these neurocognitive effects in the MIA exposed offspring, better suggesting the connection between the complement system overactivation due to inflammation and the neurotoxic and neuropsychiatric effects in the offspring (McDonald et al., 2015). In addition a MIA PolyI:C induced model study, resulted in the observation of complement components C3, C4, and C1q reported to have altered expression persisting into adulthood (Yan et al., 2024).

Despite the wide variety of studies regarding MIA, changes in behavior and in the brain structure, a further, more detailed investigation of the changes in complement gene expression has not been proposed yet. With these studies in mind, this project seeks to observe and understand the expression changes of the complement system in MIA induced rats, utilizing bioinformatics and biostatistics through R. Additionally, we will look for other co-expressed genes which can act together with these complement cascades and finally, try to understand their ontology to better understand the effects of inflammation in pregnancy and its harmful potential on the neurodevelopment of the fetuses.

1.9. Bioinformatics as a tool for biomedical research

Bioinformatics is a vast and fast-growing interdisciplinary field of study that integrates biology, computer science, mathematics and statistics to aid in the analysis and interpretation of intricate biological data. The arrival of this technology has transformed and revolutionized the biomedical sciences field, by providing the necessary tools for researchers to investigate, process and therefore draw conclusions of large pools of genetic, molecular and clinical data. The development of bioinformatics is marked back to the mid-20th century, when the use of computational power was first implemented to study biological sequences. A major turning point in bioinformatics was the creation of the Human Genome Project in the 1990s, which sought after the exploration of the human genome by mass sequencing and annotation of the entire catalogue of human genetic information by recurring to computational methodologies aimed at managing and interpreting genomic data. Due to the project's popularity and the consequent rise in the interest by the scientific community, bioinformatics have since then being constantly evolving, leading to the surge in the development of new computational methods being indispensable in numerous biomedical applications, including genomics,

proteomics, transcriptomics, metabolomics, pharmacology advancements in drug discovery and personalized medicine.

Historically, the origins of bioinformatics can be attributed to early computational biology efforts, such as the cutting-edge work of Margaret Dayhoff, who dedicated their studies in developing the first protein database in the 1960s. However, as sequencing technologies advanced, the necessity for robust computational approaches grew exponentially. The finalization of the Human Genome Project in 2003 provided a reference genome which further added to the scientific knowledge regarding genomics and therefore kickstarted a new batch of bioinformatics advancements. The interplay between biotechnology, biomedical sciences and computational sciences is tightly connected, which leads to a characteristic symbiotic relationship in a degree. When a certain field of these mentioned pillars of scientific investigation achieves ground-breaking discoveries, the resulting added knowledge is then applied to the other facets of science. This leads to the development of algorithms, databases, and high-throughput sequencing technologies which further add to the advancements in biomedical research, transforming what would be several accumulated hours of intensive lab work in the past, to a rapid process. This allows for the efficient progression in assorted areas in the biomedical research sphere, such as identification of disease-associated, biomarker discovery and the development of targeted therapies. In addition, bioinformatics also allows for the exploration of evolutionary biology, comparative genomics, and phylogenetics, helping in the study of animal models, genetic orthologs and variants, species similarity and divergence across various populations.

In contemporary biomedical sciences, bioinformatics plays an essential role in data-driven discoveries. With recent biotechnological advances, next-generation sequencing (NGS) has enabled researchers to decode entire genomes within days, vastly improving the overall understanding across an array of areas such as genetic disorders, cancer mutations and infectious diseases. In clinical research, bioinformatic tools are also utilized, facilitating the identification of specific molecular signatures, aiding in the selection and stratification of patients based on genetic profiles, leading to the development of specific drugs and therapies grounded on the results of the genetic profiling steps, thereby advancing the field of precision medicine and pharmacogenetics. Moreover, structural bioinformatics also adds to this field by aiding in the rational drug design by relying on computer-derived predictions of protein-ligand interactions, expediting a large chunk of the process of drug development. These advancements in computational tools also grant the advantage of simulating complex biological processes,

enabling researchers to test hypotheses *in silico* before deciding on conducting costly and time-intensive laboratory experiments.

The modern applications of bioinformatics have naturally expanded beyond genomics and sequencing analysis, being used for other finalities such as vaccine development. A prime example of this application is the rapid design of the mRNA COVID-19 vaccines, with bioinformatic predictions playing an important role in the development of a viable and safe vaccine. In addition to this feat, bioinformatics also plays a role in analyzing viral mutations, tracking disease outbreaks, and understanding pathogen evolution to help in the surveillance of possible epidemic outbreaks, their pathological mechanisms and therapeutic approaches. In cancer research, bioinformatics aids in the identification of novel oncogenic mutations and predict patient responses to targeted therapies, resulting in several strides for a personalized therapeutic response to cancer. The rise of synthetic biology also depends heavily on bioinformatics tools that design and optimize genetic circuits, enabling advancements in gene therapy, biotechnology and regenerative medicine. Other non-biomedical sciences utilize the benefits of bioinformatics, such as agricultural biotechnology, chemical engineering and other related fields.

Furthermore, bioinformatics is instrumental in systems biology, where it helps in the modelling of complex biological networks and pathways by predicting interaction processes, this provides useful insights into disease mechanisms and possible therapeutic targets. Recently, the integration of artificial intelligence (AI) and machine-learning (ML) within the realms of bioinformatics has been further explored in biomedical research. AI and ML are based in computer-generated predictions and constant improvement of algorithms in order to retrieve better and more accurate results. These algorithms are used to establish high fidelity predictive models and automated data analysis, with these being further enhanced. AI-driven bioinformatics is still being explored as AI remains a novelty currently, however its power and capabilities are recognized in the scientific community. A notable example of AI-driven advancements in bioinformatics is AlphaFold, an AI system developed by DeepMind, which has made significant and groundbreaking contributions to protein structure prediction. To better explain the significance of AlphaFold's advancements into this field, about 17% of the 20000 human proteins had an experimentally determined structure before the implementation of this tool. Since the development of AlphaFold, 98% of the human proteome has a predicted structure with 50% of these being labeled with high accuracy, being astronomically ahead of traditional protein structure research methods. This amazing stride into protein knowledge

unlocks several possibilities within not only biomedical research but also in science as a whole, unveiling nature's one of many mysteries in protein structure and folding with AlphaFold's achievements being recently recognized with the Nobel Prize due to its impact in structural biology.

Among the previously established subfields of bioinformatics, transcriptomics focuses on the study of RNA transcripts, providing insights into gene expression patterns under different conditions and opening up the possibility for downstream analysis. Advances in RNA sequencing (RNA-seq) and microarray data, enable researchers to analyze differential gene expression across different conditions. This analysis can lead to the identification of affected genes within the presented conditions, their interactions and mechanisms in healthy and/or disease states or expression patterns depending on the conditions. Later downstream analysis helps to understand which genes are co-expressed and therefore, which proteins are encoded. Proteomics resembles large-scale studies on the effects of said proteins, evaluating characteristics such as structure, functions, domains and interactions. Computational proteomics can identify potential biomarkers and therapeutic targets, making it fundamental for precision medicine and drug discovery. Additionally, metabolomics, the study of small molecules that act in the chemical reactions of the metabolism or that result from it, is an emerging subfield of bioinformatics, having a role in understanding the interactions between genes and the environment, such as lifestyle and diet. It can also unveil links between cellular pathways and large-scale biological mechanisms. Another important branch in the -omics family is epigenomics, the study of environment-driven gene regulation in forms of activation and repression at the chromatin standpoint. Epigenetic changes range from DNA methylation, to histone modification and when analyzed through bioinformatics, lead to discoveries regarding gene expression patterns and disease mechanisms with further integrations with other -omics approaches, leading to unmasking of how environmental factors weigh in gene regulation, aging and cancer. The integration of multi-omics data through bioinformatic platforms contributes to the discovery of exceptional insights into molecular mechanisms, paving the way for novel approaches in scientific investigation.

With the continuous advancements and evolution of computational power and algorithmic complexity, bioinformatics is set to constitute an even more prominent role in biomedical sciences, providing to be a key player in addressing critical challenges in diagnostics, treatment, and disease prevention as its capabilities help in uncovering advantageous insights into biological processes, physiopathology and molecular mechanisms. The increasing

availability of big data in the form of databases, coupled with improved analytical techniques from a biotechnology standpoint, is set to lead to more significant breakthroughs in the several -omics areas, acting interdependently allowing for the growth of systems biology and integrative biomedical research. As more biological data is generated, the role of bioinformatics is becoming exponentially important in biomedical research as it grants several tools for managing, analyzing and deriving meaningful conclusions from vast datasets.

This project explores the applications and advancements of bioinformatics in biomedical research, by utilizing proteomics tools such as reading counts form a large dataset of microarray data, normalization of the counts, exploring annotated experimental data, applying statistical methods and applying conclusions. By examining real experimental data using open-source programming language and databases, this project aims to demonstrate the indispensable role of bioinformatics in modern biomedical sciences and its potential for future breakthroughs.

2. Materials and Methods

2.1. Data and study design

To achieve a better understanding of the effects of MIA, a pre-existing Gene Expression Omnibus (GEO) dataset was used. The dataset chosen was GSE34058 (Oskvig et al., 2012) which consists of nineteen fetal brain samples of Rattus norvegicus, or common rat. These samples are divided into two groups, ten which mimic a normal control condition and nine from a maternal immune activation model. The MIA state in these samples was induced by LPS injection administrated intraperitoneally on embryonic day 15. These samples were then studied through whole genome microarray analysis, 4h after injection. This is done by extracting and purifying RNA from the cells, running it through reverse transcriptase (RT) cycles to create cDNA and coupling a fluorescent fluorophore dye to the cDNA. This is then placed in an Affymetrix microarray chip containing probes for the whole rat genome. Upon hybridization, the fluorophore in the cDNA samples is activated and emits a different fluorescence intensity depending on the success of the hybridization, e.g. weaker levels for cDNA mismatching the probes. The data preparation, analysis and further steps were all done

using the R programming language, using R Studio as an integrated development environment and other tools such as R packages.

The genes of interest chosen for this study were extracted using the HGNC Comparison of Orthology Predictions (HCOP) search tool to determine the complement system ortholog genes for Rattus norvegicus by feeding a list of the known human complement genes. With these genes defined, a more specific study can be achieved where comparisons can be made between the MIA state and the expression differences of the complement system. With this in mind, a differential gene expression analysis study was conducted using microarray data.

To evaluate the likelihood of complement gene co-expression, a Pearson Correlation test was performed utilizing the statistically significant expressed genes across the whole data set and the statistically significant expressed complement genes. This allows the creation of a list of differentially expressed genes that present a high correlation coefficient with differentially expressed complement genes. Such results open the possibility of studying possible gene relationships and/or pathways that, while possibly not being directly involved in the complement pathways, may interfere with neurodevelopment. To better understand these interrelations, Gene Ontology (GO) enrichment is conducted which computes a list of functions, locations and biological processes which the selected genes may likely act or be involved. Alongside GO analysis, a wide database search is implemented utilizing open source databases available online such as *UniProt*, *NCBI*, *GeneCard*, *STRING* and others (Figure 6).

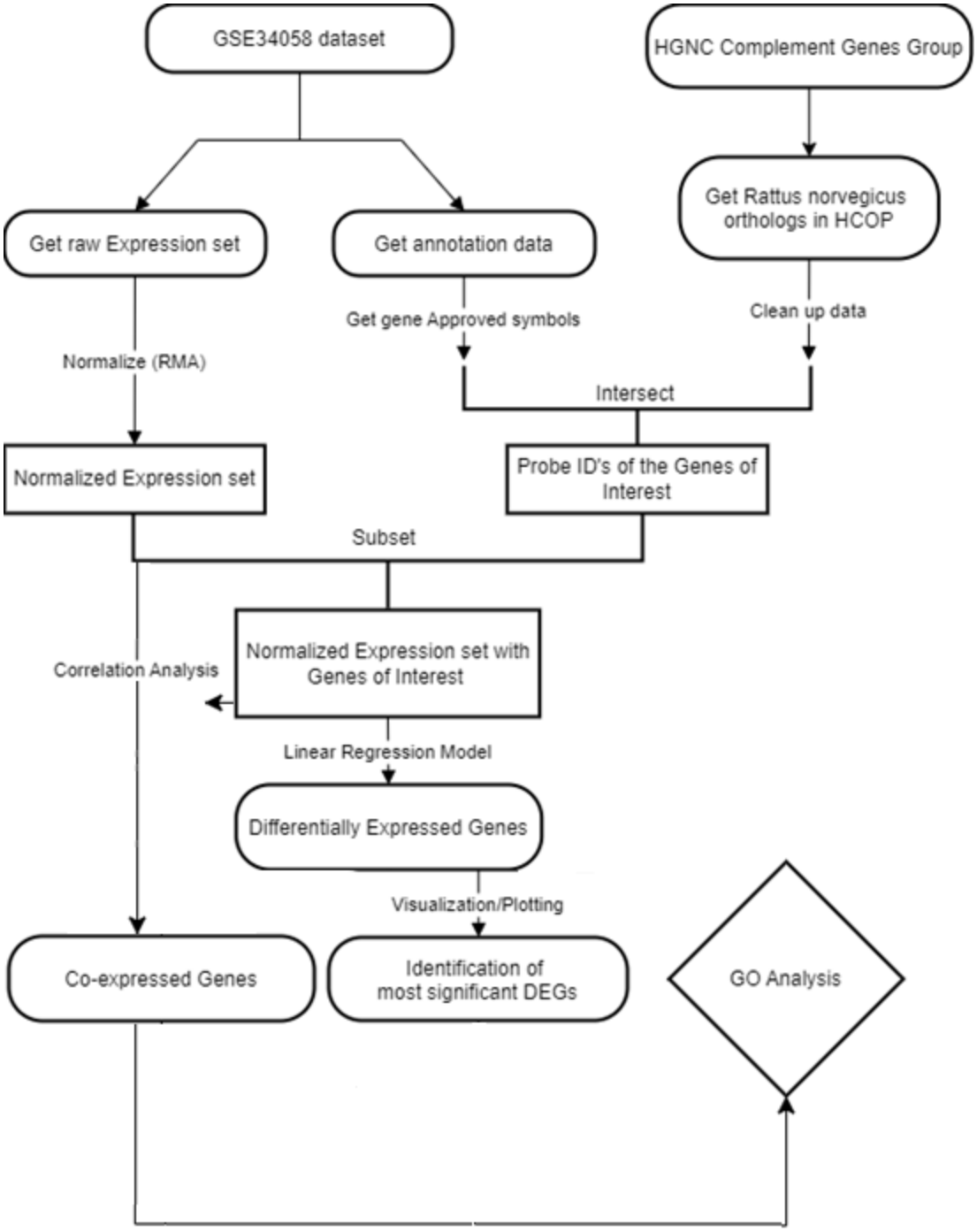


Figure 6- Study design flowchart - created in https://www.smartdraw.com/flowchart/

2.2. Data preparation

Firstly, using a list of genes extracted from the HUGO Gene Nomenclature Committee at the University of Cambridge (HGNC) website (https://www.genenames.org/) of the fifty-six human complement genes (HGNC Group ID 492), respective orthologs for the common rat

were searched utilizing the HGNC Comparison of Ortholog Predictions (HCOP) search tool (Yates et al., 2021), returning a data frame consisting of a hundred and one genes. The *tidyverse* package (Wickham et al., 2019) was used to discard repeats and genes without a defined NCBI identification number returning a number of 95 genes. Additionally, some columns were removed regarding human gene IDs, names of databases and other information not needed for the study.

To prepare the GSE34058 data from the GEO database (Barrett et al., 2013), Bioconductor was installed. The Bioconductor project contains a wide variety of packages specifically designed for bioinformatics and biomedical sciences studies. Several packages utilized in this study are supported by this project. The packages utilized in this step are *affy* (Gautier et al., 2004), *oligo* (Carvalho & Irizarry, 2010), *Biobase* (Huber et al., 2015), *GEOquery* (Davis & Meltzer, 2007), *splitstackshape* (Mahto, 2022) and *limma* (Ritchie et al., 2015). Data was downloaded into the predetermined working directory using the *GEOquery* package. The downloaded *.tar* file was unzipped using the *untar()* function. Subsequently, upon extraction of the files for each sample, they must be converted from the *.gz* format to *.cel* files so these can be compatible with the *affy* package, which reads the raw data of each *.cel* file on the directory, creating an expression set.

Expression data is summarized in an expression set which consists of three meta data sets: the expression matrix, the phenotypic data and the feature data. The expression matrix consists of rows of expression values and columns of samples. Phenotypic data refers to the phenotype of the samples, such as treatment condition, time, or in other words, covariates. Because the data provided on GEO does not contain the phenotype data, the grouping into "Control" and "MIA" was added. The addition of the *gene_assignment_003* column located in the annotation file was performed to the feature data of the expression set, as it is the column where the approved symbols for each gene are situated. Using the *oligo* package, the raw data was normalized with a standard Robust Multi-Array Average (RMA) normalization was used, computing an expression set.

To prepare the data annotation, the GSE34058 family *.soft* file was utilized. This file returns important information regarding gene assignment, number of probes and other relevant details. In this file, the gene assignment column is separated by "//", this leads to a problem when importing to R, which uses another default delimiter, this results in a single column overloaded with multiple pieces of crucial information later needed, such as the approved gene

symbol for each probe. This issue can be addressed using the *cSplit()* function in the *splitstackshape* package. Following this step, another data curation process was applied on the annotated data consisting of removing unimportant columns in the data frame. Next, the expression set was then merged with the annotated data by attributing "ID" in the *rownames* of the expression set (which contains probe IDs) and substituting it by joining the "ID" column from the annotated data using the *left_join()* function in the *dplyr* package into a combined data frame. Finally, the probe IDs were cross referenced from the annotated data to extract the respective approved symbol which is located in the *gene_assignment_003* column. This process results in an extensive but comprehensive data frame that displays not only expression levels but also gene approved symbols for further consultation.

2.3. Quality control

Quality control of the expression set was performed using the *arrayQualityMetrics* package. This package is designed specifically for microarray data, being able to read the *ExpressionSet* class and normalized data leading to a better understanding of the utility of the data when applying downstream analysis, such as differential gene expression analysis. The package initially evaluates the individual array quality by creating a MA-plot, a graph that is useful to examine background intensity and detect systematic biases. Another process to assess the data is the homogeneity between arrays which are viewed using a $\log_2$ intensities boxplot graph and a density estimate curve graph. Further quality control steps are involved in the *arrayQualityMetrics* package, such as between array comparisons and *Affymetrix* specific plots (Kauffmann et al., 2009).

To assess the similarity of the samples in question, a Principal Component Analysis (PCA) study was conducted taking the whole expression data into account. This study was performed utilizing the *prcomp()* function in the *stats* package comprised in base R. This performs a data dimensionality reduction while keeping the variability information of the data set, so it can be plotted onto a two-dimensional graph. Column sample names were added to the expression data set to then visualize in the PCA plot. Control samples were colored in blue and MIA samples in red. The results were then plotted into a scatterplot using the *plot()* function from the package *base*, while annotation was added with the *text()* function from the *graphics* package both in base R.

To find the similar probes both in the expression set and the gene annotation file, the intersection of the probe IDs from the two data sets was done, resulting in the common observations of both. With this information, subsetting the two data sets into a shorter and more concise pool of data is possible. After this step, the rat genes of interest are intersected with the gene annotation column regarding the approved symbols of the genes to further subset the data sets to only include the rat complement system genes.

2.4. Differential gene expression analysis

In order to evaluate differential gene expression values, the microarray data was analyzed through the fitting of the expression matrix *exprs* in a design matrix. The design matrix has rows consisting of samples and columns associated with model parameters, such as the expression values and the experimental conditions. For identification of the experimental conditions, these were added as binary factors with value of 0 and 1 referring to, in this project, Control and MIA. This information was extracted from the phenotypic data or *pData* entity present in the expression set. Following this procedure, the expression set can be analyzed and fitted through a statistical test of choice. Due to the categorical nature of the predictor or independent variable, Control or MIA, alongside the continuous nature of the expression data, a linear regression model was chosen as the suitable statistical test for analysis. However, when analyzing a large number of genes, measures must be taken into account for the stabilization of the data in order to derive accurate conclusions regarding the differentially expressed genes.

This was achieved by utilizing the *limma* package, as it was developed with the sole intent of analyzing microarray data through linear regression model and has become the standard practice for differential gene expression analysis. The linear model statistical test was performed for the whole expression data set, with the objective of further studying the differential gene expression of the complement genes. Data moderation was implemented following the Empirical Bayes method, implemented in the *limma* package through the *eBayes()* function, to improve variance estimation and stabilize the inference of differential expression. This ensures that the variance remains stabilized, as genes with higher expression values often present higher variance than genes with lower expressions. This approach relies on the borrowing of strength from across the different genes, leading to more stable estimates through the shrinkage of variance towards the overall mean. This computes a data frame containing information such as probe IDs, approved gene symbols, logarithmic fold change,

average expression, p-value and adjusted p-value for all the genes present in the data set. Afterwards, the results were subset to only include the complement genes which are essential for downstream analysis. By subsetting the differential expression values from the complement genes after the statistical test is performed in the whole data set, heteroscedasticity in the expression data is ensured, while technical noise is mitigated leading to a more amenable data for statistical analysis, which improves the detection of genuine biological signals and the foundation of accurate conclusions. Additionally, to better evaluate up or downregulation and better graphical analysis in the visualization step of the project, the resulting logarithmic base 10 fold change was calculated to encompass a logarithm of base 2, which is the standard practice for gene expression analysis data.

2.5. Graphs and visualization

Differential gene expression results were visualized with the *ggplot2* (Wickham et al., 2024) and *pheatmap* (Kolde, 2019) packages for the volcano plot and heatmap respectively. A volcano plot was generated to represent the relationship between the log2 fold change and statistical significance of gene expression. The log2 fold change values and p-values were extracted from the results table, and the adjusted p-values were used to highlight significantly differentially expressed genes (adjusted p-value < 0.05). The volcano plot was created using *ggplot2*, with blue color indicating significantly down-regulated genes and red indicating significantly up-regulated genes. Additionally, for better readability, gene approved symbols were displayed when having a less then 0.05 adjusted p-value alongside the *geom_text_repel()* function to ensure no text overlap.

For the heatmap, expression data for differentially expressed genes were extracted from a complement system expression set object. The heatmap was generated using the *pheatmap* package, with rows scaled to z-scores. Again, expression levels are represented in a blue-to-red gradient with the lowest z-scores being colored in dark blue and the highest being colored dark red. Additionally, all gene approved symbols were added in the rows. Both rows (genes) and columns (samples) were hierarchically clustered, and sample annotations were derived from the associated phenotype data (pData).

2.6. Correlation analysis

To investigate the correlation of gene expression levels of each gene across all the samples, Pearson Correlation analysis was performed using the *cor()* function present in the *stats* package in base R, by focusing in two gene sets: the significant genes in the whole GSE dataset and the significant genes of interest, in this case rat complement genes extracted from HCOP database. This allows for an investigation regarding gene expression patterns of the significant genes and their similarity, possibly tracing and resulting in gene clusters with similar functions or form similar families that may provide some important information regarding co-expression associated with complement system in the MIA state. This was done by subsetting the genes with an adjusted p-value of less than 0.05 in both data sets. To allow a better understanding of the which of these genes are co-expressed and how these relate with each other, hierarchical clustering was conducted using the *hclust()* function also from the *stats* package. The results of this cluster were then plotted into a dendrogram to ensure a better visualization of the similarities between the genes, with the y-axis labelled "Height" representing the dissimilarity between genes on the x-axis.

2.7. Pathway expression analysis

To observe the differential expression on the context of the complement pathway, the *Pathvisio* (Kutmon et al., 2015) program was used, loading the rat complement pathway (named WP547.gpml) from the website *Wikipathways* (Agrawal et al., 2024). The remaining complement genes extracted from HCOP that were not included in the preset pathway were manually added. To understand the function of these genes on the pathway, bioinformatic research was conducted mainly using the *Uniprot* (Bateman et al., 2023) database which usually contain gene ontology annotations upon every entry.

Upon adding all the genes to the pathway, differential gene expression data was loaded onto the program using a .csv file extracted from the differential gene expression data frame. To identify the gene products, *Pathvisio* utilizes Entrez IDs which were added to the data frame using the *AnnotationDbi* (Pagès et al., 2024) package alongside the *org.Rn.eg.db (Carlson, 2019)* package which contains the genes for the rat organism. Another column was added regarding the system code utilized by the *Pathvisio* program to read *Entrez* IDs. This consists of a column with the character "L" in each row. With the expression data loaded into the program, genes are colored depending on the expression which was changed to a blue to red

gradient with limits of -2 and 2, since the data utilized to visualize the expression is the log2 fold change column of the data frame.

2.8. Gene ontology analysis

After nominating the most significantly expressed gene in the data (Cr1l), a gene ontology enrichment analysis study was conducted utilizing the most correlated genes, with a greater than 0.8 correlation score from the whole array which resulted in a data set of 2561 genes. After subsetting the correlation matrix to only include these genes, the respective probe IDs were substituted by their gene symbols. By exporting this list to a .txt file, a list of gene symbols is acquired which was then pasted into the *G.profiler* website (https://biit.cs.ut.ee/gprofiler/gost) for Gene Ontology (GO) analysis. This website then, computes the given query outputting gene ontology results by plotting a Manhattan plot with Metabolic function, Biological processes, Cellular Components, Kegg and other results on the x-axis and the negative logarithm of the adjusted p-value which the website also calculates. Several bar plots are also computed on the "Detailed Results" tab, being separated by each type of GO results. Being able to evaluate and better understand which processes, cellular components and metabolic function are enriched or overrepresented in the data set.

2.9. Open-source Database research

Utilizing the results from the differential gene expression analysis, significant up and down regulated genes were pooled in a wide, open-source database research. This was done with the intention of finding interactions and/or paper supporting the hypothesis of the complement gene's role in MIA and its effects in neurodevelopment and neurodevelopmental disorders. Open-source databases are often updated in accordance with published experimental findings, as anyone can submit their evidence for review and then implementation. With this in mind, researching several open-source databases will reveal new advancements in the field of MIA and neurodevelopment. Although to ensure the validity of the findings, the research was expanded to include several different databases for cross-referencing and therefore, mitigate the existence of false discoveries. Additionally, peer-reviewed literature is considered as a form of curation to ensure data accuracy.

Databases can be classified into different types depending on their information. Gene databases contain information regarding the gene's expression, associated proteins, functions, orthologs and other useful data. The utilized gene databases were as follows: GeneCards which contains comprehensive gene information, expression, associated pathways and diseases, genetic variants and orthologs, alongside supporting research publications; NCBI GenBank which contains sequence storage and cross-references the information through other NCBI databases such as PubMed for supporting literature, RefSeq for reference genomic annotations, GEO for gene expression and others; Ensembl Genome Browser which provides annotated genome data for various species, providing information regarding gene structure, transcripts and phenotypes; NHGRI-EBI Catalog of human genome-wide association studies (GWAS) for studies and publications regarding the enquired gene; Rat Genome Database (RGD) for ortholog gene information including annotation, interactions, gene ontology, alongside a curated references list for literature regarding the enquired gene.

For the associated proteins research in this project, protein databases were utilized such as: Uniprot, which returns comprehensive information regarding protein sequences, structure, domains, protein-protein interactions, cellular locations and proteomics alongside functional annotations, integrating data from several different sources; Human Protein Atlas (HPA) for protein expression and localization across different tissue types, it also refers RNA expression data for different cell types and tissue sourcing through several databases, contributing to biomarker discovery and proteomics; Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) which results in a predicted protein-protein interaction network based on experimental data, computational predictions and literature mining, being useful for identifying novel functional relationships between genes/proteins;

For a better understanding of the gene/protein functions as an ecosystem, signaling and pathway databases were also utilized. With this database being the Reactome database, providing insights into biological pathways, molecular interactions and biochemical reactions, which allows multiple search entries offering a specific interaction result for the studied genes. This database also contributes to the understanding of metabolic pathways and disease mechanisms. The resulting analysis report can then be downloaded, containing useful information such as literature references for the resulting outcomes.

Additionally, disease and clinical databases were also consulted such as the Online Mendelian Inheritance to Man (OMIM) database which contains a catalog of human genes and

genetic disorders alongside its phenotypical effects. This database is useful for establishing relations with queried genes and genetic disorders, finding possible harmful mutations and genotype-phenotype correlations with also having links to related publications. Alongside OMIM, the Malacards database was also utilized to reveal the most associated diseases with the input gene. This database provides a search score for the resulting diseases, which contributes to filter out significant and unsignificant data.

3. Results

3.1. Quality control results reveal two outlier samples

After implementation of the quality control step using *arrayQualityMetrics*, a folder was created with the different quality assessment tests results. These tests include between array comparison using a heatmap and principal component analysis (PCA), array intensity distribution using boxplots and density plots, variance mean dependence and individual array quality using MA plots (see Appendix 1).

Signal intensity distribution of the arrays identified outliers, specifically on samples 8 and 17, which lead to their removal for the differential gene expression analysis. Outlier detection was performed using Kolmogorov-Smirnov statistic between each array's distribution and the distribution of the pooled data. Using the distribution values of the whole array samples, a threshold of 0.0173 was calculated, which lead to the classification of the previously mentioned samples as outliers (Figure 7A & B). Principal Component Analysis of the samples was visualized using a scatter plot, this resulted in two different clusters. These clusters indicated the two different groups (control and MIA), confirming that there is a difference between the conditions while preserving similarity in the samples of each group (Figure 7C).

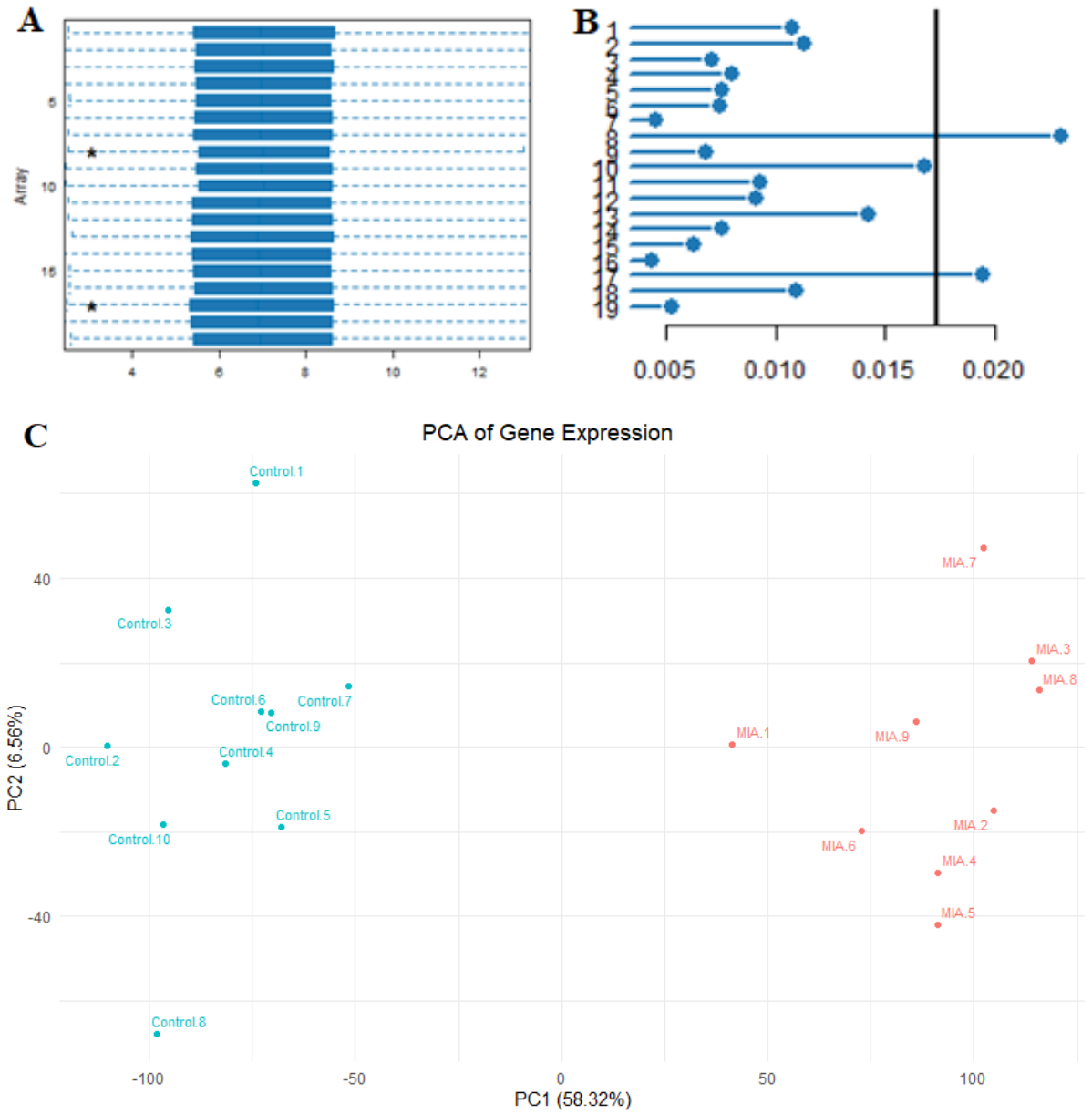


Figure 7 - A) Barplots representing signal intensity distribution of each sample, boxes number 8 and 17 show different widths being represent with an asterisk (*). B) Outlier detection of the samples through a bar chart by computing the Kolmogorov-Smirnov statistic between each array's distribution. The vertical line represents a threshold calculated based on all the distribution values, samples 8 (control) and 17 (MIA) cross this limit being deemed as outliers. C) Principal Component Analysis of the samples including the outliers, represented by a scatter plot projecting the gene expression multivariate data as a two-dimensional plane where samples cluster based on similarity. Samples 1 to 9 (control) cluster together while samples 10 to 18 (MIA) cluster together on the other side of graph. This means that samples from the same conditions are highly similar, however, samples from different conditions show low similarity.

3.2. Differential gene expression results

3.2.1. Differential gene expression visualization using a Volcano plot

Using the rat complement genes list and after fitting the data through a linear model statistical design, a volcano plot was utilized to visualize the significant changes in the differential gene expression of this data set. For a better visualization in the graph, the base 10 logarithmic fold change was changed to a base 2 logarithmic fold change. In this parameter, thresholds in the x-axis were implemented in the -0.6 and 0.6 Log2 Fold Change to define genes as “Downregulated” and “Upregulated” respectively. To further investigate the significance of the gene expression, another threshold was added in the y-axis corresponding to the negative base 10 logarithm of 0.05, the value taken under consideration for the adjusted p-value to study the statistical significance of the data set.

This resulted in five significant downregulated genes, these being Tbxa2r, Gtf3c2, C4bpa, Lcn12 and Masp2. Eleven significant upregulated genes were also identified, these being, Cr1l, C1qc, Cfh, three possible transcripts of Rps7, Cd46, C1qa, C5, C9 and Cfd. Other seven genes were identified as significant according to differential gene expression analysis but were not considered to be upregulated or downregulated due to not crossing the determined log2 fold change cut-offs. These genes were: Fcnb, Clu, C3ar1, C7, G3bp2, Ldb2 and C1qb (Figure 8).

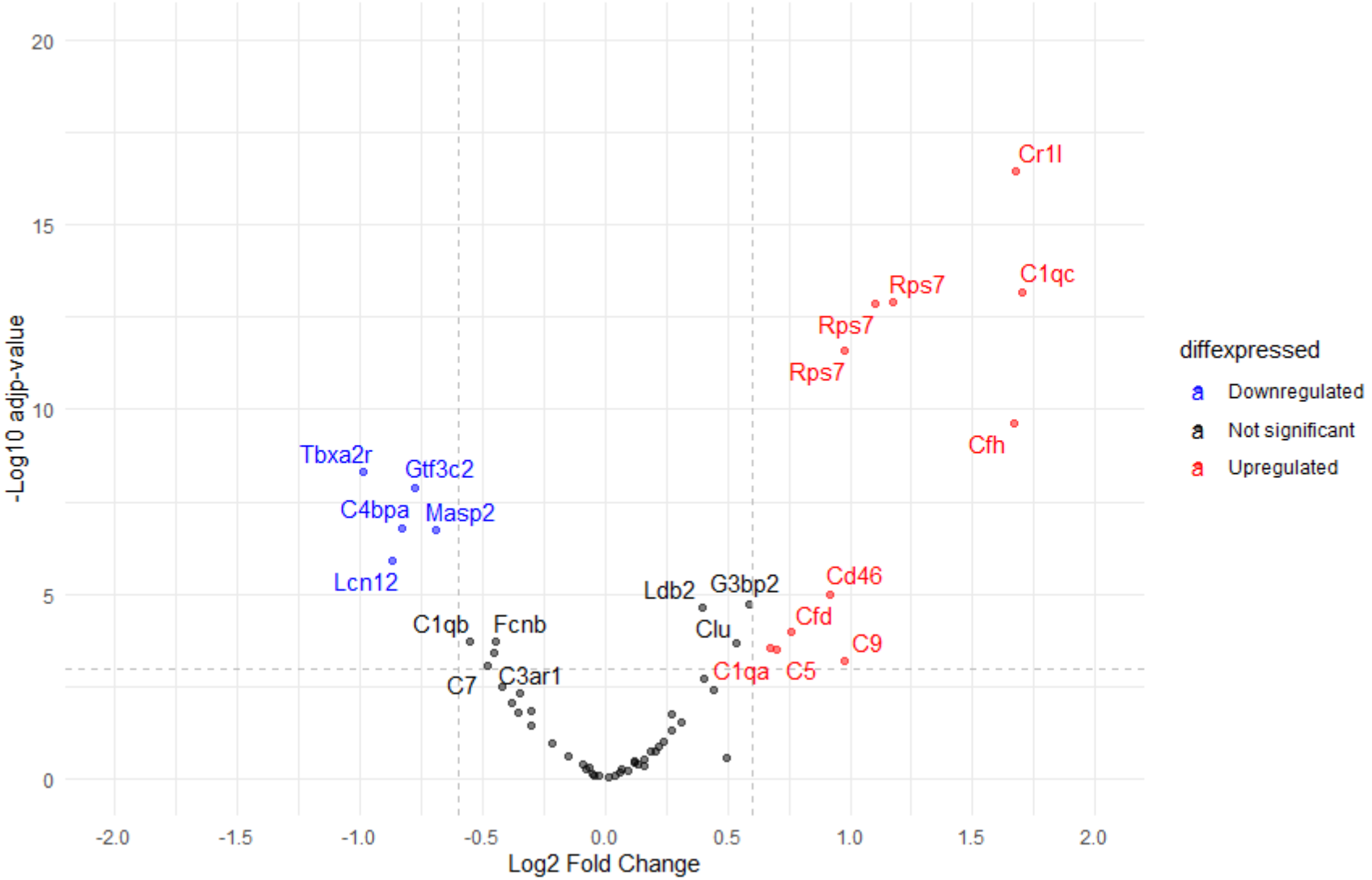


Figure 8 - Volcano plot of complement gene expression data. Gene approved symbols represent significantly expressed genes (p-value > 0.05), blue colored gene approved symbols represent statistically significant differentially downregulated genes (log_2 Fold Change < -0.5), while red colored gene approved symbols represent statistically significant differentially upregulated genes (log_2 Fold Change > 0.5). Black colored gene approved symbols represent statistically significant genes with no specific signs of down or upregulation of gene expression.

3.2.2. Differential gene expression visualization using a Heatmap

To understand the relationship between each sample and each gene, while assessing the gene expression under the different conditions, a heatmap was formulated. Hierarchal clustering was performed for the samples and genes where a clear difference is identifiable between the control groups and the MIA groups. This indicates that the complement pathway is differentially regulated in response to maternal LPS treatment, and that the gene expression patterns in this pathway can effectively separate the two groups. (Figure 9).

Clustering of the genes is also identifiable into three distinct groups. The first cluster starting from the gene C5 all the way through gene C9 which are mainly downregulated in MIA (orange box, labeled with the number 1). The second cluster which comprises the genes Cfi to C1r where no expression patterns are observed between the two experimental groups (green box, labeled with the number 2). Lastly, the third cluster where genes Cfh to C8g are notably upregulated in MIA samples (purple box, labeled with the number 3). These patterns strongly indicate that depending on the immune circumstances, groups of genes are differentially expressed.

3.3. Pathway analysis using *Pathvisio*

To better understand and identify biological processes that are behind the expression differences verified on the heatmap, pathway analysis was done with *PathVisio*, where the differential gene expression data was loaded to aid in the visualization and interpretation of the data (Figure 10). This shows the role of the complement system proteins as a concise interconnected system, divided by classical, lectin and alternative pathways converging into one final terminal pathway, and their interaction in a systems biology standpoint, providing key insights into the regulation of biological processes related to the complement system such as cell lysis, phagocytosis, activated T cell proliferation and even its effects in the coagulation pathway. Other proteins that take part in this system but are not identified in the experimental data were added from the complement system template provided by *WikiPathways* to fill in the missing information that is important in the unraveling of the complement pathway.

Results show that genes which take part in the first steps of the activation of the complement system, namely in the classical and lectin pathways, are mainly upregulated, while genes

comprising the activation of the alternative pathway show generally to be downregulated. In the classical pathway, genes that take part in the formation of the C1q complex, these being C1qa, C1qb, C1qc, C1qg, C1r and C1s, are mainly considered upregulated with the exception of C1qb which shows to be slightly downregulated. This complex is formed in the presence of an antibody-antigen complex due to the leukocytes' identification of pathogens. Interestingly, the C1s inhibiting protein, Serping1 is also shown to be slightly upregulated. Other classical complement pathway inhibiting proteins show to be upregulated, these being CD46 and Zp3r except Csmd1 that also has been described as having a role in the prevention of this pathway but shows to be downregulated. In the lectin pathway, key proteins that bind to mannose, kickstarting this side of the complement system, Colec10 and Rps7 showed to be upregulated, however next steps in the continuation of the pathway tended to be downregulated or with no considerable expression changes. These downregulated genes were the Fcnb and Masp2 genes and the non-considerably expressed genes were Mbl2 and Masp1. These two pathways converge in the activation of C2 which wasn't differentially expressed with a log2 fold change close to zero. Looking into the alternative pathway, key initiator gene Cfp showed to be downregulated, consistent with the inhibitor gene Cfh which was highly upregulated. This leads to the non-differentially expressed C3 which curiously, has its cofactors Cfd (which acts in the cleavage of C3) and Adn appear as upregulated.

All these pathways then converge into the C3 convertase enzyme, which originates the C4a, C3a, C3b and C3d mediator proteins which act downstream providing the continuation of the complement pathway, now in the terminal portion of this chain reaction. In this step, an inhibitor protein for C4a, C4bpa was shown to be downregulated, nonetheless several C3 convertase inhibiting proteins returned visible differential expression changes, these being Crry, Daf1 and Mcp which appeared as upregulated. Another protein with inhibition properties for C3 convertase, Cfi showed to be slightly downregulated. Other downstream effects of the C3 convertase is the assembly of the membrane attack complex (MAC) with key proteins interacting with each other to capacitate the formation of this structure. These being C6, C7, C8, its coactivators C8a and C8g, and C9 complement proteins. All these were labeled as downregulated in the differential expression assay, excluding C9 which was observed as upregulated. The MAC then aids in the promotion of the pathogen's cell lysis, although a gene that inhibits this process, Clu was found to be upregulated. Other mediators of the downstream effects of the C3 convertase products were described in this project, with these being Vsig4, considered slightly downregulated, which binds to C3b and other genes being C5 and C5l1

which act on C5a activation, with C5 being slightly upregulated while C5l1 resembles no change in differential expression analysis. More importantly, the downstream Cr1l and Crry genes encoding to complement system receptors from the terminal pathway appear as highly upregulated, leading to responses in muscle contraction, chemotaxis, phagocyte recruitment and more importantly, inflammation. Other membrane receptor genes C3ar1, Cr2, Zfp275 and Cd93 appear as slightly downregulated.

As the complement system also collaborates directly with the coagulation pathway, other genes that were identified in the experimental data were added to the *Pathvisio* pathway scheme. These three identified genes were the F2, Vtn and Tbxa2r genes. While their place in the coagulation pathway does not directly imply in the complement system, these were still taken into observation and addressed. F2 had its differential expression levels resulting in a slight upregulation, while Tbxa2r was considered highly downregulated. These two genes lead to changes on platelet, monocyte, lymphocyte, endothelial cells and smooth muscle cells activation. Vtn on the other hand, does not seem to act in any of the mentioned processes although it has its activation due to the Serpine1 gene which is a constituent in the coagulation pathway. Its differential gene expression results show Vtn to be considered upregulated. Lastly, Cd55 is described as having inhibitor properties for the general complement system and in this project, was shown to be slightly upregulated.

Other membrane genes which were identified in project were the Itg family of genes, mainly Itgam, Itgax, Itgad and Itgb2 genes. These are described as having key roles in the activation of T cell proliferation and therefore, their role in the immune system in events of inflammation. While Itgam and Itgax resembled low upregulation levels regarding their differential expression, the Itgad and Itgb2 genes showed close to no visible change. These results coupled with a graphical representation of the complement cascade allow for the overview of the changes in complement system related genes expression in the context of the biological processes, flow and relationships, facilitating further research into these alterations.

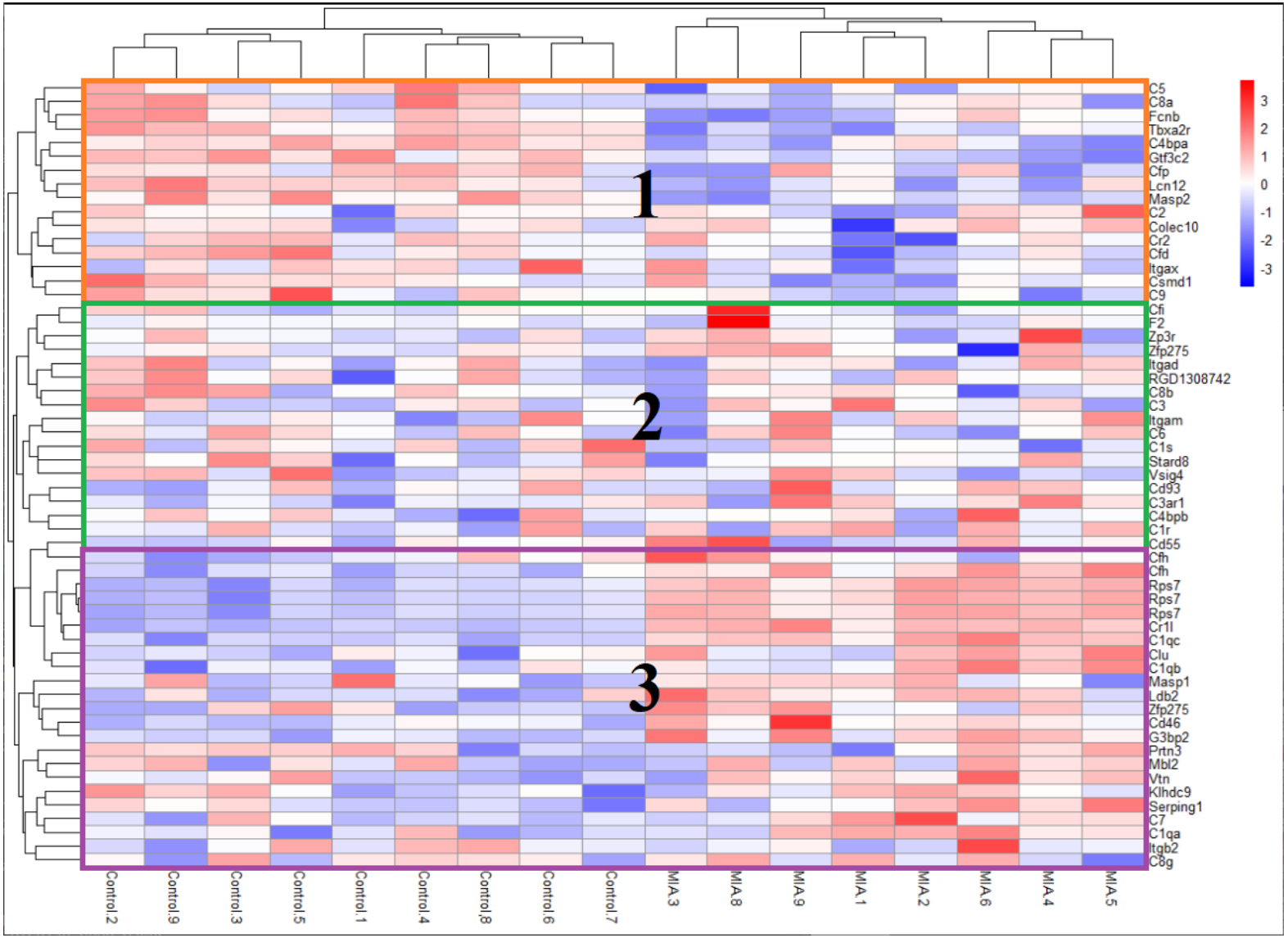


Figure 9 - Heatmap of differential complement gene expression data. The data is plotted into a grid with rows as genes and columns as samples, each observation is then colored in a blue to red gradient which indicates downregulation and upregulation respectively. On the top of the graph a dendrogram represents the hierarchical clustering performed, this further demonstrates that samples within the same condition appear similar while samples from two different conditions appear as non-similar.

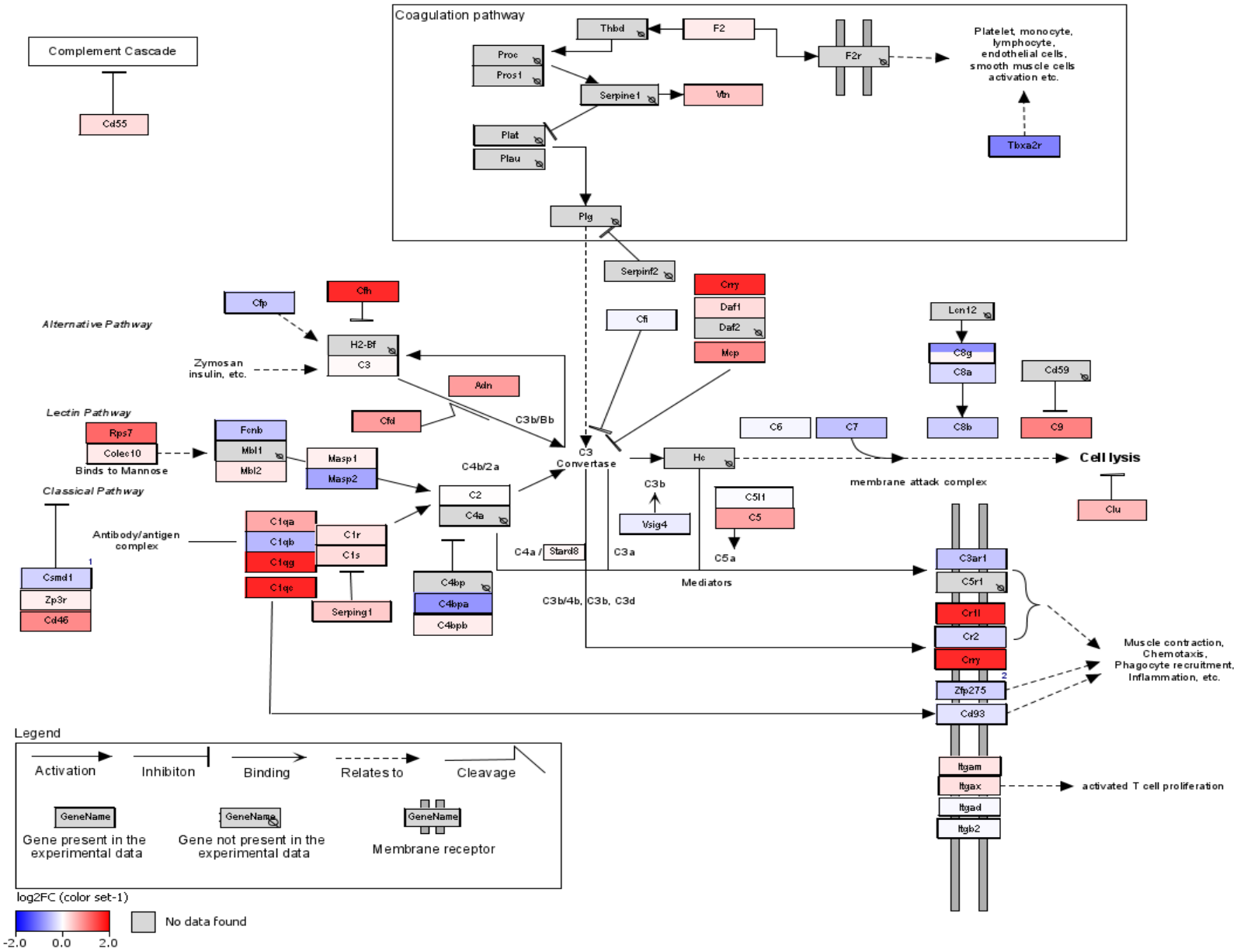


Figure 10 - Complement cascade pathway and its regulators. Genes were colored according to their differential gene expression results from prior analysis (see legend). Genes that take part in the activation of the classical complement pathway showed to be highly upregulated.

3.4. Co-expression analysis through correlation

After implementing the significant complement genes (with a pvalue < 0.05) through a Pearson Correlation pipeline based on gene expression data, results were plotted onto a dendrogram through hierarchical clustering to evaluate gene co-expression. This resulted in four clusters with height between 0 and 0.5, these being: Cluster 1 comprising of the genes Gtf3c2, C1qb, C3ar1 and Serping1; Cluster 2 including the genes Vtn, C5 and C7; Cluster 3 constituted by the genes Cr1l, C1qc, Rps7, Ldb2, Tbxa2r and G3bp2; With cluster 4 consisting of genes Cfd, Clu, C4bpa, Cd46, Fcnb and C1qa; The remaining genes were graphed as having a correlation height of about 1.5, suggesting a dissimilarity form most other genes with cluster 5 encompassing the gene Masp2; Finishing the dataset with Cluster 6 holding genes Lcn12, C8b and C9 (Figure 11).

Clusters 1 through 4 indicate a strong similarity in expression profiles, due to being categorized in lower heights, implying stronger co-expression or functional similarity between these genes. However, clusters 5 and 6 show weaker co-expression and/or more differences in expression patterns when related to the other remaining genes in the dataset. Nevertheless, when assessing the individual relationship of the constituents of these groups, genes within cluster 6 resemble high similarity within each other, suggesting functional proximity or adjacency within signaling pathways and co-expression. Although, Masp2 sits isolated in cluster 5 with no close relatives, suggesting a more different or independent role than all the other referenced genes in this correlation test.

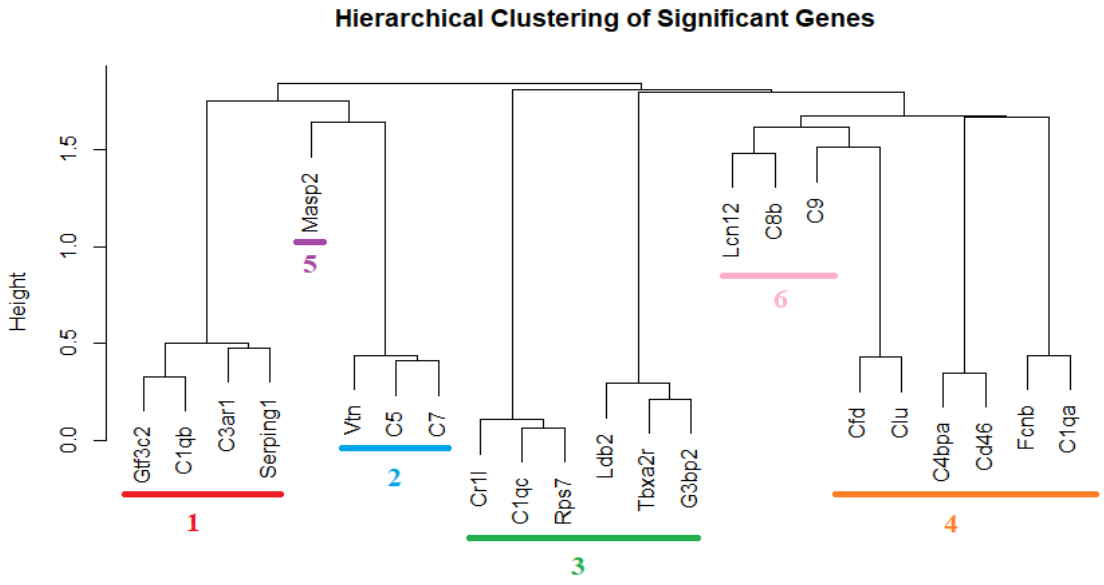


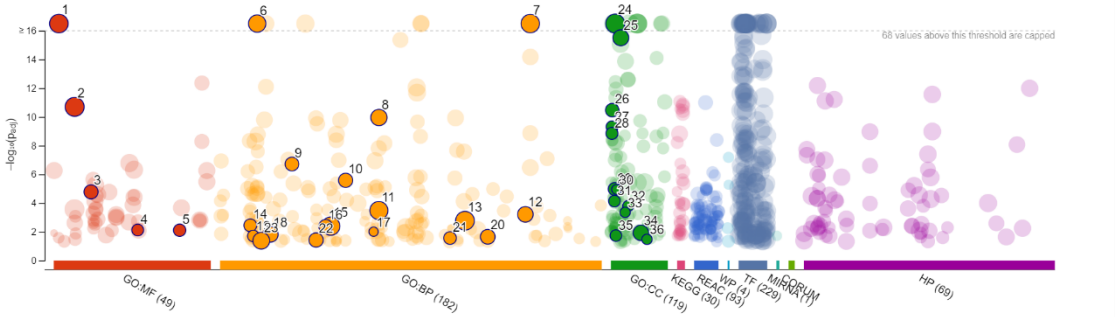
Figure 11 - Pearson correlation of significant complement genes plotted onto a dendrogram through hierarchical clustering. Six clusters are identified, cluster 1 in red (Gtf3c2, C1qb, C3ar1, Serping1), cluster 2 in blue (Vtn, C5, C7), cluster 3 in green (Cr1l, C1qc, Rps7, Ldb2, Tbxa2r, G3bp2), cluster 4 in orange (Cfd, Clu, C4bpa, Cd46, Fcnb, C1qa), cluster 5 in purple (Masp2) and cluster 6 in pink (Lcn12, C8b, C9). Clusters 1 through 4 resemble the higher gene similarity of all genes. While clusters 5 and 6 demonstrate lower gene similarity.

3.5. Gene ontology enrichment analysis

C1rl was the most significant differentially expressed gene identified. To further explore its biological role in the context of MIA, gene ontology enrichment analysis was performed using co-expressed genes. Expression of 2561 genes were identified as being significantly correlated with the expression of Cr1l and thus used for gene ontology enrichment analysis. This led to the identification of the following significant characters: 5 metabolic functions, 18 main biological processes and 13 main cellular components (Figure 6). The results show that the genes expressed alongside the complement system gene C1rl, are primarily associated with

organelle organization, organonitrogen compound metabolic process and small molecule metabolic process.

Figure 12 - Gene ontology enrichment analysis of C1rl correlated genes. Genes with a correlation score > 0.8 to C1rl, due to being the most significant complement gene, were chosen for this analysis. This was performed with the objective of studying the metabolic functions, biological processes and cellular components of these co-expressed genes alongside the complement system.



ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:MF	GO:0003824	catalytic activity	1.306×10^{-32}
2	GO:MF	GO:0005515	protein binding	1.994×10^{-11}
3	GO:MF	GO:0015631	tubulin binding	1.665×10^{-5}
4	GO:MF	GO:0044183	protein folding chaperone	7.558×10^{-3}
5	GO:MF	GO:0090079	translation regulator activity, nucleic acid binding	7.838×10^{-3}
6	GO:BP	GO:0006996	organelle organization	3.360×10^{-23}
7	GO:BP	GO:1901564	organonitrogen compound metabolic process	6.567×10^{-23}
8	GO:BP	GO:0044281	small molecule metabolic process	1.102×10^{-10}
9	GO:BP	GO:0015980	energy derivation by oxidation of organic compou...	1.926×10^{-7}
10	GO:BP	GO:0034470	ncRNA processing	2.617×10^{-6}
11	GO:BP	GO:0044271	cellular nitrogen compound biosynthetic process	3.312×10^{-4}
12	GO:BP	GO:1901135	carbohydrate derivative metabolic process	6.153×10^{-4}
13	GO:BP	GO:0071840	cellular component organization or biogenesis	1.707×10^{-3}
14	GO:BP	GO:0006352	DNA-templated transcription initiation	3.576×10^{-3}
15	GO:BP	GO:0032502	developmental process	4.067×10^{-3}
16	GO:BP	GO:0031647	regulation of protein stability	4.181×10^{-3}
17	GO:BP	GO:0043248	proteasome assembly	9.909×10^{-3}
18	GO:BP	GO:0009628	response to abiotic stimulus	1.547×10^{-2}
19	GO:BP	GO:0006626	protein targeting to mitochondrion	1.971×10^{-2}
20	GO:BP	GO:0097190	apoptotic signaling pathway	2.245×10^{-2}
21	GO:BP	GO:0065004	protein-DNA complex assembly	2.737×10^{-2}
22	GO:BP	GO:0022411	cellular component disassembly	3.769×10^{-2}
23	GO:BP	GO:0007399	nervous system development	4.226×10^{-2}
24	GO:CC	GO:0005737	cytoplasm	1.232×10^{-75}
25	GO:CC	GO:0015630	microtubule cytoskeleton	3.085×10^{-16}
26	GO:CC	GO:0000793	condensed chromosome	3.392×10^{-11}
27	GO:CC	GO:0000502	proteasome complex	4.889×10^{-10}
28	GO:CC	GO:0000428	DNA-directed RNA polymerase complex	1.400×10^{-9}
29	GO:CC	GO:0005681	spliceosomal complex	1.054×10^{-5}
30	GO:CC	GO:0005868	cytoplasmic dynein complex	1.124×10^{-5}
31	GO:CC	GO:0005657	replication fork	7.363×10^{-5}
32	GO:CC	GO:0032040	small-subunit processome	1.600×10^{-4}
33	GO:CC	GO:0030990	intraciliary transport particle	4.441×10^{-4}
34	GO:CC	GO:0045202	synapse	1.186×10^{-2}
35	GO:CC	GO:0005795	Golgi stack	1.856×10^{-2}
36	GO:CC	GO:0070847	core mediator complex	3.378×10^{-2}

version	e111_eg58_p18_f463989d
date	30/09/2024, 10:54:07
organism	rnorvegicus

g:Profiler

3.6. Open-source Database research results

Cr1l or Complement Component (3b/4b) Receptor 1-like protein was queried on several open-source databases revealing its role within regulation of complement activation and regulation of complement-dependent cytotoxicity. It acts upstream of the complement regulation pathway being part of a Complement 3b/4b receptor complex. For a better understanding of Cr1l's interactions and relationships with other genes and proteins, Reactome analysis didn't return any results as Cr1l is not identified in its database, having to rely on STRING to explore its connections. This search returned a network of several genes with Clu, C4b, C4a, ACBA7 and FCER2 connecting directly to Cr1l. Other genes included in the network but not connected to Cr1l consist of BIN1, EXOC3L2, PICALM, C3 and Cfh (Figure 13). The BioGRID database was also utilized to study the Cr1l interactors, resulting in a list of 12 genes resulting from high throughput studies, where more than 100 interactions are studied. The identified Cr1l interactor genes were: DDHD1, RMND1, CR1, SPCS2, SEC11A, SERPINB8, NFYC, ACTL6B, ACTA2, GTF2IRD1, ECM1 and HAL. The evidence for these interactors is based on Two-hybrid experiments (Luck et al., 2020), in the case of DDHD1 and on Affinity Capture relying on Mass Spectrometry for the remaining genes (Huttlin et al., 2021).

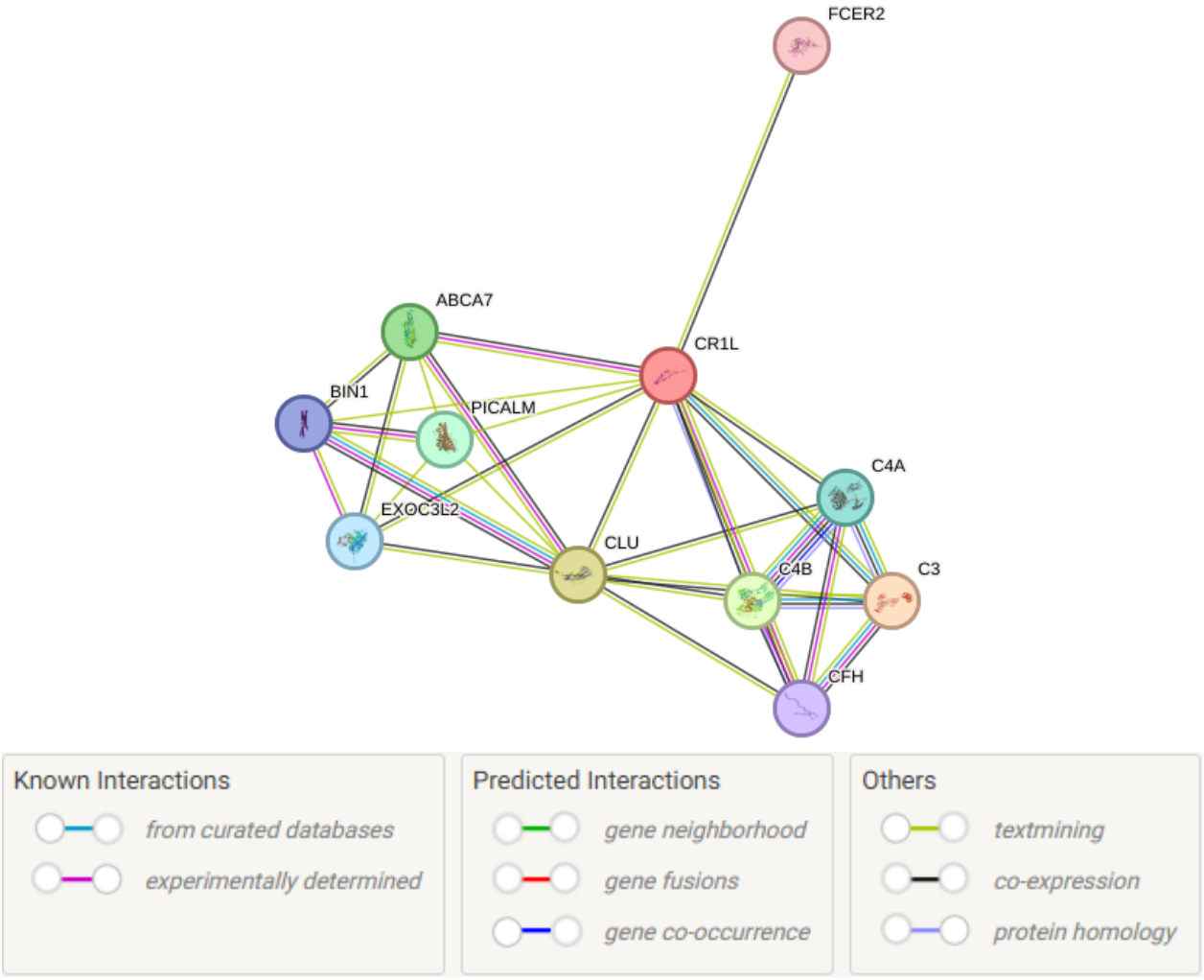


Figure 13 – Resulting STRING network depicting Cr1l relationships with other genes; nodes represent genes with the respective protein structure inside while lines represent interactions colored by source (see legend).

For Cr1l tissue-specific expression, results varied across databases with NCBI GenBank citing a study where circular RNA expression is statistically quantified in human development across 6 tissues samples (adrenal, heart, intestine, kidney, lung and stomach) at different gestational age (between 10 and 20 weeks of gestation). Nevertheless, when observing the results of Cr1l expression in this study across different tissue types, Cr1l is most predominantly expressed in the stomach at 10 weeks of development, followed by being closely expressed in the same quantity at the intestine at 20 weeks of development with the third highest level of expression being observed in the stomach at 20 weeks of development (Szabo et al., 2015). A

more robust study is also cited in the NCBI GenBank database in regards of Cr1l expression levels, relying on a genome-wide RNA-Seq analysis across 20 human tissue types (Duff et al., 2015). This study computes a higher expression level of this gene in the adult spleen, followed by the fetal liver with expression levels dropping majorly in the next most expressed tissue, the adult lung. However, according to the Human Protein Atlas database, Cr1l mRNA expression is attributed to bone marrow and lymphoid tissues, while protein expression is present mainly in the lymphoid tissues namely, bone marrow, appendix, lymph nodes, spleen and tonsils but is also identifieable in brain tissues mainly the hippocampal formation. Additionally, the Human Protein Atlas database also includes the FANTOM5 expression dataset of Cr1l micro-RNA (miRNA) sequences and the respective tissue location. In this dataset, miRNA for Cr1l are found to be expressed in several different tissue types, with the top 5 expression levels being registered in the following order: cerebral cortex, lung, kidney, thalamus, spinal cord and placenta. HPA also claims that Cr1l expression patterns cluster with genes associated with the erythroid cell type, mainly in the oxygen transport expression cluster. The UniProt database also adds information regarding Cr1l subcellular location as a secreted protein but also has evidence of being located in the membrane and the cytoplasm (Logar et al., 2004). Furthermore, UniProt also describes the protein domains, as Cr1l containing 8 sushi domains, also being named complement control proteins modules (CCP) or short consensus repeats (SCR).

In addition to harvesting information regarding Cr1l and its normal functions, disease alternatives were also researched recurring to disease databases. In accordance with OMIM, Cr1l results from a gene duplication event regarding the CR1 gene and has suffered mutations at the beginning of the first of many tandemly repeated motifs, the short consensus repeat 1 (SCR1) (Hourcade et al., 1990). The effects of these mutations are still unknown, although the GWAS database associates Cr1l with systolic blood pressure (Hoffmann et al., 2017), pulse pressure measurement, blood sedimentation (Kullo et al., 2011) and interestingly, Alzheimer's disease (Li et al., 2017). MalaCards, another disease database, supports this neurodegenerative disease claim, as also traces Cr1l with a neurodevelopmental disorder, with the top 3 results being nephronopthisis 1, Alzheimer’s disease 1 and autistic spectrum disorder. The Rat Genome Database (RGD) also establishes Cr1l’s relationship with the incidence of autistic spectrum disorder, with its evidence being inferred from sequence ontology while providing supporting publications (Kushima et al., 2018). Alongside supporting Cr1l as a probable actuator of erythrocyte blood sedimentation, NCBI GenBank also provides a publication where

the Cr1l locus is identified as being related to another neurodevelopmental disease, schizophrenia (Ripke et al., 2014).

4. Discussion

4.1. Quality of the data

The data utilized in this project comprises of 19 samples, 10 of which are control and the remaining 9 being treated with LPS, inducing MIA. After the quality control step, 2 of these samples were excluded for being outliers, reducing the sample size to 17. This raises the concern regarding sample size. In a data analysis project specially, bigger data sets yield more accurate results and low sample sizes taint the outcome of further downstream analysis. In the context of this project, the differential gene expression analysis can't be fully trusted due to the reduced sample size, which may lead to increased variability and potential overfitting of the data. This was considered when applying the linear regression model by including a moderation step relying on the Empirical Bayes method, allowing for the "shrinkage" and the stabilization of the variance and therefore, reduce heteroscedasticity. However, smaller samples have a greater false discovery rate, which makes it more challenging to discern biological signals from noise. In addition, the limited sample size may be inadequate in capturing the underlying biological variability, further complicating the interpretation of the results. To mitigate the increase of false discoveries, the statistical test was performed in the whole dataset to benefit from a robust variance estimation and then subset to only include the genes of interest, namely the complement genes. Although, moving forward, future research ought to encompass a more extensive cohort in order to bolster these results and future conclusions based on the analyses made.

4.2. Differential gene expression

Due to the nature of the experiment, this meaning the infection mediated by LPS injection, an increase in expression of classical and lectin pathway genes is expected. In the case of the alternative pathway, it has been demonstrated that this functions as an amplification mechanism, by creating a positive feedback loop with the activation of the C3 convertase alongside C3 itself. On this project, C3 didn't show to be up or downregulated, in fact the main

activators of this protein were downregulated like Cfp, Cfh and Adn, with the main causes for the occurrence of inflammation being the receptors for C4a, C3a and C5a which take part in classical and lectin pathways.

The mainly upregulated genes C1rl, C1qc, Rps7, Colec10 all take place in the first stages of the classical and lectin pathway, being the primary identifiers of the antibody/antigen complex and lectin/mannose binding proteins respectively. However, the downstream effects of the lectin pathway leading to cell lysis, are shown to be downregulated. This is consistent with the nature of the LPS injection, which is a polysaccharide, explaining the upregulation of mannose/lectin binding proteins. However, this pathogen does not cross the placenta leading to the presence of this mannose binding proteins in the fetal samples to be unknown. This method mimics bacterial infection due to its similarity to the bacterial walls, which the organism recognizes due to the multitude of bacteria that reside and possibly infect during an animal's lifetime, this makes the existence of antibacterial antibodies very likely prior to the experiment. In this case, cell lysis is unnecessary since the pathogen isn't cellular, explaining the normal expression of genes such as C6 and C7 which take part in the membrane attack complex and the downregulation of C8b and C9 which facilitate cell lysis.

4.3. Gene ontology analysis

Study of gene ontology of co-expressed genes alongside the complement system yielded a plethora of results, proving to be difficult for interpretation. This was due to the extensive list of genes chosen to be studied, with a higher number of genes, a higher number of different biological terms is expected. With the cutoff for correlation score being set to 0.8, 2561 genes were selected. Future studies should utilize a higher correlation score cutoff to decrease sample size in GO analysis and get more concise results.

The intention of conducting this ontology analysis is to evaluate the existence of complement co-expressed genes which play roles in neurodevelopment, neuron differentiation and/or other processes which may cause harm to the latter, potentiating NDDs. However, despite the wide variety of results, we can observe that biological processes such as "nervous system development", "apoptotic signaling pathway" and "developmental process" are present in the GO results graph. Nevertheless, one must pay attention to the statistical significance of these three mentioned processes being just under 0.05 for the adjusted p-value.

To better evaluate the relationship between complement system and these biological processes, one must conduct a thorough more complete study utilizing other complement genes for co-expression analysis, since this project only relied on the co-expressed genes of one complement system gene (C1rl). One can also conduct a Weighted Gene Co-expression Network Analysis to better understand and interpret which genes are co-expressed, proceed to subset the genes with higher correlation and finalize with a Functional Cluster analysis to visualize which genes are present in which processes, leading to a more comprehensive vision of the complement system in MIA and its role in the organism.

4.4. Pathway expression analysis and correlation

After applying differential gene expression analysis, the resulting data was integrated into a constructed pathway based on known molecular interactions between the described genes for further analysis. To further add to data, a co-expression analysis through a Pearson correlation test was applied to the complement gene family, this adds to the expression pattern specification providing information on the possibility of gene interplay beyond the resulting pathway. However, these results need to be double-checked through database analysis to ensure their viability and uncover hidden mechanisms that are missing in the pathway. It is also of note that the inference of gene co-expression through a correlation test is not always correct, as gene co-expression relies on a series of regulations and mechanisms that make two or more specific genes resemble a dependent expression profile. A correlation test may overlook these intricacies and only make a connection between genes that may seem co-expressed while in reality, only present similar expression profiles while not being actively related.

With these limitations in mind, a more thorough investigation was applied onto the resulting pathway and correlation analysis through the integration of free access bioinformatic knowledge present in gene databases. Starting with the classical pathway, the assembly of the C1q complex through the binding of the three distinct C1q subunits in a 1:1:1 configuration, C1qa, C1qb, C1qc/C1qg and its receptors C1s and C1r, are implied to resemble same gene expression patterns (Sontheimer et al., 2005). However in this project, differences in differential gene expression of these subunit genes was measured. Although, when investigating into the reason or possibility of this conflicting result, no differences in differential gene expression of complement C1q subunits were identified. A study with similar

methodologies with this project, (M. Kim et al., 2021), regarding complement gene expression and co-expression in the brain showed that C1qa, C1qb and C1qc alongside its receptors C1r and C1s show to resemble same expression patterns and even high scores regarding co-expression, results that are supported in the rat species by the gene expression database Bgee (Bastian et al., 2025). These evidences are conflicting with the results of this project, mainly in the Pearson correlation hierarchical dendrogram plot, which despite showing similarity regarding height, the C1q subunit genes weren't clustered together, which is expected regarding three functionally and expressively similar genes.

An identifiable cluster in the correlation dendrogram that has its genes grouped together in a way that resembles functional and differential gene expression to an extent was the genes present in cluster 6, or Lcn12, C8b and C9. The Lcn12 and C8b genes are part of a small activation pathway that converges in the stimulation of the membrane attack complex. With the addition of lipocalin 12, Lcn12, being justified as it has a relationship with the complement protein C8g, or in other instances named lipocalin 2 (Lcn2), which also has functions regarding fatty acid transportation alongside complement activation, therefore relations with complement protein C8b. These have been identified as downregulated, however the C9 gene, which also contributes to the formation of the membrane attack complex, was presented as upregulated when taking differential gene expression into account. This shows another perceived inconsistency regarding the co-expression assay (Pearson correlation test) and the pathway analysis. Looking at the literature, review articles refer these genes to be related in expression patterns and regulation (Jimenez-Duran et al., 2022), this is also established and supported in experimental data and publications (Khakzad et al., 2021;Lietzén et al., 2018;Serna et al., 2016). This grounded and reliable publications and evidence raise the question according to the validity of this study, with the necessity of a deeper, more thorough investigation needed to be done to assert these results and draw comparisons with for example, other similar datasets.

Another possible shortcoming in this pathway and correlation analysis is seen in the downstream receptors for the complement system proteins regarding the disparate differential gene expression. As previously established, these receptors should indicate similar expression levels since their functions are remotely similar, in this case, receiving signals mainly from upstream complement activation in the form of C3 convertase subproducts and collaborate in the propulsion and continuity of the signaling pathway culminating in relevant biological processes such as phagocyte recruitment, chemotaxis and inflammation. The main issue relies on the receptor genes C3ar1, Cr1l, Cr2 and Crry as these genes differ in expression levels.

Another set of genes that don't have their expression levels coincide with the preconceived expectations are the Zfp275 and Cd93, since these also play a key role in the previously mentioned continuation of the complement pathway, with in this particular case, having their activator being the C1q complex and respective subunits. In this project, a homolog present in the rat genome to the C4a gene, Stard8 is seen as non-differentially expressed and C5, which later becomes C5a, resembles upregulation. This data gives rise to the preconceived notion of that C3ar1, a receptor that is activated in the presence of these one or two ligands, would be non-differentially expressed to even maybe, slightly upregulated. Instead, the data in this project retrieves downregulation for this receptor. The same exact problem is encountered when analyzing C1q's complex expression, being upregulated, while the Zfp275 and Cd93 receptors show downregulation in their gene expression.

For the Pearson correlation results, genes that were revealed to resemble same expression patterns - seen in the volcano plot - clustered in a high familiarity fashion, as seen for example in cluster 3 with the grouping of Cr1l, C1qc and Rps7 representing the highest upregulated genes and with Ldb2, Tbxa2r and G3bp2 representing some of the determined downregulated genes, this implies that the resulting expression changes in these groups occur in a co-regulated or co-expressed way. Compellingly, other genes that show similar gene expression alterations are scattered through the different clusters, with prime examples being the downregulated C4bpa gene which resembles high familiarity with upregulated genes Cfd, Cd46 and C1qa. When observing the complement pathway, there is no inhibition relationship that would explain a higher expression of Cfd, Cd46, and C1qa to interfere in the downregulation of this C4bpa gene. This raises concern regarding the pathway's described interconnections and grouping and the validity of said systems. As the *Pathvisio* pathway seen in this project was built under a preconceived complement pathway template and the remainder of the experimental genes being consequently added with their connections being drawn based of described gene functions in the GO analysis section of the gene's respectively encoded proteins seen on the *UniProt* page. Additionally, some of the genes didn't have their differential expression changes loaded into the pathway, with the reason being unknown as *Pathvisio* doesn't allow for much moderation when importing gene expression data. This step is done by providing each gene their respective NCBI gene ID and then feeding a data frame with three columns, comprising gene expression log2 fold change, NCBI IDs and the system code for referring which database is being used regarding gene identifiers. The user then specifies the gene identifiers and system code columns, while *PathVisio* then assumes the remaining columns to be expression data with

the user not being able to verify the validity of this assumption and moderate wrong results. This can be seen in the C8g gene with resembles two different colors for differential expression, which means that *Pathvisio* considered two different NCBI IDs with different expression changes to the same gene presented in the pathway.

Interestingly, when comparing the visualization of the differential gene expression through the volcano plot with the correlation analysis dendrogram, the significant upregulated Cfh gene is missing from the Pearson correlation results. This gene is then substituted with C8b and this might be due to an error in either the code for the correlation pipeline, namely the fitting of the moderated linear model statistical test for the complement genes expression set subset. While the code took significant genes under consideration in the whole dataset's expression analysis through the moderated linear regression model, this was then correlated with the same method of applying the empirical Bayes moderated linear regression fitting with the difference being the expression set in which this statistical test was performed. This leads to statistically incorrect results as the application of the same moderation approach in two different sample sizes, impact variance and therefore, adjusted p-value leading to the incorrect interpretation as genes which weren't labeled as significant (with an adjusted p-value less than 0.05) may emerge now as significant. This is due to the previously explained variance shrinkage that is more indicated in larger datasets such as the whole experimental data and not the complement subset, as its impact becomes greater in smaller sample size due to the efforts in maintaining a mean variance value.

In general, when taking the necessary precautions *PathVisio* can be an useful tool in analyzing gene expression data in the context of a specific pathway. Although, the scarce options for data curating and moderation leave a lot to be desired. To ensure correct connections and relationships between genes, database cross referencing needs to be adopted as relying on only one may lead to false and incorrect assignments. Users can utilize databases such as Reactome to better understand and identify other key protagonists in the studied pathways that may unveil otherwise hidden regulation mechanisms that may explain differences in gene expression for the same class/groups of genes. The application of well curated expression data aids in the assessment of changes in the pathway and helps draw meaningful conclusions. As for the correlation analysis, significance and therefore variance needs to be adjusted to the correct indicated scale of the data in order to provide valid co-expression patterns that once again, can be valuable for the understanding of changes in the pathway.

In this project, *PathVisio* pathway analysis was faulty due to the inexperience of using this program. As previously mentioned, relations weren't thoroughly cross referenced leading to an incomplete pathway and a non-reliable model of the complement gene and its processes. However, gene expression data extracted from the differential gene expression analysis pipeline is set to be trustworthy, leading to interesting observations as inconsistencies and unpredicted results were determined as mentioned before. These changes should be further investigated in other projects and/or studies with an effort in trying to understand and describe why these are observed. Perhaps the integration of proteomics and even metabolomics should be sought after to provide more information regarding other possible interactors such as regulators, inhibitors, suppressors and other types of modulators in the complement pathway. Furthermore, more datasets regarding complement activation and/or inflammation should also be analyzed to provide a comparison model of complement activity for MIA-related and non MIA-related inflammation and in this way, assess if MIA provides a different complement activation profile than another mechanism for inflammation. Alternatively, a bigger sample size or multiple data sets of the same experimental design (Control vs MIA) can also be used to mitigate experimental errors that in turn, impact the results and therefore, conclusions. To summarize, results in this project regarding pathway analysis should be reviewed and reanalyzed by the implementation of more data-driven precautions and therefore, results from the previously commented methods should be looked at with attention and due diligence.

4.5. Database research on the role of Cr1l

Open-source gene databases provide invaluable resources for researchers, enabling free access to a wealth of information including genetic, proteomic and clinical data. These repositories aid and facilitate scientific discoveries, enhancing collaboration and promoting transparency in research, by allowing the databases to be updated by integrating and citing publications and their respective advances while also providing vast amounts of biological information. However, while accessibility is a significant advantage, certain precautions must be taken when utilizing these resources. Major concerns in the utilization of publicly available databases include data accuracy and reliability, as open-source databases rely on user-submitted content, opening the possibility of including errors or inconsistencies in the experiments. To remedy this, some databases refer their inclusions as verified or peer-reviewed, this allows for a curated search as one can only choose these reliable entries. Additionally, researchers must verify findings by cross-referencing through other available

databases or by consulting supporting literature. Another ethical and critical consideration is data privacy, particularly with clinical and/or genomic databases containing sensitive human genetic information. To ensure ethical and regulatory compliance, informed consent forms must be applied to patients before the submission of their data in a publicly accessible database, to protect the person's data from being used and exposed against their will. Furthermore, while many databases provide extensive annotations, these may not be up to date, necessitating regular verification by moderators and researchers to account for new discoveries. Despite these challenges, open-source databases remain an essential tool in modern biological research, as long as researchers exercise due diligence in data validation through the inclusion of recent supporting studies, ethical considerations and reliable database selection.

With these limitations in mind, results should be analyzed from a critical perspective. For example, most of the publication-based inferences are automated and therefore do not include the overall context. This is prevalent in the resulting STRING network, where most of the interactions are based on a text mining method. For an interaction to be included in the network, STRING relies on a name entity recognition algorithm based on the context of the literature. Upon defining a name entity, the algorithm reviews the context and returns a list of bad names that are explicitly mentioned in the publication but don't necessarily mean interactions. The remaining names found in the text are then taken into consideration and undergo through calibration steps (e.g. cross referencing the interaction statement into a protein complex database). This step allows for the calculation of posterior probabilities for each individual statement and the best score in each publication is considered and appears as a predicted interaction in the STRING network. Although having these moderation steps, automated text-mining will always have the probability of establishing false or incomplete inferences. Another limitation in this process is that the protein interaction statements are not cross-referenced with other publications and in that way, increase the probability of identifying wrong interactions or missing important context that is built upon other studies.

Knowing about the limitations of this method, text-mining inferences were removed from the STRING network analysis, leaving only 3 directly interacting proteins: FCER2, HEMGN and C3 (Figure 14). All of these protein interactions are labeled as co-expression with C3 having a known interaction based on a curated database, this being the Kyoto Encyclopedia of Genes and Genomes (KEGG) database with the interaction being associated with the annotated pathways for legionellosis, leishmaniasis, tuberculosis and more importantly, complement and coagulation cascades. This database provides interaction maps where, in case, the role of Cr1l

is easily visualized as a complement C3b and C4a receptor acting upstream on the regulation of the inflammatory processes of the mentioned diseases. Interestingly, across all the KEGG annotated pathways resulting from the STRING Cr1l-C3 interaction output, Cr1l (or often associated with Cr1) displays a role in the downstream signaling for the phagocytosis of pathogens. Additionally, other genes were associated in the resulting STRING search with Cr1l by being co-expressed in accordance with GEO microarray data. One of the resulting genes was FCER2, a low affinity immunoglobulin (Ig) epsilon Fc receptor, which presents an essential role for IgE regulation and production. The immunoglobulin IgE is often associated with parasitic pathogens and is secreted by eosinophils. This relationship seems plausible, considering the role of the complement system in the presence of pathogens with supporting data relying on the KEGG database's Cr1l interactions in the presence of parasitic pathogens. Another gene was labeled as co-expressed by the STRING interactions results, HEMGN, a hemogen that plays a crucial role in the regulation and differentiation of hematopoietic cells. This coincides with the disease database research where Cr1l is associated with hematopoietic cells, namely the erythrocytes and blood sedimentation disorders.

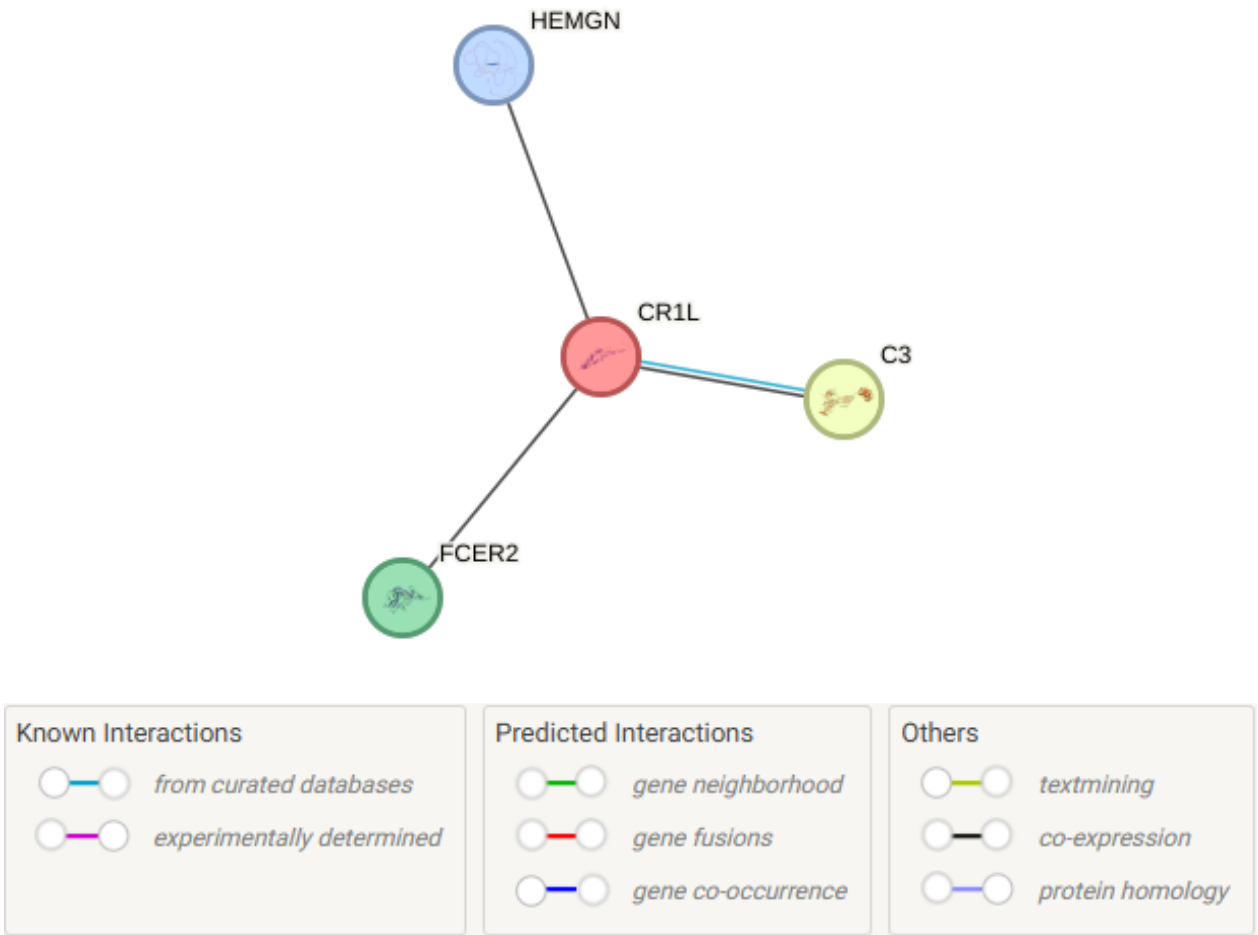


Figure 14 - Resulting STRING network depicting Cr1l relationships with other genes, with text mining results removed; nodes represent genes with the respective protein structure inside while lines represent interactions colored by source (see legend).

However, these results only seem to partially describe the effects of Cr1l since this protein is expressed in other tissues, like the brain according to HPA. Upon inspection of these results, Cr1l has a low detection rate in neuronal cells and no detection in glial cells. This seems contradictory as complement proteins are often described as being secreted by the glial cells, namely astrocytes. Nonetheless, regarding the intracellular aspect of complement proteins or complosome, these have been described as present in the neurons and act in these cell's metabolism in the event of hypoxia. This expression data is based on two FANTOM5 hippocampal tissue samples with no more brain subareas being explored. Single cell type enrichment assays provided by HPA show that Cr1l is enriched to erythroid cells although, single nuclei cluster type assays resulted in low cell type specificity while returning several hits in several different brain subareas. These conflicting remarks call for the continuation of the research regarding Cr1l's hidden role in the nervous system since it is inferred with neurodegenerative and neurodevelopmental disorders in the MalaCards disease database. Additionally, Cr1l's locus is cited in a publication as a risk for copy-number variation described in the etiology of autism spectrum disorder and schizophrenia. Another study evaluates possible risk loci for schizophrenia where, according to GeneCards, Cr1l's locus is included.

In an attempt to add information and fill out some blanks in the role of this protein, a decision was made to run the Complement receptor 1 (Cr1) protein, as it exhibits 96% homology with Cr1l (Logar et al., 2004), in Reactome and observe its related pathways and biological processes, since Cr1l didn't retrieve a response. This search resulted in the identification of 9 different biological processes where Cr1 acts, namely the transcription of RUNX1 and FOXP3 and their role in the development of regulatory T lymphocytes, regulation of the complement cascade, transcriptional regulation of RUNX1, neutrophil degranulation, innate immune system, generic transcription pathway, RNA Polymerase II transcription, gene expression (transcription) and immune system. At first glance, Cr1 presents an important role in transcription and gene expression, while Cr1l didn't seem to share these properties. When investigating further, Cr1 is mentioned as a key factor for the generation of regulatory T cells

(Tregs) – this conclusion is extracted from the first, most scored biological process by Reactome when queried for Cr1. The RUNX1 protein complexes with CBFB to control the transcription of the FOXP3 gene. FOXP3 is a transcription factor that regulates the development and function of Tregs, T lymphocytes that suppress aberrant immune responses (Kim et al., 1999). The RUNX1:CBFB complex also stimulates the transcription of the Cr1 gene that, when expressed on the surface of activated T cells, contributes to the generation of Tregs (Török et al., 2015). Cr1 also has another name as often being described as CD35. This mechanism somewhat resembles the described complosome and the CD46 receptor effects in the human activated T cells when recruiting Th1 cells. Cr1 is also described as a regulator of complement activation (RCA) protein. These regulators mainly function by two different mechanisms, by promoting irreversible dissociation of complement convertases (decay acceleration) and by acting as cofactors for Complement factor I (Cfi) mediated cleavage of C3b and C4b generating iC3b and iC4b (Zipfel & Skerka, 2009). CD55, also known as decay accelerating factor alongside Complement factor H (Cfh), membrane cofactor protein (MCP) and Cr1 work as RCA proteins with Cr1, MCP and Cfh working together as cofactors for the cleavage of C3b and C4b. Cr1l is also described as a cofactor for this same mechanism, however it exhibits binding specificity only for iC4b and not iC3b (Logar et al., 2004). C4bp is an additional cofactor for Cfi-mediated cleavage of C4b. Cr1 also displays an important role in the phagocytosis of apoptotic cells for clearance by the macrophages, possibly explaining Cr1l's phagocytic signaling capabilities. This is observed when a particular event in apoptotic cells results in the altered exposure of cell surface phospholipids, leading to the coating of these irregular membrane components by iC3b. This iC3b coating is then recognized by C1q and other opsins of the complement system, such as C3, C4 and MBL. This leads to the recruitment of macrophages by the binding of these opsins and iC3b coating to the macrophage surface receptors, Cr1 included (Mevorach et al., 1998) as this receptor also recognizes C1q (Ogden et al., 2001). This leads to the "silent" phagocytosis of the apoptotic cells by the macrophages leading to the clearance of these cells (Flierman & Daha, 2007).

Like Cr1l, the Cr1 protein is also a key protein in erythrocytes, revealing its main function as the clearance of immune complexes from the circulation. This protein is expressed in clusters in erythrocytes and acts as an adherence receptor for C3 and/or C4 fragments deposited on immune complexes, being transported to the liver and spleen, where these complexes are then processed by neighboring tissue macrophages. The Cr1 protein doesn't only bear this function within the blood tissue, as complement pathways directly interact with coagulation pathways

by acting on platelet activation and/or changing endothelial cell morphology potentiating coagulation. These different facets of Cr1 and its roles further support other databases' information regarding the different biological processes in which this protein acts. However, despite Cr1 being a close relative to Cr1l, one can't assume and draw direct conclusions regarding its functions. Cr1's functional data is merely to be taken under consideration as hypothesis and not postulations, when assessing Cr1l.

In spite of returning several interesting results across different biological processes, the main point in this project's aim is not addressed in the Reactome pathway browser research. Although being presented as expressed on neurons of the hippocampal region of the brain, Cr1l's role in that tissue is not yet specified. As previously mentioned, some complement proteins have been linked with the activation and regulation of neuronal intracellular processes in the presence of hypoxia, yet Cr1l is not described in these assessments. However, Cr1l is marked as a risk protein for neurodegenerative and neurodevelopmental disorders in several different studies. While Cr1l is not described as directly participating in this intracellular mechanism, a plausible hypothesis may be traced back to its prominent role in blood coagulation which, when dysregulated, may lead to the formation of blood clots and the blockage of blood vessels in the brain. This blockage is deemed as ischemia and leads to the depletion of oxygen and nutrients provided by the blood in neighboring cells. In the presence of hypoxia, intracellular complement in the neurons is activated leading to radical metabolic changes in order to accommodate the low oxygen levels in the vicinity. The presence of hypoxia in the fetal development, can affect the brain's development leading to a higher incidence of neurodevelopmental diseases. Nonetheless, in the events of MIA, the resulting inflammation and subsequent complement activation isn't expected to trigger abnormal blood coagulation but due to the interplay between complement and coagulation pathways, further studies have to be considered.

It's important to understand that these results are based on the search of only one protein, in this case Cr1l and Cr1, in the event of Cr1l being unavailable. While extremely informative, bigger search datasets should be used in the research of the functional pathways and interactions. Cr1l was chosen as the most correlated result following a Pearson Correlation test ran on the whole project dataset with the complement genes dataset. Alternatively, to better assess the relationship of the complement genes, specifically in the MIA state, differentially expressed genes can be analyzed recurring to the aforementioned open-source databases. This can lead to a better understanding of the MIA alterations on the complement pathway for both

down and upregulated differentially expressed genes. In addition, a cut-off point can be settled for the correlation score of the correlation matrix to study the interrelations of the most positively correlated genes when comparing with the complement gene data set. Another alternative is to engage in the multiple analysis of gene interactions by querying the most closely Cr1l related genes, as seen on the hierarchical clustering of the correlation test, mainly C1qc and Rps7. This alternative not only is able to add context and other insights into the Cr1l functions as a protein but also as a possible candidate for a mediator role in other pathways by engaging in the co-expression of other complement proteins. Although its role in neurodevelopment is still unknown and hypothetical, Cr1l presents to be a multi-faceted protein acting in many different pathways and tissues. While Cr1l's particular function remains unclear, due to the high similarity with Cr1, it is traced as being a complement receptor protein and contributing to the overall homeostasis in various pathways. For example, in the immune system, as an activator of regulatory T cells and in the clearance of apoptotic cells by macrophages and in the circulatory system by acting in the clearance of immune complexes in the blood. This homeostasis-contributing pattern alongside the several links to neurodevelopmental diseases, raises the possibility of a hidden role of Cr1l in the brain by perhaps remaining as a key factor for the system's homeostasis and normal function that should be further investigated in the future.

4.6. Complement's activity during MIA and its effects on neurodevelopment

Research on Maternal Immune Activation has significantly expanded our understanding of the links between prenatal immune challenges and neurodevelopmental disorders such as autism spectrum disorder and schizophrenia. While firstly being described as originating from IL-6 and IL-17a cytokine dysregulation (Patterson 2009) and leading to neurodevelopmental changes, MIA has been gaining some traction in the immunology research space, with several studies being published looking into etiology, brain structural malformations, biomarkers, behavioral deficits and disease mechanisms. However, not only cytokines have been found to be involved in the exacerbation of these symptoms and disorders, with pathogens that induce inflammation alongside activation of Treg cells have been found to also impact this immune activation (Xu et al). Alongside Treg cells, other cells from the immune system family have been considered as playing key roles in maternal immune activation such as microglia (Smolders et al) and macrophages (Onore et al). While not being yet known, experimentally mechanisms in the aggravation of maternal immune activation have been described as the

activation of toll-like receptors (Sheng, 2017), the TNF-alpha signaling pathway (Carpentier, 2011) alongside neutrophic factors, neurotransmitters and epigenetic mechanisms (Gilmore, 2005) (Meyer, 2008) (Basil, 2004). With these mechanisms altering brain connectivity, synaptic pruning and neurocircuitry implicating for example, in dopaminergic neurons (Hayes 2022). Another important immune system component that acts through the entire body through circulation is the complement system. Complement system is described as an accelerator for immune response, acting in the identification of pathogens through three distinct pathways, recruitment and activation of T helper cells, provoking cell lysis of the pathogen and inducing phagocytosis of apoptotic cells from the macrophages.

Naturally, when evaluating the effects of MIA, researchers also investigated the complement system due to its multi-faceted functionalities in the immune system. The complement system, mainly the C3 complement gene, also plays a role in neurodevelopmental processes such as synaptic pruning, microglial activation, and neural connectivity (Peterson 2017). The complement system's role in the immune system and neurodevelopment makes it a key candidate for scientific research and investigation of MIA and its effects on neurodevelopment and the incidence of neurodevelopmental disorders. This claim is further confirmed when studies show that maternal inflammation can induce abnormal complement activation in the brain's development with potential of dysregulating normal neuron organization, resulting in excessive or deficient synaptic pruning, decrease dendritic spine density and altered excitatory and inhibitory presynaptic neurons, impacting brain connectivity and further increasing the odds of autistic spectrum disorder and/or schizophrenia with long-term cognitive and behavioral deficits that lead into adulthood (Coiro et al 2015).

Furthermore, the utilization of the rodent model for MIA lead to significant discoveries such as the excessive expression of the complement genes in the maternal serum and the respective overactivation of the signaling pathways also provoke an exaggerated and inappropriate synaptic elimination, which in turn contributes to behavioral deficits observed in offsprings. In another study relying on schizophrenia patients, revealed that complement gene C4 is also revealed to be overexpressed in the brain and it shows to be related with synaptic pruning as well. This is also seen with activated microglia (Loyaza 2022), disrupting neurogenesis and leading to pathological synapse loss and miswiring of neural circuits, raising the hypothesis of a possible relationship between complement system and microglia activation. This was then tested in a recent study where complement and microglia seem to tandemly activate one another as microglia produce C1q that labels synapses to induce their clearance by

the microglial cells, while complement system would activate microglia through astrocyte-derived C3 inducing the MAC1-NOX2 axis or through astrocyte produced C5 which activate the C5aR1 receptor which in turn, through the activation of the MAPK, PI3K and MEK signaling pathways activating the microglia which induces synaptic pruning by the production of cytokines. Additionally, complement might act independently by recurring top the formation of the membrane attack complex provoking cell lysis in the targeted neuron (Zhao 2025).

In this project, by assessing differential gene expression and loading the results in a large complement system pathway for a better applied analysis, key changes were observed. The main evident suggestion that can be extracted from these results is the existence of a selective activation of the classical complement pathway, due to a higher expression of the C1q-related genes, in the offspring's brain following maternal immune activation (MIA), while the alternative and lectin pathways as well as downstream genes such as C2, and C3 remain largely unaffected. The observable upregulation of the complement inhibitors potentially points to a tightly regulated complement response in the event of MIA. Concerning the C1q genes, as previously established in the literature, are known to mediate synaptic pruning in neurodevelopment by tagging synapses for elimination by microglia. The presence of excessive C1q expression levels through its upregulation has been implicated in several neurodevelopmental disorders such as schizophrenia, autism spectrum disorder, and even, the neurodegenerative disease Alzheimer's. If MIA contributes to a persistent increase in C1q levels, excessive synaptic elimination can occur, contributing to a large alteration in brain connectivity in the offspring and therefore, increase incidence rates for neurodevelopmental diseases. C1q also is known as having a role in microglial activation, promoting a more inflammatory and phagocytic state, which could mean that glial cells such as microglia and/or astrocytes are being primed for activation, even if full-blown canonical inflammation isn't occurring.

Another supporting result in this project is the downregulation of the complement inhibitor protein, CUB and Sushi Multiple Domains 1 (Csmd1) gene. Csmd1 is a complement regulatory protein that inhibits the classical complement pathway by interacting with C1q preventing excessive complement activation. Given that Csmd1 is downregulated in this project's rodent MIA model, while C1q remains highly upregulated and other complement inhibitors are also upregulated, this suggests the occurrence of a complex dysregulation of complement control. Csmd1 normally restricts the C1q pathway, preventing excessive synapse targeted tagging and microglial phagocytosis. In the events of Csmd1 downregulation, resulting C1q inhibition

would be decreased leading to an upregulation of C1q and in turn, excessive microglia-driven synapse elimination. This contributes to an array of synaptic deficits, altered connectivity and cognitive/neurodevelopmental impairment in MIA offspring. Csmd1 has also been related and described to having an impact and being named a schizophrenia risk gene. With loss of Csmd1 being associated with disrupted homeostasis, cognitive deficits and increased inflammation. With this in mind, Csmd1 downregulation in this MIA model could suggest a mechanism contributing to schizophrenia- or ASD-like phenotypes. Despite limiting complement activation, Csmd1 also reveals to having anti-inflammation properties such as regulating microglial responses. With the occasion of low Csmd1, C1q-driven complement activation might make microglia more pro-inflammatory, creating a primed or sensitized inflammatory state in the brain, leading to long-term alterations in the brain's resistance to inflammation, potentially making MIA offspring more vulnerable to neuroinflammation in adolescence or adulthood. To further add to the Csmd1 findings, this gene is also found to be expressed in neural progenitor cells (NPCs) and influences neuronal differentiation and survival. When Csmd1 levels are reduced, this might impact neurogenesis, leading to altered neuronal populations with connectivity defects. Other studies have shown that Csmd1 deletion in mice, leads to altered social behaviors and increased repetitive behaviors, these of which are hallmarks of autistic-like behaviors drawing a relation with ASD incidence. If MIA is confirmed to suppress Csmd1 expression, this could be one of the mechanisms contributing to autism-like phenotypes in the rodent model.

The resulting observation of the upregulation of other complement inhibitors such as, SERPING1, Cfh, Crry and others, suggest the complement system's active attempt to suppress excessive complement activation. This could indicate a compensatory response to prevent excessive inflammation or synapse loss. If complement activation were fully unchecked and unregulated, it could lead to more successive neuroinflammation and neuronal damage, which might indicate why these regulatory genes are increasing. A key regulator of immune cell activation, Cd45, is found to be expressed in microglia, macrophages, and other immune cells (Rangaraju et al., 2018). According to studies, its upregulation in the MIA offspring model (Makinson et al., 2017), alongside C1q and decreased Csmd1, strongly suggests that microglia are activated, but in a specific, potentially regulatory manner. Whilst Cd45 is a hallmark for activated microglia, it can also have both pro- and anti-inflammatory functions, depending on the context. Given that other complement inhibitors or regulators are upregulated, this allows for the prediction of a more regulatory type of activation, rather than overt neuroinflammation.

Alongside its activation, MIA also induces microglia to be a “primed” state, this means leaving them more reactive and sensitive to future insults, like secondary infections, stress or aging. Cd45 upregulation suggests that these microglia might be in a primed or surveillance state, meaning that these are ready to respond but not necessarily causing damage yet.

These levels of high complement classical activating expression pattern of C1q coupled with the low inhibiting Csmd1 and high expression of CD45 in microglia may suggests the continuous microglia activated phagocytosis or synapses, although more supporting studies need to be performed to confirm this hypothesis. Another possible mechanism of activation is the synaptic remodeling that takes part in neurodevelopment by the microglia cells. These two hypothesis may suggest that in tandem, microglia are excessively taking part in synaptic pruning and brain connectivity remodeling which would be supported with potential cognitive deficits that are observed in neurodevelopmental disorders.

Another possible take away from CD45 upregulated levels is its ability in shutting down excessive inflammation by dephosphorylating certain key signaling molecules to limit their effects in immune activation. When assessing this property of the CD45 gene, inferences that $CD45^{+}$ upregulated microglia can therefore act as immune activation inhibitors by perhaps, upregulating complement inhibitory such as Csmd1, having a somewhat protective or homeostatic role in trying to counterbalance excess C1q and prevent neurotoxicity. This neurotoxic prevention is further explained as Cd46, a key complement regulatory protein, shows to be upregulated which in turn, inhibits complement activation by preventing excessive C3b and C4b deposition. This might suggest that the brain upregulates C1q without upregulating classical central functions executed by C3 and C3b, but instead suppressing this downstream part of the complement functionality. It also suggests that Cd46 could be limiting full complement activation, potentially preventing neurotoxicity or excessive inflammation. Overall, this potentiates the hypothesis that C1q in the maternal immune activation context is fulfilling alternative functions besides complement activation.

Looking forward, C1q’s role needs to be further studied as also, an important direction for MIA research is to integrate genetic susceptibility and environmental risk factors into models of complement dysregulation. Given that genetic variants in complement-related genes (e.g., C4, C3R) have been linked to psychiatric disorders, understanding how these genetic factors interact with maternal immune responses could help identify mechanisms that reveal higher incidences of neurodevelopmental disorders. Additionally, further more complete studies

tracking complement's activity in maternal immune activation in the rodent model could provide valuable insights into how early immune disruptions translate into later neurodevelopmental outcomes.

Overall, while the complement system is emerging as a critical mediator of MIA-induced neurodevelopmental abnormalities and neuroinflammation, much remains to be explored. Shortcomings in this project leave much yet to be discovered like the role of Cr1l, microglia in the events of MIA, pathway analysis of maternal immune activated complement and furthermore. Future research should focus on dissecting the precise mechanisms by which complement proteins contribute to neural synaptic dysfunction, leveraging novel *in vitro* such as IPSCs and *in vivo* models, and once MIA's effects are thoroughly described, a door may be opened for identifying potential therapeutic targets to mitigate the long-term effects of maternal inflammation on brain development.

4.7. Future perspectives

With the resulting observations of this project, future changes must be taken under consideration for the continuation of the study of complement system's actions and mechanisms on the maternal immune activation front and their respective impacts in the neurodevelopment, neuroinflammation and neuropsychiatric and developmental diseases profile. This implies continuing to try to categorize the differential gene expression of the complement system components and establishing a functional and pathological relationship between the already described effects of MIA in behavioral, neurogenesis and anatomical changes. Due to this project's shortcomings, additional future studies should be implemented to furnish a wide understanding of complement's role in MIA by providing more results, facilitating data discussion and comparisons.

Main changes that need to be implemented is the expansion of samples repository, as this project only had 19 samples with 2 of each condition being removed due to being flagged as outliers. This reduction in experimental data through the exclusion of outlier samples is an advised procedure that aids in the production of statistically correct conclusions, however the reduction of samples from an already small size, impairs later downstream results as a small sample size as the higher chance of including experimental errors and therefore wrong data. A suggested alternative is to introduce several other similar MIA-induced rat microarray datasets as long as the same model is applied, and the same measurement protocols are coincided (e.g.

utilizing *Affymetrix* microarray chips). Inclusion of the fetus' sex is an interesting procedure due to the existence of studies implying different neurodevelopmental changes in the case of MIA when compared by each sex. Another alluring suggestion may be the addition of different gestational day measurements, due to the already perceived notion of gestational timing of MIA impacting in several different forms. The inclusion of this step would result in the addition of another set of conditions which adds the precaution of having to change to a proper advised statistical method for the extracting of significant data.

As for the transcriptomics through differential gene expression analysis, microarray data was utilized. Another advisable choice for the change of this method is the use of RNAseq data. This is due to the microarray disadvantages that this method presents and the late resurgence of RNAseq as the golden standard for differential expression analysis and transcriptomics. Despite their widespread use, microarrays present a set of challenges that give rise to a series of limitations that hinder accuracy, sensitivity and applicability to modern transcriptomic research. The primary limitation of microarrays is their restrictive dynamic range of probes and sensitivity. As previously explained, microarrays rely on fluorescence intensity depending on the success of RNA hybridization, this may lead to signal saturation for highly expressed genes and difficulty in detecting low-abundance transcripts due to a low-level fluorescence signal. Another worrisome characteristic of microarrays is the eventuality of background noise, this is the "spilling" of fluorescence from other probes that may impact the intensity reading of neighboring probes and distort expression measurements. While the *affy* package provides a tool for background normalization, the possible existence of this confounding false data is still not desired and, if the whole microarray chip resembles a high occurrence of background noise, normalization algorithms will adjust to that value and allow for a wrong normalization and in turn, wrong results. Additionally, microarrays are inherently constrained by the requirement of pre-designed probes, meaning that only the loaded probes in a said microarray chip will result in expression data and this leads to the detection of genes that are already known. As a result, these are ineffective for identifying novel transcripts, splice variants and non-coding RNAs which limits their utility in exploratory research. Another drawback of microarrays is the inaptitude of detecting alternative splicing and transcript isoforms, since microarrays measure expression at a probe level, since these lack the resolution needed to differentiate between alternatively spliced variants of the same gene, since these hybridize with similar probes. This drawback prevents researchers from identifying novel transcripts, splice variants and non-coding RNAs as well as it also prevents form obtaining a

complete view of the transcriptome complexity. In addition, microarrays provide a semi-quantitative approach for differential expression rather than absolute quantification, as their fluorescence-based intensity measurements are relative rather than direct counts of RNA molecules.

In contrast, RNA-Seq offers numerous advantages make it the preferred method for transcriptomic analysis. As it first provides a vastly improved range and sensitivity, allowing for both low and high abundance of transcripts while unlike microarrays, RNA-Seq is not limited to pre-defined sequences, enabling the discovery of novel genes, variants and non-coding RNAs that may have an impact on MIA and/or filling the blanks between complement activation and neurodevelopmental changes. Moreover, RNA-Seq enables precise detection of transcripts, eliminates cross-hybridization artifacts and directly reads sequences which improves data accuracy and reproducibility.

Due to the conflicting results of the Pearson correlation test to assess gene co-expression, precautions need to be taken when utilizing this method. While being usually advised for this type of procedure, its application is involved with several limitations of its own. One of the issues with Pearson Correlation is the sensitivity to outliers which is particularly problematic in biological data, where variability and technical noise is common, reducing the reliability of inferred gene relationships especially for the study of co-expression. Another striking limitation is the assumption of linearity due to the methods strict measurements of linear associations while many biological interactions are non-linear and relative. These interactions include transcription factor activity, negative and positive feedback loops which resemble nonlinear relationships that cannot be captured by a correlation coefficient. This is important since, the alternative complement includes a feedback loop of its own and already mentioned possible relationships with transcription factors. Additionally, Pearson correlation assumes normal distribution of gene expression, which often result in skewed or bimodal distributions. This limitation further complicates the interpretation of co-expression networks, as spurious correlations may arise from differences in gene expression distributions rather than true biological relations. Despite apply a statistical moderation step when accounting for differential gene expression, the inclusion of a distribution assumption on top of a moderated statistical variance allows for the argument of over-preparation and transformation of the data. Since the raw statistical results were transformed in order to append to the packages and statistical tests chosen to perform the analysis. Finally, Pearson correlation fails to account for condition-specific expression patterns as a single correlation coefficient is calculated across all samples,

masking context-dependent relationships that may vary under different conditions, in this case control and MIA states. Alternative approaches rely on the choosing of another correlation test which account for non-parametric approaches, with examples being Spearman's rank correlation for capturing nonlinear associations such as the mentioned feedback loops and with Bayesian network models and partial correlation analysis to help differentiate direct from indirect interactions.

Due to the already discussed several limitations observed in *Pathvisio* program, other programs that allow for the upload of differential gene expression data for pathway visualization should be taken into consideration. With these being Cytoscape, an open-source software tool which allows for the inclusion of useful plugins that import predetermined pathway templates and allow for the upload of expression data allowing for pathway enrichment; The KEGG Mapper and DAVID functional annotation tool which also allows for differential gene expression data to be uploaded while relying on KEGG's already-submitted vast pathway repository and DAVID for clustering analysis. Alongside a change in the pathway exploring software, added measures must be taken in cross-referencing gene relations recurring to the use of publicly available databases to ensure the highest quality and reproducibility of the results.

To convey more complete results, in addition of resorting to transcriptomics alone, other multi-omics approaches may be integrated, for example proteomics which provides a look into the complement's protein functions, modifications and interactions within a biological system. Allowing for a deeper understanding of the dynamic complement protein landscape and the respective changes in the events of MIA. As proteomics tend to rely on overpriced methods that usually rely on for example mass spectrometry, x-ray crystallography and nuclear magnetic resonance spectroscopy, protein microarrays are indeed the best option as it allows for a high-throughput study of proteins while remaining cost-effective. Metabolomics are also worth of investigating, since the complement system pathway has an extensive list of interactors, cofactors and substrates that may allow for the observation of yet known intersystem relations. Metabolomics also allows for the study of the environment's impact on a list of metabolites that may reveal key discoveries in biological processes. In this case, the study of complement associated or derived metabolites may give key insights into the impact of MIA and draw relations to neurodevelopment and neuroinflammation metabolic factors. In order of adding to the expression analysis, a case for epigenomics can be made since epigenetic alterations in the genes lead to changes in expression. This approach, coupled with transcriptomics, allows for a

better understanding of gene regulation and the impact of the environment in gene expression. This is a particularly interesting field to be investigated since, MIA in some sense, provokes changes in the environment which is key in developmental biology. Moreover, the continuous expression changes in for example, microglia when becoming primed may be explained through epigenomic data. Perhaps an alteration of neuron epigenome is observed when these are merely neuronal progenitor cells and affected through neuroinflammation, to change their expression patterns which continuously perpetuates to adulthood, since schizophrenia and autism are only phenotypically addressed in adolescence despite inflammation in the developmental stages.

Additionally, *in vitro* experiments can be also important in the characterization of cell specific changes in for example, a neuron cell line or better yet iPSCs that more accurately resemble a fetal phenotype and can be differentiated into NPCs that allow for neurodevelopment experimentation assays. While *in vivo* experiments with MIA models tend to be preferable, cell culture provides a more interactive approach to determine mechanism-related changes in the events of inflammation and neurodevelopment. Cell culture and *in vitro* experiments allow for a more flexible approach in the study of inflammation, as for example, a complement system activation model can be designed by inducing the upregulation of a predetermined complement gene and proceeding to evaluate the pathway differences through single-cell RNA-seq. Cell culture also has the advantages of being more accessible since this approach doesn't resemble as many ethical barriers as the use of animal models and require less proper infrastructure and management like animal enclosures and feeding procedures. The development of iPSCs also add to the possibilities within the fetal model studies, since it has the advantage of pluripotency allowing for the research of MIA in other fetal related tissue.

5. Conclusion

In general, bioinformatics presents to be a powerful tool in the scientific community, being able to perform trustworthy statistical studies, data visualization and other forms of investigation only in one program. With most programs and databases being open-source, resources present to be widely available to conduct experiments and projects such as this one, which barely scratches the surface within the bioinformatics field.

In this project, one can verify the changes in the expression of the fetal complement system genes in the event of MIA, which leads to changes in biological processes such as inflammation, phagocyte recruitment, chemotaxis and even in muscle contraction as seen in the pathway analysis. Further along, MIA also induces differential gene expression in a wide variety of genes as proven by the correlation analysis which resulted in more than 2000 co-expressed genes when compared to one complement system gene. This is further confirmed in in GO analysis which demonstrates the wide variety of processes being activated alongside the complement system.

This results in MIA being a threat to the normal balance of the fetal organism. However, to better explore the complement system's role in neurogenesis, NDDs and neurodevelopment itself in the context of MIA, further studies must be conducted, increasing sample size, utilizing different models and different analysis methods.

6. Bibliography

Abdallah, M. W., Larsen, N., Mortensen, E. L., Atladóttir, H. Ó., Nørgaard-Pedersen, B., Bonefeld-Jørgensen, E. C., Grove, J., & Hougaard, D. M. (2012). Neonatal levels of cytokines and risk of autism spectrum disorders: An exploratory register-based historic birth cohort study utilizing the Danish Newborn Screening Biobank. *Journal of Neuroimmunology*, *252*(1–2), 75–82. https://doi.org/10.1016/j.jneuroim.2012.07.013

Akhlaghpour, H., Wiskerke, J., Choi, J. Y., Taliaferro, J. P., Au, J., & Witten, I. B. (2016). Dissociated sequential activity and stimulus encoding in the dorsomedial striatum during spatial working memory. *ELife*, *5*, e19507. https://doi.org/10.7554/eLife.19507

Arbore, G., West, E. E., Rahman, J., Le Friec, G., Niyonzima, N., Pirooznia, M., Tunc, I., Pavlidis, P., Powell, N., Li, Y., Liu, P., Servais, A., Couzi, L., Fremeaux-Bacchi, V., Placais, L., Ferraro, A., Walsh, P. R., Kavanagh, D., Afzali, B., … Kemper, C. (2018). Complement receptor CD46 co-stimulates optimal human CD8+ T cell effector function via fatty acid metabolism. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-06706-z

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S., & Soboleva, A. (2013). NCBI GEO: Archive for functional genomics data sets - Update. *Nucleic Acids Research*, *41*(D1). https://doi.org/10.1093/nar/gks1193

Bastian, F. B., Cammarata, A. B., Carsanaro, S., Detering, H., Huang, W. T., Joye, S., Niknejad, A., Nyamari, M., De Farias, T. M., Moretti, S., Tzivanopoulou, M., Wollbrett, J., & Robinson-Rechavi, M. (2025). Bgee in 2024: focus on curated single-cell RNA-seq datasets, and query tools. *Nucleic Acids Research*, *53*(D1), D878–D885. https://doi.org/10.1093/nar/gkae1118

Bucknor, M. C., Gururajan, A., Dale, R. C., & Hofer, M. J. (2022). A comprehensive approach to modeling maternal immune activation in rodents. In *Frontiers in Neuroscience* (Vol. 16). Frontiers Media S.A. https://doi.org/10.3389/fnins.2022.1071976

Datta, D., Leslie, S. N., Morozov, Y. M., Duque, A., Rakic, P., Van Dyck, C. H., Nairn, A. C., & Arnsten, A. F. T. (2020). Classical complement cascade initiating C1q protein within neurons in the aged rhesus macaque dorsolateral prefrontal cortex. *Journal of Neuroinflammation*, *17*(1). https://doi.org/10.1186/s12974-019-1683-1

Duff, M. O., Olson, S., Wei, X., Garrett, S. C., Osman, A., Bolisetty, M., Plocik, A., Celniker, S. E., & Graveley, B. R. (2015). Genome-wide identification of zero nucleotide recursive splicing in Drosophila. *Nature*, *521*(7552), 376–379. https://doi.org/10.1038/nature14475

Estes, M. L., & McAllister, A. K. (2015). Immune mediators in the brain and peripheral tissues in autism spectrum disorder. In *Nature Reviews Neuroscience* (Vol. 16, Issue 8, pp. 469–486). Nature Publishing Group. https://doi.org/10.1038/nrn3978

Estes, M. L., & McAllister, A. K. (2016). Maternal immune activation: Implications for neuropsychiatric disorders. In *Science* (Vol. 353, Issue 6301, pp. 772–777). American Association for the Advancement of Science. https://doi.org/10.1126/science.aag3194

Fatemi, S. H., Araghi-Niknam, M., Laurence, J. A., Stary, J. M., Lee, S., & Sidwell, R. W. (2004). Glial fibrillary acidic protein and glutamic acid decarboxylase 65 and 67 kDa proteins are increased in brains of neonatal BALB/c mice following viral infection in

utero [2]. In *Schizophrenia Research* (Vol. 69, Issue 1, pp. 121–123). Elsevier. https://doi.org/10.1016/S0920-9964(03)00175-0

Flierman, R., & Daha, M. R. (2007). The clearance of apoptotic cells by complement. *Immunobiology*, *212*(4–5), 363–370. https://doi.org/10.1016/j.imbio.2006.11.005

Goines, P. E., Croen, L. A., Braunschweig, D., Yoshida, C. K., Grether, J., Hansen, R., Kharrazi, M., Ashwood, P., & Van De Water, J. (2011). Increased midgestational IFN-γ, IL-4 and IL-5 in women bearing a child with autism: A case-control study. *Molecular Autism*, *2*(1). https://doi.org/10.1186/2040-2392-2-13

Graham, A. M., Rasmussen, J. M., Rudolph, M. D., Heim, C. M., Gilmore, J. H., Styner, M., Potkin, S. G., Entringer, S., Wadhwa, P. D., Fair, D. A., & Buss, C. (2018). Maternal Systemic Interleukin-6 During Pregnancy Is Associated With Newborn Amygdala Phenotypes and Subsequent Behavior at 2 Years of Age. *Biological Psychiatry*, *83*(2), 109–119. https://doi.org/10.1016/j.biopsych.2017.05.027

Guma, E., Bordignon, P. do C., Devenyi, G. A., Gallino, D., Anastassiadis, C., Cvetkovska, V., Barry, A. D., Snook, E., Germann, J., Greenwood, C. M. T., Misic, B., Bagot, R. C., & Chakravarty, M. M. (2021). Early or Late Gestational Exposure to Maternal Immune Activation Alters Neurodevelopmental Trajectories in Mice: An Integrated Neuroimaging, Behavioral, and Transcriptional Study. *Biological Psychiatry*, *90*(5), 328–341. https://doi.org/10.1016/j.biopsych.2021.03.017

Harvey, L., & Boksa, P. (2012). Prenatal and postnatal animal models of immune activation: Relevance to a range of neurodevelopmental disorders. *Developmental Neurobiology*, *72*(10), 1335–1348. https://doi.org/10.1002/dneu.22043

Hoffmann, T. J., Ehret, G. B., Nandakumar, P., Ranatunga, D., Schaefer, C., Kwok, P. Y., Iribarren, C., Chakravarti, A., & Risch, N. (2017). Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nature Genetics*, *49*(1), 54–64. https://doi.org/10.1038/ng.3715

Hourcade, D., Miesner, D. R., Bee, C., Zeldes, W., & Atkinson, J. P. (1990). Duplication and divergence of the amino-terminal coding region of the complement receptor 1 (CR1) gene: An example of concerted (horizontal) evolution within a gene. *Journal of Biological Chemistry*, *265*(2), 974–980. https://doi.org/10.1016/s0021-9258(19)40145-2

Huttlin, E. L., Bruckner, R. J., Navarrete-Perea, J., Cannon, J. R., Baltier, K., Gebreab, F., Gygi, M. P., Thornock, A., Zarraga, G., Tam, S., Szpyt, J., Gassaway, B. M., Panov, A., Parzen, H., Fu, S., Golbazi, A., Maenpaa, E., Stricker, K., Guha Thakurta, S., … Gygi, S. P. (2021). Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell*, *184*(11), 3022-3040.e28. https://doi.org/10.1016/j.cell.2021.04.011

Jacob, A., & Alexander, J. J. (2014). Complement and blood-brain barrier integrity. In *Molecular Immunology* (Vol. 61, Issue 2, pp. 149–152). Elsevier Ltd. https://doi.org/10.1016/j.molimm.2014.06.039

Jimenez-Duran, G., Kozole, J., Peltier-Heap, R., Dickinson, E. R., Kwiatkowski, C. R., Zappacosta, F., Annan, R. S., Galwey, N. W., Nichols, E. M., Modis, L. K., Triantafilou, M., Triantafilou, K., & Booty, L. M. (2022). Complement membrane attack complex is an immunometabolic regulator of NLRP3 activation and IL-18 secretion in human macrophages. *Frontiers in Immunology*, *13*. https://doi.org/10.3389/fimmu.2022.918551

Khakzad, H., Happonen, L., Tran Van Nhieu, G., Malmström, J., & Malmström, L. (2021). In vivo Cross-Linking MS of the Complement System MAC Assembled on Live Gram-Positive Bacteria. *Frontiers in Genetics*, *11*. https://doi.org/10.3389/fgene.2020.612475

Kim, J. H., Lee, S., Rho, J. K., & Choe, S. Y. (n.d.). *AML1, the target of chromosomal rearrangements in human leukemia, regulates the expression of human complement receptor type 1 (CR1) gene*. www.elsevier.com/locate/ijbcb

Kim, M., Haney, J. R., Zhang, P., Hernandez, L. M., Wang, L. kai, Perez-Cano, L., Loohuis, L. M. O., de la Torre-Ubieta, L., & Gandal, M. J. (2021). Brain gene co-expression networks link complement signaling with convergent synaptic pathology in schizophrenia. *Nature Neuroscience*, *24*(6), 799–809. https://doi.org/10.1038/s41593-021-00847-z

Kneeland, R. E., & Fatemi, S. H. (2013). Viral infection, inflammation and schizophrenia. In *Progress in Neuro-Psychopharmacology and Biological Psychiatry* (Vol. 42, pp. 35–48). https://doi.org/10.1016/j.pnpbp.2012.02.001

Knuesel, I., Chicha, L., Britschgi, M., Schobel, S. A., Bodmer, M., Hellings, J. A., Toovey, S., & Prinssen, E. P. (2014). Maternal immune activation and abnormal brain

development across CNS disorders. In *Nature Reviews Neurology* (Vol. 10, Issue 11, pp. 643–660). Nature Publishing Group. https://doi.org/10.1038/nrneurol.2014.187

Krakowiak, P., Goines, P. E., Tancredi, D. J., Ashwood, P., Hansen, R. L., Hertz-Picciotto, I., & Van de Water, J. (2017). Neonatal Cytokine Profiles Associated With Autism Spectrum Disorder. *Biological Psychiatry*, *81*(5), 442–451. https://doi.org/10.1016/j.biopsych.2015.08.007

Kremlitzka, M., Nowacka, A. A., Mohlin, F. C., Bompada, P., De Marinis, Y., & Blom, A. M. (2019). Interaction of serum-derived and internalized C3 with DNA in human B cells-a potential involvement in regulation of gene transcription. *Frontiers in Immunology*, *10*(MAR). https://doi.org/10.3389/fimmu.2019.00493

Kullo, I. J., Ding, K., Shameer, K., McCarty, C. A., Jarvik, G. P., Denny, J. C., Ritchie, M. D., Ye, Z., Crosslin, D. R., Chisholm, R. L., Manolio, T. A., & Chute, C. G. (2011). Complement receptor 1 gene variants are associated with erythrocyte sedimentation rate. *American Journal of Human Genetics*, *89*(1), 131–138. https://doi.org/10.1016/j.ajhg.2011.05.019

Kunz, N., & Kemper, C. (2021). Complement Has Brains—Do Intracellular Complement and Immunometabolism Cooperate in Tissue Homeostasis and Behavior? In *Frontiers in Immunology* (Vol. 12). Frontiers Media S.A. https://doi.org/10.3389/fimmu.2021.629986

Kushima, I., Aleksic, B., Nakatochi, M., Shimamura, T., Okada, T., Uno, Y., Morikawa, M., Ishizuka, K., Shiino, T., Kimura, H., Arioka, Y., Yoshimi, A., Takasaki, Y., Yu, Y., Nakamura, Y., Yamamoto, M., Iidaka, T., Iritani, S., Inada, T., … Ozaki, N. (2018). Comparative Analyses of Copy-Number Variation in Autism Spectrum Disorder and Schizophrenia Reveal Etiological Overlap and Biological Insights. *Cell Reports*, *24*(11), 2838–2856. https://doi.org/10.1016/j.celrep.2018.08.022

Le Friec, G., Sheppard, D., Whiteman, P., Karsten, C. M., Shamoun, S. A. T., Laing, A., Bugeon, L., Dallman, M. J., Melchionna, T., Chillakuri, C., Smith, R. A., Drouet, C., Couzi, L., Fremeaux-Bacchi, V., Köhl, J., Waddington, S. N., McDonnell, J. M., Baker, A., Handford, P. A., … Kemper, C. (2012). The CD46-Jagged1 interaction is critical for human T H 1 immunity. *Nature Immunology*, *13*(12), 1213–1221. https://doi.org/10.1038/ni.2454

Li, Q., Wineinger, N. E., Fu, D. J., Libiger, O., Alphs, L., Savitz, A., Gopal, S., Cohen, N., & Schork, N. J. (2017). Genome-wide association study of paliperidone efficacy. *Pharmacogenetics and Genomics*, *27*(1), 7–18. https://doi.org/10.1097/FPC.0000000000000250

Lietzén, N., Cheng, L., Moulder, R., Siljander, H., Laajala, E., Härkönen, T., Peet, A., Vehtari, A., Tillmann, V., Knip, M., Lähdesmäki, H., & Lahesmaa, R. (2018). Characterization and non-parametric modeling of the developing serum proteome during infancy and early childhood. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-24019-5

Logar, C. M., Chen, W., Schmitt, H., Yu, C. Y., & Birmingham, D. J. (2004). A human CR1-like transcript containing sequence for a binding protein for iC4 is expressed in hematopoietic and fetal lymphoid tissue. *Molecular Immunology*, *40*(11), 831–840. https://doi.org/10.1016/j.molimm.2003.09.010

Luck, K., Kim, D. K., Lambourne, L., Spirohn, K., Begg, B. E., Bian, W., Brignall, R., Cafarelli, T., Campos-Laborie, F. J., Charloteaux, B., Choi, D., Coté, A. G., Daley, M., Deimling, S., Desbuleux, A., Dricot, A., Gebbia, M., Hardy, M. F., Kishore, N., … Calderwood, M. A. (2020). A reference map of the human binary protein interactome. *Nature*, *580*(7803), 402–408. https://doi.org/10.1038/s41586-020-2188-x

Makinson, R., Lloyd, K., Rayasam, A., McKee, S., Brown, A., Barila, G., Grissom, N., George, R., Marini, M., Fabry, Z., Elovitz, M., & Reyes, T. M. (2017). Intrauterine inflammation induces sex-specific effects on neuroinflammation, white matter, and behavior. *Brain, Behavior, and Immunity*, *66*, 277–288. https://doi.org/10.1016/j.bbi.2017.07.016

Massarali, A., Adhya, D., Srivastava, D. P., Baron-Cohen, S., & Kotter, M. R. (2022). Virus-Induced Maternal Immune Activation as an Environmental Factor in the Etiology of Autism and Schizophrenia. In *Frontiers in Neuroscience* (Vol. 16). Frontiers Media S.A. https://doi.org/10.3389/fnins.2022.834058

Meehan, C., Harms, L., Frost, J. D., Barreto, R., Todd, J., Schall, U., Shannon Weickert, C., Zavitsanou, K., Michie, P. T., & Hodgson, D. M. (2017). Effects of immune activation during early or late gestation on schizophrenia-related behaviour in adult rat offspring. *Brain, Behavior, and Immunity*, *63*, 8–20. https://doi.org/10.1016/j.bbi.2016.07.144

Mevorach, D., Mascarenhas, J. O., Gershov, D., & Elkon, K. B. (1998). Complement-dependent clearance of apoptotic cells by human macrophages. *Journal of Experimental Medicine*, *188*(12), 2313–2320. https://doi.org/10.1084/jem.188.12.2313

Meyer, U., Feldon, J., & Yee, B. K. (2009). A review of the fetal brain cytokine imbalance hypothesis of schizophrenia. In *Schizophrenia Bulletin* (Vol. 35, Issue 5, pp. 959–972). https://doi.org/10.1093/schbul/sbn022

Ogden, C. A., Decathelineau, A., Hoffmann, P. R., Bratton, D., Ghebrehiwet, B., Fadok, V. A., & Henson, P. M. (2001). The Rockefeller University Press • 0022-1007. In *J. Exp. Med* (Vol. 194, Issue 6). http://www.jem.org/cgi/content/full/194/6/781

Pagliaro, A., Finger, R., Zoutendijk, I., Bunschuh, S., Clevers, H., Hendriks, D., & Artegiani, B. (2023). Temporal morphogen gradient-driven neural induction shapes single expanded neuroepithelium brain organoids with enhanced cortical identity. *Nature Communications*, *14*(1). https://doi.org/10.1038/s41467-023-43141-1

Rahman, T., Weickert, C. S., Harms, L., Meehan, C., Schall, U., Todd, J., Hodgson, D. M., Michie, P. T., & Purves-Tyson, T. (2020). Effect of Immune Activation during Early Gestation or Late Gestation on Inhibitory Markers in Adult Male Rats. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-58449-x

Rangaraju, S., Raza, S. A., Li, N. X. A., Betarbet, R., Dammer, E. B., Duong, D., Lah, J. J., Seyfried, N. T., & Levey, A. I. (2018). Differential phagocytic properties of CD45low microglia and CD45high brain mononuclear phagocytes-activation and age-related effects. *Frontiers in Immunology*, *9*(MAR). https://doi.org/10.3389/fimmu.2018.00405

Ripke, S., Neale, B. M., Corvin, A., Walters, J. T. R., Farh, K. H., Holmans, P. A., Lee, P., Bulik-Sullivan, B., Collier, D. A., Huang, H., Pers, T. H., Agartz, I., Agerbo, E., Albus, M., Alexander, M., Amin, F., Bacanu, S. A., Begemann, M., Belliveau, R. A., … O'Donovan, M. C. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, *511*(7510), 421–427. https://doi.org/10.1038/nature13595

Rodriguez, E., Guevara, J., Paez, A., Zapata, E., Collados, M. T., Fortoul, T. I., Lopez-Marure, R., Masso, F., & Montaño, L. F. (2008). The altered expression of inflammation-related molecules and secretion of IL-6 and IL-8 by HUVEC from newborns with maternal inactive systemic lupus erythematosus is modified by estrogens. *Lupus*, *17*(12), 1086–1095. https://doi.org/10.1177/0961203308093827

Sansom, S. N., & Livesey, F. J. (2009). Gradients in the brain: the control of the development of form and function in the cerebral cortex. In *Cold Spring Harbor perspectives in biology* (Vol. 1, Issue 2). https://doi.org/10.1101/cshperspect.a002519

Serna, M., Giles, J. L., Morgan, B. P., & Bubeck, D. (2016). Structural basis of complement membrane attack complex formation. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms10587

Sheng, J. A., & Tobet, S. A. (2024). Maternal immune activation with toll-like receptor 7 agonist during mid-gestation alters juvenile and adult developmental milestones and behavior. *Journal of Neuroendocrinology*, *36*(8). https://doi.org/10.1111/jne.13417

Smith, S. E. P., Li, J., Garbett, K., Mirnics, K., & Patterson, P. H. (2007). Maternal immune activation alters fetal brain development through interleukin-6. *Journal of Neuroscience*, *27*(40), 10695–10702. https://doi.org/10.1523/JNEUROSCI.2178-07.2007

Sontheimer, R. D., Racila, E., & Racila, D. M. (2005). *C1q: Its Functions within the Innate and Adaptive Immune Responses and its Role in Lupus Autoimmunity*.

Szabo, L., Morey, R., Palpant, N. J., Wang, P. L., Afari, N., Jiang, C., Parast, M. M., Murry, C. E., Laurent, L. C., & Salzman, J. (2015). Statistically based splicing detection reveals neural enrichment and tissue-specific induction of circular RNA during human fetal development. *Genome Biology*, *16*(1). https://doi.org/10.1186/s13059-015-0690-5

Ten, V. S., Yao, J., Ratner, V., Sosunov, S., Fraser, D. A., Botto, M., Sivasankar, B., Morgan, B. P., Silverstein, S., Stark, R., Polin, R., Vannucci, S. J., Pinsky, D., & Starkov, A. A. (2010). Complement component C1q mediates mitochondria-driven oxidative stress in neonatal hypoxic-ischemic brain injury. *Journal of Neuroscience*, *30*(6), 2077–2087. https://doi.org/10.1523/JNEUROSCI.5249-09.2010

Tiberi, L., Vanderhaeghen, P., & van den Ameele, J. (2012). Cortical neurogenesis and morphogens: Diversity of cues, sources and functions. In *Current Opinion in Cell Biology* (Vol. 24, Issue 2, pp. 269–276). https://doi.org/10.1016/j.ceb.2012.01.010

Török, K., Dezso, B., Bencsik, A., Uzonyi, B., & Erdei, A. (2015). Complement receptor type 1 (CR1/CD35) expressed on activated human CD4+ T cells contributes to generation of regulatory T cells. *Immunology Letters*, *164*(2), 117–124. https://doi.org/10.1016/j.imlet.2015.02.009

Walling, B. L., & Kim, M. (2018). LFA-1 in T cell migration and differentiation. In *Frontiers in Immunology* (Vol. 9, Issue MAY). Frontiers Media S.A. https://doi.org/10.3389/fimmu.2018.00952

Warwick, C. A., Keyes, A. L., Woodruff, T. M., & Usachev, Y. M. (2021). The complement cascade in the regulation of neuroinflammation, nociceptive sensitization, and pain. In *Journal of Biological Chemistry* (Vol. 297, Issue 3). American Society for Biochemistry and Molecular Biology Inc. https://doi.org/10.1016/J.JBC.2021.101085

West, E. E., & Kemper, C. (2023). Complosome — the intracellular complement system. In *Nature Reviews Nephrology* (Vol. 19, Issue 7, pp. 426–439). Nature Research. https://doi.org/10.1038/s41581-023-00704-1

Woods, R. M., Lorusso, J. M., Potter, H. G., Neill, J. C., Glazier, J. D., & Hager, R. (2021). Maternal immune activation in rodent models: A systematic review of neurodevelopmental changes in gene expression and epigenetic modulation in the offspring brain. In *Neuroscience and Biobehavioral Reviews* (Vol. 129, pp. 389–421). Elsevier Ltd. https://doi.org/10.1016/j.neubiorev.2021.07.015

Xu, N., Li, X., & Zhong, Y. (2015). Inflammatory cytokines: Potential biomarkers of immunologic dysfunction in autism spectrum disorders. In *Mediators of Inflammation* (Vol. 2015). Hindawi Publishing Corporation. https://doi.org/10.1155/2015/531518

Yan, S., Wang, L., Samsom, J. N., Ujic, D., & Liu, F. (2024). PolyI:C Maternal Immune Activation on E9.5 Causes the Deregulation of Microglia and the Complement System in Mice, Leading to Decreased Synaptic Spine Density. *International Journal of Molecular Sciences*, *25*(10). https://doi.org/10.3390/ijms25105480

Yates, B., Gray, K. A., Jones, T. E. M., & Bruford, E. A. (2021). Updates to HCOP: The HGNC comparison of orthology predictions tool. *Briefings in Bioinformatics*, *22*(6). https://doi.org/10.1093/bib/bbab155

Yilmaz, M., Yalcin, E., Presumey, J., Aw, E., Ma, M., Whelan, C. W., Stevens, B., McCarroll, S. A., & Carroll, M. C. (2021). Overexpression of schizophrenia susceptibility factor human complement C4A promotes excessive synaptic loss and behavioral changes in mice. *Nature Neuroscience*, *24*(2), 214–224. https://doi.org/10.1038/s41593-020-00763-8

Zheng, Y., Collins, S. L., Lutz, M. A., Allen, A. N., Kole, T. P., Zarek, P. E., & Powell, J. D. (2007). A Role for Mammalian Target of Rapamycin in Regulating T Cell Activation versus Anergy 1. In *The Journal of Immunology* (Vol. 178). www.jimmunol.org

Zipfel, P. F., & Skerka, C. (2009). Complement regulators and inhibitory proteins. In *Nature Reviews Immunology* (Vol. 9, Issue 10, pp. 729–740). https://doi.org/10.1038/nri2620