Investigating the Association between Maternal COVID-19 Infection and Modifications in the Functionality of the Immune System in Newborns

by

Solmaz Ghajar

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Epidemiology

School of Public Health

University of Alberta

Abstract:

Background: Maternal COVID-19 infection during pregnancy has raised prominent public health concerns - especially the possibility of long-term child health outcomes. Epigenetic changes induced by inflammatory environment in the placenta caused by anti-infection response in pregnant mother's body, can change the gene expression patterns of fetus cells which are crucial for healthy development of the offspring. Gene expression analysis methods can be used to assess many health outcomes, including immune response, metabolic pathways, and the development of chronic disease, into the childhood years and beyond. This study aimed to evaluate the effects of a maternal COVID-19 infection on gene expression alterations of metabolic pathways in the fetus and the newborn, with a focus on innate immune system functions.

Methods: We tested gene expression data derived from umbilical cord blood cells from infants born to mothers who had a COVID-19 infection during pregnancy (n = 8) and infants whose mothers did not have COVID-19 (n = 8). The gene expression microarray dataset (GSE195938) was previously published by Jefferson et al. (2022), and the dataset and corresponding expression for 186 gene sets were sorted by the KEGG LEGACY catalog which represents canonical biological pathways in the human body. To identify gene sets impacting continuous phenotypes that capture innate immune functions, we applied the Linear Combination Test (LCT), a gene set analysis method. LCT captures the features inherent in the data to isolate gene sets based on intrinsic expression patterns.

Results: Significant differences in patterns of gene set expression in infants born to COVID-19 mothers were noted. The analysis also resulted in identifying gene sets taking part in aminoacyl tRNA biosynthesis, asthma, and systemic lupus erythematosus which demonstrated relations to

inherent pattern expressions relating to innate immune functions phenotypes. The present study suggests that maternal COVID-19 infection may impact fetal immune development, and increase the risk of immune-mediated outcomes such as asthma. A statistically significant enrichment was also associated with gene sets responsible for amino acid and lipid metabolism and their relationship with immune signaling pathways. This finding suggests an association between maternal COVID-19 rounded to metabolic pathways associated with amino acid metabolism and pathways in folate synthesis, which are relevant to fetal growth and development.

Conclusion: In closing, maternal COVID-19 infection during pregnancy can lead to epigenetic changes in the fetus, which may have future effects on immune function and metabolic pathways. The gene sets and pathways presented here need further work to understand implications for offspring health. This analysis provides support for more studies on epigenetics as a mechanism for understanding how maternal infection may affect offspring development and health. It also underscores the importance of more prenatal care and addresses potential targets for prevention, in the interest of minimizing the future risk of autoimmune and allergic diseases, for infants born to COVID-19-infected-mothers.

Preface:

This thesis reflects Solmaz Ghajar's original research conducted under the supervision of Dr. Irina Dinu. I undertook the responsibility of the preliminary research study development, the selection of pertinent public datasets, the design of both figures and tables, conducting and authoring a literature review, and the lead author on conference abstracts and presentations. Professor Irina Dinu supported in all aspects of the project: study development and manuscript writing. The analysis in Chapter 3 was completed with Nastaran Hajizadeh Bastani and Sara Khademioureh and also the quantile normalization and Innate Immune System phenotypes calculations were done by Dr. Sumanta Ray. Professor Saumyadipta Pyne was also a supervisor on the project, participated in the study development, contributed to the analysis and interpretation of findings, and offered essential feedback and played an influential role in the manuscript writing. Figure 1.1 and Figure 1.2 were extracted from World Health Organization website. Figure 1.3 and Figure 1.4 were extracted from Shimizu et al., 2023. Table 3.1 was extracted from Gayen nee'Betal et al.,2022, the main paper we used the data for our research project.

Portions of this research have been presented as follows:

Ghajar, Solmaz. Hajizadeh Bastani, Nastaran. Khademioureh, Sara. Nezarat, Fatemeh. Pyne, Saumyadipta. Dinu, Irina (2024). Maternal COVID-19 infection During Pregnancy impacts Gene Expression in newborns. ***The Twelfth Annual Canadian Statistics Student Conference***. Poster presentation.

Ghajar, Solmaz. Hajizadeh Bastani, Nastaran. Khademioureh, Sara. Nezarat, Fatemeh. Saumyadipta Pyne. Dinu, Irina. (2023). COVID-19 During Pregnancy Alters Umbilical Cord Blood Gene Expression. ***Women and Children's Health Research Institute (WCHRI) Research Day***. Ten Minute Oral Presentation.

Acknowledgement:

I have received support from many people over the past two years. First, I want to thank my great supervisor, Prof. Irina Dinu, for all her support, encouragement, patience, and guidance throughout my MSc program. I am grateful for her advice and feedback in conducting this research and also help in establishing my future career. It has been an honor and an exceptional opportunity for me to work with Prof. Irina Dinu.

I also extend my heartfelt gratitude to my supervisory committee member, Prof. Saumyadipta Pyne, for his invaluable contributions to my MSc project. His insights, expertise, and constructive feedback and review comments have been a significant source of inspiration. I express my sincere gratitude to Dr. Morteza Hajihosseini. for his support and for contributing to the advancement of my academic journey.

I want to thank Nastaran Bastani for her sincere support and expertise during the analysis phase of the project. I am especially grateful for the time and effort they took to share their knowledge and for their helpful, positive feedback that enhanced my work.

The School of Public Health has provided an exceptional environment for graduate students to exceed their boundaries. I extend my gratitude to all members of the graduate office, as well as my colleagues, past and present faculty, and course instructors, including: Dr. Don VoaklanderDr. Roman Pabayo Pabayo.

I am thankful to the University of Alberta for significantly influencing my career path and providing a nurturing environment that I have come to regard as my second home during the past two years.

And finally, I want to thank my husband, parents and siblings. They have consistently supported me in every aspect of life, including my education, and being there for me with unconditional love.

Contents

List of Tables

List of Figures

Chapter 1: Literature Review

1.1 Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and Global pandemic:

The 2019 Coronavirus Disease (COVID-19), following the influenza pandemic of 1918, has emerged as the fifth global pandemic, posing a significant challenge to healthcare systems worldwide (Liu et al., 2020). The first case was detected in late December 2019 in Wuhan, Hubei Province, China. It swiftly spread to other countries and by late April 2020, 185 countries reported cases, affecting over 300,000 individuals. Moreover, the death toll exceeded 210,000. Therefore, The World Health Organization (WHO) declared COVID-19 a global pandemic (Clerkin et al., 2020). Since then, the number of infected individuals has surged exponentially. According to the last statistics, over 775 million cases have been reported globally to date (Figure 1.1)—this has resulted in 7 million fatalities until today (Figure 1.2) (*COVID-19 Cases | WHO COVID-19 Dashboard*, 2024.).

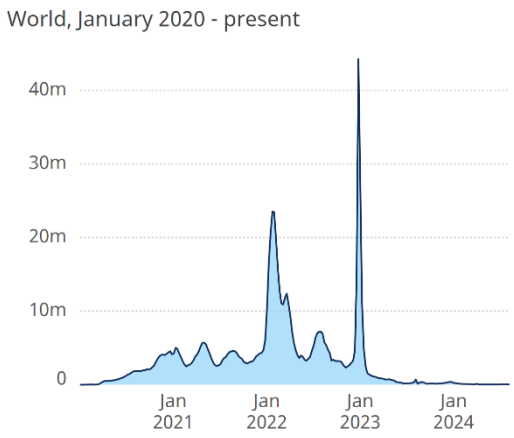

Figure 1.1 Total COVID-19 cases reported to WHO

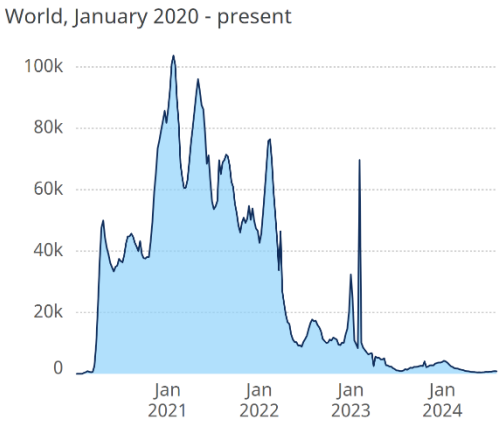

Figure 1.2 Total COVID-19 deaths reported to WHO

Source: World Health Organization

1.2 Features and pathogenesis of the SARS-CoV-2 virus:

The virus in question belongs to the Coronaviridae family, making it the seventh member known to cause disease in humans. This virus has a long, single-stranded, positive-sense RNA molecule within a lipid envelope. Although the precise origin of this virus remains undetermined, its significant genetic similarity with bat coronaviruses suggests a strong possibility that it

originated from bats (Chan et al., 2020). The virus enters the host body via respiratory droplets and traverses through the respiratory tract. It penetrates ciliated epithelial cells through the angiotensin-converting enzyme 2 (ACE2) and CD26 receptors on cell surfaces using Spike glycoprotein, which is crucial for the virus's pathogenicity. ACE2 has been identified as a cell entry receptor for SARS-CoV-2, underscoring its multifaceted role in normal physiology and viral pathogenesis (Lan et al., 2020). It is important to note that the catalytic activities of ACE2 are distinct from its role as a cellular entry receptor for viruses like SARS-CoV-2. Inhibitors targeting the catalytic site of ACE2 do not impact the virus binding process, illustrating the complex relationship between receptor activity and viral entry pathways (W. Li et al., 2005). The SARS-CoV-2 genome consists of fourteen open reading frames with approximately two-thirds of the genome dedicated to encoding sixteen non-structural proteins (nsp1–16), a central resource for the replicase composite. The remaining one-third of the genome encodes nine ancillary proteins (ORF) and four structural proteins: the nucleocapsid (N) protein, membrane (M) protein, envelope (E) protein, and spike (S) protein (Y.-Z. Zhang & Holmes, 2020). The entry of viral particles into host cells is initiated by the class I viral fusion S glycoprotein that binds to the cell membrane and then mediates fusion. The S protein is trimeric, integrated into the virion membrane in numerous copies, and imparts a distinct "corona" appearance to the virion. In virus-producing cells, the S protein of SARS-CoV-2 is cleaved by proprotein convertases, resulting in the mature virion's S protein being composed of two noncovalently associated subunits, one of which is the S1 subunit that contains the ACE2 binding domain and S2 subunit that is responsible for anchoring the S protein in the membrane. The Spike protein S1 subunit comprises four domains, the amino-terminal domain (NTD), the receptor-binding domain (RBD), and two carboxy-terminal domains (CTD1 and CTD2). Mutations in SARS-CoV-2 Spike protein's RBD have been highlighted as crucial contributions to increased interactions with ACE2, boosting virus receptor affinity, and possibly increasing the virus's infectivity compared to SARS-CoV. The three receptor-binding domains (RBDs) at the apex of the S protein trimer can assume one of two conformations: an "up" conformation that promotes access to the receptor and a "down" conformation that does not allow receptor interactions (Zhou et al., 2020). S2 adopts a dramatic change in protein design during host cell infection to promote membrane fusion and prove functional for virus infection (Fehr & Perlman, 2015). Once the virus contacts ACE2, conferring a SARS-CoV-2 host of a new cell, the SARS-CoV-2 S2 site, S2', is exposed,

and subsequent cleavage at this site, either by the host serine protease TMPRSS2 on the cell surface or by cathepsin L following ACE2-mediated endocytosis, results in fusion peptide release and formation of a fusion pore (Jackson et al., 2022). In the viral entry, TMPRSS2 is important; SARS-CoV can enter cells with low ACE2 levels if TMPRSS2 is available (Shulla et al., 2011). Upon entering the host cytosol, the viral genome leads to the translation of viral replicase proteins through ORF1a and ORF1b. These replicase proteins are then cleaved into individual non-structural proteins (nsps) by host and viral proteases. The replicase proteins reorganize the endoplasmic reticulum (ER) to produce double-membrane vesicles (DMVs), which serve as platforms for viral replication, leading to the synthesis of both genomic and sub-genomic RNAs (sgRNA). After replication is complete, the sgRNA is translated into accessory proteins and viral structural proteins, necessary for the generation of new viral particles (Harrison et al., 2020).

After introducing viral genetic material into host cells, initiating replication and producing primary proteins, this triggers an initial immune response. In 80% of cases, this results in a mild response manageable by symptomatic therapy. However, in about 20% of cases, it induces a severe immune reaction. The virus subsequently targets alveolar cells in the lungs, causing their destruction, which leads to inflammatory reactions accompanied by cytokine storms and pulmonary fibrosis (Wu & McGoogan, 2020; Zahedipour et al., 2020).

Beyond respiratory complications, there is potential for the virus to enter the bloodstream and infect tissues expressing ACE receptors. These include kidneys, and gut tissue, heart and cardiovascular system, central nervous system as well as adipose tissue and even placenta and fetal (Letko et al., 2020; Lin et al., 2020; Prochaska et al., 2020).

1.3 The symptoms of COVID-19:

The 2019 Coronavirus Disease presents with symptoms like fevers, cough, difficulty breathing, muscle aches, sore throat, diarrhea, pneumonia, acute respiratory syndrome, multiple organ failure, cytokine syndrome, which is quite serious, endothelial damage and thrombotic events (Prochaska et al., 2020). The severity of this disease can range from no symptoms in which the patients might not even realize they're carrying the virus at all to very severe cases. Often, the more severe cases are linked with older age or chronic conditions. These conditions include heart

and cardiovascular diseases, respiratory issues, kidney diseases, diabetes, and smoking (Guan et al., 2020; Jin et al., 2020).

As the disease worsens, more critical symptoms such as acute respiratory distress syndrome, a critical lung condition characterized by the accumulation of fluid in the air sacs, which reduces the oxygen levels in the blood, may develop. ARDS is often the main reason for death due to respiratory failure. Furthermore, COVID-19 doesn't just attack the respiratory system. It can lead to non-respiratory complications such as septic shock, liver and kidney injury, ocular complications, neurological issues & disseminated intravascular coagulation (Mallah et al., 2021). Additionally, and quite notably, COVID-19 has been seen to significantly impair an individual's sense of smell and taste in many cases (Gane et al., 2020).

Generally, this disease shows up with milder symptoms in children and adolescents. Yet, certain strains of the virus, like the Delta variant, can cause more severe symptoms. These children with severe COVID-19 signs show severe inflammatory responses. These are often accompanied by high fever and problems in multiple organs. Despite the low risk, there's still a possibility of death. The CDC calls this syndrome Multisystem Inflammatory Syndrome in Children, a severe condition where the immune system excessively responds to COVID-19, leading to inflammation in several organs. Criteria for MIS-C include prolonged fever, issues in multiple organ systems and lab findings of hyperinflammation (Chen et al., 2020; Siegel et al., 2021).

1.4 Transmission routes of the disease:

COVID-19, like other viruses in the coronavirus family, spreads chiefly from person to person via droplet transmission. These droplets reach others through sneezing, coughing, or even talking and can either be inhaled by others or make contact with mucous membranes in the mouth, nasal passages, or eyes. Furthermore, touching surfaces contaminated with the virus—surfaces that can stay active and detectable for about 6 days—and then touching one’s own mucous membranes may result in infection (Mallah et al., 2021; van Doremalen Neeltje et al., 2020).

A particularly worrying aspect is the transmission of COVID-19 by individuals who do not exhibit any symptoms of the illness. With a relatively long incubation period—approximately 5

days on average—infected persons can unknowingly spread the virus to others without being aware of their carrier status (McAloon et al., 2020).

1.5 The Covid-19 disease in pregnancy:

Pregnancy is considered as a critical stage characterized by significant immune and physiological changes to house a genetically foreign fetus. Physiological changes include decreased lung capacity, increased oxygen demand, and changes in heart rate. Previous studies have shown that changes in the immune system during pregnancy, promoting a shift towards a dominance of the T-helper 2 (Th2) system, which is aimed to protect the developing fetus, can result in the mother becoming more susceptible to viral infections. This is due to the fact that viral infections are typically better controlled by the Th1 system, thus exposing a vulnerability in pregnant individuals. Pregnancy increases the risk of influenza infection, in particular among pregnant women with underlying medical conditions and during the third trimester of gestation (Rasmussen et al., 2012).

According to a systematic review that includes 435 studies, pregnant women who were infected with SARS-CoV-2 had a greater risk for severe disease as indexed by higher rates of hospitalization or ICU admission compared to the general population. Cox proportional hazards analysis indicated that comorbidities, non-white ethnicity, chronic hypertension, pre-existing diabetes, age over 35 years, and pre-pregnancy BMI >30 were associated with severe COVID-19 outcomes in pregnancy. Pregnant mothers with COVID-19 are at higher risk of preterm birth and maternal mortality compared with pregnant women without COVID-19. The infants of women with COVID-19 are more likely to require admission to the neonatal unit (Allotey et al., 2020).

A recent review of obstetric cases—especially those occurring in the third trimester—revealed that 3% of pregnant women with SARS-CoV-2 required hospitalization and intensive care. There have also been instances of preterm labor as well as perinatal death linked to maternal SARS-CoV-2 infection (Prochaska et al., 2020).

1.6 Vertical transmission of disease from pregnant mother to her fetus:

The transmission of disease and viruses from a pregnant mother to her fetus (Called vertical transmission) remains the subject of ongoing debate. Some studies show that specific IgM antibodies for SARS-CoV-2 have not been detected in fetuses. Nasopharyngeal swab test results are also negative for infants born to infected mothers. This indicates no virus transmission has occurred. Furthermore, the presence of specific IgG antibodies (linked to SARS-CoV-2) in newborns suggests protective immunity transfer from mother to fetus. It is plausible that these factors are transferred from the mother's body to the fetus via the placenta which serves as a critical maternal-fetal interface; it protects against infection but is vulnerable when exposed to adverse maternal conditions. In a systematic review involving over 50,000 pregnancies, which resulted in the birth of 30,000 neonates from mothers infected with SARS-CoV-2, over 800 neonates tested positive for SARS-CoV-2 in confirmed samples. Therefore, this evidences the possibility of the vertical transmission of SARS-CoV-2 from mothers with COVID-19. Vertical transmission can occur via multiple routes, including the exchange of microorganisms in the placenta during pregnancy and exposure to blood and vaginal secretions during labor. The likelihood of Vertical transmission will largely depend on several factors such as maternal immunity and viral load in the placenta, which have been strongly associated with increased risk for vertical transmission (Musa et al., 2021). Consequently, it is not possible to eliminate entirely the possibility of disease transmission from mother to newborn. Further studies are necessary, although current statistics indicate that such transmission is rare (Chen et al., 2020; Patanè et al., 2020).

Indeed, while direct transmission through the placenta may not occur, diseases causing inflammation, immunological disturbances and coagulation abnormalities in the mother's immune system can affect placental function and, hence the fetal development. This aspect warrants careful attention (Prochaska et al., 2020).

Evidence suggests that any disease during pregnancy—especially in the third trimester—can have significant consequences for both fetus and mother like villitis, spontaneous abortion, multisystem organ damage and even death. Infections with SARS & MERS have led to maternal death, miscarriage, fetal growth restriction, premature birth, and stillbirths. Notably concerning

SARS-CoV-2, stillbirths and premature births due to maternal virus infection show higher prevalence when compared to general populations (Gee et al., 2021a).

Additionally, infectious diseases like HIV (human immunodeficiency virus) and HCV (hepatitis C virus) can trigger immune responses in mothers that subsequently alter their newborns' immune systems. These changes may result in long-term consequences for the offspring (Gabriel et al., 2019; Kamdar et al., 2020).

1.7 COVID-19 complications on the neonates

When a mother contracts COVID-19, several complications—such as inflammation, called Maternal immune activation (MIA), coagulation alterations & endothelial modifications—can occur. This not only impacts the mother but also exposes the placenta and fetus to these complications. Consequently, there are noticeable changes in both the placenta and the fetus (Sharps et al., 2020).

The presence of COVID-19 raises levels of pro-inflammatory cytokines like IFN-gamma, IL-1β, IL-8, IL-6, TNF, which disrupts the balance between anti-inflammatory & pro-inflammatory responses, and similarly affects their offspring's immune systems and leads to increase in numbers of NK cells, NKT cells, Vδ2+ γδ T cells & CD161-expressing CD8+ T cells which is higher than average in infants born to mothers with COVID-19 when compared to those born to healthy mothers. Such an inflammatory state can involve multiple organ systems and lead to other adverse outcomes, including increased risk for psychological issues like depression and bipolar disorders (Figure 1.3) (Sharps et al., 2020).

Reports indicate that COVID-19 during pregnancy may cause preterm birth and low birth weight. These conditions significantly contribute to child mortality under five years of age & complicate health with issues such as necrotizing enterocolitis, bronchopulmonary dysplasia, and respiratory distress syndrome (Gee et al., 2021b).

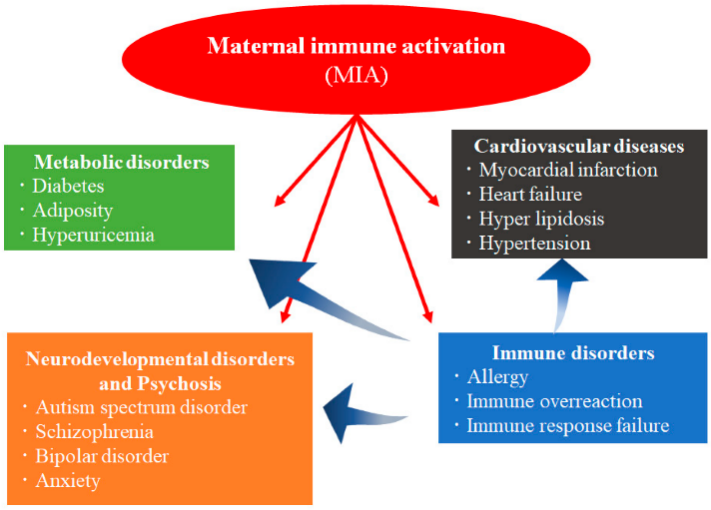

Figure 1.3 Consequences of exposure to MIA in offspring (Shimizu et al., 2023)

1.8 Impact of environment, lifestyle, and maternal diseases on child health

Studies have delved into the root causes of health problems in individuals sharing similar backgrounds, considering lifestyle & blood composition. They brought forth the "fetal origins hypothesis." According to this idea, detrimental conditions like malnutrition and illnesses during pregnancy can have lasting impacts on a child's health, increasing the likelihood of various chronic diseases later in life (Garthwaite, 2008).

For instance, malnutrition during a mother's pregnancy can contribute to future health issues in the child, such as heart problems, kidney issues, and metabolic disorders. This hypothesis gained traction after the 1918 influenza pandemic (Spanish flu). Children born to mothers affected by the disease during that time were observed to have higher rates of heart, kidney, stroke, diabetes, and other physical ailments and also autism and depression compared to peers whose mothers remained healthy. Studies indicate that the timing of exposure to illness influences both the type and severity of health issues in offspring (Almond & Mazumder, 2005; McCarthy et al., 2021).

Another example is that babies whose mothers had a urinary tract infection during pregnancy may be more likely to develop a range of lifelong health issues, such as neurocognitive disorders or asthma (Bergman & Cedar, 2013).

Research has indicated that a fetus's environment and its changes can also impact fetal gene expression which can impact the health of the child in the future. For example, chorioamnionitis—an ascending infection within the maternal pathway—induces a fetal inflammatory response by producing inflammatory mediators. This response can activate labor prematurely. Changes in amniotic fluid conditions surrounding the fetus may alter gene expression with enduring consequences for the offspring's life.

These changes include variations in antimicrobial protein production by fetal membranes and chorionic-amniotic membranes. Also noteworthy are shifts in key genes responsible for inflammatory, immune, respiratory, and nervous functions—potentially leading to asthma, allergic reactions,and neurological disorders like cerebral palsy & schizophrenia. In conclusion, it is clear that chorioamnionitis during pregnancy can disrupt both innate or acquired immune systems and future nervous system development through altered fetal gene expression (Erez et al., 2009; Gayen Nee' Betal et al., 2019; Weitkamp et al., 2016).

1.9 How the conditions during pregnancy affect the future health status of children

During pregnancy, environmental factors can alter cellular pathways and gene expression patterns in the fetus through epigenetic mechanisms which are modifications to the genome that affect gene expression without altering the underlying DNA sequence unlike mutations, including DNA methylation and chromatin modifications, often leading to changes in the epigenome and changes the regulation of genes expression in different states during the child's life. These alterations may persist into childhood, even adulthood**.** For instance, exposure to specific nutritional elements or environmental toxins during pregnancy can modify DNA methylation patterns or histone modifications in the developing fetus. These early-life epigenetic changes can have consequences extending into childhood & adulthood and increase disease susceptibility, potential epigenetic inheritance, developmental programming effects on physiological processes; metabolic disorders, behavioral and cognitive impacts, and reproductive health implications. These alterations profoundly impact health across an individual's lifespan—underscoring the importance of understanding how fetal environments influence gene regulation for long-term health outcomes. Improper regulation of genes due to these epigenetic changes has

been linked to complex disorders such as cancer, Alzheimer's disease, asthma & autism (Kazmi et al., 2019).

Examples of environmental factors impacting fetal gene expression through epigenetic mechanisms include nutritional factors like a maternal diet during pregnancy; environmental toxins such as arsenic & lead; endocrine disruptors like diethylstilbestrol (DES) and bisphenol A (BPA); behavioral factors such as maternal stress; and exposure to low-dose radiation (Giraud et al., 2020).

Here are some examples of maternal conditions impacting the gene expression of the exposed children:

The Spanish Flue Pandemic: This pandemic in 1918 had lasting effects on children exposed in utero by epigenetic modification mechanisms like DNA Methylation**,** Histone Modifications and increased the risk of chronic diseases like coronary heart disease, diabetes, kidney issues, and also suffering from cognitive or developmental impacts and Immune System Dysregulation in these children. (Garthwaite, 2008)

Maternal dyslipidemia or obesity: These children may have increased LDL cholesterol in adulthood, and higher adiposity at birth due to changes in gene expression related to lipid metabolism in offspring that may contribute to the offspring's predisposition to atherosclerosis or other metabolic disorders (Kerr et al., 2018).

High blood pressure of mother during pregnancy: Hypertensive disorders of pregnancy (HDP), including gestational hypertension (GH) and preeclampsia, have been associated with epigenetic modifications in cord blood from offspring through DNA methylation modification mechanisms and lead to lower birth weight of children and cardiovascular diseases (Kazmi et al., 2019).

Prenatal inflammation: Studies have shown that babies exposed to inflammation during pregnancy or shortly after birth had problems with their motor skills and social skills even without severe brain damage (Giraud et al., 2020).

Effects of Histological Chorioamnionitis: Epigenetic changes caused by Histological Chorioamnionitis exposure in neonates may contribute to an increased risk of long-term health problems, including cerebral palsy, developmental delay, asthma, and allergies by altered methylation patterns in genes (Fong et al., 2020).

1.10 Statement of problem

Pregnancy is recognized as a high-risk condition for SARS-CoV-2 infection due to physiological adaptations, hormonal shifts, and compromised immune function. However, knowledge about the direct impact of the virus on developing fetuses and the long-term health of newborns is still in its early stages. The existing literature lacks comprehensive information on both the short- and long-term implications of COVID-19 infection during pregnancy on newborns. Specifically, the epigenetic regulation of metabolic pathways in the context of maternal COVID-19 infection remains largely unknown.

Given the known inflammatory effects of COVID-19 on maternal physiology, it is crucial to consider how these inflammatory responses may alter the fetal environment and contribute to epigenetic modifications in the developing fetus. Inflammatory processes triggered by COVID-19 in pregnant women could potentially influence gene expression patterns in the fetus through epigenetic mechanisms. Understanding how these factors impact fetal gene expression is vital for elucidating the mechanisms underlying the effects of maternal infections and inflammatory responses on fetal development and health outcomes.

This gap in our understanding is significant, as it limits our ability to intervene effectively to optimize perinatal outcomes for both maternal and newborn health. Therefore, examining the relationship between maternal COVID-19 infection and the epigenetic regulation of metabolic pathways in offspring is a priority research need. By investigating this issue, we aim to gain a deeper understanding of how fetuses are shaped in the context of viral infection and to provide the necessary evidence to design and implement interventions that promote positive health outcomes for newborns.

1.11 Epigenetic regulations

Epigenetic regulation occurs through changes in phenotype or gene expression that are caused by mechanisms other than alterations in the underlying DNA sequence. These changes can persist throughout the cell's life, be transmitted through cell division, and even be passed down through multiple generations in a specific cell lineage. Epigenetic mechanisms provide an additional layer of control in the regulation of gene expression, allowing an organism to adapt to changing environments. This regulation is influenced by factors such as developmental stage, tissue type, environmental conditions, and disease status. Epigenetic mechanisms can alter chromatin structure, nuclear organization, and transcript stability, ultimately influencing gene expression and leading to long-term molecular and functional consequences (Figure1.4).
The mechanisms of epigenetic regulation include histone modification, DNA methylation, and expression of small non-coding RNA species.

The mechanisms through which gene expression is regulated are as follows:

1. **DNA methylation:** It is a crucial process in epigenetics that involves the DNA methyltransferase (DNMT) enzymes which modify the DNA. Methyl groups from S-adenosyl methionine are brought into the DNA to create 5-methylcytosine (5mC). DNA methylation can affect gene transcription by blocking transcription factors from binding and interfering with transcriptional machinery. DNA methylation frequently occurs at promoter CpG dinucleotides, influencing critical processes and functions. There are various roles of DNMTs, including genomic imprinting, establishing heterochromatin, gene silencing, embryonic development, and activity in de novo methylation (Arora & Tollefsbol, 2021).
2. **Post-translational modifications (PTM) of histone proteins:** These are crucial epigenetic mechanisms that consist of covalent modifications such as acetylation, methylation, ubiquitylation, and phosphorylation on specific residues of histone tails. These modifications regulate cellular processes such as transcription, repair, and replication by altering chromatin function and can either activate or repress gene expression depending on which residues are modified. Enzymes such as Histone acetyltransferases (HATs) and histone methyltransferases (HMTs) can add these groups, while histone deacetylases (HDACs) and histone demethylases (HDMs) can also remove

these modifications, allowing for the dynamic regulation of histone modifications for positive or negative regulation of gene expression (Hu et al., 2019).

3. **Noncoding RNA (ncRNA):** Non-coding RNA (ncRNA) is important in gene expression and cellular processes, and microRNAs (miRNAs) are one of the most studied types. As short RNA sequences, miRNAs can block mRNA translation into protein by binding to areas of complementary sequence and forming the miRNA-induced silencing complex (miRISC). Many mechanisms, like DNA methylation, regulate protein-coding gene transcription and microRNA expression. In addition, miRNAs can impact the expression of epigenetic regulators, such as DNA methyltransferases and histone deacetylases, and thus provide a link between the expression of proteins and the epigenome. The intersection of these pathways for the regulation of protein expression and chromatin structure creates a dynamic regulatory network between different epigenetic pathways that control patterns of gene expression either through transcriptional or post-transcriptional mechanisms (Moutinho & Esteller, 2017).

Changes in DNA methylation play a significant role in altering gene expression patterns and have been identified as major contributors to diseases. Aberrant DNA methylation can result in altered gene expression patterns, leading to the development of diseases such as cancer, autoimmune disorders, or neurodevelopmental diseases like Rett syndrome. In the field of transgenerational carcinogenesis, DNA methylation alterations caused by exposure to environmental chemicals in parents are passed on to their offspring, making their offspring more susceptible to developing tumors. This finding underlines the central role of environmentally-induced epigenetic changes on health outcomes across generations (Santos et al., 2005).

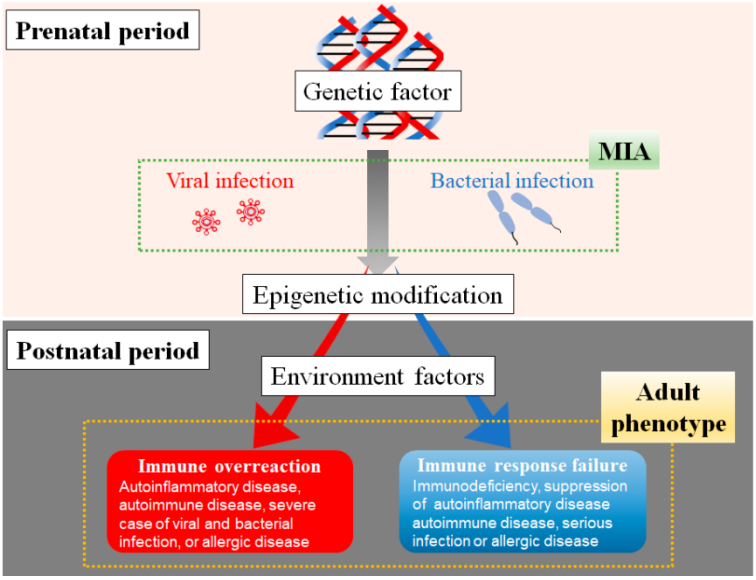

Figure 1.4 The interplay of genetic factors, maternal immune activation, and environmental influences leads to immune dysfunction after birth (Shimizu et al., 2023)

1.12 Covid-19 impacts on pregnancy and fetus:

Evidence demonstrates that neonatal immune response may be impacted by maternal SARS-CoV-2 infection. This effect is demonstrated by changes in pro-inflammatory cytokines specifically IL-6 in neonates as well as changes in neonatal immune cells; the increase in IL-6 has been associated with cognitive impairments in offspring due to maternal immune activation (MIA) (Gee et al., 2021b). Maternal COVID-19 MIA and systemic inflammation can affect epigenetic mechanisms. Beyond the effect of MIA, the potential impact of the increased psychological stress in the mothers from COVID-19 due to COVID-19 status or symptoms on the fetus could negatively affect fetal programming. Studies show that COVID-19 infection during pregnancy has gene-specific DNA Methylation changes, some specifically in stress-related genes, others with stress and inflammation-related genes, that were comparable to histological chorioamnionitis (HCA) (Hill et al., 2023). In umbilical cord blood cells of infants whose mothers had COVID-19, there were 119 genes that were differentially methylated and they were mapped to canonical pathways related to stress response, hepatotoxicity, nephrotoxicity, and cardiotoxicity. This potential change could have an effect on child development (Urday et al., 2023).

1.13 Gene Expression Patterns as Markers for Developmental Defects:

Identification of altered gene expression patterns in the developing embryo and fetus can serve as indicators for predicting developmental defects. These patterns allow for early detection, diagnostic significance, and predictive value in identifying abnormalities. They can be used as a system of markers to assess risks and understand functional consequences. Integrating gene expression data with other measures offers a comprehensive analysis for identifying and addressing potential defects in development (Taylor et al., 2019).

1.14 Methods of Gene Expression Studies

In different biological contexts, there are a variety of methods that can be used to analyze the activity of genes, and some of the most commonly used techniques in gene expression studies are:

1. Reverse Transcription Polymerase Chain Reaction (RT-PCR): RT-PCR is used to amplify and quantify RNA molecules. The method allows you to measure the expression levels of genes being studied by first converting RNA to complementary DNA (cDNA) and then amplifying the cDNA using PCR.

2. Quantitative Real-Time PCR (qPCR): qPCR is a highly sensitive method for quantitating gene expression levels. It allows for continuous monitoring of PCR amplification in real-time and provides very accurate gene expression data.

3. Microarray Analysis: Microarray technology allows researchers to analyze the expression of thousands of genes simultaneously by hybridizing cDNA or RNA samples to microarray chips that contain gene probes. This method facilitates the quantitative assessment of expression levels of multiple genes using a single experiment.

4. RNA Sequencing (RNA-Seq): RNA-Seq is a high-throughput sequencing method that is used to analyze the transcriptome of a cell or tissue type. This method provides a very comprehensive view of gene expression by allowing the sequencing and quantitation of the different RNA molecules.

5. Northern Blotting: Northern blot is the method used to detect specific RNA molecules in a sample. The method involves separating the RNA molecules using gel electrophoresis, transferring them to a membrane, and hybridizing them with labeled probes to visualize the expression of genes.

6. In Situ Hybridization: is a method that is used to localize specific RNA molecules in a cell or tissue. Through hybridization of labeled RNA probes to complementary RNA sequences in fixed samples, researchers can visualize the gene expression pattern spatial distribution.

7. Reporter Gene Assays: involve fusing a reporter gene, such as luciferase or GFP, to a gene of interest to monitor its expression. The activity of the reporter gene serves as a readout for the target gene expression level.

8. Western Blotting: Although Western blotting is primarily for protein analysis, it can also be used to validate gene expression at the protein level. By detecting specific proteins via antibodies, researchers can confirm gene expression data from RNA analysis.

9. Chromatin Immunoprecipitation (ChIP): is a technique researchers use to analyze protein-DNA interactions, such as transcription factor binding to gene promoters. By immunoprecipitating protein-DNA complexes, researchers can identify the genome regions that are involved in the control of the target gene expression. Combined with other methods, these techniques are essential tools for understanding spatial and temporal patterns of gene expression, underlying regulatory mechanisms, and functional consequences across biological systems (Burgess & Hazelton, 2000; Taylor et al., 2019).

1.15 Microarray Data method:

Microarray technology is a valuable instrument used in molecular biology for exploring gene expression profiles on a genomic scale. Microarrays are a technique used in gene expression analysis that is based on the complementarity between nucleic acid strands, allowing for the detection of specific sequences through hybridization (Drăghici, 2011).

Microarrays consist of a solid support, often a glass slide, to which cDNAs or oligonucleotides are spotted or printed in a high-density pattern. The cDNAs on the microarray represent genes or fragments of genes, allowing analysis of thousands of genes in the same experiment.

In this method, RNA is first isolated from the samples and converted into complementary DNA (cDNA). This cDNA is labeled with a fluorescent dye or radioisotope to prepare for hybridization to the microarray. Once labeled, the cDNA probes are introduced to the microarray, where they hybridize with the cDNAs present on the array, matching each probe to its complementary sequence. After hybridization, a washing step is performed to remove any unbound or nonspecifically bound probes. This process reduces background noise, ensuring more accurate data collection. The microarray is then scanned to capture the fluorescent signals emitted by the bound probes at specific spots on the array. Finally, the acquired image data is analyzed to quantify gene expression levels. By examining the fluorescence intensity at different spots on the microarray, differences in gene expression across the samples can be identified and compared. This methodology offers a systematic approach to understanding gene expression, providing valuable insights for our research (Burgess & Hazelton, 2000).

Benefits and limitations of the Microarray data method:

Microarray data methodology offers several advantages and disadvantages. One significant advantage is its high throughput capability, allowing researchers to measure the expression of thousands of genes simultaneously, which enhances understanding of gene expression patterns. Additionally, microarray analysis can uncover novel genes influenced by various conditions, potentially leading to the discovery of biomarkers and therapeutic targets. It also aids in comprehending biological processes like disease progression and responses to treatments. The technology supports comparative studies across different experimental conditions or species, enhancing insights into shared and divergent gene expression patterns. Furthermore, it facilitates functional annotation by revealing the functions of unannotated genes based on their co-expression with known genes. Microarray analysis holds promise for personalized medicine through the identification of gene signatures linked to specific diseases or drug responses. Finally, the integration of microarray data with other omics data provides a comprehensive view of biological systems.

On the downside, microarray data can be affected by high background noise and technical artifacts, which may compromise accuracy. The limited dynamic range of microarrays can lead

to inaccuracies in measuring both low- and high-abundance transcripts due to signal saturation. Cross-hybridization, where cDNAs nonspecifically bind to probes, can also skew gene expression measurements. Additionally, variations across different microarray platforms can complicate the comparison of results from diverse studies. The complexity of data analysis requires bioinformatics expertise, making the interpretation of results challenging. Validation of microarray findings with independent methods, such as qPCR, is essential to ensure accuracy. Lastly, microarrays may not encompass the entire transcriptome, potentially overlooking alternative transcripts, non-coding RNAs, and other regulatory elements. Despite these limitations, microarrays remain a vital tool for studying gene expression and provide valuable insights into complex biological processes (Dopazo et al., 2001).

However, DNA microarrays have been revolutionary in studying the gene expression patterns of development by allowing for the analysis of thousands of genes simultaneously. The implications of microarrays include the ability to profile the genome, address the identification of developmental genes, and recognize patterns of tissue-specific expression. Furthermore, microarrays have proven relevant by elucidating regulatory networks, studying gene expression across species and different stages of normal development, utilizing mutants to study developmental defects, and generating functional annotations of genes based on expression patterns.

1.17 Methods for Selecting Differentially Expressed Genes in Microarray Data Analysis

Identification of differentially expressed genes (DEGs) is a very important task in the analysis of microarray data, as it allows researchers to discover which genes display significant changes in expression between different experimental groups. There are many methods that can be used to identify DEGs in a microarray dataset. The following are some of the most known methods:

1. **Fold change analysis**: This method compares the expression levels of genes between two experimental groups by computing the ratio of the expression levels of those genes in one group relative to the other group. If the ratio meets or exceeds a predetermined threshold (usually greater than or equal to 2-fold or 1.5-fold), the gene is then considered to be differentially expressed. This method is simple and straightforward in determining which genes have the

largest difference in expression levels, but it does not take into account the size and variability of the data or whether the changes are meaningful, ultimately losing genes of interest that have steady and slight changes.

2. **The outlier ratio method:** This method identifies genes whose ratio of expression is very different than the mean ratio of expression across all genes. These genes have an expression ratio that is more extreme and different than the mean expression ratio by more than a certain number of standard deviations. This method takes into account and adjusts for the variability of the data, which consistently produces the same percentage of DEGs. However, noise in the dataset can still cause false positives during the process.

3. **Testing the significance of a statistical hypothesis:** Different statistical tests including t-tests, ANOVA, or nonparametric testing are used for hypothesis testing to determine if significant differences exist in gene expression between the experimental groups. If any gene has a p-value less than a significant level (usually $p<0.05$), the gene is then called differentially expressed. This method provides the user with statistical significance and helps to control for false positives. On the other hand, this method requires the user to make corrections for multiple comparisons due to the fact that there are a large number of genes being tested at once.

4. **Model-based approaches:** This method is statistical model-based, where models are specified to discover gene expression differences and to identify DEGs. Genes are chosen on the basis of model parameters, such as log-odds ratios or maximum likelihood estimates. This approach can take into account complex patterns in the data and can also give robust estimates of DEGs, but it requires certain assumptions regarding the data distribution and might need a considerable amount of computation.

By utilizing these methods to select differentially expressed genes in microarray data, researchers can identify important genes that contribute to biological processes, disease mechanisms, and responses to treatments. The identification of these genes can provide valuable directions for further analyses and interpretation (Drăghici, 2011).

1.18 Gene set expression analysis methods

Gene sets are groups of genes that share common functional annotations, biological pathways, or regulatory relationships. Gene sets can make it easier to analyze gene expression data because they allow you to treat the behavior of a cluster of genes as one item rather than looking at each gene separately. In microarray data analysis, gene sets help to move the field from single gene-by-gene analysis to a more biologically meaningful interpretation.

Benefits of Gene set expression analysis methods:

Gene set expression analysis (GSA) methods offer numerous advantages over traditional individual gene expression analysis techniques, enhancing our understanding of biological processes. One major advantage is their pathway-centric approach, which allows researchers to analyze collections of genes that share a specific biological function or pathway simultaneously, rather than focusing on individual genes. This fosters a deeper understanding of gene expression changes and the network relationships among genes. Additionally, GSA methods reduce subjectivity in biological interpretation by examining the mutual behavior of genes within sets, providing a clearer analytical context compared to individual gene analysis. Furthermore, GSA methods tend to have greater statistical power, enabling the detection of biologically subtle, coordinated changes in gene expression, especially when strong correlations exist among genes. These methods effectively account for the correlation structure between genes, which is vital for understanding connectivity in biological pathways. GSA also allows for the consideration of multivariate outcomes, which is essential for analyzing complex biological features. Many GSA methods utilize existing biological knowledge from sources like KEGG and Gene Ontology, enhancing the reproducibility and interpretability of results across various studies. Ultimately, GSA provides insights into the biological relevance of gene sets and pathways concerning specific biological processes or diseases, adding significant value to research efforts.

Overall, by improving biological relevance, interpretability, statistical power, and reproducibility, gene set expression analysis methods present a more holistic and insightful approach to analyzing gene expression data, enhancing our understanding of the intricate regulatory mechanisms underlying biological phenomena (Dinu et al., 2021.).

We define gene sets in microarray data analysis by incorporating functional annotations, biological pathways, and regulatory relationships. Functional annotations from databases like KEGG and Gene Ontology allow for the grouping of genes with similar functions. Additionally, gene sets can represent biological processes or pathways where genes collaborate. Analyzing changes in gene expression at the gene set level enables researchers to comprehend the collective behavior of related genes, enhancing statistical power by reducing the burden of multiple tests associated with individual genes. This analysis also facilitates the interpretation of results within the context of known biological pathways and processes. Ultimately, gene set analysis provides a biologically relevant way to examine expression changes in microarray data, helping to elucidate the underlying biological processes (Song & Black, 2008).

1.19 Molecular Signatures Database (MSigDB)

The resource known as the Molecular Signatures Database (MSigDB) consists of a collection of 34,550 gene sets organized into gene sets of related biological themes, pathways, and characteristics. These gene sets have been grouped into nine top-level collections, each with several subcollections, making it easier to identify gene sets that are closely related to the specific research question. The top-level collections include H (hallmark gene sets), C1 (positional gene sets), C2 (curated gene sets), C3 (regulatory target gene sets), C4 (computational gene sets), C5 (ontology gene sets), C6 (oncogenic signature gene sets), C7 (immunologic signature gene sets), and C8 (cell type signature gene sets). The gene sets in MSigDB may originate from a variety of sources such as online pathway databases, biomedical literature and contributions by individual experts. Each gene set page records the origin of the gene set.

Chapter 2: Methods

2.1 Study design

To demonstrate epigenetic and metabolic pathways alterations in neonates born of COVID-19-infected mothers without vertical transmission, we analyzed a gene expression dataset retrieved from the Gene Expression Omnibus (GEO) database with accession number GES 195938 (Gayen Nee' Betal et al., 2022).

The dataset was created by collecting cord blood cell samples from 16 term neonates at the time of delivery, with 8 neonates exposed to COVID-19 during pregnancy (cases). Two mothers were symptomatic at their delivery, and one of them had severe symptoms. 8 infants were served as controls with no exposure to the virus during pregnancy. In the control group, 6 were born before the COVID-19 pandemic and 2 were born to mothers who were COVID-19 antibody negative at the time of delivery. Mothers with clinical or histological chorioamnionitis, as this may impact gene expression, were excluded. Similarly, we excluded mothers vaccinated with the COVID-19 vaccine.

2.2 Microarray data normalization

Before conducting the analysis of the dataset that contains the raw data, we first prepared it via quantile normalization, one of the effective methods for this type of preparation. Some of the advantages of quantile normalization include a reduction in systematic biases and in technical variation; this way we can more clearly see the "truth" in the biological signals instead of only seeing artifacts due to data collection methods. The choice of quantile normalization rests upon these advantages, since the consequence is having enhanced quality and reliability of the data so that more accurate comparisons and analyses are possible, and it involves ranking the data for each sample, calculating the mean of the ranks across all samples, and then replacing the original value with the mean values, to achieve similar distributions across samples (Rao et al., 2008).

2.3 Statistical Analysis

2.3.1 Why Gene sets analysis approaches?

The primary approach to analyzing gene expression data focused on single-gene analysis, where the expression levels of individual genes were compared between case and control samples using methods such as fold change or t-tests. However, as mentioned earlier, using these methods results in a long list of significant genes, that is hard to interpret and it is full of biases and these biases are especially prominent when analyzing differential expression with low signal genes. The multi-gene analysis is a necessary and appropriate tool to use to capture complex biological phenomena that we cannot detect with limited one-gene tool. In contrast to one gene approaches' rigidity, high-throughput studies need control over the error due to multiple comparisons between genes, while taking precautions against false negatives. There are consequences for using stringent thresholds in individual gene analysis: false positives and limitations on our conclusions. More reliable conclusions can be reached by studying and expanding our exploration in similarity groupings or gene clusters which reflect shared biological activities. It cuts through the complexity of high-throughput gene expression research with gene sets or by studying groups of genes with shared biological characteristics (Nam & Kim, 2008).
Gene-set analysis (GSA) can be seen as a non-threshold approach to microarray study analysis. Indeed, GSA makes use of pre-defined gene sets, which are grouped together based on biological knowledge. More specifically, GSA aims to determine if any of the biological pathways are significantly perturbed even if only a small number of their coordinated gene sets are perturbed. The benefit of focusing on gene sets is more informative for understanding how a biological process functions, and provides a clearer explanation of the underlying disease mechanism (Nam & Kim, 2008).

2.3.2 Self-competitive and Self-contained GSA Methods

Gene set analysis methods are intended to deal with these problems and interpret the research data meaningfully. There are several Gene Set Analysis (GSA) methods that are used to recognize gene sets associated with specific phenotypes with different methodological assumptions, strengths, and weaknesses. They are distinguished into two hypothesis tests:
1. Competitive methods, such as GSEA, SAFE, random set methods, and GSA, compare the gene relationships inside of a gene set to the relationships outside of the set to see if the genes inside the particular set are more associated with the phenotype.
2. Self-contained methods test the relationship between the gene set of interest and the phenotype independently of the genes outside the set. Examples of self-contained GSA strategies include the Global test, MANOVA-GSA, Significance Analysis of Microarrays for Gene Sets (SAM-GS), and Linear Combination Test (LCT) (Dinu et al., 2008; Nam & Kim, 2008).

2.4 Linear Combination Test as a Gene Set Analysis Method

Linear Combination Test (LCT) is a method of Gene Set Analysis (GSA) which can be used to test the association between sets of genes and multivariate outcomes.

Traditional Gene Set Analysis (GSA) methods have several notable limitations, primarily characterized by high rates of Type II errors, often leading to false negatives. This results in missed true associations between gene sets and biological outcomes, thereby limiting the ability to identify significant pathways associated with diseases. These traditional methods tend to be less sensitive, which means they may not adequately capture relevant biological signals, hindering researchers from fully understanding the molecular pathways involved in a specific disease. Additionally, the focus on single-gene analysis often overlooks the collective behavior of gene sets, missing important interactions and correlations among genes. Furthermore, the reliance on individual gene comparisons can diminish the statistical power, making it challenging to draw meaningful conclusions from biological data.

In contrast, the Linear Combination Test (LCT) method offers several significant advantages. LCT considers the correlations among gene sets and outcomes, enabling a more comprehensive

analysis of the relationships between sets of genes and multivariate outcomes. This characteristic improves the likelihood of identifying biologically relevant relationships. Additionally, LCT is capable of handling multivariate continuous outcomes, which better reflects the complexities inherent in biological phenomena, leading to a more rigorous analysis of gene expression data. One of the key strengths of LCT is its computational efficiency, making it suitable for analyzing large-scale omics data without incurring high computational costs, which is especially important for complex study designs.

Moreover, LCT demonstrates robustness when applied to various study designs, as it can analyze binary, univariate, and multivariate continuous outcomes measured at single or multiple time points. The method employs a shrinkage covariance matrix estimator to effectively manage challenges associated with high-dimensional data, thereby enhancing the reliability of analyses. Notably, LCT excels in identifying gene signatures relevant to dynamic biological processes, such as development and disease progression, by focusing on sets of genes rather than individual genes, ensuring that changes in gene expression are accurately captured. Additionally, LCT is well-suited for case/control studies that involve multiple binary or multi-category outcomes, as it assesses linear predictors directly on the log odds scale, thereby optimizing its performance in such scenarios (Dinu et al., 2021.).

2.5 LCT method and its formula

In this study, we applied an extension of LCT for multiple continuous phenotypes considering within-gene correlations for functionally related gene sets and also incorporating correlations among multiple continuous phenotypes.

Given a microarray study on n subjects, with measures on expressions of a pre-specified set of P genes $X= (x_1,\ldots, x_p)$ and with measures on a group of q continuous phenotypes $Y=(y_1,\ldots, y_q)$. Assume that the columns in both X and Y have been centered and scaled across the n subjects. Here we want to test if there is a significant linear relationship between the set of genes X and the group of phenotypes Y. The hypothesis of interest is that the gene expressions in the pre-specified gene set X are linearly independent to the phenotypes Y. This multivariate null hypothesis can be linearly expressed and univariately stated as

H_0: There is no relationship between any linear combination of $x_1, \ldots, x_p$ and any linear combination of $y_1, \ldots, y_q$.

For the purpose of testing a linear relationship, a given linear combination of $x_1, \ldots, x_p$ can be written as $Z(X, A) = a_1x_1 +\cdots+ a_px_p$, and a given linear combination of $y_1, \ldots,y_q$ can be written as $Z(Y,B) = b_1y_1 +\cdots+ b_qy_q$ where $A \in R^p$ and $B \in R^q$ represent the coefficient vectors of the a_i's and b_j's, respectively. Given two coefficient vectors A and B of the combination coefficients, to test if Z (X, A) is associated with Z (Y, B), we focus on the question if the two combinations are correlated. This is a standard correlation test, and a standard test statistic is based on the Pearson correlation between Z (X, A) and Z (Y, B), i.e., the Pearson correlation $\rho = \rho(Z(X, A),Z(Y,B))$ is commonly used. If both X and Y are normally distributed, then the statistic

$t = \rho \sqrt{(n - 2) / (1 - \rho 2)}$ follows a student's t-distribution with n sample and degrees of freedom n -2 under the null hypothesis. The same applies approximately if the observed values are not normal, but the sample size n is large enough. For testing the null hypothesis H_0, we form the linear combinations of $x_1,\ldots, x_p$ and $y_1,\ldots, y_q$ having the highest absolute correlation, i.e., we choose coefficient vectors A and B to maximize the Pearson correlation between Z(X,A) and Z(Y,B). This yields the newly proposed version of the linear combination test (LCT) for multiple continuous phenotypes.

$$T^2 = \max_{A,B} |\rho(Z(X,A), Z(Y,B))|^2 \tag{1}$$

Let $\Sigma_{XX} = \text{cov}(X,X)$ be the covariance matrix of X, whose (i,j)th entry is $\sigma_{ij} = \text{cov}(x_i,x_j)$. Similarly, let $\Sigma_{YY} = \text{cov}(Y,Y)$ and $\Sigma_{XY} = \text{cov}(X,Y)$ be the covariance matrix of Y and the covariance matrix between X and Y. The above statistic can be written as

$$T^2 = \max_{A,B} \frac{\left(A^T \Sigma_{XY} B\right)^2}{A^T \Sigma_{XX} A \cdot B^T \Sigma_{YY} B} \tag{2}$$

When high dimensionality of X and or dimensionality of Y is concerned, deal with the singularity of Σ_{XX} and Σ_{YY} very carefully especially when the numbers of gene set size is larger than the sample size, $p > n$. One possible solution to the singularity issue is to employ the shrinkage method of Schafer and Strimmer and replace Σ_{XX} and Σ_{YY} by their shrinkage versions Σ^*_{XX} and Σ^*_{YY}. The shrinkage covariance matrix Σ^*_{XX} and its (i,j) th element σ^*_{ij} are given by σ^*_{ij} $\sqrt{\sigma ii \sigma jj}$, and γ_{ij} can be computed according to the scheme $\gamma_{ij} = 1$, if $i = j$, $\gamma_{ij} = \rho_{ij} \min(1, \max(0, 1 - \lambda^*))$, if

$i \neq j$, where ρ_{ij} is the sample correlation between x_i and x_j, and the optimal λ^* can be estimated by

$$\lambda^* = \sum_{i \neq j} \text{var}\left(\rho_{ij}\right) / \sum_{i \neq j} \rho_{ij}^2 . \tag{3}$$

Then, we apply the shrinkage pooling technique to get the shrinkage version of the test statistic.

$$T^{2*} = \max_{A,B} \frac{\left(A^T \Sigma_{XY} B\right)^2}{A^T \Sigma^*_{XX} A \cdot B^T \Sigma^*_{YY} B} \tag{4}$$

Regarding calculation of computational cost of (4), the right-hand side is a nonlinear programming problem which involves $p + q$ parameters. It is worth noting that computing the right-hand side of (4) can be computationally expensive due to maximizing directly (or indirectly) the right-hand side of (4), especially when permutation is used to obtain the p-value of the test. To address the problem of computational efficiency, we take the strategy of using two groups of normalized orthogonal bases instead of using the original observation vectors X, Y. We perform the eigenvalue decompositions for the two shrinkage covariance matrices $\Sigma^*_{XX} = UD_XU^T$ and $\Sigma^*_{YY} = VD_YV^T$ and therefore obtain two groups of orthogonal basis vectors,

$\tilde{X} = (\tilde{x}_1, \ldots, \tilde{x}_p) = (x_1 - \bar{x}, \ldots, x_p - \bar{x})UD_x^{-1/2}$ and $\tilde{Y} = (\tilde{y}_{1,\ldots,} \tilde{y}_q)\ VD_Y^{-1/2} = (y_1 - \bar{y}, \ldots, y_q - \bar{y})$;

the test statistic of (4) can be expressed in the form of a ratio.

$$T^{2*} = \max_{\alpha,\beta} \frac{\left(\alpha^T \Sigma_{\tilde{X}\tilde{Y}} \beta\right)^2}{||\alpha||_2^2 \cdot ||\beta||_2^2}, \qquad (5)$$

where $\alpha = D_X^{1/2} U^T A$, $\beta = D_Y^{1/2} V^T B$, and $\Sigma_{\tilde{X}\tilde{Y}}$ is the covariance matrix between $\tilde{X}$ and $\tilde{Y}$, with its (i,j) entry being $\mathrm{cov}(\tilde{x}_i, \tilde{y}_j)$.

The optimization problem in (5) can be solved in two steps. Firstly, for a given β, find the optimal α, which is proportional to $\Sigma_{\tilde{X}\tilde{Y}} \beta$; secondly, substitute the optimal α into (5), and find the global optimal β, which is proportional to the first eigenvector of the matrix $\Sigma^T_{\tilde{X}\tilde{Y}} \Sigma_{\tilde{X}\tilde{Y}}$ corresponding to the largest eigenvalue. We note that the value of T^{2*} equals to the largest eigenvalue of either the q × q matrix $\Sigma^T_{\tilde{X}\tilde{Y}} \Sigma_{\tilde{X}\tilde{Y}}$ or the $p \times p$ matrix $\Sigma^T_{\tilde{Y}\tilde{X}} \Sigma_{\tilde{Y}\tilde{X}}$. The cost for getting the largest eigenvalue is low, provided min (p, q) is not big.

The computational advantage is obvious when sample permutations are used to calculate the p-value of the test. Since sample permutation changes neither the correlation structure within gene sets nor the correlation structure within phenotypes, we don't need to repeat the same eigenvalue decompositions of the two shrinkage covariance matrices in (4) for the permuted data, but only for the original one. In fact, after performing the eigenvalue decompositions for

the two shrinkage covariance matrices Σ^*_{XX} and Σ^*_{YY} and creating two groups of orthogonal basis vectors $\tilde{X}$ and $\tilde{Y}$, we can perform permutations on $\tilde{Y}$ directly, instead of on the original phenol-type Y (Dinu et al., 2013; Wang et al., 2014).

2.6 Continuous phenotypes prediction

Following the normalization of the dataset, the Fisher test was used as a method of ranking genes to predict continuous phenotypes. The Fisher's ratio (FR) method of ranking is based on the gene's ability to distinguish between phenotype groups, using the variance within groups and the variance between groups (deAndrés-Galiana et al., 2016).

This method was used to predict three continuous phenotypes associated with the innate immune system including innate immune system, Regulation of innate response, and antiviral response of innate immune system. We calculated FR values to rank the genes. Genes in the top ranks were then used to build a prediction model.

2.7 KEGG LEGACY Catalog:

Before conducting gene-set analysis, we need a dataset consisting of pre-defined gene sets. For this purpose, we obtained the KEGG_LEGACY catalog, which is part of the curated gene sets within the MSigDB C2 collection, from the following link: https://www.gsea-msigdb.org/gsea/msigdb/human/genesets.jsp?collection=CP:KEGG_LEGACY

The C2 collection of the Molecular Signatures Database contains two main categories for organizing gene sets, which are based on Chemical and Genetic Perturbations (C2:CPG) and Canonical Pathways (C2:CP). The pathway gene sets are sourced from various online databases, including BioCarta, KEGG MEDICUS, Pathway Interaction Database, Reactome, SigmaAldrich, Signaling Gateway, Super Array SABiosciences, WikiPathways, and the KEGG legacy sets.

The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a vital resource for understanding the functions of biological systems at different levels, ranging from cellular processes to entire ecosystems. It achieves this by collecting and integrating molecular-level data through specialized databases, particularly leveraging extensive datasets derived from genome sequencing and high-throughput experimental techniques. The KEGG model is designed to provide detailed conceptual information about biological systems and includes 16 integrated

databases categorized into four main areas: (a) systems information, (b) genomic information, (c) chemical information, and (d) health information. The KEGG PATHWAY database falls under the category of systems information (http://www.genome.jp/kegg/). It consists of manually developed pathway maps that represent our understanding of molecular interactions, reaction networks, and relationships among various biological processes, cellular functions, and organismal systems. The KEGG PATHWAY categorizes this information into seven clusters, namely, metabolism, genetic information processing, environmental information processing, cellular processes, organismal systems, human disease, and drug development clades. The gene sets that map to the Canonical Pathways are taken directly from the KEGG pathway database, which has 186 gene sets grouped into these same seven clusters. The synthesis of information from several different sources can help improve our understanding of the molecular functions supporting biological systems.

2.8 Proposed method

Step 1

First, we extracted the data related to the measurement of genes from our cases and controls from the dataset GSE195938, and we normalized it using the quantile normalization to make it ready for analysis. Then we used The Fisher's ratio (FR) method of ranking to predict three continuous phenotypes associated with the innate immune system.

Step 2:

Then we run the univariate versions of LCT for each of the continuous phenotype's expressions considering the 186 gene sets of KEGG LEGACY catalog to find out if there is any associations between the expression of any gene sets and our intended phenotypes or not.

Step 3:

After that, we run the multivariate version of LCT for the combination of all three phenotypes, three times, and each time we assigned different weights to the values obtained for each phenotype. By assigning higher weights to the phenotype of the antiviral response of the innate

immune system, we aimed to focus more on this phenotype as the outcome of our analysis so that we could identify gene sets that play a greater role in its expression.

Table 2.1 The way of assigning weights to the phenotypes in each of the analyses.

	Phenotype		
Weighting system	Innate immune system	Regulation of innate immune response	Antiviral innate immune response
1-1-1	1	1	1
1-2-3	1	2	3
1-2-4	1	2	4

Step 4:

Our goal in assigning different weights to each phenotype in the outcome is to see if these weighting systems can help us identify a state that can better differentiate between cases and controls or not. To answer this question, we combined the P-values of cases and controls for each of the significant gene sets using the following formula:

$$\log \{(1-Pcase) * Pcontrol \} / \{ Pcase * (1-Pcontrol) \}$$

If the values in one column were greater than those in another column, it indicated that the one of the weighting systems performed better in differentiating between cases and controls.

2.9 Software and Packages

All statistical computations were conducted utilizing R software (version 4.3.2). Free R packages for conducting the LCT for continuous phenotypes can be accessed at https://sites.ualberta.ca/~yyasui/homepage.html.

Chapter3: Results

3.1 Demographic and clinical characteristics

In this study, we used gene expression data of 16 term infants. Eight infants were classified as COVID-19 group, whereby the infants were exposed to COVID-19 during pregnancy and eight infants were classified as Control group whereby the infants were born before the pandemic (n =6) or maternal COVID-19 antibody negative at the time of delivery during the pandemic (n = 2). The median time of COVID-19 infection diagnosed was 89 days (range 1–238 days) before delivery. Two mothers were symptomatic at the time of delivery, one of which experienced severe symptoms. The median gestational age of COVID-19 diagnosis was 24 weeks (range 5–37 weeks). Clinical and demographic data shown in Table 1

Table 3.1 Demographic and clinical characteristics of cases and controls (Gayen Nee' Betal et al., 2022)

	COVID-19 exposure (*N* = 8)	**Control (*N* = 8)**	***p-value***
Birth weight in Kg (mean ± SD)	3.14 ± 0.63	2.98 ± 0.56	0.6
Gestational age in weeks (mean ± SD)	38.1 ± 1.3	38.5 ± 1.4	0.6
Male sex *n* (%)	5 (62.5)	6 (75.0)	1.0
Maternal diabetes *n* (%)	1 (12.5)	0 (0)	1.0
Chronic hypertension *n* (%)	1 (12.5)	1 (12.5)	1.0
Preeclampsia *n* (%)	1 (12.5)	0 (0)	1.0
Small for gestational age *n* (%)	1 (12.5)	1 (12.5)	1.0
Healthy neonate *n* (%)	8 (100)	8 (100)	1.0

Here we see there were no significant differences in birth weight, gestational age, sex, maternal diabetes, chronic hypertension, preeclampsia, or small for gestational age between the two groups.

3.2 Univariate analysis results of results

After normalization of our gene expression data**,** the univariate version of LCT was performed for each of the three continuous phenotypes considering each of the 186 gene sets in the KEGG LEGACY catalog to evaluate their association with the expression of each of the phenotypes as outcomes.

Tables 3.2, 3.3 and 3.4 show the significant gene sets associated with the three phenotypes of innate immune system, regulation of innate immune response and antiviral innate immune response with the p-values less than or equal to 0.05.

Table 3.2 Univariate LCT results of KEGG LEGACY Gene sets for the innate immune system phenotype with p-values less than 0.05

	Gene set Name	**size**	**p-value**
1	KEGG_PROTEIN_EXPORT	23	0.0001
2	KEGG_THYROID_CANCER	29	0.0001
3	KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_HEPARAN_SULFATE	26	0.001
4	KEGG_SULFUR_METABOLISM	12	0.001
5	KEGG_OTHER_GLYCAN_DEGRADATION	15	0.003
6	KEGG_PROPANOATE_METABOLISM	32	0.003
7	KEGG_CELL_CYCLE	124	0.004
8	KEGG_DILATED_CARDIOMYOPATHY	88	0.004
9	KEGG_HYPERTROPHIC_CARDIOMYOPATHY_HCM	81	0.005
10	KEGG_NITROGEN_METABOLISM	23	0.005
11	KEGG_AMINOACYL_TRNA_BIOSYNTHESIS	40	0.006
12	KEGG_N_GLYCAN_BIOSYNTHESIS	46	0.006
13	KEGG_NON_HOMOLOGOUS_END_JOINING	13	0.006
14	KEGG_PARKINSONS_DISEASE	109	0.006

15	KEGG_PRION_DISEASES	35	0.006
16	KEGG_LYSINE_DEGRADATION	43	0.007
17	KEGG_CARDIAC_MUSCLE_CONTRACTION	72	0.008
18	KEGG_OXIDATIVE_PHOSPHORYLATION	112	0.008
19	KEGG_VALINE_LEUCINE_AND_ISOLEUCINE_DEGRADATION	43	0.008
20	KEGG_HUNTINGTONS_DISEASE	168	0.011
21	KEGG_SYSTEMIC_LUPUS_ERYTHEMATOSUS	123	0.013
22	KEGG_FOLATE_BIOSYNTHESIS	11	0.016
23	KEGG_PROXIMAL_TUBULE_BICARBONATE_RECLAMATION	21	0.02
24	KEGG_ARACHIDONIC_ACID_METABOLISM	55	0.026
25	KEGG_ABC_TRANSPORTERS	42	0.028
26	KEGG_HOMOLOGOUS_RECOMBINATION	28	0.029
27	KEGG_PORPHYRIN_AND_CHLOROPHYLL_METABOLISM	31	0.037
28	KEGG_GLYCOSYLPHOSPHATIDYLINOSITOL_GPI_ANCHOR_BIOSYNTHESIS	24	0.038
29	KEGG_BETA_ALANINE_METABOLISM	22	0.044
30	KEGG_SPHINGOLIPID_METABOLISM	38	0.048

Table 3.2 shows that the expressions of 30 gene sets of KEGG LEGACY catalog are significantly associated with the innate immune system continuous phenotype.

Table 3.3 Univariate LCT results of KEGG LEGACY Gene sets for the regulation of innate immune response continuous phenotype with p-values less than 0.05

	Gene set Name	**size**	**p-value**
1	KEGG_COMPLEMENT_AND_COAGULATION_CASCADES	67	0.017
2	KEGG_ASTHMA	25	0.019
3	KEGG_O_GLYCAN_BIOSYNTHESIS	28	0.03
4	KEGG_GLYCINE_SERINE_AND_THREONINE_METABOLISM	31	0.033
5	KEGG_ABC_TRANSPORTERS	42	0.036
6	KEGG_THYROID_CANCER	29	0.037
7	KEGG_BLADDER_CANCER	41	0.044
8	KEGG_GLYCEROLIPID_METABOLISM	47	0.045

Table 3.3 shows that the expressions of 8 gene sets of KEGG LEGACY catalog are significantly associated with the regulation of innate immune response phenotype.

Table 3.4 Univariate LCT results of KEGG LEGACY Gene sets for the antiviral innate immune response continuous phenotype with p-values less than 0.05

	Gene set Name	**size**	**p-value**
1	KEGG_AMINOACYL_TRNA_BIOSYNTHESIS	40	0.011
2	KEGG_INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION	44	0.036
3	KEGG_ASTHMA	2	0.036

According to the Table 3.4, the expressions of 3 gene sets of KEGG LEGACY catalog are significantly associated with the antiviral innate immune response continuous phenotype

3.3 Multivariate analysis results

Following running univariate LCT analysis, we run the multivariate LCT analysis for the combination of all three phenotypes with giving different weights to each phenotype in three analyses.

Tables 3.5, 3.6 and 3.7 show the significant gene sets of KEGG LEGACY catalog associated with the combination of the three phenotypes of innate immune system, regulation of innate immune response and antiviral innate immune response with the given weights of 1-1-1, 1-2-3, 1-2-4 with the p-values less or equal to 0.05.

Table 3.5 Multivariate LCT results of KEGG LEGACY Gene sets for the combination of the three phenotypes of innate immune system, regulation of innate immune response and antiviral innate immune response with the given weights of 1-1-1 with p-values less than 0.05

	Gene set Name	**size**	**p-value**
1	KEGG_ASTHMA	25	0.011
2	KEGG_ABC_TRANSPORTERS	42	0.014
3	KEGG_FOLATE_BIOSYNTHESIS	11	0.026
4	KEGG_VALINE_LEUCINE_AND_ISOLEUCINE_BIOSYNTHESIS	10	0.026
5	KEGG_GLYCINE_SERINE_AND_THREONINE_METABOLISM	31	0.027
6	KEGG_AMINOACYL_TRNA_BIOSYNTHESIS	40	0.031
7	KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_KERATAN_SULFATE	15	0.033
8	KEGG_LYSINE_DEGRADATION	43	0.033
9	KEGG_GLYCEROLIPID_METABOLISM	47	0.036
10	KEGG_N_GLYCAN_BIOSYNTHESIS	46	0.037
11	KEGG_SYSTEMIC_LUPUS_ERYTHEMATOSUS	123	0.039

According to the Table 3.5, there are 11 gene sets whose expressions are significantly associated with the combination of the three phenotypes of innate immune system, regulation of innate

immune response and antiviral innate immune response in the first weighting system with the given weights of 1-1-1

Table 3.6 Multivariate LCT results of KEGG LEGACY Gene sets for the combination of the three phenotypes of innate immune system, regulation of innate immune response and antiviral innate immune response with the given weights of 1-2-3 with p-values less than 0.05

	Gene set Name	**size**	**p-value**
1	KEGG_ASTHMA	25	0.013
2	KEGG_ABC_TRANSPORTERS	42	0.014
3	KEGG_OTHER_GLYCAN_DEGRADATION	15	0.02
4	KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_KERATAN_SULFATE	15	0.023
5	KEGG_FOLATE_BIOSYNTHESIS	11	0.027
6	KEGG_GLYCEROLIPID_METABOLISM	47	0.028
7	KEGG_LYSINE_DEGRADATION	43	0.028
8	KEGG_N_GLYCAN_BIOSYNTHESIS	46	0.031
9	KEGG_AMINOACYL_TRNA_BIOSYNTHESI	40	0.041
10	KEGG_CIRCADIAN_RHYTHM_MAMMAL	13	0.041
11	KEGG_GLYCINE_SERINE_AND_THREONINE_METABOLISM	31	0.042

According to the Table 3.6, there are 11 gene sets whose expressions are significantly associated with the combination of the three phenotypes of innate immune system, regulation of innate immune response and antiviral innate immune response in the second weighting system with the given weights of 1-2-3

Table 3.7 Multivariate LCT results of KEGG LEGACY Gene sets for the combination of the three phenotypes of innate immune system, regulation of innate immune response and antiviral innate immune response with the given weights of 1-2-4 with p-values less than 0.05

	Gene set Name	**size**	**p-value**
1	KEGG_ASTHMA	25	0.014
2	KEGG_ABC_TRANSPORTERS	42	0.018
3	KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_KERATAN_SULFATE	15	0.022
4	KEGG_LYSINE_DEGRADATION	43	0.030
5	KEGG_OTHER_GLYCAN_DEGRADATION	15	0.030
6	KEGG_VALINE_LEUCINE_AND_ISOLEUCINE_BIOSYNTHESIS	10	0.031
7	KEGG_N_GLYCAN_BIOSYNTHESIS	46	0.034
8	KEGG_GLYCINE_SERINE_AND_THREONINE_METABOLISM	31	0.035
9	KEGG_FOLATE_BIOSYNTHESIS	11	0.039
10	KEGG_SYSTEMIC_LUPUS_ERYTHEMATOSUS	123	0.039
11	KEGG_CIRCADIAN_RHYTHM_MAMMAL	13	0.040
12	KEGG_AMINOACYL_TRNA_BIOSYNTHESIS	40	0.042

According to the Table 3.6, there are 11 gene sets whose expression are significantly associated with the combination of the three phenotypes of innate immune system, regulation of innate immune response and antiviral innate immune response in the third weighting system with the given weights of 1-2-4.

3.4 Which weighting System?

To find out which weighting system is more efficient to differentiate the cases and controls, we calculate the difference of logits cases' p-values and controls' p-values.

Table 3.8 Comparing differences of logits of cases and controls p-values

			Weighting System								
			1-1-1			1-2-3			1-2-4		
	Gene set name	GS size	Case p	Control p	Diff	Case p	Control p	Diff	Case p	Control	Diff
1	KEGG_ABC_TRANSPORTERS	42	0.014	0.071	88.79	0.014	0.097	80.63	0.018	0.076	67.76
2	KEGG_AMINOACYL_TRNA_BIOSYNTHESIS	40	0.031	0.084	38.36	0.041	0.107	27.00	0.042	0.087	28.14
3	KEGG_ASTHMA	25	0.011	0.104	100.22	0.013	0.117	81.67	0.014	0.099	80.10
4	KEGG_FOLATE_BIOSYNTHESIS	11	0.026	0.065	49.30	0.027	0.076	45.33	0.039	0.077	31.41
5	KEGG_GLYCEROLIPID_METABOLISM	47	0.036	0.066	35.58	0.028	0.076	43.73	0.031	0.05	44.64
6	KEGG_GLYCINE_SERINE_AND_THREONINE_METABOLISM	31	0.027	0.32	27.60	0.042	0.332	17.73	0.035	0.337	21.02
7	KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_KERATAN_SULFATE	15	0.033	0.168	28.74	0.023	0.162	41.53	0.022	0.179	41.90
8	KEGG_LYSINE_DEGRADATION	43	0.033	0.309	23.00	0.028	0.308	27.03	0.03	0.318	24.96
9	KEGG_N_GLYCAN_BIOSYNTHESIS	46	0.037	0.065	34.78	0.031	0.057	43.02	0.034	0.068	37.31
10	KEGG_OTHER_GLYCAN_DEGRADATION	15	0.022	0.31	34.14	0.02	0.303	37.82	0.03	0.306	25.33
11	KEGG_SYSTEMIC_LUPUS_ERYTHEMATOSUS	123	0.039	0.056	34.47	0.035	0.06	37.60	0.039	0.057	34.29
12	KEGG_VALINE_LEUCINE_AND_ISOLEUCINE_BIOSYNTHESIS	10	0.026	0.809	20.84	0.028	0.797	19.50	0.031	0.799	17.83

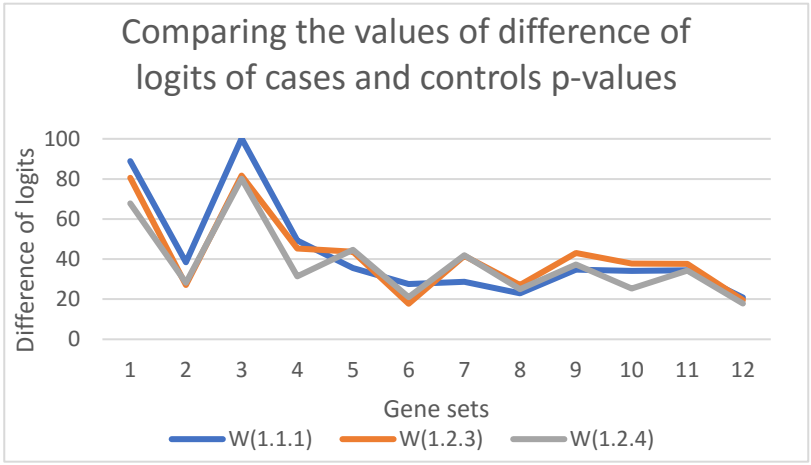

Figure 4.1 Comparing the values of difference of logits of cases and controls p-values

Figure 3.1 shows the three weighting systems perform similarly in differentiating p-values of cases and controls. Therefore, the weighting system did not perform efficiently in this regard.

Chapter 4: Discussion, Strengths and Limitations, Further studies

4.1 Discussion

The maternal-fetal interface represents a critical period in which the developing fetus is exposed to various external and internal factors that can profoundly influence its physiological development, particularly in the context of the immune system. New research is establishing a relationship between immune disorders and maternal immune activation (MIA), particularly maternal infections during pregnancy. A meta-analysis examined the influence of maternal poly(I:C) exposure during mid-gestation, which has been shown to enhance the perinatal immune response, protected up to weaning after delivery, but it did not result in evidence of a postnatal inflammatory response. Maternal infections such as COVID-19 during pregnancy are known to elicit a maternal immune response that can have ramifications for offspring health after delivery. These maternal infections have been linked to abnormal immunity after birth in children, including type 1 diabetes, allergic diseases, and neurodevelopmental disorders . Therefore, it is essential to predict the implications of maternal COVID-19 infection on long-term health outcomes for the child (Shimizu et al., 2023). Our results indicate that, Maternal COVID-19 may change multiple pathways in umbilical cord blood cells associated with some immune disorders like asthma and systemic lupus erythematosus (SLE) alongside some amino acids biosynthesis and metabolism, which in turn could be a marker of immune-mediated disorders and increase the risk of other diseases for the developing fetus.

Asthma and Systemic Lupus Erythematosus

Prior research has established that asthma, which is an allergic disease, operates by complex pathways that encompass pre-and postnatal risk factors, which are comprised of genetic factors and environmental influences from a maternal or pediatric perspective, which include infectious agents from human rhinoviruses and respiratory syncytial virus. Prenatal maternal factors such as maternal viruses/bacterial infections, maternal allergen exposures, tobacco smoke and air pollution affect epigenetic pathways that influence airway function, mucosal immune responses, systemic immune responses, the development of atopic sensitization in children. Generations of genetic factors along with the prenatal impact of maternal inflammatory factors lead to immunological hyper-sensitivity and additional allergic diseases such as asthma, eczema, and

hay fever. Our results show that MIA had an effect on the expression of Asthma pathway, which could be a marker of future development of allergic disorders in the children exposed to COVID-19 infection during gestation period (Shimizu et al., 2023).

Systemic lupus erythematosus (SLE) is an autoimmune disease beginning with genetic predisposition, where certain genetic variations influence immune system functioning, especially in genes related to toll-like receptors (TLRs). Environmental triggers, such as ultraviolet (UV) radiation and infections (e.g., Epstein-Barr virus), can activate the immune system and stimulate autoreactive B cells, leading to increased self-antigen recognition. Hyperactivation of B cells is a hallmark of SLE, resulting in the production of autoantibodies targeting various self-antigens. Dysfunction of regulatory T cells (Tregs) contributes to this B cell activation, as they fail to adequately suppress immune responses.
The formation of immune complexes occurs when these autoantibodies bind to self-antigens, which then deposit in tissues such as the kidneys and skin. This deposition activates the complement system and elicits inflammatory responses, recruiting immune cells like neutrophils and macrophages. These immune cells release reactive oxygen species and cytokines, leading to tissue inflammation and damage (Choi et al., 2012).

Another study examined a cohort of pregnant women who experienced various infections, such as viral infectious (influenza, respiratory syncytial virus (RSV) and cytomegalovirus) or bacterial infections (such as chorioamnionitis and Group B Streptococcus) or parasitic infections (Toxoplasmosis). Researchers collected biological samples from both the mothers and their newborns to analyze changes in gene expression. They utilized techniques like RNA sequencing and microarray analysis to measure the expression levels of genes in the KEGG pathways. Their findings showed that the maternal immune system in response to the infections, releases inflammatory cytokines and chemokines to defense the body against pathogens. These inflammatory mediators such as interleukins such as such IL-1, IL-2, IL-7, IL-10 and tumor necrosis factor-alpha (TNF-α) can cross the placenta influencing fetal development and modulate the expression of various genes in the developing fetal immune system. This exposure can lead to dysregulation of immune responses, creating a predisposition to conditions like asthma or systemic lupus erythematosus. Additionally, the inflammatory environment created by maternal infections may trigger epigenetic modifications, such as DNA methylation and histone

modification. These changes can potentially result in long-lasting effects on the child's immune regulation. Maternal infections can also disrupt normal developmental signaling pathways crucial for the maturation of the immune system in the fetus, leading to improper development of immune cells, such as T-helper cells, which play a significant role in allergic reactions and asthma (Osman et al., 2024; Shimizu et al., 2023; Suleri et al., 2024).

An analysis of the metabolite profiles of 7,219 patients with systemic lupus erythematosus (SLE) and 15,991 control patients was performed. Levels of metabolites related to the KEGG pathway of aminoacyl-tRNA biosynthesis were observed to differ significantly in the case cohort compared to controls that illustrates the role of aminoacyl-tRNA in the function of immune system (Yu et al., 2022).

Aminoacyl tRNA pathway is responsible for regulating the production of Aminoacyl-tRNA synthetases (ARSs). Aminoacyl-tRNA synthetases are enzymes that play a crucial role in protein synthesis by attaching the correct amino acid to its corresponding tRNA molecule. In the context of immune regulation, ARSs can act as signaling molecules, influencing the development, activation, and function of immune cells. For instance, certain ARSs, such as tryptophanyl-tRNA synthetase (WRS), can be secreted by cells and act as danger signals, triggering the release of inflammatory cytokines and type I interferons (IFNs) by immune cells. This response is particularly important in antiviral immunity, as type I IFNs play a critical role in inhibiting viral replication and activating antiviral defenses. In addition to their role in immune regulation, ARSs can also directly contribute to antiviral responses. Some ARSs, such as glutamyl-prolyl tRNA synthetase (EPRS), can be phosphorylated in response to viral infection, leading to the activation of antiviral signaling pathways. This phosphorylation event can enhance the production of type I IFNs and other antiviral factors, ultimately contributing to the control of viral infection. In autoimmune diseases, ARSs are often targeted by the immune system as if they were foreign invaders. This can lead to inflammation and damage to tissues. ARSs can also contribute to the development of autoimmune diseases by influencing immune responses, such as by attracting immune cells to the site of inflammation. Therefore, any alterations in the expression of genes related to aminoacyl-tRNA biosynthesis can contribute to development of any autoimmune diseases (Nie et al., 2019).

Amino acids biosynthesis and metabolism

Proteins are essential components required for the efficient functioning of the immune system, as they play critical roles in processes such as cell signaling, antibody production, and the formation of immune cell structures. These proteins are synthesized from amino acids, and any alterations in amino acid metabolism can lead to immune dysfunction. The COVID-19 infection has been shown to significantly disrupt amino acid biosynthesis, impacting crucial metabolic processes in the human body. Studies have shown that patients with COVID-19 often present altered levels of certain amino acids, such as glutamine, proline, tryptophan, Branched-Chain Amino Acids **(**Valine, leucine, and isoleucine) and tyrosine. These alterations in amino acid concentrations are primarily due to the systemic inflammatory response, metabolic changes associated with the viral infection, and the increased demands on the body's resources to mount an effective immune response. The dysregulation of amino acid levels can impair protein synthesis and other metabolic pathways, contributing to the clinical manifestations of the disease and potentially leading to complications such as muscle wasting and immune dysfunction (Atila et al., 2021; Calder, 2006; Masoodi et al., 2022).

In a study comparing 50 COVID-19 patients and 26 healthy controls, researchers conducted a metabolomic analysis that revealed significant alterations in metabolic pathways related to immune response and inflammation. Key pathways affected levels of cytokines such as interleukin-6 (IL-6) and tumor necrosis factor-alpha (TNF-α) in COVID-19 patients, indicating an exaggerated inflammatory response, which significantly impacts the biosynthesis of several amino acids (B.-W. Li et al., 2021; T. Zhang et al., 2022).

Our study has shown that COVID-19 infection during pregnancy can result in changes to the expression of gene sets associated with amino acid metabolism pathways, potentially impacting the immunologic functions in the newborns' bodies.

Valine, isoleucine and leucine biosynthesis

In the study investigating the effects of maternal infections during pregnancy on the expression of genes linked to the biosynthesis of branched-chain amino acids (BCAAs), a total of 200 pregnant women, who experienced documented viral or bacterial infections (such as influenza and urinary tract infections) during their pregnancy, were included in the cohort analysis. Women

were recruited from prenatal clinics and maternity wards, with a thorough screening process to confirm the presence of infections. The findings of this study indicated that maternal infections led to altered expression of genes involved in the biosynthesis pathways of these essential branched-chain amino acids (BCAAs). The gene expression profiles suggested reduced activity in the pathways responsible for the synthesis of valine, isoleucine, and leucine in the newborns which can have significant implications for the innate immune system of the child through influencing lymphocyte proliferation, metabolism, and response to infection. A deficiency in the biosynthesis of these amino acids may impair the child's immune response and make them more susceptible to infections and possibly autoimmune diseases later in life (Kumar et al., 2022; S. Zhang et al., 2017).

Branched-chain amino acids (BCAAs), which include isoleucine, leucine, and valine, play crucial roles in various physiological processes, particularly in the regulation and functioning of the immune system. **Leucine** plays a crucial role by activating the mTOR (mammalian target of rapamycin) signaling pathway, which is essential for cell growth, proliferation, and metabolism of natural killer (NK) cells of innate immunity, thereby enhancing the immune response to pathogens. Additionally, leucine influences the production of cytokines, key signaling molecules in the regulation of immune system, and enhances the secretion of pro-inflammatory cytokines (e.g., IL-2, TNF-α) from T cells, which facilitates communication among immune cells and strengthens the overall immune attack against infections.

Isoleucine serves as an energy substrate for immune cells, particularly during times of heightened immune activity, helping them respond effectively to pathogens and reduction in its production leads to immune cells dysfunction.

Valine is essential for protein synthesis in immune cells, maintaining the structural integrity and functionality of these cells during an immune response. Valine is also involved in nutrient sensing pathways that regulate metabolic processes in immune cells and helps modulate mTOR signaling, similar to leucine, though its effects may be less direct. By influencing nutrient availability, valine ensures that immune cells function efficiently in response to pathogens. The interaction among isoleucine, leucine, and valine creates a synergistic effect that supports a robust immune response. Adequate levels of all three amino acids are crucial for maintaining

immune homeostasis, aiding in the balance between pro- and anti-inflammatory responses (Allman et al., 2021; Prameswari et al., 2022).

In another study comparing the metabolomic profiles of type 1 diabetes patients with healthy individuals, it was observed that the expression of the KEGG pathway for valine, isoleucine, and leucine was increased, leading to elevated levels of these amino acids. This increase plays a role in exacerbating inflammation through enhanced activity of signaling pathways and the production of inflammatory cytokines, resulting in greater damage to pancreatic beta cells, which are responsible for insulin production. Abnormalities in amino acid levels may affect T-cell activation and differentiation. Since T_1D involves an autoimmune attack on pancreatic beta cells by activated T-cells, any alteration in the metabolism or signaling pathways influenced by these amino acids could potentially impact the autoimmune response (Holeček, 2021; Yu et al., 2022).

Regarding Glycine, serine and threonine metabolism, in a study where pregnant mice were subjected to immune activation, offsprings exhibited significant alterations in the expression profiles of genes associated with the metabolism of glycine, serine, and threonine which impacts the immune system functionality. Serine is crucial for the synthesis of phospholipids and certain neurotransmitters, both of which are essential for the proper functioning of immune cells. Insufficient serine levels can impair the development and activation of T cells and B cells, diminishing the adaptive immune response. Additionally, glycine has been found to exert anti-inflammatory effects and modulate immune responses, acting as an inhibitory neurotransmitter in the central nervous system. An imbalance or reduced availability of glycine can contribute to inflamed states within the body, making individuals more susceptible to autoimmune diseases and chronic inflammatory conditions later in life. This was demonstrated in a study of patients suffering from Crohn's disease, which is a type of inflammatory bowel disease (IBD) causes chronic inflammation of the gastrointestinal tract, and leading to a variety of symptoms and complications. patients with Crohn's disease had different metabolite signatures comparing to healthy subjects due to alteration in the expression of KEGG pathway of glycine, serine and threonine metabolism (Osman et al., 2024; Shimizu et al., 2023; Zhao et al., 2023).

Lysine degradation

In the context of maternal inflammation and its effects on offspring, the role of lysine

degradation, particularly in relation to histone modifications, is significant. A study focused on how Maternal inflammation induced by polyI:C injection in pregnant mice can lead to epigenetic changes, including the hypoa-cetylation of lysine residues on histones. This process involves the removal of acetyl groups from lysine, which typically results in a more compact chromatin structure and reduced gene expression. These epigenetic modifications can influence the expression of genes that are crucial for immune function and neurodevelopment and can lead to decreased activity of genes associated with synaptic transmission and neuronal development, which may subsequently affect proper regulation of immune responses, manifesting as either an increased susceptibility to infections or the development of autoimmune disorders. Additionally, alterations in the production of cytokines, such as interleukin-6, may lead to abnormal cytokine profiles, contributing to chronic inflammation or immune dysregulations (Tang et al., 2013).

Another dysregulated pathway is associated to ABC transporters which are a superfamily of transmembrane proteins that play a crucial role in maintaining cellular homeostasis by transporting a wide variety of substrates across cell membranes. They are essential for various physiological processes, including detoxification, nutrient uptake, and immune defense. One of the key functions of ABC transporters in the immune system is to regulate the influx and efflux of molecules that can trigger or modulate immune responses, such as cytokines and chemokines, out of cells, thereby limiting inflammation. They can also transport antigens into antigen-presenting cells, initiating an immune response.

It has been shown that the expression of related gene sets in patients with active rheumatoid arthritis (RA) is different from healthy controls. ABC transporters have been implicated in the pathogenesis of RA. MDR-ABC transporters, a specific ABC transporter, transport various endobiotics produced by the cells that play important roles in cell proliferation, cell migration, angiogenesis, and inflammation. These processes are all involved in the pathogenesis of RA. For example, ABC transporters can transport inflammatory mediators, such as cytokines and chemokines, out of cells, thereby limiting inflammation. However, in RA, the dysregulation of these transporters can lead to an imbalance in the immune response, contributing to the chronic inflammation characteristic of the disease (Shapiro, 2021).

In another study, researchers investigated the role of P-glycoprotein (P-gp) as an ABC transporter in Multiple Sclerosis (MS), as an autoimmune disease. They used mouse models and found that knockout mice, which lack P-gp, exhibited fewer clinical signs of EAE (Experimental Autoimmune Encephalomyelitis), an animal model of MS, compared to wild-type mice. These findings indicated that P-gp plays a significant role in regulating immune responses and the function of dendritic cells (DCs). Researchers observed that the function of DCs in the knockout mice was severely impaired, leading to reduced T cell responses and decreased brain inflammation. Additionally, they demonstrated that P-gp is essential for the secretion of pro-inflammatory cytokines such as TNF-α and IFN-γ, and that the addition of these cytokines could compensate for the impaired DC functions. These results suggest that ABC transporters, particularly P-gp, can act as key regulators in immune responses and autoimmune diseases like MS (Kooij et al., 2009).

Folate biosynthesis

Folate is implicated in the development of several immune-related diseases. Low levels of folate have been associated with chronic inflammatory conditions, such as inflammatory bowel disease and rheumatoid arthritis. In these diseases, inadequate folate contributes to the pathogenesis by impairing DNA stability and influencing immune cell function. Additionally, folate deficiency can lead to elevated homocysteine levels, which are linked to endothelial dysfunction and increased inflammation, further exacerbating immune-related diseases (Jones et al., 2019).

Glycosaminoglycans biosynthesis

Glycosaminoglycans (GAGs) including keratan sulfate are crucial components of the immune system, playing a vital role in modulating various aspects of immune cell function and inflammation. They act as adhesion molecules for immune cells, facilitating their migration to sites of inflammation and enhancing their recruitment by interacting with chemokines. GAGs act on antigen-presenting cells (APCs) like dendritic cells bind to antigens, facilitating their processing and presentation to T cells, initiating an adaptive immune response. They can also modulate the inflammatory response by interacting with inflammatory mediators like cytokines and chemokines, either promoting or suppressing inflammation depending on the specific GAG

and its interaction with other molecules. GAGs can bind to complement proteins, influencing the activation and regulation of the complement cascade, a crucial part of the innate immune response. Additionally, they contribute to the barrier function of epithelial tissues, preventing the entry of pathogens and maintaining tissue integrity.

In a study examining the role of GAGs in metastatic clear cell renal cell carcinoma (mccRCC), researchers found that the GAG profiles of mccRCC patients were markedly distinct from those of healthy individuals. Specifically, mccRCC patients exhibited an enrichment of keratan sulfate as well as 6-O-sulfated heparan sulfate, in their plasma. This finding suggests a potential link between altered GAG composition and the progression of mccRCC (Gatto et al., 2016).

Glycans metabolism

A study discussed the crucial roles of glycans and glycopeptides in the immune system and immune-related diseases. Glycans are sugar chains attached to proteins, and glycopeptides are combinations of glycans and peptides. Zhang et al. in their article highlighted the changes in glycosylation patterns, particularly in N-glycans and O-glycans, are associated with various immune-related diseases, including autoimmune disorders, leukemia, and infectious diseases. N-glycans, typically attached to asparagine in proteins, are implicated in autoimmune diseases like rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE). In these diseases, reduced levels of galactosyltransferase lead to increased accumulation of glycoforms of IgG (IgG0). These glycoforms, with exposed terminal GlcNAc residues, are recognized by the lectin complement pathway, contributing to inappropriate activation of the innate immune system. This study also illustrated the role of glycans in the innate immune system. Glycans act as pattern recognition receptors (PRRs) that recognize specific glycan structures on pathogens. This recognition triggers the activation of innate immune responses, leading to cytokine production and the recruitment of other immune cells. (X.-L. Zhang, 2006).

In summary, analyzing the expression of KEGG canonical pathways while accounting for innate immune system phenotypes as an outcome resulted in identification of altered expressed pathways contributing to other immune disorders and diseases like allergies, autoimmune diseases and asthma.

4.2 Strengths of the study

In the present study, our emphasis was placed on conducting an analysis at the gene set level, deliberately stepping away from the individual gene analysis traditionally performed by previous authors. This strategy of gene set level analysis presents numerous advantages when it comes to interpreting high-dimensional gene expression datasets. By focusing on specifically defined sets of functionally associated genes rather than examining single genes, gene set analysis facilitates a more biologically relevant interpretation of intricate datasets. This method enables a more expansive perspective by taking into account the collective activity of genes within biological pathways or functional modules. It has the capacity to reveal subtle, yet significantly coordinated shifts in gene expression that might not be captured if one were to analyze individual genes independently. Furthermore, gene set analysis strengthens statistical power by combining information from multiple genes, effectively diminishing noise and the variability that are inherent in high-dimensional datasets.

The importance of gene set level analysis extends to its pivotal role in identifying potential therapeutic targets and preventive approaches for a variety of diseases. By examining the collective behavior of functionally related genes within biological frameworks or processes, this analysis can pinpoint key pathways that exhibit dysregulation in disease contexts. Such dysregulated pathways present viable targets for therapeutic intervention because influencing their activity can lead to the restoration of normal cellular function and may help in curbing disease progression.

Additionally, gene set analysis can elucidate pathways linked to disease risk, including the KEGG pathways proposed in this study, thus contributing to the understanding of possible prevention strategies. A refined comprehension of the molecular mechanisms that play a role in disease development empowers researchers to devise targeted therapies aimed at reestablishing normal pathway activity or preventing disease onset by manipulating pertinent biological processes.

Beyond therapeutic targets, gene set analysis also paves the way for the identification of biomarkers that are related to disease risk or treatment response. This identification can significantly inform personalized treatment approaches and foster the advancement of precision medicine strategies. In essence, gene set level analysis is an invaluable asset in the realms of

drug discovery and development, offering deep insights into disease mechanisms that help steer the creation of innovative therapies and preventive tactics.

A particularly effective approach to performing gene set analysis is through the utilization of the Linear Combination Test (LCT). This method excels in its aptitude for efficiently capturing the inherent geometric structure of the data, allowing for the identification of gene sets that display coordinated expression patterns. LCT stands out as a reliable and robust tool, well-suited for the scrutiny of high-dimensional gene expression data. It provides a robust framework that facilitates the discovery of biologically relevant pathways or processes connected to the observed phenotypes in these datasets.

4.3 Limitations

We used the dataset from Jefferson et al.,2022. In this study, Jefferson et al.,2022 delineate several limitations that merit careful consideration. A primary concern is the small sample size, which constrains the ability to generalize the findings. Furthermore, it is essential to highlight that the study's subjects were not dividing mothers to experiencing mild SARS-CoV-2 infections group and the severe one; thus, the severity of the infection which could potentially manifest different or more pronounced immune genomic responses in the fetus has not been considered in this study.

In addition, the timing of infection during pregnancy, specifically in which week it occurs, can have different effects on gene expression in the fetus body. Within the study cohort, the timeframe between infection and cord blood collection varied considerably, spanning from 5 to 37 weeks. A more uniform approach to timing could lead to more robust findings.

Our research improves upon specific limitations present in the originally published study; however, it is worth to mention that various factors during pregnancy can significantly influence epigenetic modifications which are determinants of over or under-expression of genes and gene sets, thus affecting fetal development and potential long-term health outcomes. These factors encompass aspects such as maternal nutrition, psychological factors, environmental stressors, socio-economic challenges, lifestyle choices, health conditions, advanced maternal age, and fluctuations in the gut microbiome. It is important to note that matching cases and controls

concerning all variables that contribute to epigenetic modulation is not achievable in this study or in other research endeavors, which necessitates a careful interpretation of the findings.

Additionally, the growth and overall development of the fetus are heavily dependent on the maternal environment. The placenta is vital as it connects the mother to the fetus, facilitating essential aspects of fetal development. It provides early life immunity through the transfer of growth factors, essential nutrients, oxygen, pathogen-specific antibodies, and maternal microchimeric immune cells. Notably, the authors of the current study did not implement a strategy to eliminate the influence of maternal components in the analysis of umbilical cord blood mononuclear cells, which could potentially compromise the validity of the results.

Another limitation is the use of the LCT analytical method to obtain gene sets with differential expression. The LCT method has advantages and provides lots of information. Indeed, it cannot separate the gene sets into two groups, where gene sets are over or under-expressed. As a result, we will not be able - by using this analytical method - to predict the 'direction' of expression of the gene set itself, and how the differences in expression of these gene sets were going to allow the immune system to develop in the infant.

In this research, we utilized the KEGG collection, a well-established and extensively utilized resource for biological pathways, to evaluate the changes in canonical biological pathways resulting from maternal COVID-19 in cord blood cells. However, the KEGG database is not the sole provider of biological pathways, and there is no definitive agreement on which database is the most effective. By incorporating and synthesizing data from other databases, we can expand the breadth of our analysis and achieve a more profound understanding of the biological processes pertinent to our area of investigation. Furthermore, utilizing multiple pathway databases facilitates the cross-validation of results, thereby strengthening the validity of our research findings. It is essential to acknowledge that none of these databases may fully encapsulate the true biological mechanisms at play, as our current understanding of these processes is still developing.

4.4 Further studies

The application of 'omics technologies across various platforms—including the genome, transcriptome, proteome, metabolome, microbiome, and lipidome—combined with extensive high-dimensional population databases and specific cohorts, can yield profound insights into the long-term health impacts of maternal COVID-19 on neonates. Investigating both metabolomics and gene expression in tandem can enhance our understanding of health outcomes by offering complementary perspectives on biological processes and pathways. By analyzing the expression patterns of gene sets, researchers can elucidate the regulation and dysregulation of these pathways in the context of health and disease. This knowledge is crucial for identifying key molecular mechanisms that underpin various health conditions and for predicting disease progression or responses to treatment.

Metabolomics, which involves the thorough analysis of small molecules known as metabolites within biological samples, focuses on the end products of cellular pathways that reflect the biochemical activity occurring within cells and tissues. Through the profiling of metabolites, metabolomics captures a snapshot of an organism's metabolic state, revealing insights into its physiological condition, metabolic pathways, and interactions with environmental factors. Variations in metabolite levels can act as biomarkers for diagnosing diseases, predicting outcomes, and monitoring responses to treatment.

Integrating data at the gene set level with metabolomics enables a more holistic understanding of health outcomes. By correlating shifts in gene expression with changes in metabolite profiles, researchers can untangle complex molecular networks and pathways linked to health and disease. This integrated methodology facilitates the identification of biomarkers that possess enhanced diagnostic and prognostic accuracy, and supports the development of targeted therapeutic interventions tailored to the unique metabolic and genetic profiles of individual patients.

Prospective cohorts are particularly valuable as they allow for the longitudinal tracking of offspring health outcomes, utilizing large sample sizes and extended follow-up periods. This design provides the statistical power necessary to uncover significant associations between maternal COVID-19 infection and cardiovascular outcomes in offspring later in life, thereby bolstering the reliability and generalizability of the findings. Furthermore, prospective cohorts enable a thorough evaluation of potential confounding factors and immune-related disease risk

factors, incorporating detailed information regarding lifestyle, medical history, socioeconomic status, and genetic predispositions. This comprehensive approach aids in the identification and adjustment for potential confounders relevant to the research.

Additionally, the consideration of multiple pathway databases enhances the cross-validation of findings, thereby strengthening the robustness of research conclusions. It is essential to acknowledge that no single database may perfectly encapsulate the actual biological mechanisms at play, as our current understanding of these complex systems is still developing.

The assessment of immune disorders stemming from maternal COVID-19 necessitates further validation through both laboratory-based and live subject studies. Employing animal models such as transgenic mice or non-human primates can reinforce the in vivo relevance of the pathways that have been identified.

In summary, a thorough understanding of the complex interactions between maternal infections influences during pregnancy is crucial for pinpointing modifiable risk factors and developing effective interventions to foster optimal fetal development and enduring health. Integrating multi-omics methodologies along with longitudinal studies can significantly contribute to elucidating the mechanisms that drive maternal programming of epigenetic alterations and their subsequent consequences for the health and disease susceptibility of the offspring.

References

Allman, B. R., Spray, B. J., Mercer, K. E., Andres, A., & Børsheim, E. (2021). Markers of branched-chain amino acid catabolism are not affected by exercise training in pregnant women with obesity. *Journal of Applied Physiology*, *130*(3), 651–659. https://doi.org/10.1152/japplphysiol.00673.2020

Allotey, J., Fernandez, S., Bonet, M., Stallings, E., Yap, M., Kew, T., Zhou, D., Coomar, D., Sheikh, J., Lawson, H., Ansari, K., Attarde, S., Littmoden, M., Banjoko, A., Barry, K., Akande, O., Sambamoorthi, D., Van Wely, M., Van Leeuwen, E., … Thangaratinam, S. (2020). Clinical manifestations, risk factors, and maternal and perinatal outcomes of coronavirus disease 2019 in pregnancy: Living systematic review and meta-analysis. *BMJ*, m3320. https://doi.org/10.1136/bmj.m3320

Almond, D., & Mazumder, B. (2005). The 1918 Influenza Pandemic and Subsequent Health Outcomes: An Analysis of SIPP Data. *American Economic Review*, *95*(2), 258–262. https://doi.org/10.1257/000282805774669943

Arora, I., & Tollefsbol, T. O. (2021). Computational methods and next-generation sequencing approaches to analyze epigenetics data: Profiling of methods and applications. *Methods*, *187*, 92–103. https://doi.org/10.1016/j.ymeth.2020.09.008

Atila, A., Alay, H., Yaman, M. E., Akman, T. C., Cadirci, E., Bayrak, B., Celik, S., Atila, N. E., Yaganoglu, A. M., Kadioglu, Y., Halıcı, Z., Parlak, E., & Bayraktutan, Z. (2021). The serum amino acid profile in COVID-19. *Amino Acids*, *53*(10), 1569–1588. https://doi.org/10.1007/s00726-021-03081-w

Bergman, Y., & Cedar, H. (2013). DNA methylation dynamics in health and disease. *Nature Structural & Molecular Biology*, *20*(3), 274–281. https://doi.org/10.1038/nsmb.2518

Burgess, J. K., & Hazelton, R. H. (2000). New developments in the analysis of gene expression. *Redox Report*, *5*(2–3), 63–73. https://doi.org/10.1179/135100000101535348

Calder, P. C. (2006). Branched-Chain Amino Acids and Immunity ,. *The Journal of Nutrition*, *136*(1), 288S-293S. https://doi.org/10.1093/jn/136.1.288S

Chan, J. F.-W., Yuan, S., Kok, K.-H., To, K. K.-W., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C. C.-Y., Poon, R. W.-S., Tsoi, H.-W., Lo, S. K.-F., Chan, K.-H., Poon, V. K.-M., Chan, W.-M., Ip, J. D., Cai, J.-P., Cheng, V. C.-C., Chen, H., … Yuen, K.-Y. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-

person transmission: A study of a family cluster. *The Lancet*, *395*(10223), 514–523. https://doi.org/10.1016/S0140-6736(20)30154-9

Chen, H., Guo, J., Wang, C., Luo, F., Yu, X., Zhang, W., Li, J., Zhao, D., Xu, D., Gong, Q., Liao, J., Yang, H., Hou, W., & Zhang, Y. (2020). Clinical characteristics and intrauterine vertical transmission potential of COVID-19 infection in nine pregnant women: A retrospective review of medical records. *The Lancet*, *395*(10226), 809–815. https://doi.org/10.1016/S0140-6736(20)30360-3

Choi, J., Kim, S. T., & Craft, J. (2012). The Pathogenesis of Systemic Lupus Erythematosus – An Update. *Current Opinion in Immunology*, *24*(6), 651–657. https://doi.org/10.1016/j.coi.2012.10.004

Clerkin, K. J., Fried, J. A., Raikhelkar, J., Sayer, G., Griffin, J. M., Masoumi, A., Jain, S. S., Burkhoff, D., Kumaraiah, D., Rabbani, L., Schwartz, A., & Uriel, N. (2020). COVID-19 and Cardiovascular Disease. *Circulation*, *141*(20), 1648–1655. https://doi.org/10.1161/CIRCULATIONAHA.120.046941

COVID-19 cases | WHO COVID-19 dashboard. (2024.). Datadot. Retrieved June 9, 2024, from https://data.who.int/dashboards/covid19/cases

deAndrés-Galiana, E. J., Fernández-Martínez, J. L., & Sonis, S. T. (2016). Sensitivity analysis of gene ranking methods in phenotype prediction. *Journal of Biomedical Informatics*, *64*, 255–264. https://doi.org/10.1016/j.jbi.2016.10.012

Dinu, I., Moez, E. K., Hajihosseini, M., Leite, A. P., & Pyne, S. (n.d.). *Use of Linear Combination Test to Identify Gene Signatures of Human Embryonic Development in Single Cell RNA-Seq Experiments*. *19*(1).

Dinu, I., Potter, J. D., Mueller, T., Liu, Q., Adewale, A. J., Jhangri, G. S., Einecke, G., Famulski, K. S., Halloran, P., & Yasui, Y. (2008). Gene-set analysis and reduction. *Briefings in Bioinformatics*, *10*(1), 24–34. https://doi.org/10.1093/bib/bbn042

Dinu, I., Wang, X., Kelemen, L. E., Vatanpour, S., & Pyne, S. (2013). Linear combination test for gene set analysis of a continuous phenotype. *BMC Bioinformatics*, *14*(1), 212. https://doi.org/10.1186/1471-2105-14-212

Dopazo, J., Zanders, E., Dragoni, I., Amphlett, G., & Falciani, F. (2001). Methods and approaches in the analysis of gene expression data. *Journal of Immunological Methods*, *250*(1–2), 93–112. https://doi.org/10.1016/S0022-1759(01)00307-6

Drăghici, S. (2011). *Statistics and Data Analysis for Microarrays Using R and Bioconductor, Second Edition*. CRC Press.

Erez, O., Romero, R., Tarca, A. L., Chaiworapongsa, T., Kim, Y. M., Than, N. G., Vaisbuch, E., Draghici, S., & Tromp, G. (2009). Differential expression pattern of genes encoding for anti-microbial peptides in the fetal membranes of patients with spontaneous preterm labor and intact membranes and those with preterm prelabor rupture of the membranes. *The Journal of Maternal-Fetal & Neonatal Medicine*, *22*(12), 1103–1115. https://doi.org/10.3109/14767050902994796

Fehr, A. R., & Perlman, S. (2015). Coronaviruses: An Overview of Their Replication and Pathogenesis. In H. J. Maier, E. Bickerton, & P. Britton (Eds.), *Coronaviruses* (Vol. 1282, pp. 1–23). Springer New York. https://doi.org/10.1007/978-1-4939-2438-7_1

Fong, G., Gayen Nee' Betal, S., Murthy, S., Favara, M., Chan, J. S. Y., Addya, S., Shaffer, T. H., Greenspan, J., Bhandari, V., Li, D., Rahman, I., & Aghai, Z. H. (2020). DNA Methylation Profile in Human Cord Blood Mononuclear Leukocytes From Term Neonates: Effects of Histological Chorioamnionitis. *Frontiers in Pediatrics*, *8*, 437. https://doi.org/10.3389/fped.2020.00437

Gabriel, B., Medin, C., Alves, J., Nduati, R., Bosire, R. K., Wamalwa, D., Farquhar, C., John-Stewart, G., & Lohman-Payne, B. L. (2019). Analysis of the TCR Repertoire in HIV-Exposed but Uninfected Infants. *Scientific Reports*, *9*(1), 11954. https://doi.org/10.1038/s41598-019-48434-4

Gane, S. B. C., Kelly, C., & Hopkins, C. (2020). Isolated Sudden onset anosmia in COVID-19 infection. A novel syndrome? *Rhinology*, *58*. https://doi.org/10.4193/Rhin20.114

Garthwaite, C. (n.d.-a). *The Effect of In-Utero Conditions on Long Term Health: Evidence from the 1918 Spanish Flu Pandemic*.

Garthwaite, C. (n.d.-b). *The Effect of In-Utero Conditions on Long Term Health: Evidence from the 1918 Spanish Flu Pandemic*.

Gatto, F., Volpi, N., Nilsson, H., Nookaew, I., Maruzzo, M., Roma, A., Johansson, M. E., Stierner, U., Lundstam, S., Basso, U., & Nielsen, J. (2016). Glycosaminoglycan Profiling in Patients' Plasma and Urine Predicts the Occurrence of Metastatic Clear Cell Renal Cell Carcinoma. *Cell Reports*, *15*(8), 1822–1836. https://doi.org/10.1016/j.celrep.2016.04.056

Gayen Nee' Betal, S., Murthy, S., Favara, M., Fong, G., Chan, J. S. Y., Addya, S., Shaffer, T. H., Greenspan, J., Bhandari, V., Rahman, I., & Aghai, Z. H. (2019). Histological Chorioamnionitis Induces Differential Gene Expression in Human Cord Blood Mononuclear Leukocytes from Term Neonates. *Scientific Reports*, *9*(1), 5862. https://doi.org/10.1038/s41598-019-42205-x

Gayen Nee' Betal, S., Urday, P., Al-Kouatly, H. B., Solarin, K., Chan, J. S. Y., Addya, S., Boelig, R. C., & Aghai, Z. H. (2022). COVID-19 Infection During Pregnancy Induces Differential Gene Expression in Human Cord Blood Cells From Term Neonates. *Frontiers in Pediatrics*, *10*, 834771. https://doi.org/10.3389/fped.2022.834771

Gee, S., Chandiramani, M., Seow, J., Pollock, E., Modestini, C., Das, A., Tree, T., Doores, K. J., Tribe, R. M., & Gibbons, D. L. (2021a). The legacy of maternal SARS-CoV-2 infection on the immunology of the neonate. *Nature Immunology*, *22*(12), 1490–1502. https://doi.org/10.1038/s41590-021-01049-2

Gee, S., Chandiramani, M., Seow, J., Pollock, E., Modestini, C., Das, A., Tree, T., Doores, K. J., Tribe, R. M., & Gibbons, D. L. (2021b). The legacy of maternal SARS-CoV-2 infection on the immunology of the neonate. *Nature Immunology*, *22*(12), 1490–1502. https://doi.org/10.1038/s41590-021-01049-2

Giraud, A., Chaux, R., Allard, M.-J., Celle, M., Teyssier, G., Roche, F., Chapelle, C., Chabrier, S., Sébire, G., & Patural, H. (2020). Perinatal inflammation is associated with social and motor impairments in preterm children without severe neonatal brain injury. *European Journal of Paediatric Neurology*, *28*, 126–132. https://doi.org/10.1016/j.ejpn.2020.06.008

Guan, W., Ni, Z., Hu, Y., Liang, W., Ou, C., He, J., Liu, L., Shan, H., Lei, C., Hui, D. S. C., Du, B., Li, L., Zeng, G., Yuen, K.-Y., Chen, R., Tang, C., Wang, T., Chen, P., Xiang, J., … for 2019-nCoV, on behalf of C. M. T. E. G. (2020). *Clinical characteristics of 2019 novel coronavirus infection in China* (p. 2020.02.06.20020974). medRxiv. https://doi.org/10.1101/2020.02.06.20020974

Harrison, A. G., Lin, T., & Wang, P. (2020). Mechanisms of SARS-CoV-2 Transmission and Pathogenesis. *Trends in Immunology*, *41*(12), 1100–1115. https://doi.org/10.1016/j.it.2020.10.004

Hill, R. A., Gibbons, A., Han, U., Suwakulsiri, W., Taseska, A., Hammet, F., Southey, M., Malhotra, A., Fahey, M., Palmer, K. R., Hunt, R. W., Lim, I., Newman-Morris, V., &

Sundram, S. (2023). Maternal SARS-CoV-2 exposure alters infant DNA methylation. *Brain, Behavior, & Immunity - Health*, *27*, 100572. https://doi.org/10.1016/j.bbih.2022.100572

Holeček, M. (2021). The role of skeletal muscle in the pathogenesis of altered concentrations of branched-chain amino acids (valine, leucine, and isoleucine) in liver cirrhosis, diabetes, and other diseases. *Physiological Research*, *70*(3), 293–305. https://doi.org/10.33549/physiolres.934648

Hu, Z., Zhou, J., Jiang, J., Yuan, J., Zhang, Y., Wei, X., Loo, N., Wang, Y., Pan, Y., Zhang, T., Zhong, X., Long, M., Montone, K. T., Tanyi, J. L., Fan, Y., Wang, T.-L., Shih, I.-M., Hu, X., & Zhang, L. (2019). Genomic characterization of genes encoding histone acetylation modulator proteins identifies therapeutic targets for cancer treatment. *Nature Communications*, *10*(1), 733. https://doi.org/10.1038/s41467-019-08554-x

Jackson, C. B., Farzan, M., Chen, B., & Choe, H. (2022). Mechanisms of SARS-CoV-2 entry into cells. *Nature Reviews Molecular Cell Biology*, *23*(1), 3–20. https://doi.org/10.1038/s41580-021-00418-x

Jin, H., Liu, J., Cui, M., & Lu, L. (2020). Novel coronavirus pneumonia emergency in Zhuhai: Impact and challenges. *Journal of Hospital Infection*, *104*(4), 452–453. https://doi.org/10.1016/j.jhin.2020.02.005

Jones, P., Lucock, M., Scarlett, C. J., Veysey, M., & Beckett, E. L. (2019). Folate and Inflammation – links between folate and features of inflammatory conditions. *Journal of Nutrition & Intermediary Metabolism*, *18*, 100104. https://doi.org/10.1016/j.jnim.2019.100104

Kamdar, S., Hutchinson, R., Laing, A., Stacey, F., Ansbro, K., Millar, M. R., Costeloe, K., Wade, W. G., Fleming, P., & Gibbons, D. L. (2020). Perinatal inflammation influences but does not arrest rapid immune development in preterm babies. *Nature Communications*, *11*(1), 1284. https://doi.org/10.1038/s41467-020-14923-8

Kazmi, N., Sharp, G. C., Reese, S. E., Vehmeijer, F. O., Lahti, J., Page, C. M., Zhang, W., Rifas-Shiman, S. L., Rezwan, F. I., Simpkin, A. J., Burrows, K., Richardson, T. G., Santos Ferreira, D. L., Fraser, A., Harmon, Q. E., Zhao, S., Jaddoe, V. W. V., Czamara, D., Binder, E. B., … Relton, C. L. (2019). Hypertensive Disorders of Pregnancy and DNA Methylation in Newborns: Findings From the Pregnancy and Childhood Epigenetics

Consortium. *Hypertension*, *74*(2), 375–383. https://doi.org/10.1161/HYPERTENSIONAHA.119.12634

Kerr, B., Leiva, A., Farías, M., Contreras-Duarte, S., Toledo, F., Stolzenbach, F., Silva, L., & Sobrevia, L. (2018). Foetoplacental epigenetic changes associated with maternal metabolic dysfunction. *Placenta*, *69*, 146–152. https://doi.org/10.1016/j.placenta.2018.04.006

Kooij, G., Backer, R., Koning, J. J., Reijerkerk, A., Van Horssen, J., Van Der Pol, S. M. A., Drexhage, J., Schinkel, A., Dijkstra, C. D., Den Haan, J. M. M., Geijtenbeek, T. B. H., & De Vries, H. E. (2009). P-Glycoprotein Acts as an Immunomodulator during Neuroinflammation. *PLoS ONE*, *4*(12), e8212. https://doi.org/10.1371/journal.pone.0008212

Kumar, M., Saadaoui, M., & Al Khodor, S. (2022). Infections and Pregnancy: Effects on Maternal and Child Health. *Frontiers in Cellular and Infection Microbiology*, *12*, 873253. https://doi.org/10.3389/fcimb.2022.873253

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., & Wang, X. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, *581*(7807), 215–220. https://doi.org/10.1038/s41586-020-2180-5

Letko, M., Marzi, A., & Munster, V. (2020). Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nature Microbiology*, *5*(4), 562–569. https://doi.org/10.1038/s41564-020-0688-y

Li, B.-W., Fan, X., Cao, W.-J., Tian, H., Wang, S.-Y., Zhang, J.-Y., Lam, S. M., Song, J.-W., Zhang, C., Zhang, S.-H., Xu, Z., Xu, R.-N., Fu, J.-L., Huang, L., Jiang, T.-J., Shi, M., Wang, F.-S., & Shui, G.-H. (2021). Systematic Discovery and Pathway Analyses of Metabolic Disturbance in COVID-19. *Infectious Diseases & Immunity*, *1*(2), 74–85. https://doi.org/10.1097/ID9.0000000000000010

Li, W., Zhang, C., Sui, J., Kuhn, J. H., Moore, M. J., Luo, S., Wong, S.-K., Huang, I.-C., Xu, K., Vasilieva, N., Murakami, A., He, Y., Marasco, W. A., Guan, Y., Choe, H., & Farzan, M. (2005). Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *The EMBO Journal*, *24*(8), 1634–1643. https://doi.org/10.1038/sj.emboj.7600640

Lin, L., Lu, L., Cao, W., & Li, T. (2020). Hypothesis for potential pathogenesis of SARS-CoV-2 infection–a review of immune changes in patients with viral pneumonia. *Emerging Microbes & Infections*, *9*(1), 727–732. https://doi.org/10.1080/22221751.2020.1746199

Liu, Y.-C., Kuo, R.-L., & Shih, S.-R. (2020). COVID-19: The first documented coronavirus pandemic in history. *Biomedical Journal*, *43*(4), 328–333. https://doi.org/10.1016/j.bj.2020.04.007

Mallah, S. I., Ghorab, O. K., Al-Salmi, S., Abdellatif, O. S., Tharmaratnam, T., Iskandar, M. A., Sefen, J. A. N., Sidhu, P., Atallah, B., El-Lababidi, R., & Al-Qahtani, M. (2021). COVID-19: Breaking down a global health crisis. *Annals of Clinical Microbiology and Antimicrobials*, *20*(1), 35. https://doi.org/10.1186/s12941-021-00438-7

Masoodi, M., Peschka, M., Schmiedel, S., Haddad, M., Frye, M., Maas, C., Lohse, A., Huber, S., Kirchhof, P., Nofer, J.-R., & Renné, T. (2022). Disturbed lipid and amino acid metabolisms in COVID-19 patients. *Journal of Molecular Medicine*, *100*(4), 555–568. https://doi.org/10.1007/s00109-022-02177-4

McAloon, C., Collins, Á., Hunt, K., Barber, A., Byrne, A. W., Butler, F., Casey, M., Griffin, J., Lane, E., McEvoy, D., Wall, P., Green, M., O'Grady, L., & More, S. J. (2020). Incubation period of COVID-19: A rapid systematic review and meta-analysis of observational research. *BMJ Open*, *10*(8), e039652. https://doi.org/10.1136/bmjopen-2020-039652

McCarthy, J., Liu, D., & Kaskel, F. (2021). The Need for Life-Course Study of Children Born to Mothers With Prior COVID-19 Infection. *JAMA Pediatrics*, *175*(11), 1097. https://doi.org/10.1001/jamapediatrics.2021.2423

Moutinho, C., & Esteller, M. (2017). MicroRNAs and Epigenetics. In *Advances in Cancer Research* (Vol. 135, pp. 189–220). Elsevier. https://doi.org/10.1016/bs.acr.2017.06.003

Musa, S. S., Bello, U. M., Zhao, S., Abdullahi, Z. U., Lawan, M. A., & He, D. (2021). Vertical Transmission of SARS-CoV-2: A Systematic Review of Systematic Reviews. *Viruses*, *13*(9), 1877. https://doi.org/10.3390/v13091877

Nam, D., & Kim, S.-Y. (2008). Gene-set approach for expression pattern analysis. *Briefings in Bioinformatics*, *9*(3), 189–197. https://doi.org/10.1093/bib/bbn001

Nie, A., Sun, B., Fu, Z., & Yu, D. (2019). Roles of aminoacyl-tRNA synthetases in immune regulation and immune diseases. *Cell Death & Disease*, *10*(12), 901. https://doi.org/10.1038/s41419-019-2145-5

Osman, H. C., Moreno, R., Rose, D., Rowland, M. E., Ciernia, A. V., & Ashwood, P. (2024). Impact of maternal immune activation and sex on placental and fetal brain cytokine and gene expression profiles in a preclinical model of neurodevelopmental disorders. *Journal of Neuroinflammation*, *21*(1), 118. https://doi.org/10.1186/s12974-024-03106-7

Patanè, L., Morotti, D., Giunta, M. R., Sigismondi, C., Piccoli, M. G., Frigerio, L., Mangili, G., Arosio, M., & Cornolti, G. (2020). Vertical transmission of coronavirus disease 2019: Severe acute respiratory syndrome coronavirus 2 RNA on the fetal side of the placenta in pregnancies with coronavirus disease 2019–positive mothers and neonates at birth. *American Journal of Obstetrics & Gynecology MFM*, *2*(3), 100145. https://doi.org/10.1016/j.ajogmf.2020.100145

Prameswari, N., Irwinda, R., Wibowo, N., & Saroyo, Y. B. (2022). Maternal Amino Acid Status in Severe Preeclampsia: A Cross-Sectional Study. *Nutrients*, *14*(5), 1019. https://doi.org/10.3390/nu14051019

Prochaska, E., Jang, M., & Burd, I. (2020). COVID-19 in pregnancy: Placental and neonatal involvement. *American Journal of Reproductive Immunology*, *84*(5), e13306. https://doi.org/10.1111/aji.13306

Rao, Y., Lee, Y., Jarjoura, D., Ruppert, A. S., Liu, C., Hsu, J. C., & Hagan, J. P. (2008). A Comparison of Normalization Techniques for MicroRNA Microarray Data. *Statistical Applications in Genetics and Molecular Biology*, *7*(1). https://doi.org/10.2202/1544-6115.1287

Rasmussen, S. A., Jamieson, D. J., & Uyeki, T. M. (2012). Effects of influenza on pregnant women and infants. *American Journal of Obstetrics and Gynecology*, *207*(3), S3–S8. https://doi.org/10.1016/j.ajog.2012.06.068

Santos, K. F., Mazzola, T. N., & Carvalho, H. F. (2005). The prima donna of epigenetics: The regulation of gene expression by DNA methylation. *Brazilian Journal of Medical and Biological Research*, *38*(10), 1531–1541. https://doi.org/10.1590/S0100-879X2005001000010

Shapiro, S. C. (2021). Biomarkers in Rheumatoid Arthritis. *Cureus*. https://doi.org/10.7759/cureus.15063

Sharps, M. C., Hayes, D. J. L., Lee, S., Zou, Z., Brady, C. A., Almoghrabi, Y., Kerby, A., Tamber, K. K., Jones, C. J., Adams Waldorf, K. M., & Heazell, A. E. P. (2020). A structured

review of placental morphology and histopathological lesions associated with SARS-CoV-2 infection. *Placenta*, *101*, 13–29. https://doi.org/10.1016/j.placenta.2020.08.018

Shimizu, Y., Sakata-Haga, H., Saikawa, Y., & Hatta, T. (2023). Influence of Immune System Abnormalities Caused by Maternal Immune Activation in the Postnatal Period. *Cells*, *12*(5), 741. https://doi.org/10.3390/cells12050741

Shulla, A., Heald-Sargent, T., Subramanya, G., Zhao, J., Perlman, S., & Gallagher, T. (2011). A Transmembrane Serine Protease Is Linked to the Severe Acute Respiratory Syndrome Coronavirus Receptor and Activates Virus Entry. *Journal of Virology*, *85*(2), 873–882. https://doi.org/10.1128/JVI.02062-10

Siegel, D. A., Reses, H. E., Cool, A. J., Shapiro, C. N., Hsu, J., Boehmer, T. K., Cornwell, C. R., Gray, E. B., Henley, S. J., Lochner, K., Suthar, A. B., Lyons, B. C., Mattocks, L., Hartnett, K., Adjemian, J., Van Santen, K. L., Sheppard, M., Soetebier, K. A., MAPW1, … Raizes, E. (2021). Trends in COVID-19 Cases, Emergency Department Visits, and Hospital Admissions Among Children and Adolescents Aged 0–17 Years—United States, August 2020–August 2021. *MMWR. Morbidity and Mortality Weekly Report*, *70*(36), 1249–1254. https://doi.org/10.15585/mmwr.mm7036e1

Song, S., & Black, M. A. (2008). Microarray-based gene set analysis: A comparison of current methods. *BMC Bioinformatics*, *9*(1), 502. https://doi.org/10.1186/1471-2105-9-502

Suleri, A., Salontaji, K., Luo, M., Neumann, A., Mulder, R. H., Tiemeier, H., Felix, J. F., Marioni, R. E., Bergink, V., & Cecil, C. A. M. (2024). Prenatal exposure to common infections and newborn DNA methylation: A prospective, population-based study. *Brain, Behavior, and Immunity*, *121*, 244–256. https://doi.org/10.1016/j.bbi.2024.07.046

Tang, B., Jia, H., Kast, R. J., & Thomas, E. A. (2013). Epigenetic changes at gene promoters in response to immune activation in utero. *Brain, Behavior, and Immunity*, *30*, 168–175. https://doi.org/10.1016/j.bbi.2013.01.086

Taylor, J., Fairburn, H., Beaujean, N., Meehan, R., & Young, L. (2019). Gene expression in the developing embryo and fetus. *Bioscientifica Proceedings*. https://doi.org/10.1530/biosciprocs.5.012

Urday, P., Gayen Nee' Betal, S., Sequeira Gomes, R., Al-Kouatly, H. B., Solarin, K., Chan, J. S., Li, D., Rahman, I., Addya, S., Boelig, R. C., & Aghai, Z. H. (2023). SARS-CoV-2 Covid-19 Infection During Pregnancy and Differential DNA Methylation in Human Cord Blood

Cells From Term Neonates. *Epigenetics Insights*, *16*, 25168657231184665. https://doi.org/10.1177/25168657231184665

van Doremalen Neeltje, Bushmaker Trenton, Morris Dylan H., Holbrook Myndi G., Gamble Amandine, Williamson Brandi N., Tamin Azaibi, Harcourt Jennifer L., Thornburg Natalie J., Gerber Susan I., Lloyd-Smith James O., de Wit Emmie, & Munster Vincent J. (2020). Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. *New England Journal of Medicine*, *382*(16), 1564–1567. https://doi.org/10.1056/NEJMc2004973

Wang, X., Pyne, S., & Dinu, I. (2014). Gene set enrichment analysis for multiple continuous phenotypes. *BMC Bioinformatics*, *15*(1), 260. https://doi.org/10.1186/1471-2105-15-260

Weitkamp, J.-H., Guthrie, S. O., Wong, H. R., Moldawer, L. L., Baker, H. V., & Wynn, J. L. (2016). Histological chorioamnionitis shapes the neonatal transcriptomic immune response. *Early Human Development*, *98*, 1–6. https://doi.org/10.1016/j.earlhumdev.2016.06.001

Wu, Z., & McGoogan, J. M. (2020). Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA*, *323*(13), 1239. https://doi.org/10.1001/jama.2020.2648

Yu, X.-H., Cao, R.-R., Yang, Y.-Q., & Lei, S.-F. (2022). Identification of causal metabolites related to multiple autoimmune diseases. *Human Molecular Genetics*, *31*(4), 604–613. https://doi.org/10.1093/hmg/ddab273

Zahedipour, F., Hosseini, S. A., Sathyapalan, T., Majeed, M., Jamialahmadi, T., Al-Rasadi, K., Banach, M., & Sahebkar, A. (2020). Potential effects of curcumin in the treatment of COVID -19 infection. *Phytotherapy Research*, *34*(11), 2911–2920. https://doi.org/10.1002/ptr.6738

Zhang, S., Zeng, X., Ren, M., Mao, X., & Qiao, S. (2017). Novel metabolic and physiological functions of branched chain amino acids: A review. *Journal of Animal Science and Biotechnology*, *8*(1), 10. https://doi.org/10.1186/s40104-016-0139-z

Zhang, T., Abdelrahman, Z., Liu, Q., Wang, X., & Chen, Z. (2022). Comparisons of the immunological landscape of COVID-19 patients based on sex and disease severity by

multi-omics analysis. *Chemico-Biological Interactions*, *352*, 109777. https://doi.org/10.1016/j.cbi.2021.109777

Zhang, X.-L. (2006). Roles of Glycans and Glycopeptides in Immune System and Immune-Related Diseases. *Current Medicinal Chemistry*, *13*(10), 1141–1147. https://doi.org/10.2174/092986706776360897

Zhang, Y.-Z., & Holmes, E. C. (2020). A Genomic Perspective on the Origin and Emergence of SARS-CoV-2. *Cell*, *181*(2), 223–227. https://doi.org/10.1016/j.cell.2020.03.035

Zhao, J., Stewart, I. D., Baird, D., Mason, D., Wright, J., Zheng, J., Gaunt, T. R., Evans, D. M., Freathy, R. M., Langenberg, C., Warrington, N. M., Lawlor, D. A., & Borges, M. C. (2023). Causal effects of maternal circulating amino acids on offspring birthweight: A Mendelian randomisation study. *eBioMedicine*, *88*, 104441. https://doi.org/10.1016/j.ebiom.2023.104441

Zhou, T., Tsybovsky, Y., Gorman, J., Rapp, M., Cerutti, G., Chuang, G.-Y., Katsamba, P. S., Sampson, J. M., Schön, A., Bimela, J., Boyington, J. C., Nazzari, A., Olia, A. S., Shi, W., Sastry, M., Stephens, T., Stuckey, J., Teng, I.-T., Wang, P., … Kwong, P. D. (2020). Cryo-EM Structures of SARS-CoV-2 Spike without and with ACE2 Reveal a pH-Dependent Switch to Mediate Endosomal Positioning of Receptor-Binding Domains. *Cell Host & Microbe*, *28*(6), 867-879.e5. https://doi.org/10.1016/j.chom.2020.11.004