USING HUMAN GENETICALLY ENGINEERED STEM CELL *IN VITRO* MODELS TO REVEAL NEW INSIGHTS INTO CROHN'S DISEASE PATHOGENESIS

By

BOPING DENG

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY



School of Computer Science
College of Engineering and Physical Sciences
University of Birmingham
January 2022

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

Crohn's disease (CD) is becoming a global challenge to public health. Unfortunately, the pathogenesis of CD is still unclear, and mouse models are not optimal in studying this complex human disease. CRISPR-Cas9-based genome editing was combined with *in vitro* differentiation of human induced pluripotent stem cells (iPSCs) to generate models for exploring CD pathogenesis. Two CD risk-associated genes *NOD2* (a pattern-recognition receptor) and *LRRK2* (a large protein of undefined function that has been implicated in a wide range of biological processes) were knocked out in iPSCs from a healthy donor, and the cells were differentiated into isogenic human macrophages and intestinal organoids *in vitro*. RNA sequencing was used to explore the biological changes caused by the knockouts. Findings from the research indicated that the loss of *NOD2* led to a large defect in the response to interferon-gamma (IFN-γ) of the macrophages, whilst loss of *LRRK2* did not, though the gene was found to be significantly up regulated in macrophages treated by IFN-γ. Meanwhile, the effects from *LRRK2* deficiency were identified to focus on mitochondria in macrophages. Moreover, this work also revealed a correlation between *NOD2* (the major susceptible gene for Crohn's disease) and *LRRK2* (the major susceptible gene for Parkinson's disease (PD)) expression in macrophages treated by IFN-γ, which suggested a shared pathogenesis between CD and PD. Additionally, a pipeline for preparing single cell RNA-seq samples from CD patients' resections was established to compare those *in vitro* models with the *in vivo* situation of the disease. Monocytes/macrophages cell populations were identified in single cell RNA-seq data, which showed the subtle differences between resections from the same patient but with agreed with the IFN-g effect seen in *in vitro* macrophages. Data from CD patient's resections also showed an altered IFN-γ response and additional effects on mitochondrial

gene expression were associated with the severity of inflammation in macrophages. This finding from patients' samples supports the clinical relevance of the *in vitro* macrophage models.

Acknowledgement

As a student of the Joint PhD Training Program offered by the University of Birmingham (UoB) and Southern University of Science and Technology (SUSTech), I would like to express my heartfelt gratitude to my supervisors, Prof. Jean-Baptiste Cazier, Prof. Jonathan Frampton, Prof. Jonathan Rowe in UoB and Prof. Andrew Hutchins in SUSTech, for their supervision as well as to UoB and SUSTech for their support for my studies in both universities. Without their help and guidance, it is impossible for me to finish my PhD study.

I am also very grateful for the help from Prof. Andrew Beggs, Prof. Subrata Ghosh and all their group members on my resections' experiments and single cell RNA-sequencing work in UoB. Meanwhile, I want to thank Dr. Dongye Wang from GIBH for sharing stem cell lines with me and guiding me on macrophages' work. I am also very appreciated to Dr. Sally Cowley (University of Oxford) and Dr. Yuhan Hao (New York Genome Center), for their guidance on in vitro differentiation of macrophages and the usage of R package "Seurat" respectively.

I want to say thanks to Prof. Chris Tselepis and Dr. Mark Jeeves for their help on my annual review in UoB. And thanks a lot to everyone from Centre for Computational Biology (CCB) in UoB and Prof. Andrew Hutchins' lab in SUSTech.

Finally, I want to appreciate my parents, for their continuous supports these years. I also wish to show my great gratitude to my girlfriend, Dr. Yuwei Cong, for her long-time company and understandings on my difficulties.

List of Abbreviations

CD	Crohn’s disease
IBD	inflammatory bowel disease
UC	ulcerative colitis
NOD2	Nucleotide-binding oligomerization domain-containing protein 2
MDP	Muramyl Dipeptide
MAPK	mitogen-activated protein kinase
NF-κB	nuclear factor kappa-B
LRRK2	Leucine-rich repeat kinase 2 protein
PD	Parkinson’s disease
PBMCs	peripheral blood mononuclear cells
GWAS	Genome wide association studies
α-syn	α-synuclein
GI tract	gastrointestinal tract
CNS	central nervous system
LPS	Lipopolysaccharides
DSS	dextran sodium sulfate
IFN-γ	interferon gamma
hESCs	human embryonic stem cells
hiPSCs	human induced pluripotent stem cells
hASCs	human adult stem cells
CRISPR	Clustered regularly interspaced short-palindromic repeat
crRNAs	CRISPR RNAs
sgRNA	single-guide RNA
tracrRNA	trans-activating crRNA
PAM	protospacer adjacent motif
NHEJ	non-homologous end joining

HDR	homology-directed repair
DSBs	double-strand breaks
HR	homologous recombination
ZFNs	zinc-finger nucleases
TALENs	transcription activator-like effector nucleases
RNA-seq	RNA sequencing
DGE	differential gene expression
GSE	gene set enrichment
ORA	over-representation analysis
FCS	functional class scoring
GSEA	Gene Set Enrichment Analysis
EB	embryoid body
°C	degrees Celsius
min	minute
s	second
g	relative centrifugal force (RCF) for centrifuge

Contents

INTRODUCTION 1

1.1 Crohn's disease - a global challenge to public health 2

1.2 Current understandings on Crohn's disease pathogenesis 6

1.3 *NOD2* and Crohn's disease 9

1.4 *LRRK2* may be the link between Crohn's disease and Parkinson's disease 12

1.5 Macrophages play an important role in Crohn's disease 15

1.6 Limitations of mouse models in mimicking Crohn's disease 18

1.7 Human induced pluripotent stem cells can contribute to the *in vitro* studies of Crohn's disease 21

1.8 CRISPR-Cas9 provides capabilities for studying genetic disorders 25

1.9 RNA sequencing is a useful tool for sensing CD-related changes in response to phenotypes 29

1.10 Objectives and Aims 35

MATERIALS AND METHODS 37

2.1 Maintenance of human iPS cells 38

2.2 Mycoplasma test 39

2.3 *In vitro* differentiation of macrophages 40

2.4 *In vitro* differentiation and maintenance of intestinal organoids from hiPSCs 43

2.5 Western Blot 46

2.6 RNA extraction and sequencing 49

2.7 Analysis of bulk RNA-seq data 51

2.8 Key materials and resources 53

2.9 Ethical approval statement 59

***IN VITRO* STUDIES OF THE RELATIONSHIP BETWEEN *NOD2* DEFICIENCY AND CD PATHOGENESIS** 60

3.1 Background 61

3.2 Aims 63

3.3 Materials and Methods 64

3.3.1 Generating *NOD2* knockout hiPSCs via CRISPR-Cas9 64

3.3.2 Using immunofluorescence to verify pluripotency of the generated *NOD2* knockout hiPSCs 67

3.3.3 Karyotyping analysis of the generated NOD2 knockout hiPSCs....68
3.3.4 Characterization of hiPSC-derived macrophages....69
3.4 Results....70
3.4.1 Generation of *NOD2* knockout hiPSCs....70
3.4.2 *In vitro* differentiation of *NOD2* deficient macrophages....76
3.4.3 Untreated *NOD2* deficient macrophages may have primed down regulations in IFN-γ and LPS mediated inflammatory response....79
3.4.4 *NOD2* deficiency has surprisingly little influence on MDP-mediated inflammatory response in macrophages....90
3.4.5 *NOD2* deficiency down regulates macrophages response to IFN-γ....94
3.4.6 *NOD2* deficiency has few effects on macrophages response to LPS....107
3.4.7 *In vitro* differentiation of *NOD2* deficient intestinal organoids....115
3.4.8 *NOD2* deficiency has few effects on the IFN-γ response in intestinal organoids....118
3.5 Discussion....125
IN VITRO STUDY OF THE RELATIONSHIP BETWEEN _LRRK2_ DEFICIENCY AND CD PATHOGENESIS....131
4.1 Background....132
4.2 Aims....134
4.3 Materials and Methods....135
4.4 Results....136
4.4.1 Co-expression of *LRRK2* and *NOD2* may exist in *Homo sapiens*....136
4.4.2 *NOD2* deficiency significantly decreases *LRRK2* expression in IFN-γ treated hiPSC-derived macrophages....138
4.4.3 *In vitro* differentiation of *LRRK2* deficient macrophages....142
4.4.4 IFN-γ response is mainly intact in *LRRK2* deficient macrophages....144
4.4.5 *LRRK2* deficiency has no effect on increased *NOD2* expression in IFN-γ treated hiPSC-derived macrophages....154
4.4.6 *LRRK2* deficiency affects oxidative phosphorylation via mitochondria in IFN-γ treated macrophages....156
4.4.7 *LRRK2* deficiency leads to a heightened inflammatory response in LPS treated macrophages via interferon response....162
4.5 Discussion....171

SINGLE CELL ANALYSIS AND ORGANOIDS GENERATION FROM SURGICAL RESECTIONS HELP REVEAL CHARACTERISTICS OF MONOCYTES/MACROPHAGES FROM CD PATIENTS 176

5.1 Background 177

5.2 Aims 179

5.3 Materials and Methods 180

5.3.1 Single cell RNA-seq of resections from CD patients 180

5.3.2 Generating intestinal organoids from CD patients' resections 183

5.4 Results 184

5.4.1 Preparing samples for single cell RNA sequencing from CD resections 184

5.4.2 Identification of monocytes/macrophages from CD resections 187

5.4.3 GSEA identifies subtle differences among monocytes/macrophages responses in CD patients 195

5.4.4 GSEA helps to link the *in vitro* models with the *in vivo* condition of CD patients 198

5.4.5 Analysis of single cell RNA-seq data from CD patients' resections with severe and non-severe inflammatory pairs 202

5.4.6 Generating intestinal organoids from CD patients' resections 206

5.5 Discussion 209

GENERAL DISCUSSION 212

6.1 Summary 213

6.2 Limitations 214

6.3 Future work 217

List of References 219

Appendix 237

CHAPTER 1

INTRODUCTION

1.1 Crohn’s disease - a global challenge to public health

Crohn’s disease (CD) is a type of chronic inflammatory bowel disease (IBD), which can influence any part of the gastrointestinal tract from the mouth to the anus ("Crohn’s & Ulcerative Colitis: Know the Difference?," 2018; Waugh et al., 2013). CD is also characterized as having skip intestinal lesions, which are areas of inflamed mucosa interposed with normal-appearing tissue (Roda et al., 2020). These characteristics differentiate CD from another main type of IBD - ulcerative colitis (UC), which only affects the inner lining of the colon, with continuous lesions and superficial inflammation ("Crohn’s & Ulcerative Colitis: Know the Difference?," 2018; Roda et al., 2020; Waugh et al., 2013).

CD has long been regarded mainly as a western disease (Behzadi, Behzadi, & Ranjbar, 2015; Ng et al., 2017). After its initial recognition in 1932, CD’s incidence in the developed western world increased substantially, and recent investigations indicated the prevalence of CD had reached 0.3% of the population in many European, Oceanian and North American countries (e.g. 322 per 100,000 in Germany; 319 per 100,000 in Canada) (Ng et al., 2017). Additionally, CD seems to have a close relationship with industrialization. Although the prevalence of CD in western countries is still the highest around the world (Figure 1.1.1), it seems the incidence in these industrialized regions has reached a peak and has begun to stabilize (Ng et al., 2017). However, in newly industrialized or industrializing countries (mostly in Africa, Asia and South America), CD incidence has started to rise (Figure 1.1.2) (Ng et al., 2017).

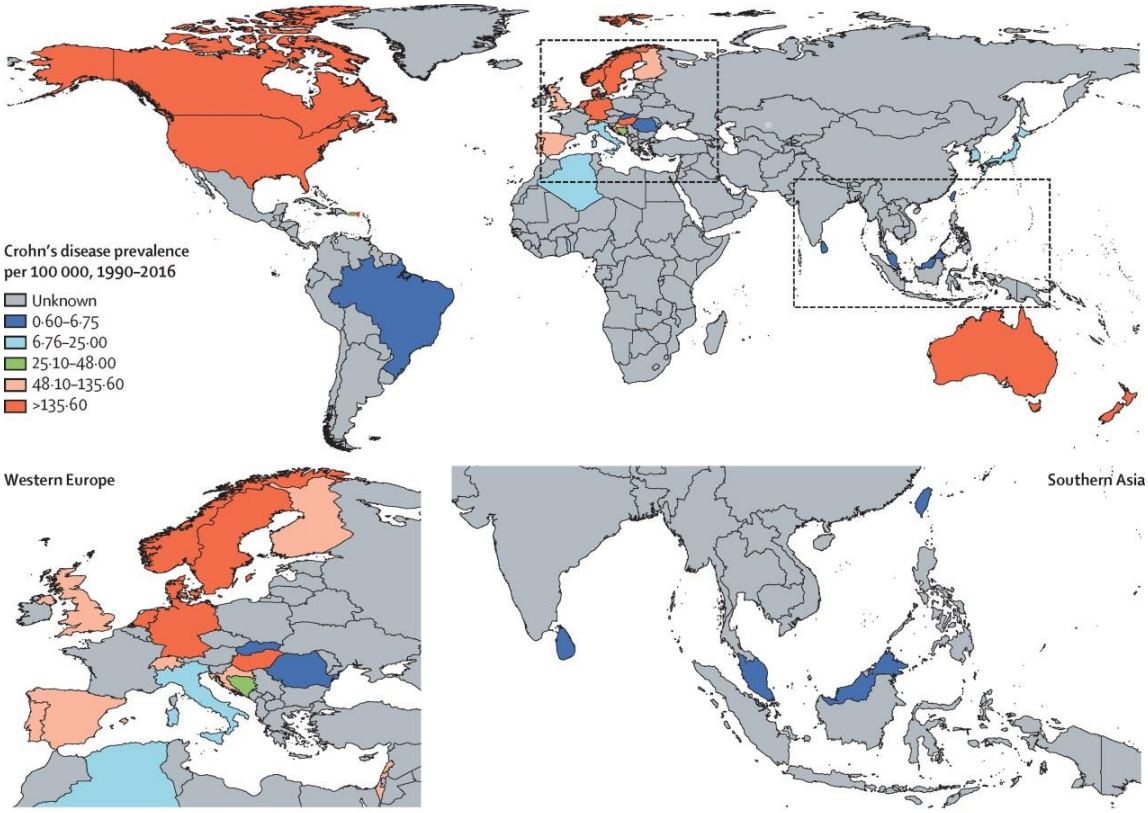


Figure 1.1.1 Worldwide Crohn's disease prevalence from 1990 to 2016

(Adjusted from (Ng et al., 2017))

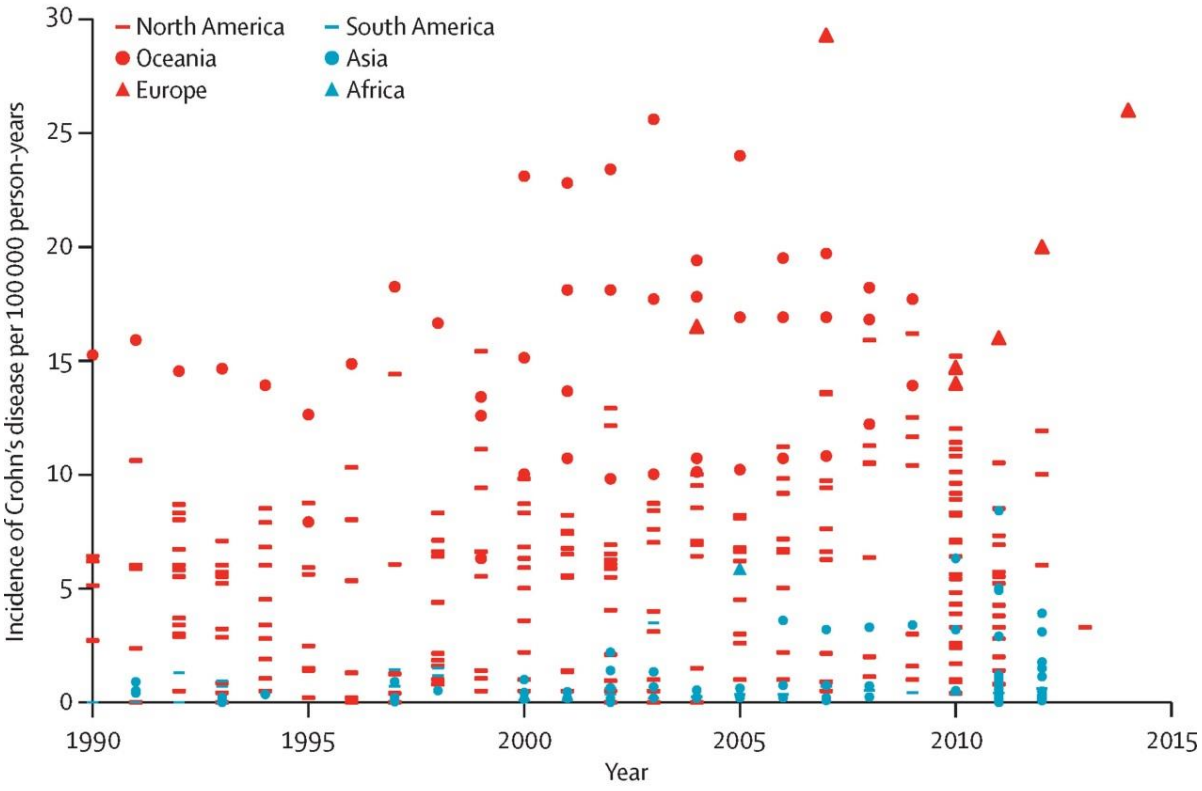


Figure 1.1.2 Worldwide Crohn's disease incidence from 1990 to 2016
(Adjusted from (Ng et al., 2017))

This increasing number of CD patients is a great burden for the healthcare system, and continuously cost expenditure for public health (Ng et al., 2017). Unfortunately, there are no approaches to cure CD at present and only palliative treatments are available. The onset of CD is around early adulthood (15-35 years old), and due to the characteristics of CD, patients can require long-term (in many cases life-long) treatment to prevent disease flares and progression, which means continuous and substantial costs for public health expenditure (Ng et al., 2017; Roda et al., 2020). Meanwhile, whilst life expectancy of IBD patients has increased because of developments of the treatments, it is still lower comparing with the healthy controls (e.g. female with IBD: 78.4; female without IBD: 86.1) (Kuenzig, Manuel, Donelle, & Benchimol, 2020). Moreover, the gap is still large on the health-adjusted life

expectancy (by evaluations of general well-being and the effect of pain) between people with and without IBD (Kuenzig et al., 2020). The serious gap can also be reflected by the complications related to CD, which include cirrhosis, portal hypertension, cholangiocarcinoma and colon cancer. These problems, in addition to being life threatening, can strongly affect a patients' quality of life (Roda et al., 2020). According to surveys, up to 50% of CD patients require surgeries within 10 years of diagnosis because of intestinal complications, and medicine is still needed even after intestinal resections (Bernstein, 2012; Roda et al., 2020). Even worse, surgeries only have a limited efficacy for CD: A third of CD patients require multiple surgeries during their life-time, and recurrence is 30% amongst those patients who have undergone operations more than once (Torres, Mehandru, Colombel, & Peyrin-Biroulet, 2017).

There is no doubt CD has become a huge challenge for worldwide public health, especially for those newly industrialized countries, which are possibly undergoing the early stages of the pattern that occurred first in the western world (Ng et al., 2017). To tackle this challenge, a priority is to pursue a better understanding of CD pathogenesis.

1.2 Current understandings on Crohn's disease pathogenesis

Unfortunately, pathogenesis of CD remains unclear, but it is widely accepted that genetic, environmental and immunological factors contribute to the risk of CD onset and progression (Roda et al., 2020; Torres et al., 2017).

Genetic variations are the basis of many human disease susceptibilities, including CD. Since the first coding variation (*NOD2* gene) related to CD risk was recognized in 2001, genome-wide association studies (GWAS) have identified over 200 CD risk-associated loci from the analysis of more than 70,000 individuals (Roda et al., 2020). A common theme has emerged from the CD risk-associated genes and loci as the corresponding genes are related to bacterial sensing, innate immunity, Th17-cell functions and intestinal mucus layer, and these help form insights into CD pathogenesis (Torres et al., 2017). However, due to the complex variances and phenotypes for CD, only 13.1% of the disease heritability is explained by genetic variations, which highlights the importance of epigenetic and other non-genetic environmental factors in disease pathogenesis (Torres et al., 2017).

Environmental factors associated with CD cover extremely wide aspects of daily life, including, but not limited to, cigarette smoking, oral contraceptives, aspirin, non-steroidal anti-inflammatory drugs, a westernized diet and its effects on microbiota diversity, extracellular environments and many others (Roda et al., 2020; Torres et al., 2017). Their effects have been well verified by studies on immigrants from low CD to high CD prevalence regions. With no changes in genetic factors, CD risk among those individuals still increases once they enter a high CD prevalence region, with a more obvious trend amongst the children of those immigrants or those who move to the CD-high environments early in life (Ng et al., 2017; Roda et al., 2020).

The immunological factors in CD pathogenesis are the downstream responses of

environmental factors, and act in relation to intestinal homeostasis, which include defects in intestinal epithelial barrier function, dysregulation of intestinal innate and adaptive immune systems (Torres et al., 2017). For instance, a general model for CD pathogenesis is for genetic and environmental factors to contribute to the disruption of the epithelial barrier, leading to luminal contents entering the lamina propria, which results in dysregulated immune responses (Roda et al., 2020; Torres et al., 2017). As most characterized effects from genetic variations and environmental factors are related to immune dysregulations, tracing the abnormalities of immune system responses with higher resolutions may be a key to reveal CD pathogenesis.

In addition, it needs to be noted that CD risk has an obvious relationship with ethnic groups as well as nations or populations. For example, surveys on worldwide populations report that Ashkenazi Jews have three-to-four times higher risk of CD, whilst African-American and Asian ancestries have the lowest risk (Torres et al., 2017). Effects of CD risk factors among different populations are also distinct. For instance, *NOD2* variants are quite common in CD patients from western countries (around 30% to 40% of CD patients had mutations in the *NOD2* gene), and identified as the strongest genetic risk association for the disease, however, those variants are not common amongst CD patients from East Asia (e.g. China, Japan), and their association with the disease were not recognized from surveys (Leong et al., 2003; Yamazaki, Takazoe, Tanaka, Kazumori, & Nakamura, 2002). However, there are hints that this may be a sampling issue due to lower numbers of CD patients and reduced systematic surveys, as a different set of mutations in *NOD2* have been positively associated with CD in some Asian populations (Ng et al., 2012). A similar thing also happens for CD associated environmental factors. As the best studied and the most modifiable environmental factor, smoking is recognized to double the risk to develop CD, and even passive smoking results in a similar effect (Roda et al., 2020; Torres et al., 2017). However, it

seems to be distinct among different populations. For instance, smoking rates in Asia and Africa are much higher than some northern European countries, but CD incidence is still lower (Torres et al., 2017). In this situation, it is necessary to take population genetic background into consideration when studies on CD pathogenesis are projected.

1.3 *NOD2* and Crohn's disease

The *NOD2* (Nucleotide-binding oligomerization domain-containing protein 2) gene, also known as *CARD15* (caspase recruitment domain-containing protein 15) and *IBD1* (inflammatory bowel disease protein 1), was the first identified susceptible gene for CD (Balasubramanian & Gao, 2017). Mutations of *NOD2* are widely distributed in CD patients from western countries (around 30% to 40%), which makes it an attractive target for the field to understand CD pathogenesis.

Associations between *NOD2* and CD risk seem somewhat paradoxical. Before *NOD2* was identified as the susceptible gene for CD, *NOD2* protein was recognized as an intracellular microbial sensor for a peptidoglycan motif called muramyl dipeptide (MDP), which is widely present in the cell walls of gram-positive and gram-negative bacteria (Eckmann & Karin, 2005). After sensing MDP, the activated *NOD2* will trigger a series of *in vivo* reactions, leading to the activation of MAPKs and NF-κB pathways. This induces lots of inflammation related biological processes such as the secretion of pro-inflammatory cytokines and autophagy (Al Nabhani, Dietrich, Hugot, & Barreau, 2017; Eckmann & Karin, 2005). Based on this, loss of function for *NOD2* was initially expected to relieve inflammation and protect against CD (Plevy, 2021). However, in reality, mutations of the gene increase the risk of CD, data derived from a large amount of clinical evidence. In this circumstance, how *NOD2* dysfunctions influence overall intestinal homeostasis becomes the key in studying *NOD2* relationships with CD pathogenesis (Plevy, 2021).

Explorations on how *NOD2* dysfunctions affect intestinal homeostasis are based around the MDP-mediated response, since *NOD2* is the only described receptor for MDP. Based on NOD-mediated the MDP-mediated response, several hypotheses have been proposed to explain why mutations in *NOD2* can increase CD risk significantly, supported by

corresponding experimental evidence. Briefly, the loss of function of *NOD2* is considered to lead to a failure to recognize MDP from bacteria and thus lead to a failure to activate the immune defense mechanisms, leading to accumulation of bacteria and an increased burden on the intestinal immune system (Eckmann & Karin, 2005). Meanwhile, mutated *NOD2* may enhance the influence from other inflammation related pathways, as MDP-mediated response via *NOD2* has been found to be capable of inhibiting other inflammatory pathways (e.g. TLRs-mediated ones) (Eckmann & Karin, 2005). Furthermore, *NOD2* variants may cause changes in cytokines' secretion when the immune cells are challenged by MDP stimulation, which interferes with the normal regulation of an inflammatory response (Eckmann & Karin, 2005). These dysregulations potentially combine and finally cause chronic mucosal inflammation *in vivo*. Figure 1.3.1 summarizes the *NOD2* function and its interactions with MDP intracellular. .

Recently, more and more evidence indicates that *NOD2* associations with CD risk are not limited to an MDP-mediated response. For instance, in zebrafish lacking the orthologue of *NOD2*, MDP-induced cellular and molecular changes are still observed (Nayar et al., 2021; Plevy, 2021), and an *in vitro* study based on human stem cell derived macrophages with mutants in *NOD2* also found NF-kB pathway is similarly activated when the cells are treated by MDP (Takada et al., 2018). It is possible that *NOD2* dysfunction results in more biological changes than just an MDP-mediated response, and uncovering those changes may contribute to new understanding of CD pathogenesis.

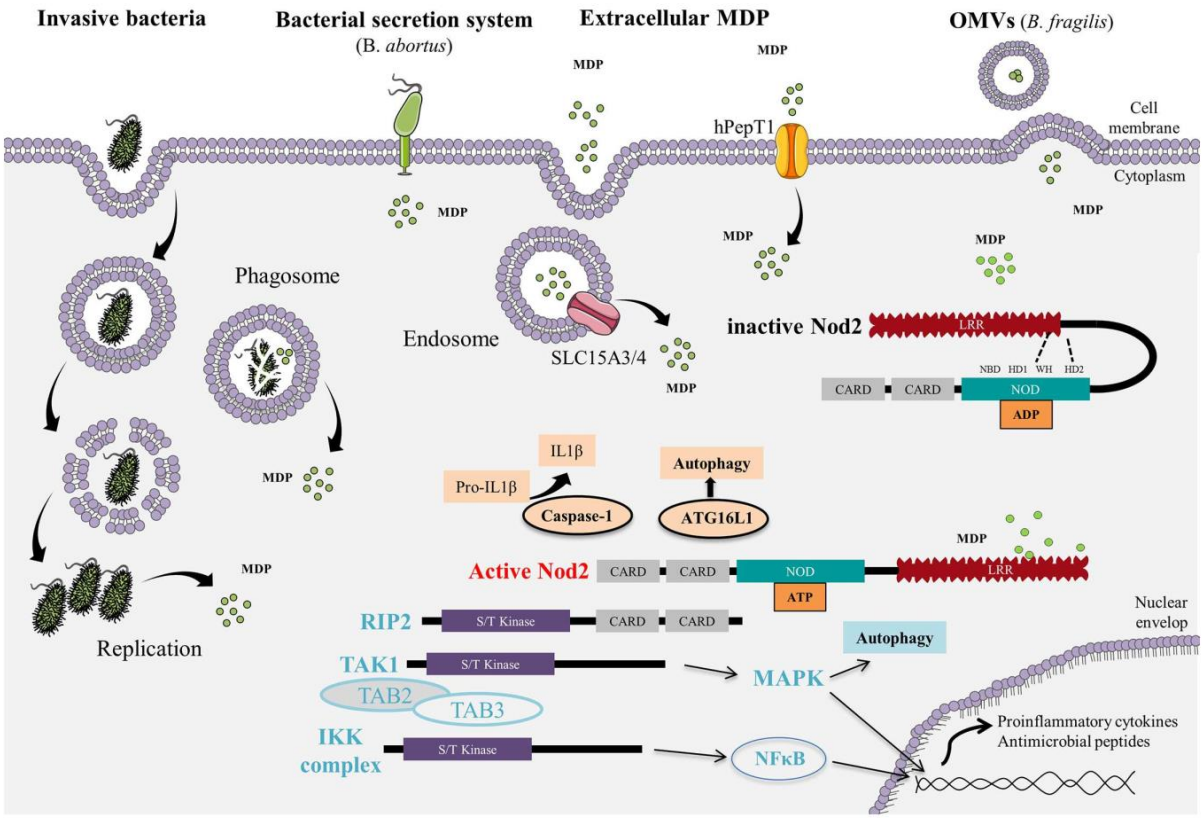


Figure 1.3.1 Mechanisms by which MDP enters into cells to trigger NOD2 signaling (Al Nabhani et al., 2017)

(Briefly, the major function of *NOD2* is recognizing MDP. MDP can enter into the cells and activate NOD2. The activated NOD2 are available to recruit proteins such as RIP2, TAK1, IKK complex …, and activate intracellular MAPK and NF-kB pathways which are related to inflammatory response.)

1.4 *LRRK2* may be the link between Crohn's disease and Parkinson's disease

LRRK2 gene codes for a protein with dual GTPase and kinase activity, and was first identified as the most frequently mutated gene in Parkinson's disease (PD). A mutation of *LRRK2*, p.G2019S substitution, has been regarded as the most common genetic determinant of PD so far (Derkinderen & Neunlist, 2018; Dzamko, 2017; H.-S. Lee, Lobbestael, Vermeire, Sabino, & Cleynen, 2021; Liu et al., 2011). Because of this, most research about *LRRK2* dysregulation focuses on neurons, and the gene's gain-of-function kinase activity has been shown to affect neurite outgrowth (Herrick & Tansey, 2021). However, *LRRK2* expression is also high in peripheral blood mononuclear cells (PBMCs), which implies its dysregulation may also have function in immune homeostasis *in vivo* (Herrick & Tansey, 2021).

Over the past decade, *LRRK2* variants have been found to be related to CD. This is mainly through Genome wide association studies (GWAS), which have identified some functional coding variants (e.g. p.G2019S, p.N2081D, p.N551K, p.N2081D) in *LRRK2* that influence CD and PD risk with the same trend (increase/decrease) among both Ashkenazi Jewish and non-Jewish Europeans (Hui et al., 2018; H.-S. Lee et al., 2021; Rivas et al., 2018). As CD is considered to be caused by the dysregulation of intestinal immune homeostasis, those evidences support the idea that *LRRK2* may also work in the immune system, and opens the gate for a possible link between CD and PD, which are taken as unrelated diseases previously.

There is also some clinical evidence that implies links between the two diseases, though the mechanism is still unclear. Epidemiological surveys among CD and PD patients report that CD patients have a increasedrisk (around 28%) to develop PD, relative to healthy controls,

although PD patients have no obvious increase (in some cases even a little decrease) in CD risk (Herrick & Tansey, 2021; H.-S. Lee et al., 2021). Although those epidemiological evidence is hard to confirm the causal association between the two diseases, combining the peak ages of the two diseases (CD: early adulthood (15-35 years old), PD: after 50 years old), it is reasonable to hypothesize that CD to PD is progressive with some shared pathogenesis (Herrick & Tansey, 2021; H.-S. Lee et al., 2021). More evidence has also implied the relationship between these two diseases, for example, some drugs used to treat CD (e.g. anti-tumour necrosis factor (TNF) agents, 5-aminosalicylic acid, azathioprine, corticosteroids) are also associated with a reduced risk to develop PD (H.-S. Lee et al., 2021). Meanwhile, molecular level work also supports this relationship. Lewy bodies, which are taken as a hallmark of PD, can be identified in the gut of PD patients up to 20 years before PD symptoms appear (H.-S. Lee et al., 2021).α-synuclein (α-syn), which is the major component of Lewy bodies, is recognized in the enteric nervous system, and found to accumulate with age (H.-S. Lee et al., 2021).In addition, α-syn is verified to be able to spread from the gut to the brain by the vagal nerve in a mouse model (H.-S. Lee et al., 2021). Moreover, many researches on the gut microbiota composition and intestinal and peripheral phenotypes among CD and PD patients also indicate the similarities between the two diseases (Herrick & Tansey, 2021). Overall, linkages between CD and PD are in accordance with the newly emerged theory of a “gut-brain axis”, which is the two-way biochemical signaling that takes place between the gastrointestinal tract (GI tract) and the central nervous system (CNS) (Chao et al., 2020; Dinan & Cryan, 2015; H.-S. Lee et al., 2021; Mayer, Knight, Mazmanian, Cryan, & Tillisch, 2014; Sudo et al., 2004; Wang & Kasper, 2014). Figure 1.4.1 summarizes the possible connections between inflammatory bowel disease (IBD) and Parkinson’s disease (PD) at present.

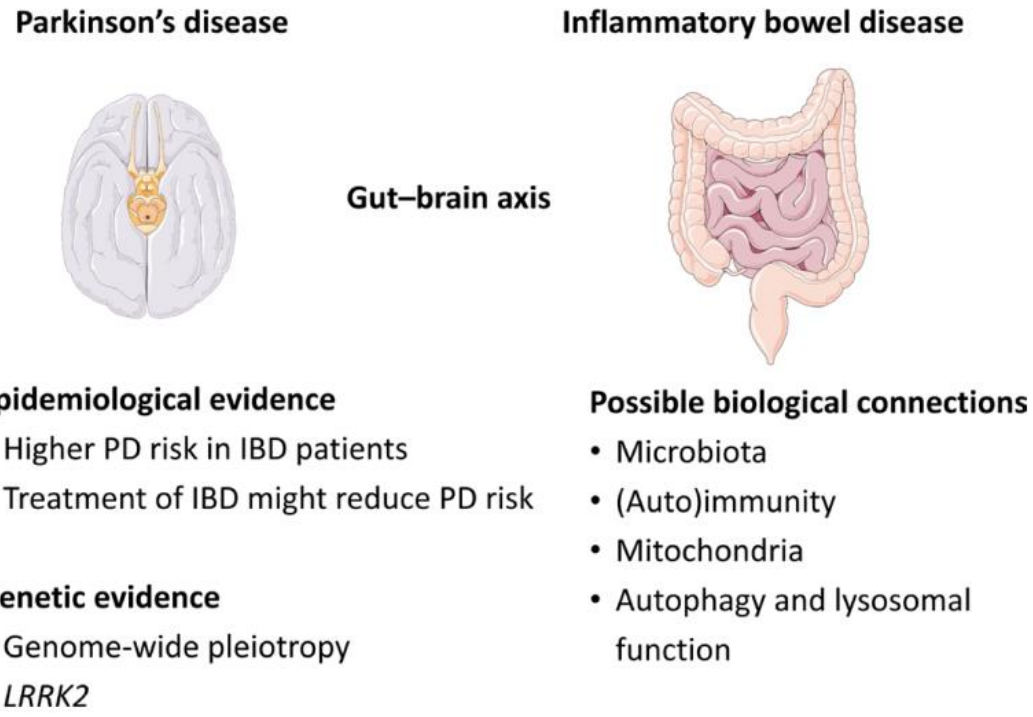


Figure 1.4.1 Possible connections between inflammatory bowel disease (IBD) and Parkinson's disease (PD) (H.-S. Lee et al., 2021)

Considering *LRRK2*'s key role in PD pathogenesis and its possible important functions in the immune system, it may sit at a crossroads between CD and PD (Herrick & Tansey, 2021). Revealing how *LRRK2* works in this linkage could be meaningful for further understanding the pathogenesis of both diseases.

1.5 Macrophages play an important role in Crohn's disease

Macrophages, cells of the innate immune system that are capable of discriminating between innocuous antigens and potential pathogens, are widely accepted as the gatekeepers of intestinal immune homeostasis (Na, Stakenborg, Seok, & Matteoli, 2019). Considering the key roles of immunological factors and intestinal homeostasis in CD pathogenesis, there is no doubt dysregulation of macrophages has a strong relationship with CD.

Macrophages in the intestines play a key role in the regulation of intestinal immune homeostasis. The intestinal macrophages are distinct from all the other tissue-resident macrophages, since most of them are derived from constant replenishment by blood-derived monocytes, which cannot self-renew, and the regulation of homeostasis is reflected by the altered monocyte-macrophage differentiation (Na et al., 2019). In the normal case, circulating monocytes enter into the lamina propria, undergo a gradual differentiation into macrophages and finally gain a more "anti-inflammatory" condition (Na et al., 2019). However, when tissue damage occurs and leads to an inflammatory response, macrophages differentiation to an "anti-inflammatory" condition is disrupted, and the cells retain their pro-inflammatory capacity (Na et al., 2019). Meanwhile, other types of immune cells also involved in a stronger immune response and can aggravate tissue damage, for instance, neutrophils, which are attracted by the damage, recruit more inflammatory monocytes to the site of inflammation, and T helper cells secret inflammatory cytokines that interact with other cell types involved in the pro-inflammatory response (Na et al., 2019). After clearance of the inflammation, the resolution stage of inflammation is established, since a long-term chronically strong immune response is harmful (Na et al., 2019). The resolution begins with the apoptosis of neutrophils, and macrophages are also equipped to detect and engulf apoptotic neutrophils to prevent secondary necrosis and exacerbated inflammation (Na et al., 2019). Macrophages also

function in clearing apoptotic intestinal epithelial cells from the damaged region by phagocytosis, and switch from a pro-inflammatory to an anti-inflammatory phenotype during efferocytosis (Na et al., 2019). Meanwhile, macrophages are also involved in contributing to re-establish the epithelial barrier, and intestinal immune homeostasis is finally restored (Na et al., 2019). *In vitro* studies of M1 and M2-like macrophages verified the potential condition transfer in macrophages *in vivo* (Galdiero, Garlanda, Jaillon, Marone, & Mantovani, 2013; Hesketh, Sahin, West, & Murray, 2017; Mosser & Edwards, 2008).

Macrophages have also been characterized to possess sub-stages even in a pro-inflammatory response to some bacterial components. The response to LPS (a major component of bacterial membranes) is representative of this characteristic. Briefly, macrophages response to LPS has associations with treatment time. The pro-inflammatory response reaches a peak at 6 hours post-treatment, and enters into a "resolution phase" at 24 hours after stimulation (Tanaka, To, O'Brien, Donnelly, & Lund, 2017). Based on data from *in vitro* studies on macrophages purified from humans and mice, changes in some genes expression (e.g. *CCR7, IL6, CSF2, IL1A*) have been recognized as markers which are capable of reflecting this progression (Schroder et al., 2012). These studies have also identified some human-specific marker genes (e.g. *CCL20, CXCL13*), which changes were only observed in human monocytes-derived macrophages (Schroder et al., 2012).

Among CD patients, many alterations related to macrophages have been reported, which include the accumulation of abnormal pro-inflammatory monocytes in intestines, altered monocyte-macrophage differentiation, and excessive cytokine secretions (Na et al., 2019). A recent study identified a special population of inflammatory macrophages from the ileum of CD patients, which were capable of differentiating into activated fibroblasts during the course of disease (Nayar et al., 2021; Plevy, 2021). Moreover, almost all current CD therapies, no

matter the classic immunosuppressive agents or the biological therapies such as anti-TNF, have been found to affect macrophage function, which further suggests a close link between dysregulation of macrophages and CD pathogenesis (Na et al., 2019).

1.6 Limitations of mouse models in mimicking Crohn's disease

Mouse models have been widely used to study the mechanisms of human diseases and identifying novel drug candidates for subsequent clinical trials (Grskovic, Javaherian, Strulovici, & Daley, 2011; Seok et al., 2013). These works assume that the results from animals can reflect human diseases, and indeed, mouse models have made great contributions to the understanding of biomedicine in recent decades (Seok et al., 2013). However, it has been hard to translate treatments in mice into successful human trials, and the condition is even worse for inflammation associated diseases (Seok et al., 2013).

Comparisons between mouse models and the clinical pathophysiology of inflammatory diseases have revealed the limitations of mouse models. For instance, a survey of the common sepsis mouse models find they can only delineate a particular type of septic response (P. Chen, Stanojcic, & Jeschke, 2014). Moreover, mouse models tend to be incapable of mimicking human inflammatory diseases, a problem that has been verified by systematic research based on transcriptome analysis (Seok et al., 2013). In that study, correlations of the gene expression levels between samples from human burns, trauma, endotoxin and their corresponding mouse models were calculated (Seok et al., 2013). High similarities were found amongst the human samples, even though the diseases are caused by different acute inflammatory stresses (Seok et al., 2013). Yet for mouse models, the correlations were quite poor, and this situation was true for both comparisons with the corresponding human disease samples and samples from mouse models with other inflammatory diseases (Seok et al., 2013). Further contrasts of temporal response patterns and regulated pathways also indicate significant differences between samples from human diseased tissues and their corresponding mouse models (Seok et al., 2013). Moreover, comparisons between published RNA-seq data of human diseased tissues and mouse models from a wide range of severe acute inflammatory diseases (e.g.

sepsis, acute respiratory distress syndrome (ARDS), infections) also support the existence of substantial differences between human responses and mouse models (Seok et al., 2013).

Many reasons may lead to a situation where mouse models are incapable of simulating human inflammatory diseases, this includes the evolutionary distance between mice and humans, diverse genetic backgrounds in human diseases and the inbred nature of mouse models, different cellular compositions between humans and mice and their distinct temporal spans for recovering from inflammation (Seok et al., 2013). As an example, mice are more resilient to inflammatory challenge than humans. The lethal dose of endotoxin for mice is 5-25mg/kg, yet this value for human is only 30ng/kg or less, which means that the endotoxin concentration capable of a massive response in humans makes little impact on mice, which has a consequent effect on the type of conclusions that can be drawn from the study of mice (Seok et al., 2013).

Specific to the two CD susceptible genes *NOD2* and *LRRK2* involved in this project, mouse models are still unsatisfactory in revealing their etiology with the disease. For the *NOD2* gene, although no CD-like symptoms can be identified from $NOD2^{-/-}$ mice, those mice show a distinct response to CD related stimuli (Wehkamp & Stange, 2005). Compared with wide type mice, $NOD2^{-/-}$ mice have more severe colitis when treated with DSS (dextran sodium sulfate), however, when treated with an endotoxin challenge, those $NOD2^{-/-}$ mice display better protection (Kobayashi et al., 2005; Pauleau & Murray, 2003). For *LRRK2*, a similar confusing situation also appears. More exacerbated DSS-induced colitis have been found in *LRRK2* knockout mice, with increased production of inflammatory cytokines (Dzamko, 2017; Liu et al., 2011). However, those *LRRK2* deficient mice are more susceptible to intestinal infection with *Listeria monocytogenes* (Dzamko, 2017; Q. Zhang et al., 2015). Finally, higher levels of susceptibility to dextran sodium sulfate (DSS)-induced colitis are

observed in transgenic mice overexpressing *LRRK2*, and *LRRK2* kinase inhibitors can be verified to ameliorate DSS-induced colitis in mice (H.-S. Lee et al., 2021; Takagawa et al., 2018). Those ambiguous results demonstrate that mouse models are not always optimal to study relationships between *NOD2/LRRK2* and CD pathogenesis, and a human *in vitro* model may better reflect the *in vivo* complex disorders in CD, with possibly novel insight for the disease.

1.7 Human induced pluripotent stem cells can contribute to the *in vitro* studies of Crohn's disease

Initially *in vitro* studies of human diseases were based on primary cells, which isolated directly from patients' tissues, and were capable to represent the disease phenotypes *in vivo* (Grandy, Tomaz, & Vallier, 2019). However, obtaining primary cells was difficult, and those cells were hard to grow *in vitro*, as they generally senesce over time, which led to limited resources of them. Immortalized or cancerous cell lines could help to overcome this obstacle, as they could be maintained in defined medium for indefinitely, and so were always capable for experiments. Yet, those cell lines would often lose their functional characteristics, and displayed a distorted or erroneous epigenome, transcriptome, proteome and even the genome itself. In this situation, scientists started to consider if it was possible to generate "normal" (untransformed, genetically normal) primary-like cells with disease characteristics *in vitro*, but could be maintained indefinitely. Human pluripotent stem cells entered into their views.

Human embryonic stem cells (hESCs) were first isolated from human blastocysts in 1998, and maintained the pluripotency to differentiate into any type of somatic cell *in vitro* (Grskovic et al., 2011; Rowe & Daley, 2019). Using defined growth factors and cellular matrices, hESCs have been successfully derived into neural cells, cardiomyocytes, pancreatic β-cells and hematopoietic cells and many other cell types (P. Chen et al., 2014; Grandy et al., 2019). Meanwhile, genetic manipulation (e.g. RNA interference or homologous recombination) enabled the introduction of disease-relevant molecular defects into hESCs, and those defects could be maintained in *in vitro*-derived cells. This gave rise to a simple experimental model: hESCs were genetically engineered to contain a disease-associated mutation, and as hESCs can be maintained in culture indefinitely they act as a source of cells.

The engineered hESCs were then differentiated to a specific cell type to use as an experimental platform to understand a human disease. These gene-edited hESC-derived somatic cells overcome some disadvantages of mouse models, and help to elucidate the molecular mechanisms of human illness (Grandy et al., 2019). In particular, hESC-derived cells are in an environment outside a complex organism (which is both an advantage and disadvantage), which allows the careful control of the external environment, for example the exclusion of medication influences or endogenous cytokines (Takada et al., 2018). However, some limitations of this model for *in vitro* studies on human diseases remains, one of which comes from hESCs themselves. A key issue is the origin of hESCs. There are continuous ethical concerns on the utilization of human embryos, and this restricts further generations of new hESC lines (e.g. federal funding from the US government limited the production of new hESC lines during the early 2000s) (Grandy et al., 2019). Limited resources for hESCs make them unavailable to study pathogenesis of human diseases under distinct population genetic backgrounds [33]. Furthermore, hESCs are hard to be applied to the thriving developing personalized medicine because of their finite origins (Grandy et al., 2019).

Human induced pluripotent stem cells (hiPSCs), which are generated from terminally differentiated somatic cells via reprogramming, inherit almost all the benefits of hESCs, and make progress to overcome the limitations of hESCs in studying human diseases *in vitro* (Takahashi et al., 2007). By introducing a defined set of transcription factors (e.g. *OCT4, SOX2, KLF4 and MYC*) into somatic cells for overexpression, the cells are capable of reprogramming back to an embryonic-like pluripotent state, and, importantly, the iPSCs retain the ability to differentiate into derivatives of all three primary germ layers *in vitro* (Takahashi et al., 2007). This characteristic makes hiPSCs a renewable cell population, and helps remove the main ethical concern from hESCs. Additionally, it is possible to generate hiPSC lines from

different populations, and so take the population genetic backgrounds into account in the *in vitro* studies of human diseases. Furthermore, hiPSCs can be generated from healthy donors and patients, and make the downstream *in vitro* studies based on true phenotypes, which indicates great prospects for the cells' applications in personalized medicine.

Another big advantage of hiPSCs for *in vitro* studies of human diseases comes from the recent technological developments of 'organoids'. Organoids are three dimensional constructs composed of multiple cell types via derived from the *in vitro* differentiation and self-organization of stem cells (M. Li & Izpisua Belmonte, 2019). Compared with traditional two-dimensional cultures, organoids are capable of simulating the architecture and functionality of native organs, to some extent. Additionally, they provide a new platform for *in vitro* studies of human diseases, as they have all the advantages of iPSCs and their *in vitro* differentiation capability, but they also can serve to model characteristics of diseased patients at the "organ" level (M. Li & Izpisua Belmonte, 2019). Development of organoid technology originated from the isolation of human adult stem cells (hASCs) that could generate into self-renewing organoids. The first isolated hASC to organoid system used adult intestinal stem cells that were isolated by the expression of *LGR5*, and those cells were verified to derive into three-dimensional structures containing four distinct intestinal cell types that could self-renew, first in mice and soon after in humans (Sato et al., 2011; Sato et al., 2009). All of these generation processes were *in vitro*, by means of cytokine-driven activation of specific pathways (J. Kim, Koo, & Knoblich, 2020). Since this initial work, many pipelines have been established to generate organoids for diverse organs, such as intestine, liver, pancreas, lung, retina, brain and kidney (M. Li & Izpisua Belmonte, 2019). However, hASCs must be from the corresponding organs to generate matching organoids, and for humans the only resources are primary tissues from clinics, simultaneously, isolation of hASCs is not easy. These

problems restrict hASC-derived organoids' application to *in vitro* studies of human diseases, as the material for research may still be limited. hiPSCs are an attractive alternatives to those hASCs, as they can be generated from many somatic cell types, such as skin, blood cells and even cells in urine (Zhou et al., 2011). Meanwhile, the capability of hiPSCs to generate three dimensional organoids (e.g. human intestinal organoids) has been confirmed *in vitro* (Spence et al., 2011), though there are still some issues that need to be solved for the hiPSC-derived organoids to be fully successful (J. Kim et al., 2020; M. Li & Izpisua Belmonte, 2019; Spence et al., 2011). For instance, hiPSC-derived organoids seem to resemble fetal tissue compared with the hASC-derived ones, which hinders them to be an optimal model for human diseases as most of the human diseases manifest during adulthoods (J. Kim et al., 2020; M. Li & Izpisua Belmonte, 2019).

Above all, hiPSCs can potentially make a contribution to the *in vitro* study of CD.

1.8 CRISPR-Cas9 provides capabilities for studying genetic disorders

Before CRISPR-Cas9 was discovered and shown effective in eukaryotic cells, sequence-specific genome editing in eukaryotes was hard work. Although some technologies (e.g. zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs)) had been developed, they were difficult to design and implement, with high costs in both time and money, and the targeting efficiency was also problematic (Cong & Zhang, 2015; Ran et al., 2013). Establishment of the CRISPR-Cas9 system makes generations of human cells with specific genome edits a relatively common event, which is an incredible progress in the field of genome engineering (Cong & Zhang, 2015; Ran et al., 2013).

Clustered regularly interspaced short-palindromic repeat (CRISPR)-Cas (CRISPR-associated) systems were discovered in prokaryotes. They are an adaptive immune system that has been identified in about half of bacterial and nearly all archaeal species to date (F. Zhang, 2019). Functions of those systems are for defending against invasions of other microbes or viruses (e.g. bacteriophages) via targeting the exogenous nucleic acids (DNA or RNA) and cleaving them (F. Zhang, 2019). CRISPR-Cas functions work in three phases: adaptation, crRNA processing, and interference (F. Zhang, 2019). In detail, a subset of Cas proteins called the “adaptation module” obtain fragments of foreign nucleic acids and insert them into a CRISPR sequence array in the bacterial genomes; afterwards, the CRISPR array is transcribed and processed into individual CRISPR RNAs (crRNAs), which are non-coding RNAs corresponding to the previously obtained exogenous sequence fragments; then, crRNAs guide the “interference module”, which is encoded by a complex comprising of Cas effector subunits or single-effector proteins, to generate cleavage on the targeting nucleic

acids, and prevent exogenous invasion (F. Zhang, 2019).

Characteristics of this defense mechanism led to scientists considering if it could be exploited to manipulate genomes precisely, especially in eukaryotic cells. After further exploration of CRISPR-Cas, this idea became a reality in 2013. By successfully transferring the CRISPR-Cas9 system into mammalian cells, this accomplishment led to this system becoming the most commonly used tool for genome editing (Cong et al., 2013; F. Zhang, 2019). This system contains two components: Cas9 protein and the single-guide RNA (sgRNA). Cas9 is a large CRISPR-associated protein identified from *Streptococcus thermophiles* (F. Zhang, 2019). It is the prototypic protein of type II CRISPR systems, which contains a DNA-cleaving HNH domain and is capable to complete the target cleavage with no needs to form multi-protein complex (Cong & Zhang, 2015; F. Zhang, 2019). sgRNA is the chimeric fusion of a mature crRNA and trans-activating crRNA (tracrRNA) (Cong & Zhang, 2015). As in CRISPR-Cas in microorganisms, crRNA in this system corresponds to the target region for DNA cleavage, and tracrRNA is necessary for maturing the crRNA in the system (F. Zhang, 2019). In addition, precise editing of the genome via this system requires a short well-conserved sequence motif at the end of target regions, which is called protospacer adjacent motif (PAM) (Hsu, Lander, & Zhang, 2014; F. Zhang, 2019). For Cas9 proteins originated from distinct microbes, the PAM sequence is different, and this diversity also determines the overall space of Cas9-mediated DNA cleavage on the genome (Hsu et al., 2014; F. Zhang, 2019). For example, the PAM sequence of Cas9 from *Streptococcus thermophiles* (StCas9) is 5’-NNAGAAW, which can only be recognized on average every 106.6bp in the human genome, whereas Cas9 from *Streptococcus pyogenes* (SpCas9) has shorter and simpler PAM sequence (5’-NGG), which exists in the human genome every 12.7bp on average (F. Zhang, 2019). After identifying the target region in the genome, Cas9

first creates blunt double-strand breaks (DSBs) 3bp upstream of the PAM, then the cleavage is repaired by non-homologous end joining (NHEJ) or homology-directed repair (HDR) pathway, depending on the requirements of the experiments (Cong & Zhang, 2015; F. Zhang, 2019). In general, NHEJ is enough to generate gene knockouts, because it is an error prone process that introduces insertion/deletion (indels) into the target site (Cong & Zhang, 2015; Ran et al., 2013). However, for more precise editing, for example introducing defined modifications at the target locus, the HDR pathway has to be used. This process requires an exogenously designed repair templates for the DSBs, and they should be integrated with chromosomal DNA via homologous recombination (HR), which occurs at a lower frequency than NHEJ (Cong & Zhang, 2015; Ran et al., 2013).

Although the CRISPR-Cas9 system is simpler than any other previous technologies (ZFNs and TALENs) for genome editing, obstacles still exist, however, these problems have aided by progress in computational biology and genetic engineering. A general pipeline for genome editing of mammalian cells by CRISPR-Cas9 system has been established. It starts from target selection. Through the help of many CRISPR design tools developed by computational approaches (e.g. http://tools.genome-engineering.org), this step can be as convenient as inputting the sequence of the target region, and then the tools can return candidates for sgRNA sequences, with predictions for off-target sites for assessing the designs (Ran et al., 2013). The next step is constructing sgRNA-expressing plasmids. In this step, oligo pairs corresponding to the designed sgRNA are synthesized, and annealed and ligated into plasmids (Ran et al., 2013). For ease of use, those plasmids (e.g. pSpCas9(BB)) have been commercialized, with Cas9 and the remainder of sgRNA except for the target (Ran et al., 2013). Construction of sgRNA-expressing plasmids can be simplified by introducing the target into this invariant scaffold (Ran et al., 2013). Through genetic engineering technologies,

the plasmids can be transfected into competent cells (e.g. Stbl3) for amplification, and Sanger sequencing is able to validate the successful construction of sgRNA-expressing plasmids (Cong & Zhang, 2015). Then, those plasmids need to be transfected into mammalian cells. Many commercialized reagents (e.g. Lipofectamine 2000) have been produced for this aim, and work quite well (Cong & Zhang, 2015). Subsequently, transcription and translation of the constructed plasmids work intracellularly by the aid of RNase III from the host, and this combination leads to editing of the host genome (Cong & Zhang, 2015).

As a human disease with a complex genetic background, CRISPR-Cas9 is a helpful tool for intracellular sequence-specific genome editing, and can provide more capabilities to reveal how those disorders trigger the disease. Furthermore, the technology may also help in developing new therapies for CD, such as correcting genetic disorders intracellularly, which have been in clinical trials for many other human diseases (Cox, Platt, & Zhang, 2015; F. Zhang, 2019).

1.9 RNA sequencing is a useful tool for sensing CD-related changes in response to phenotypes

RNA sequencing (RNA-seq), is a powerful technique that can rapidly and relatively easily quantify the whole cell transcriptional output in response to changes in the environment of inside the cell (Stark, Grzelak, & Hadfield, 2019). Compared with other approaches to measure gene expression (e.g. quantitative Real-time PCR, microarray), data from RNA-seq is more comprehensive, with higher quality, a higher dynamic range, and with less experimental burden and lower price (Stark et al., 2019). Meanwhile, large data generated from RNA-seq promotes usage of computational methods in biological research, which helps to resolve many difficulties, and contributes to a better understanding of the complexities of nature. Because of these advantages, in just over a decade, RNA-seq has become an indispensable tool for a wide range of research fields (Stark et al., 2019). Figure 1.9.1 summarizes the workflow of RNA sequencing from experiments to data analysis.

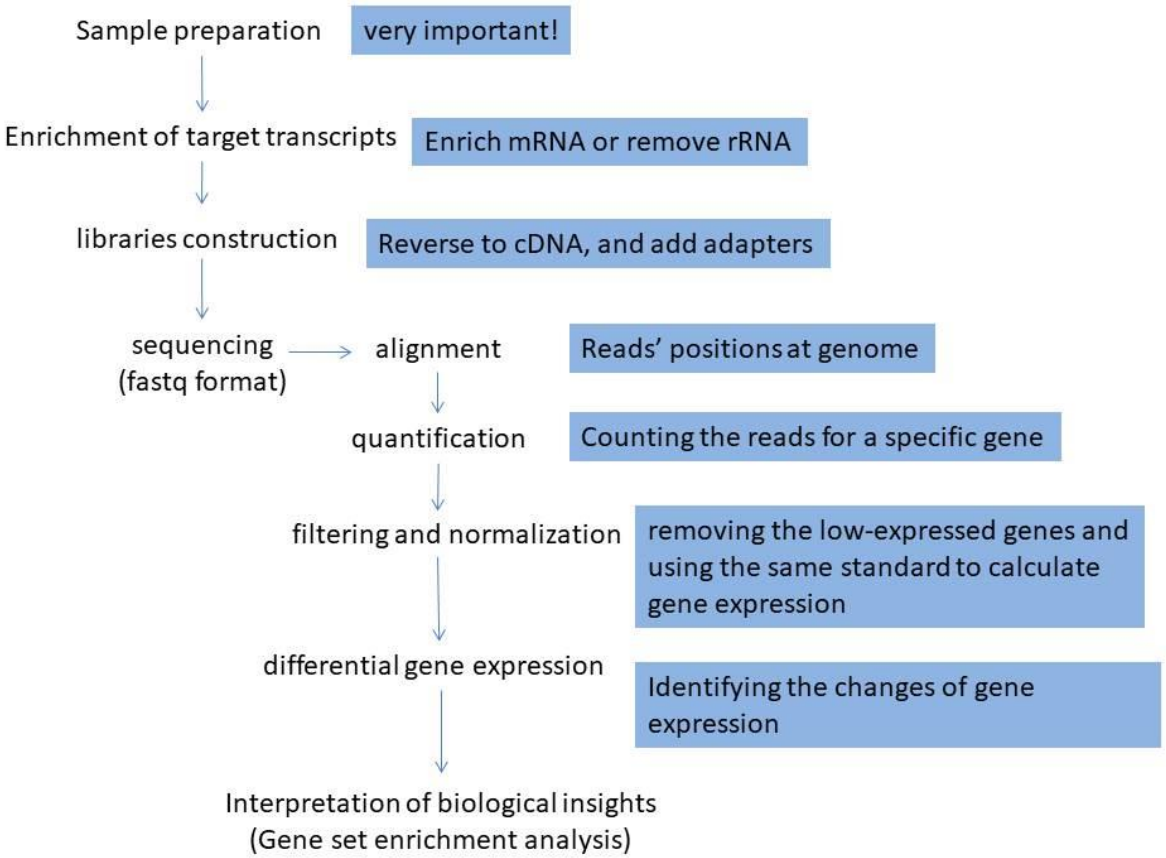


Figure 1.9.1 Workflow of RNA sequencing

Sample preparation is the initial step for RNA-seq, and high-quality extracted RNA without severe degradation is required for good data. This reflects true biological change. Some commercialized reagents (e.g. RNAzol) have been developed to optimize RNA extraction. It is also necessary to plan biological replicates for each sample, for helping understand the inherently stochastic variance in every biological system (Stark et al., 2019). Next, according to the needs of the experiment, target transcripts are selected from the total extracted RNA. In some cases, for analysis of gene expression, enrichment of transcripts containing a poly(A) tail is enough as all the protein-coding regions of transcriptome can be selected via this method (Stark et al., 2019). However, non-coding RNAs are also potentially interesting, whilst in other situations ribosomal RNAs in the samples need to be depleted (Stark et al., 2019). Selected RNAs are fragmented and reverse transcribed into cDNAs (Stark

et al., 2019). Subsequently, adaptor-ligated libraries are constructed and sequenced on a high-throughput short-read platform (often Illumina platforms) for generating the sequencing data (Stark et al., 2019). At this stage, depth and length of the reads, and sequencing mode (single-end or paired-end) are set depending on the requirements of the experiment. Generally, 10-30 million reads per sample is enough for RNA-seq, and paired-end sequencing with longer read lengths can help accurately map the reads to reference genomes or transcriptomes more accurately, with fewer ambiguities in the results (Stark et al., 2019). Although the process from transcript selection to sequencing seems to be complicated, thanks to the rapid development of industries in this field, this work can be accomplished using standardized workflows, by commercialized kits and equipment.

Data files obtained from RNA-seq are base-called sequencing reads (usually in FASTQ format). Those files are huge, containing millions to billions of nucleotides, and the analysis of them must rely on computational methods. There are four basic steps for the computational analysis of RNA-seq data, which are the alignment of sequencing reads, quantification of transcript abundance, filtering and normalization, and differential expression modeling (Stark et al., 2019). First, raw sequence reads are mapped to the reference genome (or transcriptome), which has been annotated by large genome projects such as GENCODE, HGP and Ensembl (Collins, Morgan, & Patrinos, 2003; Frankish et al., 2019; Frankish et al., 2021; Harrow et al., 2006; Howe et al., 2021; Hubbard et al., 2002), and the results of mapping are converted into genomic coordinates (Stark et al., 2019). Next, reads with location information are assigned to genes or transcripts, and the abundance of reads for each gene or transcript is quantified at the same time. Results of this step are summarized in an expression matrix for all the samples in the experiments, with generally each row as a gene or transcript and each column as a sample (Stark et al., 2019). Considering different read depths between the samples and distinct

gene/transcript lengths or GC contents in each sample, the expression matrix needs to be normalized based on those differences. Meanwhile, at this stage, genes or transcripts with uniformly low read abundance are detected and removed from the matrix, as they are often noisy and not informative for downstream differential gene expression (DGE) analysis (Stark et al., 2019). Afterwards, changes of gene/transcript expression levels in the experiments are identified via DGE analysis, with the help of statistical models (Stark et al., 2019).

DGE results are just a set of values from statistics, and how to interpret and generate them into biological insights becomes the challenge for the next steps (Bu et al., 2021; Subramanian et al., 2005). In the beginning, this work relies on the DE genes. Expressions of those genes are significantly different in the experiments, and the values (e.g. adjusted p-value, fold change) from statistics in DGE results contribute to their identifications (Maleki, Ovens, Hogan, & Kusalik, 2020). Those DE genes are taken as the key from experiments, and then need to be interpreted using previously existing knowledge. For a small number of DE genes, it is possible to individually interpret each gene's response. However, when there are a large number of identified DE genes, for example hundreds or thousands, interpreting them one by one is impractical. Furthermore, in reality, biological processes are often driven by changes in a relatively large group of genes which share common functions or attributes, but interpretations on single genes can disregard this aspect and so affect the reliability of the results (Maleki et al., 2020). In this situation, gene set enrichment (GSE) analysis emerges as a powerful analysis tool. This analysis first defines sets of genes related to specific biological processes from the current knowledge bases (e.g. MSigDB, GeneSigDB, GeneSetDB), then observes whether genes with expression levels' changes in the experiments are significantly enriched to the genes from each set (Maleki et al., 2020).

Many methods for GSE analysis have been developed. The initial one is

over-representation analysis (ORA), which is the most commonly used method (Bu et al., 2021). This method is the natural extension of one-by-one interpretations of DE genes (Bu et al., 2021). It still relies on the identified DE genes, but the interpretations are based on evaluating whether those genes are overrepresented in a specific gene set via statistical analyses (Bu et al., 2021). Compared with the previously single-gene analysis, this method has made great progress, however, it has some limitations. The most apparent one is ORA uses identified DE genes, which are from user-defined thresholds, and the changes of thresholds may produce distinct results (Maleki et al., 2020). Meanwhile, simple filtering by the thresholds also makes ORA ignore all the genes which are filtered out, which also contributes to the biological processes (Maleki et al., 2020). Furthermore, ORA takes all the identified DE genes as the same, but there are still some important differences (e.g. fold change) among them which can affect the interpretations (Maleki et al., 2020). To help the deficiencies of ORA, functional class scoring (FCS) method, which is represented by Gene Set Enrichment Analysis (GSEA), was developed (Maleki et al., 2020; Subramanian et al., 2005). This method does not require the identifications of DE genes, and instead uses all the information from DGE results for the analysis (Bu et al., 2021; Maleki et al., 2020). The core of this method is an enrichment score, which is calculated by defined rules to reflect how significantly the genes vary with respect to variable expression levels in the experiments enriched to the set of genes for a specific biological process (Bu et al., 2021; Maleki et al., 2020). Expressive genes are put in order by some ranking metric, such as fold change, from high to low, for introducing the neglected differences in ORA into the analysis, and help to identify if the biological process is up or down regulated (Bu et al., 2021; Maleki et al., 2020). Additionally, more elements are taken into consideration for biological interpretation of DGE results, for instance, diversity of distinct genes' contributions to specific pathways, and this

leads to GSE method based on the pathway topology (Bu et al., 2021; Maleki et al., 2020), rather than just changes in a gene for some defined arbitrary threshold.

Because of the characteristics of RNA-seq, this technology is available to identify subtle biological changes behind phenotypes, which makes it useful for studying CD pathogenesis in disease models. Considering the complexity of CD, generally none of those models have obvious phenotype changes, for example there are no mouse models with a knockout of CD susceptible genes that directly generate intestinal inflammation *in vivo* (Pizarro, Arseneau, Bamias, & Cominelli, 2003). This indicates subtle changes caused by genetic deficiencies, with few effects on phenotypes may be the keys to CD pathogenesis. RNA-seq can help to sense those biological changes behind phenotypes possibly associated with CD.

1.10 Objectives and Aims

The general approach of this project is to gain new insight into CD pathogenesis via human genetically engineered stem cell *in vitro*-derived models. In detail, two representative CD susceptible genes - *NOD2* and *LRRK2* are chosen, and how their deficiencies deregulate the function of macrophages, which are the gatekeepers of intestinal immune homeostasis, is aimed to be found out. The effects of those two genes' deficiencies on intestinal organoids are also studied. Using CRISPR-Cas9 and hiPSCs *in vitro* differentiation, the disease models are isogenic and are suited to uncovering the biological changes from genetic deficiencies, and RNA-seq is utilized to sense those changes. Furthermore, the pipeline for preparing single cell RNA-seq (scRNA-seq) samples from CD patients' resections is optimized, and the monocyte/macrophages are explored in the generated data. This research overcomes some limitations of mouse models and primary cell-based studies on CD, and help to reveal new insights into CD pathogenesis.

Below are the aims of each chapter in the thesis, and Figure 1.10.1 is the abstract of the PhD work.

1. In Chapter 3:
    a. Finding out the effects of *NOD2* deficiency on hiPSC-derived macrophages and intestinal organoids (especially for the non-MDP related influence) (Chapter 3).
    b. Based on those findings, it is possible to explore how dysregulations affect intestinal immune homeostasis, and their possible relationship with CD pathogenesis.
2. In Chapter 4:

a. Finding out why the deficiency of *LRRK2*, a key determinant of PD, can contribute to CD pathogenesis.
b. It is hopefully to know whether the inter-relationship between the two diseases can be uncovered by *in vitro* model based studies.

3. In Chapter 5:
   a. Establishing a pipeline for preparing scRNA-seq samples from CD patients' resections.
   b. Using this pipeline to perform an isogenic comparison between monocytes/macrophages with distinct inflammatory severity to understand CD at higher resolution.
   c. This work can help to assess if the findings from *in vitro* models have practical significance for the clinic.

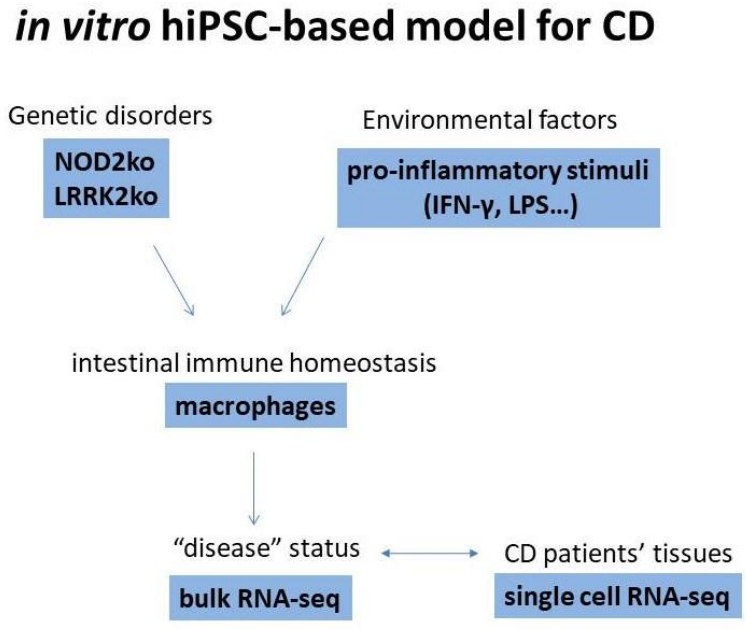


Figure 1.10.1 Abstract of the PhD work
(To summarize this project, it it for studying if NOD2ko/LRRK2ko macrophages meet pro-inflammatory stimuli such as IFN-γ/LPS, will they have some "disease" status? and whether those status can be recognized in CD patients?)

CHAPTER 2

MATERIALS AND METHODS

2.1 Maintenance of human iPS cells

The human iPS cell line used for all the experiments was a published one generated from the urine cells of a 24-year old Caucasian female, which was marked as “S0730” in the original paper (Zhou et al., 2011). Exome of the cell line was sequenced, and none of the CD-related loci was identified in *NOD2* and *LRRK2* gene. The cell line was obtained from the Guangzhou Institutes of Biomedicine and Health (Guangzhou, China).

Human iPS cells were maintained in mTeSR™1 (STEMCELL Technology, #85850) medium using 6-well plates (Corning, #3516) coated by Corning® Matrigel® hESC-Qualified Matrix (Corning, #354277), and the culture media was replaced every day (Braam et al., 2008; Ludwig et al., 2006). ReLeSR™ (STEMCELL Technology, #05872) was used to detach the cells for passage and cryopreservation, and the Accutase solution (Sigma-Aldrich, #A6964) was used for detachment when the single cell dissociation was required. Cryopreserved cells were stored in CryoStor® CS10 (STEMCELL Technology, #07930) by Nunc™ Biobanking and Cell Culture Cryogenic Tube (ThermoFisher SCIENTIFIC, #375418). The tubes containing the cells were put into Scientific™ Mr. Frosty™ (ThermoFisher SCIENTIFIC, #5100-0001) with isopropanol, and gradient frozen in -80 ℃ freezer. Cryovials were moved into liquid nitrogen for long-term storage. Methods to thaw the human iPSCs was similar to the other human cell lines, first immediately put the cryovials containing the frozen cells into 37 ℃ water bath for quick melting, then the contents were transferred into a 15 ml tube, and ~2 ml mTeSR™1 was added to wash the cells. After centrifuged (200 g, 5 mins, room temperature) and carefully discarded the supernatant, cells were resuspended in mTeSR™1 and separated into the 6-well plates coated by Matrigel for further maintenance.

2.2 Mycoplasma test

Mycoplasma tests of hiPSCs were by TransDetect PCR Mycoplasma Detection Kit (Transgen Biotech, #FM311-01) before *in vitro* differentiation. 40 μl of culture supernatant from the cells with 80% confluence was incubated at 95 ℃ for 10 minutes. PCR reactions were set up as Table 2.2.1 with both negative and positive control, followed by the amplification program (94 ℃ for 4 mins; 94 ℃ for 30 s, 60 ℃ for 30 s, 72 ℃ for 30 s, 35 cycles; 72 ℃, 5 mins) on BIO-RAD T100 Thermal Cycler. The PCR products were detected by electrophoresis on 1.5% agarose gel.

Component	Volume
Template	2 μl
Myco Primer Mix	0.4 μl
TransDetect® Myco PCR SuperMix（2x）	10 μl
MycoFree Water	7.6 μl
Total volume	20 μl

Table 2.2.1 Components of PCR reactions

2.3 *In vitro* differentiation of macrophages

EB (embryoid body)-based method with cytokines was used for the *in vitro* differentiation of macrophages. The method was adjusted from published protocol on producing macrophages from hiPSCs long-term and efficiently for possible applications (e.g. drug screening) in pharmaceutical industries (Buchrieser, James, & Moore, 2017; Flynn et al., 2015; Gutbier et al., 2020; Karlsson et al., 2008; van Wilgenburg, Moore, James, & Cowley, 2014; Wilgenburg, Browne, Vowles, & Cowley, 2013). Figure 2.3.1 was the general view of whole protocol.

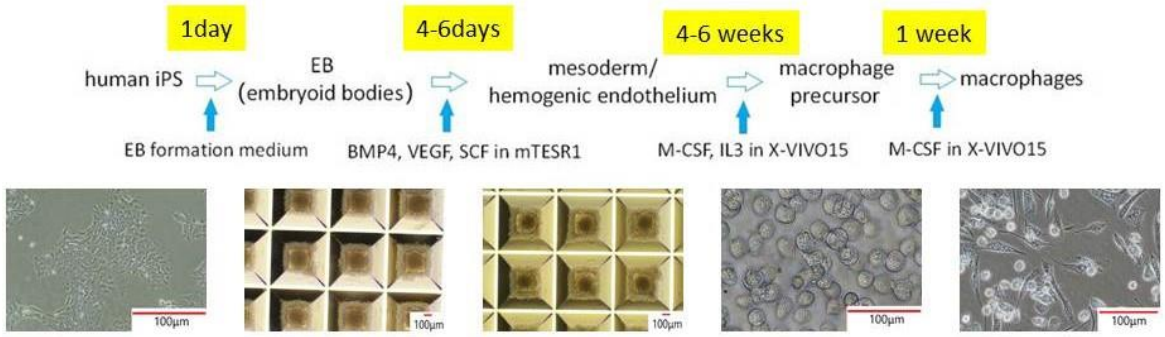


Figure 2.3.1 Protocol of the *in vitro* differentiation of macrophages from hiPSCs

For details, cultured hiPSCs were detached and dissociated into single cells, then passed through a 40 µm Cell Strainer (Corning, #352340) to remove clumps. AggreWell™800 (STEMCELL Technology, #34815) was washed by Anti-Adherence Rinsing Solution (STEMCELL Technology, #07010) with centrifuge (2000 g, 5 min, room temperature) to remove bubbles, then 1 ml AggreWell™ EB Formation Medium (STEMCELL Technology, #05893) with 10 µM Y-27632 (STEMCELL Technology, # 72302) was added into each well of the plate which was designed to be used. Cell concentration of the dissociated hiPSCs was

counted by Countstar® BioTech Automated Cell Counter, and around 5x10^6 cells were suspended in 1 ml AggreWell™ EB Formation Medium with 10 μM Y-27632, then transferred into a well of AggreWell™800 and mixed with the 1 ml medium already in that well by pipette. Next, AggreWell™800 with the cells was centrifuged (100 g, 3 mins, room temperature), and incubated at 37 ℃ with 5% CO2 and 95% humidity for 24 hours.

After 24 hours incubation, EBs formed. Then the medium was 75% changed (first discarded half by pipette, then added half, repeated once) every day, avoiding touching EBs in the micro wells. Medium for the EBs was based on mTeSR™1, with 50 ng/ml BMP4 (PeproTech, AF-120-05ET), 50 ng/ml VEGF (PeproTech, AF-100-20) and 20 ng/ml SCF (PeproTech, AF-300-07).

After 4-6 days' culture, EBs were available for harvested. 1ml wide bore tip (Corning, TF-1005-WB-R-S) was used to harvest EBs from micro wells by pipette. For removing the single cells and debris, EBs with the medium passed the 37 μm Reversible Strainer (STEMCELL Technology, #27250) with arrow up in a 50 ml PP Centrifuge Tube (Corning, #430829). Micro wells were washed by DMEM/F12 (Hyclone, SH30023.01), and the liquid also passed the same strainer for fully collection of EBs. The strainer was reversed with arrow down in a new 50 ml PP Centrifuge Tube, then the EBs hanging on the filter membrane were resuspended by "Monocytes" medium and separated into 6-well plates by pipette with 1ml wide bore tips. For the concentration, EBs from one well of AggreWell™800 corresponded to a 6-well plate, and 3 ml of "Monocytes" medium was used for each well of the plate. "Monocytes" medium was made by 500 ml X-VIVO15 (Lonza, 04-418Q), 5 ml GlutaMax (ThermoFisher Scientific, #35050-061), 5 ml Penicillin-Streptomycin Solution (Hyclone, SV30010) and 500 ul 2-Mercaptoethanol (55mM), with 100 ng/ml M-CSF (R&D Systems, 216-MC; ThermoFisher Scientific, PHC9501) and 25 ng/ml IL-3 (R&D Systems, 203-IL;

PeproTech, AF-200-03).

Every 7-10 days, two thirds of the “Monocytes” medium was changed in each well containing EBs. Floating “monocytes” (actually macrophage precursors) were observed after 4-6 weeks, and were capable to be collected when changing the medium. Replaced medium was collected in 50 ml tubes, then centrifuged (200 g, 10 mins, room temperature), discarded supernatant and resuspended in “Macrophages” medium. 1-2x10^6 cells were mixed with 2-3 ml “Macrophages” medium and transferred into a well of 6-well plate. After one week’s incubation at 37 ℃ with 5% CO2 and 95% humidity, the macrophages could be used for further experiments. “Macrophages” medium was made by 500 ml X-VIVO15, 5 ml GlutaMax, 5 ml Penicillin-Streptomycin Solution with 100 ng/ml M-CSF.

2.4 *In vitro* differentiation and maintenance of intestinal organoids from hiPSCs

STEMdiff™ Intestinal Organoid Kit (STEMCELL Technology, #05140) was used for deriving and maintaining the intestinal organoids from hiPSCs. Figure 2.4.1 summarized the process of organoids' in vitro differentiation (Spence et al., 2011; Tsai et al., 2017).

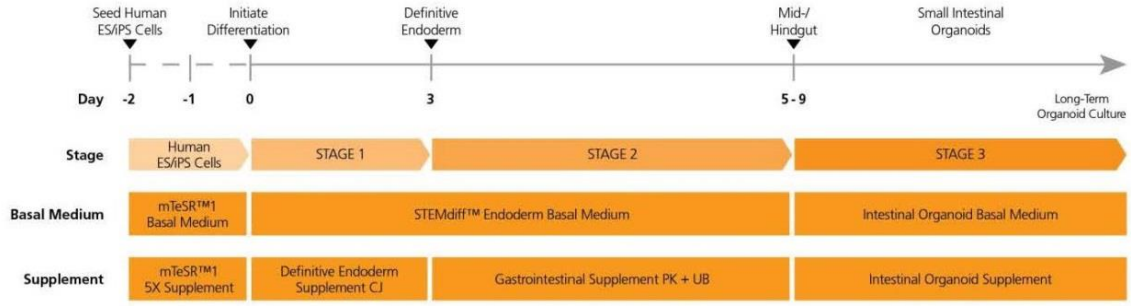


Figure 2.4.1 Summary of the organoids' in vitro differentiation

(From Technical manual of STEMdiff™ Intestinal Organoid Kit)

hiPSCs were digested into clumps by ReLeSR™, and the size of the clumps were around 50-200 μm. One fourth of the cell clumps from a well of 6-well plate (~80% confluence) were transferred to a well of 24-well plate (Corning, #3524) coated by Corning® Matrigel® hESC-Qualified Matrix, then cultured with 1 ml mTeSR™1. The 24-well plate containing the cells was incubated at 37 ℃ with 5% CO2 and 95% humidity for 24 hours.

After 24 hours, medium was fully changed by 0.7 ml "DE medium" per well, following by another 24 hours' incubation. Then the medium was fully changed every day for another two days, by 0.5 ml "DE medium" per well. The contents of "DE medium" were 1.98 ml STEMdiff™ Endoderm Basal Medium (STEMCELL Technology, #05111) with 20 μl

STEMdiff™ Definitive Endoderm Supplement CJ (STEMCELL Technology, #05113) (Day 0-Day 2).

From Day 3, medium was fully changed into 0.5 ml “MH medium” per well, following by 24 hours’ incubation. Everyday fully medium change (0.5ml “MH medium” per well) continued to Day 9, and the old medium was collected from Day 5. “MH medium” was the mixture of 2.94 ml STEMdiff™ Endoderm Basal Medium, 30 μl Supplement PK (STEMCELL Technology, #05141) and 30 μl Supplement UB (STEMCELL Technology, #05142).

Free-floating spheroids could be observed from day 5, and were available to be harvested from the old medium. Medium containing the spheroids was transferred to a 15 ml tube and centrifuged (300 g, 5 mins, room temperature). Before the centrifuge, amount of spheroids could be estimated in a clean 24-well plate under microscope. The supernatant was carefully aspirated and discarded, and the spheroids were mixed with Corning® Matrigel® Growth Factor Reduced (GFR) Basement Membrane Matrix, Phenol Red-Free (Corning, #356231) on ice (~50 spheroids per dome, each dome needed 50 ul Matrigel). The spheroids were embedded into center of the wells (24-well plate) by pipette, and incubated at 37 ℃ for 30 mins to make the Matrigel dome solidify. 0.5-1 ml “OGM medium” per well was then added into the plate, and the plate was incubated at 37 ℃ with 5% CO2 and 95% humidity. Medium (0.5-1 ml) was changed fully every two or three days, and the organoids were available to be passaged after 7-10 days’ culture. All the organoids used in my thesis were originated from spheroids collected at Day 8. Composition of the “OGM medium” were 9.7 ml STEMdiff™ Intestinal Organoid Basal Medium (STEMCELL Technology, #05143), 200 ul STEMdiff™ Intestinal Organoid Supplement (STEMCELL Technology, #05144) and 100 ul GlutaMax.

Passaging the organoids required the Cultrex Organoid Harvesting Solution (R&D

Systems, #3700-100-01). Organoids were collected by the Harvesting Solution and transferred into 15 ml tube using 3 ml Pasteur pipettes (NEST Biotechnology, #318212). The 15 ml tube containing organoids was incubated on ice for 30 mins with moderate shaking to depolymerize the Matrigel. Afterwards, the same amount of cold DPBS (Hyclone, SH30028.02) with Harvesting Solution was added into the tube, then the tube was centrifuged (300 g, 5 mins, 4℃) and the supernatant was carefully aspirated and discarded. Organoids at the bottom of the tube were resuspended in cold GFR Matrigel (50 ul per dome), and passaged in 1:4 to 1:8 ratio by the same method as embedding spheroids. The passaged organoids were still cultured in "OGM medium" and incubated at 37 ℃ with 5% CO2 and 95% humidity, with complete medium change every two or three days.

Methods to cryopreserve and thaw the organoids were similar to the ones for hiPSCs. For cryopreservation, CryoStor® CS10 was used instead of the cold Matrigel to resuspend the collected organoids, and the organoids were separated into cryovials. Cryovials containing the organoids were gradient frozen in -80 ℃ freezer, then transferred to liquid nitrogen for long-term storage. For thawing, cryovials with frozen organoids inside were immediately put into 37 ℃ water bath for quick melting, then the contents were transferred into a 15 ml tube, and ~2 ml DMEM/F-12 was added to wash the organoids. After centrifuged (200 g, 5 mins, room temperature) and carefully discarded the supernatant, organoids were resuspended in cold GFR Matrigel (50 ul per dome), and embedded in 24-well plate. "OGM medium" was used for culturing the organoids, with complete medium change every two or three days.

2.5 Western Blot

For protein samples' collection, cells were detached by TrypLE™ Express Enzyme (1X), phenol red (ThermoFisher Scientific, #12605010), and washed by DPBS with centrifuge (200 g, 5 mins, room temperature). Supernatants were discarded and the cell lysis was by 2X Loading buffer directly (around 35 ul for each well of 6-well plate). Contents of 2X Loading buffer were 4% SDS, 5% 2-mercaptoethanol, 20% glycerol, 0.004% bromophenol blue and 0.125 M Tris HCl, pH 6.8. 8% SDS-PAGE (stacking gel: 5%, separating gel: 8%) was used on the samples, and around 30ul cell lysates were loaded onto each lane. Electrophoresis apparatus was filled with 1X running buffer (25 mM Tris base, 192 mM glycine, 0.1% SDS, and adjusted to pH 8.3). Gel was run at 65 V for 40 mins and then 125 V for 70 mins to separate the proteins. Prestained Protein Ladder, 10 to 180 kDa (ThermoFisher Scientific, # 26616) was used to mark sizes of the proteins.

After electrophoresis, filter papers, sponges and gel were soaked in 1X transfer buffer (Glycine2.9 g, Tris 5.8 g, SDS 0.37 g, 200 ml absolute alcohol and 800 ml ddH_2O) for 10 mins. 0.45 μm PVDF membrane was soaked in absolute alcohol for 10 mins. "Sandwich" structure was prepared as Figure 2.5.1, and any air bubbles in the "sandwich" were gently removed by roller or pipette to prevent their inhibitions on gel-to-membrane protein transfer. The "sandwich" was placed into transfer cassette on ice, and wet transfer with 1X transfer buffer was used at 110 V for 90 mins at 4 ℃.

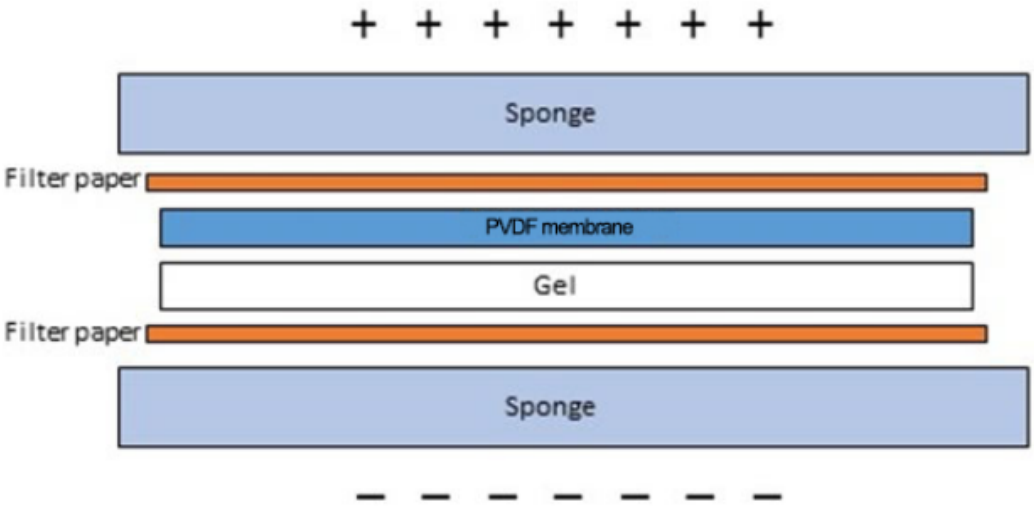


Figure 2.5.1 “Sandwich” structure for protein transfer

(From “BENCHTOP WESTERN BLOT PROTOCOL”, NOVUS BIOLOGICALS)

After transfer, membrane was stained by Ponceau red for 30 seconds, and rinsed by ddH_2O to visualize the protein bands. Then the membrane was washed by 1X TBST (1X TBS+1‰Tween 20) for several times until clear. Afterwards, the membrane was incubated in the blocking solution (5% non-fat dry milk diluted by 1X TBST) with constant rocking for 1h at room temperature. The primary antibody was diluted by 1% non-fat dry milk in the ratio of 1:500 (for internal reference, 1:5000), and the membrane was incubated in the primary antibody solution with gentle rocking at 4 ℃ overnight. Next, the membrane was washed by 1X TBST three times (each time for 10 minutes) with gentle rocking, and incubated in the secondary antibody solution (1:5000 in 1X TBST) for 90 mins at room temperature with gentle rocking. The membrane was washed by 1X TBST for three times as previous, and incubated in Pierce™ ECL Western Blotting Substrate (ThermoFisher Scientific, #32106) for 1min. Protein bands on the membrane were detected by Tanon 6100c chemiluminescent imaging system and the corresponding software GelCap.

List of the antibodies:

STAT1 antibody (working dilution: 1:500) (Cell Signaling Technology, 9172S)

Anti-beta Actin antibody (working dilution: 1:5000) (Abcam, ab8227)

Goat Anti-Rabbit IgG(H+L), HRP Conjugate (working dilution: 1:5000) (Transgen Biotech, # HS101-01)

2.6 RNA extraction and sequencing

RNAzol® RT (Molecular Research Center, RN190) was used for the total RNA extraction (Chomczynski, 1993; Chomczynski & Sacchi, 1987; Chomczynski, Wilfinger, Kennedy, Rymaszewski, & Mackey, 2010; Wilfinger, Mackey, & Chomczynski, 1997). Samples were collected from culture plates directly by 1 ml RNAzol and transferred into 1.5 ml tubes (Corning, MCT-150-C-S), without any washing steps. 0.4 ml RNase free water (Hyclone, SH30538.02) was added into each tube containing the samples, following by ~15 seconds vigorous shaking. The mixture was stored on ice for 15 mins, then centrifuged at 4 ℃, 12000 g for 15 minutes. Stratification was obvious after centrifuge, and 1 ml of the supernatant (~75%) was transferred into a 2 ml tube (Corning, MCT-200-C-S). Subsequently, 1 ml isopropanol was mixed with the supernatant in the new tube. New tube containing the mixture was stored on ice for 10 mins, and centrifuged at 4 ℃, 12000 g for 10 minutes. White RNA pellets could be seen at this time, and would be more visible after washing the pellets twice. The washing step required 0.6 ml 75% ethanol, with centrifuge settings at 4 ℃, 8000 g for 3 minutes, and discarded the supernatant.

The isolated RNA was stored in 1 ml 95% ethanol at -80 ℃. Before using the samples, tubes were centrifuged at 4 ℃, 8000 g for 3 minutes, and the supernatant was discarded. RNA pallets were resuspended in RNase free water, and vortexed at room temperature for 5 minutes for solubilization. RNA qualities were tested by Nanodrop and Agilent 2100 Bioanalyzer, and the samples with enough amount and good quality (mainly evaluated the RNA degradation by RIN value) would be used for library constructions. All the tips, tubes and reagents were RNase-free. Face masks were worn during the experiments to prevent the possible RNase effects from mouth.

The service of RNA sequencing was from Novogene, China. Briefly, mRNA was

enriched from the samples using poly-T oligo-attached magnetic beads, and library constructions were by NEBNext® Ultra™ RNA Library Prep Kit for Illumina (NEB, E7530). Sequencing platform was Illumina Hiseq, and the samples were 150 bp pair-end (PE150) sequenced. Around 6 G raw data in fastq format was generated for each sample.

2.7 Analysis of bulk RNA-seq data

For high quality data for the downstream analysis, raw data was filtered by removing low quality reads and reads containing adapter and poly-N and generated into clean data. Sequencing reads' alignment and assembly, quantification of transcript abundance were executed on Linux (version 2.6.32-754.35.1.el6.x86_64 x86_64) system, by the software STAR (version 2.7.8a) (Dobin et al., 2013) and RSEM (version 1.3.1) (B. Li & Dewey, 2011) separately. GRCh38 genomes and "Homo_sapiens.GRCh38.93.filtered.gtf" were used as the reference (Haeussler et al., 2019).

After obtaining the gene expression level, analyses were performed in R (version 4.0.4). Gene expression matrix was generated by R package "tximport" (version 1.18.0) (Soneson, Love, & Robinson, 2015), and the counts were normalized by GC content with the R package "EDASeq" (version 2.24.0) (Risso, Schwartz, Sherlock, & Dudoit, 2011). Any genes which had no normalized counts and less than 10^1.5 in at least two samples were taken as low expressed genes to be filtered out. Detection of differentially expressed genes was by R package "DESeq2" (version 1.30.1) (Love, Huber, & Anders, 2014). Gene set analysis was by R package "clusterProfiler" (version 3.18.1) (Wu et al., 2021; G. Yu, Wang, Han, & He, 2012), and the gene set database used for the analysis was downloaded from http://download.baderlab.org/EM_Genesets/current_release/Human/symbol/ (Gillespie et al., 2022; Rodchenkov et al., 2020).

For drawing the plots, correlation distance among the samples was calculated by R package "amap" (version 0.8-18) (Caussinus, Fekri, Hakam, & Ruiz-Gazen, 2003; Caussinus, Hakam, & Ruiz-Gazen, 2002), and the heatmaps were drawn by R package "gplots" (version 3.1.1) (Warnes, Bolker, Bonebakker, Gentleman, & Huber, 2016). Bar plots of the gene expression level were drawn by GraphPad Prism (version 8.0). MA plots were drawn by R

package “ggpubr” (version 0.4.0) (Kassambara, 2020) and R package “ggplot2” (version 3.3.5) (Wickham, 2011). All the figures from gene set analysis were drawn by R package “enrichplot” (version 1.10.2) (G. Yu, 2021).

In the analysis of the bulk RNA-seq data, obvious outlier samples were removed from among the samples. These outliers might cause a bias in the analysis, and subsequently lead to incorrect results. Two conditions together were used to define sample outliers: first, the outliers had a large correlation distance from all the other biological replicates with similar treatments; second, distribution of the outliers’ transcript counts (always presented in density plots) was evidently distinct from others.

2.8 Key materials and resources

	Source	Identifier
Reagents, kits and materials		
mTeSR™1	STEMCELL Technology	Cat #85850
6-well clear TC-treated plate	Corning	Cat #3516
Matrigel® hESC-Qualified Matrix	Corning	Cat #354277
ReLeSR™	STEMCELL Technology	Cat #05872
Accutase solution	Sigma-Aldrich	Cat #A6964
CryoStor® CS10	STEMCELL Technology	Cat #07930
Nunc™ Biobanking and Cell Culture Cryogenic Tube	ThermoFisher SCIENTIFIC	Cat #375418
Scientific™ Mr. Frosty™	ThermoFisher SCIENTIFIC	Cat #5100-0001
TransDetect PCR Mycoplasma Detection Kit	Transgen Biotech	Cat #FM311-01
40 μm Cell Strainer	Corning	Cat #352340
AggreWell™800	STEMCELL Technology	Cat #34815
Anti-Adherence Rinsing Solution	STEMCELL Technology	Cat #07010
AggreWell™ EB Formation Medium	STEMCELL Technology	Cat #05893
10 μM Y-27632	STEMCELL Technology	Cat #72302
BMP4	PeproTech	Cat #AF-120-05ET
VEGF	PeproTech	Cat #AF-100-20
SCF	PeproTech	Cat #AF-300-07
Axygen®1ml wide bore tip	Corning	Cat #TF-1005-WB-R-S

37 μm Reversible Strainer	STEMCELL Technology	Cat #27250
50 ml PP Centrifuge Tube	Corning	Cat #430829
DMEM/F12	Hyclone	Cat #SH30023.01
X-VIVO15	Lonza	Cat #04-418Q
GlutaMax	ThermoFisher Scientific	Cat #35050-061
Penicillin-Streptomycin Solution	Hyclone	Cat #SV30010
M-CSF	R&D Systems	Cat #216-MC
M-CSF	ThermoFisher Scientific	Cat #PHC9501
IL-3	R&D Systems	Cat #203-IL
IL-3	PeproTech	Cat #AF-200-03
STEMdiff™ Intestinal Organoid Kit	STEMCELL Technology	Cat #05140
STEMdiff™ Endoderm Basal Medium	STEMCELL Technology	Cat #05111
STEMdiff™ Definitive Endoderm Supplement CJ	STEMCELL Technology	Cat #05113
STEMdiff™ Gastrointestinal Supplement PK	STEMCELL Technology	Cat #05141
STEMdiff™ Gastrointestinal Supplement UB	STEMCELL Technology	Cat #05142
24-well clear TC-treated plate	Corning	Cat #3524
Corning® Matrigel® Growth Factor Reduced (GFR) Basement Membrane Matrix,	Corning	Cat #356231
STEMdiff™ Intestinal Organoid Basal Medium	STEMCELL Technology	Cat #05143
STEMdiff™ Intestinal Organoid Supplement	STEMCELL Technology	Cat #05144
Cultrex Organoid Harvesting Solution	R&D Systems	Cat #3700-100-01

3 ml Pasteur pipettes	NEST Biotechnology	Cat #318212
DPBS, 1X, without calcium, magnesium	Hyclone	Cat #SH30028.02
TrypLE™ Express Enzyme (1X), phenol red	ThermoFisher Scientific	Cat #12605010
Prestained Protein Ladder, 10 to 180 kDa	ThermoFisher Scientific	Cat #26616
Pierce™ ECL Western Blotting Substrate	ThermoFisher Scientific	Cat #32106
RNAzol® RT	Molecular Research Center	Cat #RN190
Axygen® 1.5 mL Snaplock tube	Corning	Cat #MCT-150-C-S
Axygen® 2.0 mL Snaplock tube	Corning	Cat #MCT-200-C-S
RNase free water	Hyclone	Cat #SH30538.02
NEBNext® UltraTM RNA Library Prep Kit for Illumina	NEB	Cat #E7530
(pSpCas9(BB)-2A-Puro (PX459) V2.0	Addgene	Cat #62988
EndoFree Maxi Plasmid Kit	TIANGEN	Cat #DP117
TIANamp Genomic DNA Kit	TIANGEN	Cat #DP304
SURVEYOR Mutation Detection Kit	Transgenomic	Cat #706020
Wright-Giemsa Stain solution	Solarbio	Cat #G1020
Stain Buffer (FBS)	BD Biosciences	Cat #554656
MACS Tissue Storage Solution	Miltenyi	Cat #130-100-008
Primocin™	Invivogen	Cat #ant-pm
60 mm TC-treated Culture Dish	Corning	Cat #430166
1x red blood cell lysis buffer	Biolegend	Cat #420301

gentleMACS C Tubes	Miltenyi	Cat #130-096-334
Multi Tissue Dissociation Kit 1	Miltenyi	Cat #130-110-201
gentleMACS Octo Dissociator with Heaters	Miltenyi	Cat #130-096-427
70 μm Cell Strainer	Corning	Cat #352350
Chromium Single Cell 3′ GEM, Library & Gel Bead Kit v3	10X Genomics	Cat #PN-1000092
IntestiCultTM Organoid Growth Medium (Human)	STEMCELL Technologies	Cat #06010
Antibodies		
STAT1 antibody	Cell Signaling Technology	Cat #9172S
Anti-beta Actin antibody	Abcam	Cat #ab8227
Goat Anti-Rabbit IgG(H+L), HRP Conjugate	Transgen Biotech	Cat #HS101-01
SSEA-4 antibody	Abcam	Cat #ab16287
Tra-1-60 antibody	Merck	Cat #MAB4360
OCT4 antibody	ThermoFisher Scientific	Cat #PA5-27438
NANOG antibody	R&D Systems	Cat #AF1997
Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	ThermoFisher Scientific	Cat #A21202
Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 555	ThermoFisher Scientific	Cat #A21428
Donkey Anti-Goat IgG NorthernLights™ NL557-conjugated Antibody	R&D Systems	Cat #NL001

FITC Mouse Anti-Human CD14	BD Biosciences	Cat #561712
FITC Mouse IgG2a, κ Isotype Control	BD Biosciences	Cat #556652
Software		
Linux (version 2.6.32-754.35.1.el6.x86_64 x86_64)	Linux team	www.linux.org
R (version 4.0.4)	R Development Core Team	http://www.R-project.org
GraphPad Prism (version 8.0)	GraphPad Software	https://www.graphpad.com/scientific-software/prism/
FlowJo (version 7.6)	FlowJo LLC	https://www.flowjo.com/
Cell Ranger (version 3.0.2)	10x Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/
CRISPOR (version 4.99)	Maximilian Haeussler	http://crispor.tefor.net
STRING (version 11.5)	Christian von Mering lab	https://string-db.org/
STAR (version 2.7.8a)	Alexander Dobin	https://github.com/alexdobin/STAR
RSEM (version 1.3.1)	Bo Li & Colin Dewey	https://github.com/deweylab/RSEM
R package “tximport” (version 1.18.0)	Bioconductor	https://bioconductor.org/packages/release/bioc/html/tximport.html
R package “EDASeq” (version 2.24.0)	Bioconductor	https://bioconductor.org/packages/release/bioc/html/EDASeq.html
R package “DESeq2” (version 1.30.1)	Bioconductor	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
R package “clusterProfiler” (version 3.18.1)	Bioconductor	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
R package “amap” (version 0.8-18)	CRAN	https://cran.r-project.org/web/packages/amap/index.html

R package “gplots” (version 3.1.1)	CRAN	https://cran.r-project.org/web/packages/gplots/index.html
R package “ggpubr” (version 0.4.0)	CRAN	https://cran.r-project.org/web/packages/ggpubr/index.html
R package “enrichplot” (version 1.10.2)	Bioconductor	https://bioconductor.org/packages/release/bioc/html/enrichplot.html
R package “Seurat” (version 4.0.2)	CRAN	https://cran.r-project.org/web/packages/Seurat/index.html
R package “SingleR” (version 1.4.1)	Bioconductor	https://bioconductor.org/packages/release/bioc/html/SingleR.html
R package “ggplot2” (version 3.3.5)	CRAN	https://cran.r-project.org/web/packages/ggplot2/index.html
R package “EnhancedVolcano” (version 1.8.0)	Bioconductor	https://www.bioconductor.org/packages/release/bioc/html/EnhancedVolcano.html
Online resources		
single cell RNA-seq data from CD patients	GEO	GEO: GSE134809

2.9 Ethical approval statement

For generation and usage of the human iPSC lines, the iPSC line used in this study was previously generated under the guidance of the ethics committee of the Guangzhou Institutes of Biomedicine and Health (Guangzhou, China) and the University of Natural Resources and Life Sciences (in Vienna, Austria).

Recruitment of CD patients' resections was under the ethical approval of the "TRAFIC" project (Tissue remodeling and fibrostenosis in Crohn's disease - a study of molecular signals and biomarkers), which was issued on the 13/06/2019 by Institute of Translational Medicine, University of Birmingham, UK.

CHAPTER 3

IN VITRO STUDIES OF THE RELATIONSHIP BETWEEN *NOD2* DEFICIENCY AND CD PATHOGENESIS

3.1 Background

NOD2 protein has a tripartite structure with two caspase recruitment domains (CARD) at the N-terminal end, a centrally located nucleotide binding and oligomerization domain (NOD) and a C-terminal set of ten leucine-rich repeats (LRRs) (Eckmann & Karin, 2005). Among the composition of the protein, LRRs are thought to be involved in ligand recognition, and the three main CD related variants (R702W, G908R, 1007fs) all alter the gene product within or close to the LRRs (Eckmann & Karin, 2005). Meanwhile, as described in section 1.3, *NOD2* was initially recognized as an intracellular microbial sensor for muramyl dipeptide (MDP). The structural based evidence combining with further experiments laid the foundations to suggest that dysregulations of MDP-mediated response is important when studying how *NOD2* deficiency functions in CD pathogenesis.

However, as mentioned in section 1.3, it seems *NOD2* associations with CD risk are more than just as an MDP-mediated response. Surveys from CD patients also support this idea, since only mutations in *NOD2* have such a strong relationship with CD risk, but for other receptors (e.g. *TLR2, TLR4*) of other bacterial components (e.g. PGN, LPS), which can cause inflammatory response similar to MDP, their mutations make much less contribution, if any, to increased CD risk (Arnott et al., 2004; L. Chen, Lin, Zhan, & Lv, 2012; Cheng et al., 2015; Gewirtz, 2007). In general, disabilities of other receptors for bacterial components can cause the accumulation of bacterial burden in the intestines, but it is surprising why *NOD2* is so special among them. This suggests that *NOD2* deficiency may cause some unrecovered effects not directly related to MDP-mediated response.

Meanwhile, for current research of the link between *NOD2* and CD risk, the main explanatory model assumes that effects from *NOD2* deficiency are shared in distinct cell types or tissues, which originate from inability to recognize MDP (Rochereau et al., 2021; Trindade

& Chen, 2020). Yet, cell type or tissue specific heterogeneity is general in all organisms, and for better understandings of the linkages between *NOD2* deficiency and CD pathogenesis, it is necessary to take this heterogeneity into consideration.

3.2 Aims

The aim of this chapter is to study the possible dysregulations caused by *NOD2* deficiency besides MDP-mediated response, and how those dysregulations influence intestinal immune homeostasis. *In vitro* hiPSC-derived macrophages and intestinal organoids were used as models for this exploration. Ultimately, the aim is to reveal possible abnormalities in the macrophages that can lead to insight about CD.

3.3 Materials and Methods

3.3.1 Generating *NOD2* knockout hiPSCs via CRISPR-Cas9

To generate the *NOD2* knockout hiPSC line, a middle exon (exon 4) of *NOD2* gene was selected as the target for introducing frame shift mutations. sgRNAs were designed by the online tool (http://crispor.tefor.net) (version 4.99), with "UCSC Dec2013 (hg38 analysis set)+SNPs:dbSNP148 Kaviar" as the reference genome (Concordet & Haeussler, 2018). PAM sequence was NGG, corresponding to the commercialized plasmid to be used (pSpCas9(BB)-2A-Puro (PX459) V2.0 (Addgene, #62988)) for experiments (Ran et al., 2013). Two sgRNAs with the fewest off-targets (Figure 3.3.1) from the design were synthesized. Both of them were constructed as sgRNA-expressing plasmids, transformed into competent cells for amplification, and extracted by EndoFree Maxi Plasmid Kit (TIANGEN, #DP117). The plasmids were first transfected into 293T cells to test the cleavage efficiency of sgRNAs. Genomic DNA were extracted by TIANamp Genomic DNA Kit (TIANGEN, #DP304), and the cleavage was detected by SURVEYOR Mutation Detection Kit (Transgenomic, #706020). In the experiments, two sgRNAs had the similar cleavage efficiency, so the one with fewer off-targets (36/rev) in hiPSCs was chosen.

Position/ Strand	Guide Sequence + *PAM* + Restriction Enzymes + Variants Only G- Only GG- Only A-	MIT Specificity Score	CFD Spec. score	Predicted Efficiency Show all scores Doench '16	Mor.-Mateos	Outcome Out-of-Frame	Lindel	Off-targets for 0-1-2-3-4 mismatches + next to PAM
36 / rev	GCGAGACTGAGCAGACACCG TGG Enzymes: *BstDSI, MnlI, BseDI, PspPI, Bme18I, DseDI, HpyCH4III* Cloning / PCR primers	86	93	73	65	62	85	0 - 0 - 0 - 5 - 60 0 - 0 - 0 - 0 - 2 65 off-targets
76 / fw	CGCTTCCTCAGTACCTATGA TGG Cloning / PCR primers	83	89	41	30	62	87	0 - 0 - 2 - 8 - 69 0 - 0 - 0 - 1 - 2 79 off-targets

Figure 3.3.1 Results of sgRNA design

The wild type hiPSCs were transfected by constructing a 36/rev-expressing plasmids, and 1ug/ml puromycin treatment for 24 hours were used to select cells with successful transfection. Genomic DNA of the cells was extracted for detecting cleavage by SURVEYOR assay (Qiu et al., 2004), T/A cloning and Sanger sequencing. Cells with NHEJ were selected as candidates, and two rounds of CRISPR-Cas9 screening were performed to ensure the target gene was biallelic knockout. Successfully generated cell line was cultured in mTeSR™1 for identifications of pluripotency markers (by immunofluorescence, which would be described in section 3.3.2), karyotype (by G-banding method, which would be described in section 3.3.3) and mycoplasma (methods in section 2.2). Maintenance of the *NOD2* knockout cells was the same as the wild type hiPSCs (see section 2.1 for more details). Alignment and analysis of Sanger sequencing results were by BioEdit (version 7.2.5).

partial sequence of NOD2 exon 4 which was input into the computational tool for design of sgRNA:

TGCCACATGCAAGAAGTATATGGCCAAGCTGAGGACCACGGTGTCTGCTCAGTCTC

GCTTCCTCAGTACCTATGATGGAGCAGAGACGCTCTGCCTGGAGGACATATACACA

GAGAATGTCCTGGAGGTCTGGGCAGATGTGGGCATGGCTGGACCCCCGCAGAAGA

GCCCAGCCACCCTGGGCCTGGAGGAGCTCTTC

Primers for identifying NOD2 knockout:

Primer-1: ACGGCAGTCACTGATGTCTT

Primer-2: ACAGTGTCCGCATCGTCATT

Components for PCR reactions:

Component	Volume
5×PrimeSTAR Buffer (Mg2+ plus)	10 μl
dNTP Mixture (2.5 mM)	4 μl
PrimeSTAR HS DNA Polymerase (2.5 U/μl)	0.5 μl
ddH_2O	32.5 μl
Primer-1(15 pmol)	1 μl
Primer-2(15 pmol)	1 μl
Templete	1 μl
Total volume	50 μl

PCR program for identifying NOD2 knockout:

98 ℃ for 5mins;

98 ℃ for 10s, 55 ℃ for 5s, 72 ℃ for 30s, 30 cycles;

72 ℃ for 10mins

3.3.2 Using immunofluorescence to verify pluripotency of the generated *NOD2* knockout hiPSCs

For immunofluorescence, cells were washed by DPBS for twice, then fixed by 4% paraformaldehyde for 10 minutes at room temperature. Next, the paraformaldehyde was discarded and the cells were washed by DPBS for twice, then treated by 0.2% Triton X-100 for 30 minutes (permeabilization) (S. Chen et al., 2020). After discarding the 0.2% Triton X-100, cells were washed by DPBS for twice, and blocked by 3% BSA for 1 hour at room temperature, then incubated with diluted primary antibodies overnight at 4℃ (S. Chen et al., 2020). After being washed by DPBS for three times, cells were incubated with fluorescent dye conjugated secondary antibodies for 2 hours at room temperature in a dark environment (S. Chen et al., 2020). Nuclei were stained by 5 ng/ml DAPI, and the images were acquired by fluorescence microscope with OLYMPUS DP73 camera.

Antibodies used in identifying pluripotency markers:

SSEA-4 (Abcam, ab16287) (working dilution: 1:60);

Tra-1-60 (Merck, MAB4360) (working dilution: 1:60);

OCT4 (ThermoFisher Scientific, PA5-27438) (working dilution: 1:200);

NANOG (R&D Systems, AF1997) (working dilution: 1:60);

Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 488 (ThermoFisher Scientific, A21202) (working dilution: 1:2000);

Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 555 (ThermoFisher Scientific, A21428) (working dilution: 1:100);

Donkey Anti-Goat IgG NorthernLights™ NL557-conjugated Antibody (R&D Systems, NL001) (working dilution: 1:200)

3.3.3 Karyotyping analysis of the generated NOD2 knockout hiPSCs

For karyotyping by G-banding, cells cultured in the 6-well plate were washed by DPBS and treated by 100 ng/ml colcemid for 1 hour (metaphase arrest), then the cells were collected by Accutase solution and resuspended in 10 ml hypotonic solution with incubation at 37℃ for 15 mins. After centrifuge at room temperature for 200g, 3mins, the cells were resuspended in 1 ml fixation solution for 30 minutes in a dark environment (S. Chen et al., 2020). With centrifuge at room temperature for 200g, 3mins, the supernatant was discarded with only around 100 μl solution left. The cells were resuspended in the 100 μl solution, and added on the glass slide dropwise, and tried at room temperature. Chromosomes were stained by Giemsa, and observed under OLYMPUS BX41/BX2 for analysis. Components of hypotonic solution were 200 ml dH_2O with 0.56 g KCl and 0.5 g sodium citrate (S. Chen et al., 2020). Fixation solution was the mixture of methanol and acetic acid at the ratio 3:1 (S. Chen et al., 2020).

3.3.4 Characterization of hiPSC-derived macrophages

Wright-Giemsa Staining was by Wright-Giemsa Stain solution (Solarbio, G1020) in the plates directly. After staining for 3 minutes and washed by DPBS for several times, cells were observed under microscope to identify the morphology.

The expression of CD14 protein in the hiPSC-derived macrophages was identified by flow cytometry with isotype control. The cells were detached by 0.5 mM EDTA, washed by the cold Stain Buffer (FBS) (BD Biosciences, #554656) for twice, and resuspended the cells' concentration to 2 x 10^7 cells/ml by the cold Stain Buffer (FBS). Two clean tubes were prepared, and 50 μl aliquots of the of the cell suspension, 30μl Stain Buffer (FBS) and 20μl pre-diluted antibody were added into each tube. The tubes were incubated on ice for 45 mins with no lights, and the cells were washed twice by 1 ml Stain Buffer (FBS) for removing the unbound antibodies. The cell pellets in each tube were resuspended by 0.5 ml Stain Buffer (FBS), and analyzed by flow cytometry with FlowJo (version 7.6) as soon as possible. The antibodies used in the experiments were FITC Mouse Anti-Human CD14 (BD Biosciences, #561712) and FITC Mouse IgG2a, κ Isotype Control (BD Biosciences, #556652), which were both pre-diluted reagents. Under the microscope, the hiPSC-derived macrophages are also observed to phagocytize the ink when the ink drops were added into the plates (not shown in the results).

3.4 Results

3.4.1 Generation of *NOD2* knockout hiPSCs

Genomic DNA of the cell clones after puromycin treatment was extracted. After PCR amplification, DNA was sequenced to select the clones with possibly biallelic frame shift mutations. For those selected clones, T/A cloning of the PCR products was performed to confirm the results. Unfortunately, in the first round of knockout, no clones had frame shift mutations caused by biallelic NHEJ, and a clone ($NOD2^{T\ ins/wt}$) with monoallelic T insertion was generated (Figure 3.4.1.1 and Figure 3.4.1.2).

The $NOD2^{T\ ins/wt}$ cells were cultured and knocked out for the second round using the same sgRNA. Clones with biallelic frame shift mutations were selected as in the first round, and the required *NOD2* knockout hiPSCs (GT insertion was generated on wild type alleles) were obtained (Figure 3.4.1.3 and Figure 3.4.1.4).

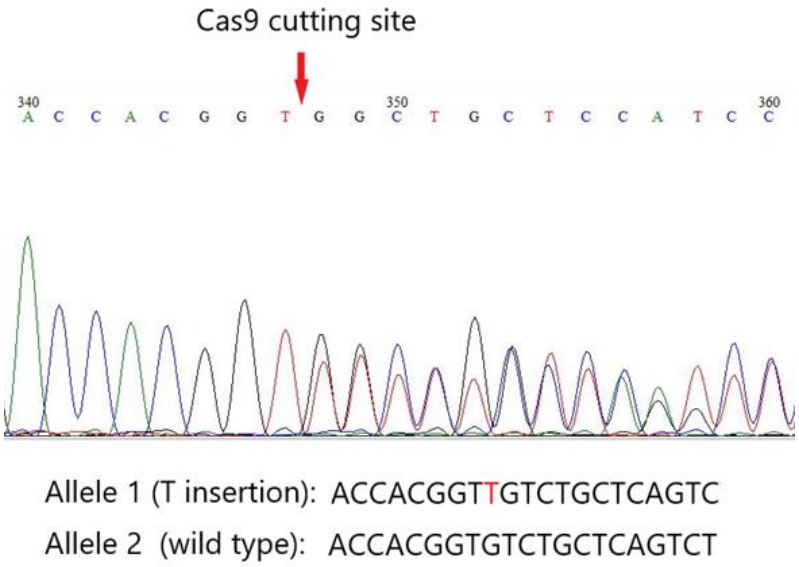


Figure 3.4.1.1 Sanger sequencing results of PCR products after first round knockout

(Cas9 cutting site was marked, and it was obvious that there was a monoallelic T insertion after the cutting site, which formed the following nested peaks)

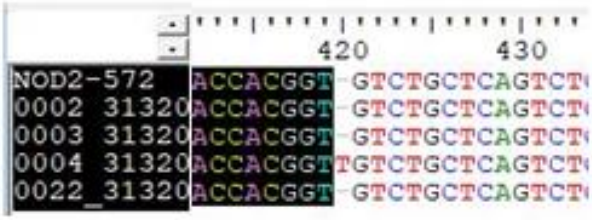


Figure 3.4.1.2 Sanger sequencing results of T/A cloning in the first round

(NOD2-572 was the reference sequence from human GRCh38 genome. Sequencing results of T/A cloning were aligned to the reference, and confirmed the monoallelic T insertion (Cas9 cutting site was just after the shadow region))

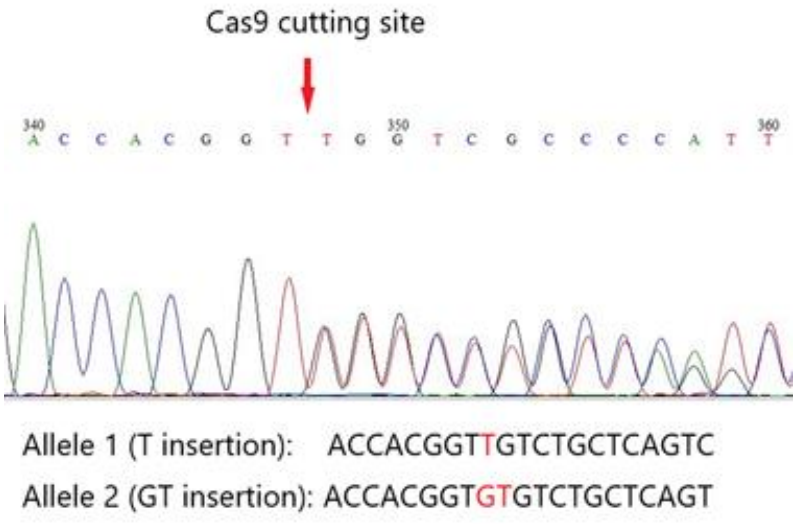


Figure 3.4.1.3 Sanger sequencing results of PCR products after second round knockout

(Cas9 cutting site was marked, and it was obvious that the wild type allele in Figure 3.4.1.1 was replaced by a GT insertion, with the nested peaks after the cutting site)

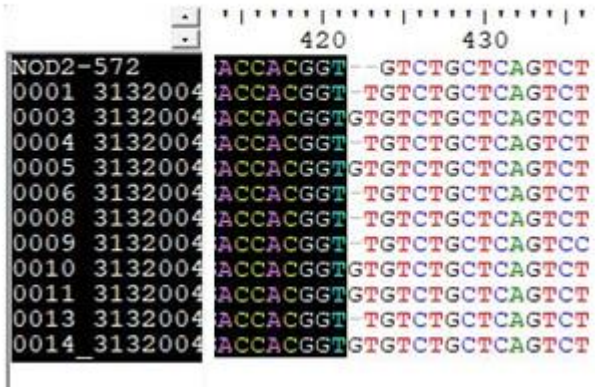


Figure 3.4.1.4 Sanger sequencing results of T/A cloning in the second round

(Sequencing results of T/A cloning were aligned to the reference (NOD2-572), and confirmed the T/GT insertion in the generated cell line (Cas9 cutting site was just after the shadow region))

Pluripotency of the generated *NOD2* knockout hiPSCs was confirmed by immunofluorescence of four markers (*SSEA-4*, *Tra-1-60*, *OCT4*, *NANOG*), and the results indicated the cells maintained pluripotency after genome editing (Figure 3.4.1.5).

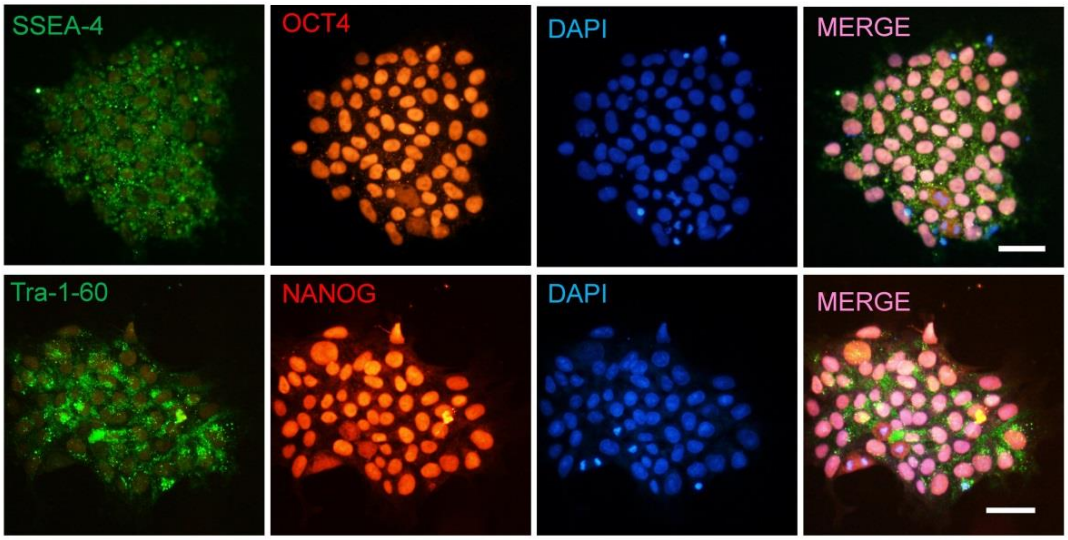


Figure 3.4.1.5 Immunofluorescence results of the pluripotency markers

(Four markers were separated into two batches (*SSEA-4*, *OCT4*; *Tra-1-60*, *NANOG*) for identifying the pluripotency of the cells. Nuclei were stained by DAPI in both batches, and the three images in each batch were merged as in the last one)

Karyotype of the generated *NOD2* knockout hiPSCs was identified by G-banding method, and the results indicated the resulting *NOD2* knockout cells had a normal karyotype (Figure 3.4.1.6).

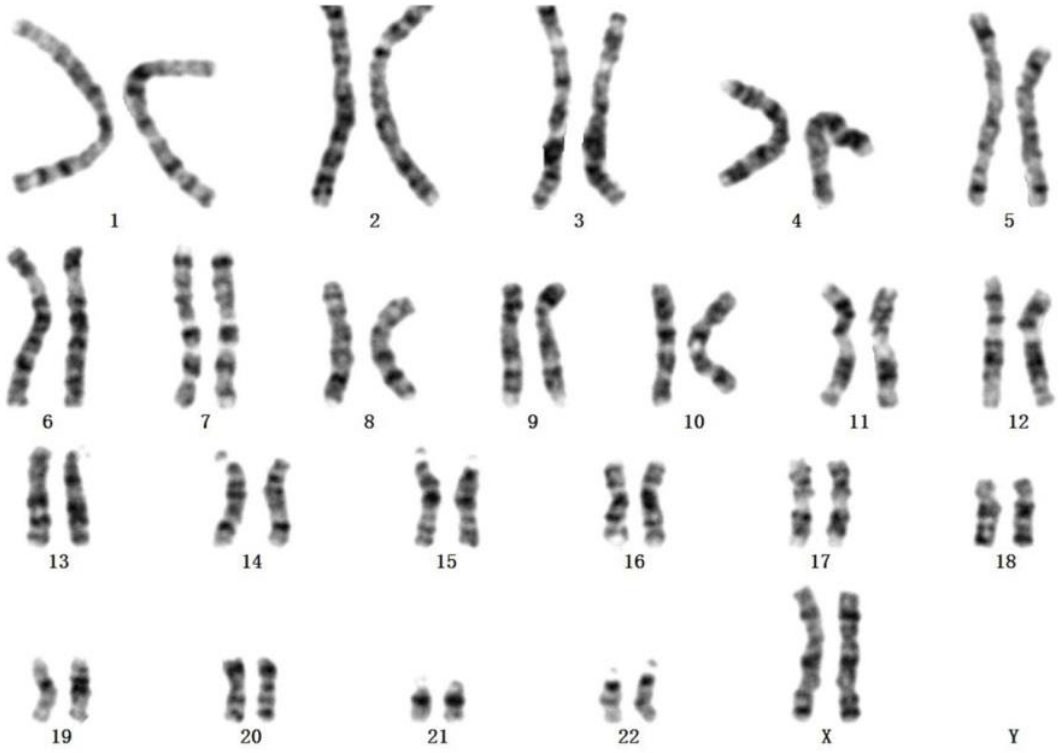


Figure 3.4.1.6 Results of karyotyping by G-banding

Mycoplasma test was performed with both positive and negative controls. The results indicated that there was no mycoplasma contamination in the cultured *NOD2* knockout hiPSCs. Wild type and *LRRK2* knockout hiPSCs were also tested at the same time, and both of them had no mycoplasma contaminations (Figure 3.4.1.7).

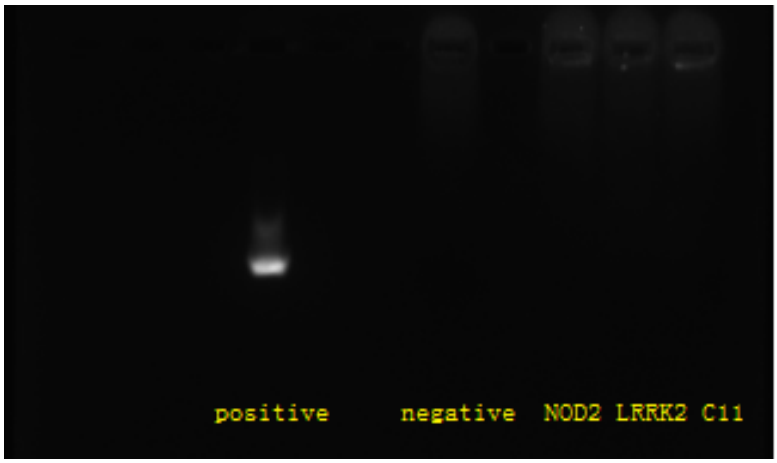


Figure 3.4.1.7 Results of mycoplasma test

(From left to right: positive control, negative control, *NOD2* knockout hiPSCs, *LRRK2* knockout hiPSCs, wild type hiPSCs. Comparing with the positive control, mycoplasma was detected in neither of the hiPSC lines)

3.4.2 *In vitro* differentiation of *NOD2* deficient macrophages

NOD2 deficient macrophages were differentiated from the *NOD2* knockout hiPSCs *in vitro* (Figure 3.4.2.1).

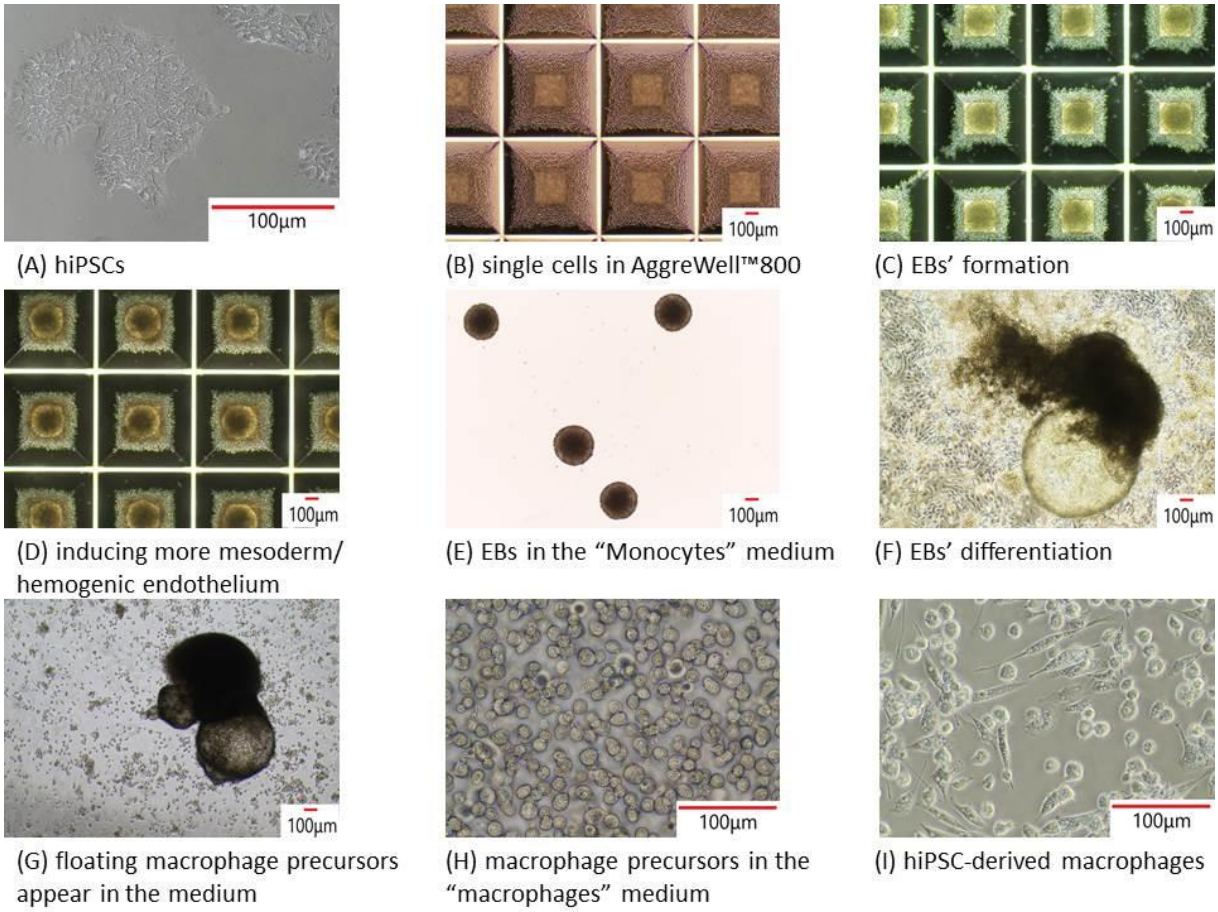


Figure 3.4.2.1 In vitro differentiation of *NOD2* deficient macrophages
(Figure (A) to (I) displayed the process of the in vitro differentiation. *NOD2* knockout hiPSCs were cultured in mTeSR™1 (A), then dissociated into single cells and spun into the micro wells of AggreWell™800 containing AggreWell™ EB Formation Medium (B). After 24 hours incubation, EBs formed (C) and the medium started to be 75% changed every day for 4 to 6 days (D). Next, EBs were collected into "Monocytes" medium with wide bore tips (E). With two third of medium change every 7 to 10 days, EBs differentiation could be observed (F). After changing medium regularly for 4 to 6 weeks, floating macrophage precursors appeared, and "cleared up" the debris and small dead cell in the supernatant previously (G). The macrophage precursors were collected into "macrophages" medium (H). After 7 days culture, precursors adhered to the plate and completed the further differentiation into NOD2 deficient macrophages (I))

Wild type hiPSCs were differentiated into macrophages *in vitro*, for comparison with *NOD2* deficient ones (Figure 3.4.2.2).

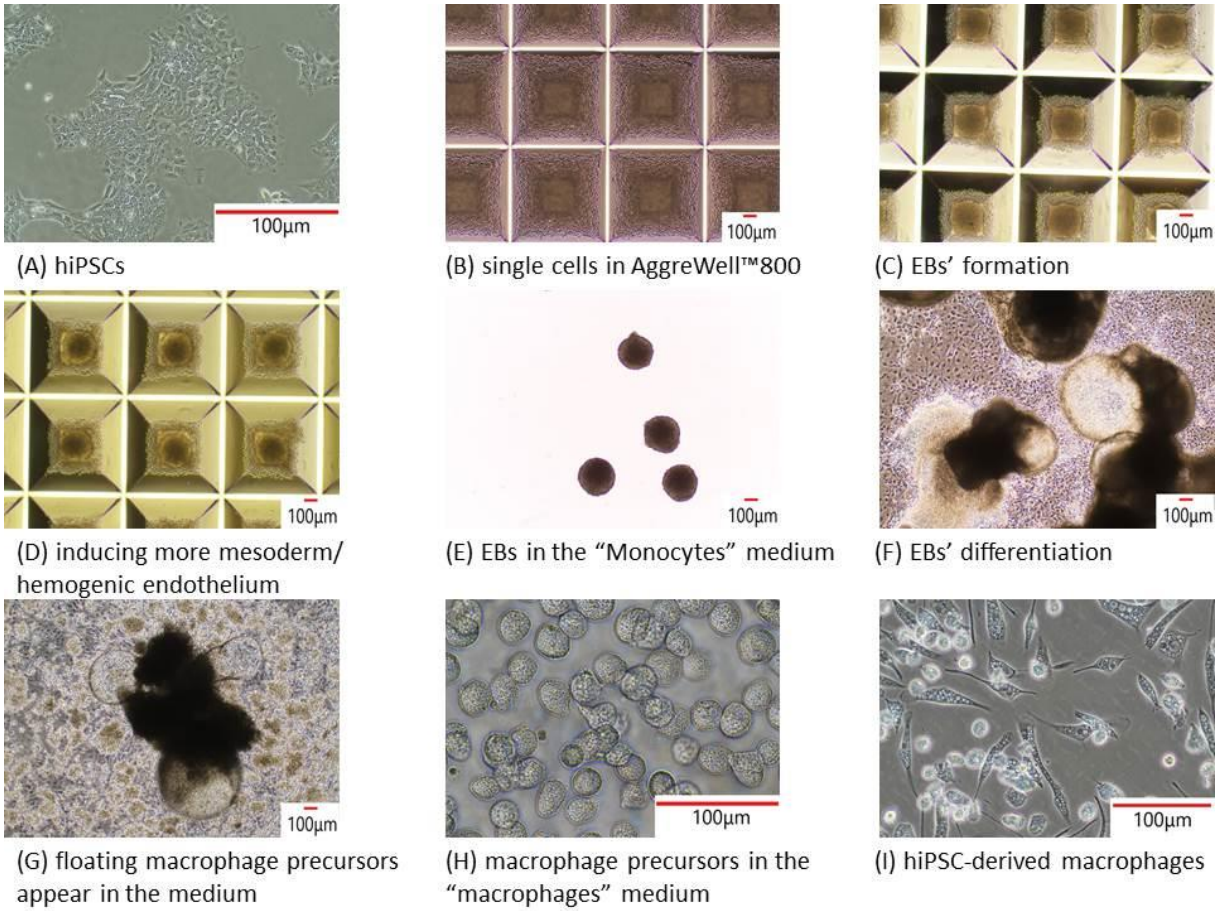


Figure 3.4.2.2 *In vitro* differentiation of wild type macrophages

(Figure (A) to (I) displayed the process of the in vitro differentiation. The wild type hiPSCs followed the same stages as the *NOD2* knockout hiPSCs, and derived into wild type macrophages *in vitro*.

Some wild type macrophages were stained by Wright-Giemsa solution, and the results indicated *in vitro* differentiated macrophages had the similar morphology as the natural macrophages (Figure 3.4.2.3).

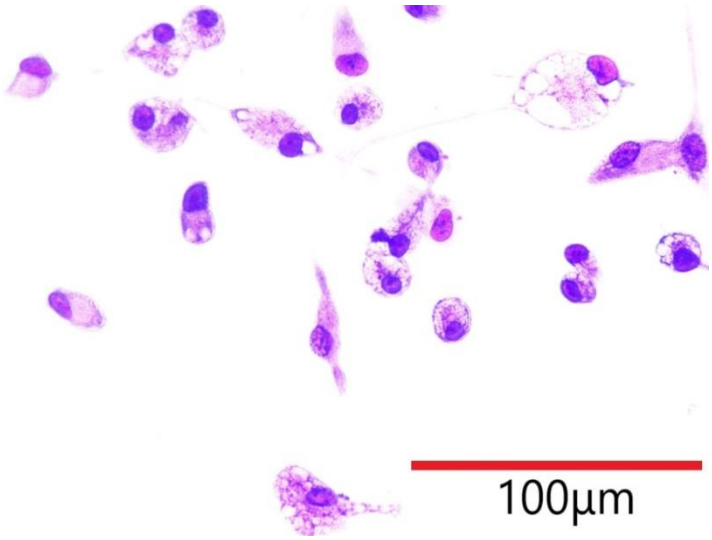


Figure 3.4.2.3 Wright-Giemsa staining of wild type hiPSC-derived macrophages
(The nuclei and large phagocytic vesicles of macrophages were clearly displayed in the figure.)

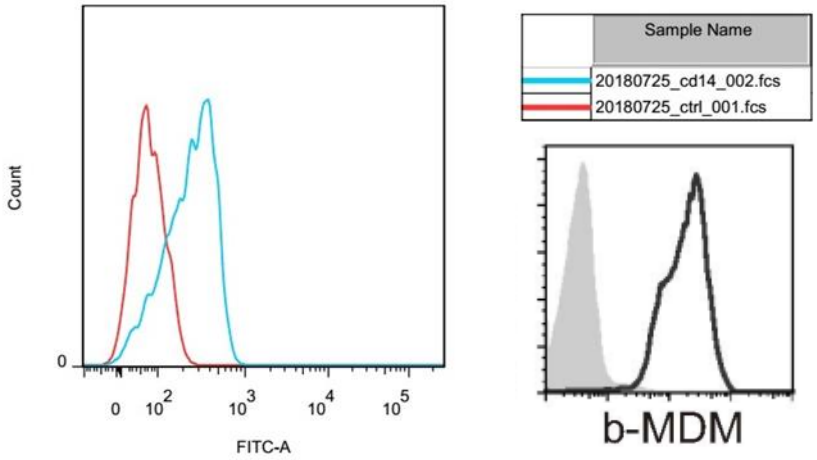


Figure 3.4.2.4 CD14 expression in wild type hiPSC-derived macrophages
(The expression of CD14 protein was identified by flow cytometry, blue curve was CD14 and red curve was its isotype control. This confirmed the CD14 expression, and the comparison between CD14 and isotype control was similar as the situation in the blood-derived macrophages (b-MDM) from the literature (Wilgenburg et al., 2013).)

3.4.3 Untreated *NOD2* deficient macrophages may have primed down regulations in IFN-γ and LPS mediated inflammatory response

To reveal how *NOD2* deficiency affects macrophages, RNA-seq based studies were started using macrophages without any treatment. This will help confirm that the in vitro-derived macrophages are a good model for macrophage biology and identify the impact of NOD2 knockout on in vitro macrophages, if any. Before doing so, the morphology of macrophages with and without *NOD2* deficiency were compared first, and it was hard to identify any differences between them (Figure 3.4.3.1).

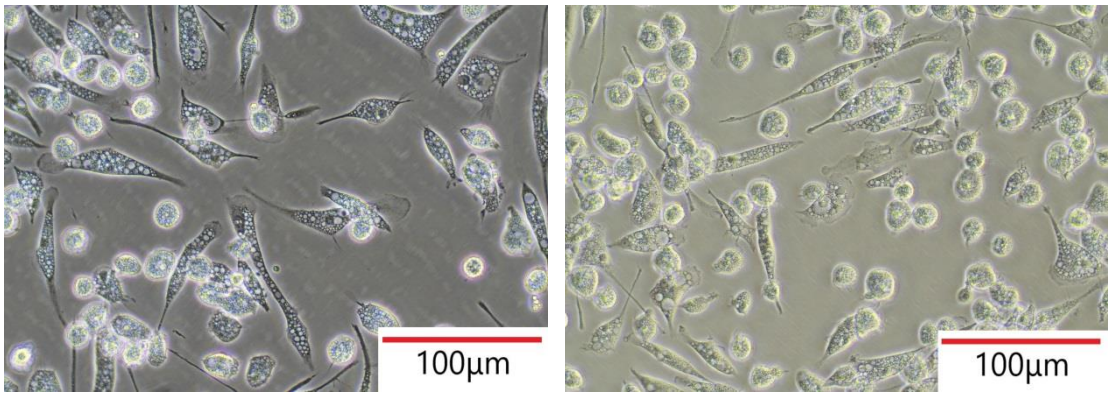


Figure 3.4.3.1 Comparison between the morphology of the wild type and *NOD2* deficient macrophages

(left: wild type *in vitro*-derived macrophages;

right: NOD2 deficient *in vitro*-derived macrophages)

Total RNA of wild type and *NOD2* deficient macrophages were extracted for bulk RNA sequencing, with nine (wildtype_untreated_rp1-9) and three (nod2ko_untreated_rp1-3) biological replicates respectively. After generating the expression matrix, correlation distances between the samples were calculated, and the results were displayed by heat map (Figure

3.4.3.2). The samples " wildtype_untreated_rp9" and " nod2ko_untreated_rp1" might be the outliers, and the distributions of their transcript counts (presented in density plots) also supported this idea (Figure 3.4.3.3). In this situation, the two outliers were removed from the downstream analysis.

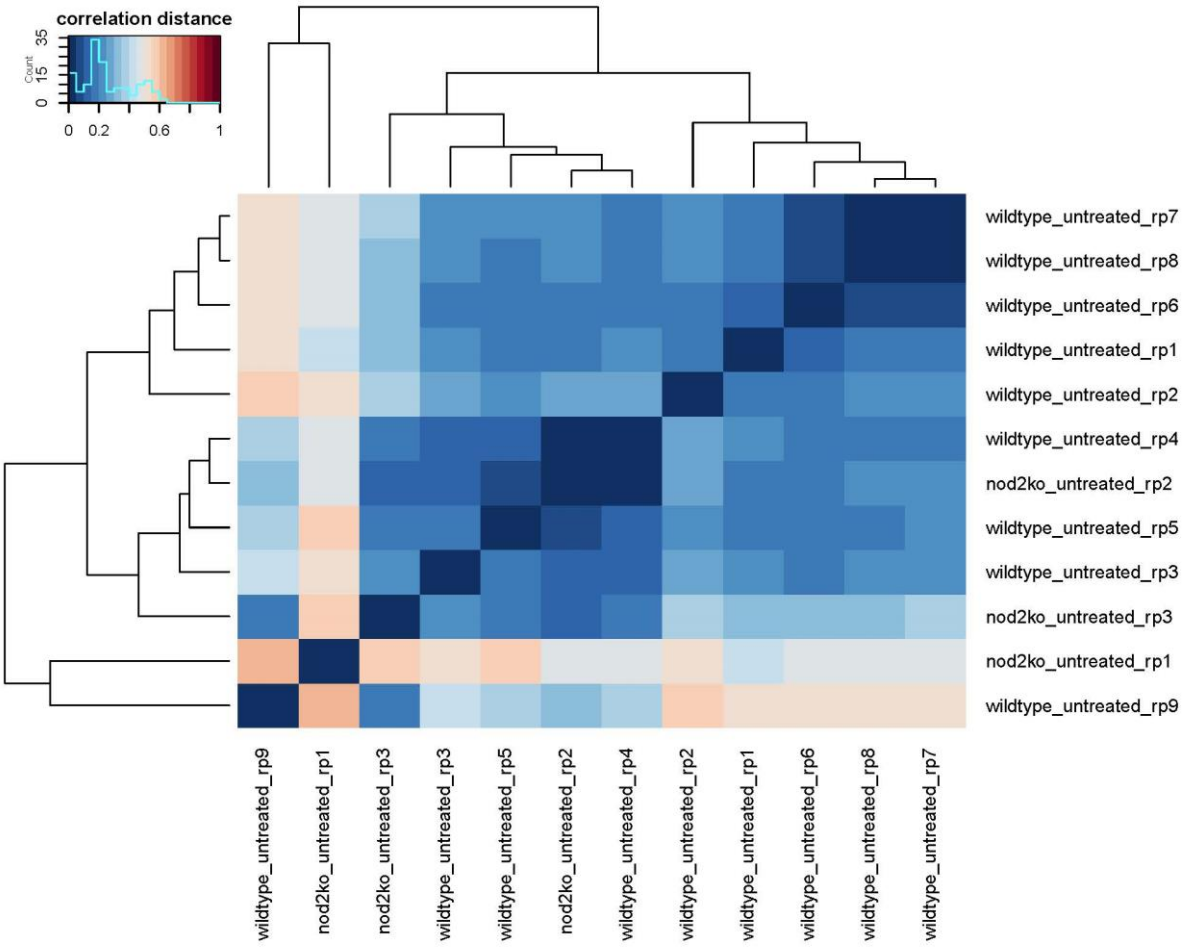


Figure 3.4.3.2 Correlation distances between the wild type and *NOD2* deficient macrophages

(Samples' correlation distance was displayed by heat map, and the colour change from blue to red represents the correlation distance from small to large between 0 and 1. 0 meant the two samples were the same and 1 meant they were distinct. Dendrogram was computed and reordered based on row and column means respectively. The heatmap was drawn by "heatmap.2" function from R package "gplots" (version 3.1.1))

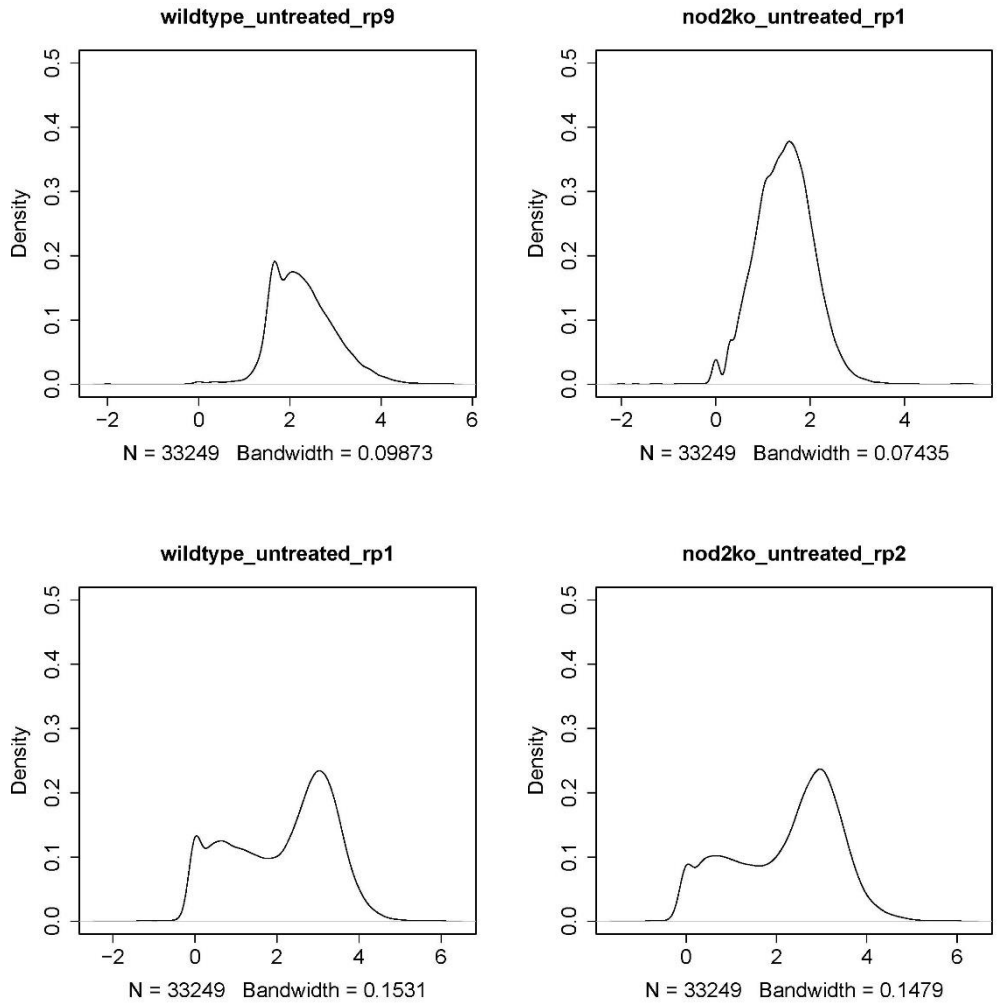


Figure 3.4.3.3 Density plots of transcript counts in the possible outliers and their biological replicates

(In the figure, the upper two plots were the possible outliers, and the corresponding plots in the bottom contained the typical curves of their biological replicates. It was obvious that the curves in the upper plots were abnormal, and are evidence of a quality control failure in the RNA-seq)

After DGE analysis by R package "DESeq2" (version 1.30.1), DE genes in the *NOD2* deficient macrophages were selected by the criteria "p.adj < 0.01" (Bonferroni-Hochberg multiple test corrected P-value) and "fold change>2". 582 up regulated and 778 down regulated genes were identified, and marked by red and blue points separately in the MA plot (Figure 3.4.3.4).

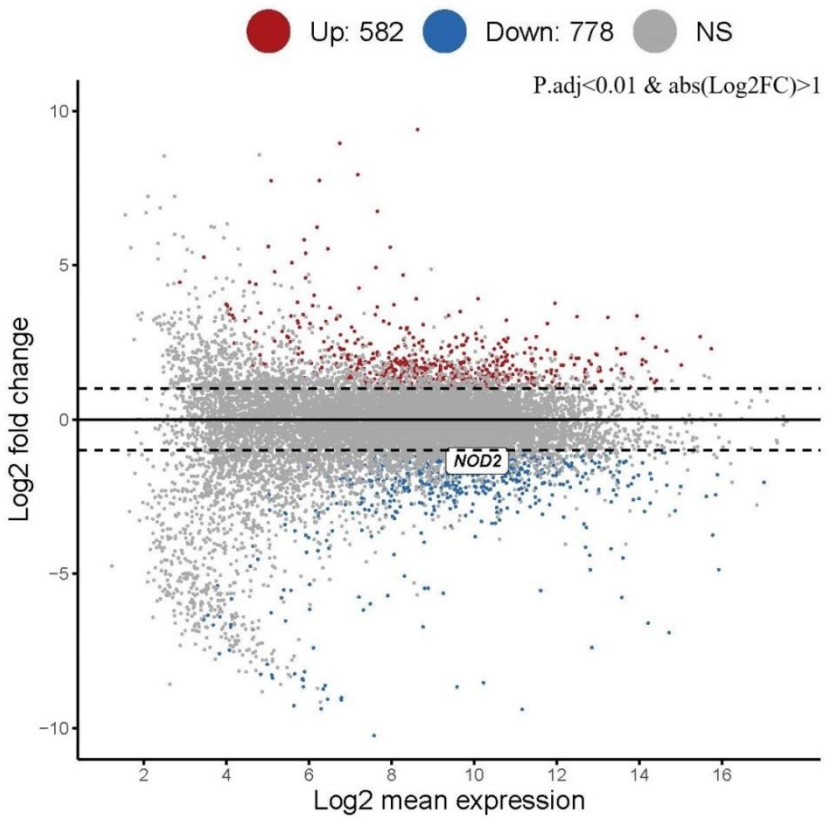


Figure 3.4.3.4 MA plot for the results from DGE analysis of *NOD2* deficient macrophages

(Data from wild type macrophages was taken as the reference in the DGE analysis. Each point in the plot was a gene. x-axis was the log2 transformation of genes' mean expression among wild type and *NOD2* deficient macrophages, y-axis was the log2 transformation of gene expression fold change in *NOD2* deficient macrophages comparing with the wild type. "NS" meant non-significant genes and are marked in grey, "Up" and "Down" meant identified up and down regulated genes which are marked in red and blue respectively)

Over-representation analysis (ORA) on those up and down regulated genes were performed respectively to reveal the implications among them. Top 20 enriched terms with the lowest adjusted p-value were displayed by bar plots (Figure 3.4.3.5 and Figure 3.4.3.6). The results indicated *NOD2* deficiency might lead to some significant biological changes on the macrophages. For up regulated genes, enriched terms with the lowest adjusted p-value concentrated on the protein translation-related processes. Meanwhile, the down regulated genes were significantly enriched in terms associated with the response to interferon-gamma (IFN-γ), as well as response to other types of interferons.

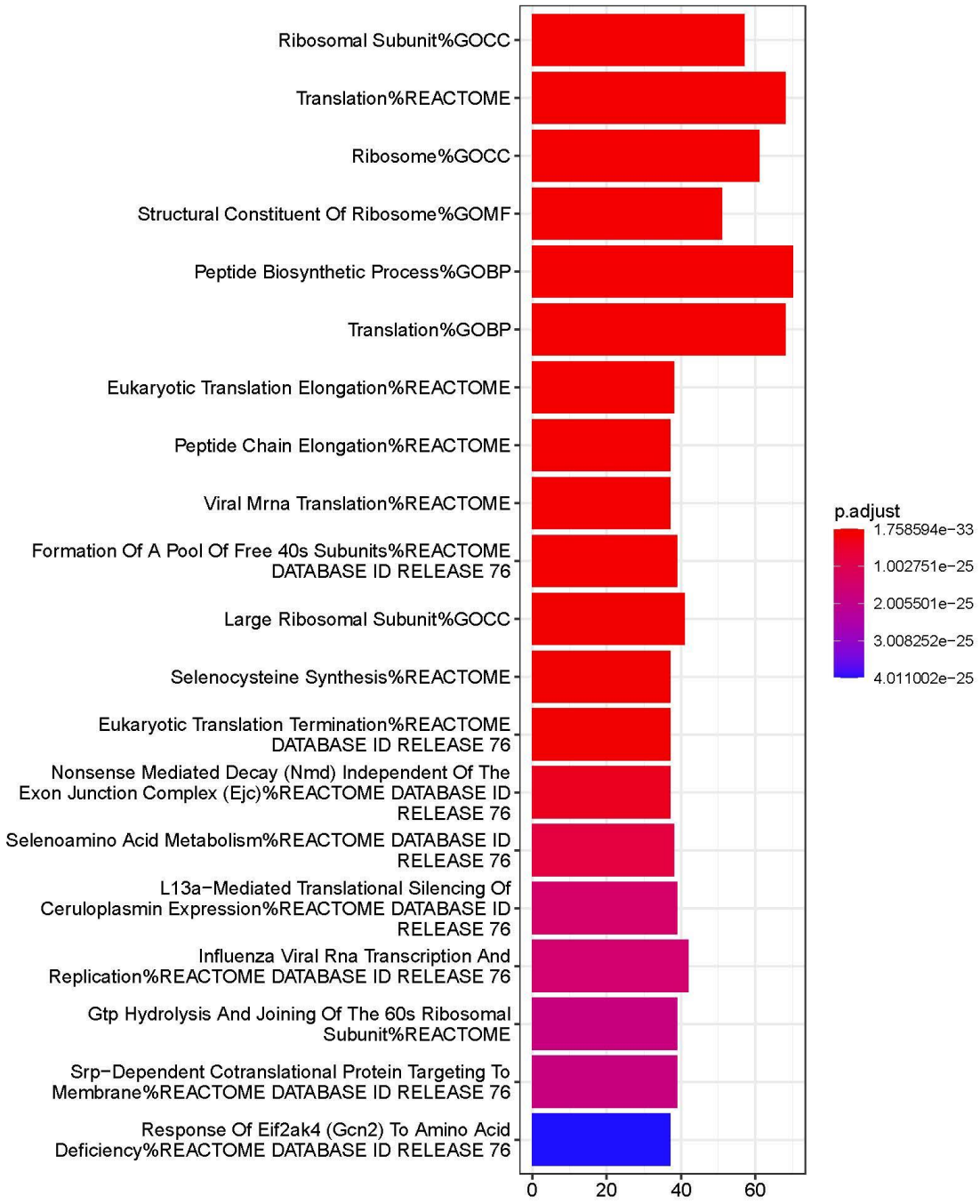


Figure 3.4.3.5 ORA results of up regulated genes in untreated *NOD2* deficient macrophages

(The bar plot displayed the top 20 enriched terms with the lowest adjusted p-value. Value of x-axis was the number of genes enriched to the term, adjusted p-value for each term corresponded to the colour in the legend)



Figure 3.4.3.6 ORA results of down regulated genes in untreated *NOD2* deficient macrophages

(The bar plot displayed the top 20 enriched terms with the lowest adjusted p-value. Value of x-axis was the number of genes enriched to the term, adjusted p-value for each term corresponded to the colour in the legend)

Functional class scoring analysis were also performed by GSEA, based on the whole gene expression information. Surprisingly the results indicated a down regulation of cellular response to IFN-γ in *NOD2* deficient macrophages (Figure 3.4.3.7), as well as the IFN-γ Signaling (Figure 3.4.3.8). This was surprising as the macrophages had not been treated with IFN-γ, and suggested *NOD2* deficient macrophages may have primed down regulation of IFN-γ mediated inflammatory response. GSEA results also identified the down regulation of response to other types of interferons as ORA (Figure 3.4.3.9). Moreover, the analysis identified the down regulation of response to lipopolysaccharide (LPS, an exogenous pro-inflammatory stimulus) (Figure 3.4.3.10) in those untreated *NOD2* deficient macrophages, which indicated the possibly primed down regulation of LPS mediated inflammatory response. Meanwhile, the overall down regulation of inflammatory response was truly identified from those untreated *NOD2* deficient macrophages (Figure 3.4.3.11).

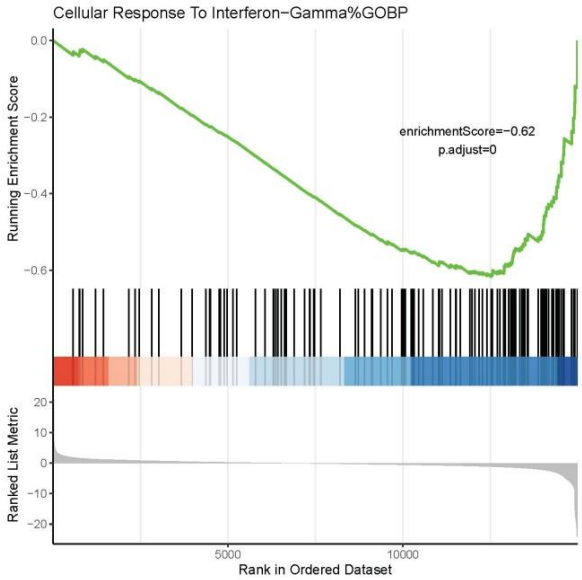


Figure 3.4.3.7 Down regulation of "Cellular Response To Interferon-Gamma" (GO: 0071346) in untreated *NOD2* deficient macrophages

(In the plot which displayed the results of GSEA, the green curve represented the changes of enrichment score along with the expressing genes ranked by fold change from high to low. Each black line was an expressing gene from the specific gene set. Colour changes from red (fold change > 0) to blue (fold change < 0) and the ranked list metric displayed the distributions of expressing genes' fold change with log2 transformation)

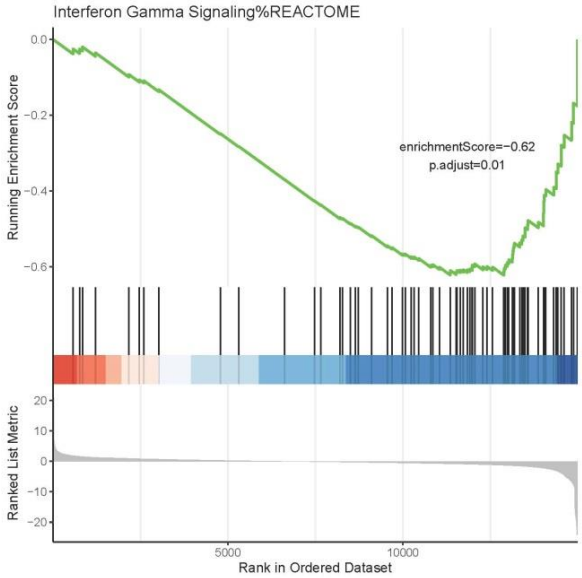


Figure 3.4.3.8 Down regulation of "Interferon Gamma Signaling" (REACTOME: R-HSA-877300.3) in untreated *NOD2* deficient macrophages

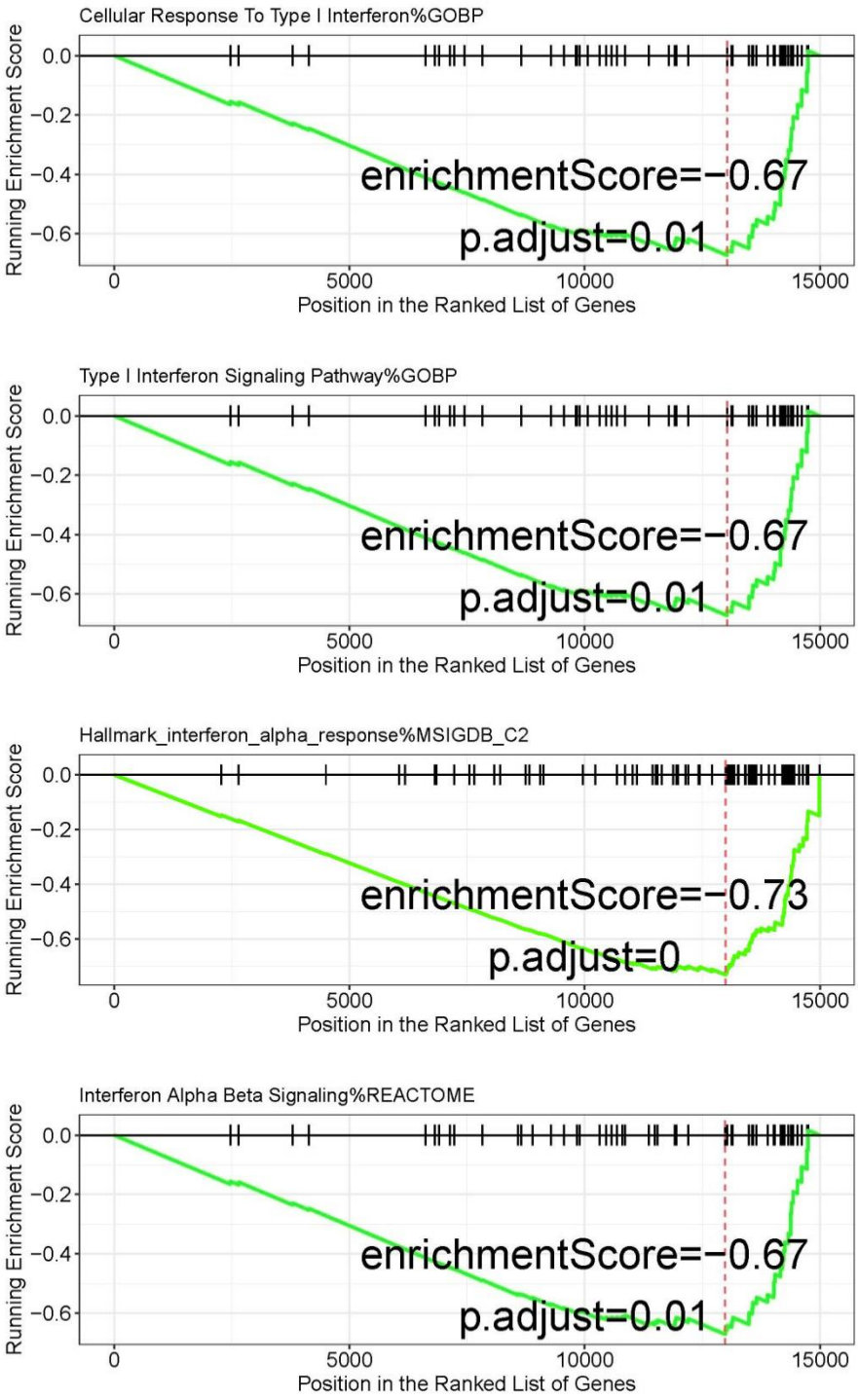


Figure 3.4.3.9 Down regulations of response to other types of interferons in untreated *NOD2* deficient macrophages

("Cellular Response To Type I Interferon" (GO: 0071357);

"Type I Interferon Signaling Pathway" (GO: 0060337);

"Hallmark interferon alpha response" (MSIGDB_C2);

"Interferon Alpha Beta Signaling" (REACTOME: R-HSA-909733.4))

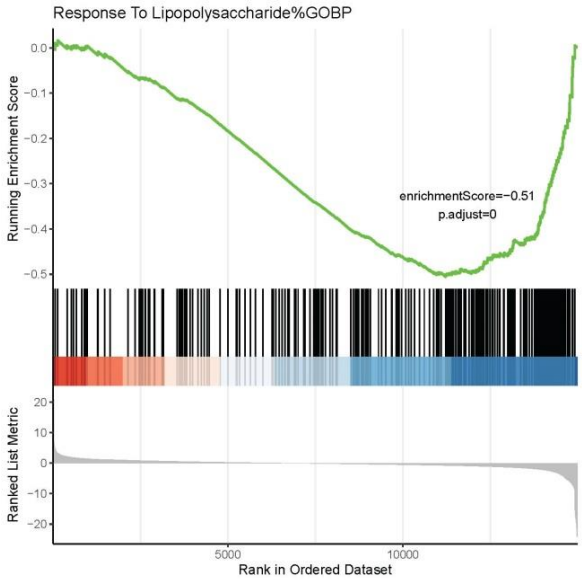


Figure 3.4.3.10 Down regulation of "Response To Lipopolysaccharide (GO: 0032496)" in untreated *NOD2* deficient macrophages

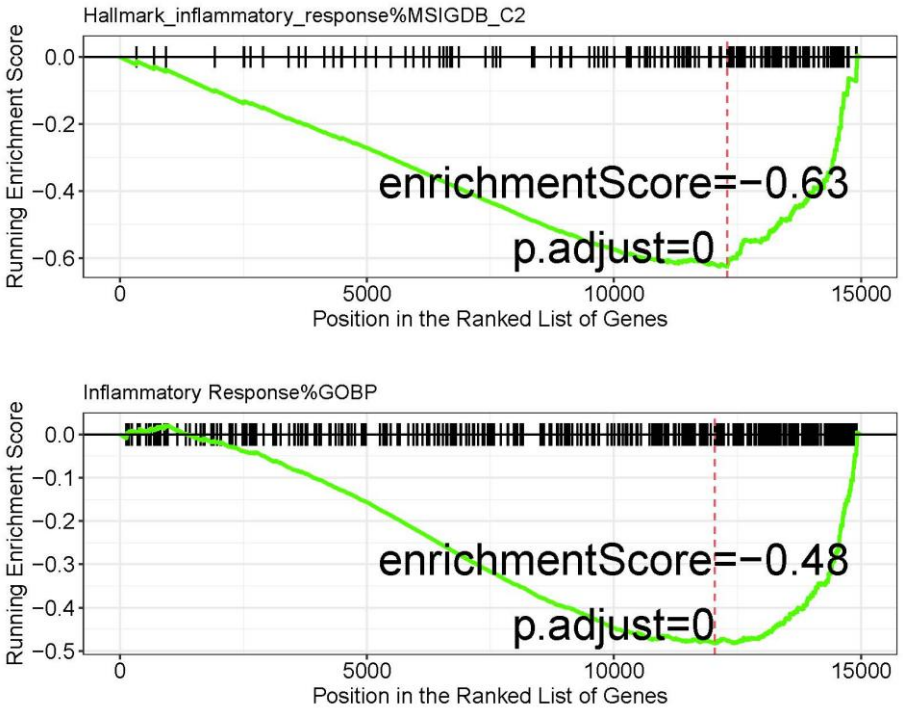


Figure 3.4.3.11 Primed down regulation of inflammatory response in *NOD2* deficient macrophages

("Hallmark inflammatory response" (MSIGDB_C2);

"Inflammatory Response" (GO: 0006954))

3.4.4 *NOD2* deficiency has surprisingly little influence on MDP-mediated inflammatory response in macrophages

As *NOD2* has been recognized as the receptor of MDP for quite a long time, it was very possible *NOD2* deficiency could make huge influence in the MDP treated macrophages. For verifying the hypothesis, *NOD2* deficient and wild type macrophages were stimulated with 1ug/ml MDP (Invivogen, tlrl-mdp), and collected samples in a time course for bulk mRNA sequencing. For reasons that are unclear, MDP-treated samples tended to have lower overall RNA quality. It is possible that residual MDP somehow interferes with RNA extraction or purification, although there are no reports of this effect in the literature. Due to the relatively low quality of extracted RNA, data from only a limited selection of samples was obtained, which were: wild type macrophages with 6 hours treatment (three biological replicates, wildtype_mdp_6h_rp1-3); 12 hours treatment (one biological replicate, wildtype_mdp_12h_rp1); 24 hours treatment (one biological replicate, wildtype_mdp_24h_rp1), and *NOD2* deficient macrophages with 12 hours treatment (one biological replicate, nod2ko_mdp_12h_rp1); 24 hours treatment (one biological replicate, nod2ko_mdp_24h_rp1).

The correlation distance between those samples were calculated, and no obvious outliers amongst them were found (Figure 3.4.4.1).

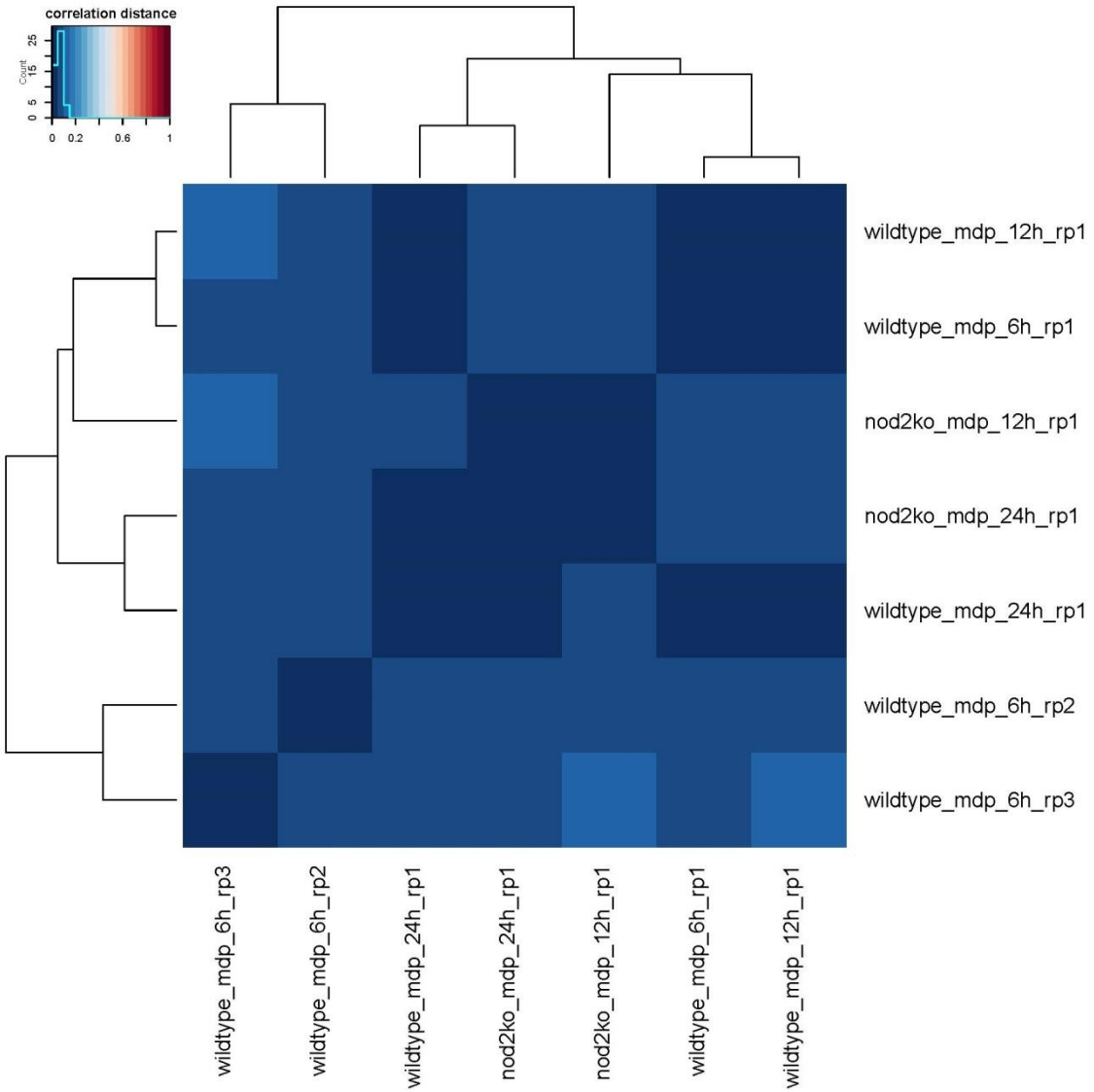


Figure 3.4.4.1 Correlation distance of wild type and *NOD2* deficient macrophages with MDP treatment

It was obvious that the hiPSC-derived macrophages could strongly respond to MDP, via the DGE analysis between wild type macrophages with and without MDP treatment (Figure 3.4.4.2). Meanwhile, GSEA results indicated antimicrobial response in those macrophages could be significantly up regulated after treated by MDP (Figure 3.4.4.3).

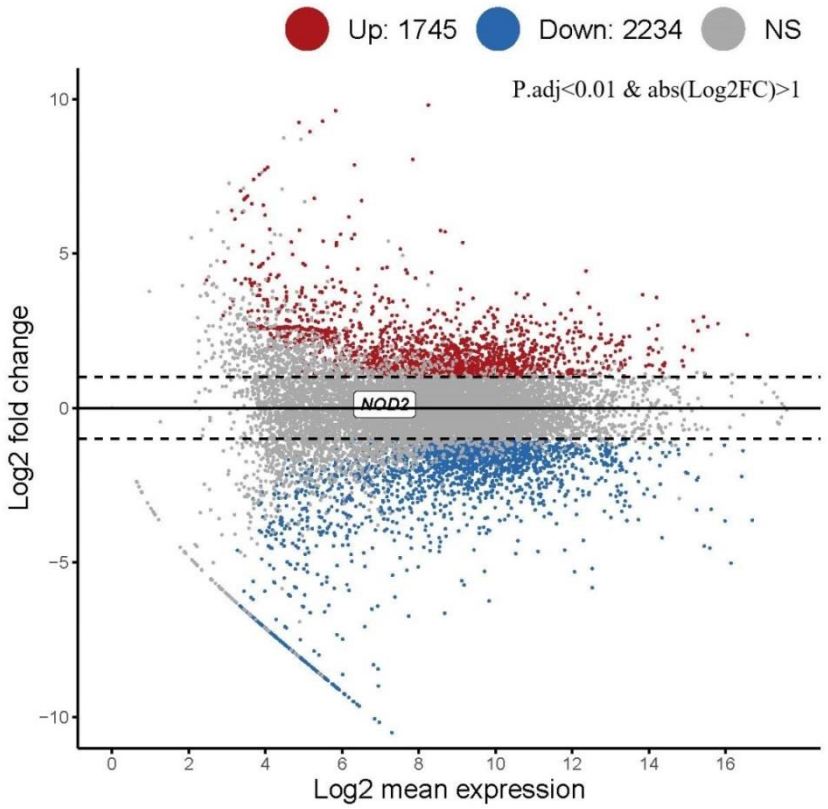


Figure 3.4.4.2 DE genes from the comparison between MDP treated and untreated wild type macrophages

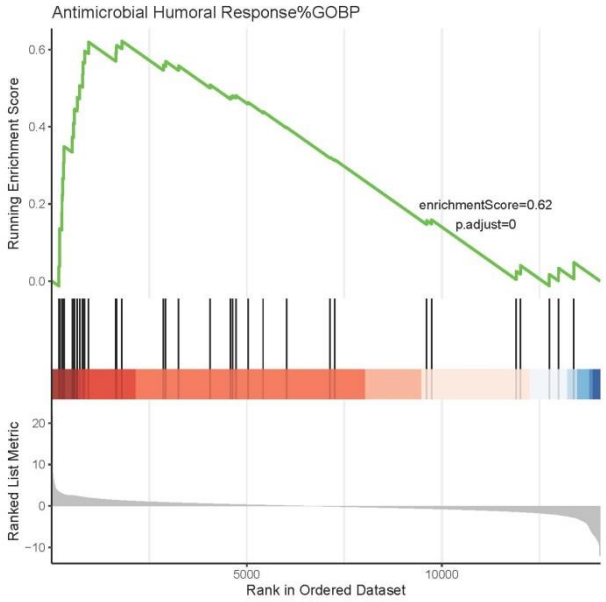


Figure 3.4.4.3 Up regulation of "Antimicrobial Humoral Response" (GO: 0019730) when treating wild type macrophages with MDP

Next, DGE analysis were done between the MDP treated *NOD2* deficient and wild type macrophages, and the DE genes were identified based on the criteria which was previously used to define significantly different ("p.adj < 0.01" and "fold change>2") (Figure 3.4.4.4). Surprisingly, only a few DE genes were identified (13 up regulated and 55 down regulated). Moreover, GSEA on the DGE results was performed, and none of the significantly enriched terms were found. This suggested *NOD2* deficiency made little influence on MDP-mediated inflammatory response in the hiPSC-derived macrophages.

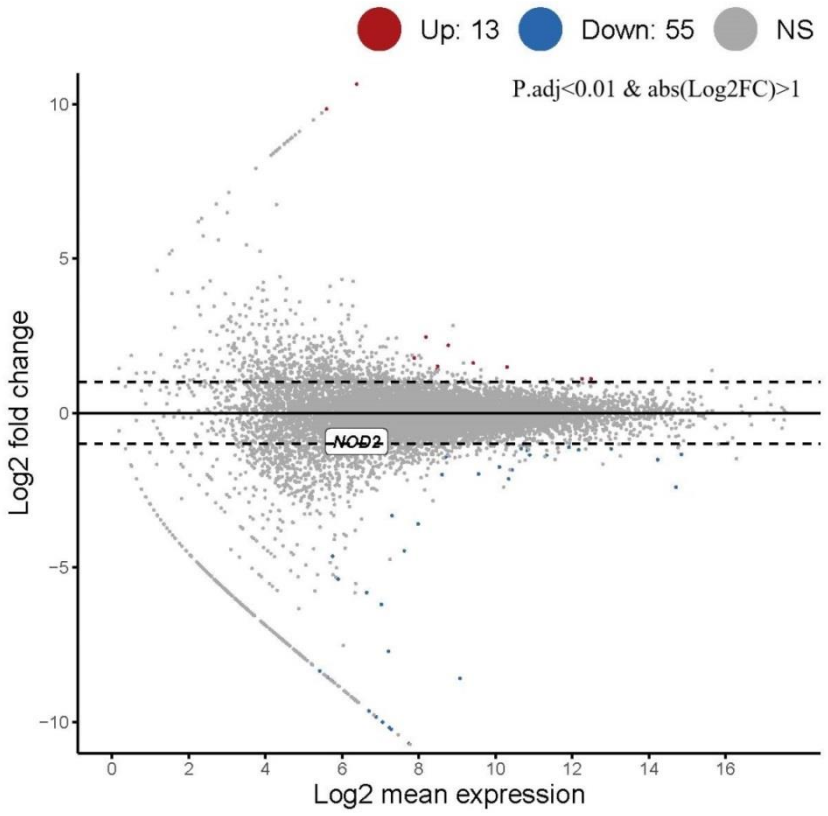


Figure 3.4.4.4 DE genes from the comparison between MDP treated *NOD2* deficient and wild type macrophages

3.4.5 *NOD2* deficiency down regulates macrophages response to IFN-γ

Due to the surprising results in section 3.4.4, it became interesting to explore the IFN-γ and LPS mediated inflammatory response further. These response of the macrophages to these was identified to have a possibly pre-primed down regulation, even in the untreated *NOD2* deficient macrophages. By doing so, it can verify whether those inflammatory responses in macrophages were truly affected by *NOD2* deficiency.

The experiments started from the IFN-γ mediated inflammatory response. In those experiments, *NOD2* deficient macrophages were collected, and treated with 100 ng/ml recombinant human IFN-γ (Peprotech, #300-02-100). Meanwhile, to exclude the effects from treating time, the experiments included a short time course (6 hours, 12 hours and 24 hours). Wild type macrophages with the same treatment were taken as the reference. As in untreated macrophages, the morphology between wild type and *NOD2* deficient macrophages with IFN-γ stimulations were compared, and no obvious differences were found (Figure 3.4.5.1, Figure 3.4.5.2 and Figure 3.4.5.3).

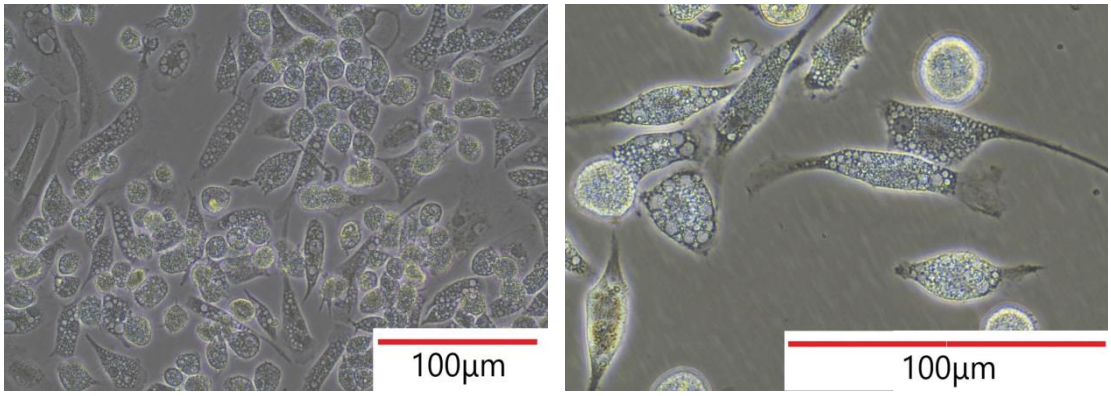


Figure 3.4.5.1 Comparison between the morphology of the wild type and *NOD2* deficient macrophages with 100ng/ml IFN-γ treatment for 6 hours

(left: wild type macrophages; right: *NOD2* deficient macrophages)

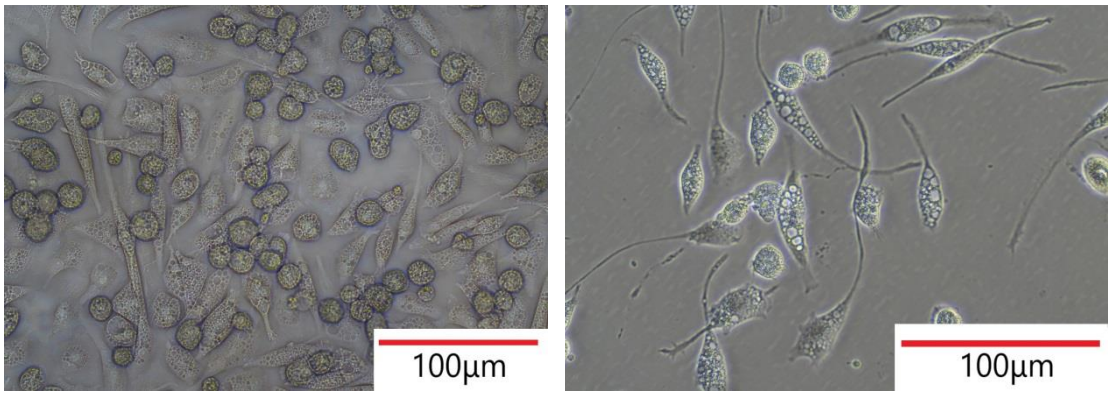


Figure 3.4.5.2 Comparison between the morphology of the wild type and *NOD2* deficient macrophages with 100ng/ml IFN-γ treatment for 12 hours

(left: wild type macrophages; right: *NOD2* deficient macrophages)

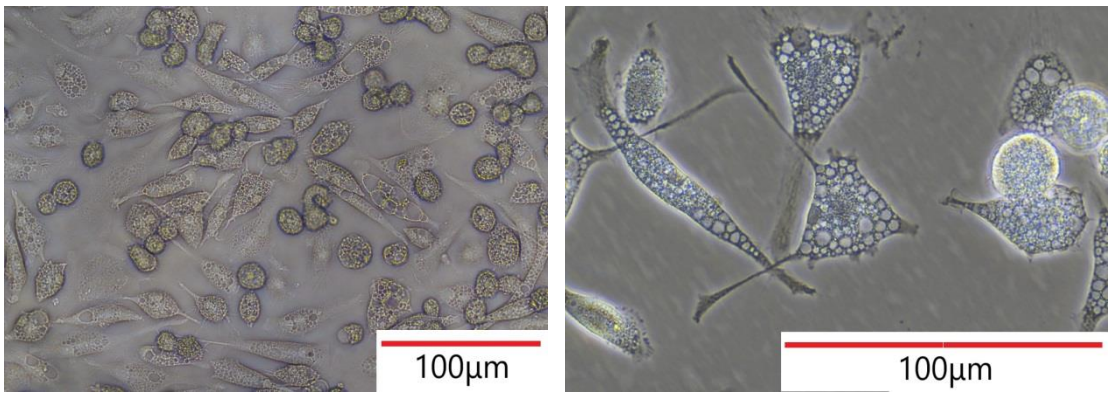


Figure 3.4.5.3 Comparison between the morphology of the wild type and *NOD2* deficient macrophages with 100ng/ml IFN-γ treatment for 24 hours

(left: wild type macrophages; right: *NOD2* deficient macrophages)

Macrophages with IFN-γ treatment were collected for bulk mRNA sequencing. For each point in the time course, six (wildtype_ifng_6h_rp1-6, wildtype_ifng_12h_rp1-6, wildtype_ifng_24h_rp1-6) and three (nod2ko_ifng_6h_rp1-3, nod2ko_ifng_12h_rp1-3, nod2ko_ifng_24h_rp1-3) biological replicates from wild type and *NOD2* deficient macrophages were obtained respectively. By calculating the correlation distance between the samples, none of the outliers were found (Figure 3.4.5.4). Meanwhile, the correlation distance between the samples indicated there might be general and obvious differences between *NOD2* deficient and wild type macrophages with IFN-γ stimulation, and the differences had few associations over the time of treatment. DGE analysis and DE genes identification further confirmed the obvious changes caused by *NOD2* deficiency in IFN-γ treated macrophages (Figure 3.4.5.5), and verified length of treating time made few differences on IFN-γ mediated inflammatory response (Figure 3.4.5.6 and Figure 3.4.5.7).

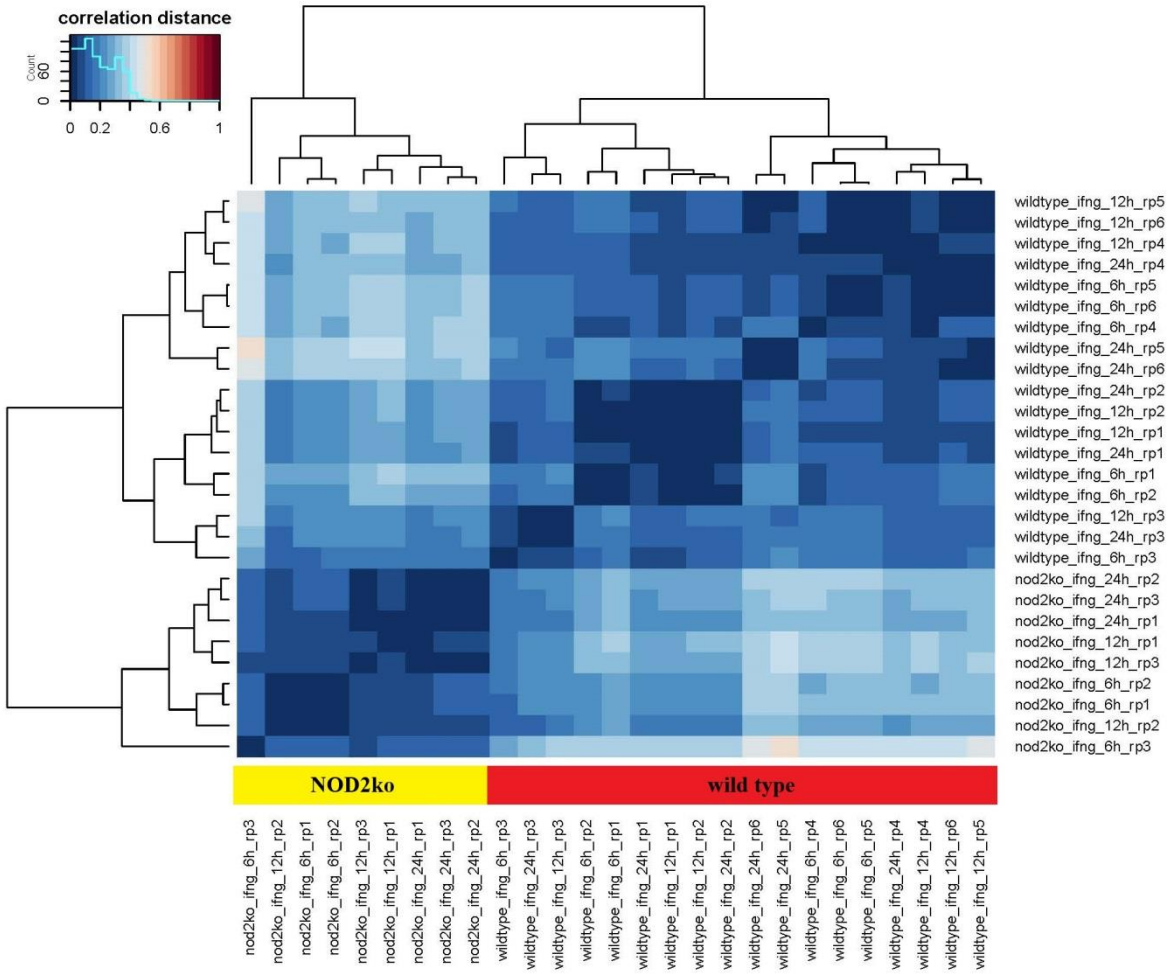


Figure 3.4.5.4 Correlation distances between the wild type and *NOD2* deficient macrophages with IFN-γ stimulation

(Samples' correlation distance was displayed by heat map. Light blue and dark blue "rectangles" could be seen in the plot, which meant IFN-γ treated *NOD2* deficient and wild type macrophages separated obviously in gene expression levels)

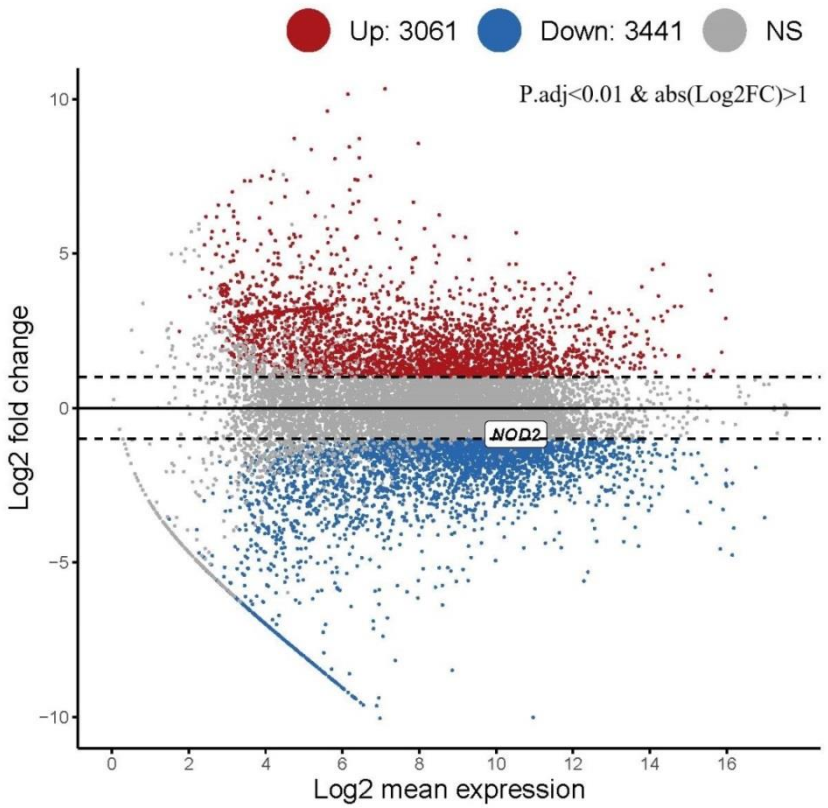


Figure 3.4.5.5 DE genes in *NOD2* deficient macrophages with IFN-γ treatment

(In the MA plot, DE genes were selected by the criteria "p.adj < 0.01" and "fold change>2". 3061 up regulated and 3441 down regulated genes were identified)

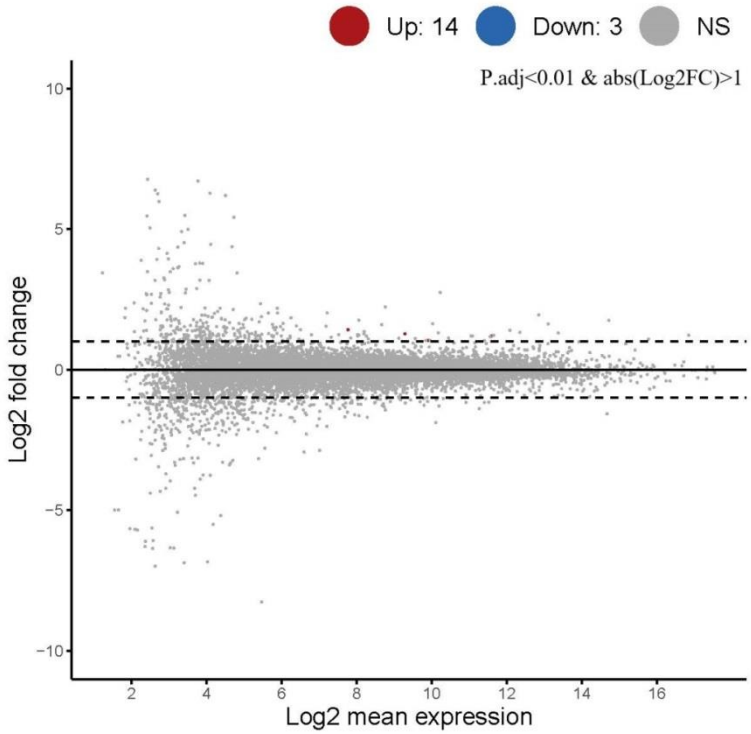


Figure 3.4.5.6 DE genes from the comparison between wild type macrophages with 12 hours and 6 hours IFN-γ treatment

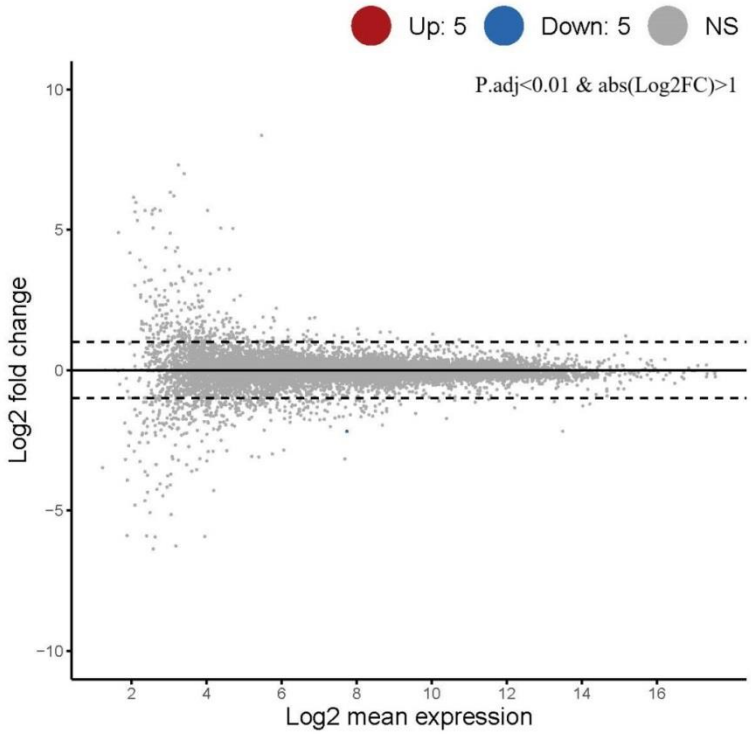


Figure 3.4.5.7 DE genes from the comparison between wild type macrophages with 24 hours and 12 hours IFN-γ treatment

GSEA was performed based on the comparison between *NOD2* deficient and wild type macrophages with IFN-γ treatment, and the results identified the down regulation of IFN-γ pathway in IFN-γ treated *NOD2* deficient macrophages (Figure 3.4.5.8). Meanwhile, from GSEA specifically based on the KEGG pathway database, JAK-STAT signaling pathway was identified as being down regulated (Figure 3.4.5.9). The results from GSEA also indicated the down regulation of inflammatory response in IFN-γ treated *NOD2* deficient macrophages (Figure 3.4.5.10). These data showed *NOD2* deficiency could truly down regulated IFN-γ mediated inflammatory response in macrophages, which was in concert with the findings from untreated cells in section 3.4.3.

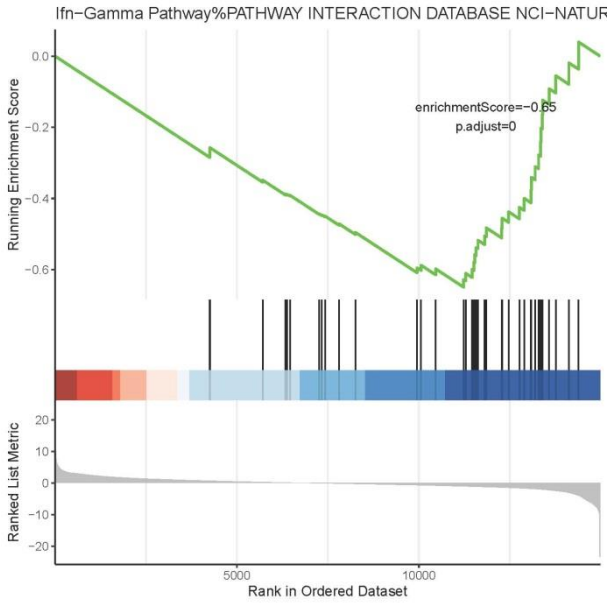


Figure 3.4.5.8 Down regulation of "Ifn-Gamma Pathway" in IFN-γ treated *NOD2* deficient macrophages

(Gene set of "Ifn-Gamma Pathway" was from PATHWAY INTERACTION DATABASE NCI-NATURE CURATED DATA)

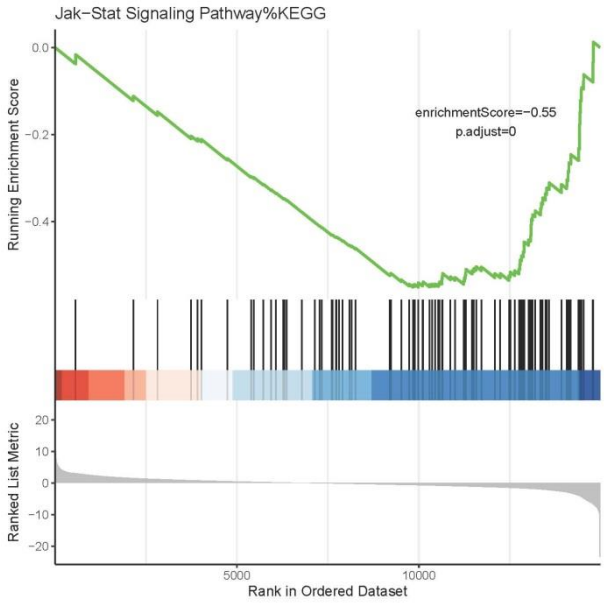


Figure 3.4.5.9 Down regulation of "Jak-Stat Signaling Pathway" (KEGG: hsa04630) in IFN-γ treated *NOD2* deficient macrophages

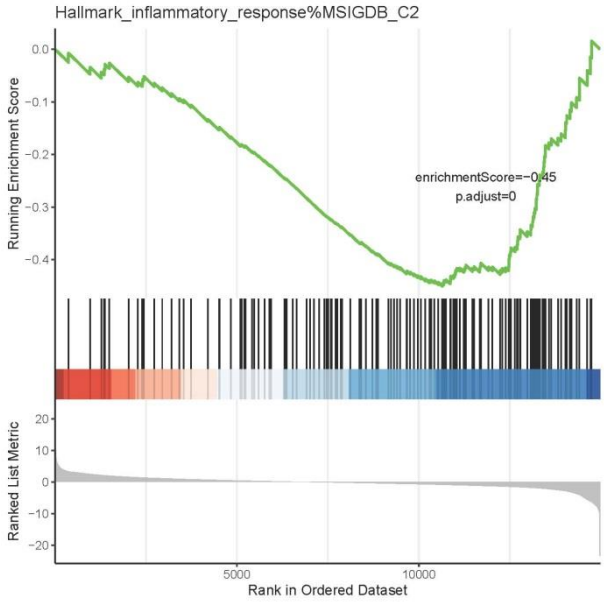


Figure 3.4.5.10 Down regulation of "Hallmark_inflammatory_response" in IFN-γ treated *NOD2* deficient macrophages

(Gene set "Hallmark_inflammatory_response" was from "C2" collection of Molecular Signatures Database (MSIGDB))

Some key genes involved in IFN-γ pathway (*JAK1*, *JAK2*, *STAT1*, *STAT3*, *IRF1*, *IRF9*) were chosen, and their expression levels in macrophages from untreated to different treating time were viewed (Figure 3.3.5.11). The results indicated those genes were indeed all down regulated in IFN-γ treated *NOD2* deficient macrophages compared to wild-type macrophages.

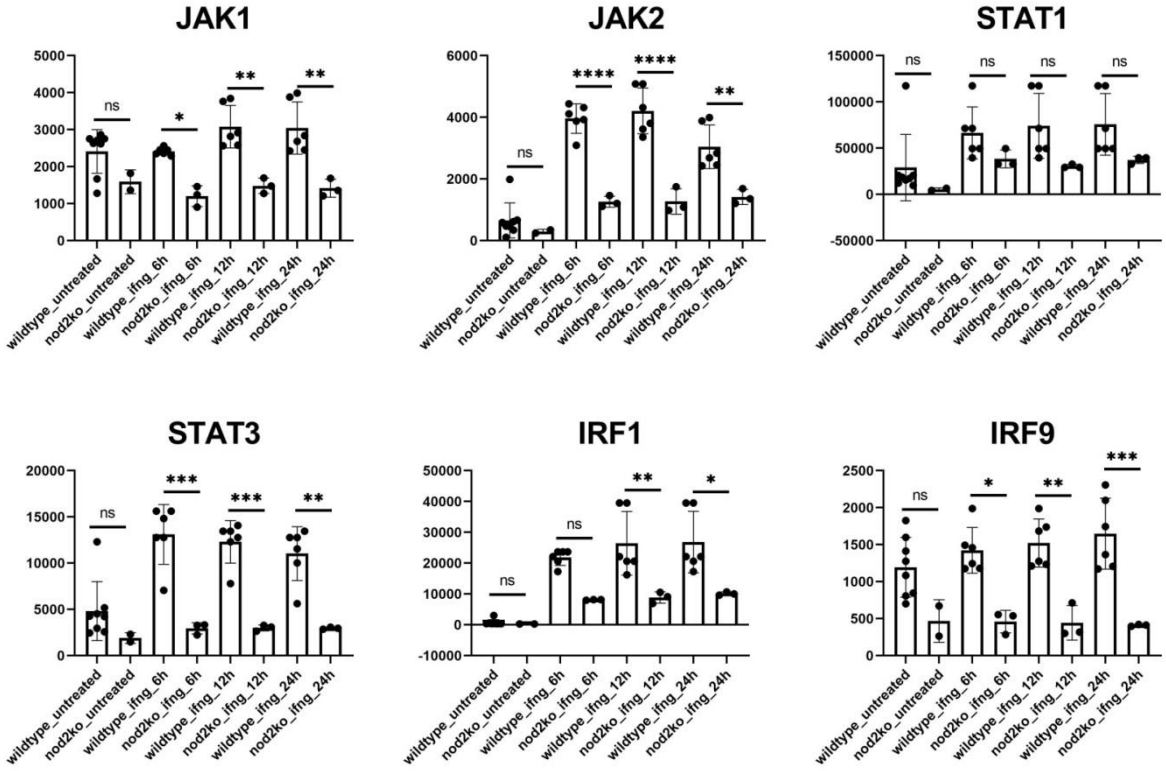


Figure 3.4.5.11 Expression levels of some key genes involved in IFN-γ pathway in wild type and *NOD2* deficient macrophages

(In each bar plot, x-axis was the sample condition, and the value of y-axis was the gene's averaged counts (post-normalization) among the replicates for the condition. Error bar (standard deviation) was also displayed in the plots.

Below are the meanings of sample conditions in the plots:

wildtype_untreated: wild type macrophages without any treatments;

wildtype_ifng_6/12/24h: wild type macrophages with 6/12/24 hours 100 ng/ml IFN-γ treatment;

nod2ko_untreated: *NOD2* deficient macrophages without any treatments;

nod2ko_ifng_6/12/24h: *NOD2* deficient macrophages with 6/12/24 hours 100 ng/ml IFN-γ treatment)

Furthermore, the expression of genes which encoded IFN-γ receptor (*IFNGR1* and *IFNGR2*) was also compared in wild type and *NOD2* deficient macrophages (Figure 3.4.5.12). For the untreated cells, expression levels of both genes were the same. For the IFN-γ treated cells, expression levels of *IFNGR1* were obviously down regulated comparing with the untreated cells, and the trend was more significant for *NOD2* deficient macrophages. For *IFNGR2* expression, no obvious differences were detected.

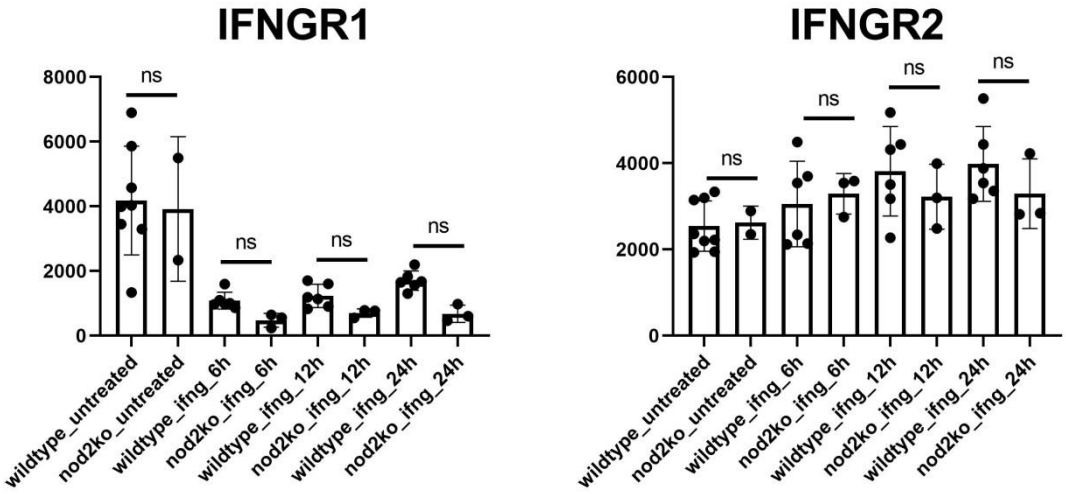


Figure 3.4.5.12 Expression levels of IFN-γ receptor genes

As the supplement and verification of RNA-seq analysis, expression levels of *STAT1* protein were identified via western blotting in wild type and *NOD2* deficient macrophages treated by 100ng/ml IFN-γ for 24 hours. β-Actin protein was used as internal reference, and could indicate the differences of loading protein amount between the samples. It was obvious that expression levels of *STAT1* protein was lower in *NOD2* deficient macrophages compared to the wild type ones, though the total amount of loading protein between the samples had some differences (based on the bands of β-Actin protein) (Figure 3.4.5.13).

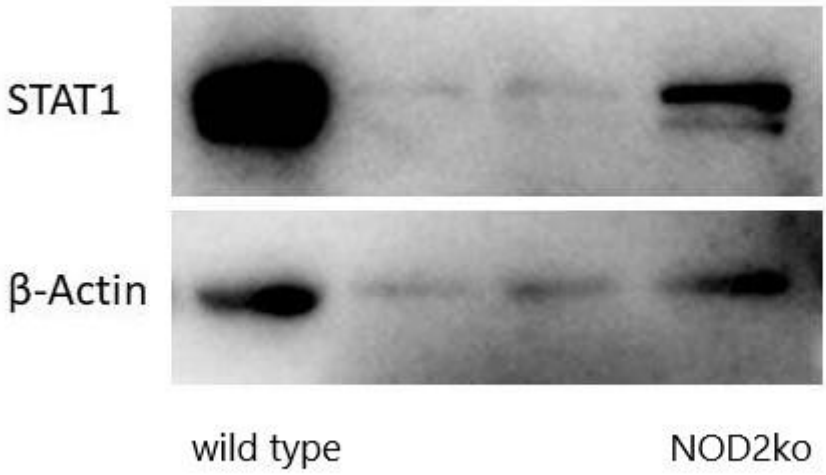


Figure 3.4.5.13 Expression levels of *STAT1* protein in wild type and *NOD2* deficient macrophages with 24 hours 100ng/ml IFN-γ treatment

(wild type: wild type macrophages with 24 hours 100 ng/ml IFN-γ treatment;

NOD2ko: *NOD2* deficient macrophages with 24 hours 100 ng/ml IFN-γ treatment.

The middle two lanes were samples from other project)

In addition, DGE analysis was done between *NOD2* deficient macrophages with and without IFN-γ treatment, and used the results to perform GSEA. Cellular response to IFN-γ was significantly enriched and up regulated when *NOD2* deficient macrophages were treated by IFN-γ (Figure 3.4.5.14), as well as the up regulation of IFN-γ signaling (Figure 3.4.5.15). These results illustrated the effect of the loss of *NOD2* on macrophages. The IFNg response was impaired, and down regulated the cells' response to IFN-γ, although the full pathway was not ablated.

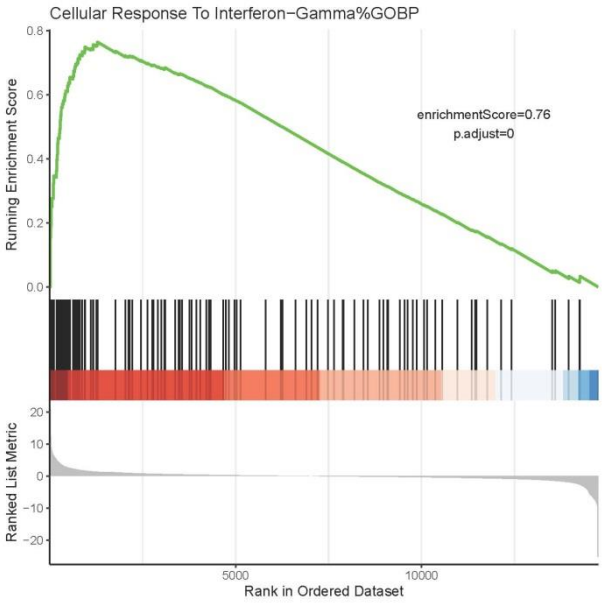


Figure 3.4.5.14 Up regulation of "Cellular Response To Interferon-Gamma" (GO: 0071346) when *NOD2* deficient macrophages were treated by IFN-γ

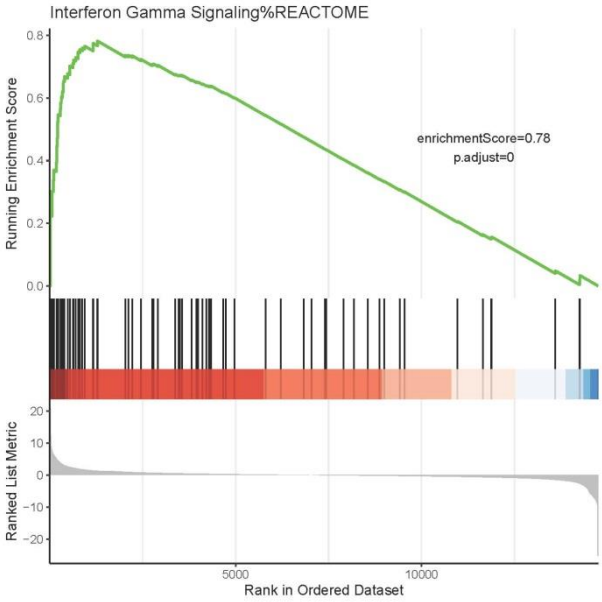


Figure 3.4.5.15 Up regulation of "Interferon Gamma Signaling" (REACTOME: R-HSA-877300.3) when *NOD2* deficient macrophages were treated by IFN-γ

3.4.6 *NOD2* deficiency has few effects on macrophages response to LPS

In this section, LPS mediated inflammatory response in macrophages was the focus of research, to verify whether it was truly affected by *NOD2* deficiency in macrophages. *NOD2* deficient macrophages were treated with 1 ug/ml LPS (Sigma-Aldrich, #L6529). Additionally, as described in section 1.5, macrophages pro-inflammatory response to LPS has sub stages, to address this a time course was used (6 hours, 12 hours and 24 hours) in the experimental design. The same treatment was made on wild type macrophages as reference. As previous, morphology was compared between the LPS treated wild type and *NOD2* deficient macrophages, and no differences were identified (Figure 3.4.6.1, Figure 3.4.6.2 and Figure 3.4.6.3)

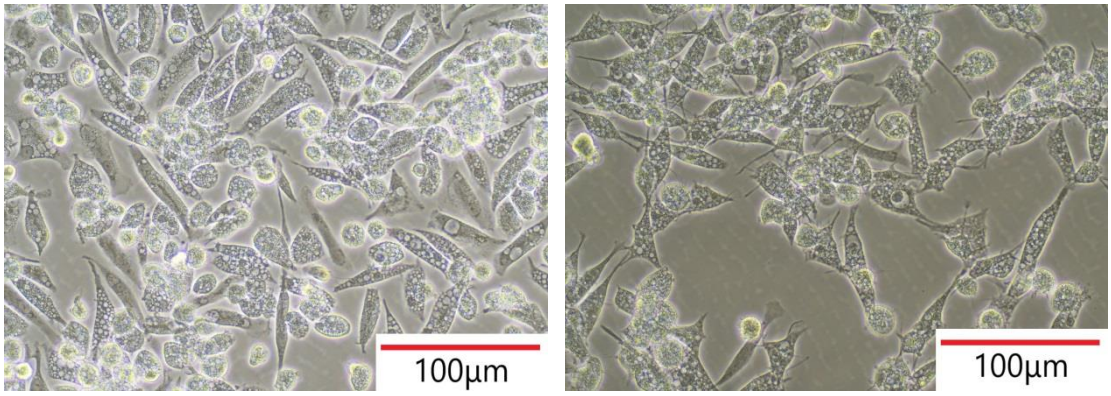


Figure 3.4.6.1 Comparison between the morphology of the wild type and *NOD2* deficient macrophages with 1ug/ml LPS treatment for 6 hours

(left: wild type macrophages; right: *NOD2* deficient macrophages)

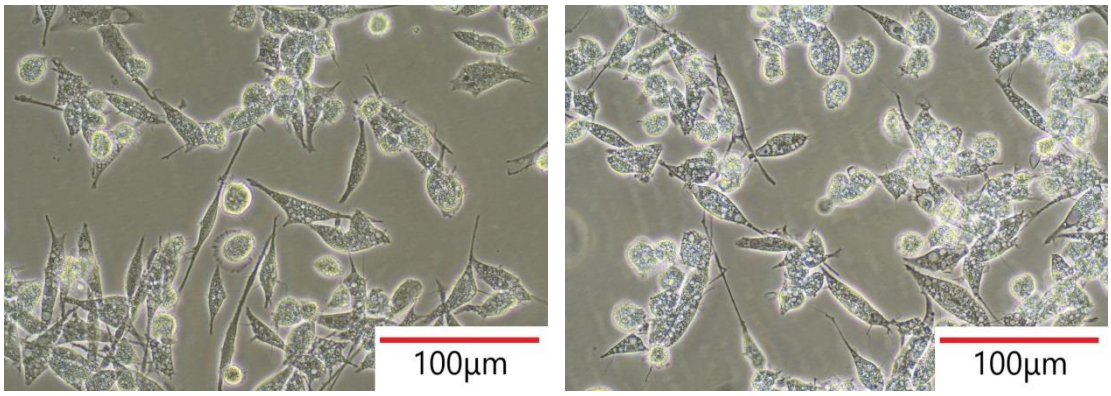


Figure 3.4.6.2 Comparison between the morphology of the wild type and *NOD2* deficient macrophages with 1ug/ml LPS treatment for 12 hours

(left: wild type macrophages; right: *NOD2* deficient macrophages)

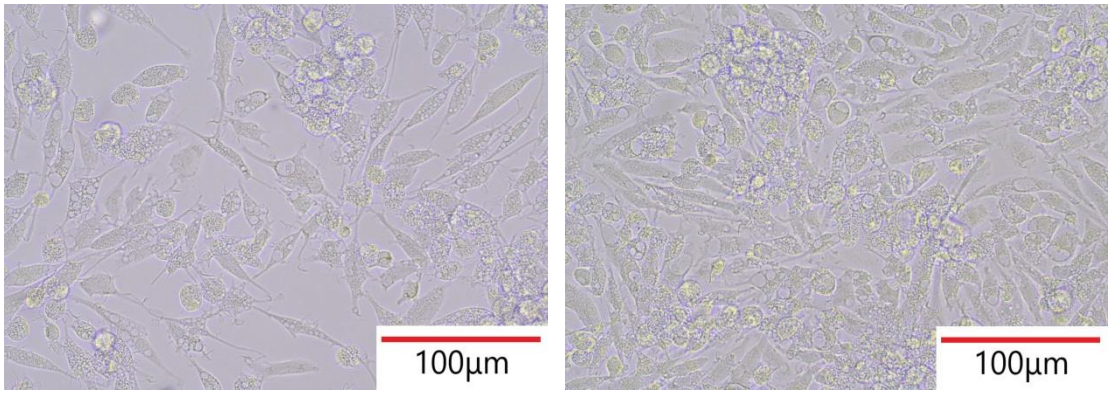


Figure 3.4.6.3 Comparison between the morphology of the wild type and *NOD2* deficient macrophages with 1ug/ml LPS treatment for 24 hours

(left: wild type macrophages; right: *NOD2* deficient macrophages)

Next, macrophages with LPS stimulation were collected for bulk mRNA sequencing. For each point in the time course, three biological replicates from wild type (wildtype_lps_6h_rp1-3, wildtype_lps_12h_rp1-3, wildtype_lps_24h_rp1-3) and *NOD2* deficient (nod2ko_lps_6h_rp1-3, nod2ko_lps_12h_rp1-3, nod2ko_lps_24h_rp1-3) macrophages were obtained respectively. After calculating the correlation distance between samples, no outliers were found (Figure 3.4.6.4). Meanwhile, from the correlation distance, no obvious differences between LPS treated wild type and *NOD2* deficient macrophages were found. DGE analysis was done and the analysis identified DE genes in *NOD2* deficient macrophages treated by LPS, and the number of DE genes was quite low (Figure 3.4.6.5). This suggests that the LPS response in the *NOD2* knockout and WT macrophages is similar, and *NOD2* loss has little or no effect on LPS response. GSEA was also performed on the data, and only limited terms associated with response to interferons, which existed in the untreated *NOD2* deficient macrophages were enriched but not significantly (p.adjust > 0.01) (Figure 3.3.6.6). Overall, this suggests that *NOD2* deficiency has few effects on LPS mediated inflammatory response in macrophages, which is distinct from the findings from untreated macrophages. This mismatch is very possible to be caused by the limited biological replicates, since after QC of RNA-seq data, only two replicates from untreated *NOD2* deficient macrophages left in section 3.4.3.

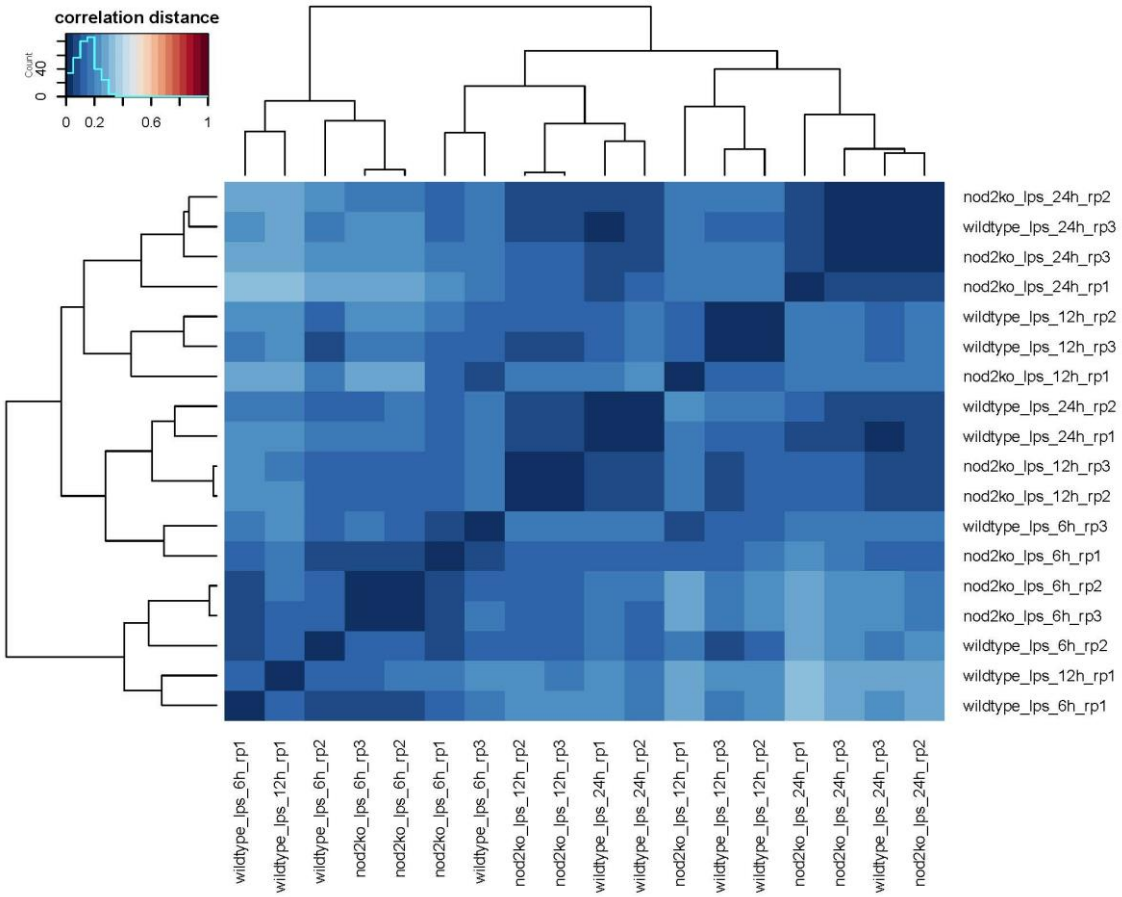


Figure 3.4.6.4 Correlation distances between the wild type and *NOD2* deficient macrophages with LPS stimulation

(Samples' correlation distance was displayed by heatmap. *NOD2* deficient and wild type samples mixed in the plot, which suggested there were no obvious differences between them)

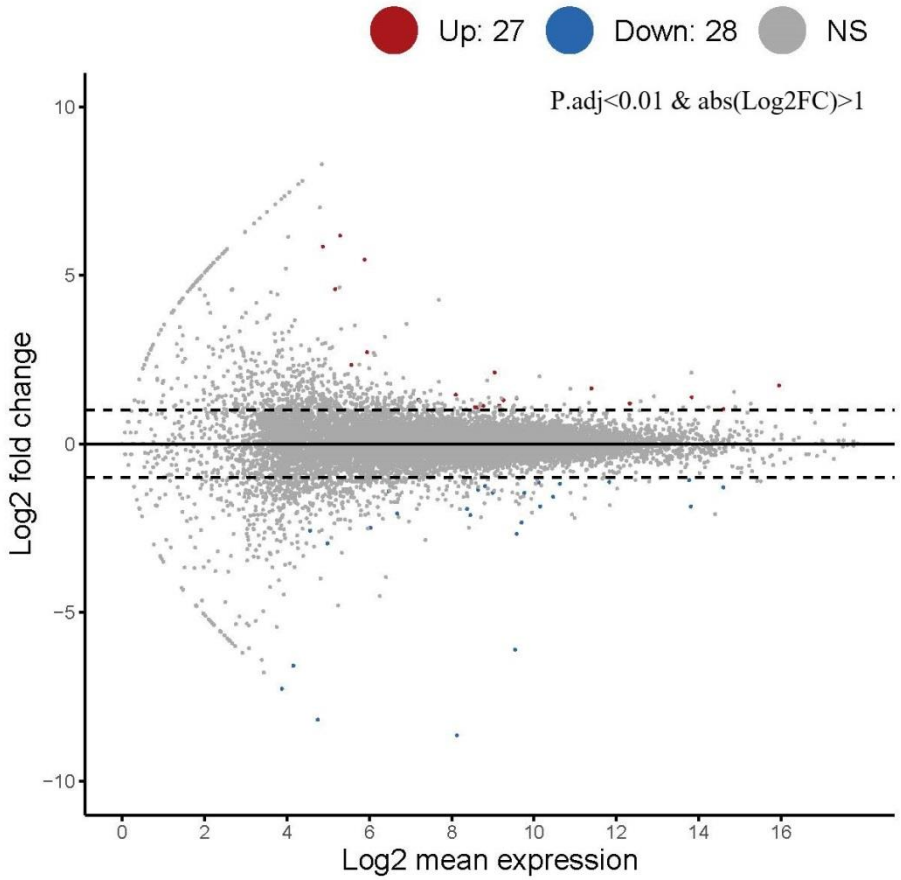


Figure 3.4.6.5 DE genes in *NOD2* deficient macrophages with LPS treatment

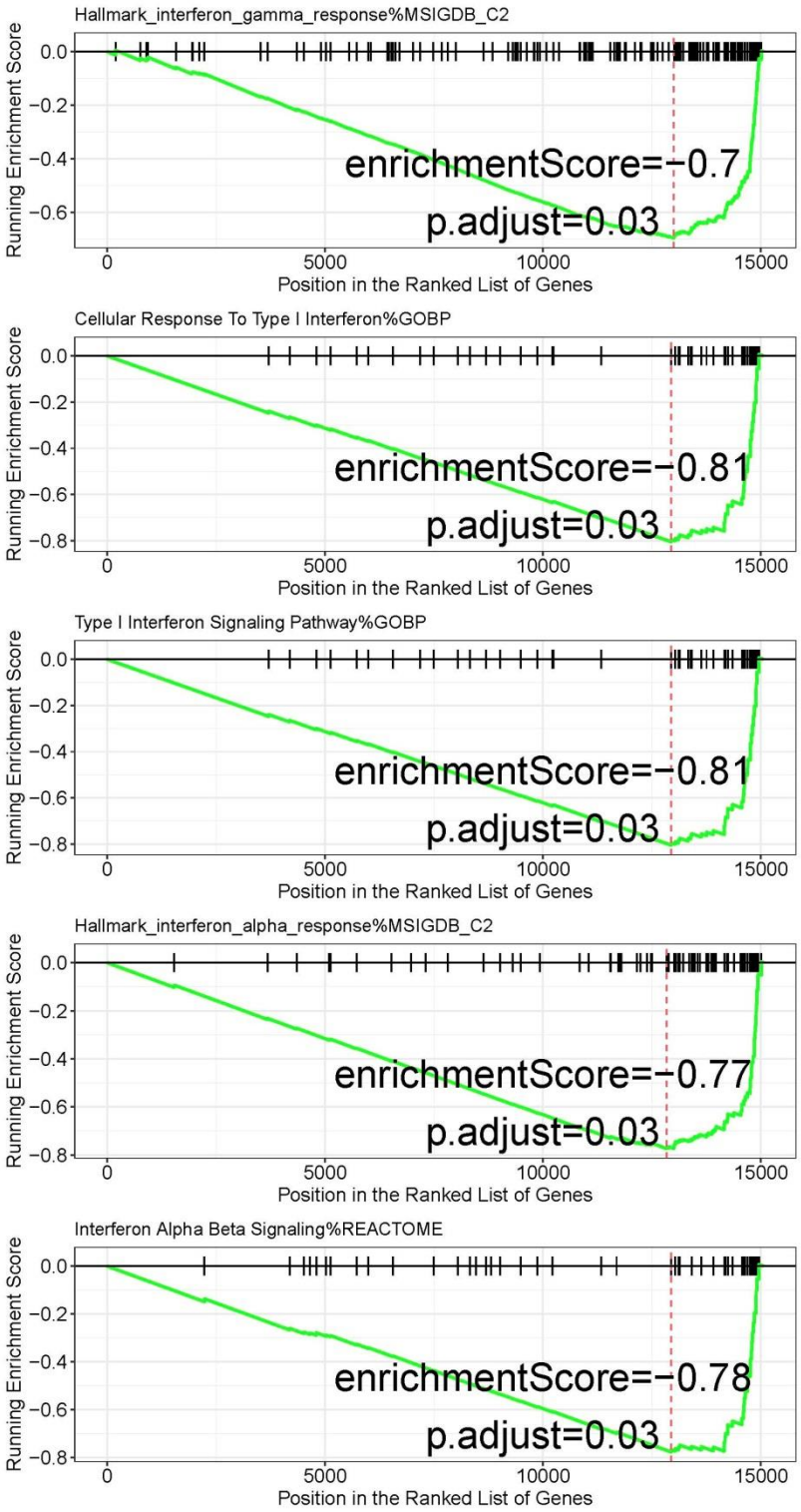


Figure 3.4.6.6 Down regulations of response to interferons in LPS treated *NOD2* deficient macrophages

(overlapped with what had observed in untreated *NOD2* deficient macrophages, but not significant (p.adjust > 0.01))

The expressions of specific genes marks the sub stages of the LPS response (*CCL20*, *CCR7*, *IL6*, *CSF2*, *IL1A*) in the LPS-treated macrophages (Figure 3.4.6.7). These results indicated that all genes had similar expression levels at each time point in both wild type and *NOD2* deficient macrophages. It further verifies *NOD2* deficiency has few effects on LPS mediated inflammatory response in macrophages. Moreover, changes of those genes with time followed the same trend as observed in human monocyte-derived macrophages, which reached the peak at 6 hours, and were down regulated at 24 hours. This shows that the hiPSC-derived macrophages share the same characteristics with the natural human macrophages in LPS mediated inflammatory response.

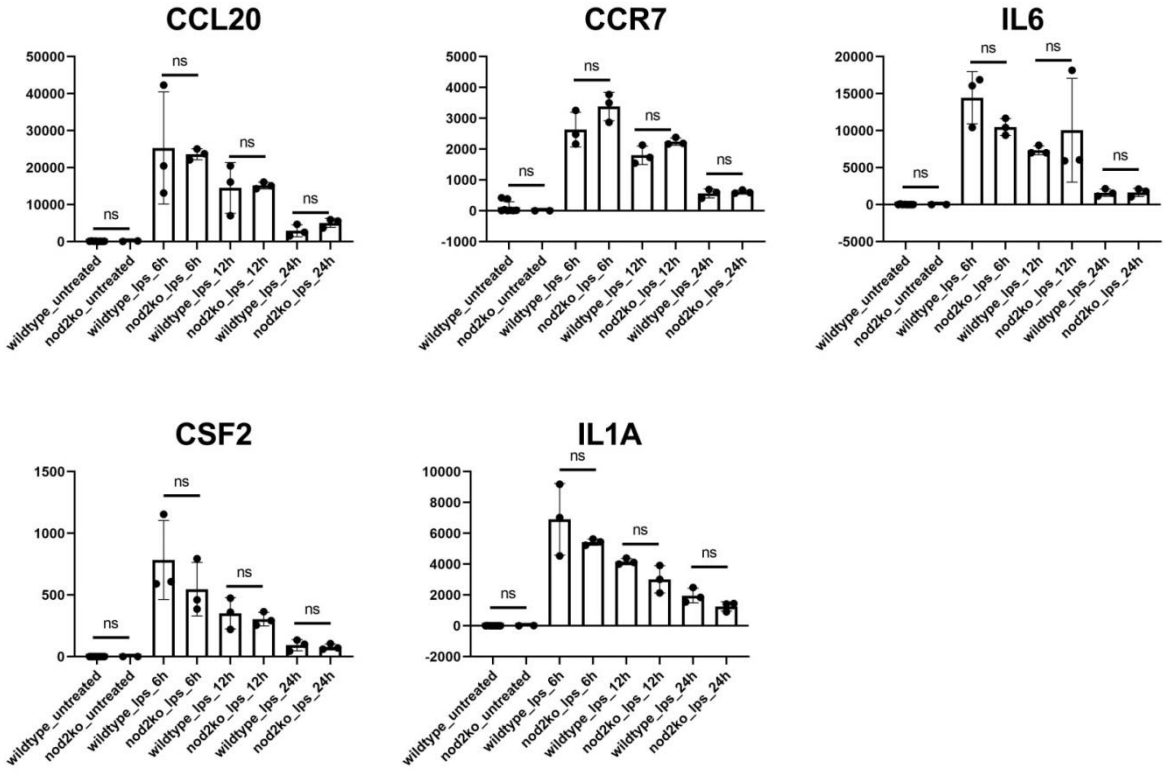


Figure 3.4.6.7 Expressions of the genes which could reflect the transfer of sub stages in LPS treated wild type and *NOD2* deficient macrophages

(wildtype_untreated: wild type macrophages without any treatments

wildtype_lps_6/12/24h: wild type macrophages with 6/12/24 hours 1 μg/ml LPS treatment;

nod2ko_untreated: *NOD2* deficient macrophages without any treatments;

nod2ko_lps_6/12/24h: *NOD2* deficient macrophages with 6/12/24 hours 1 μg/ml LPS treatment)

3.4.7 *In vitro* differentiation of *NOD2* deficient intestinal organoids

Considering that *NOD2* had such a large impact on the macrophages, it was reasonable to wonder whether the same situation was true for other tissues. Hence the intestinal organoids, which could be generated from hiPSCs *in vitro*, were used in the experiments. In the experiments, *NOD2* deficient intestinal organoids were obtained successfully from the hiPSC line generated in section 3.4.1 (Figure 3.4.7.1), as well as the wild type intestinal organoids as the reference (Figure 3.4.7.2).

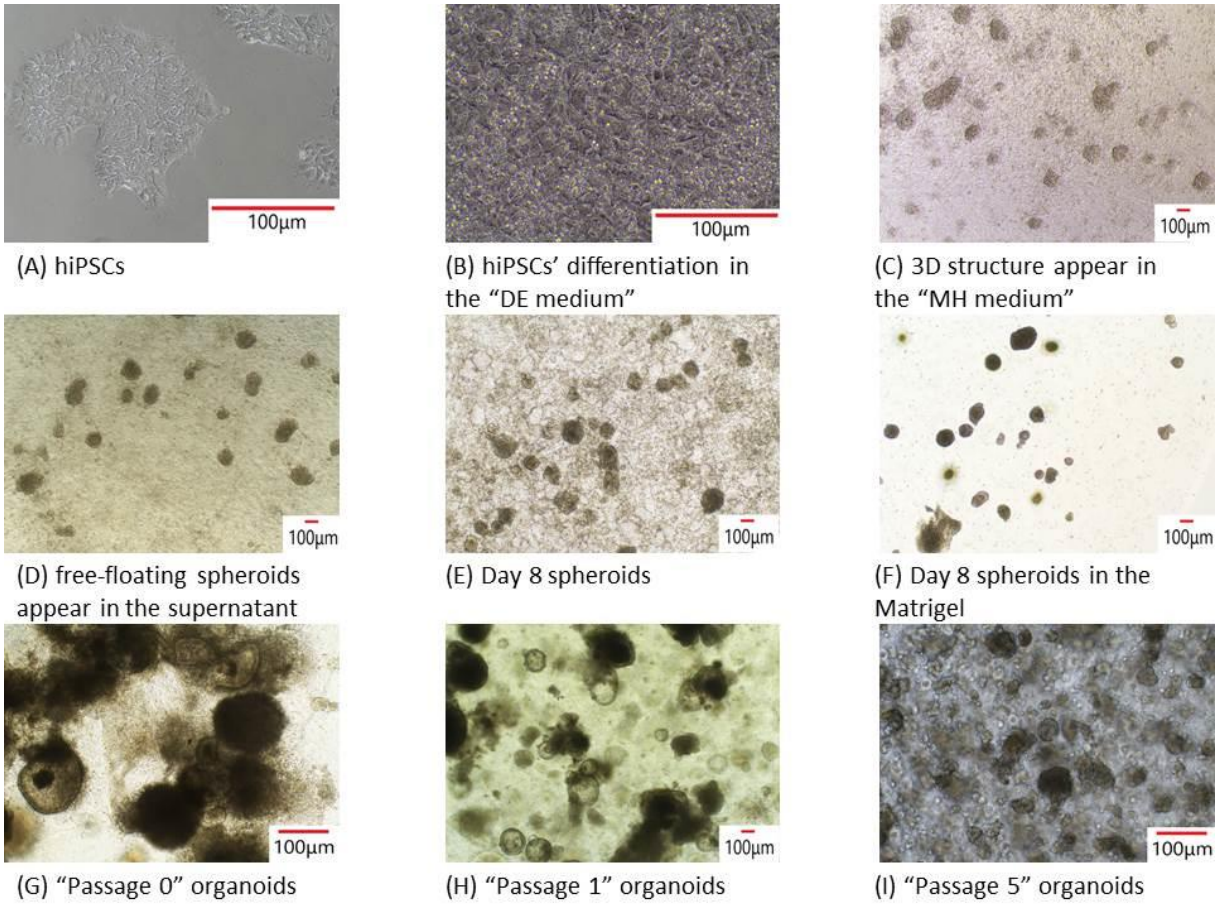


Figure 3.4.7.1 In vitro differentiation of *NOD2* deficient intestinal organoids

(*NOD2* deficient hiPSCs were cultured in mTeSR™1 in 24 well plate (A). Medium was fully changed every day by the "DE medium" for 3 days (Day 0-Day 2) (B). From Day 3, medium for the culture was changed into "MH medium", with everyday fully medium change for another 7 days (Day 3-Day 9). 3D structure appeared in the well from Day4 of the experiments (C). As early as Day 5, free-floating spheroids could be observed in the supernatant, and were available to be collected (D). For these experiments, spheroids from Day 8 were collected (E), and mixed with Growth Factor Reduced Matrigel to form the domes in 24 well plate (F). "OGM medium" was added into the well for culturing the spheroids, with fully medium change every three or four days. After culturing the spheroids for 7-10 days, they became bigger and darker with a more obvious 3D structure, which meant the spheroids are able to be passaged. At this stage the spheroids were defined as "Passage 0" organoids (G). Those passaged spheroids were still maintained in "OGM medium", and could be passaged again after another 7-10 days' culture. Passage of the organoids was regular, and after 2-3 rounds of passage, the stable organoids' line which could be long-time maintained as the cell lines formed. (H) shows "Passage 1" organoids, and (I) shows the organoids at "Passage 5".)

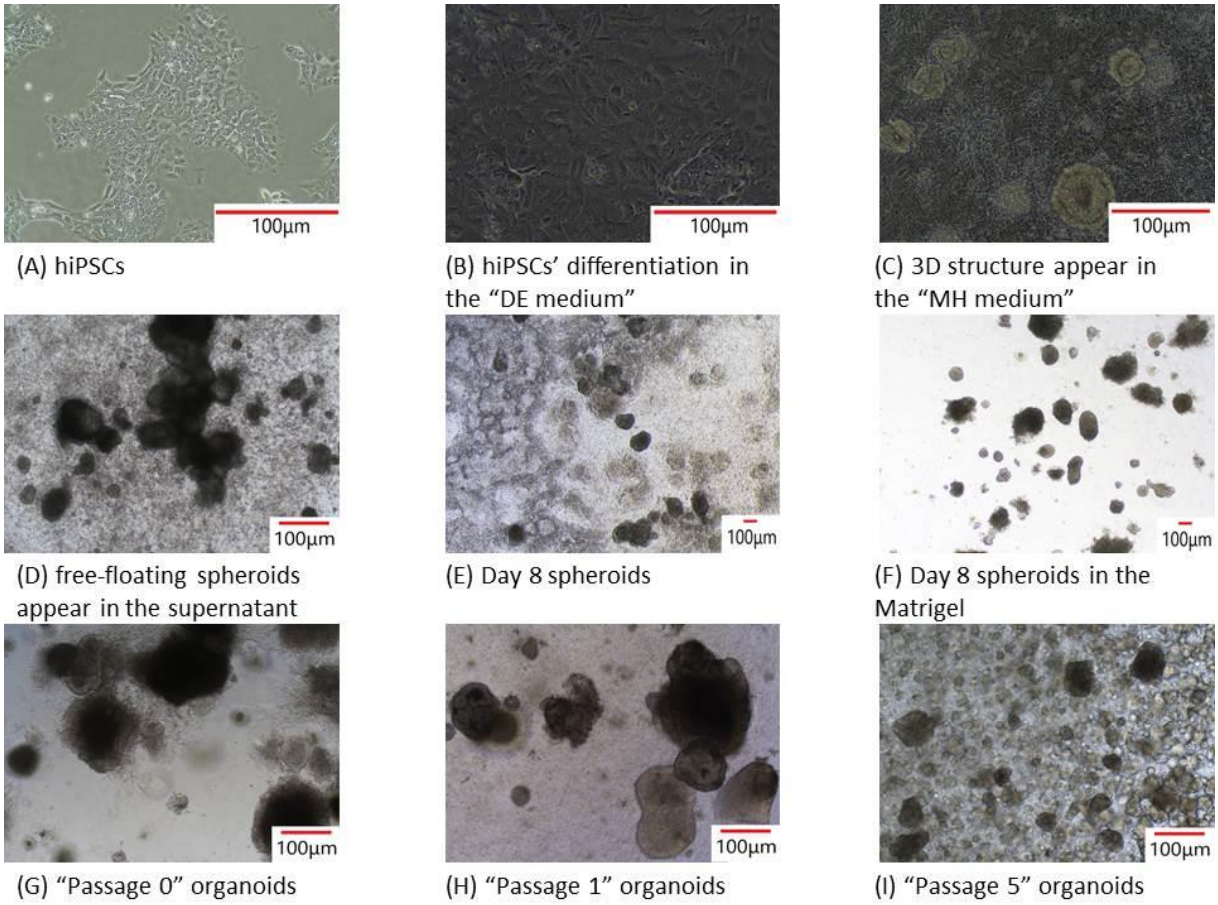


Figure 3.4.7.2 In vitro differentiation of wild type intestinal organoids

(Figure (A) to (I) displayed the process of the in vitro differentiation. The wild type hiPSCs followed the same stages as the NOD2 deficient hiPSCs, and derived into wild type intestinal organoids *in vitro*.

3.4.8 *NOD2* deficiency has few effects on the IFN-γ response in intestinal organoids

To reveal whether *NOD2* deficiency made the same large impact in intestinal organoids as in macrophages, in the experiments, *NOD2* deficient intestinal organoids were treated by IFN-γ, with the same concentration (100 ng/ml) and the same time course (6 hours, 12 hours, 24 hours) as the macrophages. Samples were collected for RNA extraction and bulk mRNA sequencing, and two biological replicates from each time point were obtained (nod2ko_ifng_6h_rp1-2; nod2ko_ifng_12h_rp1-2; nod2ko_ifng_24h_rp1-2). Corresponding samples were also obtained from wild type intestinal organoids (wildtype_ifng_6h_rp1-2; wildtype_ifng_12h_rp1-2; wildtype_ifng_24h_rp1-2). Meanwhile, four biological replicates were collected respectively from untreated *NOD2* deficient and wild type intestinal organoids (nod2ko_untreated_rp1-4; wildtype_untreated_rp1-4).

Calculation of the correlation distance between the samples found no likely outliers (Figure 3.4.8.1). Meanwhile, from the dendrogram, it was visible that macrophages with and without IFN-γ treatment clustered respectively, and in each cluster samples mixed with each other. This suggested that hiPSC-derived intestinal organoids had response to IFN-γ stimuli, however, *NOD2* deficiency might not have the huge impact in those organoids as it was in macrophages.

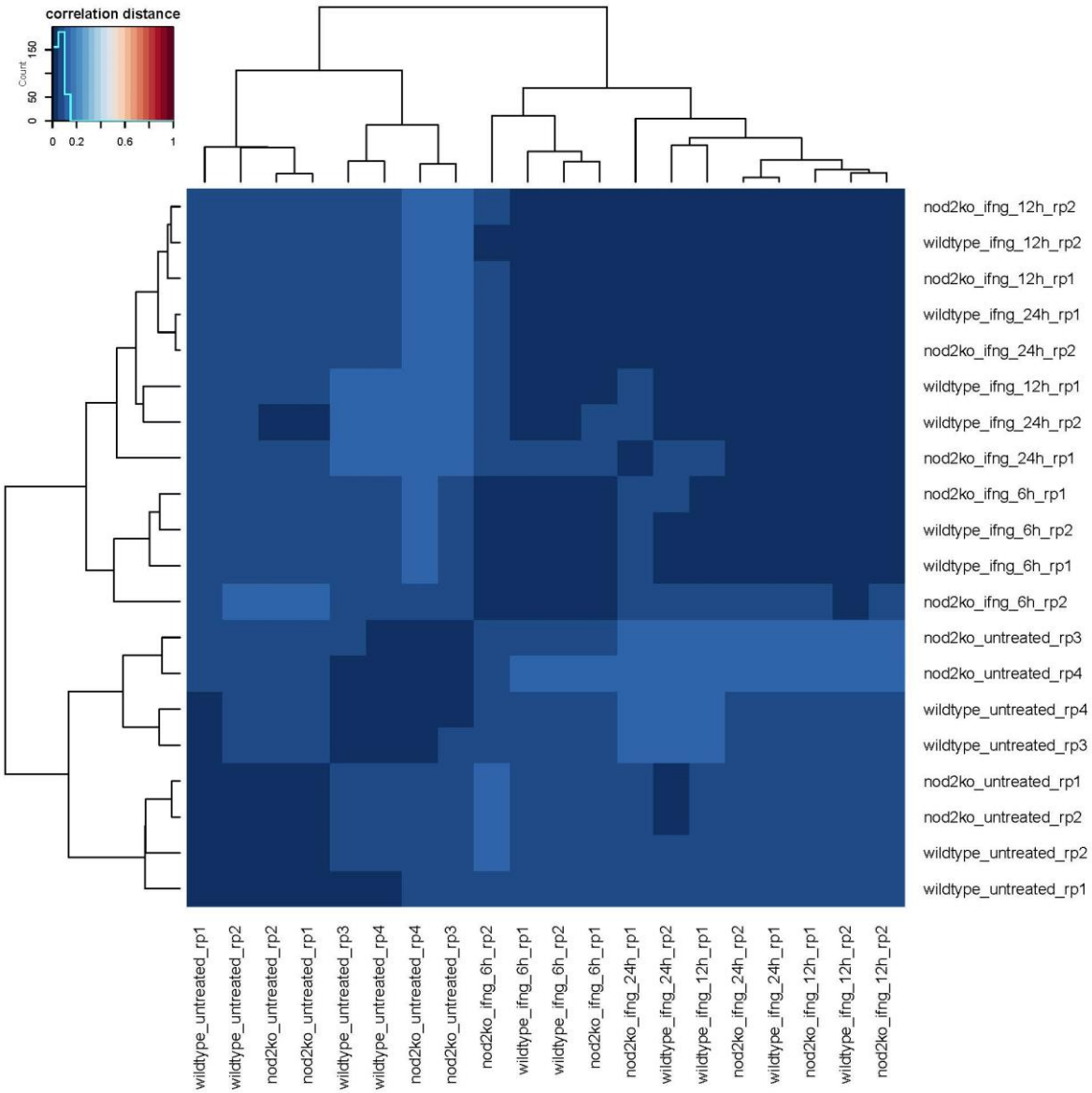


Figure 3.4.8.1 Correlation distances of the hiPSC-derived intestinal organoids with and without IFN-γ treatment

hiPSC-derived intestinal organoids were verified to have a significant response to IFN-γ stimulation via GSEA between the wild type intestinal organoids with and without IFN-γ. In the results, cellular response to IFN-γ and the associated signaling were significantly up regulated when the organoids (wild type) were treated with IFN-γ (Figure 3.4.8.2 and 3.4.8.3).

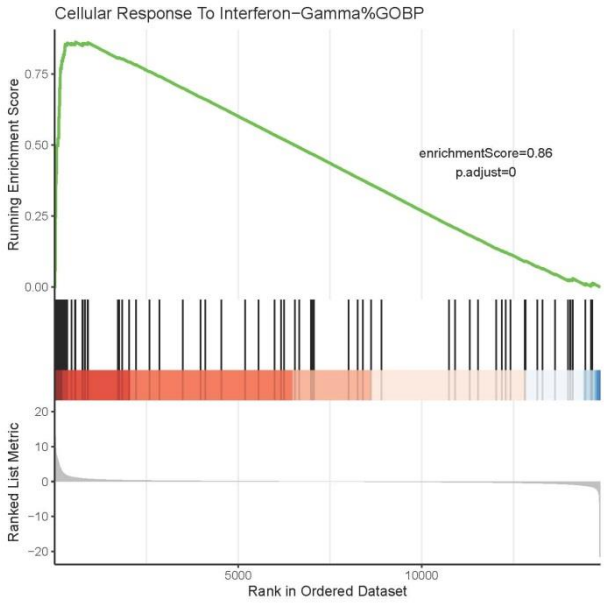


Figure 3.4.8.2 Up regulation of "Cellular Response To Interferon-Gamma" (GO: 0071346) when the organoids (wild type) were treated by IFN-γ

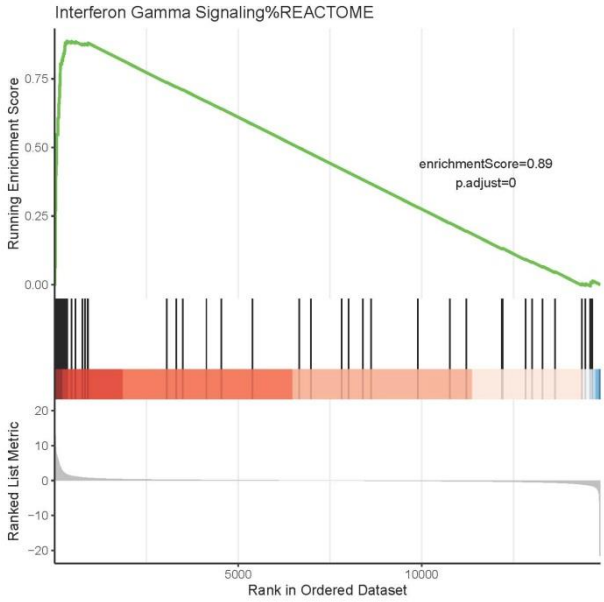


Figure 3.4.8.3 Up regulation of "Interferon Gamma Signaling" (REACTOME: R-HSA-877300.3) when the organoids (wild type) were treated by IFN-γ

Next, DGE analysis was done between the *NOD2* deficient and wild type intestinal organoids with IFN-γ treatment, and only identified small amount of DE genes (Figure 3.4.8.4). GSEA was performed on the DGE results, and none of the terms were significantly enriched (p.adjust< 0.05). This clearly suggested that *NOD2* deficiency had no huge impact in those organoids as it was in macrophages

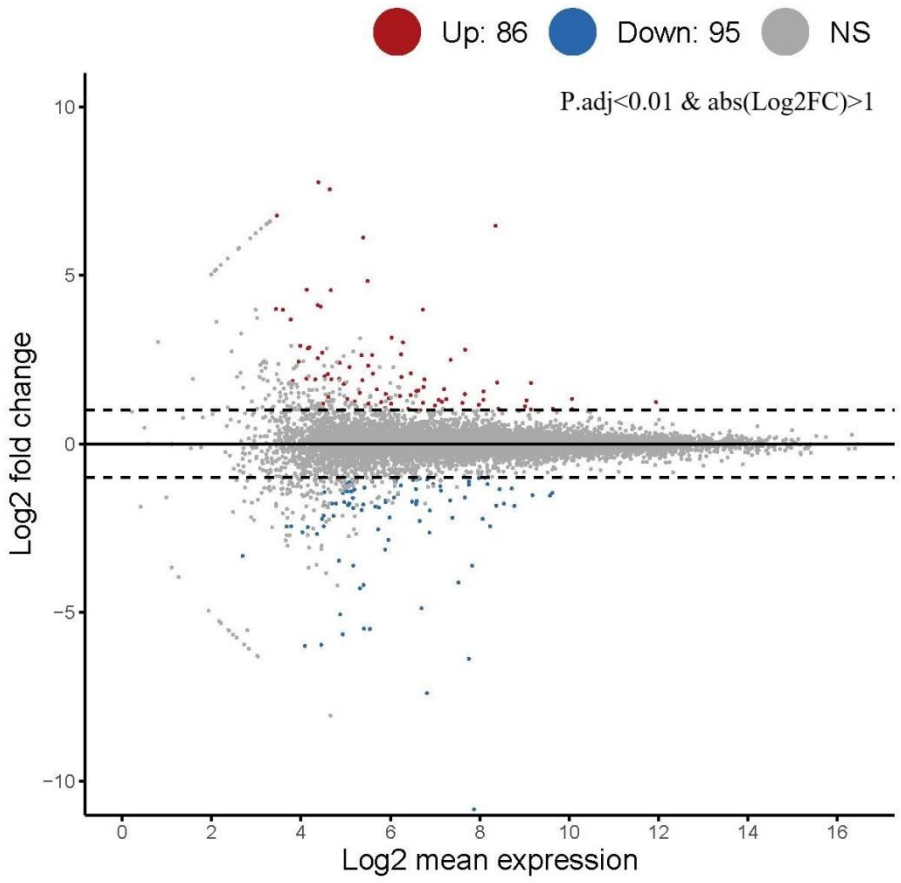


Figure 3.4.8.4 DGE analysis between *NOD2* deficient and wild type intestinal organoids with IFN-γ treatment

(86 up regulated and 95 down regulated genes were identified, and the results were displayed by MA plot)

NOD2 deficient and wild type intestinal organoids without any treatments were also compared between each other, and only small amount of DE genes were identified (Figure 3.4.8.5). GSEA did not identify any significantly enriched terms (p.adjust< 0.05) in those untreated intestinal organoids. This suggested that primed effects of *NOD2* deficiency on IFN-γ mediated inflammatory response in macrophages did not exist in intestinal organoids, and further supported the idea that *NOD2* had no effects on IFN-γ stimulation of intestinal organoids.

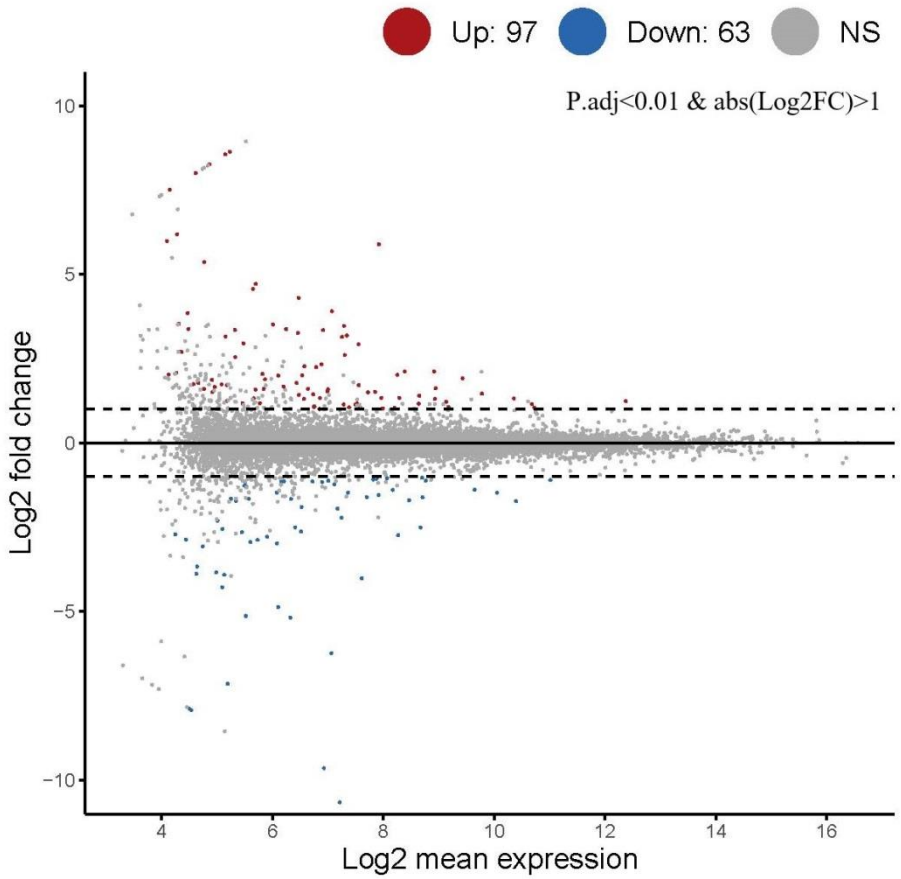


Figure 3.4.8.5 DGE analysis between untreated *NOD2* deficient and wild type intestinal organoids

(97 up regulated and 63 down regulated genes were identified, and the results were displayed by MA plot)

In addition, the expressions of a series of genes known to respond to IFN-γ treatment (*JAK1, JAK2, STAT1, STAT3, IRF1, IRF9*) (Figure 3.4.8.6) and IFN-γ receptor encoding genes (*IFNGR1* and *IFNGR2*) were viewed specifically (Figure 3.4.8.7). These genes had almost same expression levels between wild type and *NOD2* deficient macrophages, and did the same alterations after IFN-γ treatment. The results further demonstrated that loss of *NOD2* did not alter IFN-γ responses in organoids, which is in direct contrast to its effect on macrophages. Meanwhile, those gene expressions also indicated length of treating time made few differences on IFN-γ mediated inflammatory response in those organoids, which was same as the macrophages.

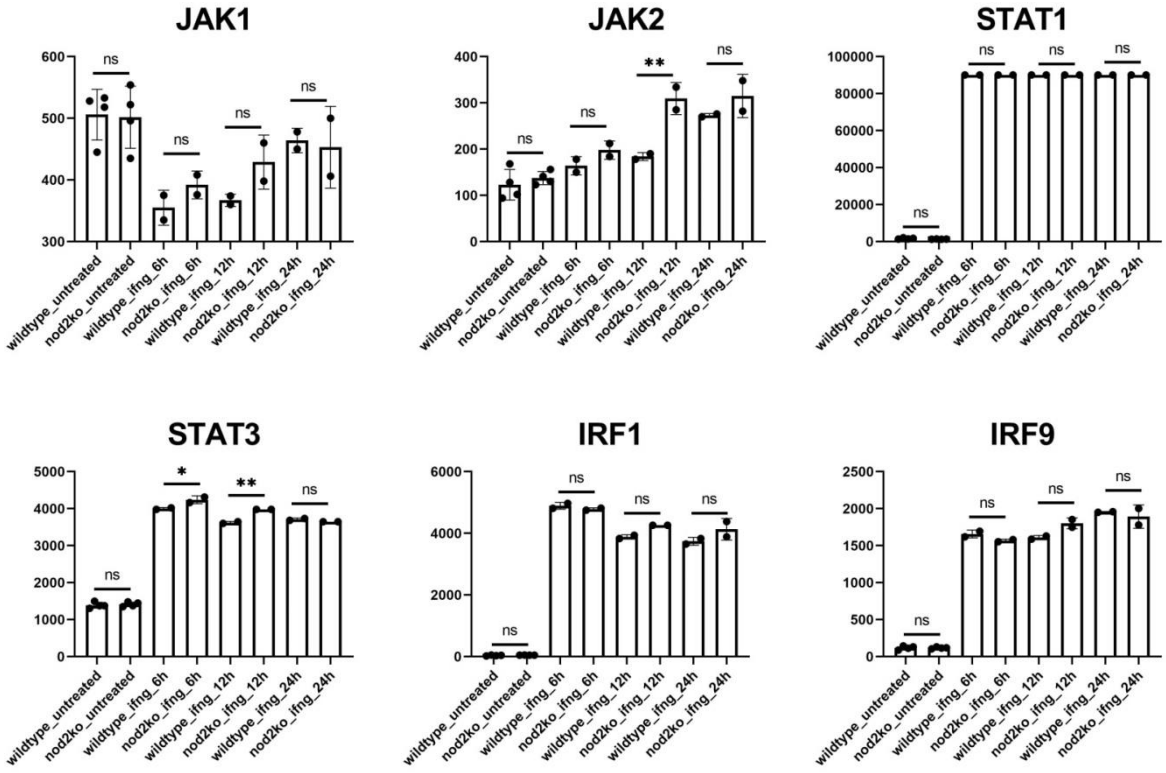


Figure 3.4.8.6 Expression levels of IFN-γ pathway involved key genes in wild type and *NOD2* deficient intestinal organoids

(wildtype_untreated: wild type intestinal organoids without any treatments;
wildtype_ifng_6/12/24h: wild type intestinal organoids with 6/12/24 hours 100 ng/ml IFN-γ treatment;
nod2ko_untreated: *NOD2* deficient intestinal organoids without any treatments;
nod2ko_ifng_6/12/24h: *NOD2* deficient intestinal organoids with 6/12/24 hours 100 ng/ml IFN-γ treatment;)

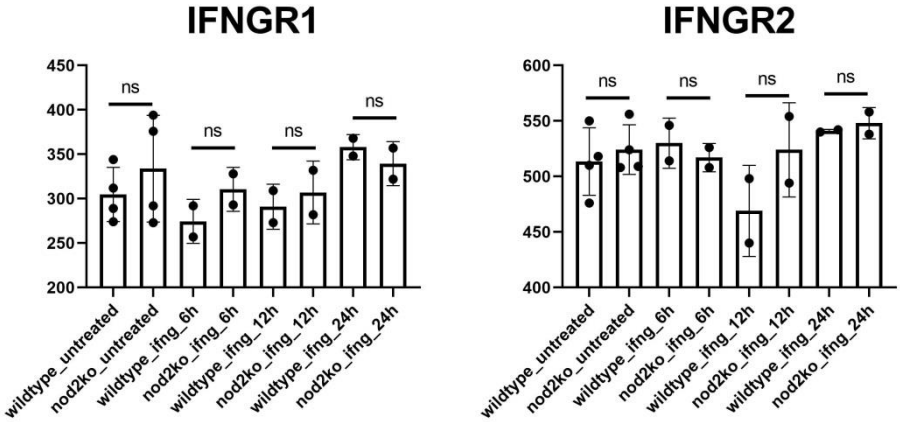


Figure 3.4.8.7 Expression levels of IFN-γ receptor encoding genes in wild type and *NOD2* deficient intestinal organoids

3.5 Discussion

The aim of this chapter was to study the possible dysregulations caused by *NOD2* deficiency and how those dysregulations influence intestinal immune homeostasis by *in vitro* hiPSC-derived macrophages and intestinal organoids. This work is hopefully to contribute to some novel insights between *NOD2* deficiency and CD pathogenesis, besides the MDP-mediated inflammatory response.

For this aim, a hiPSC line from a healthy Caucasian background with no CD-related mutations in both *NOD2* and *LRRK2* gene (from exome sequencing) was used in the experiments, and *NOD2* knockouts were successfully generated from the line by introducing frame shift mutations into exon of the gene via CRISPR-Cas9. Those *NOD2* deficient hiPSCs were differentiated into macrophages *in vitro*, and compared with the wild type controls to reveal the possible differences in gene expression. Interestingly, after removing the outliers among the samples, both ORA and GSEA methods detected down regulation of interferon response in those *NOD2* deficient macrophages without any treatment, especially for IFN-γ. Additionally, down regulation of a LPS response was also detected in the GSEA. This indicates *NOD2* deficiency may introduce primed down regulations of IFN-γ and LPS mediated inflammatory response into macrophages, and leads to reduced basal level of inflammatory response in the cells. Based on those findings, it is also reasonable to hypothesize *NOD2* deficiency can affect IFN-γ and LPS mediated inflammatory response, in addition to the MDP mediated one.

Further experiments started from MDP, which, to date, is the only recognized *NOD2* ligand (Al Nabhani et al., 2017; Rochereau et al., 2021). This bacterial component is thought to be the key of CD pathogenesis caused by *NOD2* dysfunctions (Al Nabhani et al., 2017; Rochereau et al., 2021). MDP was used to stimulate the hiPSC-derived macrophages with and

without *NOD2* deficiency. However, surprisingly, the data indicated that the MDP response in wild type macrophages was nearly identical to that in *NOD2* knockout macrophages. This matched some other works which indicated MDP mediated inflammatory response could still be observed in the absence of *NOD2* (Nayar et al., 2021; Plevy, 2021; Takada et al., 2018) (section 1.3).

Lacking an effect from MDP, a second pro-inflammatory stimulus, LPS, was used in the experiments. LPS binds to the toll-like receptor, *TLR4* (Lu, Yeh, & Ohashi, 2008), and activates a pro-inflammatory pathway (Lu et al., 2008; Nijland, Hofland, & Van Strijp, 2014). In LPS treated macrophages, none of obvious changes on morphology were observed between cells with and without *NOD2* deficiency. For the RNA sequencing data, few DE genes were detected between wild type and *NOD2* deficient macrophages either, and the GSEA only identified the existed biological changes of IFN-γ response in the untreated cells, but without the same high significance. When checking the differences with time series, there were still no obvious changes in the expression of genes which could reflect the sub stages' transfer of LPS mediated pro-inflammatory response in macrophages. All these results from experiments indicated *NOD2* deficiency had few effects on macrophages inflammatory response to LPS, which did not match the hypothesis generated from those untreated samples.

Finally, IFN-γ was used in the experiments. IFN-γ is the only member of the type II class of interferons, and this cytokine is critical for both innate and adaptive immune response, plays important role in activating macrophages *in vivo* (Gray & Goeddel, 1982; Schoenborn & Wilson, 2007; Schroder, Hertzog, Ravasi, & Hume, 2004). Meanwhile, current researches have recognized the intracellular IFN-γ response was mainly via activation of the JAK-STAT pathway (A. J. Lee & Ashkar, 2018). In IFN-γ treated macrophages, none of obvious changes on morphology between cells with and without *NOD2* deficiency were observed, but RNA

sequencing data truly generated some surprising findings, since big differences were identified. GSEA indicated activation of IFN-gamma pathway (mainly JAK-STAT pathway) was significantly down regulated in the *NOD2* deficient macrophages, leading to the reduced inflammatory response. It seemed *NOD2* deficiency did not affect IFN-γ receptors in macrophages, since no differences were identified from the expression levels of *IFNGR1* and *IFNGR2* (the genes encode IFN-γ receptors) in untreated macrophages with and without *NOD2* deficiency. Although in the IFN-γ treated macrophages, *IFNGR1* expression was slightly lower in *NOD2* deficient macrophages. Those slight differences could hard be the main contributors of huge distinctness in IFN-γ response, as normally such slight differences needed a process of accumulation to make great varieties. Moreover, by analyzing the *NOD2* deficient macrophages with and without IFN-γ treatment, besides verifications of the huge effects on macrophages response to IFN-γ which was caused by *NOD2* deficiency, it also indicated that the effects were just down regulation, but not the complete destruction of the IFN-γ mediated inflammatory pathway. Essentially, IFN-γ was not as potent in the *NOD2* deficient macrophages.

As the findings from IFN-γ treated macrophages were so surprising, it was necessary to validate it from other aspects. In this situation, *STAT1*, which is a major mediator of IFN-γ signaling (Bromberg, Horvath, Wen, Schreiber, & Darnell, 1996; Qing & Stark, 2004), and the expression was down-regulated in *NOD2* deficient macrophages treated with IFN-γ (from the analysis of RNA sequencing data), was chosen. From the Western blot, *STAT1* protein was indeed lower in *NOD2* deficient macrophages, compared to the wild type macrophages, which corresponded to the results from RNA-seq. This supports the idea that IFN-γ signaling, is somehow impaired in the *NOD2* deficient macrophages.

Those surprising findings from IFN-γ treated macrophages also led to the queries on

whether this abnormal response to IFN-γ was cell-type, or general in other somatic cells with *NOD2* deficiency. Using *in vitro* derived intestinal organoids, when wild type and *NOD2* deficient organoids were treated by IFN-γ, the organoids had a normal response to IFN-γ stimulation, but intriguingly, none of the differences in response were detected between the wildtype and *NOD2* knockout macrophages. This result confirms that the NOD2-mediated abnormal IFN-γ response is not a generic problem, and whilst it may also be present in other immune cells, it does not exist in intestinal cells and I only observed it in macrophages.

It is interesting to discuss the links between *NOD2* dysfunction and IFN-γ response in Crohn's disease, as analysis of *NOD2* has previously been focused on MDP-mediated pathways. However, something similar appears to occur in Blau syndrome, which is a juvenile-onset systemic granulomatosis known to be related to *NOD2* mutations, the disease associated mutation has been revealed to affect macrophages response to IFN-γ, and causes a more severe inflammatory response (Takada et al., 2018). This is the opposite phenotype seen here, although there is argument concerning whether the Blau syndrome mutation is a gain or loss of function for *NOD2* (Dugan et al., 2015; Kanazawa et al., 2005; Takada et al., 2018). Interestingly, in the research from Takeda et al., (2018), a trend for mild up regulation of an inflammatory response was already evident in *NOD2* mutant macrophages without any pro-inflammatory stimuli (Takada et al., 2018). The work based on the *NOD2* knockout macrophages also identifies dysregulations in the response to IFN-γ in the untreated cells. Furthermore, detailed analysis of RNA-seq data revealed *NOD2* deficiency functions directly on the IFN-γ pathway, which is a cell-type specific effect without recognition in the intestinal organoids.

Additionally, the findings from *NOD2* knockout macrophages are diverse. For example, a study about Blau syndrome used a mutant NOD2 in iPSCs and generated *in vitro*

macrophages (Takada et al., 2018). Both this study and that study used hiPSC-derived macrophages as disease models, however in that study the *in vitro* derived macrophages were immortalized by using a lentiviral vector to introduce 3 transgenes (*cMYC, BMI1*, *MDM2*) (Takada et al., 2018). This process of transformation may impact cell type control as the cells adopt characteristics of cancer. However, some comparisons are nonetheless possible. The response to IFN-γ is down regulated in the *NOD2* knockout macrophages, but the same response is found to be up regulated in their cells with a *NOD2* mutation. A reason for this phenomenon can be complex, including but not limited to different methods for the macrophages' derivation (their method skips the EB formation step), differences between the immortalized hiPSC-derived macrophage cell lines (their work) and the unprocessed macrophages (our work), influence from the genetic background (all the cells used in their work has a Japanese background, whilst ours are Caucasian), and the key one in mind, which is variants in the same gene can have distinct influence for a disease. The last possible reason has been verified in some GWAS analysis on human diseases (Hui et al., 2018; H.-S. Lee et al., 2021), and in this case, it is not unexpected that a heterozygous mutation can causes diverse responses, compared to the knockout of a gene.

Since *NOD2* is recognized as a key susceptibility gene for CD, it is possible that these effects on macrophage response to IFN-γ also exist in the actual disease. In the experiments, the defective IFN-γ response was much more obvious than the macrophages' response to bacterial components (MDP, LPS), which has been recognized as the key in *NOD2* mediated CD pathogenesis. As a cytokine predominantly produced by NK cells or innate lymphoid type 1 cells, IFN-γ is recognized as an endogenous pro-inflammatory stimulator of macrophages (A. J. Lee & Ashkar, 2018). Considering the critical role of macrophages in maintaining intestinal immune homeostasis (section 1.5), it is possible that dysregulations of IFN-γ

response affect this process, perhaps via hindering the macrophages' ability to set a the threshold for the transition from pro-inflammatory to anti-inflammatory conditions, thus leading to chronic intestinal inflammation in CD. Moreover, the *in vitro* derived macrophages are available to isolate effects caused by distinct stimulation, since the cells are free of all the *in vivo* environments. Those results help to reveal the importance of cytokines in CD pathogenesis, which are not induced by the abnormal response to bacterial components, but functions on the cells (at least macrophages) directly. Certainly, clinical evidence to support the findings from *in vitro* models is necessary, which will be discussed in Chapter 5.

CHAPTER 4

IN VITRO STUDY OF THE RELATIONSHIP BETWEEN *LRRK2* DEFICIENCY AND CD PATHOGENESIS

4.1 Background

Leucine-rich repeat kinase 2 (*LRRK2*) protein, is a 286 kDa multi-domain protein with a catalytic core possessing dual activities of GTPase and kinase, has been found to be related to both CD and PD risk (Dzamko, 2017; H.-S. Lee et al., 2021). As mentioned in section 1.4, mutations of this gene have been identified as one of the greatest genetic contributors to Parkinson's disease (PD), with overwhelming number of studies in neurons, and the gene's disorders in immune homeostasis has attracted more and more attention, as a possible link between CD and PD. However, the brain has been considered as a 'sterile' organ, which has no microbes and few immune cell types (Ratajczak et al., 2018). In this situation, studying why *LRRK2* dysregulations can increase both PD and CD risk with some possibly shared pathogenesis should start from further understanding how *LRRK2* functions in maintaining intestinal immune homeostasis.

There are quite a few studies on the links between *LRRK2* and CD from the perspective of molecular mechanisms, there is no clear model and there are conflicting results. Around a sixfold increase of *LRRK2* expression has been found in the comparison between severely inflamed and non-severely inflamed areas of intestinal biopsies from the same CD patient, and differences in the gene expression have been specifically identified in the immune cell types such as T lymphocytes, B lymphocytes, neutrophils, macrophages and dendritic cells, and have no relationship with intestinal epithelial cells (Dzamko, 2017; Gardet et al., 2010). In the mouse models described in section 1.6, although studies on *LRRK2* deficient mice has identified mice that have distinct susceptibility to some specific gut microbes that caused intestinal infection (Dzamko, 2017; Q. Zhang et al., 2015), it is still not clear whether *LRRK2* expression induces or reduces intestinal inflammation. In the DSS-induced models (the gold standard model for studying colitis in rodent models), both knockout and overexpression of

LRRK2 in the mice lead to more severe colitis (Dzamko, 2017; H.-S. Lee et al., 2021; Liu et al., 2011; Takagawa et al., 2018).

It is quite a novel topic to study the molecular mechanisms of the link between CD and PD, and most of the work still rests on observations. Biochemical analysis of PBMCs from both CD and PD patients has identified a shared higher expression of *LRRK2* relative to the healthy controls, and in DSS models, alterations in the brain has been found, along with intestinal inflammation (Herrick & Tansey, 2021). Considering the key roles *NOD2* and *LRRK2* play in CD and PD respectively, it is reasonable to suggest that the two genes may have some uncovered relationship, with shared problems in both diseases.

4.2 Aims

The aim of this chapter is to explore why *LRRK2* deficiency, a key determinant of PD, can affect intestinal immune homeostasis and contribute to CD pathogenesis via the *in vitro* derived macrophage model cells. Meanwhile, since the models are isogenic, hopefully it is able to find out whether there are inter-relationships between CD and PD on their key susceptible genes - *NOD2* and *LRRK2* respectively.

4.3 Materials and Methods

The *LRRK2* knockout hiPSCs used in these experiments were generated from the same wild type cell line as the *NOD2* knockout, with a frame shift mutation on exon 9 of the gene (2bp deletion, homozygous). This gene-edited line was published as "C11-LRRK2-KO1" (S. Chen et al., 2020). The cell line was obtained from the Guangzhou Institutes of Biomedicine and Health (Guangzhou, China). Maintenance of this cell line was the same as the wild type hiPSCs (see section 2.1 for more details).

Mycoplasma test of the *LRRK2* knockout hiPSCs followed the methods in section 2.2, and the result was negative (Figure 3.3.1.7).

For *in vitro* differentiation of macrophages, details of the methods were in section 2.3. For RNA extraction, sequencing and analysis, methods followed the descriptions in section 2.6 and 2.7. Analysis of the genes co-expression was by STRING (version 11.5) (https://string-db.org/) (Mering et al., 2003; Szklarczyk et al., 2010; Szklarczyk et al., 2021; Szklarczyk et al., 2016).

4.4 Results

4.4.1 Co-expression of *LRRK2* and *NOD2* may exist in *Homo sapiens*

Although STRING is mainly the database for protein-protein interaction, analysis from its online version (https://string-db.org/) (version 11.5) can also be used to predict whether the genes correspond to interesting proteins that have co-expression or not, based on statistical analysis of evidence from a large number of experiments (Szklarczyk et al., 2021). Since there is a relationship between *NOD2* and LRRK2 in CD and PD, these two genes were selected for study. Interestingly, results of the analysis indicated co-expression between *LRRK2* and *NOD2* had higher confidence in *Homo sapiens* compared with any other organisms (Figure 4.4.1.1). This implied relationship between *NOD2* and *LRRK2* might exist in humans more possibly comparing with any other organisms.

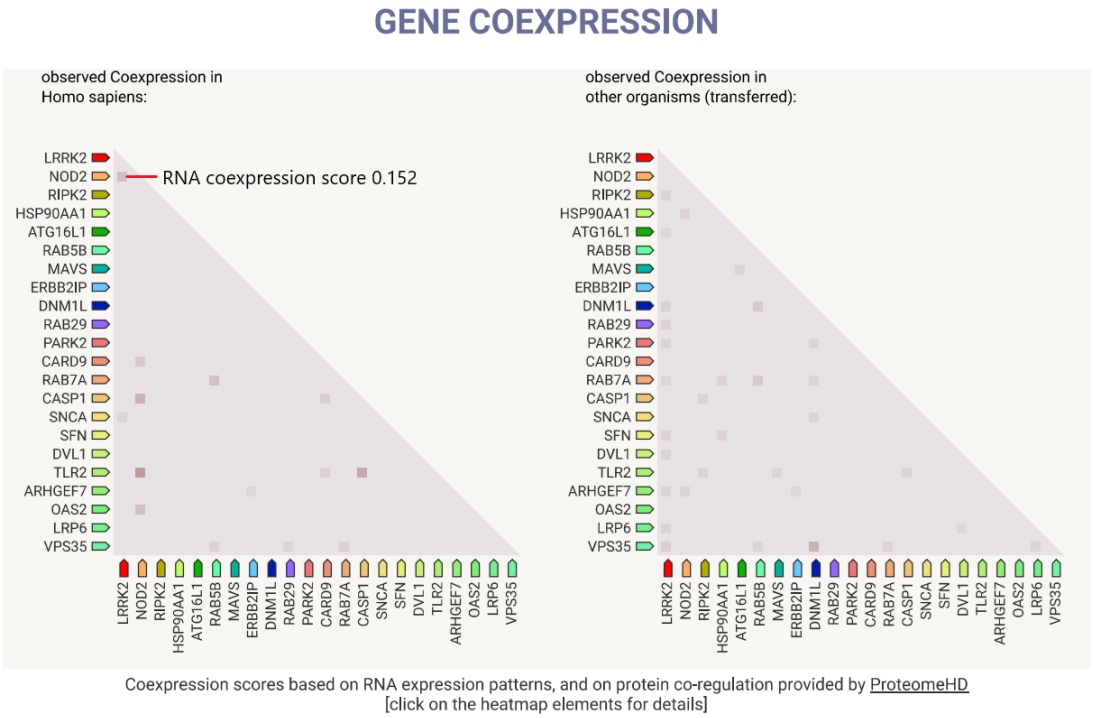


Figure 4.4.1.1 Co-expression between *LRRK2* and *NOD2* gene in *Homo sapiens*

(left: *Homo sapiens*; right: other organisms.

Rows and columns were the genes corresponding to the proteins which were predicted to be involved in the interaction between *NOD2* and *LRRK2*. RNA coexpression score indicated the confidence of the coexpression between the genes, and higher score meant higher possible confidence of the genes' coexpression. All the scores over 0.1 were marked out in the heatmap.)

4.4.2 *NOD2* deficiency significantly decreases *LRRK2* expression in IFN-γ treated hiPSC-derived macrophages

To verify co-expression between *NOD2* and *LRRK2*, expressions of those two genes were compared in the RNA sequencing data from wild type hiPSC-derived macrophages and intestinal organoids (Figure 4.4.2.1). As expected, co-expression existed between those two genes, as when *LRRK2* expression increased, expression of *NOD2* increased as well. Additionally, this phenomenon was obvious in macrophages treated with IFN-γ, which indicated those macrophages might be a good candidate to explore the relationship between the two genes.

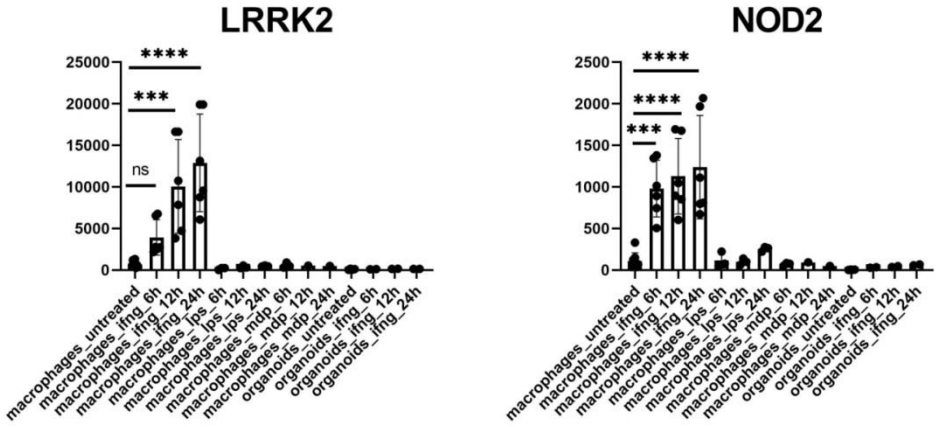


Figure 4.4.2.1 Expression levels of *LRRK2* and *NOD2* genes in the RNA-seq data from wild type hiPSC-derived macrophages and intestinal organoids

(macrophages_untreated: wild type macrophages without any treatments;
macrophages_ifng_6/12/24h: wild type macrophages with 6/12/24 hours 100 ng/ml IFN-γ treatment;
macrophages_lps_6/12/24h: wild type macrophages with 6/12/24 hours 1 μg/ml LPS treatment;
macrophages_mdp_6/12/24h: wild type macrophages with 6/12/24 hours 1 μg/ml MDP treatment;
organoids_untreated: wild type intestinal organoids without any treatments;
organoids_ifng_6/12/24h: wild type intestinal organoids with 6/12/24 hours 100 ng/ml IFN-γ treatment)

Next, data from the IFN-γ treated *NOD2* deficient macrophages, which was obtained in Chapter 3, to explore an interplay between NOD2 and LRRK2. Expression of *LRRK2* in those cells were compared relative to the wild type reference, and obvious up regulation of *LRRK2* expression after IFN-γ treatment was found. This phenomenon existed in wild type macrophages, but could not be detected in the *NOD2* deficient macrophages (*LRRK2* expression only had slight increase after IFN-γ treatment in those *NOD2* deficient macrophages) (Figure 4.4.2.2). This suggests *LRRK2* may be downstream of *NOD2*, or, alternatively, that deregulation of *LRRK2* is a consequence of the IFN-γ dysregulation in the *NOD2* knockout macrophages. The predicted off-targets for the sgRNA used in generating the NOD2 knockout hiPSC line was also checked (section 3.4.1), and no computationally predicted off-targets were found in the *LRRK2* gene (Appendix). Moreover, this phenomenon seemed to be specific to IFN-γ treated hiPSC-derived macrophages, as in both macrophages with other pro-inflammatory stimulation (LPS) and intestinal organoids treated by IFN-γ, none of obvious differences in *LRRK2* expression between wild type and *NOD2* deficient cells/tissues were detected (Figure 4.4.2.3 and 4.4.2.4). In this situation, it was reasonable to consider whether the surprising down regulation of IFN-γ response in *NOD2* deficient macrophages (section 3.4.4) was related to *NOD2*'s effects on *LRRK2* expression.

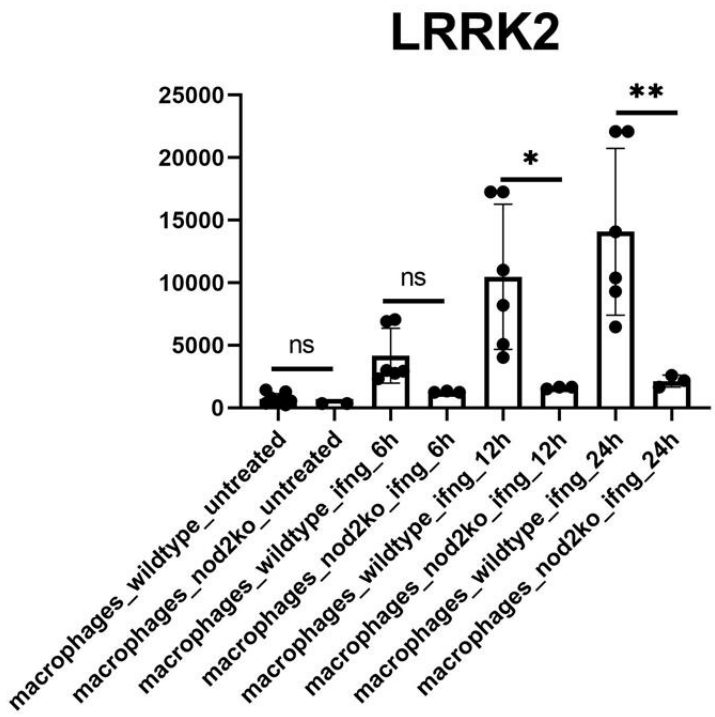


Figure 4.4.2.2 *LRRK2* expression in IFN-γ treated wild type and *NOD2* deficient macrophages

(macrophages_wildtype_untreated: wild type macrophages without any treatments;
macrophages_wildtype_ifng_6/12/24h: wild type macrophages with 6/12/24 hours 100 ng/ml IFN-γ treatment;
macrophages_nod2ko_untreated: *NOD2* deficient macrophages without any treatments;
macrophages_nod2ko_ifng_6/12/24h: *NOD2* deficient macrophages with 6/12/24 hours 100 ng/ml IFN-γ treatment)

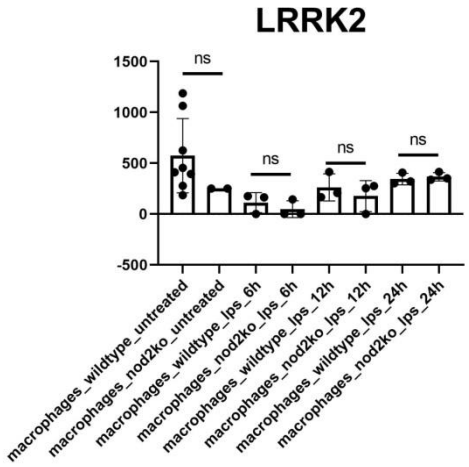


Figure 4.4.2.3 *LRRK2* expression in LPS treated wild type and *NOD2* deficient macrophages

(macrophages_wildtype_untreated: wild type macrophages without any treatments
macrophages_wildtype_lps_6/12/24h: wild type macrophages with 6/12/24 hours 1 μg/ml LPS treatment;
macrophages_nod2ko_untreated: *NOD2* deficient macrophages without any treatments;
macrophages_nod2ko_lps_6/1/2/24h: *NOD2* deficient macrophages with 6/12/24 hours 1 μg/ml LPS treatment)

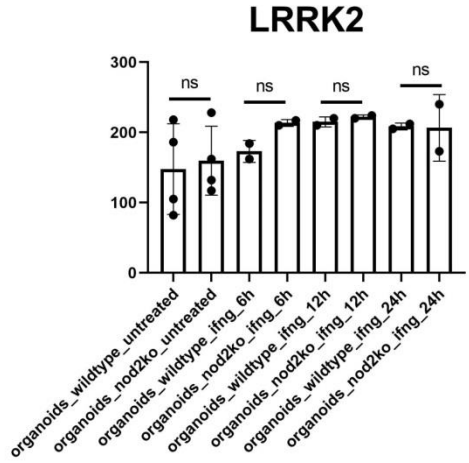


Figure 4.4.2.4 *LRRK2* expression in IFN-γ treated wild type and *NOD2* intestinal organoids

(organoids_wildtype_untreated: wild type intestinal organoids without any treatments;
organoids_wildtype_ifng_6/12/24h: wild type intestinal organoids with 6/12/24 hours 100 ng/ml IFN-γ treatment;
organoids_nod2ko_untreated: *NOD2* deficient intestinal organoids without any treatments;
organoids_nod2ko_ifng_6/12/24h: *NOD2* deficient intestinal organoids with 6/12/24 hours 100 ng/ml IFN-γ treatment)

4.4.3 *In vitro* differentiation of *LRRK2* deficient macrophages

To verify whether the surprising down regulation of IFN-γ response in *NOD2* deficient macrophages (section 3.4.4) was related to *NOD2*'s effects on *LRRK2* expression identified in section 4.4.2, the *LRRK2* deficient macrophages were generated *in vitro*, to see if the IFN-γ mediated inflammatory response in macrophages could be affected in the absence of *LRRK2*. The *LRRK2* knockout hiPSCs were used to generate *LRRK2* deficient in vitro-derived macrophages, using the protocol described previously (Buchrieser et al., 2017; Flynn et al., 2015; Gutbier et al., 2020; Karlsson et al., 2008; van Wilgenburg et al., 2014; Wilgenburg et al., 2013) (section 2.3) (Figure 4.4.3.1).

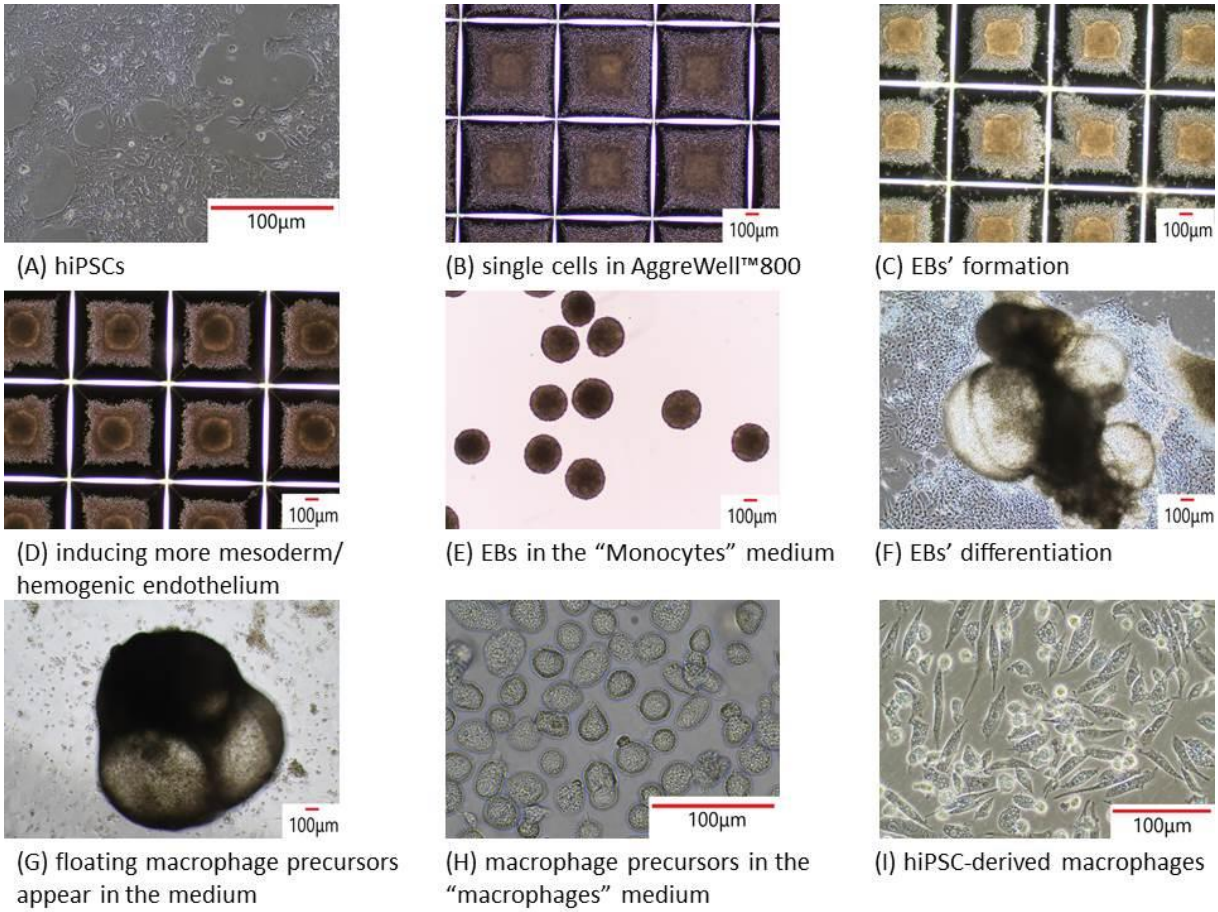


Figure 4.4.3.1 *In vitro* differentiation of *LRRK2* deficient macrophages

(Figure (A) to (I) displayed the process of the *in vitro* differentiation. The *LRRK2* knockout hiPSCs followed the same stages as the wild type and *NOD2* knockout hiPSCs in section 3.3.2, and successfully generated the *LRRK2* deficient macrophages *in vitro*.)

4.4.4 IFN-γ response is mainly intact in *LRRK2* deficient macrophages

As previous, firstly the untreated *LRRK2* deficient macrophages were compared with wild type ones. There were essentially no identifiable morphological differences identified between the two lines (Figure 4.4.4.1).

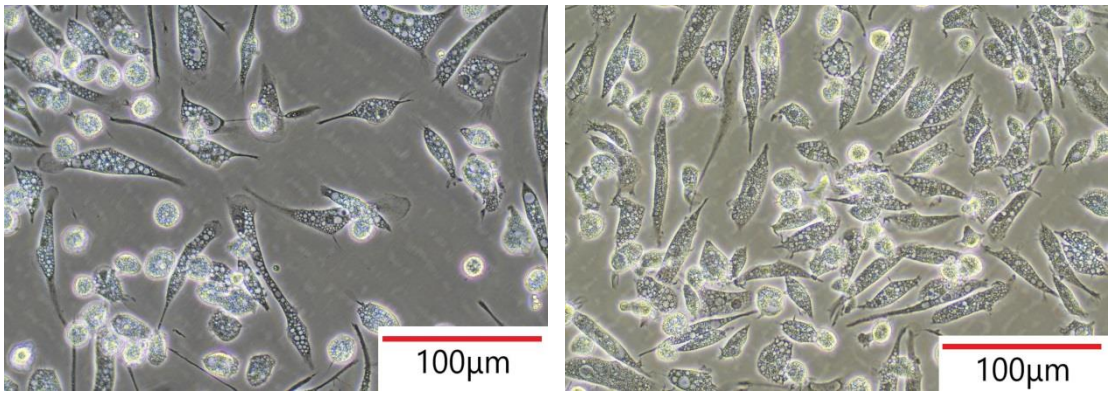


Figure 4.4.4.1 Comparison between the morphology of the untreated wild type and *LRRK2* deficient macrophages
(left: wild type macrophages; right: *LRRK2* deficient macrophages)

For RNA sequencing, total RNA of *LRRK2* deficient macrophages was isolated for bulk RNA sequencing, from three biological replicates (lrrk2ko_untreated_rp1-3). Correlation distances with the wild type macrophages were calculated (Figure 4.4.4.2), and no likely outliers were found. The heatmap of correlation distances did not indicate major differences between wild type and *LRRK2* deficient macrophages, suggesting that the IFN-γ response is not affected in those *LRRK2* deficient macrophages. By DGE analysis between the untreated *LRRK2* deficient and wild type macrophages, there was only identified a very limited number of DE genes (36 up regulated and 90 down regulated) (Figure 4.4.4.3). Meanwhile, GSEA did

not identified terms seemingly very related to immune response, with only limited ones related to mitochondria (Figure 4.4.4.4). This indicates that the knockout of *LRRK2* does not appear to substantially affect wildtype *in vitro* derived macrophages.

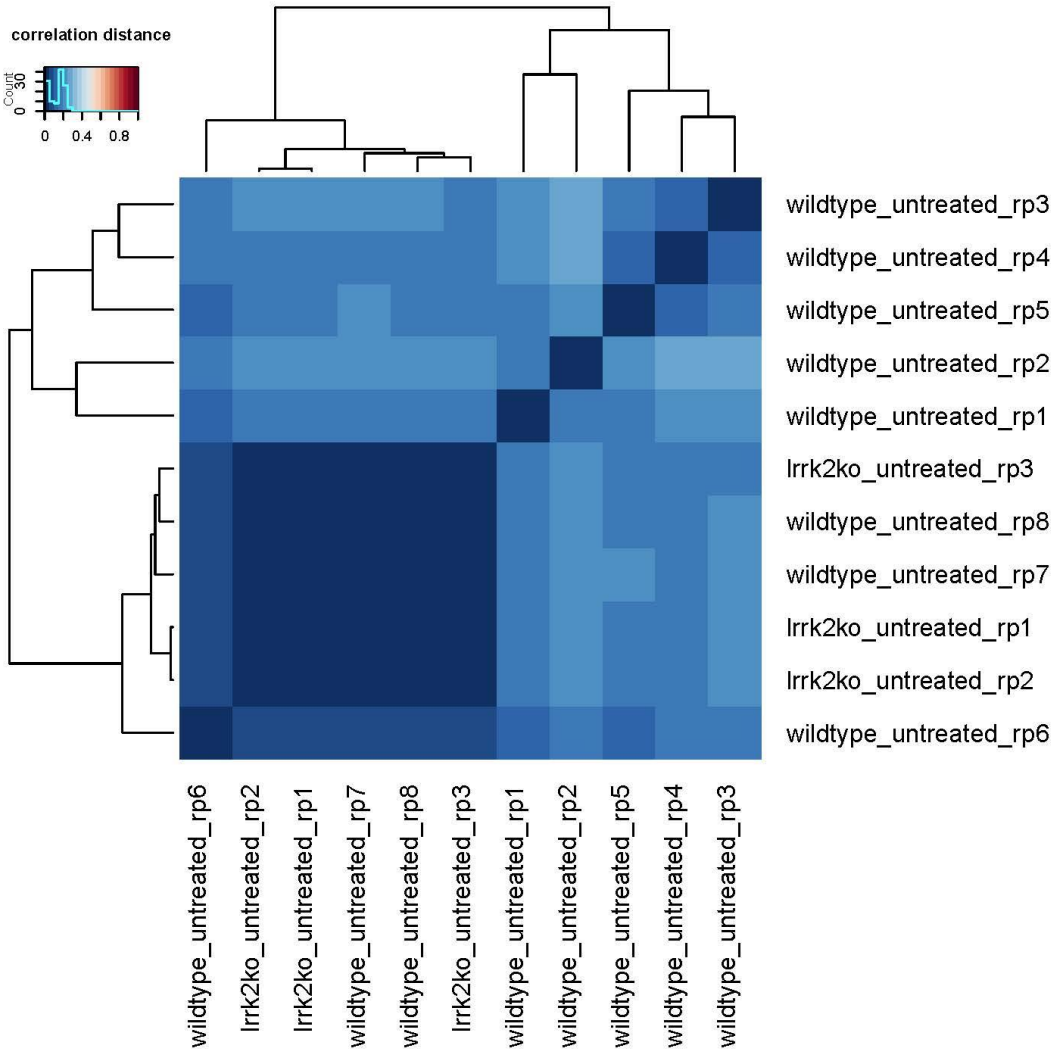


Figure 4.4.4.2 Correlation distances between the wild type and *LRRK2* deficient macrophages

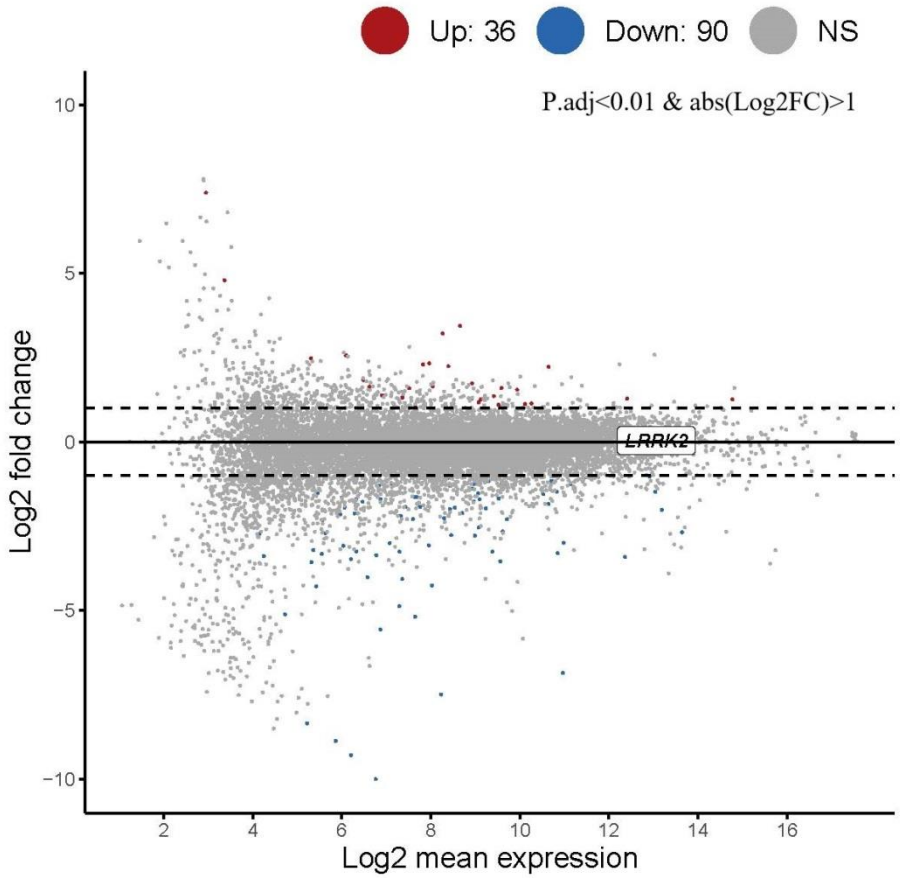


Figure 4.4.4.3 DE genes between untreated *LRRK2* deficient and wild type macrophages (DE genes were selected by the criteria "p.adj < 0.01" and "fold change>2" after DGE analysis, and the results were displayed by MA plot)

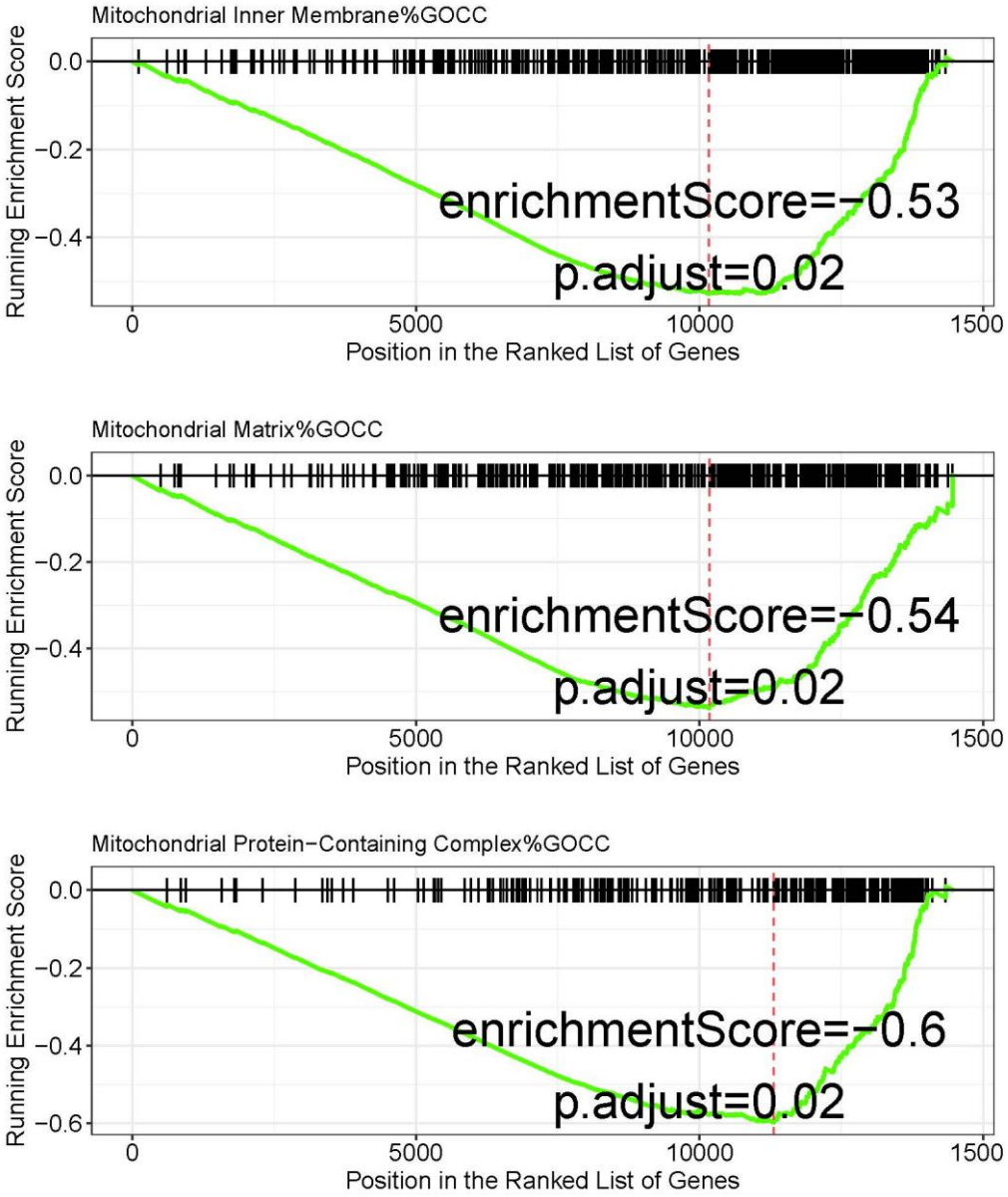


Figure 4.4.4.4 Limited terms related to mitochondria were not significantly enriched (p.adjust > 0.01) **in the comparison between untreated *LRRK2* deficient and wild type macrophages**

("Mitochondrial Inner Membrane" (GO: 0005743);

"Mitochondrial Matrix" (GO:0005759);

"Mitochondrial Protein-Containing Complex" (GO: 0098798))

Next, more *LRRK2* deficient macrophages were collected, then treated by 100 ng/ml recombinant human IFN-γ in the same time course as section 3.4.4 (6 hours, 12 hours and 24 hours) for bulk RNA sequencing. For each time point, three biological replicates were generated (lrrk2ko_ifng_6h_rp1-3, lrrk2ko_ifng_12h_rp1-3, lrrk2ko_ifng_24h_rp1-3). By calculating correlation distance between the samples (Figure 4.4.4.5), "lrrk2ko_ifng_12h_rp2" was a substantial outlier, outside of other related samples from the dendrogram. However, despite the correlation distance between this sample and the others was not high, the density plot of transcript counts for this sample (Figure 4.4.4.6) was normal. Hence, his sample "lrrk2ko_ifng_12h_rp2" was not excluded from the downstream analysis.

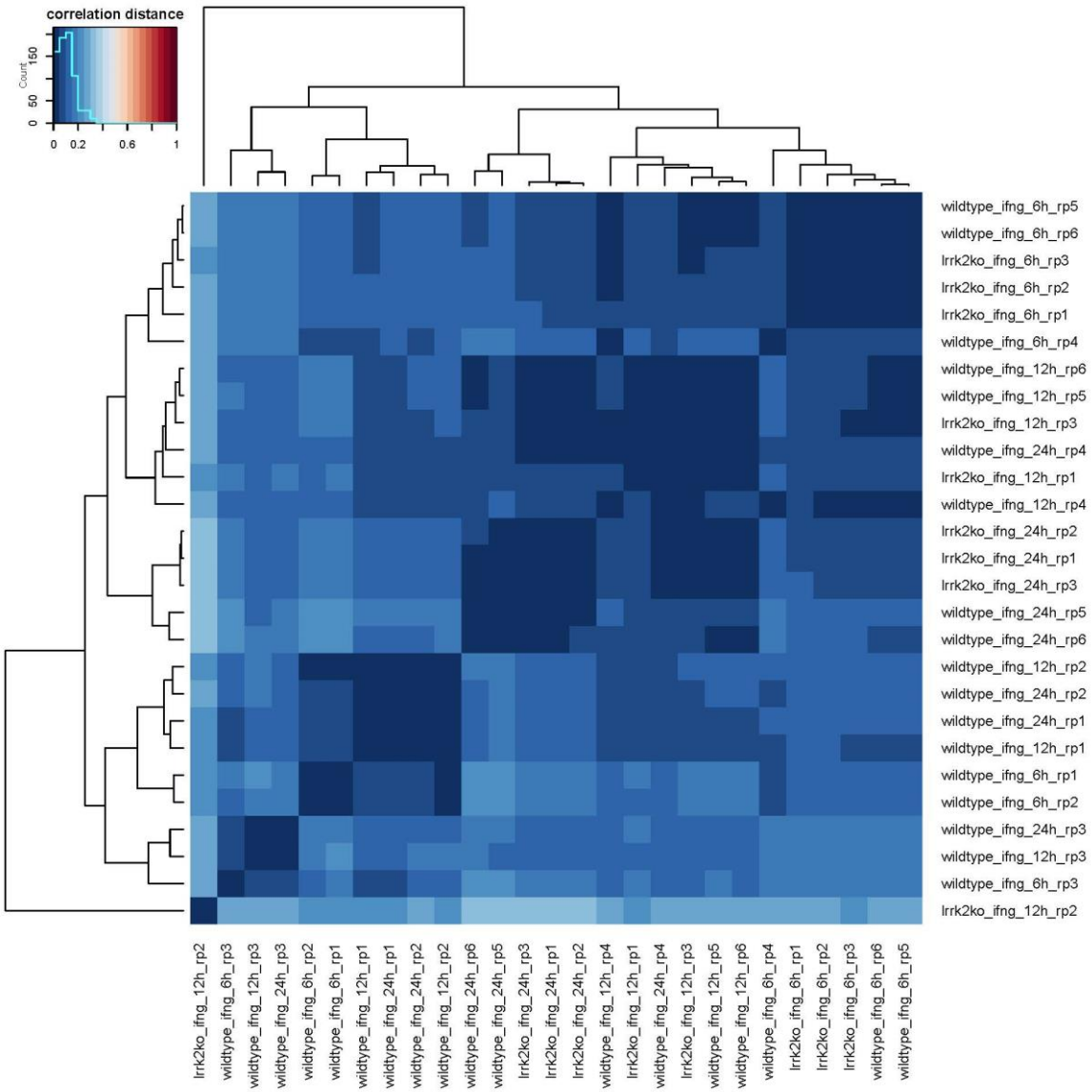


Figure 4.4.4.5 Correlation distances between the wild type and *LRRK2* deficient macrophages with IFN-γ treatment

(Samples' correlation distance was displayed by heat map. Most *LRRK2* deficient samples mixed with the wild type samples. However, it seemed "lrrk2ko_ifng_12h_rp2" was outside from all the other samples.)

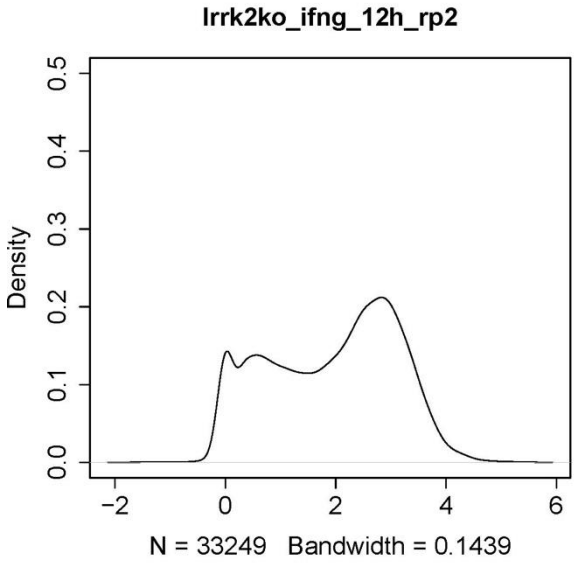


Figure 4.4.4.6 Density plot of transcript counts for "lrrk2ko_ifng_12h_rp2"

DGE analysis was done between *LRRK2* deficient and wild type macrophages with IFN-γ treatment. 157 up and 237 down regulated genes were identified by the criteria "p.adj < 0.01" and "fold change>2" (Figure 4.4.4.7), which indicated some, but not huge differences existed in those IFN-γ treated *LRRK2* deficient macrophages. In both ORA and GSEA, effects of the response to IFN-γ were not identified (section 4.4.6). Furthermore, there were no differences on the expression of key genes involved in IFN-γ pathway (*JAK1, JAK2, STAT1, STAT3, IRF1, IRF9*) (Figure 4.4.4.8). Overall these data support the notion that *LRRK2* loss has little or no effect on the IFN-γ response.

To further verify whether *LRRK2* involved in the surprising down regulation of IFN-γ response in *NOD2* deficient macrophages (section 3.4.4), the overlap down regulated genes between IFN-γ treated *NOD2* and *LRRK2* deficient macrophages were identified (Figure 4.4.4.9), and ORA was performed on them. In the results, the significantly enriched terms were not related to IFN-γ response (Figure 4.4.4.10). This suggests *LRRK2* has little relationships with the surprising down regulation of IFN-γ response in *NOD2* deficient macrophages (section 3.4.4).

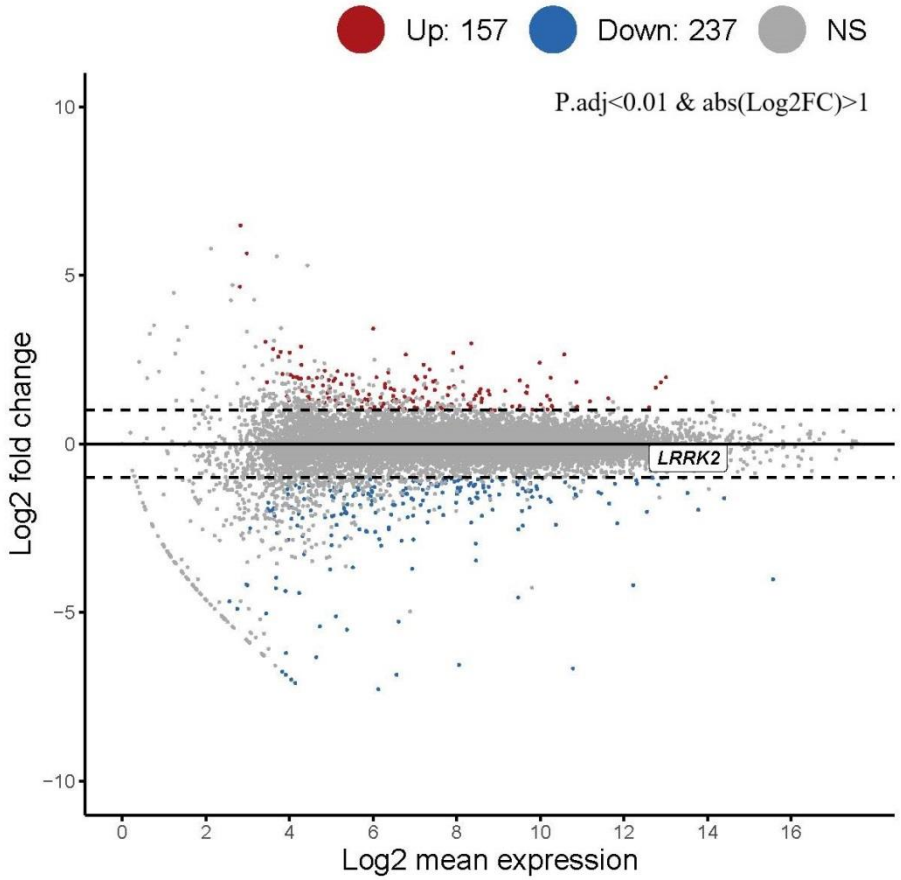


Figure 4.4.4.7 DE genes in *LRRK2* deficient macrophages with IFN-γ treatment

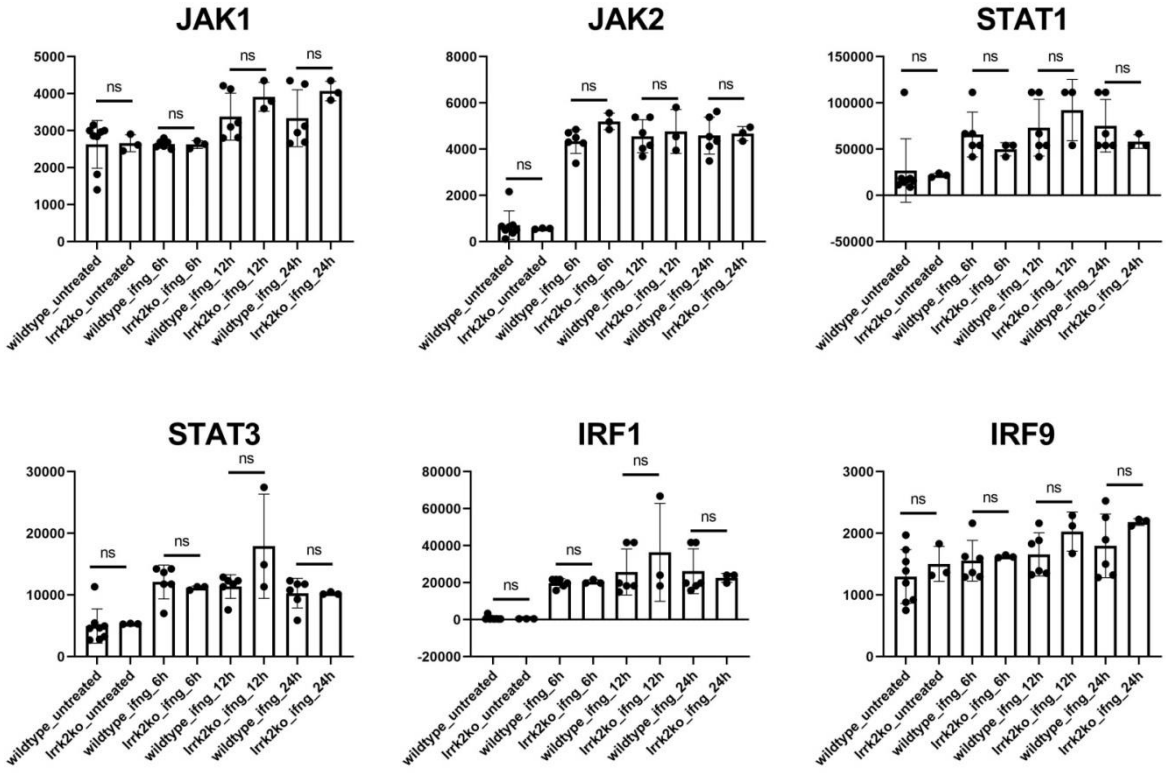


Figure 4.4.4.8 Expression levels of some key genes involved in IFN-γ pathway in wild type and *LRRK2* deficient macrophages

(wildtype_untreated: wild type macrophages without any treatments;

wildtype_ifng_6/12/24h: wild type macrophages with 6/12/24 hours 100 ng/ml IFN-γ treatment;

lrrk2ko_untreated: *LRRK2* deficient macrophages without any treatments;

lrrk2ko_ifng_6/12/24h: *LRRK2* deficient macrophages with 6/12/24 hours 100 ng/ml IFN-γ treatment)

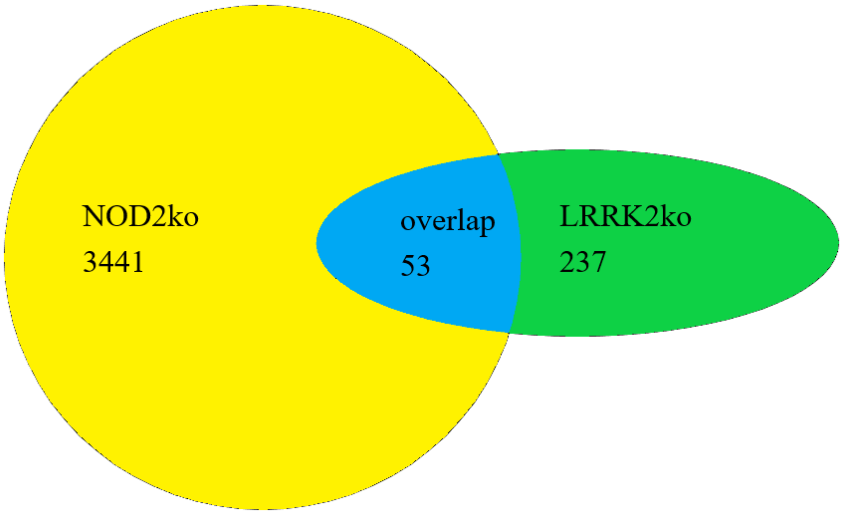


Figure 4.4.4.9 Overlap down regulated genes between IFN-γ treated *NOD2* and *LRRK2* deficient macrophages

(Comparing with wild type reference, 3441 down regulated genes were identified from IFN-γ treated *NOD2* deficient macrophages. For IFN-γ treated *LRRK2* deficient macrophages, this value was 237. 53 genes were overlapped between IFN-γ treated *NOD2* and *LRRK2* deficient macrophages)

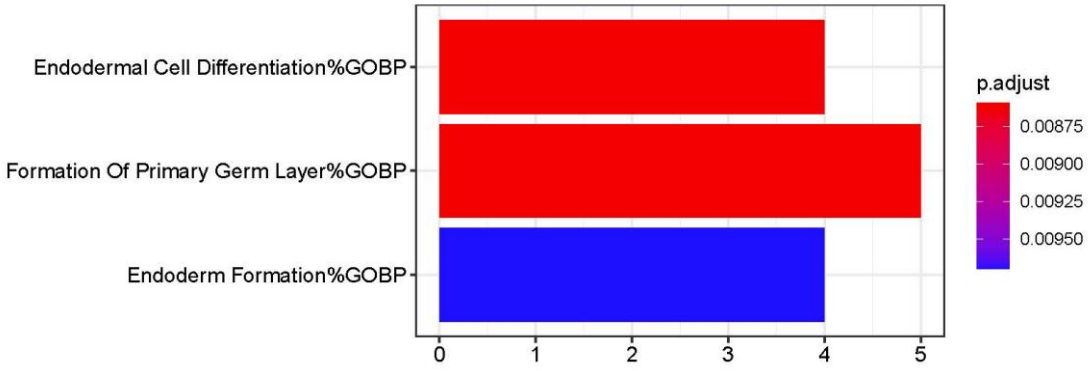


Figure 4.4.4.10 ORA of the overlap down regulated genes (p.adjust ≤ 0.01)

4.4.5 *LRRK2* deficiency has no effect on increased *NOD2* expression in IFN-γ treated hiPSC-derived macrophages

It was attractive to know whether *LRRK2* deficiency could make effects on *NOD2* expression in IFN-γ treated macrophages, as *NOD2* deficiency made on *LRRK2* expression (section 4.4.2). For this aim, *NOD2* expression in those IFN-γ treated *LRRK2* deficient macrophages were viewed, comparing with the wild type reference (Figure 4.4.5.1). The result indicated *LRRK2* deficiency had no effects on *NOD2* expression when macrophages were treated by IFN-γ. Considering all of the macrophages in the experiments described here were isogenic, this suggested the relationship between *NOD2* and *LRRK2* in IFN-γ treated macrophages was not bidirectional, *NOD2* could affect *LRRK2*, but *LRRK2* had almost no influence on *NOD2*.

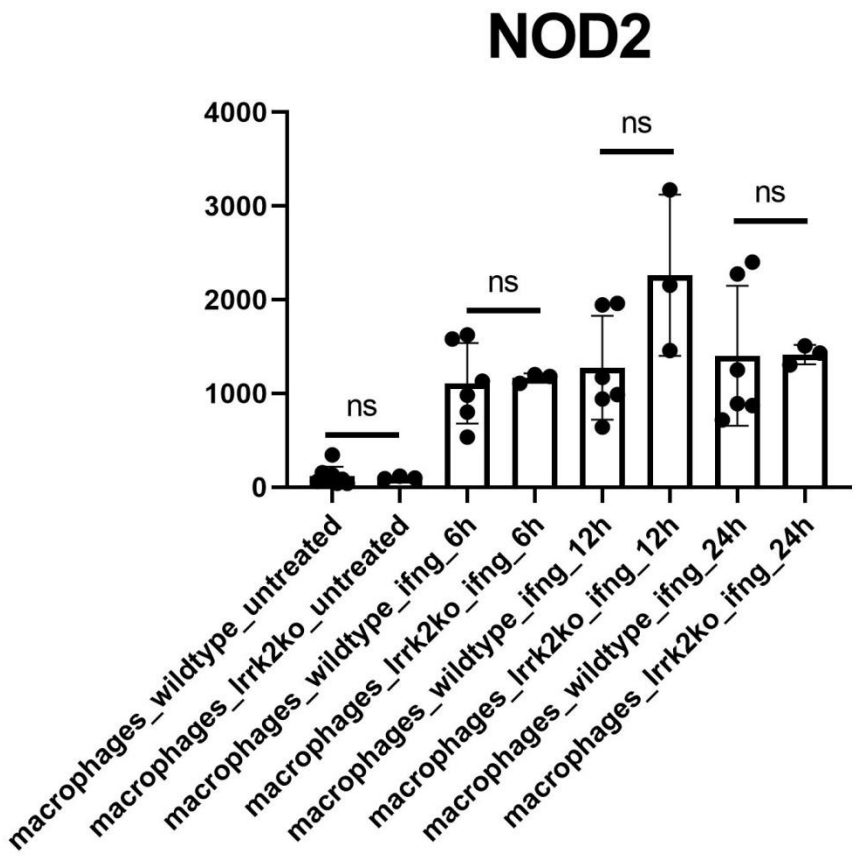


Figure 4.4.5.1 *NOD2* expression in untreated and IFN-γ treated wild type and *LRRK2* deficient macrophages

(macrophages_wildtype_untreated: wild type macrophages without any treatments;
macrophages_wildtype_ifng_6/12/24h: wild type macrophages with 6/12/24 hours 100 ng/ml IFN-γ treatment;
macrophages_lrrk2ko_untreated: *LRRK2* deficient macrophages without any treatments;
macrophages_lrrk2ko_ifng_6/1/2/24h: *LRRK2* deficient macrophages with 6/12/24 hours 100 ng/ml IFN-γ treatment)

4.4.6 *LRRK2* deficiency affects oxidative phosphorylation via mitochondria in IFN-γ treated macrophages

Although *LRRK2* deficiency did not make big differences on IFN-γ mediated inflammatory response in macrophages, it was still meaningful to further explore what effects it could bring to the macrophages, as 157 up and 237 down regulated genes were identified from the DGE analysis relative to the wild type reference. To interpret the differences between IFN-γ treated *LRRK2* deficient and wild type macrophages, both ORA and GSEA were performed. In the ORA results between IFN-γ treated LRRK2 deficient and wild type macrophages, significant enrichment of inflammatory response from the up regulated genes was identified (Figure 4.4.6.1). At the same time, significant enrichments of mitochondrial related terms were also identified from the down regulated genes (Figure 4.4.6.2). In the GSEA results, unlike untreated macrophages, down regulation of oxidative phosphorylation in IFN-γ treated *LRRK2* deficient macrophages was significantly identified (Figure 4.4.6.3), with many terms related to mitochondria and cellular energy supply chain significantly affected (Figure 4.4.6.4). Meanwhile, TNFα signaling via NF-kB pathway was identified to be up regulated (Figure 4.4.6.5). Those results indicated *LRRK2* deficiency could affect oxidative phosphorylation via mitochondria in IFN-γ treated macrophages, and led to some more severe inflammation.

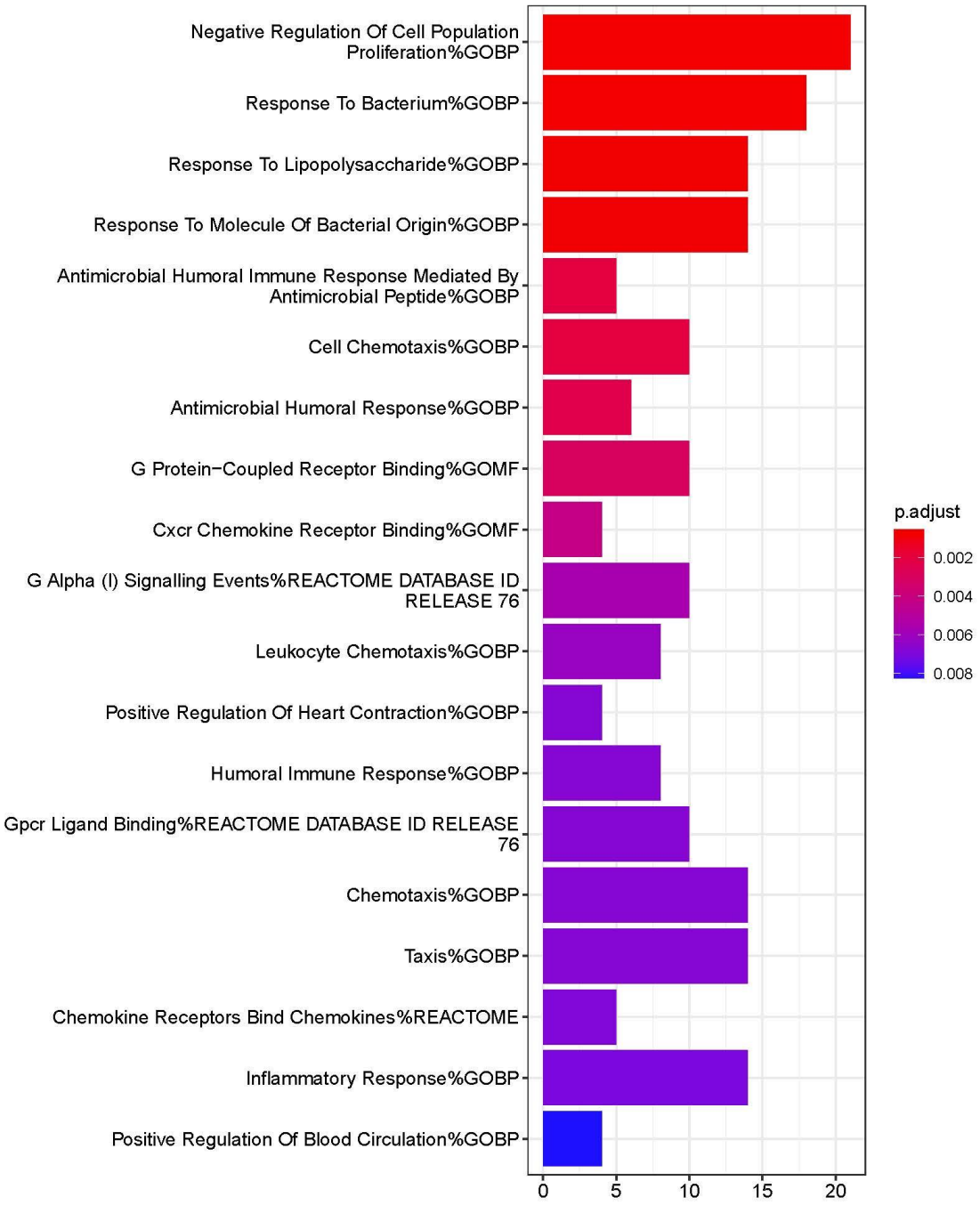


Figure 4.4.6.1 ORA results of up regulated genes in IFN-γ treated *LRRK2* deficient macrophages

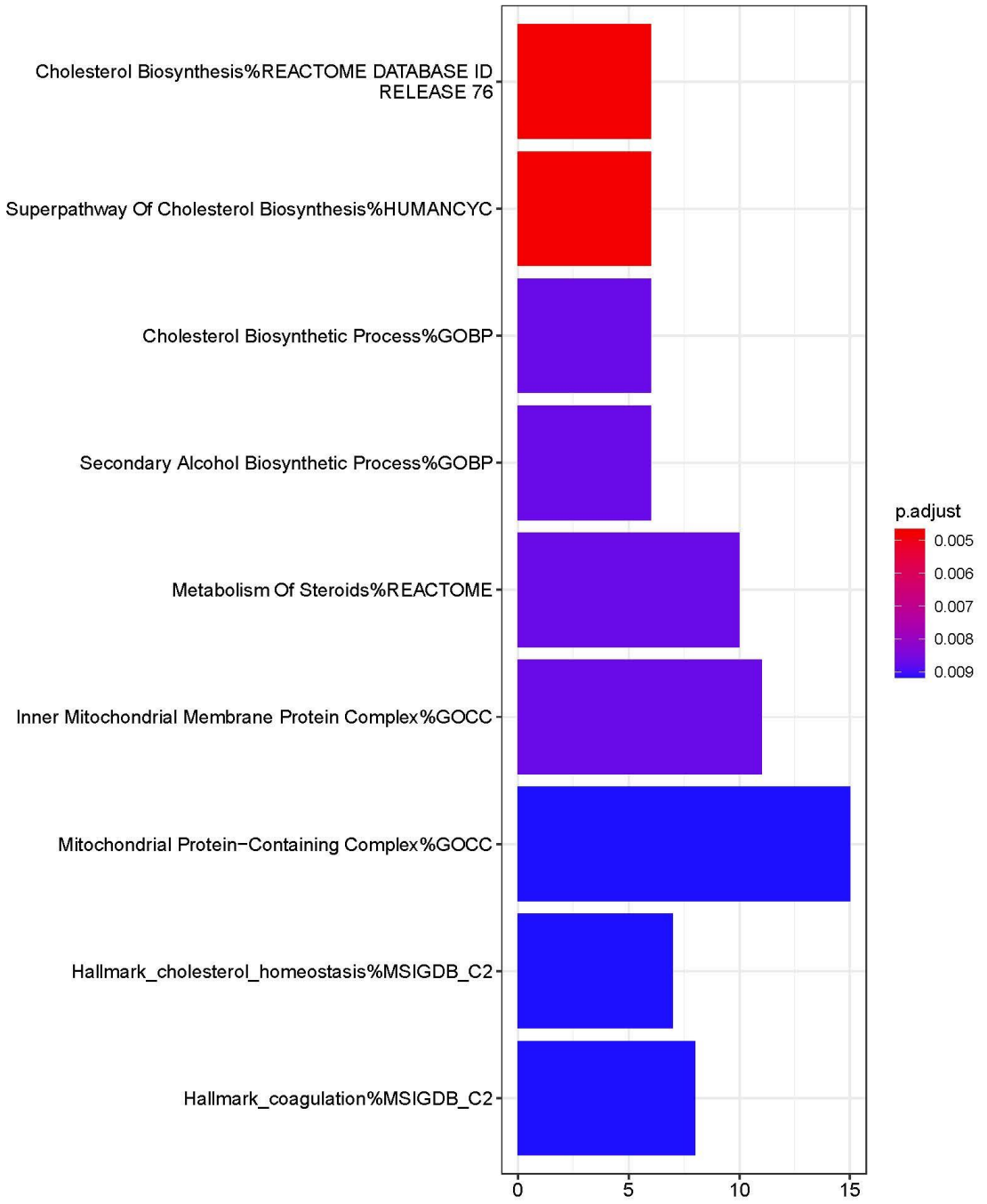


Figure 4.4.6.2 ORA results of down regulated genes in IFN-γ treated *LRRK2* deficient macrophages

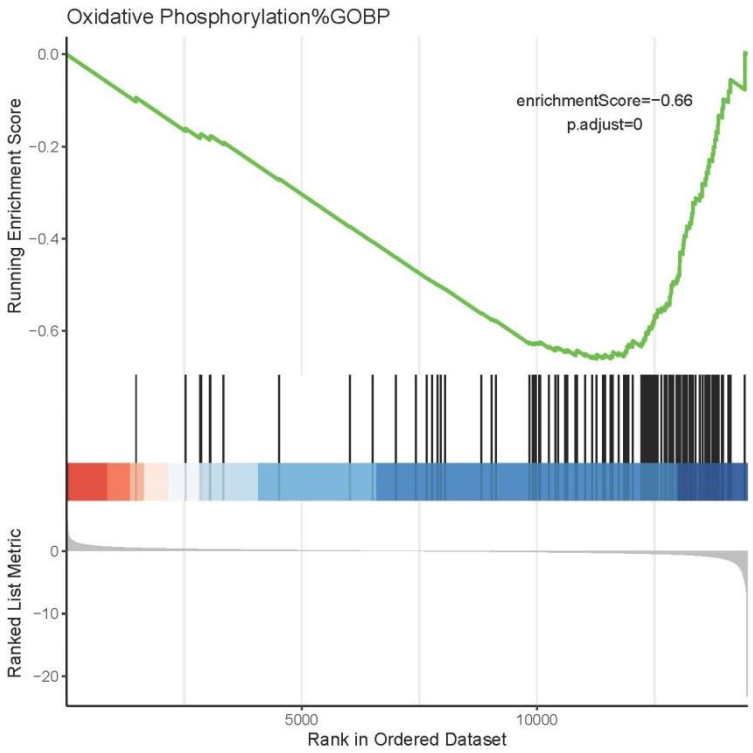


Figure 4.4.6.3 Down regulation of "Oxidative Phosphorylation" (GO: 0006119) in IFN-γ treated *LRRK2* deficient macrophages

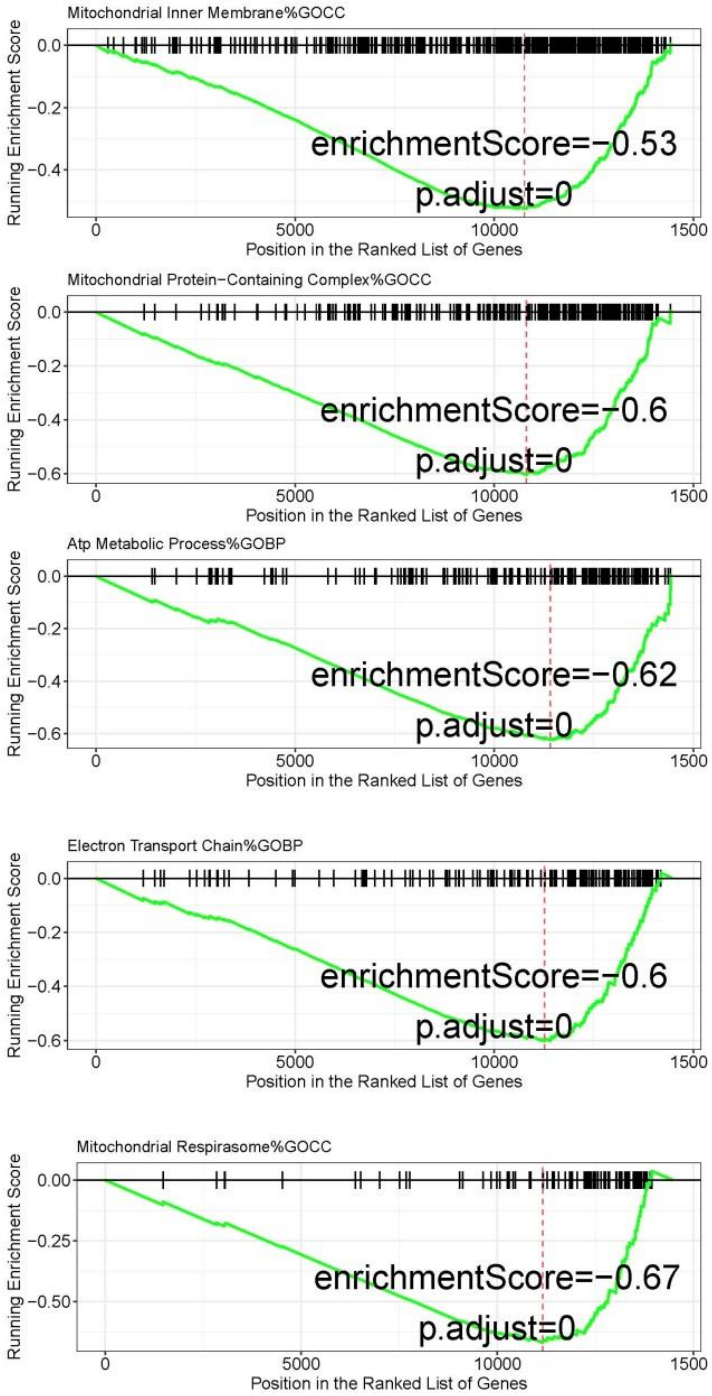


Figure 4.4.6.4 Down regulations of mitochondria and cellular energy supply chain related terms in IFN-γ treated *LRRK2* deficient macrophages (p.adjust ≤ 0.01)
("Mitochondrial Inner Membrane" (GO: 0005743);
"Mitochondrial Protein-Containing Complex" (GO: 0098798);
"Atp Metabolic Process" (GO: 0046034);
"Electron Transport Chain" (GO: 0022900);
"Mitochondrial Respirasome" (GO: 0005746))

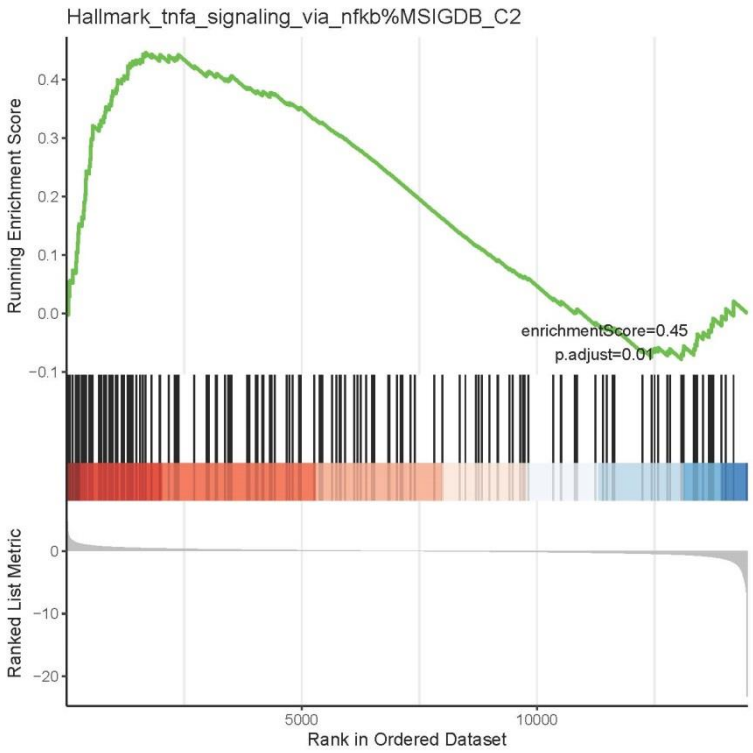


Figure 4.4.6.5 Up regulation of "Hallmark_tnfa_signaling_via_nfkb" in the IFN-γ treated *LRRK2* deficient macrophages

(Gene set "Hallmark_tnfa_signaling_via_nfkb" was from "C2" collection of Molecular Signatures Database (MSIGDB))

4.4.7 *LRRK2* deficiency leads to a heightened inflammatory response in LPS treated macrophages via interferon response

Based on theachieved RNA sequencing data from wild type macrophages, in addition to IFN-γ, it was reasonableto explore whether *LRRK2* deficiency made any differences in macrophages response to LPS. Hence, *LRRK2* deficient macrophages were treated by 1 ug/ml LPS, with the same time course (6 hours, 12 hours and 24 hours) as in Chapter 3, and samples were collected for bulk RNA-seq. Three biological replicates were prepared for each time point (lrrk2ko_lps_6h_rp1-3, lrrk2ko_lps_12h_rp1-3, lrrk2ko_lps_24h_rp1-3). After calculating the correlation distance between samples, none of possible outliers were identified (Figure 4.4.7.1). Moreover, from the dendrogram, *LRRK2* deficient macrophages were found to cluster more closely with themselves, nor mix with the corresponding wild type macrophages, which meant there might be some significant differences between the LPS treated wild type and *LRRK2* deficient cells.

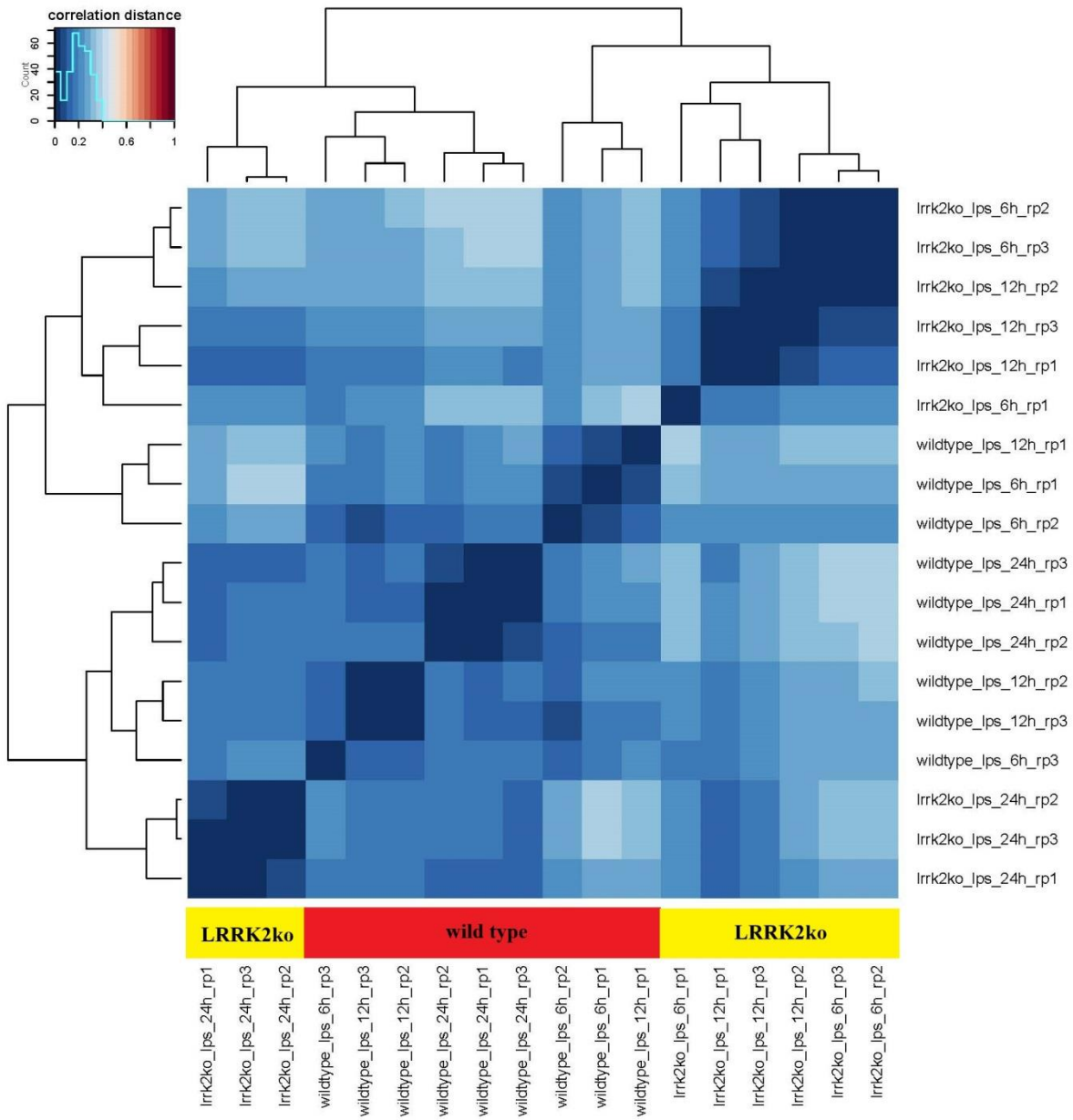


Figure 4.4.7.1 Correlation distance of wild type and *LRRK2* deficient macrophages with 1ug/ml LPS stimulation

Next, DGE analysis was done between all the *LRRK2* deficient and wild type macrophages with LPS treatment, and identified 339 up and 111 down regulated genes by the criteria "p.adj < 0.01" and "fold change>2" (Figure 4.4.7.2), which meant there might be

some differences between the LPS treated *LRRK2* deficient macrophages. Those differences were interpreted by ORA and GSEA. In the ORA results, up regulated genes in LPS treated *LRRK2* deficient macrophages were found to significantly enrich in terms associated with response to interferons (Figure 4.4.7.3). Meanwhile, in GSEA results, the significant up regulations of those terms were also found (Figure 4.4.7.4). As response to interferons directly involved in the cellular inflammatory response, it was possible that the inflammatory response in LPS treated macrophages was affected by *LRRK2* deficiency.

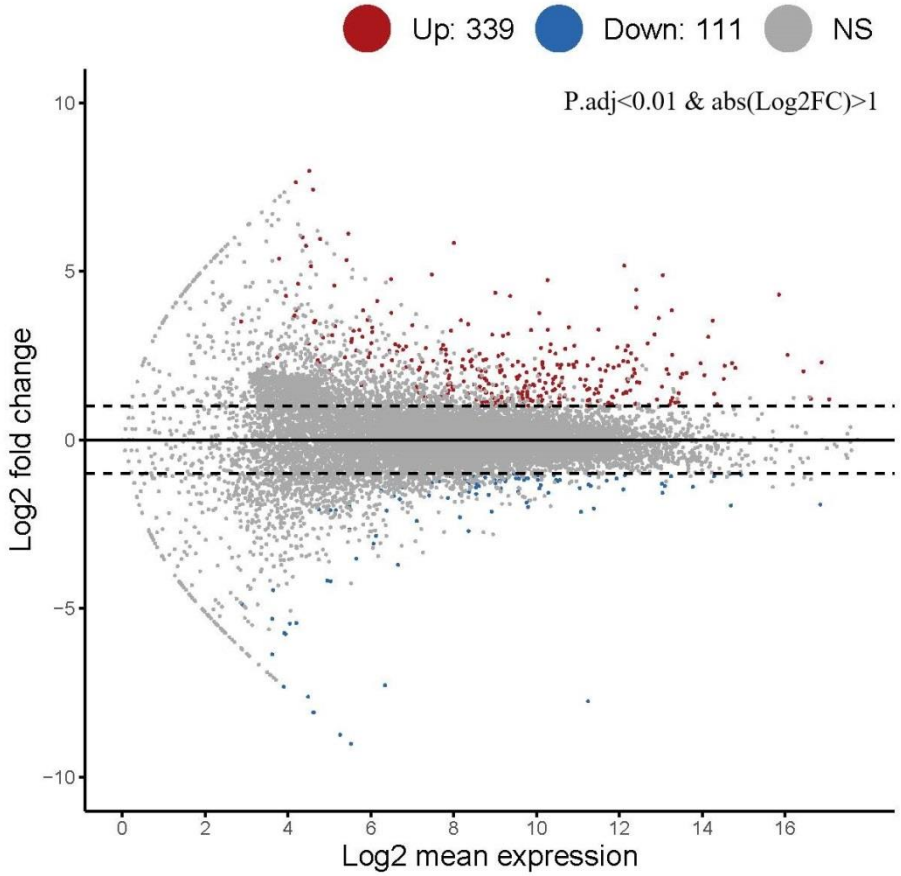


Figure 4.4.7.2 DE genes identified from the comparison between LPS treated *LRRK2* deficient and wild type macrophages

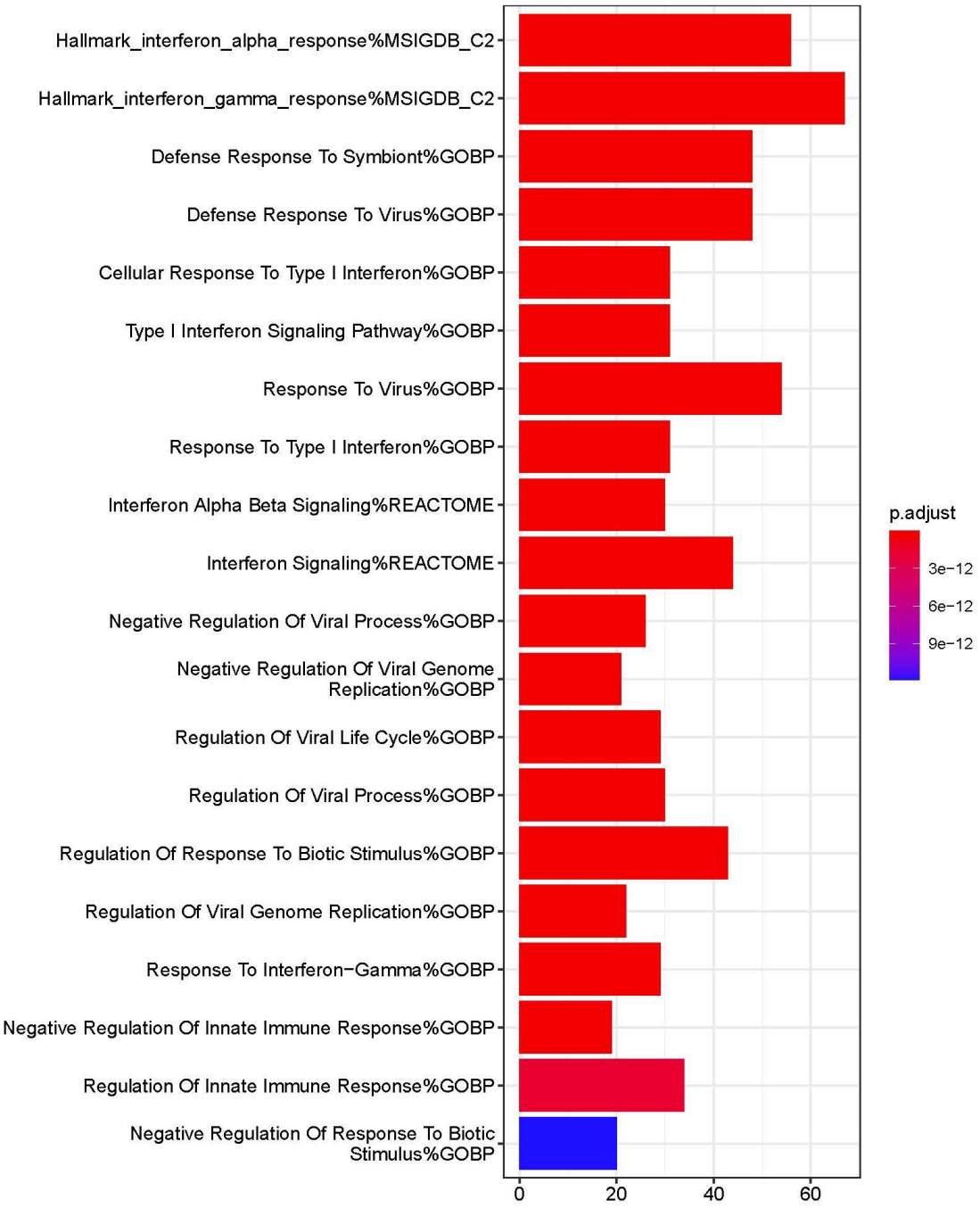


Figure 4.4.7.3 ORA results of up regulated genes in LPS treated *LRRK2* deficient macrophages

(The bar plot displayed the top 20 enriched terms with the lowest adjusted p-value.)

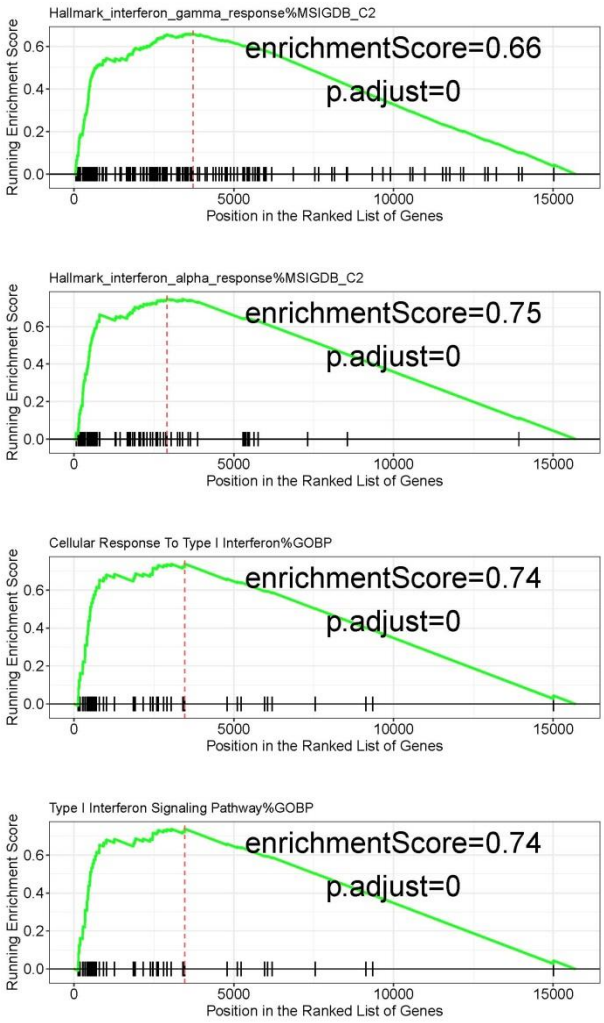


Figure 4.4.7.4 Up regulations of the terms associated with response to interferons in LPS treated *LRRK2* deficient macrophages (p.adjust ≤ 0.01)

("Hallmark_interferon_gamma_response" (MSIGDB_C2);

"Hallmark_interferon_alpha_response" (MSIGDB_C2);

"Cellular Response To Type I Interferon" (GO: 0071357);

"Type I Interferon Signaling Pathway" (GO: 0060337))

Expressions of the genes which could reflect the transfer of sub stages in LPS treated macrophages (*CCL20*, *CCR7*, *IL6*, *CSF2*, *IL1A*) were viewed (Figure 4.4.7.5), and those genes' expressions indicated the sub stages' transfer might be affected by the *LRRK2* deficiency.

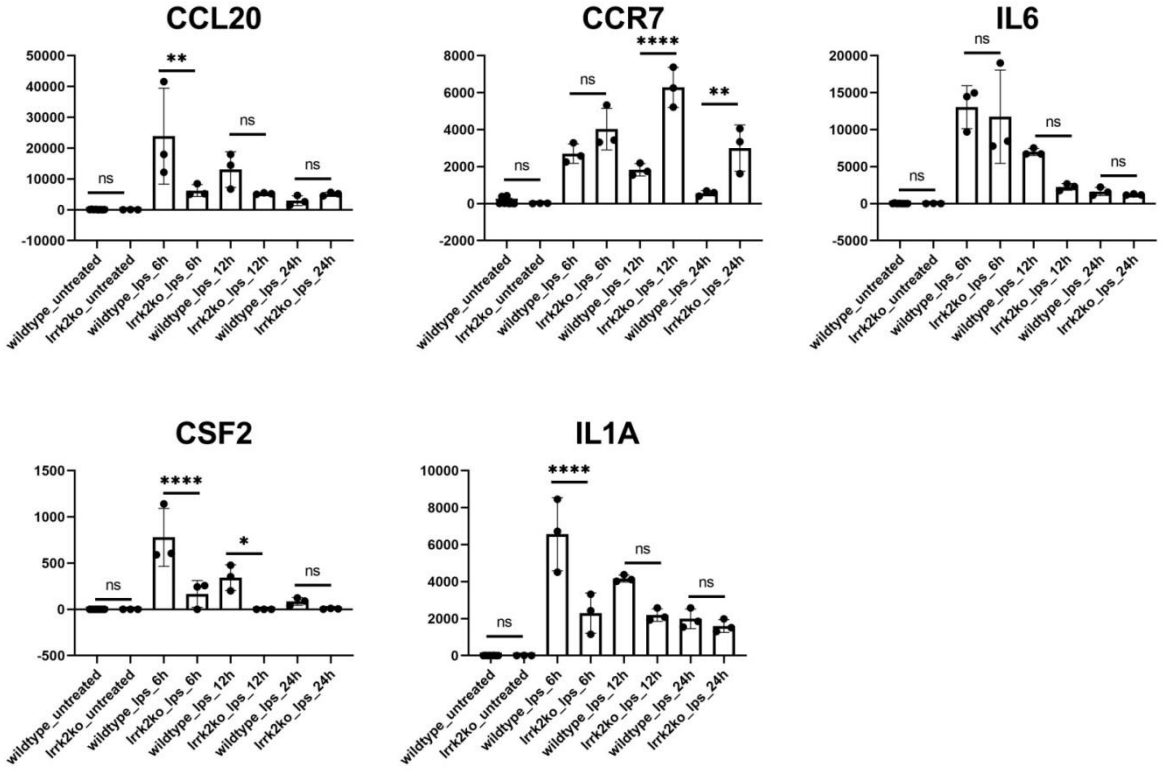


Figure 4.4.7.5 Expressions of the genes which could reflect the transfer of sub stages in LPS treated wild type and *LRRK2* deficient macrophages

(wildtype_untreated: wild type macrophages without any treatments
wildtype_lps_6/12/24h: wild type macrophages with 6/12/24 hours 1 μg/ml LPS treatment;
lrrk2ko_untreated: *LRRK2* deficient macrophages without any treatments;
lrrk2ko_lps_6/12/24h: *LRRK2* deficient macrophages with 6/12/24 hours 1 μg/ml LPS treatment)

Next, the macrophages after 24 hours of LPS treatment were further studied, as they could directly reflect the results of sub stages' of LPS treated macrophages. After DGE analysis, more DE genes were identified between *LRRK2* deficient and wild type macrophages (561 up regulated and 361 down regulated, by the criteria "p.adj < 0.01" and "fold change>2") (Figure 4.4.7.6). This was more than the time points were merged to measure DGEs. In the ORA and GSEA results, besides enrichments of response to interferons, significant up regulations of inflammatory response with TNFα signaling via NF-kB pathway were identified (Figure 4.4.7.7 and 4.4.7.8). This means *LRRK2* deficiency can lead to a heightened inflammatory response in LPS treated macrophages via up regulations of interferon response.

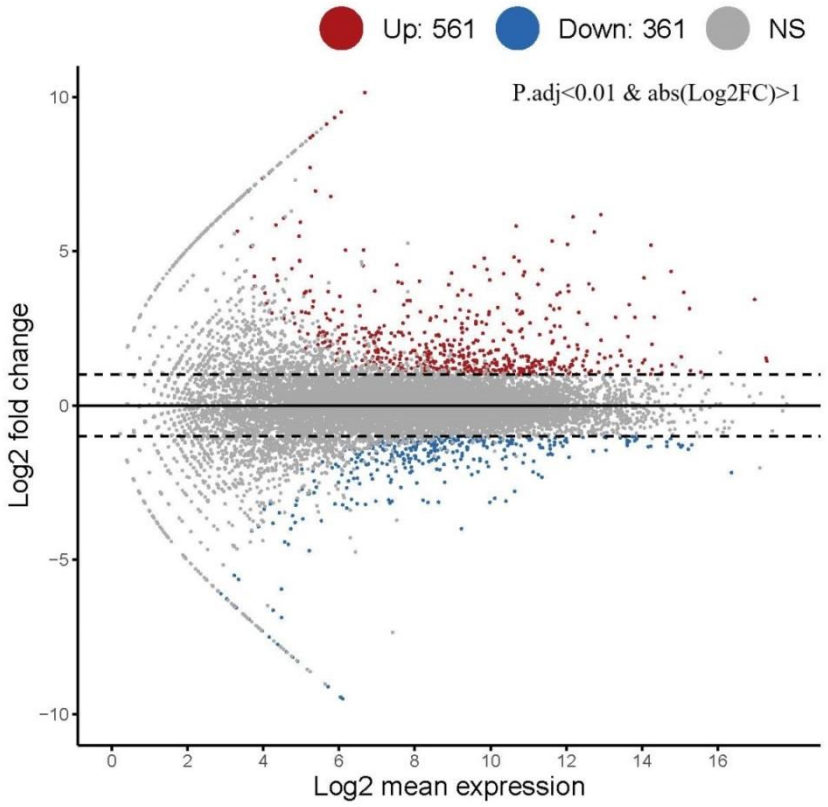


Figure 4.4.7.6 DE genes identified from the comparison between *LRRK2* deficient and wild type macrophages with 24 hours LPS treatment

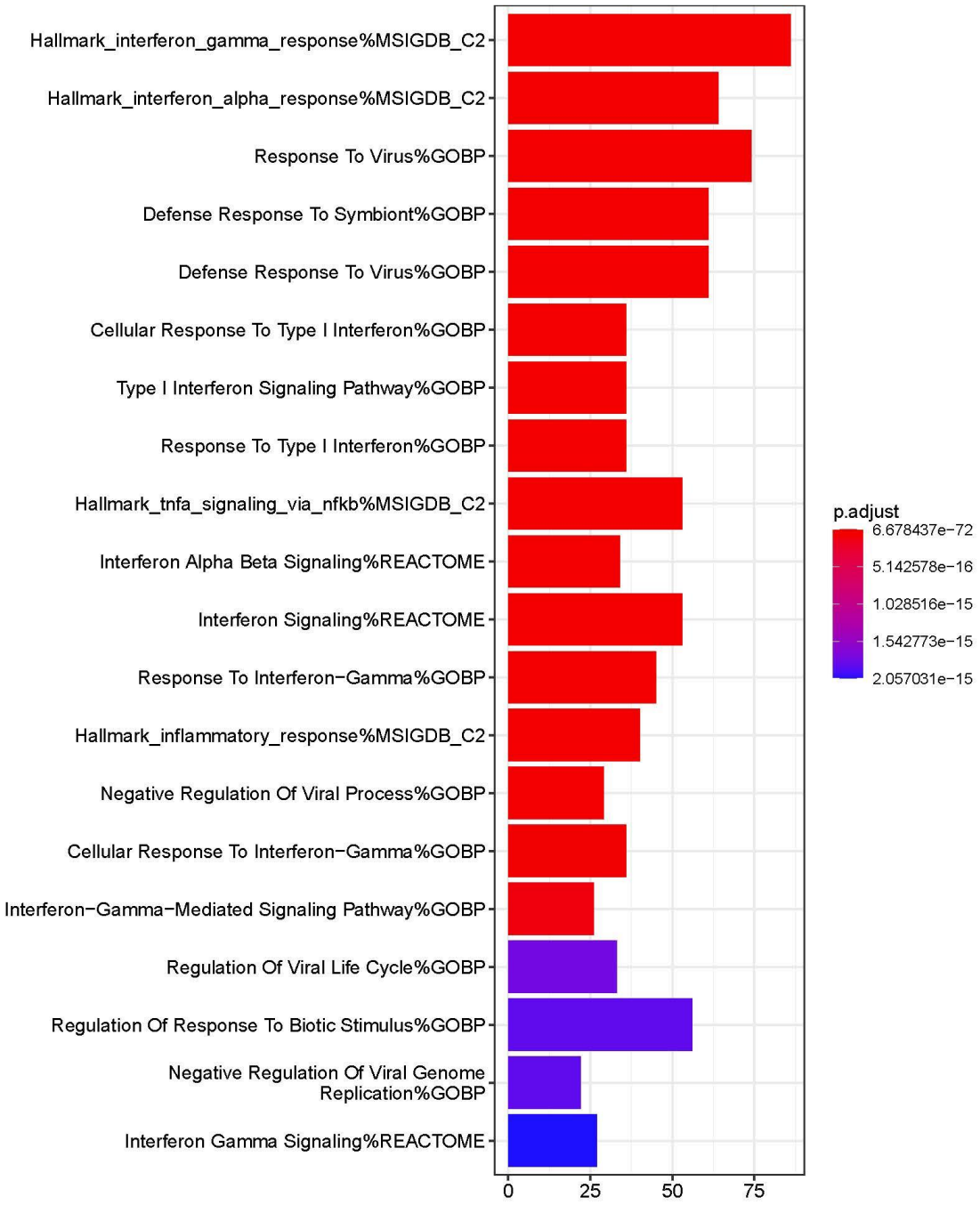


Figure 4.4.7.7 ORA results of up regulated genes in *LRRK2* deficient macrophages with 24 hours LPS treatment

(The bar plot displayed the top 20 enriched terms with the lowest adjusted p-value.)

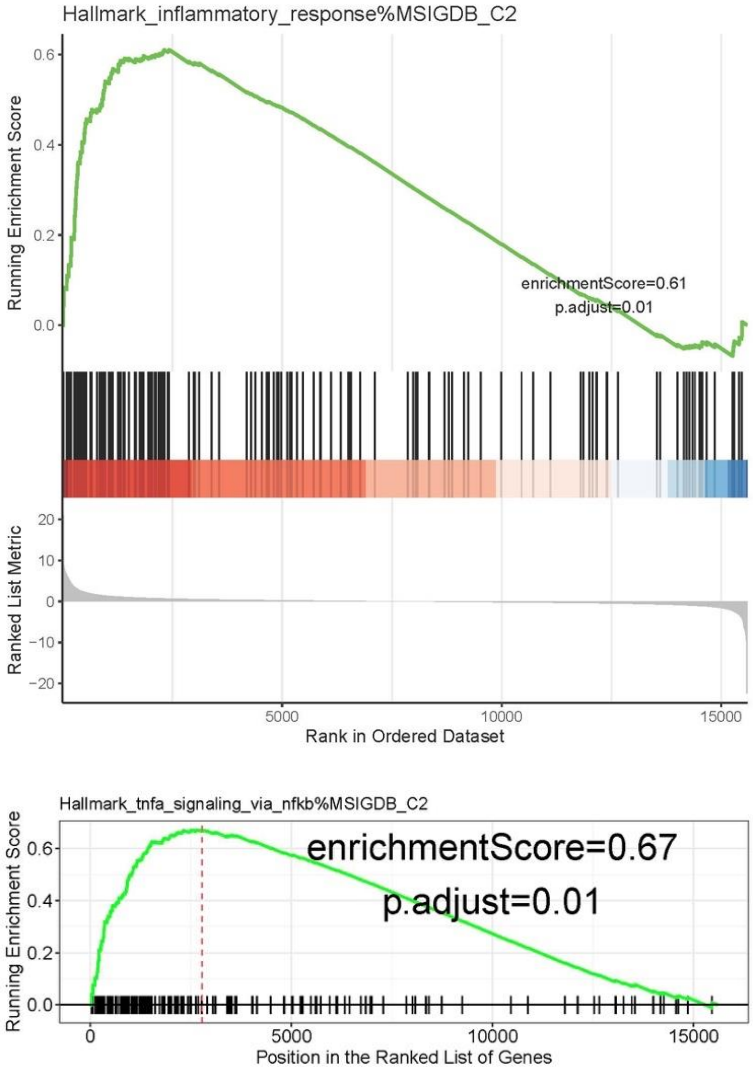


Figure 4.4.7.8 Up regulations of "Hallmark_inflammatory_response" and "Hallmark_tnfa_signaling_via_nfkb" in the *LRRK2* deficient macrophages with 24 hours LPS treatement

(Gene sets "Hallmark_inflammatory_response" and "Hallmark_tnfa_signaling_via_nfkb" was from "C2" collection of Molecular Signatures Database (MSIGDB))

4.5 Discussion

The aim of this chapter was to explore the possible inter-relationships between *NOD2* and *LRRK2*, which was the key susceptible gene for CD and PD respectively, by *in vitro* models. From this study, hopefully new insights into the shared pathogenesis between CD and PD can be generated, as well as why a key determinant of PD, *LRRK2*, can affect intestinal immune homeostasis.

For the aim, firstly, by retrieving the current knowledge, a possible link between *NOD2* and *LRRK2* was posited. STRING database (Szklarczyk et al., 2021) helped to recognize the co-expression of those two genes might exist in *Homo sapiens* more possibly comparing with any other organisms. Meanwhile, a review of the literature about the relationship between *NOD2* and *LRRK2* suggested it had not been reported before. Next, comparison between the expression of those two genes in the RNA sequencing data from chapter 3 was done. Co-expression of *NOD2* and *LRRK2* was particularly significant in IFN-γ treated macrophages, as increase of *LRRK2* and *NOD2* accompanied with each other. Additionally, the loss of *NOD2* significantly decreased *LRRK2* expression in those IFN-γ treated macrophages, which was not caused by the possible off-targets during the gene editing, and specific in IFN-γ treated macrophages among the current dataset. This indicated IFN-γ treated macrophages might be a good candidate for exploring possible relationships between *NOD2* and *LRRK2*, and the effects on *LRRK2* expression in those IFN-γ treated *NOD2* deficient macrophages also made us wonder whether it involved in the abnormal IFN-γ response which had beenobserved in *NOD2* deficient macrophages (section 3.4.4).

To clarify this issue, *in vitro* differentiated macrophages from the *LRRK2* knockout hiPSCs were generated from the same cell line as *NOD2* knockout, and hence had the same genetic background. IFN-γ was used to stimulate those *LRRK2* deficient macrophages. RNA

sequencing identified some biological changes, however, overall a much fewer number of DE genes were identified comparing with what had been found from *NOD2* deficient macrophages in section 3.4.4. This suggested loss of *LRRK2* had only relatively small effects on IFN-γ response in macrophages, and suggested *LRRK2* was not involved in the abnormal IFN-γ response which was observed in *NOD2* deficient macrophages (section 3.4.4). This was further verified by analysis of the overlapped down regulated genes between IFN-γ treated *NOD2* and *LRRK2* deficient macrophages, as ORA of those genes did not significantly enrich terms related to IFN-γ response.

Although *LRRK2* deficiency only made relatively small effects on the IFN-γ response in macrophages, it was still attractive to know what those effects were exactly. Through the aid of GSEA between IFN-γ treated *LRRK2* deficient macrophages and wild type reference, significantly enriched terms related to down regulation of oxidative phosphorylation were identified, with changes on mitochondria and cellular energy supply chain. Up regulation of TNFα signaling via the NF-kB pathway was identified at the same time. This suggested the small effects caused by *LRRK2* deficiency in IFN-γ treated macrophages were down regulation of oxidative phosphorylation via mitochondria, and those effects led to more inflammation in the cells. Additionally, from these generated RNA-seq data, the relationship between *NOD2* and *LRRK2* in those IFN-γ treated macrophages was found to be non-bidirectional, loss of *NOD2* could affect *LRRK2* expression, but loss of *LRRK2* had almost no influence on *NOD2* expression.

The experiments also explored whether *LRRK2* deficiency could made any differences in macrophage response to LPS. In the experiments, *LRRK2* deficiency was found to lead to some differences, which could be partially recognized from dendrograms of the samples (Figure 4.4.7.1). Further interpretation on those differences revealed that up regulation of

interferon responses, which were associated with more severe inflammation among *in vitro* studies. Sub stages of LPS mediated pro-inflammatory response in macrophages were found to be possibly affected by *LRRK2* deficiency by comparing expressions of the marker genes with wild type reference. This implied that taking into account treatment time might help to reveal differences. Consequently, only the 24 hours post-stimulation time point was considered. Those macrophages at the start of "resolution phase" in LPS mediated pro-inflammatory response were compared. Not surprisingly, analysis of those cells directly identified the up regulation of an inflammatory response with TNFα signaling via NF-kB pathway in the *LRRK2* deficient macrophages, which suggests *LRRK2* deficiency could contribute to LPS-dependent inflammatory responses in macrophages. This suggested *LRRK2* deficiency could lead to a heightened inflammatory response in LPS treated macrophages via effects on interferon response.

As mentioned in section 1.4 and 4.1, current knowledge suggests a relationship between Crohn's disease and Parkinson's disease, and hypothesized that *LRRK2* might play an important role in the shared pathogenesis (Herrick & Tansey, 2021; Ikezu et al., 2020; Rolli-Derkinderen et al., 2020). However, there is still only circumstantial evidence at the molecular level that can help to explain this association. Through the aid of the isogenic *in vitro* macrophage models, the correlations between *NOD2* and *LRRK2* genes were systematically explored (two significantly associated risk factor genes for CD and PD) in the hiPSC-derived macrophages after IFN-γ treatment, and identified a one-way influence from *NOD2* to *LRRK2*, which corresponded to the progressive process from CD to PD. Interestingly, the reverse situation was not true, and *LRRK2* appeared to have little effect on *NOD2*, and the responses were unrelated. This relationship could be explored further by co-culturing the hiPSC-derived macrophages with neurons to address the role of microglia

(macrophage-like cells) in the brain. It was also possible to adopt protocols to further differentiate macrophages into microglia-like cells *in vitro*, as microglia exist in the brain as tissue-resident macrophages (Claes, Van den Daele, & Verfaillie, 2018; Haenseler et al., 2017).

To better understand the relationship, it is necessary to know the details of how *LRRK2* deficiency affects IFN-γ treated macrophages. IFN-γ can significantly up regulate *LRRK2* expression in macrophages, leading to the hypothesis that the gene may play an important role in the IFN-γ-dependent inflammatory response (Gardet et al., 2010). However, recent studies have indicated that *LRRK2* might not play such an important role, as it was found to have no effect on any primary functions of macrophages (e.g. phagocytosis) (H. Lee et al., 2020). TheRNA-seq based approach, which looks more at regulatory processes rather than phenotypic outcome provides more direct evidence to refute this hypothesis, as *LRRK2* deficiency had only a modest effect on the IFN-γ response. It seemed that the loss of *LRRK2* was mainly related to mitochondrial functions, which led to down regulation of oxidative phosphorylation and affected the inflammatory response. This effect was not huge, and required accumulation for some time to observe a clear influence on *in vivo* homeostasis. Moreover, *LRRK2* deficiency was found to be able to up regulate inflammatory responses in macrophages treated by LPS (a bacterial component), and the changes were also related to distinct responses to interferon.

According to the evidence from this chapter, it was reasonable to generate new ideas on the links between the pathogenesis in CD and PD. Those links could be explained by *LRRK2* deficiency caused dysregulations in activated macrophages. It is possible either for CD or PD, the dysregulations make slight effects on pro-inflammatory response in macrophages/microglia, via functioning on mitochondria and interferon response. Those

effects accumulate long-term, then cause disorders leading to acute disease. Theobserved unidirectional progress from CD to PD could be explained by the inter-relationships between their key susceptible gene, *NOD2* and *LRRK2* respectively. In macrophages/microglia, if *NOD2* is mutated (relatively common in CD patients), *LRRK2* expression will also be affected, and thus lead to an accumulation of the effects on pro-inflammatory responses in the brain. However, *LRRK2* mutations could hardly make differences on *NOD2* expression in macrophages/microglia.

CHAPTER 5

SINGLE CELL ANALYSIS AND ORGANOIDS GENERATION FROM SURGICAL RESECTIONS HELP REVEAL CHARACTERISTICS OF MONOCYTES/MACROPHAGES FROM CD PATIENTS

5.1 Background

With the improved understanding on CD, more and more evidence indicates that personal difference, such as genetic background, is very important for the study of CD. For instance, the same risk factors have distinct influence on different populations, as described in section 1.2. Indeed, a lot of CD therapies, even the most effective ones, for example Anti-TNF, do not achieve the same benefit in different patients, and the diversity to the specific therapies has been verified to be related with cellular compositions of inflamed tissues from CD patients (Martin et al., 2019; Nayar et al., 2021). Traditional methods which merge the samples from individuals to isolate a general understanding of the disease is not ideal for CD, as pathogenesis for CD may be diverse for different patients. In this situation, studying CD related inflammations in an isogenic background is helpful to better understand the disease.

Studying inflammation in CD without effects from genetic background is not easy. Unlike *in vitro* models in Chapter 3 and 4 which have matching negative controls (wild type), no negative reference can be found amongst CD patients, and the alternative is comparing tissues from the same patient with distinct inflammatory severity. However, those research materials are extremely limited, as they can only be obtained from biopsies or surgical resections from CD patients. Normal (non-diseased) tissues are (of course), even harder to acquire, and in CD patients the comparison is instead between tissues with different levels of inflammation (by pathological examination). This has the benefit of the two samples having the same genotype, but this is also a disadvantage as any CD-associated mutations are present in cells from both tissues. There are further problems, these patient-tissue studies face two further obstacles: first, whilst experiments on those tissues can generate detailed data, they are extremely hard, often impossible, to repeat, as the same patient is rarely available for further study. Secondly, as the samples are both in inflammatory conditions from the same patients,

differences between them can be quite subtle, and hard to identify.

Single cell RNA sequencing (scRNA-seq), which is a natural extension of traditional bulk RNA sequencing, can be applied to resections and can overcome some of these obstacles. This technology identifies the heterogeneity among the cells, and is capable of revealing cell diversity by the generation of a large amount of data for many cells from a single sample (Aldridge & Teichmann, 2020; Stark et al., 2019; Tang et al., 2009). Additionally, as a high-resolution novel technology, scRNA-seq is sensitive to even relatively small biological changes in specific cell types, which is helpful in identifying imperceptible differences among the samples. Finally, as the above bulk RNA-seq data shows, the impact of a genetic change may not lead to the same effect in different cells. For example, loss of *NOD2* had a major impact on the IFN-γ response in macrophages, but it did not have any impact in intestinal organoids. Single cells RNA-seq can isolate individual cell types in complex human tissues and so identify the impact on specific cell types.

Since experiments on CD patients' resections are essentially a single shot for any one patient, it is necessary to find a solution which can maintain the characteristics of patients for a longer time to enable subsequent follow up experiments, or for deposit in biobanks to contribute to further understandings on disease. Intestinal organoids derived from adult stem cells can help on this issue, as described in section 1.7.

5.2 Aims

The aim of this chapter is to establish a pipeline for preparing scRNA-seq samples from CD patients' resections. Corresponding to the *in vitro* studies in Chapter 3 and 4, the target are the monocytes/macrophages in those samples, and the subtle differences between the cells that are associated with distinctly inflammatory severities and CD progression are expected to be isolated. This work will be used to determine the relevance of the findings from in vitro macrophages with the phenotype of macrophages in clinical samples.

5.3 Materials and Methods

5.3.1 Single cell RNA-seq of resections from CD patients

Resections from CD patients were collected immediately after the surgeries, then transferred to pathologists for examination and isolated into: severe inflamed ("stricture") and non-severe inflamed ("proximal") pieces for lab research. Recruitment of CD patients' resections was under the ethical approval of the "TRAFIC" project (Tissue remodeling and fibrostenosis in Crohn's disease - a study of molecular signals and biomarkers), which was issued on the 13/06/2019 by Institute of Translational Medicine, University of Birmingham, UK.

After arriving at the lab, resections were weighted first, and put into a 15 ml Falcon tube containing MACS Tissue Storage Solution (Miltenyi, #130-100-008) with 100 ug/ml Primocin™ (Invivogen, ant-pm). They were stored at 4 ℃ overnight.

In the next morning, resections were put into 60 mm TC-treated Culture Dish (Corning, #430166) with some 1x red blood cell lysis buffer (Biolegend, #420301) inside, and cut into 2-4 mm^2 pieces by sterile scalpel and forceps at room temperature. Fat in the pieces when they could be observed were dissected out and discarded in this step. All the processes were performed in the laminar flow hood.

Next, small pieces of resections in the dishes were transferred into gentleMACS C Tubes (Miltenyi, #130-096-334) which contained around 5 ml enzyme mix. The automatic dissociation program was applied to the sample preparation, with specific settings based on the tissue types. For the resections from CD patients, Multi Tissue Dissociation Kit 1 (Miltenyi, #130-110-201) was used, and the amount of enzymes in the experiments corresponded to the manual of the kit. In this experiment, enzyme mix in each gentleMACS C

Tube consisted of 4.7 ml serum-free RPMI 1640, 200 µl Enzyme D, 100 µl Enzyme R and 25 µl Enzyme A. The digestion was by gentleMACS Octo Dissociator with Heaters (Miltenyi, #130-096-427), with the ready-to-use gentleMACS Program "37C_h_TDK1".

After the program finished, C tubes containing the samples were detached from the machine, and stored at room temperature for 5 minutes to settle down the non-dissociated tissues. The supernatant was taken out, filtered by Falcon® 70 µm Cell Strainer (Corning, #352350), and spun down at 300 g, room temperature for 7 mins. Supernatant was discarded and the cells were re-suspended in 1ml 1X DPBS (without Calcium, Magnesium and Phenol Red). 10 µl cell suspension was stained by the equal volume of Trypan blue for calculating the cell viability by BIO-RAD TC20^{TM} Automated Cell Counter. Only the samples which had over 80% live cells could proceeded to the single cell RNA-seq.

For the samples which could not meet the requirements of live cell concentration, fluorescence activated cell sorting (FACS) was used to enrich for live cells. Cells were stained by 1ug/ml propidium iodide (PI) at room temperature for 5 mins, and the large nozzle was used in cell sorting. PI negative cells were collected and re-suspended in 1ml 1X DPBS (without Calcium, Magnesium and Phenol Red), then 10 µl cell suspension was taken out for calculating the cell viability. For maximum recovery of the cells, all the tubes used in the enrichment steps were FBS-coated.

Samples which had over 80% live cell proportion were loaded onto a 10X Genomics Chromium system, and subsequent steps were processed using a Chromium Single Cell 3′ GEM, Library & Gel Bead Kit v3 (10X Genomics, PN-1000092). 5,000 cells per sample were targeted for recovery. Quality of reversed transcribed cDNA was checked by Agilent High Sensitivity D5000 ScreenTape, and then the cDNA were used for libraries construction for sequencing. Samples were 150bp pair-end sequenced by Illumina Miseq platform, with the

sequencing depth 20,000 reads per cell.

BCL files generated from Illumina sequencers were demultiplexed into FASTQ format by Cell Ranger (version 3.0.2) (Zheng et al., 2017). FASTQ files were aligned to GRCh38 ("refdata-cellranger-GRCh38-3.0.0" from Cell Ranger) and generated into the single cell gene expression matrix (barcode counting, UMI counting) by the same software. "raw_feature_bc_matrix" in the results were used for the downstream analysis. R package "Seurat" (version 4.0.2) was used for identifying the monocytes/macrophages from resections' data, and drawing the related plots (Butler, Hoffman, Smibert, Papalexi, & Satija, 2018; Hao et al., 2021; Satija, Farrell, Gennert, Schier, & Regev, 2015; Stuart et al., 2019). Data filtering was by the criteria "nFeature_RNA>50 & nFeature_RNA<6000 & percent.mt<20", to remove the possibly empty droplet and doublet with the low quality/dying cells. "SCT" method was used for data normalization (Hafemeister & Satija, 2019), "CCA" method was used for data integration, "umap-learn" method was used for dimension reduction, and "DESeq2" method was used for identifying the variable genes between two conditions. For cell type annotations, R package "SingleR" (version 1.4.1) (Aran et al., 2019) was used with the "HumanPrimaryCellAtlasData" as the reference. Volcano plot was drawn by R package "EnhancedVolcano" (version 1.8.0) (Blighe, Rana, & Lewis, 2021). GSEA was performed by R package "clusterProfiler" (version 3.18.1) with the same gene set database as the bulk RNA-seq. GSEA plots were drawn by R package "enrichplot" (version 1.10.2) (Wu et al., 2021; G. Yu, 2021; G. Yu et al., 2012).

Other single cell RNA-seq data from CD patients' resections was downloaded from GEO: GSE134809 (Martin et al., 2019). The pipeline for the analysis was similar to the analysis pipeline used here, only the filtering criteria used here was different from the publication "nCount_RNA>800 & percent.mt<25".

5.3.2 Generating intestinal organoids from CD patients' resections

The remaining cells after scRNA-seq were also used for generating intestinal organoids (Mahe et al., 2013; Sato et al., 2011). The cells were mixed with Corning® Matrigel® Growth Factor Reduced (GFR) Basement Membrane Matrix, Phenol Red-Free, and embedded in the 24-well plates with 50 μl mixture per dome, with around 30mins incubation at 37 ℃ to solidify the domes. 500 μl IntestiCult™ Organoid Growth Medium (Human) (STEMCELL Technologies, #06010) with 100 μg/ml Primocin™ per well was used to culture the domes. The domes were incubated at 37℃ with 5% CO2 and 95% humidity, with fully medium change without Primocin every three or four days. After one or two weeks, three dimensional structures could be observed in the successful wells that generated organoids, and those tissues could be passaged as hiPSC-derived intestinal organoids. For several rounds of passage, a resection derived intestinal organoids line was established, and was capable of being to be maintained (culture, passage, freeze and thaw) as hiPSC-derived intestinal organoids.

5.4 Results

5.4.1 Preparing samples for single cell RNA sequencing from CD resections

Weight of the CD resections from severe inflamed (“stricture”) and non-severe inflamed (“proximal”) regions (Figure 5.4.1.1) from the same patient were 0.9291 g and 0.7448 g respectively. The “stricture” sample was the one more analogous to the disease condition, and the “proximal” sample was the one more analogous to the healthy condition. However, separation of those two samples was only by pathological examination, and the samples were both from the same CD patient’s intestinal resections.

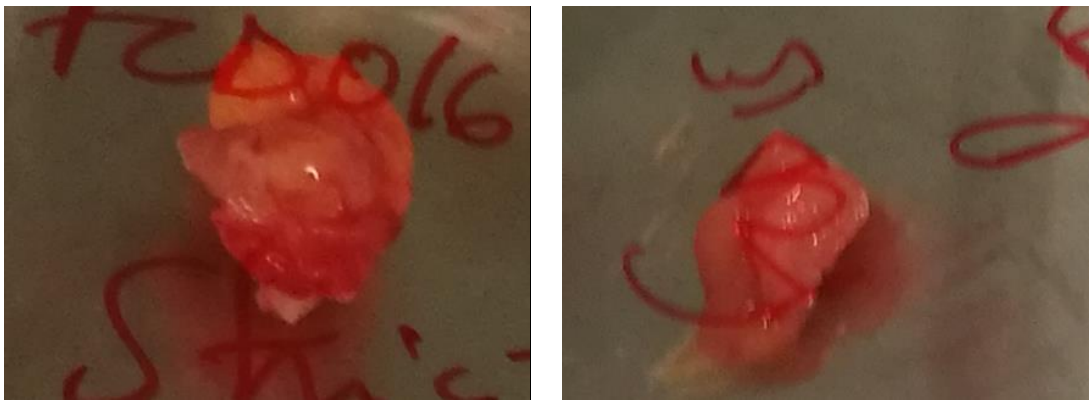


Figure 5.4.1.1 Surgical resections from the same CD patient for scRNA-seq
(left: “stricture” (severe inflammation) sample;
right: “proximal” (non-severe inflammation) sample)

After first round of digestion, cell viabilities were 68% and 65% for “stricture” and “proximal” samples respectively, which did not meet the requirements for single cell RNA-seq. In this situation, live cells were enriched using FACS. Meanwhile, debris in the cell suspension could be observed clearly in the samples (Figure 5.4.1.2).

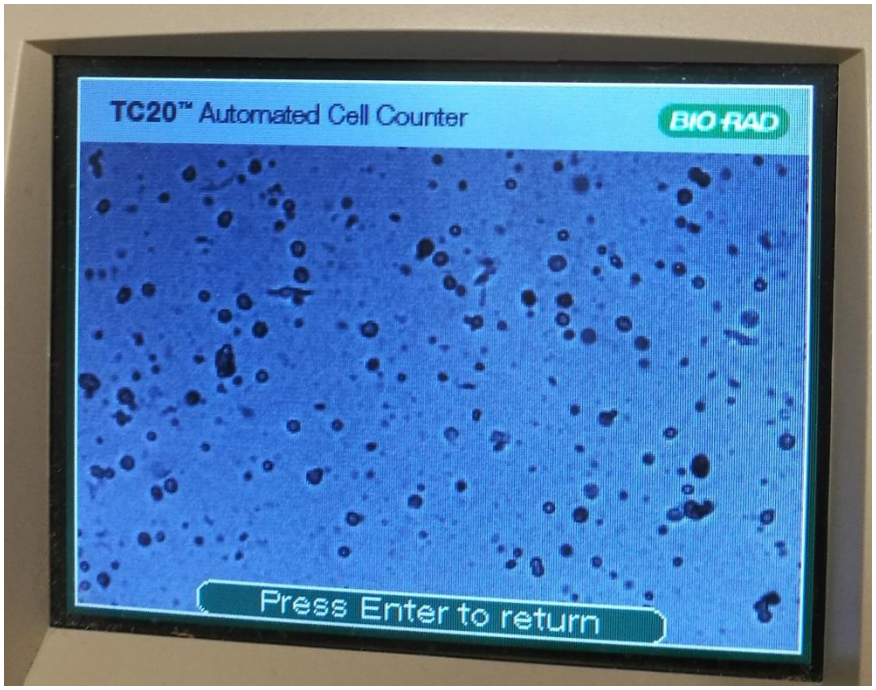


Figure 5.4.1.2 Cell counter Screen when identifying the cell viability of "stricture" sample after first round of digestion

With the help of FACS, cell viabilities of "stricture" and "proximal" samples increased to 88% and 90% respectively, which were qualified to be loaded onto the 10X machine for single cell RNA-seq. The total cell concentration were 1.2x10^6/ml and 2.0x10^6/ml respectively. Debris was removed in the cell suspension of both samples (Figure 5.4.1.3 and 5.4.1.4).

Qualities of cDNA were qualified for sequencing, and the cells were subjected to single cell RNA-seq data as section 5.3 described.

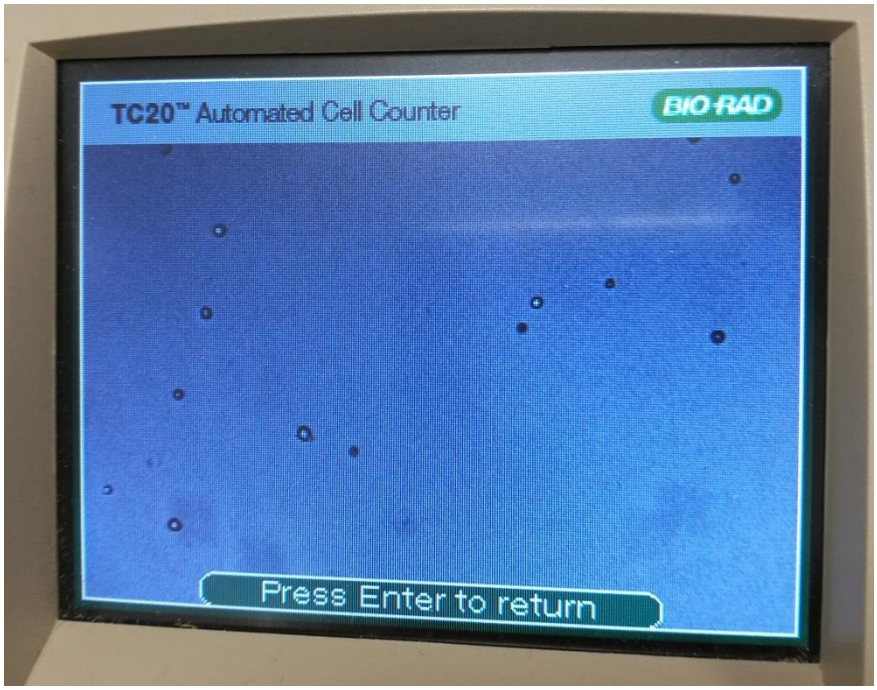


Figure 5.4.1.3 Cell counter Screen when identifying the cell viability of "stricture" sample after FACS

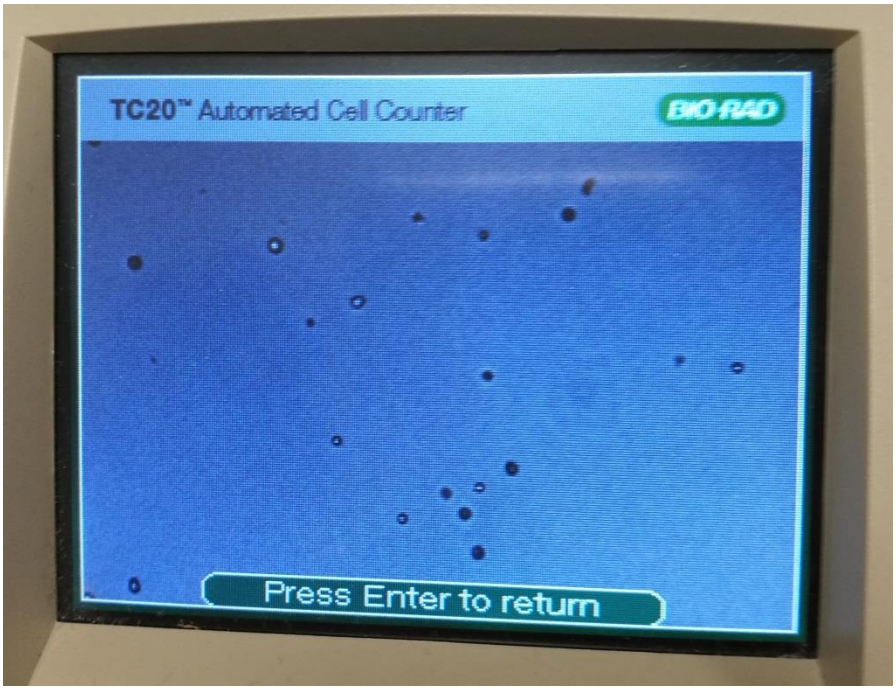


Figure 5.4.1.4 Cell counter Screen when identifying the cell viability of "proximal" sample after FACS

5.4.2 Identification of monocytes/macrophages from CD resections

Single cell RNA sequencing data from the CD resections were generated into the single cell gene expression matrix by Cell Ranger, and the QC figures showed read counts (nCount_RNA), number of genes (nFeature_RNA), and the percentage of reads that map to the mitochondrial genome (percent.mt) of the calling single cells (Figure 5.4.2.1 and Figure 5.4.2.2). Only the cells with “nFeature_RNA>50 & nFeature_RNA<6000 & percent.mt<20” were kept for downstream analysis. After data integration by “CCA” method, cells were projected into a UMAP, and using Louvain clustering (threshold=1.2) 23 clusters were identified (number 0-22) were identified from the CD resections (Figure 5.4.2.3).

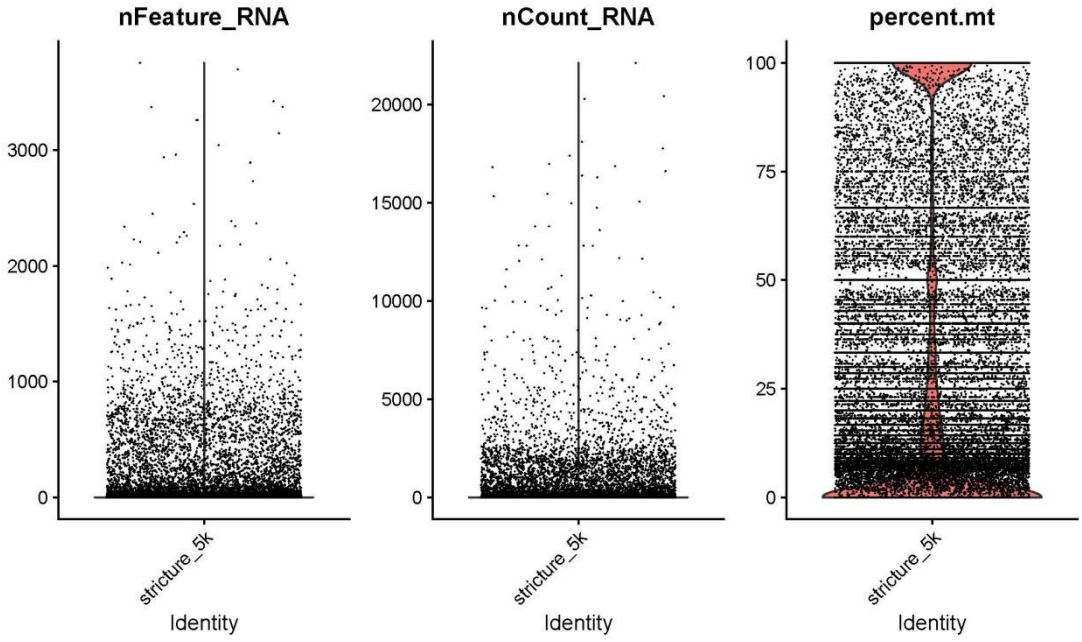


Figure 5.4.2.1 QC figure for "stricture" sample

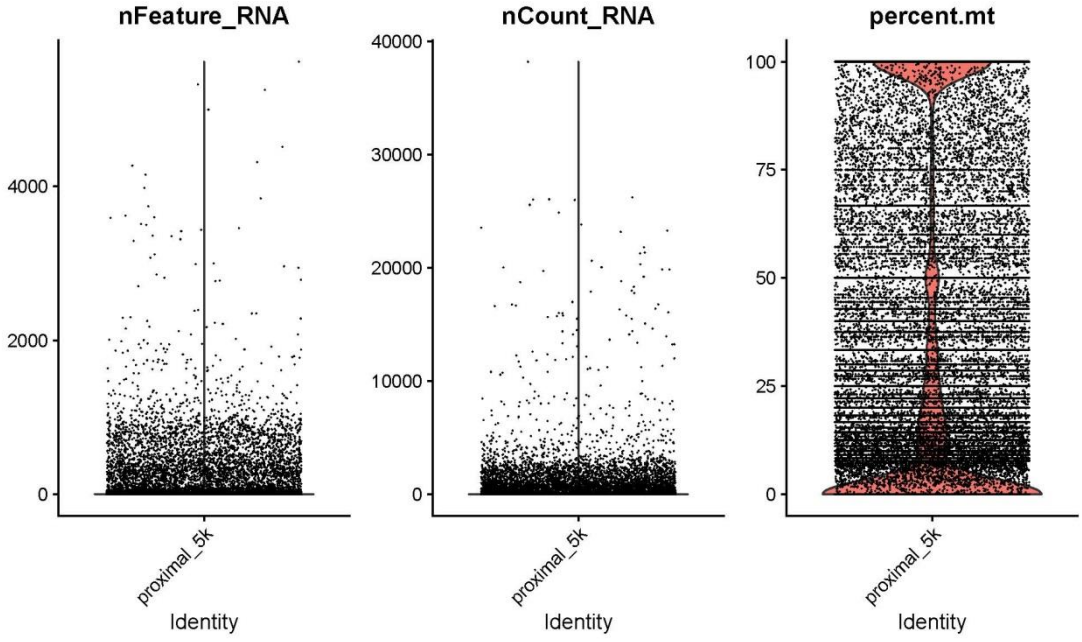


Figure 5.4.2.2 QC figure for "proximal" sample

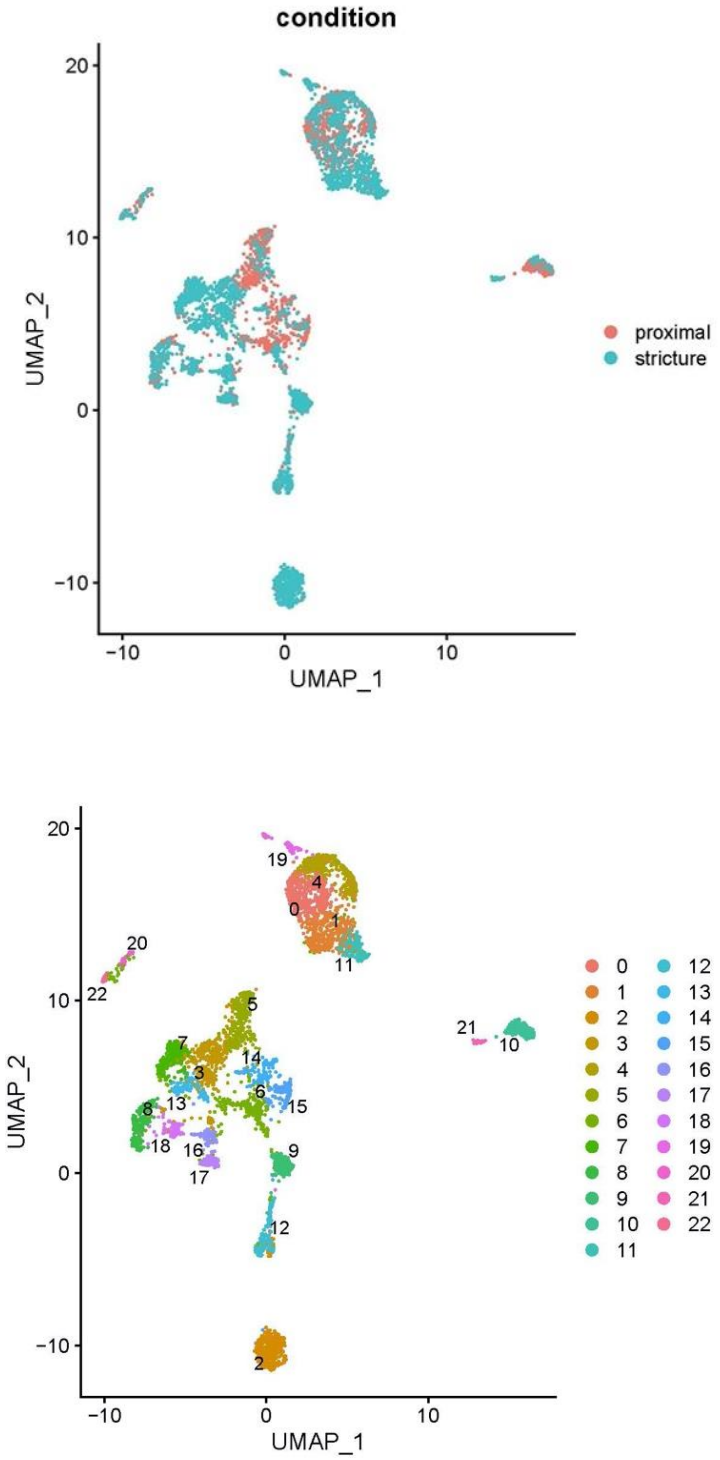


Figure 5.4.2.3 Clusters identified from CD resections

(The upper plot displayed the cells' distribution for each sample in the integrated data. As mentioned in section 5.3, "stricture" and "proximal" were the inflamed and uninflamed samples from the same CD patient; the lower plot displayed the identified clusters in the whole data)

Cell type annotations for each cluster were first by reference-based method, via R package "SingleR" (Aran et al., 2019), and cluster "12" was identified as monocytes/macrophages (Figure 5.4.2.4). Quality control of cluster "12" indicated that the monocytes/macrophages were highly expressive, which excluded possible effects from low-grade expression in the downstream analysis (Figure 5.4.2.5). Meanwhile, Expression levels of marker genes (*CD14*, *CD68*, *MRC1*) for monocytes/macrophages were also viewed, and the results indicated that those genes were highly expressed in cluster "12", compared with others (Figure 5.4.2.6). In this situation, cluster "12" was identified as monocytes/macrophages. the expression of the target genes previously identified in Chapter 3 and Chapter 4, *NOD2* and *LRRK2*, were compared, and their expression in "proximal" and "stricture" samples was both low, but were detected in cluster "12" (monocytes/macrophages) (Figure 5.4.2.7). The other clusters in the results were annotated as specific cell types based on the results from R package "SingleR" and the expression of some specific marker genes such as CD8A(Figure 5.4.2.8).

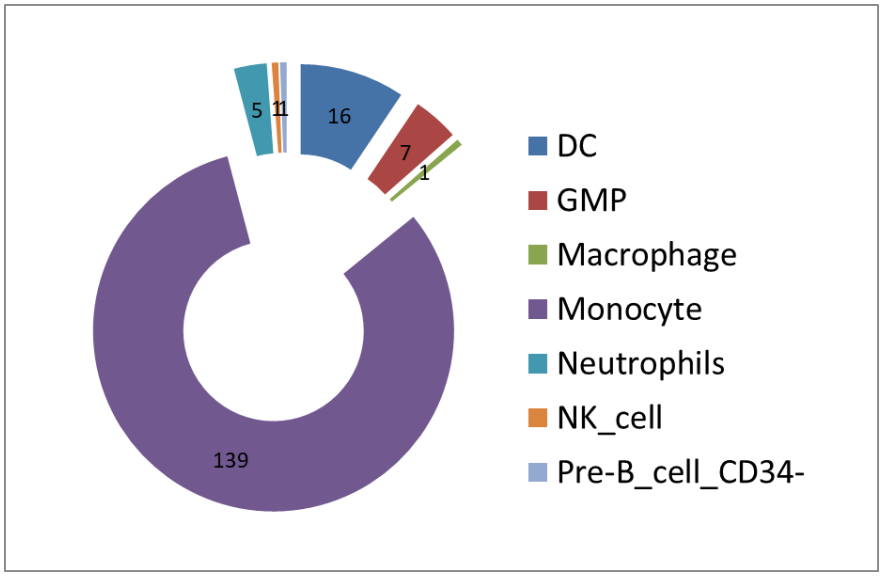


Figure 5.4.2.4 Identified cell types in cluster "12"

(The results were displayed by doughnut. Cell numbers of each type were marked on the ring.)

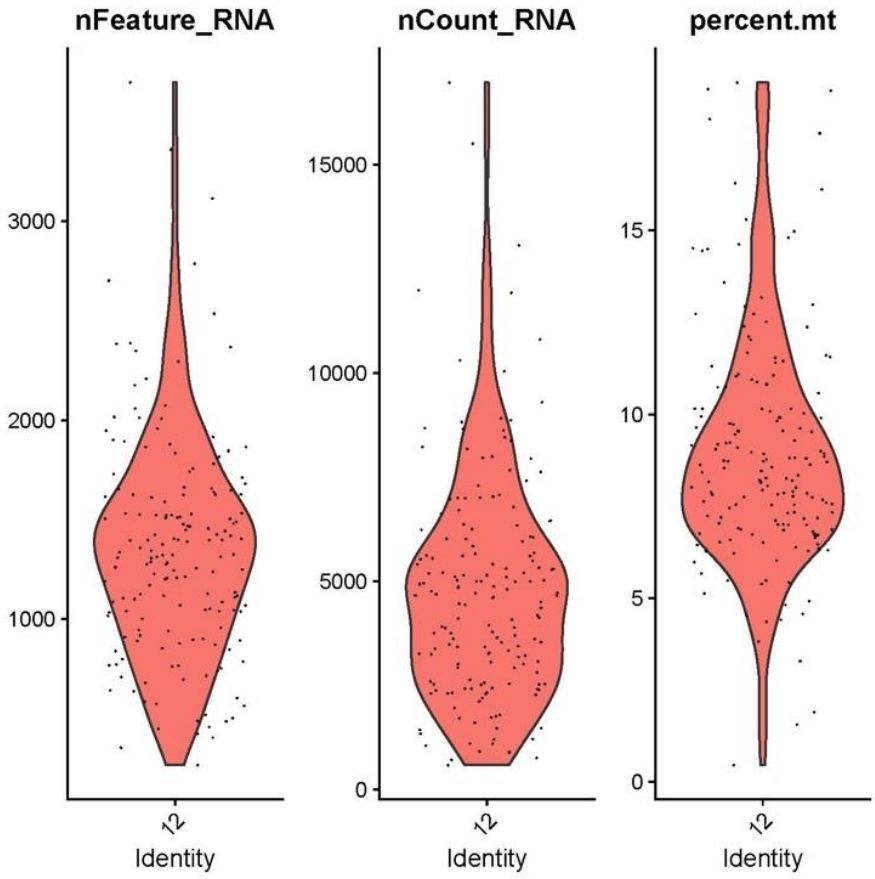


Figure 5.4.2.5 QC figure for the cells in cluster "12"

(Values of "nFeature_RNA" and "nCount_RNA" in the cells from cluster "12" were in a reasonable range, not too high or too low, with acceptable "percent.mt".)

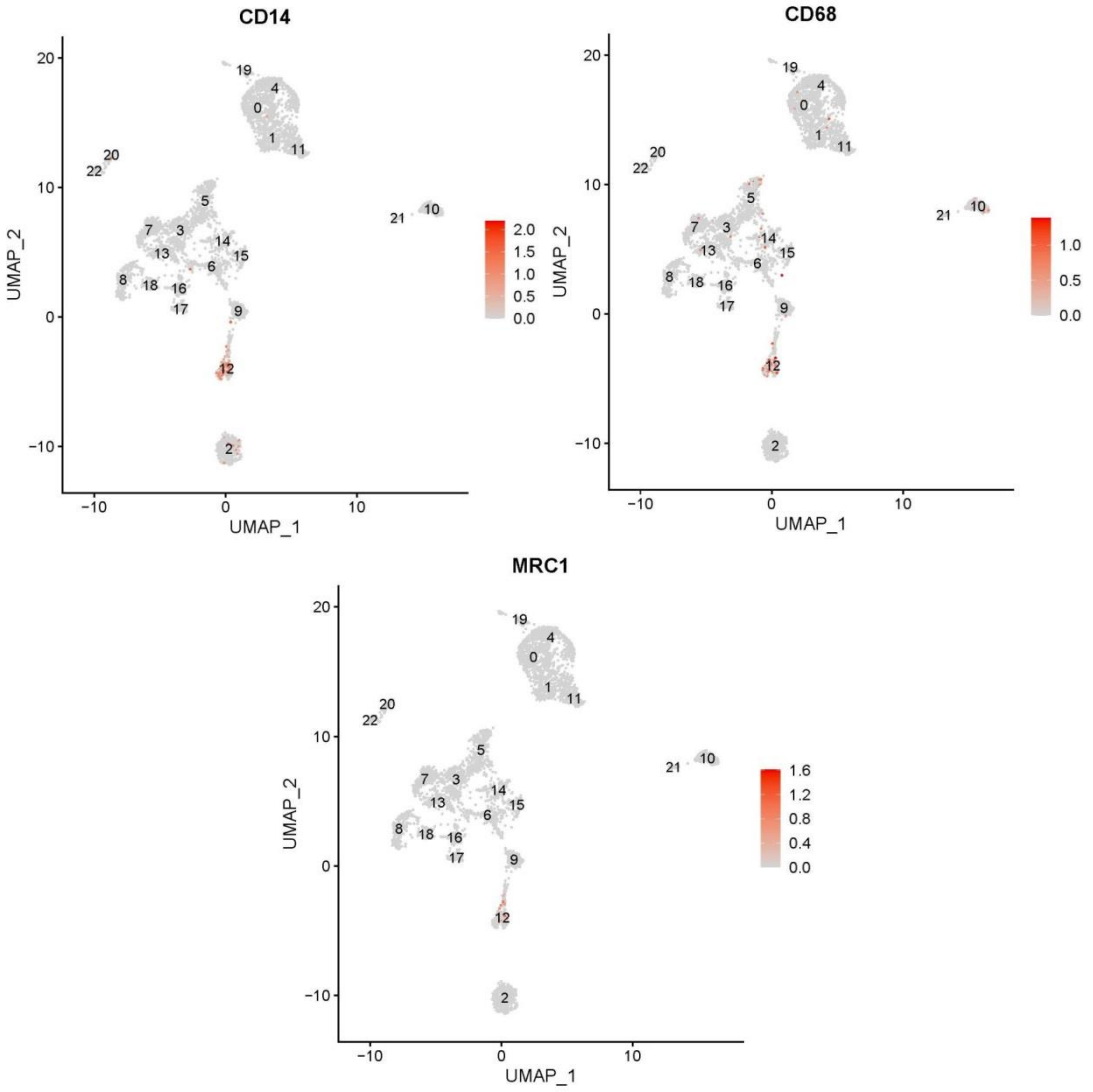


Figure 5.4.2.6 Expression levels of monocytes/macrophages marker genes (CD14, CD68, MRC1) in the resections' data

(log transferred data was used for drawing the plots)

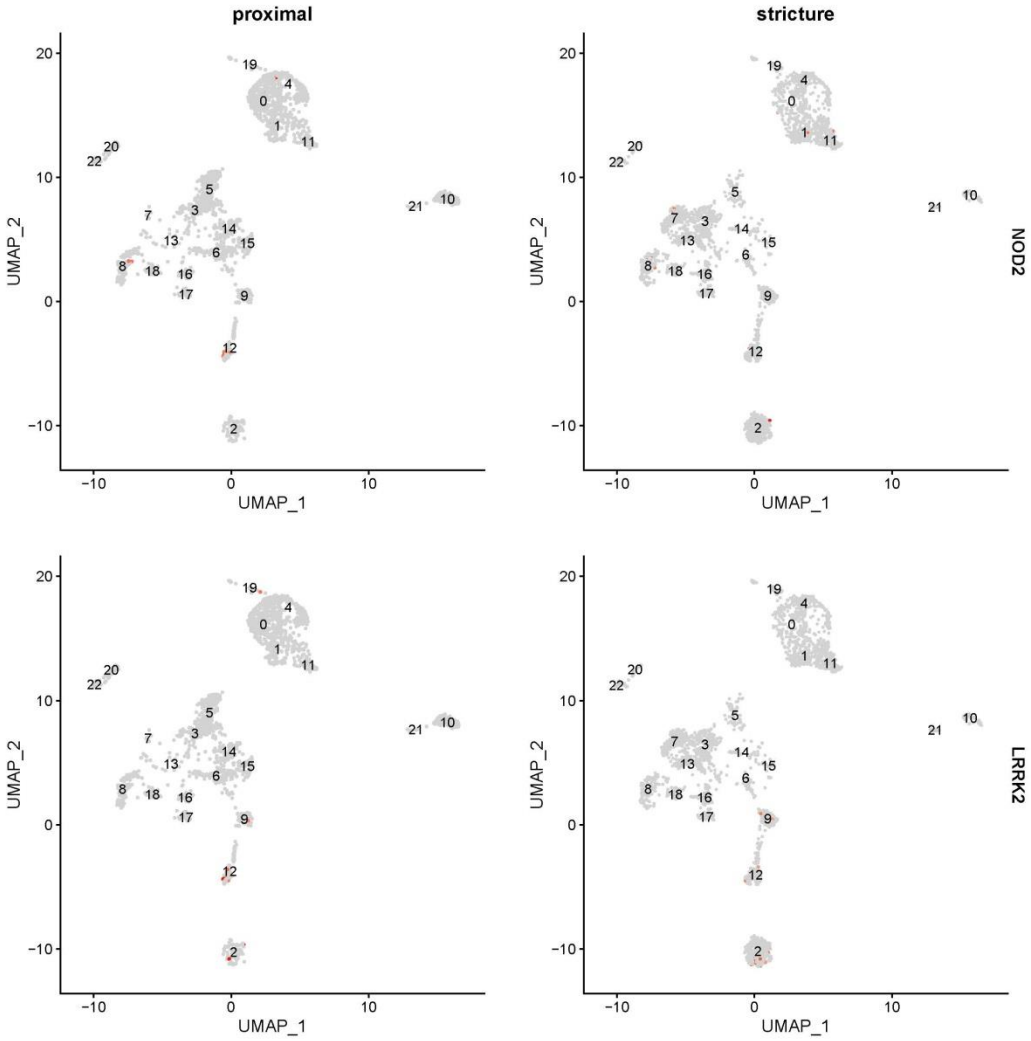


Figure 5.4.2.7 Expressions of *NOD2* and *LRRK2* in the "proximal" and "stricture" samples

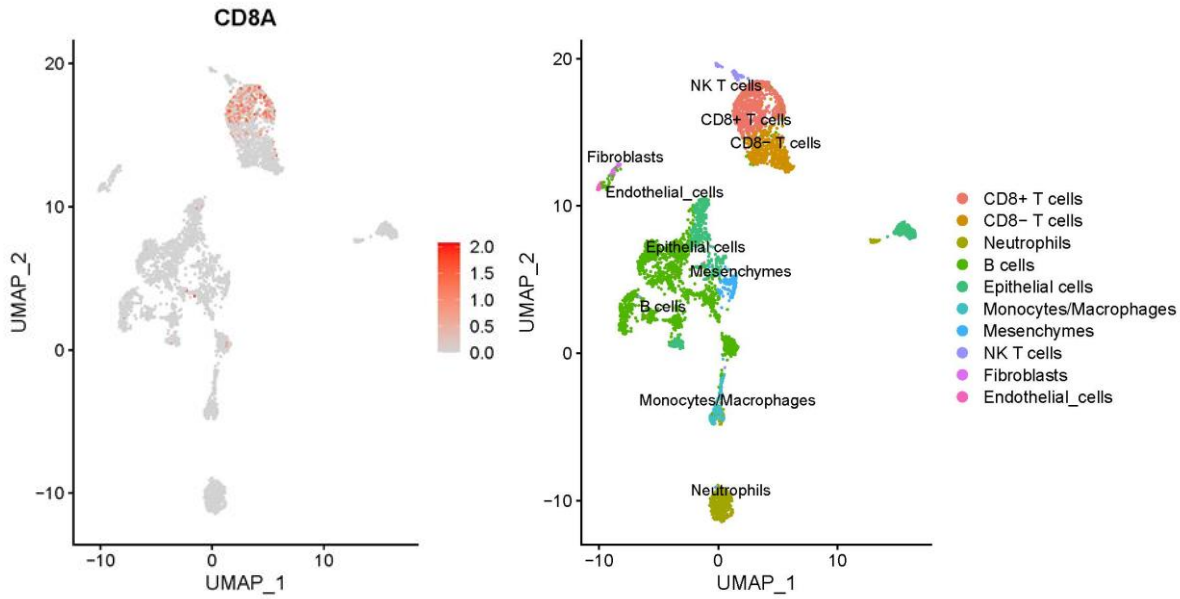


Figure 5.4.2.8 Specific cell types of each cluster

(left: CD8A expression in the resection samples;

right: cell type annotation of each cluster)

5.4.3 GSEA identifies subtle differences among monocytes/macrophages responses in CD patients

Next, DGE analysis was performed between the cells from cluster “12” of “stricture” and “proximal” samples, to identify the differences of gene expression in severe inflamed monocytes/macrophages comparing with the non-severe inflamed ones. DE genes were selected by less strict criteria (“avg_log2FC > 1” and “p_val_adj < 0.05”) than the in vitro studies (Figure 5.4.3.1). However, almost no DE genes were identified, which meant ORA could not be used to interpret the differences between the samples. This was not such a surprising result, as identifying DGE genes from scRNA-seq can be challenging, and the samples came from the same patient so the differences between them were subtle. In this situation, it was reasonable to try to interpret these subtle differences via GSEA, as the method did not require the thresholds and used the whole information from gene expression. By GSEA, as expected, significant up regulation of inflammatory response in the “stricture” sample was identified, corresponding to the sample condition examined by the pathologists (Figure 5.4.3.2), which proved GSEA was capable to interpret the subtle differences among monocytes/macrophages from CD patients. In this situation, it is okay to rely on GSEA to reveal the factors which changed accompanied with the severity of intestinal inflammation, and those factors might play important role in CD pathogenesis.

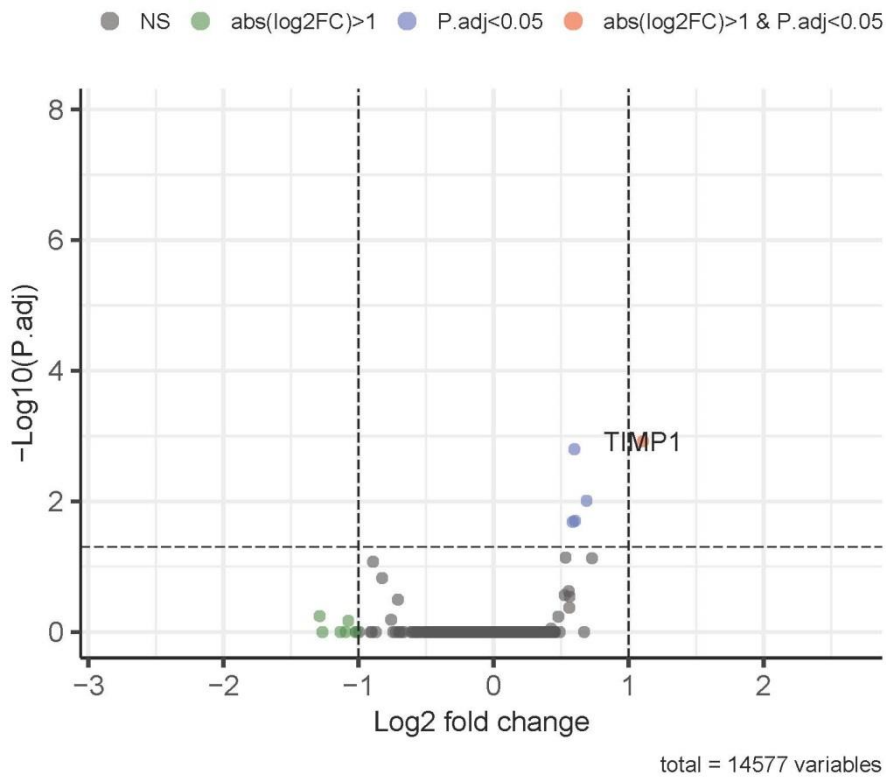


Figure 5.4.3.1 DE genes identified from comparisons between severe and non-severe inflamed monocytes/macrophages

(The results are displayed by volcano plot, and only one gene "*TIMP1*" was left after selection by the criteria "avg_log2FC > 1" and "p_val_adj < 0.05")

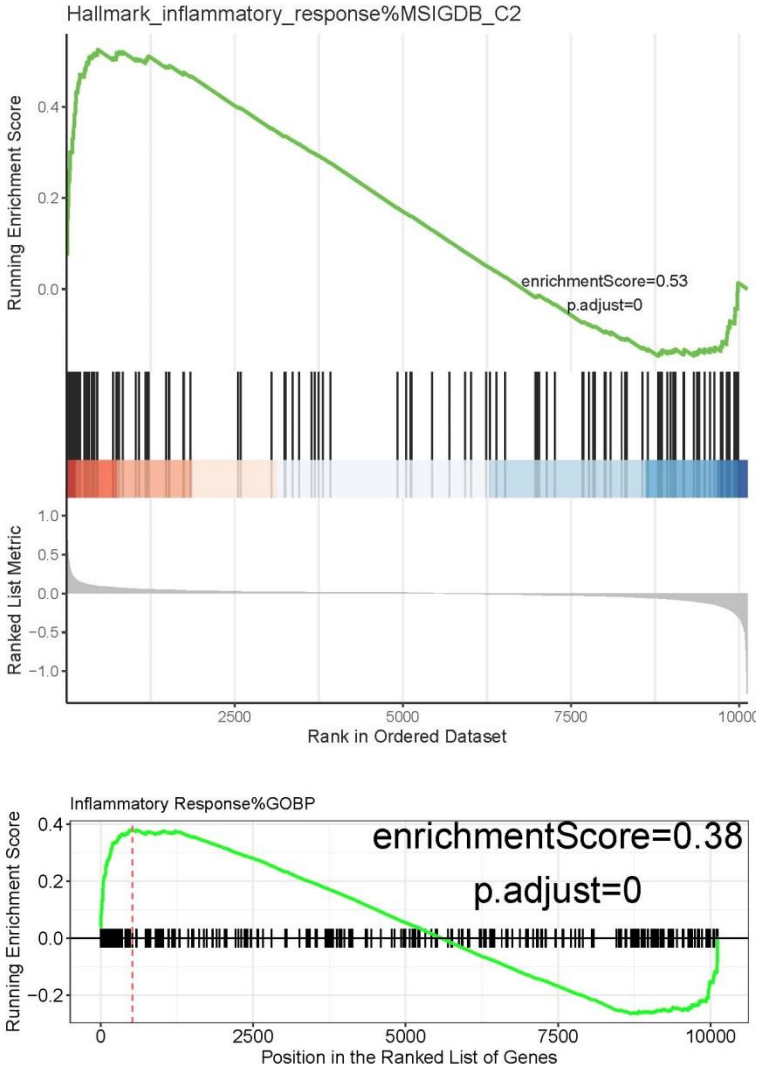


Figure 5.4.3.2 Up regulations of inflammatory response in monocytes/macrophages from "stricture" (severe) sample (p.adjust < 0.01)

("Hallmark inflammatory response (MSIGDB_C2)";

"Inflammatory Response (GO: 0006954)")

5.4.4 GSEA helps to link the *in vitro* models with the *in vivo* condition of CD patients

Next, it was meaningful to reveal the factors which changed accompanied with the severity of intestinal inflammation from the GSEA results comparing the "stricture" (more severe) sample with the "proximal" (less severe). A significant down regulation of the term associated with response to IFN-γ (Figure 5.4.4.1) and the corresponding signaling pathway (Figure 5.4.4.2) were identified. As the "stricture" sample had a more severe inflammatory condition comparing with "proximal" one, this result indicated that the down regulation of IFN-γ response in monocytes/macrophages was not contradictory with more severe intestinal inflammation in vivo. This matched the findings from *in vitro* studies of *NOD2* deficient macrophages, as in those macrophages with the most important CD risk factor (*NOD2* mutations), there was a clear down regulation of IFN-γ response genes.

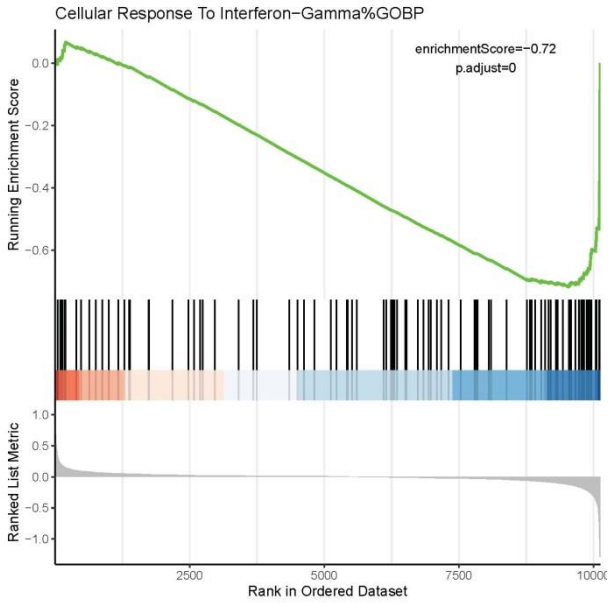


Figure 5.4.4.1 Down regulation of "Cellular Response To Interferon-Gamma" (GO: 0071346) in the "stricture" (severe) sample

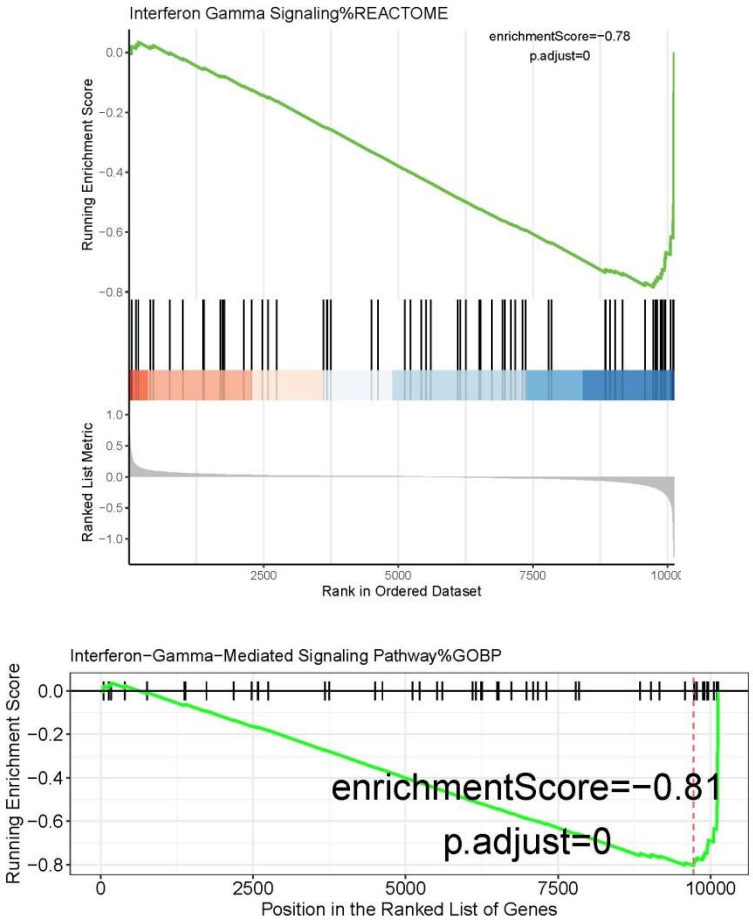


Figure 5.4.4.2 Down regulations of terms related to signaling pathway in the "stricture" (severe) sample

("Interferon Gamma Signaling" (REACTOME: R-HSA-877300.3);

"Interferon-Gamma-Mediated Signaling Pathway" (GO:0060333))

From the GSEA results, down regulations of oxidative phosphorylation in the “stricture” (severe) sample (Figure 5.4.4.3) with the effects on mitochondria and cellular energy supply chain (Figure 5.4.4.4) were also identified. The results indicated down regulation of oxidative phosphorylation via effects on mitochondria existed in monocytes/macrophages from CD resections with more severe inflammation. This matched the findings from *in vitro* studies of IFN-γ treated *LRRK2* deficient macrophages..

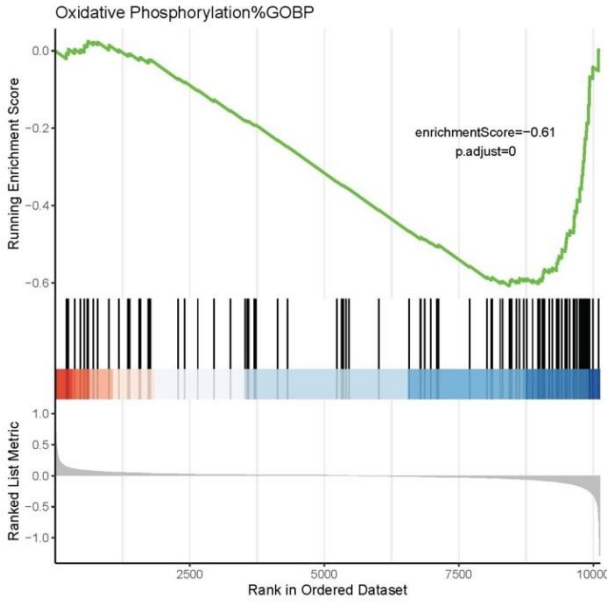


Figure 5.4.4.3 Down regulation of “Oxidative Phosphorylation” (GO: 0006119) in the monocytes/macrophages from “stricture” (severe) sample

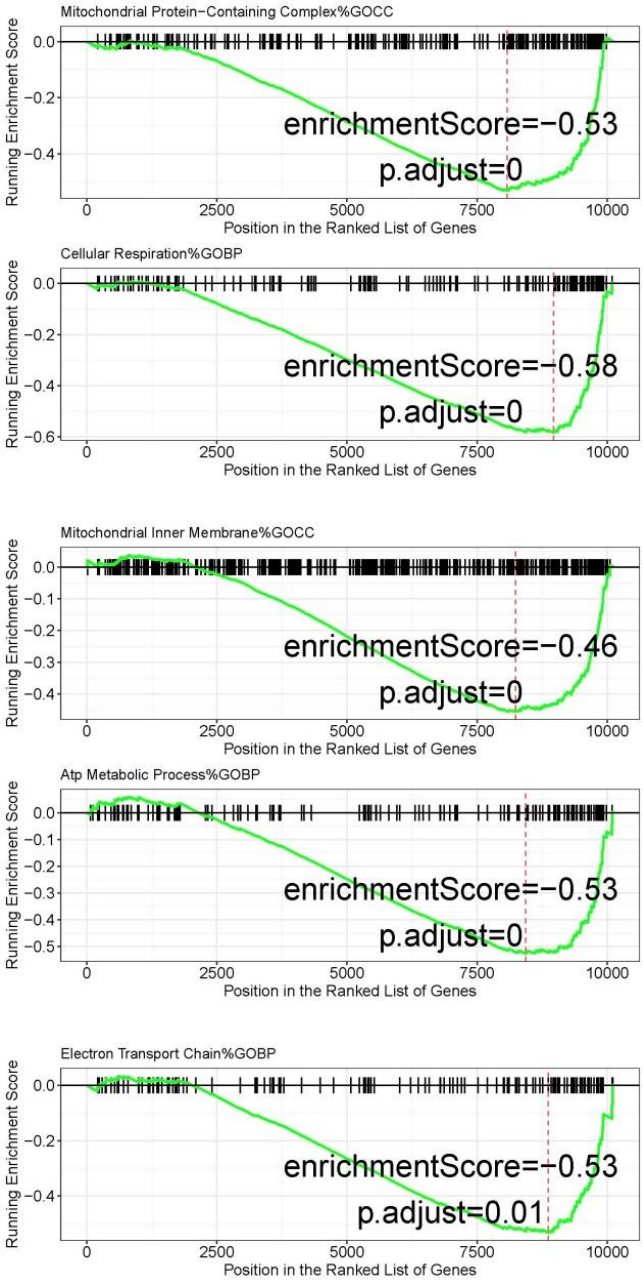


Figure 5.4.4.4 Down regulations of mitochondria and cellular energy supply chain related terms in the monocytes/macrophages from "stricture" (severe) sample (p.adjust ≤ 0.01)

("Mitochondrial Protein-Containing Complex" (GO: 0098798);

"Cellular Respiration" (GO: 0045333);

"Mitochondrial Inner Membrane" (GO: 0005743);

"Atp Metabolic Process" (GO: 0046034);

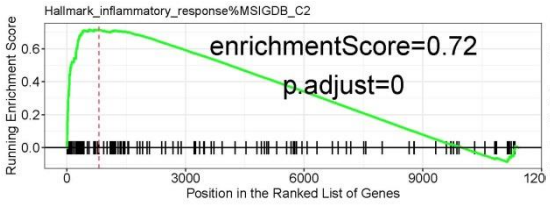
"Electron Transport Chain" (GO: 0022900))

5.4.5 Analysis of single cell RNA-seq data from CD patients' resections with severe and non-severe inflammatory pairs

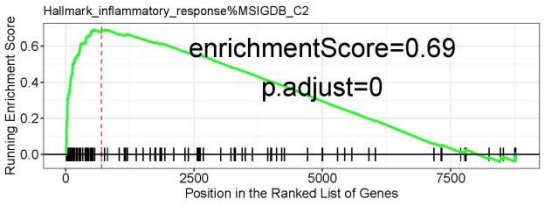
Single cell RNA-seq data from 11 paired resections (patient 1-11) from the same patient with severe ("infl") and non-severe ("uninfl") inflammation was downloaded from GEO: GSE134809 (Martin et al., 2019). Data was filtered using the criteria "nCount_RNA > 800 & percent.mt < 25" as in the original paper, and the same pipeline as section 5.4.2 - 5.4.4 was used to identify monocytes/macrophages, then GSEA was performed to interpret the differences between the monocytes/macrophages between severe and non-severe inflammatory samples. Less strict criteria for those samples was used, which took the GSEA results with "p.adjust < 0.05" as the significantly enriched ones. Up regulation of inflammatory responses in "infl" (severe) sample were successfully identified in 10/11 paired resections (except patient 6), which further confirmed GSEA was capable to interpret the subtle differences among monocytes/macrophages from CD patients (Figure 5.4.5.1). For enrichments of terms associated with response to IFN-γ, they were identified in 8/11 paired resections (patient 2, 3, 4, 5, 7, 8, 9, 10), however, down regulation was only identified in one paired resections (patient 8) (Figure 5.4.5.2). For enrichment of terms associated with oxidative phosphorylation, mitochondria and cellular energy supply chain, they were identified in 8/11 paired resections (patient 2, 4, 5, 6, 8, 9, 10, 11), with three (patient 4, 5, 10) up regulations and five (patient 2, 6, 8, 9, 11) down regulations (Figure 5.4.5.3). These results indicated both the effects on response to IFN-γ and oxidative phosphorylation, mitochondria and cellular energy supply chain were associated with inflammatory severity in monocytes/macrophages from CD patients, yet, the modes of influence were distinct. This suggests personal differences must be taken into considerations when studying those factors'

correlations with intestinal inflammation.

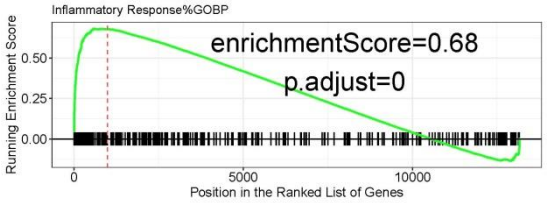
Patient 1



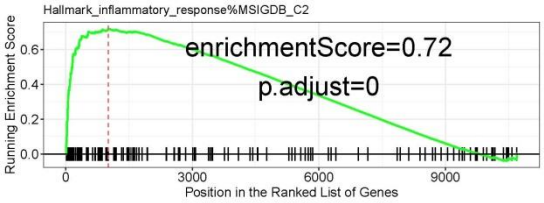
Patient 2



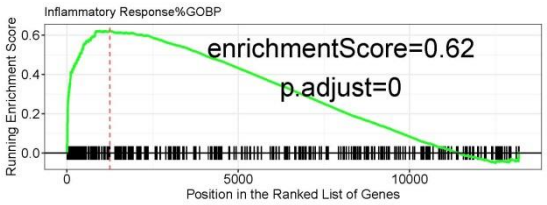
Patient 3



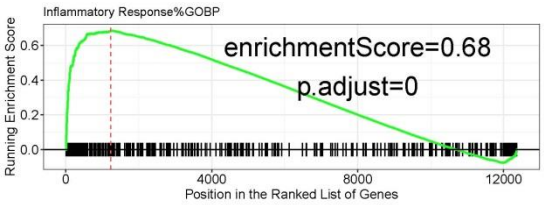
Patient4



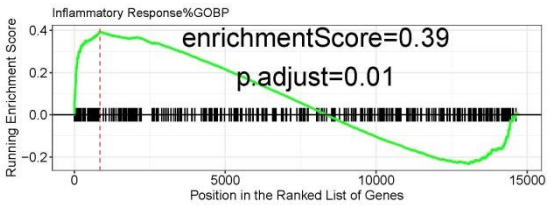
Patient 5



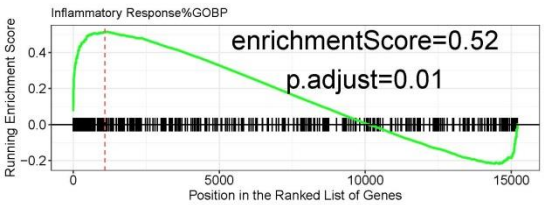
Patient 7



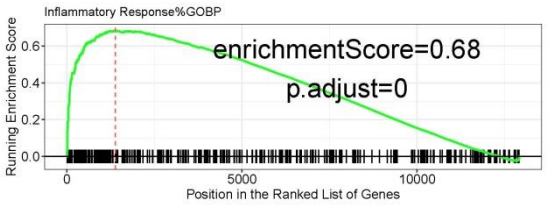
Patient 8



Patient 9



Patient 10



Patient 11

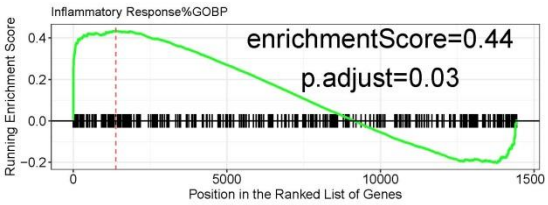
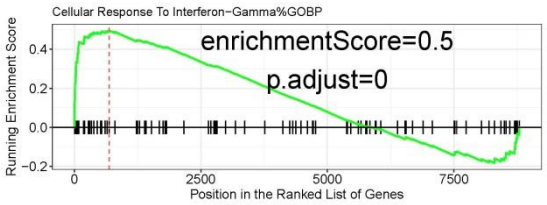
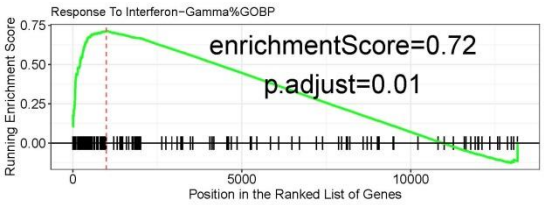


Figure 5.4.5.1 Enriched terms related to inflammatory response from GSEA between "infl" (severe) and "uninfl" (severe) sample in each patient

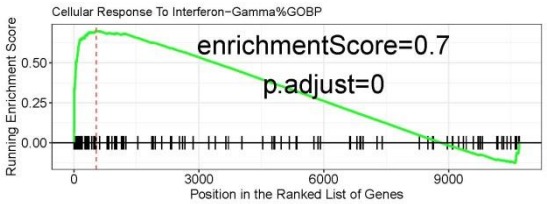
Patient 2



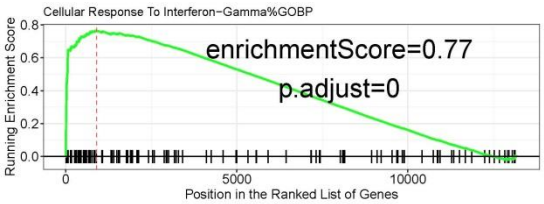
Patient 3



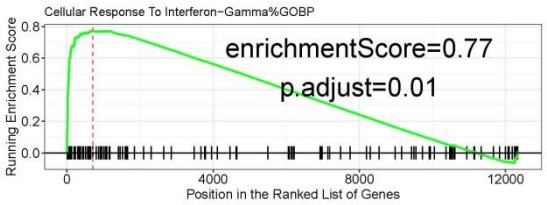
Patient 4



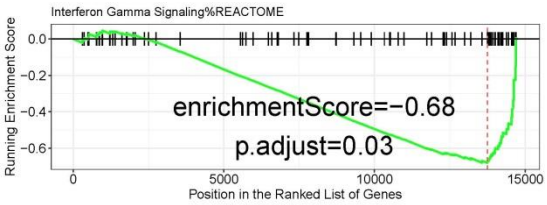
Patient 5



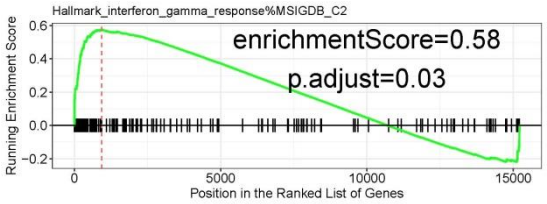
Patient 7



Patient 8



Patient 9



Patient 10

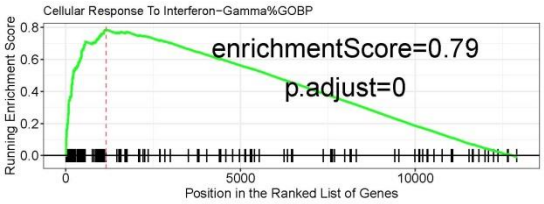
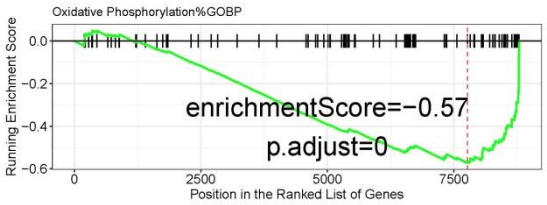
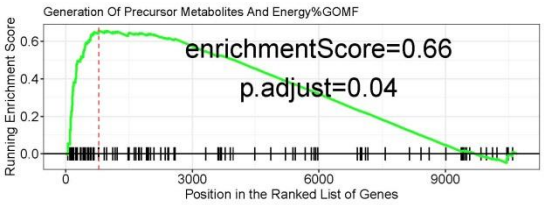


Figure 5.4.5.2 Enriched terms related to IFN-γ response from GSEA between "infl" (severe) and "uninfl" (severe) sample in each patient

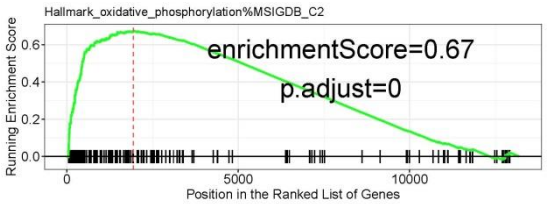
Patient 2



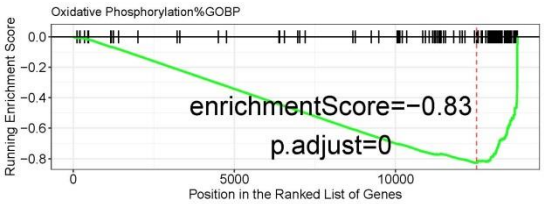
Patient 4



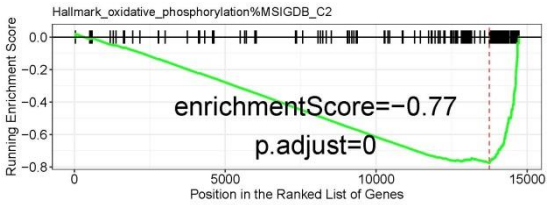
Patient 5



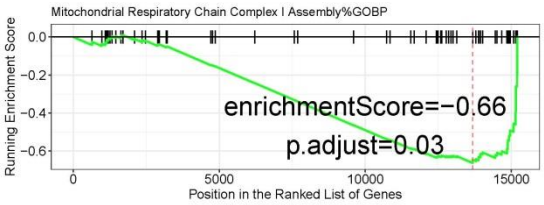
Patient 6



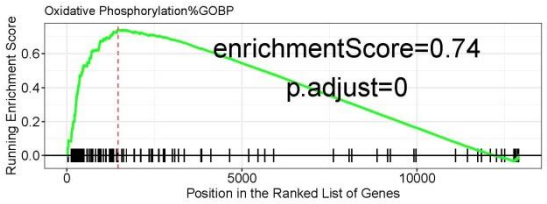
Patient 8



Patient 9



Patient 10



Patient 11

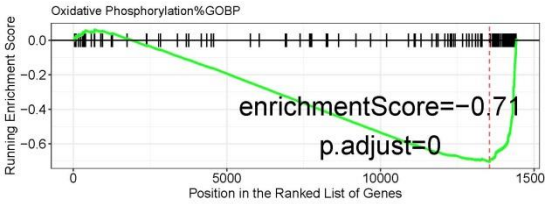


Figure 5.4.5.3 Enriched terms related to oxidative phosphorylation, mitochondria and cellular energy supply chain from GSEA between "infl" (severe) and "uninfl" (severe) sample in each patient

5.4.6 Generating intestinal organoids from CD patients' resections

For maintaining the characteristics of patients and accumulate resources for the biobanks of CD, intestinal organoids lines were tried to be generated from the remaining cells after scRNA-seq. Both the "stricture" and "proximal" samples were used for the trials on generating organoids, however only the "stricture" (severe inflammation) sample was able to generate a successful organoid line (Figure 5.4.6.1). For the "proximal" (non-severe inflammation) sample, although three dimensional tissues appeared during the culture, the number of them was much fewer than the "stricture" sample. Finally, however, it was not possible to establish stable organoids line from the proximal cells (Figure 5.4.6.2).

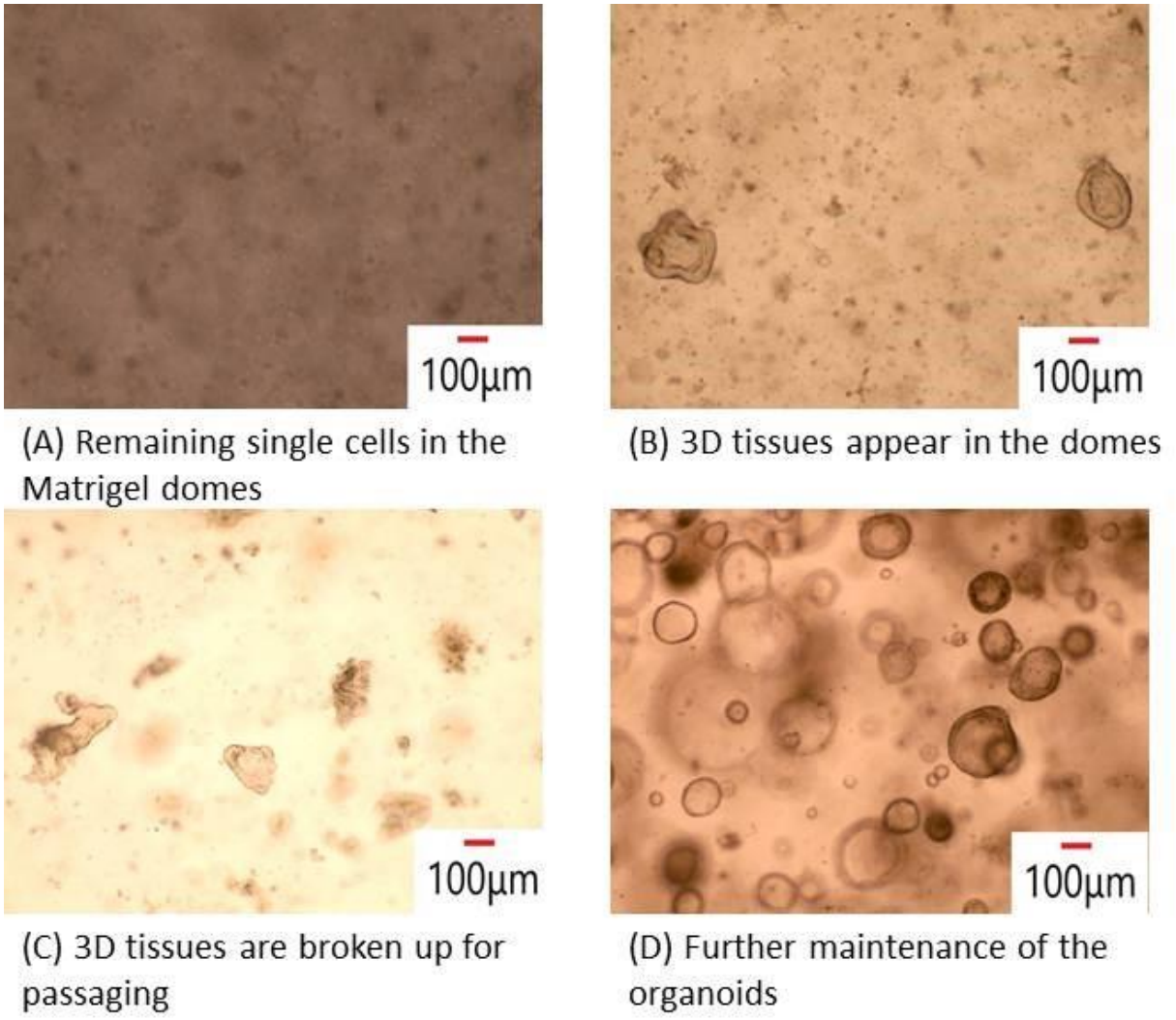


Figure 5.4.6.1 Generation of intestinal organoids from "stricture" sample

(For generating intestinal organoids, remaining single cells after scRNA-seq were embedded in Matrigel domes in 24-well plate, and cultured by medium with antibiotics (A). After one or two weeks culture with fully medium change every three or four days, 3D tissues appeared in the dome (B). Maintaining those tissues for several further days could make them grow larger in size, and capable for passage as hiPSC-derived intestinal organoids. Those tissues were broken up for passaging (C), and still maintained in the same medium. Organoids which grew well were as displayed in (D), with many "circles" be observed, and after several passage, intestinal organoids line from "stricture" sample was established successfully.)

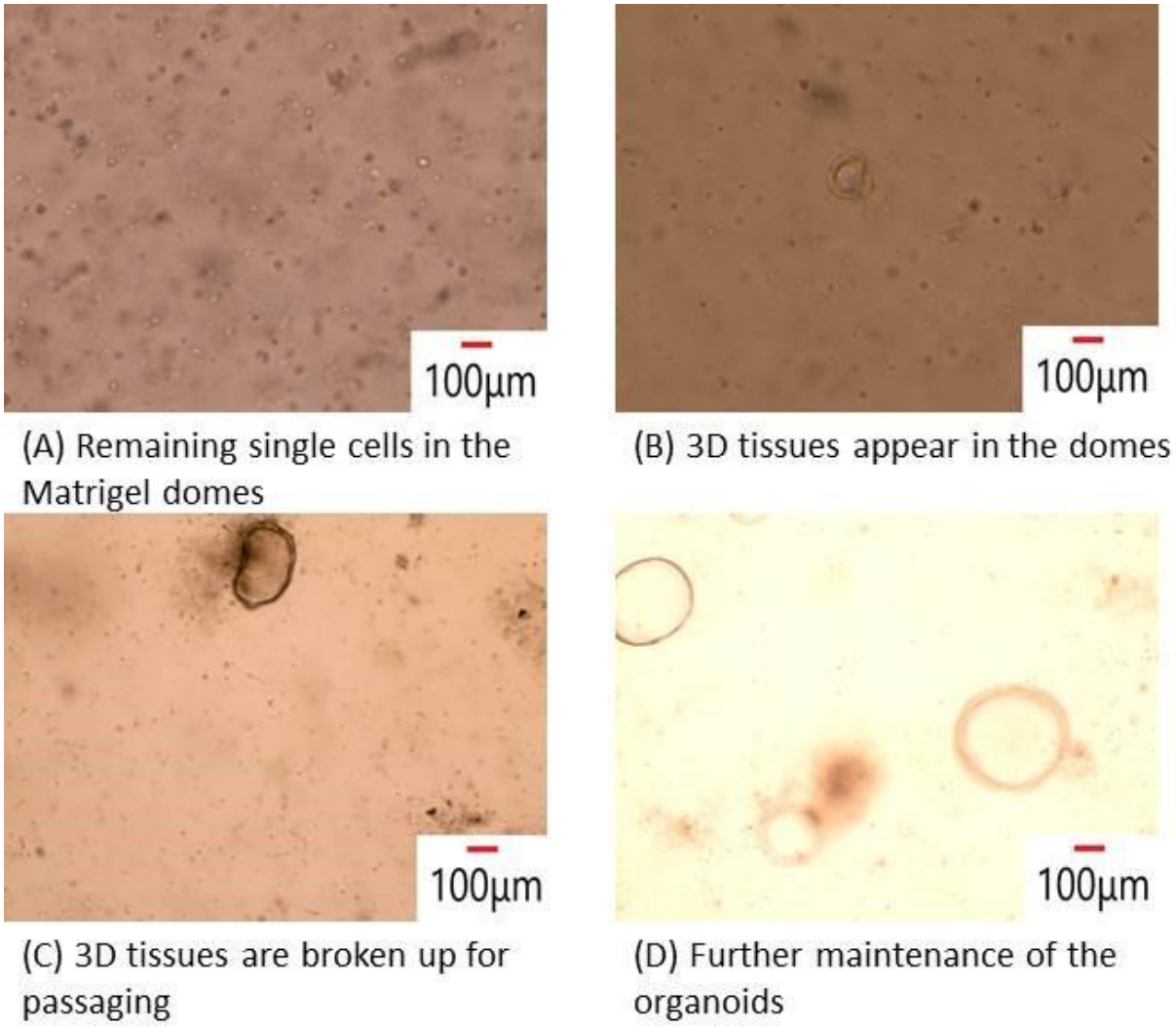


Figure 5.4.6.2 Generation of intestinal organoids from "proximal" sample

(For the figures from (A) to (D), they displayed the same process as Figure 5.4.7.1, just changed the original materials into remaining single cells from "proximal" sample. It could be found that in (D) observed "circles" were much fewer than in "stricture" sample, and ultimately no stable organoid line was generated from the "proximal" sample.)

5.5 Discussion

In this chapter, a pipeline for preparing scRNA-seq samples from CD patients' resections was established, and the generated data helped to assess the clinical meaning of the findings from *in vitro* macrophages, for possibly bridging *in vitro* models with clinical data in the future.

For this aim, the pipeline in laboratory for dissociating the CD patients' resections into single cells was established successfully, with better operability, good cell viability and less noise for the data analysis (removing debris). By this pipeline, single cell RNA sequencing data from a pair of resections from a CD patient was obtained, with distinct severity of inflammation (from pathological examination). Because the findings in Chapter 3 and 4 were mainly generated from *in vitro* derived macrophages, macrophages became the focus from those *in vivo* samples, and they were identified relying on computational analysis. Since macrophages in the intestines might experience differentiation from monocytes (Na et al., 2019), and considering the large number of similarities between those two cell types, it would be reasonable to merge those two cell types in the analysis.

The monocytes/macrophages from resections were successfully identified, and the cells from more severe inflammatory tissue were compared with the ones from less severe tissue to reveal the factors which changed accompanied with the severity of intestinal inflammation. In the analysis, GSEA was able to identify subtle differences among monocytes/macrophages responses in CD patients, yet ORA did not work. GSEA identified down regulation of IFN-γ response in monocytes/macrophages from resection with more severe inflammation. This meant that the down regulation of IFN-γ response in macrophages, as described in section 3.4.4 in the *NOD2* knockout macrophages, could be related to the severity of intestinal inflammation in CD. Meanwhile, down regulation of oxidative phosphorylation with the

effects on mitochondria in monocytes/macrophages was also identified from resection with more severe inflammation. This illustrated effects on mitochondria in macrophages, which were identified in section 4.4.6, really associated with the severity of intestinal inflammation in CD. Furthermore, the same style of scRNA-seq data (paired resections from the same CD patient with distinct severity of inflammation) was collected from online resources (Martin et al., 2019). Analysis of those data confirmed GSEA was able to identify subtle differences among monocytes/macrophages responses in CD patients. In addition, it also indicated both IFN-γ response and effects on mitochondria were related to the severity of intestinal inflammation in CD, which might play important role in the disease pathogenesis. Yet, it was hard to clarify how those factors affect the intestinal inflammation, as the influence had obvious person-to-person differences.

There is still small amount of scRNA-seq data from CD patients' resections (Huang et al., 2019; Martin et al., 2019; Nayar et al., 2021), which limits the further understandings on the disease. In my opinion, it is mainly caused by two reasons. First is the scarcity of research materials, since CD is impossible to be cured by surgeries (Bernstein, 2012; Roda et al., 2020), only in the case treatments by drugs cannot work well patients will choose surgeries. Second is the difficulty to achieve good cell viability, unlike cancer samples, cells in those resections from CD patients are under seriously inflammatory condition and too many of the cells have dead *in vivo*. The biggest contribution of this chapter is to establish a pipeline for to prepare scRNA-seq samples from CD patients' resections. Although the trials cost much time and were accompanied by many failures, finally the pipeline was succesfully generated and it continues to work well in cooperative groups in University of Birmingham. Meanwhile, the pipeline includes the generation of intestinal organoids lines from the remaining cells after scRNA-seq, which is useful for depositing CD patients' biobanks. Additionally, by analyzing

those scRNA-seq data from ours and others experiments, the links between *in vitro* models with the *in vivo* condition of CD patients are built up, and the works verify some of the findings from *in vitro* models really have clinical meaning.

CHAPTER 6

GENERAL DISCUSSION

6.1 Summary

Using human genetically engineered stem cell *in vitro* models, some new insights into CD pathogenesis was revealed, which would be hard to isolate from mouse models, immortalized/cancerous cell lines or primary cells from CD patients. Meanwhile, CD patients' resections were studied, and the data generated was used to assess the clinical relevance of the findings from the *in vitro* models.

For the aims on *NOD2*, the work found *NOD2* deficiency could lead to down regulation of IFN-γ response in the hiPSC-derived macrophages, and this characteristic was primed in macrophages even in the absence of treatment. In this situation, dysregulations of IFN-γ response can be recognized as the "diseased" status in the *NOD2* deficient macrophages. For the aims on *LRRK2*, this "disease" status was recognized as dysregulations of oxidative phosphorylation via mitochondria in IFN-γ treated macrophages.

Using the pipeline generated in the laboratory for preparing scRNA-seq samples from CD patients' resections, both two "diseased" statuses above were verified to be related to the severity of inflammation in CD. Data analysis of the online resources further supported the relationship.

Additionally, the work suggests that the co-expression between *NOD2* and *LRRK2* in IFN-γ treated macrophages, and the co-expression was recognized to be one-way by the knockout of those two genes (*NOD2* deficiency is available to decrease *LRRK2* expression, but *LRRK2* deficiency has no effects on *NOD2* expression). This phenomenon is possible to be associated with the hypothesis that CD and PD are linked, and the idea is supported by clinical evidences (described in section 1.4).

6.2 Limitations

Although this project has helped to achieve the above milestones, there remain some limitations.

First, for the *in vitro* studies, NOD2 and LRRK2 were mutated via introducing frameshift mutations into the genome. This operation can more obviously reflect the effects caused by genetic disorders, however, those circumstances which completely destroy the gene's function is almost impossible in reality. Among CD patients, mutations in the disease risk genes tend to be SNPs, and the gene can partially maintain its function, with not huge changes from the normal case. More and more evidence indicates the necessity of those mutation-based studies on CD, as statistical analysis reveals that even mutations in the same gene can have distinct influence on the disease risk (Hui et al., 2018; H.-S. Lee et al., 2021). Importantly, one limitation of this study should be emphasized. This work is a full loss of function in *NOD2* and *LRRK2*, and the precise SNP mutations that are associated with CD were not used. This may lead to some subtlety in the effect of *NOD2* on hinder the more precise understandings of the disease, and it is possible that the phenotype seen here is exaggerated over the *in vivo* CD phenotypes. Nonetheless, the fact that a similar pattern of disruption (response to IFN-γ and mitochondria) can be seen in the macrophages from CD patients (scRNA-seq data) is encouraging.

Second, this work lacks validation of the *NOD2* knockout at the protein level. Although the RNA sequencing data has indirectly verified the knockout of *NOD2* in *in vitro* differentiated macrophages (*NOD2* was recognized to be downregulated in both untreated, MDP treated and IFN-γ treated *NOD2* deficient macrophages, which was displayed in Figure 3.4.3.4, Figure 3.4.4.4, Figure 3.4.5.5 respectively), it is still important to validate *NOD2* knockout at protein level. Actually a lot of work was done, but it was not possible to obtain a

convincing Western blot result. *NOD2* was not expressed in hiPSCs (from the literature review and from the RNA-sequencing data generated in this study), so the validation of its knockout had to be by other somatic cells which could be *in vitro* differentiated. By literature review, *NOD2* was known to primarily express in the peripheral blood leukocytes and some intestinal epithelial cells (Ferrand et al., 2019). Analysis of the RNA sequencing data indicated macrophages treated with IFN-γ for 24 hours had the highest expression of *NOD2* (Figure 4.4.2.1), though the expression was still quite low compared to other genes, which were not highly expressive (e.g. *LRRK2*). In this situation, IFN-γ was applied to the wild type and *NOD2* deficient macrophages, and quantified the amount of *NOD2* protein in each sample via western blot. However, the target band for *NOD2* could still not be detected in wild type macrophages, due to the problematic specificity of the antibody. In this situation, it may be useful to generate custom *NOD2* antibodies. Moreover, some other protein-level detections such as IP (Immunoprecipitation) may also be useful for validation.

Third, the inter-relationship between *NOD2* and *LRRK2* expression identified in this research is only verified with *in vitro* derived macrophages, and evidence from *in vitro* derived microglia-like cells is lacking. Potentially this can be achieved from co-culture of the *in vitro*-derived macrophages with *in vitro*-derived neuronal cells (Claes et al., 2018; Haenseler et al., 2017; Takata et al., 2017). Meanwhile, evidence for the PD patients' macrophages and microglia is also lacking.

Fourth, the CD resection-based work was limited in size, lack of clinical information and centered around the transcriptome. Additionally, somatic mutations of the CD resections were not analyzed, but whole-genome sequencing of samples from IBD patients indicated their existence and effects recently (Olafsson et al., 2020).

Fifth, the work described here did not make full use of the intestinal organoids, for both

hiPSC-derived ones and the resection-derived ones. Immature conditions of those hiPSC-derived intestinal organoids restrict them to be a very helpful model for the research of CD pathogenesis (J. Kim et al., 2020; M. Li & Izpisua Belmonte, 2019).

6.3 Future work

The above limitations help to guide prospective future work.

First, it is reasonable to generate hiPSC lines with specific CD risk mutations. This precise gene editing can be completed via homologous recombination (HR) with designed oligo sequences (Cong & Zhang, 2015; Ran et al., 2013). Then more precise understanding of the relationship between those CD risk-associated genes and disease pathogenesis can be obtained from the platform for *in vitro* differentiation.

Second, the experiments on IFN-γ treated macrophages should be repeated in microglia-like cells, and it can help strengthen the ideas on the pathogenesis between CD and PD. This works could rely on an established method (Claes et al., 2018; Haenseler et al., 2017). Meanwhile, data from PD patients' macrophages and microglia from other studies or online resources should be pursued, to verify the link between CD macrophages in the intestine and PD microglia in the brain.

Third, the sample size for single cell analysis of resections from CD patients should be expanded, and new technologies such as single-cell based genomics need to be introduced into the research. For example the somatic mutations in the resections can be recognized by DNA sequencing at single-cell level, and the study can answer the question whether the findings from this work are associated with the possibly somatic mutations. Meanwhile, for more detailed understandings of CD, the clinical information such as what treatments those patients have experienced needs to be taken into consideration in the experiments..

Fourth, the hiPSC-derived intestinal organoids should be made more mature *in vitro,* to help apply them as a good disease model for CD. Current methods to mature those organoids rely on transplanting them into immunocompromised mice (Frum & Spence, 2020; Jung et al., 2018; G.-A. Kim, Spence, & Takayama, 2017; Rodansky, Johnson, Huang, Spence, &

Higgins, 2015; Singh, Poling, Spence, Wells, & Helmrath, 2020; Watson et al., 2014). Although it works, these chimeras lead to more noise for further research, and lose the advantage of those “organs in a dish”, since mice are still required in the experiments. Some trials on the complete *in vitro* maturation of those organoids has been performed, and achieved some progress, though the organoids generated in those trails were still not as mature as the ones from transplantation method (Jung et al., 2018). However, thanks to the detailed understandings derived from the development of those organoids (Q. Yu et al., 2021), hopefully a method which is able to mature the hiPSC-derived intestinal organoids completely *in vitro* can be established soon.

Fifth, in this project, the *in vitro* models are only used to explore CD pathogenesis, however, these models (hiPSC-derived macrophages or intestinal organoids) could also be used to develop novel treatments for CD. For example, based on the findings from *in vitro* differentiated macrophages, it is possible to test small molecules and other biologics on the *NOD2* or *LRRK2* deficient macrophages, and discover candidates which can significantly reverse the “diseased” status of the cells. It is also possible to target those candidates to develop novel therapies for CD.

List of References

Al Nabhani, Z., Dietrich, G., Hugot, J.-P., & Barreau, F. (2017). Nod2: The intestinal gate keeper. *PLoS pathogens, 13*(3), e1006177.

Aldridge, S., & Teichmann, S. A. (2020). Single cell transcriptomics comes of age. *Nature communications, 11*(1), 1-4.

Aran, D., Looney, A. P., Liu, L., Wu, E., Fong, V., Hsu, A., . . . Abate, A. R. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nature immunology, 20*(2), 163-172.

Arnott, I. D. R., Nimmo, E. R., Drummond, H. E., Fennell, J., Smith, B. R. K., MacKinlay, E., . . . O'Sullivan, M. (2004). NOD2/CARD15, TLR4 and CD14 mutations in Scottish and Irish Crohn's disease patients: evidence for genetic heterogeneity within Europe? *Genes & Immunity, 5*(5), 417-425.

Balasubramanian, I., & Gao, N. (2017). From sensing to shaping microbiota: insights into the role of NOD2 in intestinal homeostasis and progression of Crohn’s disease. *American Journal of Physiology-Gastrointestinal and Liver Physiology, 313*(1), G7-G13.

Behzadi, P., Behzadi, E., & Ranjbar, R. (2015). The incidence and prevalence of Crohn’s disease in global scale. *SOJ Immunol, 3*(2), 1-6.

Bernstein, C. N. (2012). The natural history of inflammatory bowel disease. In *Crohn's Disease and Ulcerative Colitis* (pp. 343-363): Springer.

Blighe, K., Rana, S., & Lewis, M. (2021). EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. In.

Braam, S. R., Zeinstra, L., Litjens, S., Ward-van Oostwaard, D., Van Den Brink, S., Van Laake, L., . . . Passier, R. (2008). Recombinant vitronectin is a functionally defined

substrate that supports human embryonic stem cell self-renewal via αVβ5 integrin. *Stem cells, 26*(9), 2257-2265.

Bromberg, J. F., Horvath, C. M., Wen, Z., Schreiber, R. D., & Darnell, J. E. (1996). Transcriptionally active Stat1 is required for the antiproliferative effects of both interferon alpha and interferon gamma. *Proceedings of the national academy of sciences, 93*(15), 7673-7678.

Bu, D., Luo, H., Huo, P., Wang, Z., Zhang, S., He, Z., . . . Guo, J. (2021). KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic acids research*.

Buchrieser, J., James, W., & Moore, M. D. (2017). Human induced pluripotent stem cell-derived macrophages share ontogeny with MYB-independent tissue-resident macrophages. *Stem cell reports, 8*(2), 334-345.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., & Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature biotechnology, 36*(5), 411-420.

Caussinus, H., Fekri, M., Hakam, S., & Ruiz-Gazen, A. (2003). A monitoring display of multivariate outliers. *Computational Statistics & Data Analysis, 44*(1-2), 237-252.

Caussinus, H., Hakam, S., & Ruiz-Gazen, A. (2002). Projections révélatrices contrôlées recherche d'individus atypiques. *Revue de Statistique Appliquée, 50*(4), 81-94.

Chao, Y.-X., Gulam, M. Y., Chia, N. S. J., Feng, L., Rotzschke, O., & Tan, E.-K. (2020). Gut–Brain Axis: Potential Factors Involved in the Pathogenesis of Parkinson's Disease. *Frontiers in Neurology, 11*, 849.

Chen, L., Lin, M.-J., Zhan, L.-L., & Lv, X.-P. (2012). Analysis of TLR4 and TLR2 polymorphisms in inflammatory bowel disease in a Guangxi Zhuang population.

World Journal of Gastroenterology: WJG, 18(46), 6856.

Chen, P., Stanojcic, M., & Jeschke, M. G. (2014). Differences between murine and human sepsis. *Surgical Clinics, 94*(6), 1135-1149.

Chen, S., Luo, Z., Ward, C., Ibañez, D. P., Liu, H., Zhong, X., . . . Wang, D. (2020). Generation of two LRRK2 homozygous knockout human induced pluripotent stem cell lines using CRISPR/Cas9. *Stem cell research, 45*, 101804.

Cheng, Y., Zhu, Y., Huang, X., Zhang, W., Han, Z., & Liu, S. (2015). Association between TLR2 and TLR4 gene polymorphisms and the susceptibility to inflammatory bowel disease: a meta-analysis. *PloS one, 10*(5), e0126803.

Chomczynski, P. (1993). A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. *Biotechniques, 15*(3), 532-534.

Chomczynski, P., & Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical biochemistry, 162*(1), 156-159.

Chomczynski, P., Wilfinger, W., Kennedy, A., Rymaszewski, M., & Mackey, K. (2010). RNAzol® RT: a new single-step method for isolation of RNA. *Nature methods, 7*(12), 4-5.

Claes, C., Van den Daele, J., & Verfaillie, C. M. (2018). Generating tissue-resident macrophages from pluripotent stem cells: Lessons learned from microglia. *Cellular immunology, 330*, 60-67.

Collins, F. S., Morgan, M., & Patrinos, A. (2003). The Human Genome Project: lessons from large-scale biology. *Science, 300*(5617), 286-290.

Concordet, J.-P., & Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic acids research,*

46(W1), W242-W245.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., . . . Marraffini, L. A. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science, 339*(6121), 819-823.

Cong, L., & Zhang, F. (2015). Genome engineering using CRISPR-Cas9 system. In *Chromosomal mutagenesis* (pp. 197-217): Springer.

Cox, D. B. T., Platt, R. J., & Zhang, F. (2015). Therapeutic genome editing: prospects and challenges. *Nature medicine, 21*(2), 121-131.

Crohn's & Ulcerative Colitis: Know the Difference? (2018). https://www.webmd.com/ibd-crohns-disease/crohns-colitis-difference

Derkinderen, P., & Neunlist, M. (2018). Crohn's and Parkinson disease: is LRRK2 lurking around the corner? *Nature Reviews Gastroenterology & Hepatology, 15*(6), 330-331.

Dinan, T. G., & Cryan, J. F. (2015). The impact of gut microbiota on brain and behaviour: implications for psychiatry. *Current Opinion in Clinical Nutrition & Metabolic Care, 18*(6), 552-558.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics, 29*(1), 15-21.

Dugan, J., Griffiths, E., Snow, P., Rosenzweig, H., Lee, E., Brown, B., . . . Davey, M. P. (2015). Blau syndrome–associated NOD2 mutation alters expression of full-length NOD2 and limits responses to muramyl dipeptide in knock-in mice. *The Journal of Immunology, 194*(1), 349-357.

Dzamko, N. L. (2017). LRRK2 and the immune system. *Leucine-Rich Repeat Kinase 2 (LRRK2),* 123-143.

Eckmann, L., & Karin, M. (2005). NOD2 and Crohn's disease: loss or gain of function?

Immunity, 22(6), 661-667.

Ferrand, A., Al Nabhani, Z., Tapias, N. S., Mas, E., Hugot, J.-P., & Barreau, F. (2019). NOD2 expression in intestinal epithelial cells protects toward the development of inflammation and associated carcinogenesis. *Cellular and molecular gastroenterology and hepatology, 7*(2), 357-369.

Flynn, R., Grundmann, A., Renz, P., Hänseler, W., James, W. S., Cowley, S. A., & Moore, M. D. (2015). CRISPR-mediated genotypic and phenotypic correction of a chronic granulomatous disease mutation in human iPS cells. *Experimental hematology, 43*(10), 838-848.

Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., . . . Armstrong, J. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic acids research, 47*(D1), D766-D773.

Frankish, A., Diekhans, M., Jungreis, I., Lagarde, J., Loveland, J. E., Mudge, J. M., . . . Barnes, I. (2021). GENCODE 2021. *Nucleic acids research, 49*(D1), D916-D923.

Frum, T., & Spence, J. R. (2020). hPSC-derived organoids: models of human development and disease. *Journal of Molecular Medicine*, 1-11.

Galdiero, M. R., Garlanda, C., Jaillon, S., Marone, G., & Mantovani, A. (2013). Tumor associated macrophages and neutrophils in tumor progression. *Journal of cellular physiology, 228*(7), 1404-1412.

Gardet, A., Benita, Y., Li, C., Sands, B. E., Ballester, I., Stevens, C., . . . Xavier, R. J. (2010). LRRK2 is involved in the IFN-γ response and host response to pathogens. *The Journal of Immunology, 185*(9), 5577-5585.

Gewirtz, A. T. (2007). TLRs in the Gut. III. Immune responses to flagellin in Crohn's disease: good, bad, or irrelevant? *American Journal of Physiology-Gastrointestinal and Liver*

Physiology, 292(3), G706-G710.

Gillespie, M., Jassal, B., Stephan, R., Milacic, M., Rothfels, K., Senff-Ribeiro, A., . . . Gong, C. (2022). The reactome pathway knowledgebase 2022. *Nucleic acids research, 50*(D1), D687-D692.

Grandy, R., Tomaz, R. A., & Vallier, L. (2019). Modeling disease with human inducible pluripotent stem cells. *Annual Review of Pathology: Mechanisms of Disease, 14*, 449-468.

Gray, P. W., & Goeddel, D. V. (1982). Structure of the human immune interferon gene. *Nature, 298*(5877), 859-863.

Grskovic, M., Javaherian, A., Strulovici, B., & Daley, G. Q. (2011). Induced pluripotent stem cells—opportunities for disease modelling and drug discovery. *Nature reviews Drug discovery, 10*(12), 915-929.

Gutbier, S., Wanke, F., Dahm, N., Rümmelin, A., Zimmermann, S., Christensen, K., . . . Geering, B. (2020). Large-scale production of human iPSC-derived macrophages for drug screening. *International journal of molecular sciences, 21*(13), 4808.

Haenseler, W., Sansom, S. N., Buchrieser, J., Newey, S. E., Moore, C. S., Nicholls, F. J., . . . Allen, N. D. (2017). A highly efficient human pluripotent stem cell microglia model displays a neuronal-co-culture-specific expression profile and inflammatory response. *Stem cell reports, 8*(6), 1727-1742.

Haeussler, M., Zweig, A. S., Tyner, C., Speir, M. L., Rosenbloom, K. R., Raney, B. J., . . . Gonzalez, J. N. (2019). The UCSC genome browser database: 2019 update. *Nucleic acids research, 47*(D1), D853-D858.

Hafemeister, C., & Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome biology, 20*(1),

1-15.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck Iii, W. M., Zheng, S., Butler, A., . . . Zager, M. (2021). Integrated analysis of multimodal single-cell data. *Cell*.

Harrow, J., Denoeud, F., Frankish, A., Reymond, A., Chen, C.-K., Chrast, J., . . . Swarbreck, D. (2006). GENCODE: producing a reference annotation for ENCODE. *Genome biology, 7*(1), 1-9.

Herrick, M. K., & Tansey, M. G. (2021). Is LRRK2 the missing link between inflammatory bowel disease and Parkinson's disease? *npj Parkinson's Disease, 7*(1), 1-7.

Hesketh, M., Sahin, K. B., West, Z. E., & Murray, R. Z. (2017). Macrophage phenotypes regulate scar formation and chronic wound healing. *International journal of molecular sciences, 18*(7), 1545.

Howe, K. L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., . . . Bhai, J. (2021). Ensembl 2021. *Nucleic acids research, 49*(D1), D884-D891.

Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell, 157*(6), 1262-1278.

Huang, B., Chen, Z., Geng, L., Wang, J., Liang, H., Cao, Y., . . . Wang, H. (2019). Mucosal profiling of pediatric-onset colitis and IBD reveals common pathogenics and therapeutic pathways. *Cell, 179*(5), 1160-1176.

Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., . . . Down, T. (2002). The Ensembl genome database project. *Nucleic acids research, 30*(1), 38-41.

Hui, K. Y., Fernandez-Hernandez, H., Hu, J., Schaffner, A., Pankratz, N., Hsu, N.-Y., . . . Li, X. (2018). Functional variants in the LRRK2 gene confer shared effects on risk for Crohn's disease and Parkinson's disease. *Science translational medicine, 10*(423).

Ikezu, T., Koro, L., Wolozin, B., Farraye, F. A., Strongosky, A. J., & Wszolek, Z. K. (2020).

Crohn's and Parkinson's Disease-Associated LRRK2 Mutations Alter Type II Interferon Responses in Human CD14+ Blood Monocytes Ex Vivo. *Journal of Neuroimmune Pharmacology, 15*(4), 794.

Jung, K. B., Lee, H., Son, Y. S., Lee, M.-O., Kim, Y.-D., Oh, S. J., . . . Kim, D.-S. (2018). Interleukin-2 induces the in vitro maturation of human pluripotent stem cell-derived intestinal organoids. *Nature communications, 9*(1), 1-13.

Kanazawa, N., Okafuji, I., Kambe, N., Nishikomori, R., Nakata-Hizume, M., Nagai, S., . . . Sakurai, Y. (2005). Early-onset sarcoidosis and CARD15 mutations with constitutive nuclear factor-κB activation: common genetic etiology with Blau syndrome. *Blood, 105*(3), 1195-1197.

Karlsson, K. R., Cowley, S., Martinez, F. O., Shaw, M., Minger, S. L., & James, W. (2008). Homogeneous monocytes and macrophages from human embryonic stem cells following coculture-free differentiation in M-CSF and IL-3. *Experimental hematology, 36*(9), 1167-1175.

Kassambara, A. (2020). Package 'ggpubr'. In.

Kim, G.-A., Spence, J. R., & Takayama, S. (2017). Bioengineering for intestinal organoid cultures. *Current opinion in biotechnology, 47*, 51-58.

Kim, J., Koo, B.-K., & Knoblich, J. A. (2020). Human organoids: model systems for human biology and medicine. *Nature Reviews Molecular Cell Biology, 21*(10), 571-584.

Kobayashi, K. S., Chamaillard, M., Ogura, Y., Henegariu, O., Inohara, N., Nuñez, G., & Flavell, R. A. (2005). Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. *Science, 307*(5710), 731-734.

Kuenzig, M. E., Manuel, D. G., Donelle, J., & Benchimol, E. I. (2020). Life expectancy and health-adjusted life expectancy in people with inflammatory bowel disease. *Cmaj,*

192(45), E1394-E1402.

Lee, A. J., & Ashkar, A. A. (2018). The dual nature of type I and type II interferons. *Frontiers in immunology, 9*, 2061.

Lee, H.-S., Lobbestael, E., Vermeire, S., Sabino, J., & Cleynen, I. (2021). Inflammatory bowel disease and Parkinson's disease: common pathophysiological links. *Gut, 70*(2), 408-417.

Lee, H., Flynn, R., Sharma, I., Haberman, E., Carling, P. J., Nicholls, F. J., . . . Wade-Martins, R. (2020). LRRK2 is recruited to phagosomes and co-recruits RAB8 and RAB10 in human pluripotent stem cell-derived macrophages. *Stem cell reports, 14*(5), 940-955.

Leong, R. W. L., Armuzzi, A., Ahmad, T., Wong, M. L., Tse, P., Jewell, D. P., & Sung, J. J. Y. (2003). NOD2/CARD15 gene polymorphisms and Crohn's disease in the Chinese population. *Alimentary pharmacology & therapeutics, 17*(12), 1465-1470.

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics, 12*(1), 1-16.

Li, M., & Izpisua Belmonte, J. C. (2019). Organoids—preclinical models of human disease. *New England Journal of Medicine, 380*(6), 569-579.

Liu, Z., Lee, J., Krummey, S., Lu, W., Cai, H., & Lenardo, M. J. (2011). The kinase LRRK2 is a regulator of the transcription factor NFAT that modulates the severity of inflammatory bowel disease. *Nature immunology, 12*(11), 1063-1070.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology, 15*(12), 1-21.

Lu, Y.-C., Yeh, W.-C., & Ohashi, P. S. (2008). LPS/TLR4 signal transduction pathway. *Cytokine, 42*(2), 145-151.

Ludwig, T. E., Bergendahl, V., Levenstein, M. E., Yu, J., Probasco, M. D., & Thomson, J. A.

(2006). Feeder-independent culture of human embryonic stem cells. *Nature methods, 3*(8), 637-646.

Mahe, M. M., Aihara, E., Schumacher, M. A., Zavros, Y., Montrose, M. H., Helmrath, M. A., . . . Shroyer, N. F. (2013). Establishment of gastrointestinal epithelial organoids. *Current protocols in mouse biology, 3*(4), 217-240.

Maleki, F., Ovens, K., Hogan, D. J., & Kusalik, A. J. (2020). Gene set analysis: challenges, opportunities, and future research. *Frontiers in genetics, 11*, 654.

Martin, J. C., Chang, C., Boschetti, G., Ungaro, R., Giri, M., Grout, J. A., . . . Greenstein, A. J. (2019). Single-cell analysis of Crohn's disease lesions identifies a pathogenic cellular module associated with resistance to anti-TNF therapy. *Cell, 178*(6), 1493-1508.

Mayer, E. A., Knight, R., Mazmanian, S. K., Cryan, J. F., & Tillisch, K. (2014). Gut microbes and the brain: paradigm shift in neuroscience. *Journal of Neuroscience, 34*(46), 15490-15496.

Mering, C. v., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., & Snel, B. (2003). STRING: a database of predicted functional associations between proteins. *Nucleic acids research, 31*(1), 258-261.

Mosser, D. M., & Edwards, J. P. (2008). Exploring the full spectrum of macrophage activation. *Nature reviews immunology, 8*(12), 958-969.

Na, Y. R., Stakenborg, M., Seok, S. H., & Matteoli, G. (2019). Macrophages in intestinal inflammation and resolution: a potential therapeutic target in IBD. *Nature Reviews Gastroenterology & Hepatology, 16*(9), 531-543.

Nayar, S., Morrison, J. K., Giri, M., Gettler, K., Chuang, L.-s., Walker, L. A., . . . Merad, M. (2021). A myeloid–stromal niche and gp130 rescue in NOD2-driven Crohn's disease. *Nature, 593*(7858), 275-281.

Ng, S. C., Shi, H. Y., Hamidi, N., Underwood, F. E., Tang, W., Benchimol, E. I., . . . Chan, F. K. L. (2017). Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *The Lancet, 390*(10114), 2769-2778.

Ng, S. C., Tsoi, K. K. F., Kamm, M. A., Xia, B., Wu, J., Chan, F. K. L., & Sung, J. J. Y. (2012). Genetics of inflammatory bowel disease in Asia: systematic review and meta-analysis. *Inflammatory bowel diseases, 18*(6), 1164-1176.

Nijland, R., Hofland, T., & Van Strijp, J. A. G. (2014). Recognition of LPS by TLR4: potential for anti-inflammatory therapies. *Marine drugs, 12*(7), 4260-4273.

Olafsson, S., McIntyre, R. E., Coorens, T., Butler, T., Jung, H., Robinson, P. S., . . . Dawson, C. (2020). Somatic evolution in non-neoplastic IBD-affected colon. *Cell, 182*(3), 672-684.

Pauleau, A.-L., & Murray, P. J. (2003). Role of nod2 in the response of macrophages to toll-like receptor agonists. *Molecular and cellular biology, 23*(21), 7531-7539.

Pizarro, T. T., Arseneau, K. O., Bamias, G., & Cominelli, F. (2003). Mouse models for the study of Crohn's disease. *Trends in molecular medicine, 9*(5), 218-222.

Plevy, S. (2021). A molecular connection hints at how a genetic risk factor drives Crohn’s disease. In: Nature Publishing Group.

Qing, Y., & Stark, G. R. (2004). Alternative activation of STAT1 and STAT3 in response to interferon-γ. *Journal of Biological Chemistry, 279*(40), 41679-41685.

Qiu, P., Shandilya, H., D'Alessio, J. M., O'Connor, K., Durocher, J., & Gerard, G. F. (2004). Mutation detection using Surveyor™ nuclease. *Biotechniques, 36*(4), 702-707.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature protocols, 8*(11), 2281-2308.

Ratajczak, M. Z., Pedziwiatr, D., Cymer, M., Kucia, M., Kucharska-Mazur, J., & Samochowiec, J. (2018). Sterile inflammation of brain, due to activation of innate immunity, as a culprit in psychiatric disorders. *Frontiers in psychiatry, 9*, 60.

Risso, D., Schwartz, K., Sherlock, G., & Dudoit, S. (2011). GC-content normalization for RNA-Seq data. *BMC bioinformatics, 12*(1), 1-17.

Rivas, M. A., Avila, B. E., Koskela, J., Huang, H., Stevens, C., Pirinen, M., . . . Ganna, A. (2018). Insights into the genetic epidemiology of Crohn's and rare diseases in the Ashkenazi Jewish population. *PLoS genetics, 14*(5), e1007329.

Rochereau, N., Roblin, X., Michaud, E., Gayet, R., Chanut, B., Jospin, F., . . . Paul, S. (2021). NOD2 deficiency increases retrograde transport of secretory IgA complexes in Crohn’s disease. *Nature communications, 12*(1), 1-13.

Roda, G., Ng, S. C., Kotze, P. G., Argollo, M., Panaccione, R., Spinelli, A., . . . Danese, S. (2020). Crohn’s disease. *Nature Reviews Disease Primers, 6*(1), 1-19.

Rodansky, E. S., Johnson, L. A., Huang, S., Spence, J. R., & Higgins, P. D. R. (2015). Intestinal organoids: a model of intestinal fibrosis for evaluating anti-fibrotic drugs. *Experimental and molecular pathology, 98*(3), 346-351.

Rodchenkov, I., Babur, O., Luna, A., Aksoy, B. A., Wong, J. V., Fong, D., . . . Wrana, M. (2020). Pathway Commons 2019 Update: integration, analysis and exploration of pathway data. *Nucleic acids research, 48*(D1), D489-D497.

Rolli-Derkinderen, M., Leclair-Visonneau, L., Bourreille, A., Coron, E., Neunlist, M., & Derkinderen, P. (2020). Is Parkinson’s disease a chronic low-grade inflammatory bowel disease? *Journal of neurology, 267*(8), 2207-2213.

Rowe, R. G., & Daley, G. Q. (2019). Induced pluripotent stem cells in disease modelling and drug discovery. *Nature Reviews Genetics, 20*(7), 377-388.

Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., & Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nature biotechnology, 33*(5), 495-502.

Sato, T., Stange, D. E., Ferrante, M., Vries, R. G. J., Van Es, J. H., Van Den Brink, S., . . . Siersema, P. D. (2011). Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology, 141*(5), 1762-1772.

Sato, T., Vries, R. G., Snippert, H. J., Van De Wetering, M., Barker, N., Stange, D. E., . . . Peters, P. J. (2009). Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature, 459*(7244), 262-265.

Schoenborn, J. R., & Wilson, C. B. (2007). Regulation of interferon-γ during innate and adaptive immune responses. *Advances in immunology, 96*, 41-101.

Schroder, K., Hertzog, P. J., Ravasi, T., & Hume, D. A. (2004). Interferon-γ: an overview of signals, mechanisms and functions. *Journal of leukocyte biology, 75*(2), 163-189.

Schroder, K., Irvine, K. M., Taylor, M. S., Bokil, N. J., Le Cao, K.-A., Masterman, K.-A., . . . Fairbairn, L. (2012). Conservation and divergence in Toll-like receptor 4-regulated gene expression in primary human versus mouse macrophages. *Proceedings of the national academy of sciences, 109*(16), E944-E953.

Seok, J., Warren, H. S., Cuenca, A. G., Mindrinos, M. N., Baker, H. V., Xu, W., . . . Hennessy, L. (2013). Genomic responses in mouse models poorly mimic human inflammatory diseases. *Proceedings of the national academy of sciences, 110*(9), 3507-3512.

Singh, A., Poling, H. M., Spence, J. R., Wells, J. M., & Helmrath, M. A. (2020). Gastrointestinal organoids: a next-generation tool for modeling human development. *American Journal of Physiology-Gastrointestinal and Liver Physiology, 319*(3), G375-G381.

Soneson, C., Love, M. I., & Robinson, M. D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research, 4*.

Spence, J. R., Mayhew, C. N., Rankin, S. A., Kuhar, M. F., Vallance, J. E., Tolle, K., . . . Zorn, A. M. (2011). Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro. *Nature, 470*(7332), 105-109.

Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA sequencing: the teenage years. *Nature Reviews Genetics, 20*(11), 631-656.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck Iii, W. M., . . . Satija, R. (2019). Comprehensive integration of single-cell data. *Cell, 177*(7), 1888-1902.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., . . . Lander, E. S. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the national academy of sciences, 102*(43), 15545-15550.

Sudo, N., Chida, Y., Aiba, Y., Sonoda, J., Oyama, N., Yu, X. N., . . . Koga, Y. (2004). Postnatal microbial colonization programs the hypothalamic–pituitary–adrenal system for stress response in mice. *The Journal of physiology, 558*(1), 263-275.

Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., . . . Bork, P. (2010). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic acids research, 39*(suppl_1), D561-D568.

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., . . . Bork, P. (2021). The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic acids research, 49*(D1), D605-D612.

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., . . . Bork, P.

(2016). The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic acids research*, gkw937.

Takada, S., Kambe, N., Kawasaki, Y., Niwa, A., Honda-Ozaki, F., Kobayashi, K., . . . Hotta, A. (2018). Pluripotent stem cell models of Blau syndrome reveal an IFN-γ–dependent inflammatory response in macrophages. *Journal of Allergy and Clinical Immunology, 141*(1), 339-349.

Takagawa, T., Kitani, A., Fuss, I., Levine, B., Brant, S. R., Peter, I., . . . Strober, W. (2018). An increase in LRRK2 suppresses autophagy and enhances Dectin-1–induced immunity in a mouse model of colitis. *Science translational medicine, 10*(444).

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell, 131*(5), 861-872.

Takata, K., Kozaki, T., Lee, C. Z. W., Thion, M. S., Otsuka, M., Lim, S., . . . Malleret, B. (2017). Induced-pluripotent-stem-cell-derived primitive macrophages provide a platform for modeling tissue-resident macrophage differentiation and function. *Immunity, 47*(1), 183-198.

Tanaka, A., To, J., O'Brien, B., Donnelly, S., & Lund, M. (2017). Selection of reliable reference genes for the normalisation of gene expression levels following time course LPS stimulation of murine bone marrow derived macrophages. *BMC immunology, 18*(1), 1-12.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., . . . Siddiqui, A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nature methods, 6*(5), 377-382.

Torres, J., Mehandru, S., Colombel, J.-F., & Peyrin-Biroulet, L. (2017). Crohn's disease. *The*

Lancet, 389(10080), 1741-1755.

Trindade, B. C., & Chen, G. Y. (2020). NOD1 and NOD2 in inflammatory and infectious diseases. *Immunological Reviews, 297*(1), 139-161.

Tsai, Y.-H., Nattiv, R., Dedhia, P. H., Nagy, M. S., Chin, A. M., Thomson, M., . . . Spence, J. R. (2017). In vitro patterning of pluripotent stem cell-derived intestine recapitulates in vivo human development. *Development, 144*(6), 1045-1055.

van Wilgenburg, B., Moore, M. D., James, W. S., & Cowley, S. A. (2014). The productive entry pathway of HIV-1 in macrophages is dependent on endocytosis through lipid rafts containing CD4. *PloS one, 9*(1), e86071.

Wang, Y., & Kasper, L. H. (2014). The role of microbiome in central nervous system disorders. *Brain, behavior, and immunity, 38*, 1-12.

Warnes, M. G. R., Bolker, B., Bonebakker, L., Gentleman, R., & Huber, W. (2016). Package 'gplots'. *Various R programming tools for plotting data*.

Watson, C. L., Mahe, M. M., Múnera, J., Howell, J. C., Sundaram, N., Poling, H. M., . . . Sun, Y. (2014). An in vivo model of human small intestine using pluripotent stem cells. *Nature medicine, 20*(11), 1310-1314.

Waugh, N., Cummins, E., Royle, P., Kandala, N. B., Shyangdan, D., Arasaradnam, R., . . . Johnston, R. (2013). Faecal calprotectin testing for differentiating amongst inflammatory and non-inflammatory bowel diseases: systematic review and economic evaluation. *Health technology assessment (Winchester, England), 17*(55), xv-xix.

Wehkamp, J., & Stange, E. F. (2005). NOD2 mutation and mice: no Crohn's disease but many lessons to learn. *Trends in molecular medicine, 11*(7), 307-309.

Wickham, H. (2011). ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics, 3*(2), 180-185.

Wilfinger, W. W., Mackey, K., & Chomczynski, P. (1997). Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity. *Biotechniques, 22*(3), 474-481.

Wilgenburg, B. v., Browne, C., Vowles, J., & Cowley, S. A. (2013). Efficient, long term production of monocyte-derived macrophages from human pluripotent stem cells under partly-defined and fully-defined conditions. *PloS one, 8*(8), e71098.

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., . . . Zhan, L. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation, 2*(3), 100141.

Yamazaki, K., Takazoe, M., Tanaka, T., Kazumori, T., & Nakamura, Y. (2002). Absence of mutation in the NOD2/CARD15 gene among 483 Japanese patients with Crohn's disease. *Journal of human genetics, 47*(9), 469-472.

Yu, G. (2021). enrichplot: Visualization of Functional Enrichment Result. In.

Yu, G., Wang, L.-G., Han, Y., & He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: a journal of integrative biology, 16*(5), 284-287.

Yu, Q., Kilik, U., Holloway, E. M., Tsai, Y.-H., Harmel, C., Wu, A., . . . He, Z. (2021). Charting human development using a multi-endodermal organ atlas and organoid models. *Cell, 184*(12), 3281-3298.

Zhang, F. (2019). Development of CRISPR-Cas systems for genome editing and beyond. *Quarterly Reviews of Biophysics, 52*.

Zhang, Q., Pan, Y., Yan, R., Zeng, B., Wang, H., Zhang, X., . . . Liu, Z. (2015). Commensal bacteria direct selective cargo sorting to promote symbiosis. *Nature immunology, 16*(9), 918-926.

Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., . . . Zhu, J.

(2017). Massively parallel digital transcriptional profiling of single cells. *Nature communications, 8*(1), 1-12.

Zhou, T., Benda, C., Duzinger, S., Huang, Y., Li, X., Li, Y., . . . Hao, L. (2011). Generation of induced pluripotent stem cells from urine. *Journal of the American Society of Nephrology, 22*(7), 1221-1228.

Appendix

Predicted off-target condition of NOD2 deficient hiPSC line

(offtargetSeq: sequence of the off-target region;

chrom: which chromosome the off-target region on;

start: position of the off-target region start on the human genome;

end: position of the off-target region end on the human genome;

strand: off-target region is on forward strand (+) or reverse strand (-);

locusDesc: description of the off-target region)

offtargetSeq	chrom	start	end	strand	locusDesc
GTGAGACAGAAAAGAC ACCGCGG	chr5	1.79E+08	1.79E+08	-	intergenic:RP11-798K23.1-U4
AGGAGTCTGAGCAGAC ACCATGG	chr5	1.78E+08	1.78E+08	+	intergenic:RN7SL646P-CTB-26E19.1
GGGAGGCTGAAAAGAC ACCGCGG	chr10	6840188	6840210	-	intron:LINC00707
GGGAGGCTGAGCAAAC ACCAAGG	chr2	16471076	16471098	+	intergenic:AC010880.1-AC104623.2
GGGAGACAGAGCAAAC ACCTGGG	chr22	27776929	27776951	+	intron:MN1
GGGAGCCTGTGCAGAA ACCGAGG	chr18	22711193	22711215	-	intergenic:RP11-739L10.1-RP11-370A5.1
TAGAGGCTGAGCAGACA CTGTGG	chr17	76668146	76668168	-	intergenic:RNU6-227P-RP11-318A15.2/MXRA7
GGGAGACAGAGGAGAC ACCAAGG	chr4	78174028	78174050	-	intergenic:FRAS1-AC093897.1
GCGAGACTGAGAAAAA ACCCTGG	chr1	99406871	99406893	-	intergenic:LPPR4-RP11-413P11.1
GCAACACTGGGCAGAC ACCTGGG	chr9	81257992	81258014	-	intergenic:AL354682.1-TLE1
GGGTGCCTGAGCAGAC ACCAAGG	chr18	27053381	27053403	-	intron:CHST9
GGGGACCTGAGCAGAC ACCGTGG	chr8	1.41E+08	1.41E+08	+	intergenic:RP11-642A1.1-AGO2
GCGACACTGAACAGATA CCTGGG	chr1	1.17E+08	1.17E+08	+	intron:VTCN1
GCAAGTCTGAGCACACA CCTCGG	chr7	27210794	27210816	+	intergenic:HOTTIP/HOTTIP_4-RP1-170O19.17
GCGGGAGGGAGCAGAC ACCTCGG	chr11	69953590	69953612	+	intergenic:RP11-643C9.2-RP11-626H12.1
GCGAGCCTGAGCGGGC ACCTCGG	chr19	8592878	8592900	-	intron:ADAMTS10
GAAATATTGAGCAGACA CCGAGG	chr4	1.27E+08	1.27E+08	-	intergenic:RP11-318I4.1-Y_RNA
GGGAGACTGGGGAGAC ACCTTGG	chr16	85419111	85419133	+	intergenic:RP11-680G10.1-AC092377.1

GTGGGCCTGAGCAGAC ACTGAGG	chr20	8667587	8667609	-	intron:PLCB1
GAGCGACAGGGCAGAC ACCGTGG	chr2	47886996	47887018	+	intron:FBXO11
GTGATACTGAGCAGACA ACTTGG	chr20	7549991	7550013	-	intergenic:MIR8062-RN7SL547P
GGGAGGGTGAGCAGAC ACAGAGG	chrX	98900142	98900164	+	intergenic:RP11-485F13.1-AL109750.1
GGGAGAGTGAGCAGGC ACAGTGG	chr3	1.4E+08	1.4E+08	-	exon:NMNAT3
GGGGGCCTGAGCAGAC ACAGAGG	chr10	1.01E+08	1.01E+08	+	exon:KAZALD1
CCGAGGCTGATCAGACA CCCAGG	chr1	31218961	31218983	+	intron:NKAIN1
GAGAGAGAGAGCTGAC ACCGTGG	chr5	1.6E+08	1.6E+08	-	intergenic:AC008703.2-AC008691.1
CAGAAACTGAGCAGAC GCCGCGG	chrY	6911505	6911527	-	intron:TBL1Y
GCTAGTCTGAGGAGACA CAGAGG	chr7	94760414	94760436	-	intergenic:RNU6-956P-RN7SKP129
GGGAGACTGAGCAGAA TCCTAGG	chr5	54577452	54577474	+	intergenic:SNX18-CTD-2591A1.1
GCTGGACTGAGCTGGCA CCGAGG	chr22	37144311	37144333	-	intron:IL2RB
GCCAGCCAGAGCAGAC ACAGGGG	chr16	86225366	86225388	+	intergenic:RP11-514D23.3-RP11-514D23.2
GGGAGACAGAGCAGAG ACCTTGG	chr1	1.57E+08	1.57E+08	-	intergenic:RP11-85G21.2-FCRL5
GCGGGTCTGAGCAGTCA TCGGGG	chr9	1.34E+08	1.34E+08	-	intergenic:RXRA-RP11-473E2.2
GTGAGAGTGAGCAGAC AGCATGG	chrX	72718576	72718598	+	intergenic:PHKA1-U3
GGCAGACTGAGCACAC ACTGAGG	chr10	1.23E+08	1.23E+08	-	intergenic:BUB3-RP11-282I1.1
ACGGGACGGAGCAGAC ACGGGGG	chr1	1289508	1289530	-	intron:SCNN1D
ACGGGACGGAGCAGAC ACGGGGG	chr1	1289965	1289987	-	intron:SCNN1D
TCGAGACTCAGCACTCA CCGAGG	chr6	1.48E+08	1.48E+08	+	intergenic:SAMD5-RP11-307P5.1
GGGAAACTGAGCAGAC CCAGGGG	chr20	32970347	32970369	+	intergenic:EFCAB8-SUN5
GGGCGACTGACCAGAC ACAGAGG	chr16	66977264	66977286	+	intergenic:CES3-CES4A
TCGCGGCTGAGCAGACG CCGGGG	chr14	68816384	68816406	-	intergenic:ZFP36L1-ACTN1
CCAAGACTGAGCAGAC CCTGTGG	chr18	11489727	11489749	-	intergenic:RP11-128P17.3-LINC01255
GCGACACGGAGCAGAG ACCCAGG	chr10	1.3E+08	1.3E+08	+	intergenic:EBF3-LINC00959
GCGAGGCTGAGCGGGC ACGGAGG	chr12	67069666	67069688	+	intergenic:RP11-123O10.4-RP11-123O10.3
GTGAGGCTGAGCAGAC AGAGCGG	chr10	95336575	95336597	-	exon:SORBS1

CAGAGACTGAGCAGAGACTGAGG	chr11	1.22E+08	1.22E+08	+	intergenic:SORL1-RNU6-256P
GCGAGACTGCTCTGACACCATGG	chr18	38028689	38028711	+	intergenic:MIR4318-RP11-687D19.1
TCGTGACTGAGCAGCCCCCGGGG	chr22	49457348	49457370	+	intergenic:RP1-34P24.3-CTA-722E9.1
GCAAGAATGACCAGACACGGTGG	chr11	39507526	39507548	-	intergenic:RNU6-99P-AC027806.1
GCCAGACTGATCACACACTGAGG	chr22	34094162	34094184	+	intron:LL22NC03-86D4.1
GCCACACTGAGCAGACTCCTTGG	chr1	25652363	25652385	-	intergenic:RP1-187B23.1-MAN1C1
GGGAGACTGAGCAGACCCACTGG	chr2	29356675	29356697	+	intron:ALK
GGGAGACTGGGCAGACAGAGTGG	chr2	23414287	23414309	+	intron:KLHL29
GCCAGACTGTGCACACAGCGTGG	chrX	39462633	39462655	+	intergenic:RP11-157D23.2-MIR3937
GCGAGACTTGGGAGACACGGGGG	chr19	12657414	12657436	-	intron:MAN2B1
GCCAGAGTGAGCAGTCACAGGGG	chr18	44509004	44509026	+	intron:LINC01478
GAGAGACTGAGCATACCCAGGGG	chr9	89614806	89614828	-	intergenic:GADD45G-UNQ6494
TAGAGACTGAGCAGACAGGGTGG	chr17	56448456	56448478	+	intron:ANKFN1
GCGAGACTCCGCCGCCACCGGGG	chr1	1231729	1231751	-	intron:SDF4
GCCAGACTGAGGAGACCCGGCGG	chr14	1.05E+08	1.05E+08	+	intergenic:RP11-477I4.4-LINC00638