University of Southampton

FACULTY OF MEDICINE

HUMAN DEVELOPMENT AND HEALTH

IMMUNO-GENOMIC PROFILING OF PAEDIATRIC INFLAMMATORY BOWEL DISEASE

by

Tracy Francisco Antonio Coelho

Thesis for the degree of Doctor of Philosophy

[September, 2019]

Supervisory Team:

Prof Sarah Ennis, Prof Anthony P Williams, Prof R.M. Beattie, Dr Yifang Gao

TABLE OF CONTENTS

SUPPLEMENTARY MATERIAL

LIST OF TABLES

LIST OF FIGURES

Research Thesis: Declaration of Authorship

I, **Tracy Francisco Antonio Coelho** declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Thesis title: Immuno-genomic Profiling of Paediatric Inflammatory Bowel Disease

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:

Coelho T, Andreoletti G, Ashton JJ, et al. Immuno-genomic profiling of patients with inflammatory bowel disease: a systematic review of genetic and functional in vivo studies of implicated genes. Inflamm Bowel Dis 2014;20:1813-9.

Coelho T, Andreoletti G, Ashton JJ, et al. Genes implicated in thiopurine-induced toxicity: Comparing TPMT enzyme activity with clinical phenotype and exome data in a paediatric IBD cohort. Sci Rep 2016;6:34658.

Signature: Tracy A F Coelho Date: 04/10/2019

PUBLICATIONS

1. **Coelho T**, Andreoletti G, Ashton JJ, Batra A, Afzal NA, Beattie, Ennis S. Immuno-genomic profiling of patients with inflammatory bowel disease: a systematic review of genetic and functional in vivo studies of implicated genes. Inflammatory Bowel Disease. 2014.
2. Andreoletti G, Ashton JJ, **Coelho T**, Willis C, Haggarty R, Gibson J, et al. Exome analysis of patients with concurrent pediatric inflammatory bowel disease and autoimmune disease. Inflamm Bowel Dis. 2015;21(6):1229-36.
3. Ashton JJ, **Coelho T**, Ennis S, Batra A, Afzal NA, Beattie RM. Presenting phenotype of paediatric inflammatory bowel disease in Wessex, Southern England 2010-2013. Acta Paediatr 2015;104:831-7.
4. **Coelho T**, Andreoletti G, Ashton JJ, Batra A, Afzal NA, Gao Y, Williams A.P, Beattie, Ennis S. Genes implicated in thiopurine-induced toxicity: Comparing TPMT enzyme activity with clinical phenotype and exome data in a paediatric IBD cohort. Scientific Reports. 2016.
5. Ashton JJ, **Coelho T**, Ennis S, et al. Endoscopic Versus Histological Disease Extent at Presentation of Paediatric Inflammatory Bowel Disease. J Pediatr Gastroenterol Nutr 2016;62:246-51.
6. Ashton JJ, Andreoletti G, **Coelho T**, Haggarty R, Batra A, Afzal NA, et al. Identification of Variants in Genes Associated with Single-gene Inflammatory Bowel Disease by Whole-exome Sequencing. Inflamm Bowel Dis. 2016;22(10):2317-27.
7. Ashton JJ, Colquhoun CM, Cleary DW, **Coelho T**, Haggarty R, Mulder I, et al. 16S sequencing and functional analysis of the fecal microbiome during treatment of newly diagnosed pediatric inflammatory bowel disease. Medicine (Baltimore) 2017;96:e7347.
8. Andreoletti G, Shakhnovich V, Christenson K, **Coelho T**, Haggarty R, Afzal NA, et al. Exome Analysis of Rare and Common Variants within the NOD Signaling Pathway. Sci Rep 2017;7:46454.
9. Edwards A, Gao Y, Allan RN, Ball D, de Graaf H, **Coelho T**, et al. Corticosteroids and infliximab impair the performance of interferon-gamma release assays used for diagnosis of latent tuberculosis. Thorax 2017;72:946-9.
10. Ashton JJ, Cullen M, Afzal NA, **Coelho T**, Batra A, Beattie RM. Is the incidence of paediatric inflammatory bowel disease still increasing? Arch Dis Child 2018;103:1093-4.
11. Ashton JJ, Borca F, Mossotto E, **Coelho T**, Batra A, Afzal NA, et al. Increased prevalence of anti-TNF therapy in paediatric inflammatory bowel disease is associated with a decline in surgical resections during childhood. Aliment Pharmacol Ther. 2019;49(4):398-407.

Oral Presentations:

1. Genetics of IBD-PhD. Invited speaker. Annual Winter Meeting BSPGHAN. Stratford-upon-Avon. 2015.
2. Functional clustering of paediatric IBD patients: hypo-inflammatory immune signatures of innate signalling in a treatment naïve cohort. Annual ESPGHAN meeting, Geneva. 2018.

ACKNOWLEDGEMENTS

Now that the end of this long and arduous journey seems to be in sight, it's time to thank all the people who have made it possible for me to get to this point.

First of all, I would like to thank all my supervisors, Profs Sarah Ennis, Mark Beattie, Tony Williams and Yifang Gao for all their hard work, patience, support and valuable advice during my time as a PhD student. I am very lucky indeed to have supervisors with such a fantastic mix of skills, disciplines and exceptional wisdom. I am ever so grateful to Mark for introducing me to the world of academia, encouraging me to sign up for the PhD and for being a constant source of motivation and support. I admire Sarah for her hard work, discipline, dedication and her supervisory skills. I have certainly learnt a lot working with her over the last few years. Tony's wisdom in immunology has no boundaries and I hope I have helped myself sufficiently to his intellectual generosity. Yifang has been such a wonderful person and friend through all these years. I thank her for all the support, help and a reassuring approach, especially when the experiments in the laboratory did not go as per plan.

I am hugely indebted to my funding body, Crohn's in Childhood Research Association (CICRA), for giving me this life time opportunity to do research. I thank all our volunteers, patients and their families for their understanding, eagerness and participation in this research.

I thank Rachel Haggarty, senior research nurse for her unfailing and consistent support all through my PhD. Her admirable qualities including kindness, politeness, genuine sincerity and care towards patients and their families has been an indispensable asset to the research study. Special thanks to my clinical colleague and friend, Dr Akshay Batra who has been such a wonderful person, always reassuring me during stressful times and lending a patient ear to all frustrations and difficulties that I experienced over the last few years. I wish to thank Dr James Ashton for his help during this research, including recruitment, advice, publications and support. I also wish to extend my gratitude to Dr Nadeem Afzal for his esteemed opinions, support and advice. I thank my co-PhD colleagues Gaia Andreoletti and Enrico Mossotto whole heartedly, who have worked with me patiently and contributed to this research. I further acknowledge the fruitful collaboration with Prof Dieter Riethmacher and Dr Eva Riethmacher, which added another dimension to the research done for this thesis. I wish to thank Dr Kathy Porter and her team from the tissue bank (particularly Jenna Watt), Nikki Graham from the DNA laboratory and Dr Annie Griffins from the WISH laboratory. I would also like to thank Javier Garcia-Frutos for his friendly approach and admin support.

Unfortunately, there are too many people who have helped me complete my research and PhD that I cannot name them all. To everyone I haven't named, I want to send my heartfelt gratitude for your help and support.

Lastly, but most importantly, I would like to thank my wife Rochelle for her never ending love, support and patience to get me through this PhD. I am also thankful to my little boys Rhys and Zach who have been ever so understanding and patient through this whole journey. From now on, I will be able to spend more quality time with my family.

FUNDING

This research was supported by Crohn's in Childhood Research Association (**CICRA**).

Abbreviations

AHC	Adult healthy controls
ALT	Alanine transaminase
AJ	Adherens junctions
ANNOVAR	Annotate Variation
5-ASA	5-aminosalicyclic acid
ATG16L1	Autophagy-related 16-like 1
AZA	Azathioprine
BAFF	B-cell activating factor
BAM	Binary alignment/map
BWA	Burrows-Wheeler Aligner
BWT	Burrows-Wheeler transformation
CADD	Combined Annotation Dependent Depletion
CARD	Caspase activation and recruitment domain
CD	Crohn's disease
CIAP	Cellular inhibitors of apoptosis
CNV	Copy number variation
CRP	C-reactive protein
DAMPs	Damage-associated molecular patterns
DANN	Deleterious annotation of genetic variants using neural networks
DAPPLE	Disease Association Protein-Protein Link Evaluator
DC	Dendritic cell
DNA	Deoxyribonucleic acid
ECM	Extra-cellular matrix
EDTA	Ethylenediaminetetraacetic acid
EEN	Exclusive enteral nutrition
ELISA	Enzyme-linked immunosorbent assay
eQTL	Expression quantitative trait locus
ESR	Erythrocyte sedimentation rate
EVS	Exome Variant Server
FAK	Focal adhesion kinase
FMT	Faecal microbial transplant
GATK	Genome Analysis Toolkit
GCLP	Good Clinical Laboratory Practice
GERP	Genomic Evolutionary Rate Profiling
GGT	Gamma-glutamyl transpeptidase
GI	Gastro-intestinal
GOF	Gain of function
GRAIL	Gene Relationships Across Implicated Loci
GWAS	Genome wide association studies
HUGO	Human genome organisation
HGNC	HUGO Gene Nomenclature Committee
HC	Hierarchical clustering

IBD	Inflammatory bowel disease
IBDU	IBD unclassified
IKK	Inhibitor of kappa B kinase
ILC	Innate lymphoid cells
IL-6	Interleukin-6
IL-10	Interleukin-10
IL-1β	Interleukin-1 beta
IQR	Interquartile range
JAKs	Janus kinases
JAMs	Junctional adhesion molecules
LD	Linkage disequilibrium
LOD	Logarithm of the odds
LPS	Lipopolysaccharide
LRR	Leucine-rich repeat
MAF	Minor allele frequency
MAPK	Mitogen-activated protein kinases
MDP	Muramyl dipeptide
MFI	Median fluorescence intensity
MHC	Major histocompatibility complex
6-MP	6-Mercaptopurine
MSCs	Mesenchymal stromal cells
MOCOS	Molybdenum cofactor sulfurase
NF-kB	Nuclear factor kappa B
NGS	Next generation sequencing
NK cells	Natural killer cells
NOD	Nucleotide-binding oligomerization domain
NRES	National Research Ethics Service
OCT	Optimal Cutting Temperature
PAMPs	Pathogen associated molecular patterns
Pam3CSK4	Synthetic tri-acylated lipoprotein
PBMC	Peripheral blood mononuclear cells
PCA	Principal component analysis
PCDAI	Paediatric Crohn's Disease Activity Index
PHC	Paediatric healthy controls
PIBD	Paediatric IBD
POSTN	Periostin (gene)
PRR	Pathogen recognition receptor
PUCAI	Paediatric Ulcerative Colitis Activity Index
RIP2	Receptor interacting response protein 2
RNA	Ribonucleic acid
SAM	Sequence alignment/map
SD	Standard deviation
SEM	Standard error of the mean
SKAT-O	Sequence kernel association optimal unified test
SNP	Single nucleotide polymorphism

SOP	Standard operating procedure
STAT	Signal transducer and activator of transcription
TBNK cells	T cells, B cells, Natural killer cells
TGF-β	Transforming growth factor-beta
6-TGN	6-Thioguanine nucleotide
TJ	Tight junction
TLR	Toll-like receptor
TNF-α	Tumour necrosis factor- alpha
TPMT	Thiopurine S-methyl transferase
T-reg	Regulatory T cells
Tri-DAP	L-Ala-g-D-Glu-mDAP
UHS	University Hospital Southampton
USP21	Ubiquitin-specific protease 21 gene
UC	Ulcerative colitis
VCF	Variant call format
WBC	White blood cell
WES	Whole exome sequencing
WGS	Whole genome sequencing
WISH	Wessex Immunology Sciences Hub
WTCCC	The Wellcome Trust Case Control Consortium
XIAP	X-linked inhibitor of apoptosis
ZO	Zona occludentes

ABSTRACT

INTRODUCTION

Inflammatory bowel disease is a complex polygenic disease with a multi-factorial aetiology. Several immune signalling pathways regulated by multiple genes have been implicated in the pathogenesis of the disease. The current categorisation of IBD is based on the clinical phenotype which is blind to the immunological phenotype. The central aim of this project was to assess the functional integrity of key immune signalling pathways in paediatric patients with IBD through immunological assays and examine the relationship between altered immune function and patient-specific genetic mutations. The secondary aim of this project was to explore the potential for patient stratification based on their immuno-genomic profiles in order to facilitate the application of targeted therapeutics in IBD.

HYPOTHESES

1. Of the multiple immune pathways implicated and confirmed by genetic studies, individual patients present with compromised function of one or more specific immune signalling pathways.
2. Aberrant immune pathway function identified by immunoassays can be correlated with individual mutation profiles across genes within these pathways.
3. Integrated immuno-genomic profiling can stratify patients into novel subtypes that will permit better mechanistic and therapeutic evaluation, and facilitate targeting of therapies to molecularly stratified patients.

OVERVIEW OF METHODS

In this thesis, a multiplex-based immunological assay was developed and optimised to assess the molecular integrity of biological pathways implicated in IBD. Based on preliminary experiments assessing the suitability of different cell activation systems, extensively published literature and practicalities, cryopreserved (frozen) PBMCs were used for conducting immuno-assays. A standard operating procedure for conducting assays was developed following optimisation experiments in control samples. The optimised assay was conducted in a treatment naïve paediatric IBD cohort of twenty-two patients alongside ten adult and ten paediatric controls. The cytokine data generated were subjected to hierarchical clustering, an unsupervised machine-learning application to assess patient stratification based on immunological profiles. Clusters generated were then interrogated

against exome sequencing data to identify genomic signals which may be contributing to the clustering pattern.

In addition to the immunological assays investigating innate immune signalling in IBD, the research work presented in this thesis includes the assessment of periostin protein as a potential biomarker of disease activity, immuno-histochemistry of periostin expression in the GI mucosa and the interrogation of exome sequencing data in the context of clinical outcomes and plasma levels of periostin.

The thesis also includes the applicability of genomic analysis to assess thiopurine-induced drug toxicity in paediatric patients with IBD.

MAIN RESULTS

A systematic review of published literature across a panel of genes implicated in IBD, conducted at the start of my PhD project showed a paucity of immuno-genomic studies, thereby signposting a clear direction for my research integrating immunological work-up with exome sequencing data.

Following extensive dose-finding and optimisation experiments conducted in healthy volunteers, a standard operating procedure was established using cryopreserved PBMCs. In the optimised assay, PBMCs were selectively stimulated for NOD2, TLR1-2 and TLR4-mediated pathways using MDP, Pam3CSK4 and LPS respectively. Effector responses of IL-10, IL-1β, IL-6 and TNF-α were assessed through multiplex assays to evaluate corresponding NF-κB, MAPK and NLRP3-inflammasome pathways.

A significant difference in immune induction was observed between the paediatric and adult healthy controls for two out of the twelve cytokine responses (TLR4-induced IL-10 and TNF-α). Therefore, only age-matched paediatric controls were used in the final analysis to interpret biological variations in paediatric patients. Combined innate immune responses in patients across twelve effector responses were significantly reduced compared to paediatric controls ($p=0.003$) and were driven primarily by 'hypofunctional' TLR responses. Application of hierarchical clustering to the immunological data generated eight distinct patient clusters based on their induced immune profile. Further interrogation of the clustering pattern using a whole gene pathogenicity score 'GenePy', identified a statistical excess of mutations in the TNF-α signalling pathway in one immune profile-based cluster, with a significant contribution from the *USP21* gene.

In the periostin study, an inverse correlation was observed between plasma levels of periostin and disease activity for Crohn's disease, with significantly higher levels observed during remission rather than active disease, suggesting a more prominent reparative role rather than inflammatory for this protein in IBD. The utility of plasma periostin as a biomarker of inflammation in IBD could not established as there were no significant differences in the plasma levels of this protein observed between patients and paediatric controls.

In the study investigating thiopurine drug-induced toxicity, a highly pathogenic novel variant was identified in the *TPMT* gene, demonstrating the strength of next generation sequencing as a powerful tool in identifying pathogenic variants, not detected through standard genotyping used for predicting thiopurine drug toxicity. A significant association was observed between the *MOCOS* gene and TPMT biochemical enzyme activity, highlighting the importance of genes other than *TPMT* in thiopurine-induced toxicity. Although NGS has the ability to detect rare or novel variants in the *TPMT* and other implicated genes, there is no clear advantage of genomic testing over the biochemical test in predicting toxicity to thiopurine drugs.

CONCLUSIONS

This research supports an immunological profiling led genomic analysis approach to a complex polygenic disorder. It provides a rationale for patient stratification based on their immuno-genomic profiles, signposting new mechanistic insights with a futuristic goal of personalised therapy. Currently, there is an unmet need for novel groupings in IBD based on the immunological profiles of individual patients. However, with the emergence of robust high throughput technologies coupled with the integration of multi-omic data, patient-stratification based on the molecular patterns of the disease looks increasingly possible. Judicious integration of multi-omic data will pave the way for personalised treatment approaches in IBD, leading to improved health outcomes and reducing the morbidity associated with this debilitating condition.

THESIS OUTLINE

The thesis is structured into 4 sections followed by final conclusions and summary.

Section I includes a general introduction to inflammatory bowel disease and an overview of the developments in understanding the genetic and immunological basis of the disease.

Section II (Chapters 2-5)

- **Chapter 2** is a systematic review of literature (published in 2014), detailing functional studies in genotyped individuals with IBD across a panel of implicated genes. The systematic review aimed at exploring the strengths of integrating genomics with functional studies, which is the central theme of my project.
- **Chapter 3** details the existing research setup, patient recruitment, procurement of biological specimens, generic methods for processing of samples for immunological assays, whole exome sequencing and the basic pathway for the computational analysis of raw sequenced data.
- **Chapter 4** discusses the optimisation of immunological methods in healthy controls and the development of a standard operating procedure (SOP) for conducting immunological assays in patients.
- **Chapter 5** describes immunological assays conducted on a treatment naïve paediatric IBD cohort and non-IBD controls. This chapter finally integrates the cytokine profiles observed in patients through immunological assays with the exome data, aiming to understand the relationship between altered immune function and patient-specific genetic mutations.

Section III includes **chapter 6**, focusing on the role of periostin protein in IBD. This chapter assesses the utility of periostin protein as a biomarker of disease activity in IBD, examines the protein expression in colonic tissue in IBD and interrogates genetic mutations against the plasma levels and surgical outcomes in paediatric patients with IBD.

Section IV includes **chapter 7**, examining the utility of next-generation sequencing (NGS) for predicting toxicity and clinical response to thiopurine drugs in paediatric patients with IBD. This work, which was published in 2016, investigates the genes implicated in thiopurine toxicity against biochemically-measured levels of the drug-metabolising enzyme (thiopurine methyl-transferase, TPMT), clinical response to the drug and adverse effects. The research focuses on the *TPMT* gene and a panel of 15 other genes implicated in thiopurine toxicity.

Section V includes the final conclusions and future direction.

Section I

Chapter 1 THE IMMUNOLOGY AND GENETICS OF INFLAMMATORY BOWEL DISEASE

This chapter presents an introduction to inflammatory bowel disease (IBD) followed by an outline of the normal immune mechanisms operating within the bowel mucosa. The chapter details the complex immunological pathways and mechanisms implicated in the pathogenesis of IBD, and provides an overview of the genetic advances, which have led to a better understanding of the molecular basis of IBD. This chapter also includes a description of the overarching influence of environmental factors, modulating immune responses in genetically susceptible individuals.

1.1 INTRODUCTION TO INFLAMMATORY BOWEL DISEASE

Inflammatory bowel disease (IBD) is a chronic inflammatory condition with a complex multifactorial pathogenesis. The two distinct phenotypes of IBD include Crohn's disease (CD) and ulcerative colitis (UC), which can present with a diverse but quite often overlapping signs and symptoms. Due to overlapping features between the two phenotypes, a clear distinction is not always possible in which case the term IBDU (IBD unclassified) is used[1,2]. As IBD manifests as a chronic relapsing condition with a considerable morbidity, it can have a major impact on the individual's quality of life and ability to work. Crohn's disease can affect any part of the gastro-intestinal tract from the mouth to the anus and generally has a segmental distribution with areas of macroscopically unaffected mucosa. UC usually involves the rectum and the inflammation extends proximally in a contiguous fashion to involve the colon. Inflammatory changes seen in Crohn's disease can involve all the layers of the bowel wall whereas in UC the inflammation is typically restricted to the mucosa.

Although the precise mechanisms underlying the pathogenesis of IBD are not fully understood, dysregulated mucosal immune responses to the gut microorganisms in genetically predisposed individuals have been shown to be the key operating mechanisms[3,4]. The GI tract plays an important role in maintaining immune tolerance to the commensal bacteria, self-antigens and the allergenic food proteins through a finely balanced network of immune checkpoints effected through innate and adaptive immunity, and epithelial barrier mechanisms. Dysregulation of these immune checkpoints can result

in altered immune responses with the bowel mucosa resulting in persistent epithelial injury, ulceration and other inflammatory changes seen in IBD.

Similar to other chronic immune-mediated conditions, IBD has been classified as a prototypical complex polygenic disease, in which biological heterogeneity arises from intricate interactions between the host and the environment[5]. Despite the major advances over the past two decades, disease heterogeneity creates enormous challenges for researchers and clinicians, leading to treatment dilemmas, potentially preventable disease-associated morbidity and several unaddressed clinical and research questions. The current binary categorisation of IBD into CD and UC although clinically useful, is historical and arbitrary based on descriptive characteristics including the clinical presentation, endoscopic and histological findings[6,7]. Whilst there are defined diagnostic criteria for CD and UC, there is frequently a histological overlap and subsequent re-classification of the disease sub-type particularly in children[8]. There is also a significant overlap in the genetic susceptibility to IBD, with patients frequently having both disease sub-types in their family history[9]. Furthermore, diverse genetic mutations can present with similar clinical phenotypes adding further challenges to the understanding of the genetic nature of the disease[10].

As per the Paris classification[11], paediatric onset IBD includes patients developing the disease before age 17 years. Paediatric onset IBD differs from adult onset IBD in many aspects including the disease type, location, disease course, response to treatment and genetically attributable risks[12]. A key feature of paediatric IBD is the impact of disease on growth and puberty[13]. There is a predominance of pan-enteric disease in paediatric patients compared to ileo-caecal involvement seen commonly in adults. Upper GI disease is also more common in children compared to adult onset disease[12,13]. From the genetic perspective, although there is no difference in the common risk variants across the various age groups, patients developing the disease in childhood are more likely to harbour rare variants with high penetrance as well as a higher burden of common risk variants[14,15]. Based on the understanding that monogenic forms of IBD present with a severe phenotype in the early years of life, several sub-categories have been suggested[14]. These include early onset IBD (<10 years, EOIBD), very early onset IBD (<6 years, VEOIBD), infantile IBD (<2 years) and neonatal IBD (<28 days)[14]. Although this sub-classification is useful from a clinical perspective, it does not provide a robust platform to guide diagnostics and an individualised treatment approach[14,16].

A better disease classification based on the underlying immunological and genetic phenotype would not only enhance our understanding of the disease but also lead to identification of sub-groups of patients who would benefit from tailored therapy and interventions. With the emergence of innovative and transforming tools for analysis of big data or multi-omic data, stratification of distinct groups of patients (also called pheno-mapping) is increasingly applied, especially in the fields of oncology, cardiology and diabetes[17,18]. The application and integration of multi-omic analysis including genomics, transcriptomics, proteomics, metabolomics in IBD is also gathering pace. Generation of reliable predictive models of disease course, clinical response to treatment, drug safety profile and complications risks through a judicious interpretation of multi-omic data holds promise for futuristic implementation of personalised medicine in IBD[5,19].

1.1.1 Incidences and geographical trends

The incidence rates of paediatric IBD (pIBD) vary between various populations ranging from 0.3 to 11.4/100000/year with the highest rates reported in North America, Scandinavia and Australasia[20]. The variability in geographical trends reflects the multifactorial nature of the disease, including population genetics and environmental factors. The incidence of IBD has increased sharply over the recent years in all age groups in most Western populations[21 22]. A systematic review in 2011 on the epidemiology of pIBD looking at 139 studies reported an increasing trend in 77% of the studies analysed[20]. Approximately 25% of IBD cases present in childhood or adolescence[23]. In the paediatric age group, approximately 4% of patients present before age 5 years and 18% before age 10 years, with a peak incidence in adolescence[15]. Studies have shown that the increased incidence of pIBD is paralleled by a proportionate increase in the incidence of adult IBD[24]. Scottish data over the last 40 years has shown a consistent rise in the incidence with a 76% increase since the mid-1990s[25]. A study published in 2014 demonstrated an increased incidence of pIBD of almost 50% in the Wessex region of Southern England over the last decade[26]. The reasons for this increase are not entirely clear. Although it may partly be due to earlier diagnosis and improved referral patterns, it is likely that it does represent a true increase in incidence in keeping with the overall global trends[26].

1.1.2 Family history

There is an increased risk of developing IBD among family members, with 5-23% of patients having a first degree relative affected by the disease[27]. The familial nature of IBD

was recognised more than fifty years ago, based on early epidemiological data, including difference in prevalence figures among different ethnicities, familial clustering of cases and concordance among monozygotic twins[27-29]. A higher proportion of patients with CD have a family history of IBD compared to patients with UC[30]. The concordance rates in CD are up to 60% among monozygotic twins compared to up to 20% in UC. Among dizygotic twins, the concordance rates are 12% and 5% in CD and UC respectively[28,31]. As will be discussed in the subsequent sections, the early observations of familial aggregation and findings of observed concordance in twin studies, established a clear genetic disease component and paved the way to a successful era of molecular genetics.

1.1.3 Clinical presentation

The classical presentation of Crohn's disease in childhood includes a constellation of abdominal pain, diarrhoea, poor appetite and weight loss. Abdominal pain is the most common single symptom at presentation and is often periumbilical but may localise to the right lower quadrant or diffusely to the lower abdomen in colonic disease. Other clinical features include blood in stools, impairment of linear growth and puberty, perianal lesions and mouth ulcers. Compared to ulcerative colitis, the initial symptoms of Crohn's disease can be subtle and variable given the diffuse and diverse anatomic localisation of the disease. The classical symptoms of UC and IBDU include chronic bloody diarrhoea, abdominal pain and tenesmus. The period of illness prior to presentation is relatively short (median of 3-4 months) in UC compared to CD (6 months), presumably due to the presence of blood in stools, resulting in earlier presentation to healthcare[32]. Extra-intestinal manifestations of IBD are seen in both CD and UC. Most commonly involved organs include the joints (arthralgias & arthritis), skin (erythema nodosum, pyoderma gangrenosum), ocular (episcleritis, uveitis and rarely orbital myositis), hepato-biliary (primary sclerosing cholangitis) and the renal system[33,34].

1.1.4 Overview of treatment

The broad approaches to treatment in IBD are induction of remission followed by maintenance therapy by employing a multi-pronged approach including nutritional, pharmacological and where appropriate surgical interventions. Treatments used in IBD vary depending on the age group, extent and severity of the disease. The general principles underpinning management of this complex condition include induction of a sustained disease remission by down regulation of inflammatory responses in order to control

symptoms, normalising growth and pubertal development, promoting normal social development and avoiding long term disease related complications[35]. Figure **1.1** provides a broad overview of the treatments used in paediatric IBD.

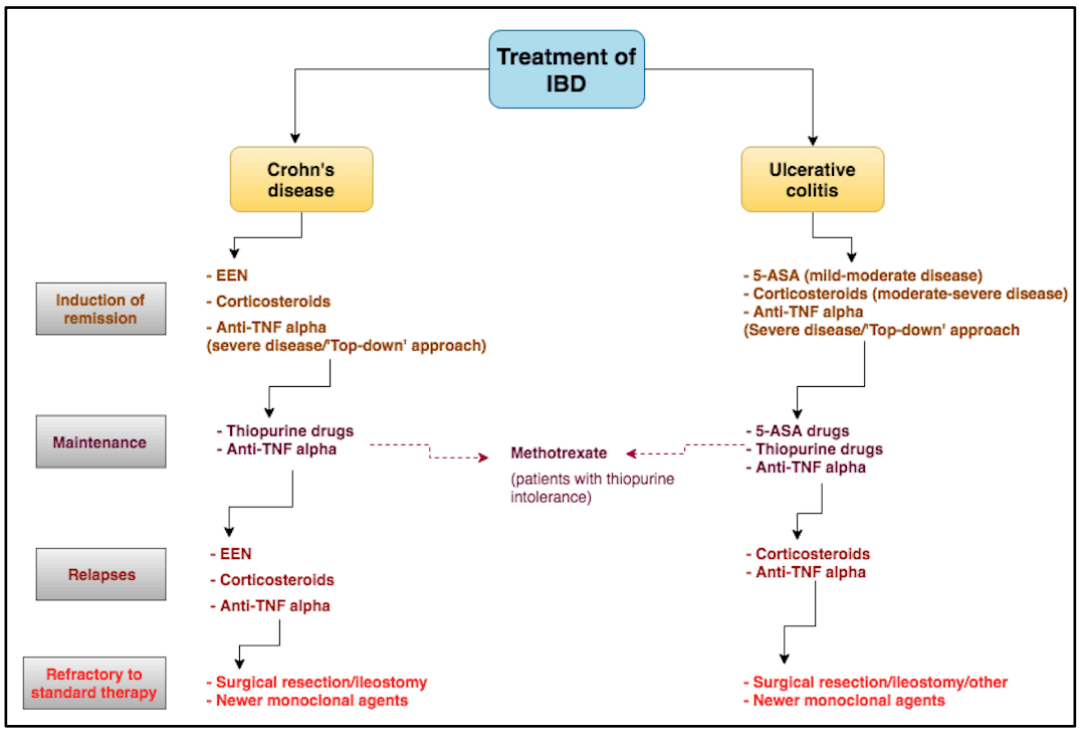


Figure 1.1 Treatment approach in Paediatric IBD

*The figure outlines the broad principles of treatment in paediatric IBD. There are some differences in the management of Crohn's disease (CD) compared to ulcerative colitis (UC) such as the use of exclusive enteral nutrition (EEN) in CD, but not UC. The treatment algorithm includes induction of remission, maintenance therapy, treatment of relapses and management of cases showing poor response to standard therapy. (**Abbreviations**: Anti-TNF, anti-tumour necrosis factor alpha, 5 ASA, 5-aminosalicyclic acid). Adapted from ECCO/ESPGHAN guidelines on the management of Crohn's disease and ulcerative colitis.*

The initial goal of treatment is safe and rapid induction of remission with medications and nutritional support. For paediatric CD, exclusive enteral nutrition (EEN) is the first line of treatment, taken in the form of liquid drinks for a minimum of six weeks with complete exclusion of other foods. The remission rates with EEN are approximately 80% in appropriately selected cases[36]. Treatment with intravenous or oral corticosteroids is considered in individuals not responding to EEN, non-compliant patients and in those unable to tolerate EEN for various reasons including stricturing disease involving the small bowel[37].

Thiopurines drugs including azathioprine or 6-mercaptopurine are commonly used as maintenance therapy for patients with CD and UC[38]. As per current guidelines, the

indications for initiating treatment with thiopurines include patients who relapse within six months of diagnosis, following two or more courses of steroids in twelve months, extensive disease at diagnosis, steroid-dependent patients and to enhance the efficacy of anti-TNF alpha agents by reducing the immunogenicity of biological agents[37,39]. Methotrexate, a dihydrofolate reductase (DHFR) inhibitor can be used as maintenance therapy in patients intolerant to thiopurines[37]. Biological agents such as anti-TNF alpha antagonists (infliximab, adalimumab and their biosimilars) are widely used as a step-up approach for patients failing to respond to steroids, frequently relapsing disease, patients with stricturing or penetrating disease and severe peri-anal manifestations[37]. In recent years, there has been increasing evidence of the benefits of 'top-down' approach with the early use of anti-TNF agents to induce early histological remission, reduce disease progression and IBD-related complications[40,41]. Although some centres are now routinely adopting a top-down strategy, there are limited long-term data on its benefits and safety[19,41]. Newer classes of biologics including vedolizumab ($\alpha4\beta7$ integrin antagonist) and ustekinumab (anti-IL12/IL23) can be considered in patients with IBD refractory to anti-TNF alpha therapy, although there are limited data on its efficacy and safety in the paediatric age group[42-44].

In UC, the mainstay of treatment is corticosteroids for inducing remission in moderate-severe disease and 5-aminosalicylates (5-ASA) such as mesalazine for mild-moderate disease. Maintenance of remission requires the long-term use of 5-ASA drugs either alone or in combination with thiopurines. Anti-TNF alpha agents are also widely used in UC patients failing to respond to first line therapy[42,43].

Despite the remarkable responses to established treatments such as anti-TNF alpha agents in some patients, primary resistance and inadequate treatment responses continue to pose significant problems in up to 40% of patients[45,46]. Current treatment approaches in IBD are largely guided by clinical phenotypes, disease progression and course, with no clinical emphasis to understand the true immunological basis of the disease in individual patients. Treatment failures result in escalation of treatment strategies, often necessitating invasive surgical procedures, with consequential life-long morbidity[9]. A better understanding of the molecular pathogenesis of the disease on an individualised basis would result in better clinical outcomes. Clinicians have become increasingly aware of the need to characterise and classify the disease phenotype and where possible assess the genetic and immunological factors particularly in the younger child, thereby establishing a framework for evaluation of emerging therapies in stratified groups[47]. Targeting specific checkpoints

of immune dysregulation in individual patients or subgroups to provide tailored treatment has been increasingly recognised, offering better theragnostic precision, minimising adverse side-effects and improving the overall health outcomes[9,48].

1.2 An outline of mucosal immuno-physiology in health

The human intestinal tract covers a very large surface area of approximately 200-400 square metres, harbouring an estimated hundred trillion (10^{14}) microbes. Whilst these are largely commensal bacteria performing numerous metabolic functions essential to the host, pathogenic bacteria that invade the mucosal environment pose a constant threat of invasive disease[49,50]. The GI mucosa has a complex, but a tightly regulated immune system shaped by millions of years of co-evolution between the host and approximately one thousand or more distinct microbial species[49,51]. The GI immune system has evolved several tiers of defence to prevent translocation of microbial elements across the bowel wall thereby maintaining a delicate and dynamic balance between tolerance of innocuity and vigilance of microbial threats. The intestinal epithelium comprising a single layer of tightly connected columnar cells along with the overlying mucus represents the first line of immune defence and provides a physical barrier separating the essentially sterile host from the intestinal microorganisms[50]. The epithelium consists of specialized cells: the enterocytes which are the absorptive cells, goblet cells, neuro-endocrine cells, Paneth cells and Microfold cells (M cells), all of which are derived from common precursor stem cells residing in the crypts**.** The Paneth cells and Microfold cells (M cells), which along with other immune cells within the sub-epithelial lamina propria such as the dendritic cells, macrophages, granulocytes and innate lymphoid cells constitute the innate immune system of the GI mucosa. Goblet cells secrete glycosylated mucins that form a mucus matrix over the epithelium thereby forming the first level of defence against any microbial invasion. Figure **1.2.**

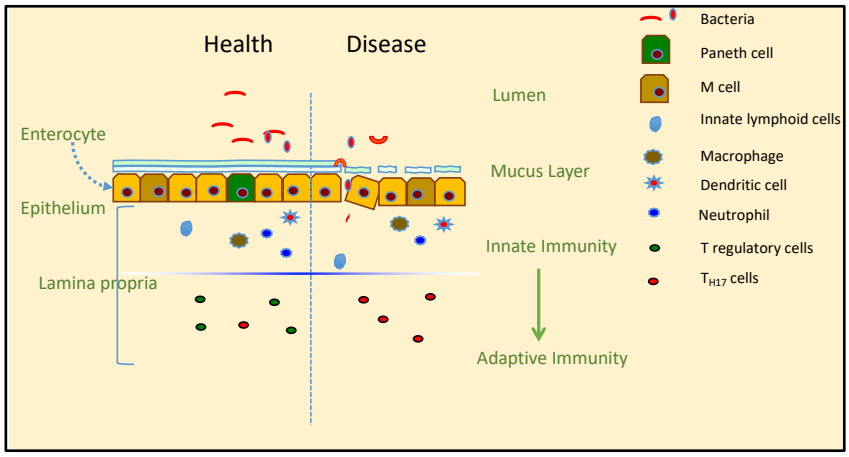


Figure 1.2 Intestinal epithelial barrier and other elements of defence

The small bowel mucosa is lined by columnar epithelium with an overlying mucus layer and an underlying sub-epithelial lamina propria. The epithelial layer comprises the enterocytes, goblet cells, M cells, Paneth cells and the stem cells. Various types of immune cells can be found throughout the lamina propria including macrophages, various types of T-cells, dendritic cells, innate lymphoid cells, neutrophils and others. Bacterial antigens detected by cells of the innate line of defence (Paneth cells, M cells and others) initiate and orchestrate immune responses by releasing extra-cellular mediators that recruit other immune cells including those of the adaptive immune system.

As part of routine vigilance, the Paneth cells and the M cells interact with the luminal microbes, continually sampling microbial proteins via specialised transmembrane or cytosolic receptors, called pathogen recognition receptors (PRRs). The microbial proteins are transported across these cells to the sub-epithelial space through a process known as transcytosis for further scrutiny by the other cells of the innate immune system[52]. The PRRs on immune cells have the ability to detect and recognise specific microbial products known as pathogen associated molecular patterns (PAMPs)[53]. Following detection of PAMPs, the receptors undergo conformational changes recruiting adaptor proteins and downstream signalling, resulting in the activation of inflammatory cascades and production of mediators of inflammation such as cytokines and chemokines. If the initial defence provided by the epithelial barrier is breached or evaded by a pathogenic organism, the first cells that respond include the phagocytic granulocytes and macrophages, which will ingest and kill microbes through a variety of mechanisms. In contrast to the phagocytic cells, the dendritic cells play a key role to take up transported proteins and process them further for presentation to the adaptive immune system. Figure **1.3.**

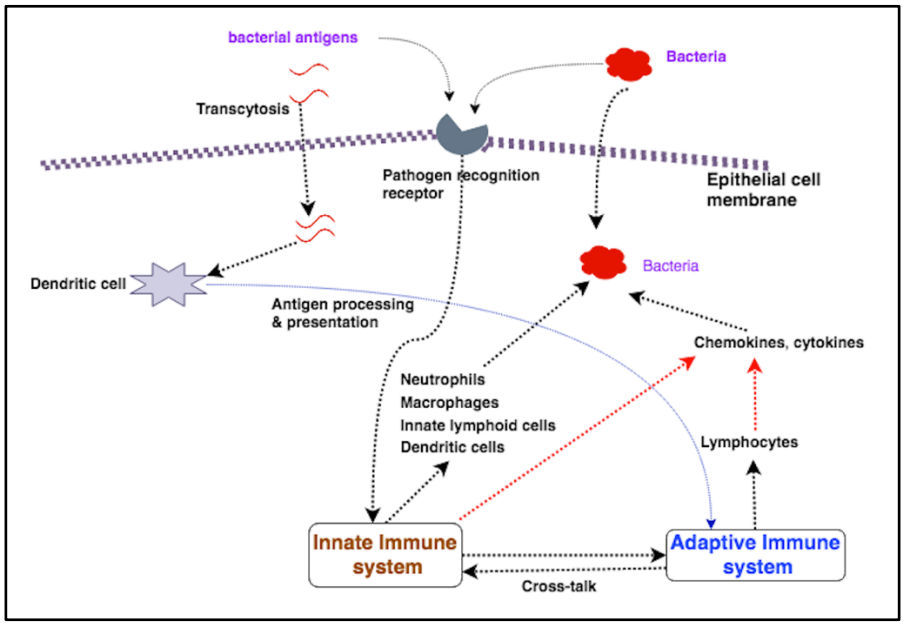


Figure 1.3 Mucosal immune mechanisms

Pathogen recognition receptors (PRRs) can detect bacterial antigens and other pathogen associated molecular patterns (PAMPs), resulting in recruitment of innate immune cells such as the neutrophils, macrophages, dendritic cells and innate lymphoid cells. Innate immune responses can result in phagocytosis and killing of the microorganism which have breached the epithelial barrier. Activation of innate signalling pathways results in production of pro-inflammatory cytokines. Innate immune pathways closely interact with adaptive immune mechanisms for further orchestration of the immune responses. Additionally, microbial antigens breaching the epithelial cell wall via transcytosis are processed by dendritic cells for presentation to the adaptive immune system. Adaptive immune responses are mediated primarily via lymphocytes through production of cytokines and chemokines.

The adaptive immune responses are induced when the innate defence mechanisms are overwhelmed by invading pathogens. The adaptive immune system interacts closely with the mediators and effectors of the innate immune system for a well-coordinated and effective immune response[54]. Adaptive immune responses are mediated through production of chemokines and cytokines, primarily via $CD4^+$ T-helper lymphocytes (Th cells or helper T cells)[55]. Historically, two main types of helper T cells have been described, Th1 and Th2 cells. Th1 cells offer protection against infectious pathogens and the immune responses mediated by these cells are characterised by the secretion of cytokines including interferon-γ, tumour necrosis factor-α and IL-12[55,56]. Th2 cells classically function to provide immunity against parasites and typically produce cytokines which include IL-4, IL-5 and IL-13[55,57]. Until a few years ago, it was a well-accepted paradigm that intestinal inflammation in CD is mediated by a Th1 response and in UC via Th2 response. In recent years, several other Th subtypes have been identified as drivers of inflammation in IBD

including Th17, Th9, Th22 and the regulatory T cells (Tregs)[55]. Evidently, with the emerging role of other T cell lineages, the concept of Th1 and Th2 responses in CD and UC respectively, as the primary pathways of inflammation has been largely superseded[58]. In health, a finely tuned homeostasis exists between the Tregs and the pro-inflammatory Th cells. An overzealous Th response or an inadequate Treg response can tip the balance towards an undesired inflammatory response[4].

In addition to the primary cells of the innate and adaptive immune system, non-immune cells of the mucosa including the epithelial cells, sub-epithelial myofibroblasts, adipocytes and stromal cells play a crucial role in shaping and fine-tuning immune responses in the GI mucosa[59].

1.3 AN OVERVIEW OF THE PATHOGENESIS OF IBD

Inflammatory bowel disease has a complex biology involving multiple genetic, immunological and environmental factors. Figure **1.4**. Over the last decade or so, with the technological advancements such as genome wide association studies (GWAS) and next generation sequencing (NGS), a more defined albeit complex interplay between genes, host immunity and the resident microbiota has emerged[4]. This has prompted a multitude of functional studies to assess the complexities of immune networks and identify potential triggers for IBD[60,61]. The collective findings of these studies, have identified mechanisms involved in the pathogenesis of IBD, including disruption of the epithelial barrier integrity, immune surveillance and effector responses, leading to immune dysregulation and inflammation[62]. There has been recent interest in the role of the exposome in triggering the onset and disease flare-ups. The exposome is the summation of all environmental factors an individual is exposed to during lifetime[63]. Multiple theories have been proposed to explain the environmental risk factors that may influence the immune system or the gut microbiome, resulting in an abnormal immune response in the gut. One of the predominant theories include the 'hygiene hypothesis', which attributes the increasing incidence of IBD to a lack of exposure to enteric pathogens in childhood. Improved hygiene and western diets may lead to an increased susceptibility to an inappropriate immune response upon exposure to new antigens later in life[64]. Given the scarcity of prospective studies and trials, clear identification of specific environmental factors within individuals is challenging. Although the strength of association between IBD and the various environmental factors vary among individuals and populations, a crucial denominator is the genetic make-up and its overarching influence over the immune function[65,66].

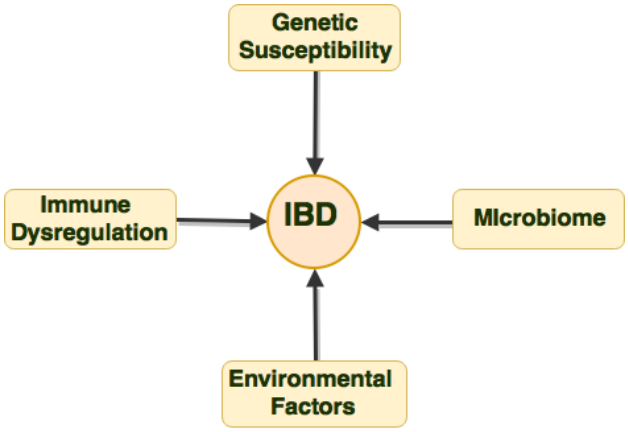


Figure 1.4 An aetiological model of IBD

The figure shows multifactorial triggers for the pathogenesis of IBD, including immune dysregulation in a genetically susceptible individual with contributions from the host microbiome and the environment.

1.4 A HISTORICAL JOURNEY THROUGH THE GENETICS OF IBD

Since the original description of Crohn's disease by Burrill B. Crohn in 1932[67], there have been several lines of epidemiological indicators implicating genetic predisposition in the pathogenesis of IBD. The most notable being familial aggregation of cases, geographical heterogeneity and ethnicity, and most importantly the high concordance rates among monozygotic twins[68]. In the 1980s several studies confirmed the early findings of familial clustering of CD and suggested a positive family history ranging between 8 to 25%. At that time, CD was considered more familial than UC[69]. Large studies were carried out in the late 80s in the UK and Scandinavia, which showed an increased concordance in twin studies[68,70]. These early studies progressed towards family-based linkage analysis in the 90's to identify specific susceptibility genes.

1.4.1 Linkage studies

Observed familial clustering of IBD led to the hypothesis that the susceptibility genes in IBD could be tracked via linkage analysis studies. Linkage is a method to identify chromosomal regions that co-segregate with the presence or absence of a trait in pedigrees with multiple affected members. Linkage between a trait and a specific segment of the chromosome will be present when relatives that share the common trait co-inherit the specific chromosomal region more often than that expected by chance[71]. In essence, this means searching for excessive allele sharing in affected family members[72].

To map disease genes to specific chromosomal regions, several hundred genetic markers from defined loci across the whole genome are genotyped in family members with the disease. The segregation of markers is followed through a suitable number of affected and unaffected members drawn from several families. The results may indicate marker loci that are tightly linked to the disease thereby tracking the sub-chromosomal location for the disease gene. The convention is to use the logarithm of the likelihood ratio, called the lod score (logarithm of the odds) developed by Newton Morton, which is a ratio of the observed co-segregation with disease versus that expected if the marker is unlinked to the locus[73]. Positive lod scores favour the presence of linkage, whereas negative scores suggest that linkage is less likely. By convention, a lod score of greater than 3.0 is normally taken to be the threshold of statistical significance for linkage[74,75].

Seven loci (IBD 1-7) on chromosomes 16q, 12, 6p, 14q, 5q, 19 and 1p were identified and named according to their date of reporting and independent replication through eleven published linkage studies[71,76-86]. The discovery of nucleotide-binding oligomerisation domain-containing-2 (*NOD2*) gene on IBD locus 1(chromosome 16, lod score ≥ 3.6) in 2001 was a major breakthrough in the history of IBD genetics[87]. Three comparatively common variants were identified, two missense and one frameshift mutation (described in subsequent sections). To date, the *NOD2* gene is regarded as a single largest genetic risk factor for Crohn's disease[72,87].

1.4.2 Genome-wide association studies

Genome-wide association studies typically look for statistically significant differences in allele frequencies between a large number of independent diseased individuals and population controls across hundreds of thousands of genetic markers spread throughout the genome[88]. With the completion of the Human Genome project in 2003[89] and the subsequent International Hapmap Project in 2005[90], gene discovery through genome wide association studies in complex diseases including IBD acquired an accelerated pace[91]. These studies have had great impact in providing insights into the understanding of the disease and its complex biology. GWAS were made possible through the identification of single nucleotide polymorphisms (SNPs), which are sites in the genome sequence where individuals differ by a single nucleotide base. The human genome contains about 3 billion base pairs with SNPs positioned typically over distances of 300 base pairs of DNA[92]. Sequencing of the human genome introduced a new paradigm into genomic research, enabling the generation of GWAS[91]. Subsequent to the technological advances facilitated

by the HapMap project, a huge amount of effort was put in through substantial international collaborative ventures, which then led to the detection and validation of more than ten million SNPs across the human genome[91,92]. It was then established that much of this variation across the human genome could be efficiently captured by a subset of "tag" SNPs via the phenomenon of linkage disequilibrium (LD) among neighbouring SNPs. Figure **1.5.** SNPs typically have two alleles (biallelic), which means within a population there are two commonly occurring possibilities for a SNP location. The frequency of a SNP is defined in terms of the minor allele frequency (MAF) or the frequency of the less common allele[93]. Current GWAS are typically powered to characterise variants of MAF ≥ 0.05 (those that are present in more than 5 % of the population) and do not include the contributions from rare variants[4].

Figure 1.5 Rationale for using tag SNPs to identify haplotypes

A short stretch of DNA is shown with 3 'tag' SNPs indicated with arrows within 4 haplotypes occurring in different people within the population. Most of the sequence is identical in all the 4 DNA segments except at the bases indicated by arrows where the variation occurs. Genotyping just the 3 SNPs is sufficient to identify these haplotypes uniquely. The tag SNPs can be used as surrogate markers for identifying haplotypes. Adapted from The International HapMap Project[89].

GWAS requires application of a high throughput genotyping platform, nowadays typically analysing one million or more SNP markers in the genome of several thousand cases and controls[94]. To carry out a genome-wide association study, two groups of people are investigated, including participants with the phenotype of interest and control subjects without the phenotype of interest (matched to the participant group, particularly as regards ethnicity, as well as other factors dependent on the trait of interest). GWAS compares common genetic variants in large numbers of affected cases to those in unaffected controls to determine whether an association exists. In simplistic terms, if certain variants are found at a significantly different frequency in people with the phenotype of interest compared to controls and withstand a multiple testing correction, the variants are said to be associated with the phenotype.

Using the HapMap data, which map linkage disequilibrium across the human genome, representative SNPs are selected that will differentiate or tag the common haplotypes at

each locus. The 'tag' SNPs are then genotyped in disease cases and controls using microarrays and the allele frequencies for each SNP are compared in the two groups. Once genotyping is complete, SNP data is subjected to a number of quality-control checks and those SNPs that fail this rigorous quality control check are removed from further consideration. Each SNP that survives this check is then tested for association with the disease. A Manhattan plot is typically used in GWAS and it plots the negative log of the p-value against chromosomal position. As an example of a Manhattan plot, figure **1.6** shows genome-wide scans for seven complex human conditions published by The Wellcome Trust Case Control Consortium (WTCCC) in 2007[91]. This landmark study included a joint GWAS undertaken in a British population examining approximately 2000 cases for each of the seven conditions and a shared set of 3000 controls. The conditions assessed included bipolar disorder, coronary artery disease, Crohn's disease, hypertension, rheumatoid arthritis, type 1 diabetes and type 2 diabetes.

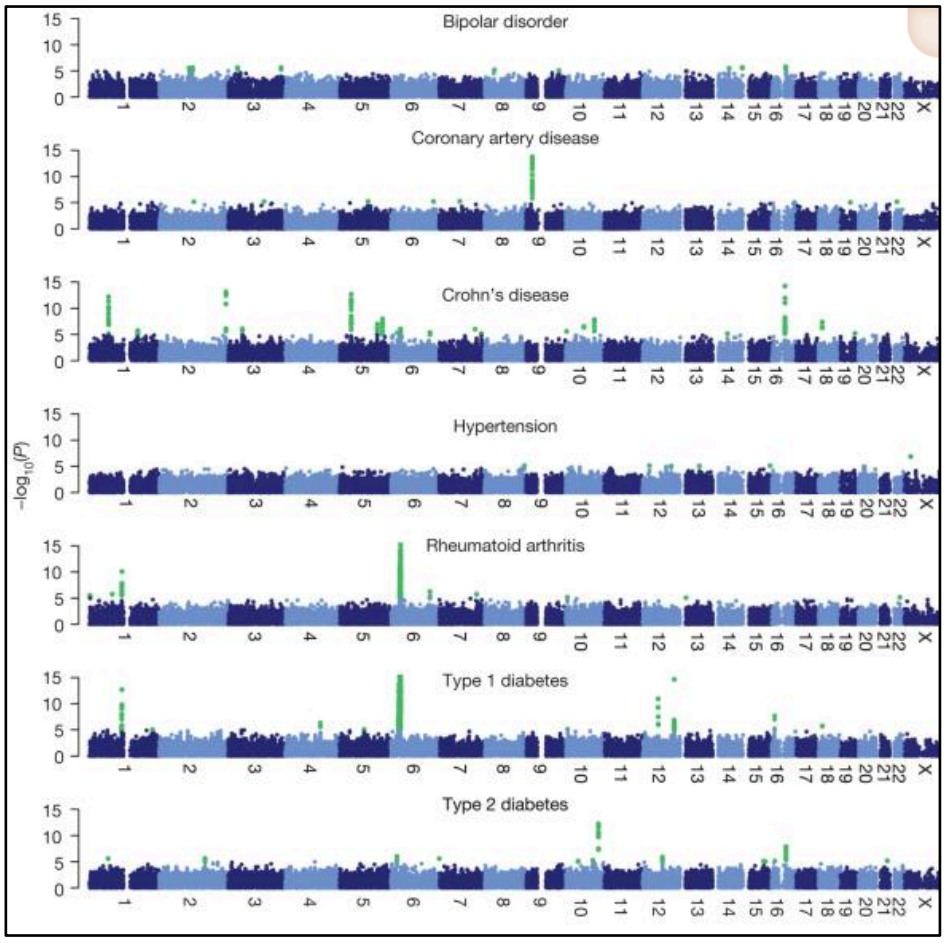


Figure 1.6 An example of a Manhattan plot in a GWAS

Figure from the GWAS published by the Wellcome Trust Case Control Consortium (WTCCC) in 2007. For each of the seven conditions examined, the negative logarithm of the association p value derived by logistic regression for each SNP is displayed along the Y-axis and the chromosomal

position along the X-axis. For clarity, the chromosomes are shown in alternating colours. Each dot represents a SNP and SNPs with p values <1 x 10^{-5} are highlighted in green. Statistically significant SNPs tower above the lower level scatter of dots, giving a visual impression similar to a Manhattan skyline.

A major challenge in interpreting the results of GWAS is the possibility of false positives due to multiple testing. Typically, more than one million SNPs are analysed in a GWAS and often several phenotypes in the same data are assessed for genotype-phenotype association[93]. Because of the number of statistical tests that are performed, there is a high false positive rate. Therefore, the results are subjected to multiple-testing correction and depending on the study design, genome wide statistical significance is set at p values of approximately 0.5 x 10^{-8} [94]. The genotype test statistics are calculated for each variant and referenced against statistics expected under the null hypothesis of no disease association. SNPs associated with disease at an appropriate statistical threshold are genotyped in an independent replication dataset of cases and controls to ensure that a genotype-phenotype association observed in a GWAS represents a credible association and not an artefactual finding due to uncontrolled biases. Even so, replication of the association in an independent cohort is desirable to confirm findings[95,96]. Many if not all variants identified through GWAS are "associated" and not causal for disease. Determining the causal factors can be extremely challenging through GWAS and will require detailed fine-scale mapping and mechanistic studies[97].

1.4.3 GWAS in IBD

There have been at least 18 GWAS in IBD to date and 5 meta-analyses, which have successfully identified approximately 235 loci, thereby unravelling fundamental aspects of IBD biology[60,98,99]. The international IBD Genetics Consortium performed a meta-analysis of previously published IBD genome-wide association scans and used the Immunochip platform for replicating the initial findings[60]. The immunochip genotyping chip (Illumina; San Diego, California) was originally developed when it became increasingly obvious that many immune-related diseases shared genetic susceptibility loci. The study published in 2012 included data from more than 75000 individuals of European ancestry, providing sufficient statistical power to identify 71 new associations for a total of 163 loci. Some of the most significant findings included: 30 loci were CD specific, 23 UC specific and notably, 110 of the 163 loci were associated with both CD and UC. This degree of sharing

of genetic risk strongly suggests that many of the biological mechanisms involved in CD and UC are overlapping[60,100].

Another significant finding of the meta-analysis was that about 70% (113/163) of the IBD loci show a striking association with other complex diseases. Prior to this meta-analysis, several other published reviews examined the shared susceptibility genes between different immune-mediated and inflammatory diseases[101-103]. In this meta-analysis[60], type1 diabetes mellitus (DM) shared the largest number of loci with IBD, partially driven by the large number of known type1 DM associations. Seven other immune-mediated diseases showed strong enrichment overlap, with largest being ankylosing spondylitis and psoriasis. IBD loci were also markedly enriched for genes involved in primary immuno-deficiencies. An overlap with loci for complex mycobacterial disease was also detected. IBD associations were found in 7/8 loci identified by GWAS for leprosy, including 6 cases where the same SNP was implicated. These findings have generated an interest regarding the role of mycobacteria in IBD although the relationship is likely to be complex[60,100].

In 2015, a trans-ancestry GWAS was performed, with genome-wide data or Immunochip data from an extended cohort of 86,640 European individuals and Immunochip data from 9,846 individuals of East Asian, Indian or Iranian descent. A further 38 loci were identified increasing the number of known IBD loci to 200[98]. Although the majority of risk-associated loci were shared across populations, genetic heterogeneity between divergent populations was observed at several established risk loci driven by differences in allele frequency (*NOD2*) or effect size (*TNFSF15* and *ATG16L1*). For example, the major European risk variants in *NOD2* and *IL23R* were not identified in individuals of East Asian ancestry. Two of the three risk variants at *TNFSF15-TNFSF8* had much larger effects on IBD risk in East Asians compared to Europeans despite similar allele frequencies in the two populations[98]. Another association study published in 2015, investigating chronic inflammatory conditions including IBD, psoriasis, ankylosing spondylitis and primary sclerosing cholangitis identified 6 new loci for CD and 3 shared risk loci with the other conditions[104]. Additionally, a recently published GWAS in IBD identified 25 new susceptibility loci, of which, 6 were associated with CD, 5 with UC and 14 risk loci shared between the two conditions. Three of the risk loci harboured integrin genes, which encode proteins in pathways that have been identified as important therapeutic targets in IBD[99].

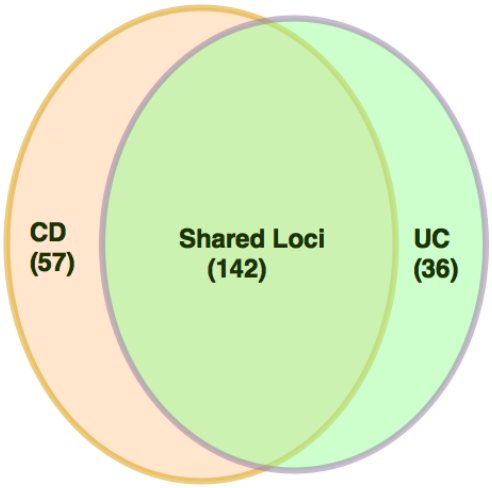


Figure 1.7 Genomic loci identified in IBD through GWAS

The risk loci identified through GWAS in IBD are indicated in this venn diagram. As seen in the diagram, 60% of risk loci are shared between the two main forms of IBD.

Heritability can be calculated from GWAS studies from the additive genetic variance attributable to all identified SNPs, known as the SNP effects[105]. Heritability is defined as the proportion of phenotypic variance that can be attributed to genetic variance[106]. The genomic loci described so far can explain 13.1% and 8.2% of variance in disease heritability for CD and UC respectively[98]. Despite initial high hopes and success, GWAS has not been able to account for a substantial proportion of the missing heritability in IBD. One of the primary reasons for the missing heritability in GWAS is that they work well for common variants, but are underpowered to assess rare variants. Rare variants with high penetrance and common variants with low penetrance may not be detected[107]. Another possible explanation for missing heritability is that GWAS scrutinises SNP data only whereas other forms of genetic variations such as CNVs (copy number variants) can account for some heritability. Additionally, GWAS have had limited success in identifying interactions between genes, referred to as 'epistasis', which can play an important role in the pathogenesis of complex diseases[108].

1.4.4 Next generation sequencing

Next generation sequencing (NGS) involves massively parallel sequencing of DNA molecules, allowing for sequencing throughput several orders greater than conventional Sanger sequencing[109]. NGS reads provide direct information for each nucleotide covered and no prior knowledge of potential variant sites is required, enabling the identification of rare and novel variants. The advancements in NGS technology has allowed for rapid and cost-effective sequencing, allowing investigation of rare disorders, including those with

incomplete penetrance, and unidentified biological causes[110]. Establishing an extensive catalogue of human genetic variation within populations provides a baseline for the exclusion of common and presumably less benign variation, when looking at data derived from an individual sample[111].

In recent years, there has been considerable discussion about the optimal approaches in the application of NGS for the investigation of Mendelian and complex genetic conditions[112]. For many years, the cost of whole genome sequencing (WGS) remained prohibitively expensive making it a less practical option compared to whole exome sequencing (WES)[113]. However, with decreasing sequencing costs and rapid technological advances in the field, it is likely that there will be a shift increasingly towards WGS, both as a diagnostic and research tool[113,114]. The exome comprises all the exons (protein-coding units), which constitute about 1-2% of the genome[72,115]. Approximately 85% of the disease-causing mutations are typically found in the protein-coding regions and canonical splice sites, supporting the idea that exon-focused sequencing will yield the most functionally interesting variants[116,117]. A targeted sequencing approach therefore is a potentially efficient strategy in identifying rare functional variants, given the prohibitive cost of sequencing the whole genome and the computational challenges of interpreting the sequence variations outside the coding regions. The utility of whole exome sequencing was first demonstrated by Ng et al. in 2009[118] on 12 individuals, and has since been demonstrated to be an exceptionally useful tool in a diverse spectrum of undiagnosed genetic conditions[119-121]. With the power to detect novel variants from a small number of individuals, WES is proving invaluable in the diagnosis of rare diseases of unknown etiology. WES can be a powerful research tool to complement GWAS in unravelling rare variants thereby contributing positively to the heritability of IBD[72,117,121]. A striking example where exome sequencing has proven invaluable is the detection of *XIAP* gene mutations, presenting in a child less than two years old with a refractory form of IBD unresponsive to conventional treatment. This patient was successfully treated with haemopoetic stem cell transplant[122]. In a recently published study by Girardelli *et al*, a rare homozygous *NOD2* mutation (c.G1277A; p.R426H) was identified through WES in an infant presenting with a severe IBD-like phenotype with poor response to standard therapy[123]. The commonly described *NOD2* mutations in IBD are located in the leucine-rich repeat (LRR) domain of the gene and typically lead to loss of function[124]. However, in the index patient described in the study[123], the homozygous variant was detected in the nucleotide-binding domain (NBT). Mutations in the NBT domain of *NOD2* are frequently

found in patients with Blau syndrome (BS), which is inherited as an autosomal dominant condition and clinically characterised by eye, skin and joint involvement, without intestinal manifestations. Mutations in the NBT domain of *NOD2* in BS typically cause gain of function (GOF)[125]. Functional studies conducted in the patient demonstrated GOF as seen in BS, but presented with an infantile IBD-phenotype without the clinical diagnostic markers of BS. In another study, highly pathogenic mutations in the conserved region of the tripartite motif containing 22 gene (*TRIM22*), were identified through WES in three patients presenting with a severe infantile form of IBD[126]. Two of these patients had homozygous variants and one patient had compound heterozygous variants in *TRIM22*. Functional studies showed that the variants disrupted the ability of the gene to regulate NOD2-dependent pro-inflammatory pathway signalling.

Although WES has been an appealing alternative to whole genome sequencing (WGS), more recent studies have suggested that WGS provides a higher diagnostic yield compared to WES[127,128]. A major advantage of WGS is its ability to sequence all parts of the genome, including the non-coding DNA and enable more reliable detection of structural variants[112]. Additionally, WGS provides more complete coverage of the exome, providing greater variant detection sensitivity[128].

1.5 THE IMMUNO-GENETIC ARCHITECTURE OF IBD

Candidate genes implicated in IBD highlight the interplay between several cellular mechanisms and immune pathways that are crucial for maintaining gastro-intestinal (GI) homeostasis. These mechanisms broadly include the barrier function of the epithelium, innate immune regulation of microbial invasion, epithelial restitution, mesenchymal cell responses, autophagy, endoplasmic reticulum stress, and the various effectors and regulators of adaptive immune response[3,4]. Mutations in the key regulatory genes result in disruption of the carefully balanced homeostasis that exists between the GI immune system and the complex microbial milieu of the gut. The adverse outcome of this imbalance is inflammation of the gut resulting in IBD. Figure **1.8** outlines a brief overview of the genes implicated in the various IBD-related processes. The molecular details of the mucosal mediators, immune cells, signalling pathways and the genes regulating these pathways at the various points of defence in the gut will be discussed in the subsequent sections.

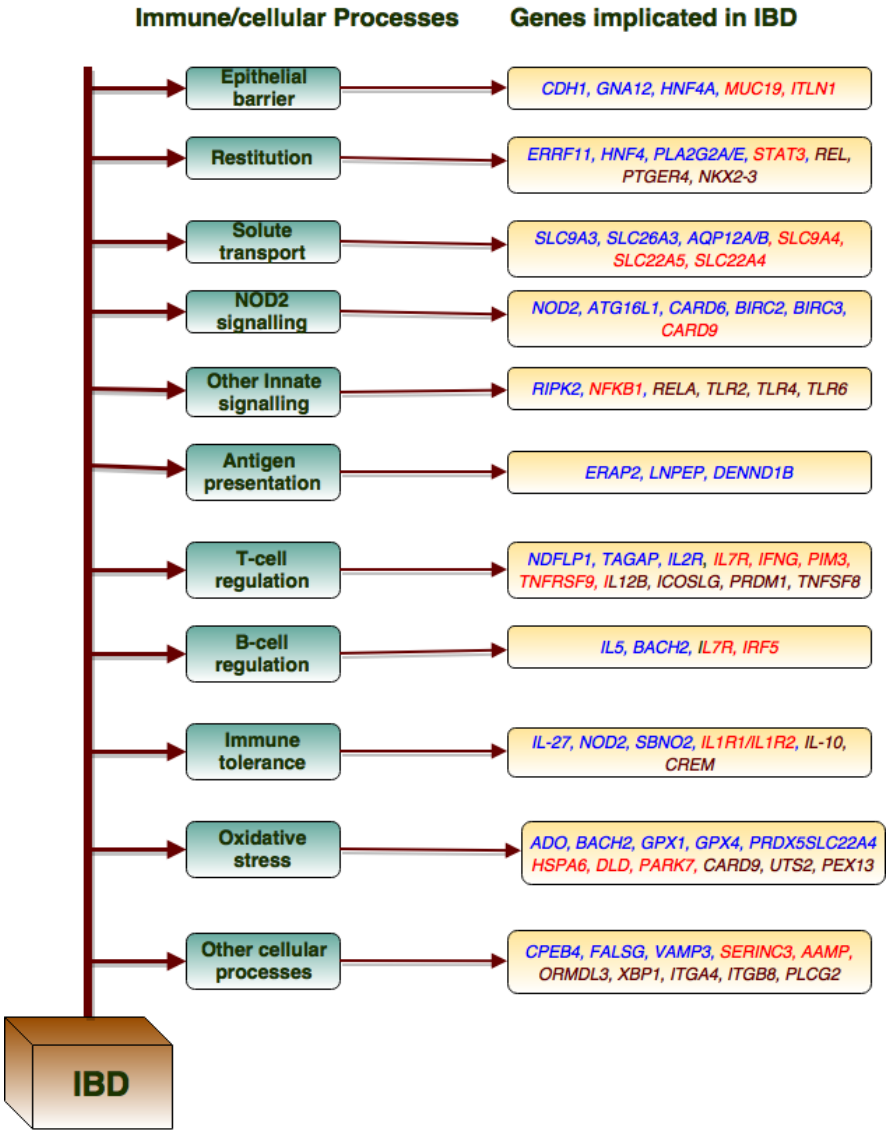


Figure 1.8 Cellular processes and genes implicated in IBD

Intestinal homeostasis involves a finely co-ordinated balance between the epithelial, innate and Adaptive immune processes. The boxes on the left denote the various cellular/immune processes which play an important role in the causation of IBD, with the respective candidate genes on the right. Text colour indicates whether the genes are associated with CD (blue), UC (red) or both (brown). The first line of defence includes the mucus and epithelial barrier (at the top), followed by the innate mechanisms of immune defence and then the adaptive immunity. More recently, the role of several other genes has become the focus of IBD therapeutics including the integrin genes (ITGA4, ITGB8).

1.5.1 Epithelial barrier and other luminal elements of defence

The intestinal epithelium is a single layer of columnar cells with an overlying layer of mucus separating the sub-epithelial lamina propria from the hostile microbial medium of the gut. The colon has a dual mucus layer; the inner layer has properties that restrict

bacterial motility and adhesion to the epithelium[50,129]. The inner mucus layer is formed from sheets of MUC2 mucin organised into stacked layers in a stratified manner, which physically separates the epithelium from the bacteria. *Muc2*-null (*Muc2*$^{-/-}$) mice do not have any protective mucus and develop spontaneous colitis. A number of specific changes in the mucus layer have been observed in patients with UC including depletion of recognisable goblet cells, reduced MUC2 synthesis and secretion, decreased sulfation of mucin leading to reduced viscosity and impaired glycosylation with consequent shortening of the oligosaccharide chains; all leading to an increased susceptibility to erosions and inflammation[130,131].

The epithelial cells associate with each other through a series of intercellular junctions, the most important being the tight junctions (TJ) and adherens junctions (AJ). TJs are circumferential and continuous structures that form a permeability barrier at the apical end of the cell. Some of the protein families found at the tight junctions include the occludins, claudins, zona occludentes (ZO) proteins and junctional adhesion molecules (JAMs). AJs reside immediately sub-adjacent to the TJs and are formed of cadherins, catenin proteins and others[129]. See Figure **1.9.**

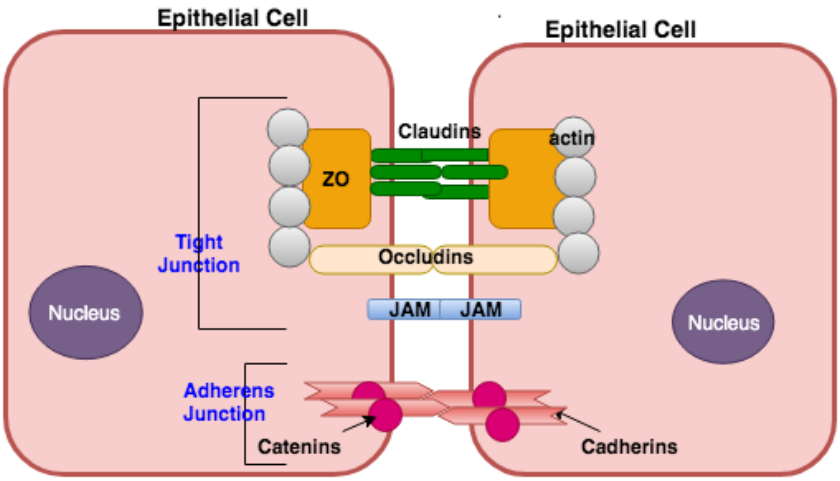


Figure 1.9 Inter-cellular epithelial junctions

The tight junctions (TJ) and adherens junctions (AJ) are multi-protein junctional complexes, which function to limit the permeability through para-cellular epithelial spaces. The TJs which are located towards the apical surface of the epithelial cell are constituted by proteins including claudins, occludins, zona occludentes (ZO) and junctional adhesion molecules (JAM). The TJ structure is anchored to the actin component of the cytoskeleton, which serves to keep the adjacent cells together. The AJ is located at the basal membrane and includes cadherins, catenins and other proteins within its structure.

Epithelial barrier integrity is crucial for intestinal homeostasis in the context of IBD and GWAS have identified several genes associated with epithelial integrity, these include *CDH1, GNA12, HNF4A, ERRRFI1, MUC3A, MUC19, ITLN1 and PTPN2*[4,132]. *CDH1* is encoded on chromosome 16 in the region flanking IBD-1 locus and encodes for E-cadherin, a principal protein at the adherens junction. *CDH1* gene polymorphisms have been found to be associated with Crohn's disease and intestinal biopsies from patients carrying the mutant alleles show inappropriate protein localisation and cytosolic accumulation[133]. The *PTPN2* gene located on the short arm of chromosome 1, acts as a negative regulator of signalling by interferon-gamma (IFN-γ), which is a pro-inflammatory cytokine known to play an important role in the pathogenesis of IBD. *PTPN2* exerts a protective effect by restricting the capacity of IFN-γ to increase epithelial permeability and prevents induction of the pore-forming protein, claudin-2. The protective effect of *PTPN2* has been demonstrated in patients with CD through functional studies[134]. Likewise, GWA studies have implicated the role of other candidate genes involved in epithelial barrier integrity in IBD[4]. It is believed that in addition to allowing increased access of bacteria and their products to the underlying lamina propria thereby triggering an inflammatory response, decreased epithelial barrier function may also have a critical role in the persistence of diarrhoea in IBD through luminal loss of fluid and electrolytes[134].

1.5.2 Checkpoints in the innate immune system

In a healthy gut, a well-balanced mutual relationship between the gut microbiota, the intestinal epithelium and the intestinal immune cells, provides a homeostatic environment. It is now widely accepted that a break-down of this homeostasis due to inherent 'weaknesses' at the immune checkpoints leads to inflammatory changes seen in IBD[4,60].

1.5.2.1 Innate immune cells in the intestinal mucosa

The epithelial barrier is supported by a well-evolved mucosal immune system, which is tasked to prevent any microbial intrusion whilst remaining tolerant of the commensal bacteria. Paneth cells of the intestinal epithelium, dendritic cells (DCs), macrophages, innate lymphoid cells (ILCs) and neutrophils are key cellular components of the innate mucosal response in the event of a pathogenic incursion[4,50] (Figures 1.2 & 1.3). The Paneth cells, which reside at the base of the crypts secrete anti-microbial peptides (AMPs) in response to bacterial triggers and also inflammatory mediators to recruit other immune cells[135]. The intestinal DCs are specialised antigen presenting cells, forming a critical

interface for monitoring the mucosal environment and sending signals to initiate appropriate adaptive immune responses[136].

Likewise, resident macrophages play a key role in maintaining mucosal homeostasis through their phagocytic and bactericidal activities, and by reducing the production of pro-inflammatory cytokines. Although intestinal macrophages are derived from monocytes, they display a distinct functional phenotype. Characteristically, they do not express CD14 and are largely refractory to bacterial stimulation in terms of production of pro-inflammatory cytokines, however they retain their bactericidal and phagocytic function. Moreover, they express anti-inflammatory cytokines such as IL-10, promote differentiation of regulatory T cells and suppress DC-triggered immune responses. As will be discussed in the subsequent sections, mutations in the genes encoding for IL-10 and its receptors IL-10R, have been associated with very early onset IBD presenting with a severe disease phenotype[137]. CD14 present on mononuclear cells and polymorphonuclear phagocytes normally acts as a high-affinity receptor for bacterial complexes such as lipopolysaccharides (LPS) initiating a pro-inflammatory cytokine response. Lack of CD14 expression therefore serves to dampen inflammatory responses to luminal bacteria and promotes a state of homeostasis[50,138].

The innate lymphoid cells (ILCs) are a novel population of innate lymphocytes, which are selectively enriched at the mucosal sites and play an important role in regulating and modifying immune responses[139]. The ILCs are immediate and effective producers of potent cytokines implicated in the pathogenesis of IBD, including TNF-α, IL-5, IL-13, IL-17, IL-22, IFN-γ and others. A careful scrutiny of the known IBD susceptibility loci have shown that the genes encoding these key inflammatory cytokines are either expressed or closely linked to the ILC function[139]. Based on their cytokine profile and expression of specific cell surface markers, ILCs are currently classified into three distinct functional categories. Type 1 ILCs express T-bet, a Th1-type transcription factor and produce IFGγ; Type 2 ILCs produce Th2 cytokines such as IL-5 and IL-13 and are dependent on GATA-binding protein-3 (GATA3) and retinoic acid receptor-related orphan receptor-α (ROR-α); and type 3 include subtypes that express transcription factor ROR-γt and produce IL-17 and/or IL-22. These three subtypes closely parallel the major helper T cell subsets in terms of their signature cytokines[50,140]. A study reported an increased frequency of type 1 ILCs in the inflamed intestinal mucosal tissue of patients with CD[141]. Type 3 ILCs play an important role in regulating tissue repair in the intestine, particularly in the context of IBD.

Studies have shown reduced number of type 3 ILCs in the intestinal tissue of patients with IBD compared to non-IBD controls[142,143].

Neutrophils also play a significant role in the resolution of inflammation through various mechanisms, for example by synthesising anti-inflammatory mediators such as lipoxin A_4. They have elegant defense mechanisms to eliminate microbes that have translocated across the epithelium. During the inflammatory response, neutrophils contribute to the recruitment of other immune cells, but also facilitate mucosal healing by releasing mediators, necessary for the resolution of inflammation[4,144].

1.5.2.2 Matricellular proteins and mesenchymal stromal cells (MSCs)

The extra-cellular matrix is a complex network around tissues consisting of matricellular proteins which are important in regulating and maintaining a balance of immune responses within the mucosa and surrounding tissues. These include a diverse group of structurally unrelated proteins such as osteonectin, periostin, tenascins, thrombospondins, transforming growth factor-beta 1 (TGF-β1) and others[145,146]. Some of the key functions of the matricellular proteins include interacting with specific cellular receptors to initiate signal transduction, binding to cytokines and related molecules, orchestrating as well as inhibiting immune responses in a cell-specific manner and regulating tissue repair. The role of matricellular proteins such as periostin has been described in a wide range of neoplastic, inflammatory and fibrotic conditions[147-149]. In IBD, there is a growing body of evidence to suggest its role in mucosal inflammation, tissue remodelling and repair[150,151]. Several studies have demonstrated that periostin mediates activation of inflammatory cascades such as the NF-κB following its interaction with αv-integrins, which are trans-membrane receptors facilitating adhesion of immune cells to the ECM[150]. In tumour biology, studies have shown that periostin promotes cancer cell survival, tumour growth, invasion and metastasis through the activation of PI3K/AKT pathway[152-154]. Periostin has also been associated with the development of fibrosis through dysregulated mesenchymal remodelling and repair in a number of conditions including idiopathic pulmonary fibrosis, scleroderma, renal interstitial fibrosis, myocardial fibrosis and others[155,156]. The role of periostin in intestinal fibrosis has however not been previously described.

Mesenchymal stromal cells (MSCs) are abundantly found in the bowel wall and together with the extra-cellular matrix form the connective tissue network beneath the epithelium. Originally, these cells were conceptualized as primarily structural, but recent evidence has highlighted the broad importance of these cells in diverse cellular mechanisms in health

and disease[3,59,157]. Accumulating evidence supporting the wide-ranging immuno-modulatory functions of MSCs led to several pre-clinical studies in animal models of colitis, which further prompted clinical trials to assess the utility and feasibility of MSCs in human IBD. Although the data on its safety and feasibility are encouraging, there are inconsistent reports on its efficacy[158-160]. Nevertheless, MSC-based therapy represents a promising approach for adopting cell-based therapy in treatment-resistant IBD[161].

1.5.2.3 Toll-like receptors in the intestinal epithelium

Toll-like receptors are cell surface pathogen recognition receptors, acting as sensors for extracellular microbial elements, as well as endocytosed ligands. Although TLRs are expressed on the cell surface, some such as TLR3 and TLR7/8/9 that are responsive to intracellular viral RNA and unmethylated bacterial DNA are displayed on the surface of endosomes. Engagement of TLRs with their respective ligands initiate innate cell-signalling pathways, such as the NF-κB and other pathways depending upon the respective ligands[162] (Figure **1.10**). Up to 13 TLRS have been described in mammals; 10 in humans[163]. All TLRs have the same basic structure characterized by an extracellular domain containing leucine-rich repeats (LRRs) arranged in a horse-shoe or crescent-shaped solenoid structure and a cytoplasmic tail that contains a conserved region called the Toll/IL-1 receptor (TIR) domain.

The expression profiles of TLRs vary between tissues, but they are predominantly expressed in tissues involved in immune function; peripheral blood leukocytes, spleen, lungs, gastrointestinal tract and others. Most TLRs are expressed in the human gastrointestinal tract, at least at the mRNA level. In the human small bowel, TLR3, TLR4 and TLR5 are expressed on the basolateral surfaces of villus enterocytes. In the human colon, TLR3 and TLR5 are abundantly expressed, whereas TLR2 and TLR4 expression is low[164]. TLR9 is expressed on the apical membrane as well as the basolateral membrane of the intestinal epithelium. Interestingly, it has been demonstrated that activation through the apical membrane determines tolerance, whilst activation via the basolateral membrane initiates the NFκB pathway. The differential spatial expression patterns therefore could be a key regulatory mechanism to distinguish commensal bacteria from pathogenic organisms[53,165]. Most of the studies on the function of TLRs in the intestinal epithelium have been derived from animal models and cell lines. In a study by Cario et al, intestinal mucosal biopsies from patients with IBD were examined for expression of TLRs using immunofluorescence histochemistry. TLR3 was observed to be significantly down

regulated in active CD, but not UC. In contrast, TLR4 was strongly up regulated in both CD and UC, and TLR2/TLR5 expression remained unchanged in IBD. These data suggest that alterations in TLR expression in patients with IBD may be linked to the pathogenesis of IBD[166].

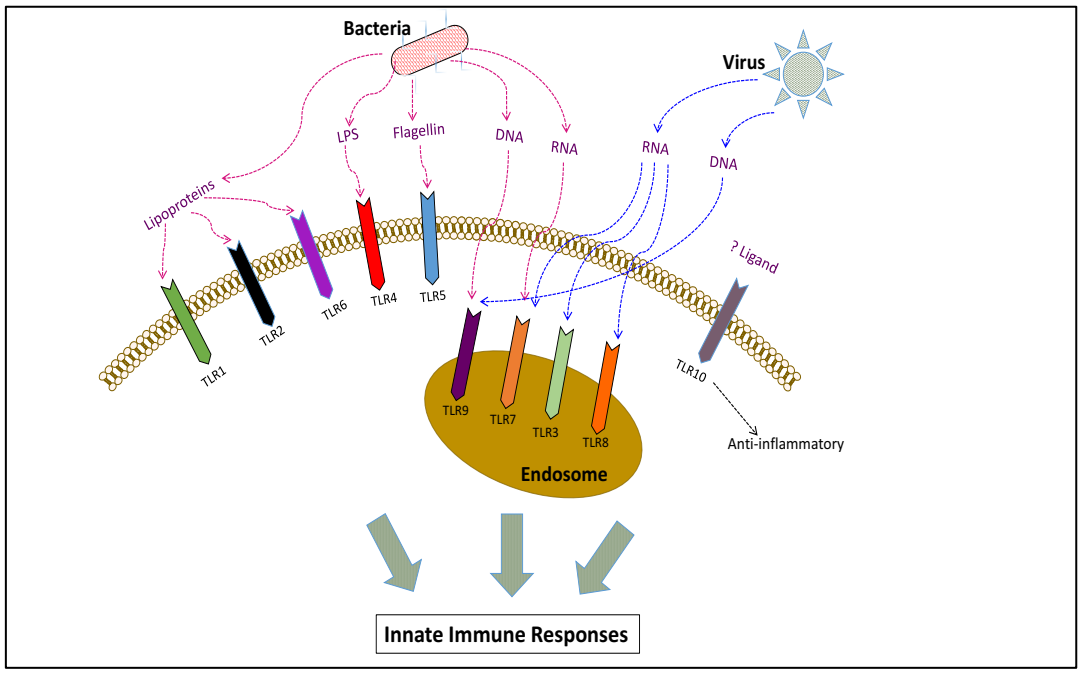


Figure 1.10 Distribution patterns of TLRs in the small intestine and the colon

Well-characterised toll-like receptors on the cell-surface membrane include TLR1, TLR2, TLR4, TLR5, TLR6 and TLR10. TLR3, TLR7, TLR8 and TLR9 are expressed on the surface of endosomes and detect intra-cellular viral RNA and unmethylated bacterial DNA. TLR1, TLR2 and TLR6 sense bacterial lipoproteins, TLR4 detects lipopolysaccharides (LPS) and TLR5 detects flagellin. TLR9 recognises endosomal CpG DNA (synthetic DNA oligonucleotides containing CpG motif), as well as viral DNA. TLR3/7/8 detect viral RNA within the cytosol. Each TLR following detection of its respective ligand will initiate a cascade of interactions downstream, resulting in the activation of cell-signalling pathways.

Mutations in several TLR genes have been implicated in the pathogenesis of IBD. The most well-studied among the TLRs for their association with IBD include TLR2 and TLR4. In a Japanese study, Wang et al showed that variants in the *TLR2* gene were associated with a severe disease phenotype in UC including steroid-dependency, however no association was observed between *TLR2* variants and the risk of developing UC[167]. Several studies have evaluated the association of genetic variants in *TLR4* and IBD, although the results from some studies have been inconclusive[168-170]. In an Australian study, the *TLR4* variant allele Asp299Gly was associated with CD phenotype limited to the colon[169]. In another study, two variant alleles D299G and T399I were shown to be

associated with both UC and CD in a Caucasian population through meta-analysis, although a genotype-phenotype correlation could not be established[168].

1.5.2.4 NOD-like receptors and innate immune responses

NOD-Like receptors (NLRs) have been shown to play a key role in innate defence responses against cytosolic pathogens, sensing a wide range of bacterial stimuli, toxins and damage-associated molecular patterns (DAMPS)[171]. More than twenty NLRs have been described in mammals based on their N-activation domains involved in signal transduction[172]. However, among the NLRs, only NOD1 and NOD2 functions have been well characterised in the GI epithelium. While NOD1 recognises the dipeptide D-glutamyl-meso-diaminopimelic acid (iE-DAP) moiety derived from most gram-negative, NOD2 senses N-acetyl muramyl dipeptide (MDP), a peptidoglycan derived from both gram-positive and gram-negative bacteria. Stimulation of the NLRs results in the activation of inflammatory pathways such as the NF-κB pathway, resulting in the transcription of several inflammatory mediators[53].

1.5.2.4.1 NOD2 and IBD

NOD2, also known as caspase recruitment-domain containing protein 15 (CARD15), is largely localised to the cytosolic compartment. It is expressed in macrophages, dendritic cells, Paneth cells on the intestinal epithelium and even in T-cells. It is still not entirely clear which of the intestinal cells mediate the risk conferred by the *NOD2* variants[173]. The NOD2 protein has 3 major domains; a C-terminal domain with leucine-rich repeat motifs (LRR), nucleotide-binding oligomerisation domain (NBT) and the N-terminal caspase activation and recruitment domain (CARD). NOD2 senses its ligand via the LRR terminus. Typically, activation by MDP induces NOD2 oligomerisation through the NBT domain and binding of the receptor-interacting serine/threonine-protein kinase 2 (RIP2) through the CARD terminus[174]. Figure **1.11.**

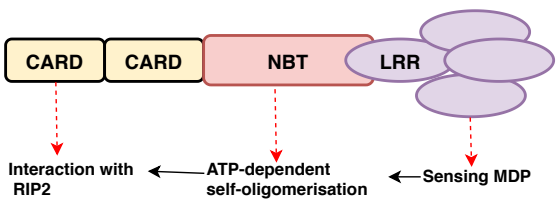


Figure 1.11 Basic structure of the NOD2 protein

NOD2 has three main domains; Leucine-rich repeat (LRR)- a C-terminal domain, Nucleotide Binding domain (NBT) and the N-terminal caspase recruitment domain (CARD). LRR mediates

*MDP sensing, NBT mediates ATP-dependent self-oligomerisation and the CARD domain mediates protein-protein interaction with RIP2, an adaptor protein in the downstream signalling post-NOD2 stimulation. (**Abbreviations**: MDP- Muramyl dipeptide; RIPK2- Receptor-interacting serine/threonine-protein kinase 2)*

RIP2 undergoes polyubiquitination mediated by the inhibitor of apoptosis proteins (IAPs), which include cellular inhibitor of apoptosis protein 1 (cIAP1), cellular inhibitor of apoptosis protein 2 (cIAP2) and X-linked inhibitor of apoptosis protein (XIAP)[175,176]. Polyubiquitinated RIP2 then serves as a scaffold for binding TGF-β-activated kinase 1 (TAK1) protein complex, which further downstream leads to the activation of the nuclear factor kappa B (NF-κB) and mitogen-activation protein kinases (MAPK) pathways. Activation of these pathways results in the generation of inflammatory cytokines as part of immune defence[174]. Figure **1.12.**

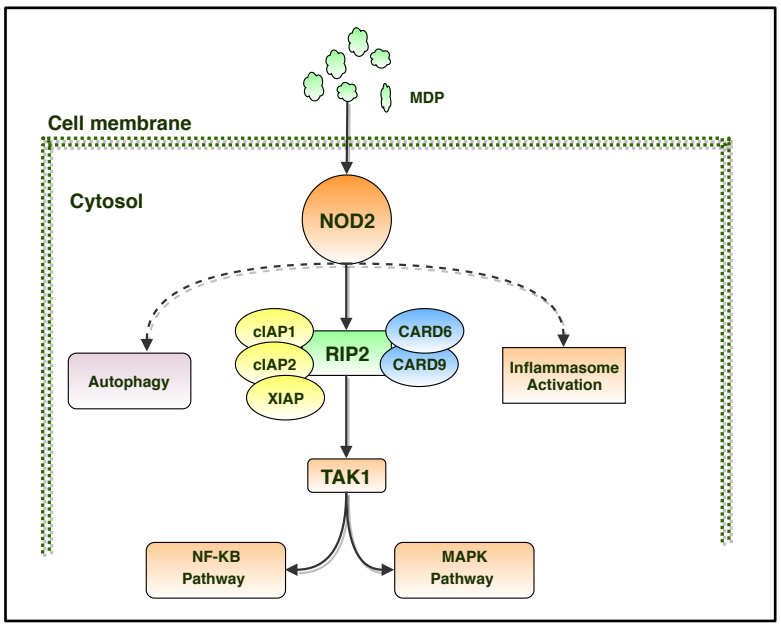


Figure 1.12 NOD2 activation and the downstream signalling pathways

*MDP upon entering the cytosol stimulates NOD2, which then complexes with an adaptor protein RIPK2 for downstream signalling. RIPK2 interaction via the inhibitor of apoptosis proteins (cIAP1, cIAP2, XIAP) and CARD6 or CARD9 results in the activation of TAK1 protein kinase. The inhibitor of apoptosis proteins (IAPs) mediate polyubiquitination of RIP2, which is an essential step for the activation of TAK1. Further downstream, TAK1 activates the two inflammatory pathways, NF-κB and MAPK, resulting in the production of inflammatory cytokines. NOD2 also recruits the autophagy proteins to the plasma membrane at the bacterial entry site, promoting degradation of PAMPs. NOD2 stimulation also plays an important role in inflammasome activation. (**Abbreviations**: CARD - Caspase recruitment and activation domain; cIAP- cellular inhibitor of apoptosis protein; MAPK- mitogen activated protein kinases; MDP- Muramyl*

dipeptide; RIPK2- Receptor-interacting serine/threonine-protein kinase 2; TAK1- TGF-β-activated kinase 1; XIAP- X-linked inhibitor of apoptosis protein)

Currently, it is still debated how loss-of-function variants in *NOD2* determine the risk of CD. It has been proposed that impaired NOD2 functioning results in a defective interaction between the mucosal immune system and the resident microbiota, causing persistent intestinal inflammation following bacterial invasion[53]. NOD2 stimulation plays a critical role in the induction of autophagy, as well as inflammasome activation. NOD2 also has a place in promoting Th17 responses by inducing expression of IL-23[4].

One of the most remarkable achievements in the genetics of IBD came in 2001 when fine mapping of the IBD-1 locus on chromosome 16 by a French group identified the leucine-rich repeat (LRR) variants in the *NOD2* gene (nucleotide-binding oligomerization domain-containing-2) as conferring susceptibility to Crohn's disease. In this study, a positional cloning strategy using linkage analysis followed by linkage disequilibrium mapping in 179 CD patients and 261 unaffected relatives was used to identify three independent associations for Crohn's disease; SNP 8(R702W), SNP 12 (G908R) and SNP 13 (1007fs)[87]. These three variants account for more than 80% of the identified germline variants. A number of rare variants have been described subsequently, all of which almost exclusively localize to the LRR region[177,178]. Although the frequency of *NOD2* variants differs among various populations, a meta-analysis estimated that carrying two *NOD2* variants has a 98% specificity of having a complicated course in patients with CD[179]. *NOD2* has since been consistently replicated in GWAS implicating its role in IBD[60,98].

1.5.2.4.2 NOD1 and IBD

NOD1, also known as caspase recruitment domain-containing protein 4 (CARD4), is an intracellular pathogen recognition receptor like NOD2. It is structurally similar to NOD2 except for the amino-terminal, which consists of only one CARD domain. In contrast to NOD2, NOD1 receptor is constitutively expressed on epithelial cells throughout the gastrointestinal tract[180]. Following interaction with specific dipeptides from gram-negative bacteria, the NOD1 receptor undergoes conformational changes, which initiate an inflammatory signalling cascade via the NF-κB and MAPK pathways[53]. Studies have implicated the role of NOD1 in response to Helicobacter pylori and other gram-negative enteric pathogens, which typically avoid recognition by TLRs[181]. Although *NOD1* genetic mutations have been implicated in IBD, data from different studies have shown conflicting findings[180,182].

1.5.2.5 Autophagy and IBD

Autophagy is believed to be one of the mechanisms by way of which antigens are handled and processed for subsequent presentation to the T-cells via MHC-II surface molecules[183]. Through autophagy, the cell tends to auto-digest damaged intra-cellular organelles or intra-cellular bacteria by formation of an 'isolation membrane'. This sequestered auto-phagosome is then marked for degradation by fusion with lysosomes. Autophagy is a crucial process in innate immunity whereby viruses, bacteria and parasites are engulfed within a membrane for a lethal hit. This layer of defence is particularly important when the microbial invader has breached the cell surface and bypassed Toll-like receptor (TLR) 'stop and search' surveillance. NOD2 recruits the protein autophagy-related 16-like 1 (ATG16L1) to the plasma membrane at the bacterial entry site. Cells with mutant *NOD2* are incapable of this directive ATG16L1 recruitment and consequently fail to entrap pathogens through autophagy. As a result, bacteria tend to persist abnormally in the mucosal cells causing inflammation.

Strong evidence of association with CD has been established through GWAS in the autophagy genes, *ATG16L1* and immunity-related GTPase family M protein (*IRGM*)[60,184,185]. The SNP rs2241880 in *ATG16L1*, a missense mutation leading to a threonine to alanine substitution at position 300 (T300A), has been shown to be strongly associated with CD risk[185]. Patients with CD homozygous for this risk variant exhibit a number of impairments in their bacterial handling; failure of autophagosome formation, defective antigen presentation to T cells, inefficient Paneth cell granule exocytosis, increased inflammatory cytokine expression by the Paneth cells, generation of increased levels of reactive oxygen species (ROS) and inflammasome over-activation[184,185]. *IRGM* is the only human gene with innate immunity related GTPases, which are necessary for gamma-interferon mediated resistance to pathogens. IRGM expression is clearly a requirement during initiation of autophagy for an efficient clearance of the bacteria. Reduced function of this gene could therefore lead to persistence of intracellular pathogens, resulting in chronic tissue damage and inflammation[186,187].

1.5.2.6 Endoplasmic reticulum stress and IBD

Endoplasmic reticulum stress is a phenomenon wherein there is an imbalance in the cellular processes that regulate protein processing. A number of stimuli including infections and non-specific imbalances in cellular homeostasis can lead to ER stress, leading to accumulation of mis-folded or unfolded proteins. To be able to cope with this

stress, eukaryotic cells have evolved a mechanism called unfolded protein response (UPR) to restore cellular homeostasis. UPR is achieved through cessation of protein translation, degradation of abnormal proteins and increasing production of molecular chaperones involved in protein folding. UPR is primarily cytoprotective, however it can trigger apoptosis after sustained ER stress. Perturbation in the UPR pathways has been suggested as a mechanism in the pathogenesis of IBD[188]. *XBP1* gene on 22q12 is critical in the protein quality control mechanisms of the UPR pathway. Hypomorphic variants of *XBP1* have been implicated as risk factors for IBD[189].

1.5.3 Adaptive immunity and crucial signalling pathways in IBD

1.5.3.1 Th17/IL-23 cell signalling axis

The discovery of Th17/IL-23 axis marked a milestone in the understanding of IBD-associated inflammation and prompted the development of potent therapeutic targets for IBD[3]. The axis represents a critical portal for cross-talk between the adaptive and innate immune mechanisms. Th17 cells, which produce highly potent pro-inflammatory cytokines such as IL-17, are abundantly found in the mucosa of patients with active IBD[190,191]. It is believed that Th1, Th2, Tregs and Th17 cells are all derived from a common precursor naïve $CD4^+$ T cells and differentiation into their respective T-cell lineages occurs following stimulation by their specific effector cytokines. The development of Th17 subsets from naïve T-cells is induced by the cytokines TGF-β and IL-6[192]. **Figure 1.13**. The amplification and stabilization of the differentiation process requires IL-21 and IL-23. Retinoic orphan receptor gamma t (RORϒt) has been identified as a critical regulator and a key transcription factor for this process, as is evident from reports that elevated expression of RORϒt has been found in the lamina propria of CD patients[193]. Th17 cell differentiation is also heavily influenced by the gut microbiota, which could be one of the reasons why these cells show a preferential homing to the GI tract. Th17 cells are characterised by RORϒt and IL-23R expression. Recently, it has been reported that CD161 could be a potential surface marker for disease promoting Th17 cells as high levels of these subsets were found in the inflammatory infiltrate of CD lesions[194]. The key feature of Th17 cells is the production of IL-17 cytokines. IL-17 is also produced by other cells including cytotoxic T-cells, neutrophils, eosinophils, NK cells and macrophages. The IL-17 cytokine family includes 6 members: IL-17A, IL-17B, IL-17C, IL-17D, IL-17E and IL-17F. IL-17A and IL-17F are the more extensively characterized members with highly potent pro-inflammatory effects[195]. The primary pro-inflammatory effects are mediated through

induction TNF-α, IL-1β, chemokines (CXCL8, CXCL1, CXCL10), GM-CSF, G-CSF, IL-6 and metalloproteases. Due to expression of receptor IL-17RA on a wide range of cells, the inflammatory backlash is seen on a wide range of cellular targets including the epithelium, endothelium, macrophages etc. Several other pro-inflammatory cytokines can amplify the effects of IL-17 A and F, especially TNF-α and IL-1β. In addition to the IL-17 cytokines, Th17 cells also produce other effectors such as IL-21, IL-22 and IL-9. IL-21 has a role in the differentiation of Th17 cells and can amplify Th17 cells response through a positive feedback mechanism. IL-22 and IL-9 other hand, can have ambivalent responses producing both pro-inflammatory and anti-inflammatory responses[196]. Several genes in the Th17 pathway have been linked with IBD susceptibility, including IL23R, TNFSF15, STAT3, IL12B, CCR6 and JAK2[197].

IL-23 is a heterodimeric potent cytokine which shares a molecular subunit p40 with IL-12 and is produced by a wide variety of immune cells. IL-23 is not crucial for the development of Th17 cells, but it is critically important for enhancing their survival and amplifying inflammatory effector responses. Mechanistic studies combined with the evidence derived from multiple GWAS hits in the IL-23/Th17 axis has generated intense interest in targeting this molecular pathway for the treatment of resistant IBD. Ustekinumab, a monoclonal antibody antagonizing the p40 subunit shared by IL-23/IL-12, has shown excellent clinical results and is now widely used in adult age group-based IBD protocols, particularly after failure of anti-TNF α therapy[44,198].

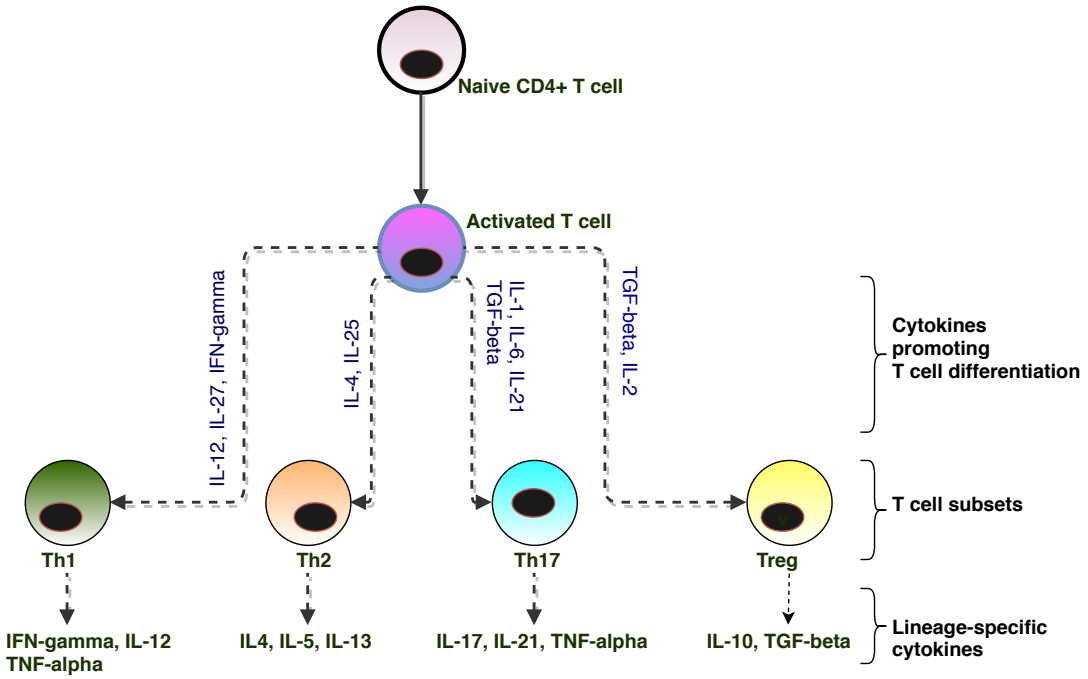


Figure 1.13 Development of the T cell subsets

The T-cell subsets shown in the diagram arise from a common precursor naïve $CD4^+$ T cell. The T cell differentiation pathway is influenced by specific stimulatory conditions. Differentiation into their respective T-cell lineage occurs following stimulation with their respective effector cytokines. The pattern of cytokine expression as indicated in the figure characterises individual T-cell subsets and their specific roles in host immunity.

1.5.3.2 IL-10 signalling pathway in IBD

Interleukin-10 is secreted by a wide variety of cells and over the last many years, it has been identified as crucial anti-inflammatory cytokine essential for maintaining gut homeostasis. IL-10 restricts the secretion of pro-inflammatory cytokines such as TNF-α and IL-12[199]. IL-10 signalling is required for limiting the expansion of Th17 cells, which are pro-inflammatory mercenaries[200]. The IL-10 receptor (IL-10R) has two α sub-units and two β subunits, which are encoded by IL-10RA and IL-10RB respectively. A downstream signalling cascade ensues following engagement by IL-10 with its receptor, resulting in activation of the receptor associated janus tyrosine kinases JAK1 and TYK2. This leads to phosphorylation and nuclear translocation of STAT-3 (signal transducer and activator of transcription 3), resulting in induction of STAT-3 dependent genes[199,201]. Glocker *et al*[199]in 2009 identified mutations in genes encoding the IL-10R subunit proteins in patients with very early onset IBD. The group identified three distinct homozygous mutations through linkage analysis and candidate gene sequencing in 4 out of 9 patients with infantile-onset IBD. This subset of patients presented with features of a very severe form of IBD at a very young age (< 1 year), with poor response to conventional therapy and requiring surgical treatment. One of the patients underwent an allogeneic stem cell transplant with reported success[199]. Subsequently, further studies have been reported in patients with homozygous mutations in the IL-10 signalling with a reported sustained remission after bone marrow transplantation[201-203].

1.5.3.3 JAK-STAT portal in IBD

The Janus kinase/signal transducer and activator of transcription (JAK/STAT) pathway constitutes a major portal for vital cellular processes including cell growth, differentiation, proliferation and several immune mechanisms[204]. Various cytokines and effectors communicate through this pathway to orchestrate an appropriate cellular response through target gene expression[205]. In mammals, JAKs (Janus kinases) represent a family of four non-receptor tyrosine kinases, JAK1, JAK2, JAK3 and TYK2. The JAKs were identified between 1989 and 1993, and named after Janus, the two-faced Roman God of gates,

doorways and duality[206]. This is because the JAK proteins have both a catalytic and a pseudo-kinase domain. The JAK molecules are located at the intra-cellular tail of the transmembrane cytokine receptors. Interaction with their respective cytokines results in the phosphorylation of the JAKS and the cytokine receptor chains, forming binding sites for the STAT (signal transducers and activators of transcription) proteins[206]. STATs comprise a family of seven structurally and functionally related proteins: STAT1, STAT2, STAT3, STAT4, STAT5A, STAT5B and STAT6. Once activated, STATs dimerize (either as homo or heterodimers) and translocate to the nucleus and modulate the expression of target genes[207,208]. Figure **1.14.**

Given its critical role in the intracellular signalling, aberrations in this crucial pathway have been implicated in several human diseases, including IBD[204,206].GWAS have implicated several genes in JAK-STAT pathway as candidate genes for IBD[60]. Some of these include JAK2, TYK2, STAT3 and STAT4, with genetic variants associated with an increased risk of developing IBD. The expression of JAK/STAT family proteins has been widely investigated in IBD through functional studies, particularly the role of STAT3. High STAT3 activity and expression has been shown to correlate with disease activity in patients with IBD[60,197].

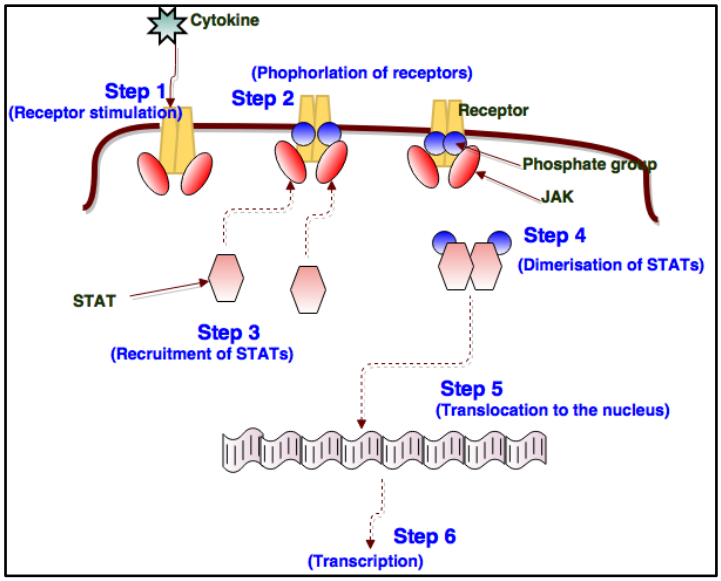


Figure 1.14 A schematic representation of JAK-STAT pathway

Figure describes the sequential steps in the JAK-STAT pathway:1) stimulation of the specific cell surface receptor by its respective cytokine, 2) activation of the JAKs resulting in phosphorylation of the tyrosine residues of the receptor, 3) receptor phosphorylation induces recruitment of the STATs to the receptor followed by their phosphorylation by JAKs, 4) STATs undergo dimerization (homodimers or heterodimers), 5) dimers translocate to the nucleus followed by 6) transcription. Within the nucleus, they bind to specific promoter regions of their target genes and modulate gene

*transcription. (**Abbreviations:** JAK- Janus kinases, STAT- signal transducer and activators of transcription)*

1.6 ENVIRONMENTAL FACTORS IN IBD

Although genetic factors have been highlighted as the key determinants of the risk of IBD, there has been an increasing evidence to support the role of environmental factors in the pathogenesis of IBD. The globally observed trends of the rising incidence of IBD is a clear indication that the cumulative risk in any given population exceeds that which can be explained by genetic drift[65,66,209].

1.6.1 Microbiome

An area that offers particular promise to the understanding of IBD pathogenesis with consequent implications on therapeutics is the interaction between an individual's genome and the environment[210]. Intestinal dysbiosis with reduced microbial diversity and stability has been increasingly reported in the pathogenesis of IBD[4,211]. Although clear differences in the composition of gut microbiota have been identified between diseased and healthy individuals, no single micro-organism has so far been consistently implicated in gut dysbiosis in IBD[212]. Disease state has been associated with a decrease in the abundance of several taxa[65,213]. Microbiome studies have demonstrated that IBD is associated with a reduced composition of certain microbial species such as bacteroides, firmicutes, ruminococcaceae and bifidobacterium, and an increased presence of Escherichia coli and fusobacterium in the gut lumen[211]. It is however difficult to decipher whether the altered composition of the microbiome is the trigger to disease pathogenesis or if it's the result of an extensive inflammatory response. Furthermore, even amongst healthy individuals the exact composition of the microbial flora is heavily dependent on the diet consumed and several other environmental factors. Even within families sharing both genetics and a living space, marked differences have been observed in the microbiome and also within the same individual at different time points.

Several clinical studies have been conducted on the feasibility of faecal microbial transplantation (FMT) as a treatment option for IBD[214]. The concept underpinning this approach is based on the notion that FMT reduces dysbiosis in the bowel through alteration of the microbiome using stool from a healthy donor and transplanting into the bowel of patients with IBD[215]. A systematic review and meta-analysis of studies investigating the role of faecal microbiota transplantation in IBD demonstrated overall clinical remission

rates of 36% and 50.5% for UC and CD respectively. However, the long-term follow-up, durability and safety profile data were unclear[216].

1.6.2 Perinatal factors

The rate of Caesarean section has more than doubled over the last two decades[217] and studies have shown that the rise in IBD probably parallels the changes in the mode of delivery[65,218,219]. The basic principle underpinning the philosophy relates directly to the influence of the mode of delivery to the composition of the intestinal microbiota. Babies born through a Caesarean section acquire a skin-type microbiota, whereas babies born through a vaginal delivery acquire vaginal microbiota[65,220]. A study investigating the microbial composition of the gastrointestinal tract in 7 year olds showed significantly reduced clostridial species in children born through a Caesarean delivery[221]. Although individual studies investigating the relationship between the method of delivery and the risk of IBD have yielded conflicting results, the pooled results of a meta-analysis suggested that delivery through a Caesarean section was a risk factor for adult and Paediatric onset Crohn's disease but not ulcerative colitis[218].

1.6.3 Use of antibiotics

There is increasing evidence that exposure to antibiotics in early life is associated with the risk of developing IBD[222]. However, the results from various epidemiological studies have shown disparate data[223,224]. Alteration of the gut microbiota, resulting in dysregulated immune responses in genetically susceptible individuals has been proposed as the underlying mechanism[225]. In a meta-analysis including fifteen observational studies, an association between antibiotics exposure and the development of Crohn's disease was demonstrated, but not with UC[225]. Paradoxically, in another study conducted in Asia-Pacific, including 442 patients and 940 age-matched controls, antibiotic exposure before age 15 was associated with a protective effect on IBD[224].

1.6.4 The hygiene hypothesis

The hygiene hypothesis is based on the observation of an increased incidence of IBD coincident with improvements in sanitation. This hypothesis fits conceptually with the observation that living and growing up in an environment with reduced exposure to microbes, results in ineffective priming of the immune system and consequently impaired immune responses[226]. As an example, supportive of hygiene hypothesis, early-life hygiene

factors could be contributory to the decreasing incidence of intestinal tuberculosis, but an increased incidence of IBD in migrant populations translocated to industrialised countries[227]. Other factors have also been examined as proxy markers of environmental exposures in early life including *helicobacter pylori* (*H. pylori*) infection, family size, birth order, urban upbringing and pet exposure[228]. For example, *H. pylori* infection and colonisation has been associated with household overcrowding and poor hygiene facilities. IBD is more prevalent in developed countries where *H. pylori* infection is less common[229]. *H. pylori* infection may be protective against IBD by upregulating the function of regulatory T cells[230]. Children raised with fewer siblings may have fewer exposures to infections necessary to prime the immune system to subsequent pathogenic challenges. In one population-based study, CD patients were shown to live in smaller households with fewer siblings[231]. Children raised in an urban setting tend to have a more 'hygienic' upbringing than those living in rural locations such as farms, although a robust link with IBD has not been demonstrated through clinical studies[232].

1.6.5 Dietary factors

Several observational studies have attempted to explore the role of dietary ingredients as risk factors for IBD. Diet is considered as an essential environmental factor influencing immune-mediated mechanisms in the host by directly impacting on the composition of the gut microbiome as well as through mechanisms involving intestinal permeability[63]. However, understanding the precise role of dietary factors has been limited by difficulties in accurately capturing food intake and the complexity of interactions between various foods[65,233]. Dietary ingredients with the best available evidence to date include dietary fats, gluten, maltodextrin, animal proteins and emulsifiers[234-237]. Studies in animal models have shown reduced mucin expression in the intestine, increased permeability and dysbiosis with high fat and high sugar diets. Dietary fat accelerated ileitis, whilst high gluten intake enhanced intestinal inflammation by inducing increased intestinal permeability[65,235,237]. Certain dietary elements including carbazoles (found in cruciferous vegetables), tryptophan-enriched proteins (vegetables, fish) and soluble fibres (complex carbohydrates) are protective through their anti-inflammatory properties[238]. The role of exclusive enteral nutrition in CD is well-established. The key mechanisms promoting mucosal healing with enteral nutrition include correction of dysbiosis, enhancement of epithelial barrier integrity, downregulation of pro-inflammatory cytokines and modulation of gastro-intestinal immune responses[65,239,240].

1.6.6 Geographical factors

A higher incidence of IBD has been observed with increasing latitude of residence. In two large, prospective studies in women enrolled in the Nurse's Health Study I (NHS I, 1976) and the NHS (II, 1989) in the USA, a higher incidence of IBD was observed in women residing in the northern latitudes[241]. An association of IBD with higher latitudes has also been reported in European studies. A leading suggested explanation for the latitudinal gradient is the reduced exposure to sunlight with consequent lower vitamin D levels at higher latitudes[65,242,243].

1.6.7 Effects of migration

Although the overall incidence of IBD world-wide is increasing, global trends indicate lower incidence of IBD in developing countries compared to Western countries[244]. Potential factors contributing to this observation include the hygiene hypothesis, sun exposure and dietary factors influencing the gut microbiome[65]. There is good evidence that moving to areas of high-incidence increases the risk of IBD, particularly UC[245-247]. A large population study of South Asian immigrants to the UK indicated that IBD incidence among first-generation immigrants was much higher when compared with the population in South Asia[245]. Subsequent follow-up studies showed the incidence rates to be even higher[246]. In a Canadian study, the disease phenotype among the first-generation immigrants was milder compared to Caucasians, however in the second-generation immigrants the disease phenotype was as severe and also showed an increased perianal involvement in patients with CD[246,247].

1.7 SUMMARY

The aetiology of IBD is complex involving multiple genetic, immunological and environmental factors. Over the last two decades, remarkable progress has been achieved in the understanding of the immunological mechanisms and the genes implicated in the pathogenesis of IBD. In excess of 230 genomic loci have been identified through GWAS and other genetic studies, which harbour genes altering susceptibility to IBD. This gene-set is enriched for those encoding proteins of the innate and adaptive immune system, epithelial barrier function, endoplasmic reticulum stress, autophagy and other cellular processes. Despite the success of GWAS in IBD, a substantial proportion of heritability remains unexplained. The missing heritability could be due to rare variation not detected through GWAS, limited ability of GWAS in identifying epistasis and environmental

factors contributing to disease. Advancements in NGS have increasingly proven successful in identifying rare causal variants and taking our understanding of IBD to the next level. However, the precise identification of the aberrant immune mechanisms in individual patients remains a formidable challenge. Also, the functional significance of the multiple genetic variants detected remains elusive. Current global trends indicate an increasing incidence in IBD, however the cumulative risk of developing the disease in any given individual cannot be explained by genetic variants alone. The overarching influence of environmental factors on the gut microbiome, consequently modulating the immune responses in a genetically susceptible individual plays a key role in the pathogenesis of this complex disease. With the emergence of robust high throughput technologies, multi-omic profiling of individual patients for a futuristic personalised therapy remains an ongoing research endeavour. However, given the molecular complexity of IBD, the greatest challenge lies in the integrative analysis across the multi-level omics data to define meaningful clinical models to guide personalised therapy.

Section II

Chapter 2 A SYSTEMATIC REVIEW OF IMMUNOLOGICAL AND GENETIC STUDIES IN IBD

This chapter is a systematic review of functional studies conducted alongside genetic studies from a selected panel of genes of known biological importance in IBD. The review which was published in 2014, explores the strengths of integrating genetics with functional studies to assess the causality of the multitude of genetic variants implicated in IBD. The review highlights the paucity of mechanistic studies investigating the functional importance of potential candidate genes and variants in the pathogenesis of IBD. The findings of this systematic review were important from my own research perspective, providing a direction towards conducting research in the field of immunogenomics of IBD. A further update of the literature review up to December 2018 is also included in this chapter.

2.1 Introduction

With the recent advances in genomic medicine, in excess of 230 loci have been implicated in the aetiology of IBD[3]. However, a clear interpretation of the functional relevance of the many newly discovered genetic variants is limited by a palpable lack of mechanistic studies to investigate the functional impact of these variants. Although several candidate genes identified through genetic studies can be mapped on to known biological pathways, a substantial fraction of these genes (>40%) are poorly characterised at the functional level[248]. The main objective of this chapter is to present a systematic review of published literature, detailing functional studies conducted concurrently with work-up to investigate genes of known biological importance in IBD. The secondary objective is to explore the potential of integrating genomics with functional studies in assessing the causality of the ever-expanding catalogue of variants identified through genetic studies.

2.2 METHODS

2.2.1 Prioritisation of genes for conducting a systematic search

This study was conducted in February 2014 at which time approximately 163 IBD genomic loci were known, following publication of a widely cited meta-analysis of GWAS in IBD[60]. Methods to select genes for our systematic search were adapted from this paper published in 2012, in which five network connectivity tools were used to prioritise candidate genes within the genomic loci. These included GRAIL (Gene Relationships Across Implicated Loci), DAPPLE (Disease Association Protein-Protein Link Evaluator), three different sources of eQTLs (expression quantitative trait locus), coding SNPs and Co-expression network analysis. GRAIL identifies relationships between genes in different disease associated loci through text searching of abstracts in the published scientific papers; DAPPLE examines disease-associated single nucleotide polymorphisms (SNPs) by compiling protein-protein interaction data from various sources; cSNP indicates where associated SNPs directly alter the protein, or are in strong linkage disequilibrium with known SNPs which alter the protein; eQTL denotes genes where associated SNPs are correlated with an alteration in protein expression levels and co-expression refers to genes for which expression patterns are linked to known inflammatory processes[60]. Given the individual limitations of each of the tools utilised for functional clustering of genes, we adopted a consensus approach, requiring genes to be simultaneously implicated by at least two of the *in silico* methods. This approach narrowed our analysis to seventy-one candidate genes. Figure **2.1** and table **A.1** in Appendix A).

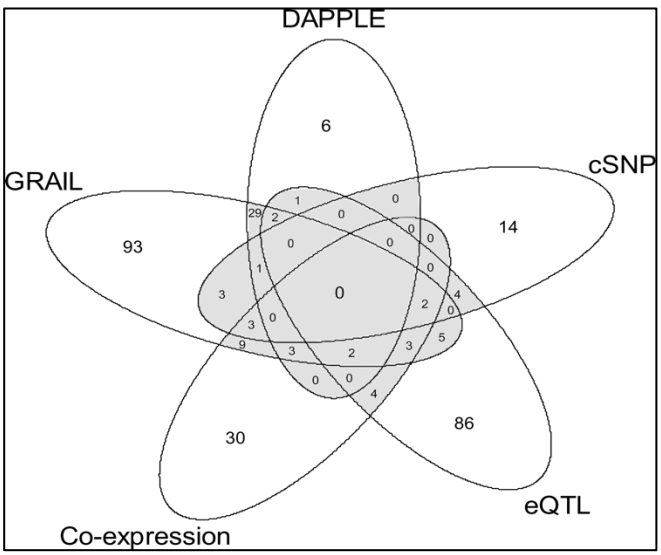


Figure 2.1 Selection of genes for the systematic search

For the selection of genes, a method employing five network connectivity tools was adapted from a previously published meta-analysis of genome-wide association scans in IBD. The Venn diagram shows the five bioinformatic tools, including GRAIL, DAPPLE, cSNP, Co-expression and eQTL. For the systematic search, genes were selected if identified by at least two out of the five network connectivity tools. Based on this method, seventy-one genes out of three hundred were prioritised as shown in the shaded region. (Abbreviations: cSNP, coding single nucleotide polymorphisms; DAPPLE, Disease Association Protein-Protein Link Evaluator; eQTL, expression quantitative trait locus; GRAIL, Gene Relationships Across Implicated Loci).

2.2.2 Systematic search

Two authors (Tracy Coelho and James Ashton) independently conducted an electronic search through ovidsp using MEDLINE and EMBASE from 1996 to February 2014 for each of the 71 genes prioritised for the systematic review. A uniform search strategy was adopted for all the selected genes combining primary and secondary search terms using Boolean operators. The primary search terms used for all genes included "IBD" or "inflammatory bowel disease" or "Crohn's" or "CD" or "ulcerative colitis" or "UC". The secondary terms specific to each of the genes included the approved gene name as on 'HUGO Gene Nomenclature Committee', gene symbol, synonyms or gene abbreviations. The electronic search was conducted using a combination of all primary and secondary search terms per gene. No language restrictions were applied. An example is shown in table **2.1**.

Table 2.1 An example of systematic search on *ATG16L1* gene

1. IBD*.tw.
2. Inflammatory bowel disease*.tw.
3. Crohn*.tw.
4. CD.tw.
5. UC.tw.
6. Ulcerative colitis.tw.
7. 1 or 2 or 3 or 4 or 5 or 6
8. ATG16L1.tw.
9. autophagy related 16-like 1.tw.
10. "APG16 autophagy 16-like (S. cerevisiae)".tw.
11. APG16L.tw.
12. "ATG16 autophagy related 16-like (S. cerevisiae)".tw.
13. "ATG16 autophagy related 16-like 1 (S. cerevisiae)".tw.
14. ATG16L.tw.
15. ATG16A.tw.
16. FLJ10035.tw.
17. WDR30.tw.
18. 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17
19. 7 and 18

An example of systematic search for the gene ATG16L1 is shown in this table. Numbers 1-6 indicate all publications retrieved using the primary search terms and number 7 combines all the primary search terms from 1-6. Number 8 includes approved gene symbol, number 9 approved gene name, 10-14 previous names of the gene, 15-17 include synonyms for the gene and number 18 combines all the search terms from 8-17. Number 19 finally includes all publications retrieved by combining the primary search terms for IBD (number 7) and search terms for the ATG16L1 gene (number 18).

2.2.3 Selection of studies

Only those genetic studies with concurrent mechanistic/immunological work-up to assess the functional impact of genetic variants in patients with IBD were included. Studies using animal models or cell lines without human data, review articles without original data and conference abstracts were excluded. The titles and abstracts of articles were screened, assessed for their relevance and duplicates were excluded. Selected papers identified after the initial screen, were retrieved in full text and the reference lists scanned to look for other relevant publications. At each stage of the study selection, the eligibility of publications was assessed independently by the two reviewers. Any disagreement was resolved through discussion and a consensus approach.

2.3 RESULTS

A structured search across the seventy-one prioritised genes identified 35142 potentially eligible citations between the two reviewers, describing genetic and functional studies in IBD. Eight out of the seventy-one genes had publications meeting the inclusion criteria. These included sixty-seven studies, of which the *NOD2* gene had forty-one immuno-genetic studies followed by *IL-10* with eleven eligible studies. As most of the immuno-genetic studies on IL-10 included genes encoding IL-10 receptor sub-units (*IL10RA* and *IL10RB*), studies describing these genes were also included in the final analysis. The other genes with eligible publications were *ATG16L1* (7 studies), *SLC22A4* (3 studies), *IL23R* (2 studies), and *SLC11A1, CCL2* and *STAT3* had one each. See figure **2.2.**

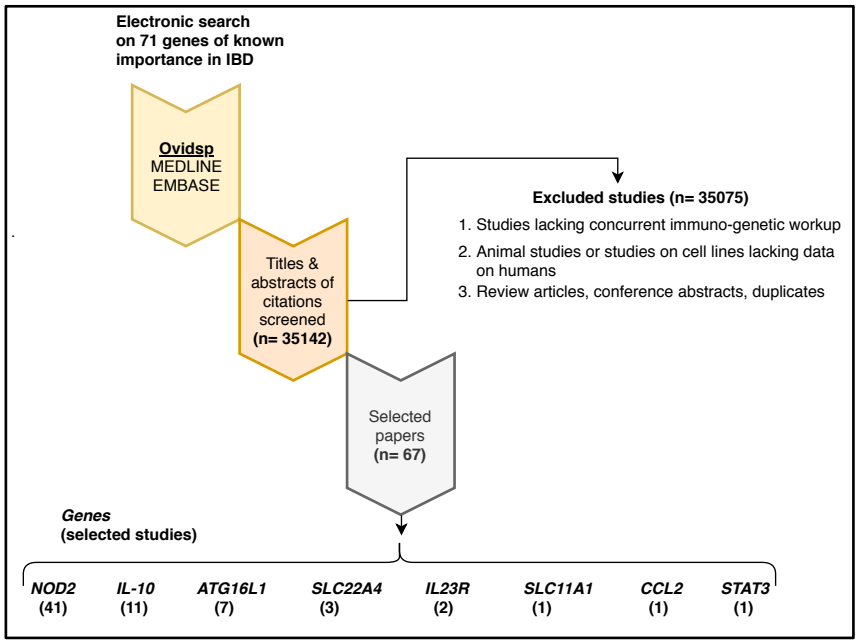


Figure 2.2 Systematic search on seventy-one prioritised genes

An electronic search conducted using MEDLINE and EMBASE on the seventy-one prioritised genes identified 35142 publications of which only 67 studies met the inclusion criteria. Eligible studies were identified for only eight out of the seventy-one genes. The number of studies selected per gene are shown in the lower section of the figure.

An overview of all the studies identified on *NOD2*, *IL-10* and *IL-10* receptor genes is presented in table **2.2**. Out of 1670 manuscripts screened for the *NOD2* gene, 19% included genetic studies without functional work-up and 3.7% were immunological studies without genetic analysis. Of the 2354 publications screened for *IL-10* and *IL-10* receptor genes, 12.3% (289) were immunological and 2.3% genetic studies, although only 11 studies (0.5%) met the eligibility criteria which included concurrent immuno-genetic work up. Approximately 6% and 36% of the studies on *NOD2* and *IL-10* genes respectively were conducted as pre-clinical experiments in animal models.

Table 2.2 Electronic search on NOD2, IL-10 and IL-10 receptor genes

Studies	*NOD2*	*IL-10/IL-10RA/IL10RB*
Total number of publications screened	1670	2354
* Selected Studies	41 (2.5 %)	11 (0.46 %)
Functional studies in humans (no genotyping)	62 (3.7 %)	289 (12.3 %)
Human genetic studies including GWAS and meta-analysis (no functional studies)	315 (18.8 %)	55 (2.3 %)
Studies on animal models (genetic or functional)	98 (5.8 %)	845 (35.9 %)
Studies on cell lines (genetic or functional)	22 (1.3 %)	22 (1 %)
Reviews (No original data)	81 (4.8 %)	59 (2.5 %)
Publications not pertinent to our review and duplicate studies	1051 (63 %)	1073 (45.5 %)

Table 2.2 presents a synopsis of all studies identified through a systematic search for NOD2, IL-10 and the genes encoding the IL-10 receptor sub-units. The greyed-out selection includes studies meeting the eligibility criteria, requiring studies to report both genetic and functional data concurrently in IBD patients.

2.3.1 Selected studies on the *NOD2* gene

Forty-one manuscripts were included in the final qualitative analysis for the *NOD2* gene, detailing genetic and immunological work up together, for patients with IBD. Details of these selected studies are presented in table **A.2** in Appendix A. All the selected studies on the *NOD2* gene were reported from the year 2002 onwards, driven by the landmark discovery of the *NOD2* gene in 2001 as the first gene conferring susceptibility to IBD[87,249,250]. The NOD2 functions an intra-cellular pathogen-recognition receptor (PRR) through recognition of muramyl dipeptide (MDP), triggering an inflammatory cascade via the NF-κB pathway. The majority of functional studies on NOD2 therefore focus on the assessment of cytokines induced via this inflammatory pathway[124,251-256]. Some studies have shown impaired NF-κB activation in response to MDP by detecting reduced phosphorylation levels in nuclear extracts of cells stimulated by MDP[257,258]. Rahman *et al* performed studies on regulatory T-cells isolated from patients with Crohn's disease harbouring mutations within the *NOD2* gene, patients with the wild type allele and healthy volunteers. The study demonstrated that regulatory T-cells when activated by MDP are normally protected from apoptosis, however the protective effect was not observed in

individuals with *NOD2* mutations[259]. A study by Hedl *et al* in 2007 observed that monocyte-derived human macrophages when pre-treated with NOD2 stimulants such as MDP show a reduced induction of pro-inflammatory cytokines upon re-stimulation with MDP, and stimulation with TLR2 and TLR4 agonists. The cytokines analysed in this study included TNF-α, IL-8, and IL-1β. In health, macrophages exhibit cross-tolerance to PRR stimulation, which is reflective of their ability to adapt to chronic stimulation. However, NOD2-stimulated macrophages isolated from individuals with Crohn's disease homozygous for Leu1007insC *NOD2* mutations were deficient in their ability to cross-tolerize to subsequent treatment with TLR2 and TLR4 ligands[255]. On the same lines, NOD2/TLR crosstalk has also been demonstrated by work from other groups[251,260-266]. Some studies have proposed cross-tolerance and synergy between NOD1 and NOD2 signalling, suggesting that *NOD2* mutations could lead to down -regulation of NOD1 signalling[267]. More recent studies have indicated impaired bacterial handling and autophagy induction in individuals with *NOD2* mutations. Cooney *et al* showed that individuals with homozygous *NOD2* mutations are unable to recruit ATG16L1 to the plasma membrane, resulting in defective engulfment of invading bacteria by autophagosomes. Their findings indicated that patients harbouring *NOD2* mutations have impaired mechanisms involving autophagy induction, bacterial trafficking and antigen presentation[268].

In terms of the methods used for genetic analysis, thirty-eight of the selected studies performed genotyping for *NOD2* variants, three studies performed candidate gene sequencing and none employed next generation sequencing. Immunological work-up was conducted using blood specimens in thirty-four studies, gastro-intestinal (GI) tissue in five studies and a combination of peripheral blood cells and colonic tissue in two studies.

2.3.2 Review of studies on *IL-10*, *IL-10RA* and *IL-10RB* genes

From a total of 2354 publications, only eleven manuscripts, focusing on IL-10 related genes met the inclusion criteria. See **A.2** in Appendix A. for the selected publications. The immuno-genetic studies conducted to assess the functional integrity of IL-10-mediated signalling focused on the functional capacity of IL-10 to limit the induction of IL-6, IL-8, IL-1β, TNF-α and other pro-inflammatory cytokines. The studies highlighted the importance of IL-10 as a potent anti-inflammatory and regulatory cytokine, with mutations resulting in the abrogation of IL-10-dependent negative feedback regulation with a consequent hyper-secretion of pro-inflammatory cytokines. Koss *et al* (2000) investigated

the influence of biallelic polymorphisms in the genes encoding TNF-α, lymphotoxin-alpha and IL-10 on cytokine production patterns. IL-10 and TNF-α cytokine induction was analysed following activation of whole blood specimens with lipopolysaccharide (LPS), a TLR4 receptor stimulant. It was shown that there were differences in the effects of genetic mutations on cytokine production patterns in individuals with Crohn's disease and ulcerative colitis compared to controls[269]. Gasche *et al* in 2003 screened patients with Crohn's disease and healthy controls for mutations in the *IL10RA* gene. They also assessed the ability of monocytes in these individuals to inhibit TLR4-mediated IL-10 production. The study data indicated that individuals carrying mutations in the *IL10RA* gene were less sensitive to IL-10 mediated TNF-α production[270]. Van der Linde *et al* identified a point mutation (Gly15Arg) in the leader sequence of *IL-10* following genotyping of *IL-10* alleles in seventeen sibling pairs with CD and seventy-five healthy controls. The functional impact of these variants was examined by stimulating peripheral blood mononuclear cells with LPS or phorbol ester, and then quantifying IL-10 induction from the cellular supernatants by ELISA or Western blotting. The activity of recombinant immature wild type or mutated *IL-10* was also tested in vitro in a proliferation assay with LPS-stimulated human monocytic cell line (HL60 cells). Patients harbouring the point mutation showed impaired IL-10 secretion with a consequent reduced anti-inflammatory activity [271]. In line with previously published data, Glocker *et al* in 2009, performed genetic linkage analysis and candidate gene sequencing on two independent consanguineous families with children presenting with very early onset severe IBD. Six additional patients who presented with severe IBD-like disease in the first year of life were also screened for mutations in *IL-10RA* and *IL-10RB*. Three distinct homozygous mutations were identified in four out of the nine patients. Immunological assays using PBMCs were conducted to assess the impact these mutations *in vitro*. Patients carrying these mutations showed deficient STAT3 phosphorylation on stimulation with IL-10 and consequently reduced induction of STAT3-dependent genes. Increased secretion of TNF-α and other pro-inflammatory cytokines was observed in the patients with the IL-10 receptor mutations, consistent with the hypothesis that the IL-10 dependent negative feedback mechanisms which down regulate hyper-inflammatory responses are disrupted in these patients[199]. In line with this concept and findings, other studies have also reported defective down-regulation of pro-inflammatory immune responses in patient harbouring mutations in the IL-10 associated genes[203,272-274].

Of the eleven selected studies, whole exome sequencing was used in one study[273], candidate gene sequencing in six studies[199,203,271,272,274,275]and SNP genotyping in four

studies[269,270,276,277]. Specimens analysed for the immunological studies included blood samples such as PBMCs, monocytes, whole blood or serum in eight studies and GI tissue in addition to blood samples in three studies.

2.3.3 Immuno-genomic update since 2014

As an update to the systematic review which was published in 2014, a further electronic search was conducted using MEDLINE via ovidsp from January 2014 to December 2018. Given the substantial rise in the number of candidate genes in IBD over the last few years and for practical reasons, the online search was conducted using broad terms and Boolean operators, without including specific gene names. See table **2.3**.

Table 2.3 Combination of keywords for the systematic review update

1	IBD.tw.
2	Inflammatory bowel disease.tw.
3	Crohn's.tw.
4	ulcerative colitis.tw.
5	1 or 2 or 3 or 4
6	gene.tw.
7	genes.tw.
8	genomics.tw.
9	genetics.tw.
10	immune.tw.
11	immunology.tw.
12	6 or 7 or 8 or 9 or 10 or 11
13	5 and 12

Table 2.3 displays the search strategy used for the systematic review update for period between January 2014 and December 2018. Numbers 1-4 include broad search terms for IBD and number 5 combines searches from numbers 1-4 together. Numbers 6-1l include keywords for a broad search on genes, genetics and immunology, and number 12 represents a combination of these searches. Finally, number 13 combines publications containing keywords from both number 5 and number 12.

A total of 2231 publications were identified using the above search strategy. Animal studies not including human data, reviews and conference abstracts were excluded. The online search for the update was conducted by a single reviewer (TC). A total of thirty-four manuscripts were identified as meeting the eligibility criteria. See table **2.4** for an overview of the selected studies.

Table 2.4 Selected studies for the systematic review update

Implicated mechanisms	Genes	Number of publications
Epithelial		
Apical transport	*SLC26A3, SLC03A1*	2
Junctional proteins	*MAG13* and other genes encoding tight-junction proteins	1
Others	*TCN2*	1
Innate immunity		
NOD2 & associated proteins	*NOD2, TRIM22, XIAP*	4
Toll-like receptor	*TLR2, TLR5*	2
Complement	*MBL*	1
Cellular processes		
Autophagy	*ATG16L1, ATG16L2*	5
Oxidative processes	*Genes encoding NADPH oxidases (CYBA, CYBB, NCF1, NCF2, NCF4, RAC1, RAC2)*	1
Adaptive immunity	*PTPN22 & PTPN22, FOXP3,* genes encoding Th1 & Th17 cellular pathways	3
Innate & adaptive	*FCGR3A, ICOSLG, SHIP1*	3
Cytokines	*IL6R, IL10, IFNA4 & IFNA10, TNFA, TRAIL, TNFSF15*	7
Matricellular proteins	*MMP9*	1
Others	*SMAD3, HLA-DQB1,* other gene panels	3

The selected studies are shown against the corresponding implicated mechanisms and the genes involved. In the bottom row, the 'other gene panels' included studies investigating multiple genes alongside gene expression within the colonic tissue.

The selected studies investigated the various pathogenetic mechanisms in IBD including epithelial barrier integrity, cellular processes such as autophagy, innate immunity, adaptive immune mechanisms, multiple cytokines across the immune system and others. The most commonly investigated candidate genes were the autophagy genes, *ATG16L1* and *ATG16L2* with a combined total of five publications meeting the inclusion criteria. There were two publications each for the *NOD2*, *XIAP* and the toll-like receptor genes, and three for the *TNFA* and associated genes.

Several studies have implicated defective autophagy as a causative mechanism in IBD, particularly Crohn's disease. This was evident from the studies identified in the published systematic review as well as the update. Autophagy is a cellular recycling mechanism which results in the degradation of dysfunctional cellular components including bacterial products for an orderly intra-cellular clearance. This is achieved through engulfment and isolation of the cellular material within a double-membraned vesicle, known as

autophagosome followed by fusion with lysosomes and degradation of the cellular debris[278]. In a study by Wolfkamp *et al*, patients with IBD (CD=65, UC=34) and 8 healthy controls were genotyped for three of the CD-associated genes *ATG16L1*, *IRGM* and *NOD2*. Phagocytic activity of monocytes and granulocytes was measured by flow cytometry which showed a significantly higher percentage of actively phagocytic monocytes in CD patients, but not UC when compared to controls. Enhanced phagocytosis was observed in patients harbouring variants in the *ATG16L1* and *NOD2* genes. The authors proposed that enhanced phagocytosis leads to an increased accumulation of bacterial products within the cell, however defective autophagy in the context of mutant genes leads to poor intra-cellular clearance and therefore increased inflammatory responses[279]. Ma *et al* demonstrated that a SNP rs11235604 in the autophagy gene *ATG16L2* is functionally associated with CD in the Chinese population. The SNP was genotyped in 363 patients with CD and 486 healthy controls. The SNP was observed in a significantly higher proportion of patients compared to controls ($p=0.04$). A markedly reduced mRNA expression in T cells was observed among patients compared to controls ($p<0.001$) and in individuals harbouring the SNP ($p=0.005$)[280]. In a healthy state, the turnover of ATG16L1 is tightly regulated depending on basal apoptotic stimuli such as caspase-3 activity maintaining a balance between cell survival and apoptosis. In a study by Murthy *et al*, it was shown that caspase-3 can accelerate cleavage of ATG16L1 in the presence of common risk alleles including Thr300Ala (T300A), resulting in reduced basal autophagy and therefore increased inflammatory cytokine production. The authors therefore proposed that the therapeutic inhibition of caspase-3 activation could potentially restore autophagy and GI homeostasis through stabilisation of ATG16L1 production and activity[281].

Among the cytokines, selected immuno-genetic studies described the role of IL-6, IL-10, interferon alpha and TNF-mediated signalling in IBD causing inflammation through dysregulated cytokine production. IL6 is a pro-inflammatory cytokine which engages with its receptor IL6R resulting in the activation of inflammatory cascades. Parisinos *et al* recently demonstrated that individuals carrying the variant rs2228145 in *IL6R* have a 2-fold elevation of soluble IL6R (s-IL6R), reduced activation of IL6R-mediated signalling and consequently a reduced risk of developing IBD (OR= 0.876, p= 0.0003). The effects observed were similar to what would be expected if a monoclonal antibody was used to antagonise IL6R, thereby paving the way for potential therapeutic agents designed to treat IBD by blocking IL6R-mediated signalling[282]. Xiao *et al* identified two heterozygous variants in *IFNA4* and *IFNA10* genes in a Chinese cohort using exome sequencing and

further validated by Sanger sequencing. These variants were significantly associated with CD and patients carrying these variants had significantly lower levels of interferon alpha 10 (IFNA10). Interferons have several functions including immuno-modulatory, anti-proliferative and anti-viral, and these functions were shown to be impaired in the context of these variants[283]. Hedl *et al* investigated the interactions between tumour necrosis factor ligand superfamily member 15 (TNFSF15) and death receptor-3 (DR3), a pathogen recognition receptor in macrophages isolated from patients with IBD. Human macrophages expressed DR3 and the TNFSF15: DR3 interaction was observed to amplify downstream signalling via the MAPK and NF-κB pathways. Macrophages from individuals carrying the risk allele rs6478108 showed an enhanced cytokine production through a gain of function, thereby defining yet another disease-relevant mechanism in individuals with IBD[284]. For further details on the selected studies, see **A.3** in Appendix A.

2.4 DISCUSSION

The systematic review identified a large number of genetic and immunological studies investigating the role of several candidate genes of biological importance in IBD. There was however a clear paucity of studies integrating immunological work-up with genetic findings and *vice versa*. The majority of selected studies (approximately 61%) focused on the role of the *NOD2* gene in IBD. The first IBD susceptibility locus (IBD1) was identified in 1996 on chromosome 16 through a genome-wide association scan[76] and replicated thereafter through several other studies, most notably including the analysis of large pooled data sets by the IBD International Genetics Consortium, demonstrating a strong linkage to this locus[285]. Two independent groups in 2001 subsequently identified the *NOD2* gene within this locus[87,249], following which several genetic and immunological studies were conducted investigating the role of this gene and other associated genes in IBD.

A significantly larger number of animal model studies for *IL-10* compared to *NOD2* (36% v/s 6%) were identified in this review. This was possibly due to the extensive use of *IL-10* knock out models for experimental colitis. Animal knockout models have contributed significantly towards understanding the molecular basis of IBD across a diverse range of genes and biological pathways. Although the functional impact of induced or spontaneous genetic mutations in defined biological processes can be assessed through experimental models, the findings may not effectively mirror the consequences of human genetic variants on immune function. Similarly, transformed cell lines are good experimental models for transfecting in a gene when the behaviour of the cell itself is not of interest.

However, this can be a major pitfall whilst trying to establish the actual functional impact caused by genetic variants in human disease. Also, transfection of cell lines can lead to over-expression of genes resulting in an unpredictable outcome[286]. Therefore, for a better understanding of the role of candidate genes in disease pathogenesis, mechanistic studies in human cohorts may be more meaningful.

There could be several reasons for the lack of combined immuno-genetic studies in IBD. IBD is a complex polygenic disease due to multiple mutations at the implicated loci in a given individual. Several genes encoding key immunological pathways may be affected with varying severity of contributory impact, thereby making it extremely challenging to design functional studies bespoke to the genetic findings in individual patients. Other reasons for the paucity of functional studies to follow up and investigate implicated genetic variants include the relative lack of enthusiasm to invest in conducting functional work in a GWAS variant where there is no proven causality. Also, given the diverse nature of genomics and immunology on the academic front, conducting immuno-genetic/genomic research requires a robust collaborative set-up across the specialities.

Our study had some limitations. Given the multiple number of genes prioritised, the biodiversity of their physiological roles and the heterogeneity of studies identified, a meta-analysis of studies was not performed for this systematic review. The review was limited to seventy-one genes based on the eligibility criteria. It is therefore possible that there are immuno-genetic studies in published literature on genes not prioritised for this review. Some of the commonly implicated genes such as *IRGM, ERAP2, MUC19, CDH1* and some others were not identified for the review based on the nature of the study design. Moreover, with the relentless pace of advancements in IBD genomics over the last few years, the list of candidate genes and loci has been rapidly increasing[60,287,288]. Given the heterogeneity of the genetic and functional methods used, selection of publications was subject to reviewer bias. This was however minimised through a standardised assessment, discussion and consensus at all stages of the review. A substantial number of studies evaluated a panel of genes for genetic profiling, rather than a single gene resulting in an unavoidable overlap between studies across the genes under assessment. For example, some of the selected studies were included for both *ATG16L1* and *NOD2* gene, as the studies evaluated both the genes from an immuno-genomic angle given that the two genes work closely together[268,289].

Comparing the four-year update with the previously published systematic review, it was evident there have been more studies investigating the functional impact of genetic

findings in IBD. Nevertheless, the efforts to further elucidate pathogenetic mechanisms in IBD through functional studies need to continue at an accelerated pace for a better understanding of the disease. Although the technological advancements in genomics have provided significant insights into the molecular basis of IBD through identification of several candidate genes and genomic variants, there is a specific need to verify and establish the causality of the implicated variants before they can direct clinical management. A collaborative approach integrating genomics, immunology and allied specialties will go a long way in establishing the causality of candidate genes, disease risk-prediction, disease progression and thereby enhance the prospects of a personalised therapeutic approach in IBD for the foreseeable future.

Chapter 3 RESEARCH SET UP AND GENERIC METHODS

This chapter details the recruitment of patients to the study, research set-up, processing of biological samples and generic methods. Patients for my PhD project were selected from the main study, the 'Genetics of Paediatric IBD'. Specific methods will be described in the respective chapters.

3.1 INTRODUCTION TO THE 'GENETICS OF PAEDIATRIC IBD' STUDY-SOUTHAMPTON

The Genetics of Paediatric IBD study Southampton is one of the largest in the UK recruiting children diagnosed with IBD for in-depth nucleic acid sequencing. The study is a registered portfolio member of the NIHR Clinical Research Network (CLRN ID 11158) which funds full-time research nurse support to facilitate recruitment of patients and family members. To date, in excess of 500 paediatric patients diagnosed with IBD aged 0-18 years, in addition to their parents, and affected relatives have been recruited. Since May 2014, following a substantial amendment to the protocol, adult IBD have also been recruited to the study. The research study is conducted at the University of Southampton Genomic Informatics group in collaboration with the Regional Paediatric Gastroenterology Unit, Southampton Children's Hospital, University Hospital Southampton (UHS) who sponsor the study, Immunology group based at the Wessex Investigational Sciences Hub (WISH) laboratory and the NIHR Biomedical Research Centre (BRC), Southampton.

3.1.1 Regional paediatric gastroenterology service

The Paediatric Medical Unit of UHS hosts the regional paediatric gastroenterology unit for Wessex (population 3.5 million), as part of a full range of tertiary paediatric gastroenterology services. Led by Professor Mark Beattie, this unit has a national reputation for clinical excellence and is accredited for training in paediatric gastroenterology and nutrition. The unit undertakes in excess of 1000 endoscopic procedures per year and approximately 60-70 new cases of IBD are diagnosed each year. All children are diagnosed using the Porto diagnostic criteria[2] and treated according to national and international guidelines[35,37,42,43].

3.1.2 Genomic Informatics Group

The Genomic Informatics Group within Human Genetics, led by Professor Ennis is based in the Duthie Building on the UHS site. Bioinformatic pipelines for analyses of NGS data including whole exome, whole genome and RNA sequencing are actively maintained to integrate advances in software, updates in reference datasets/sequences and optimised annotation and mapping techniques. The group is experienced in method development and analysis of exome data for various diseases including IBD[290-293]. The Genomic Informatics Group has a high-speed connection to the University of Southampton Iridis supercomputing facility (http://cmg.soton.ac.uk/iridis) and benefits from substantial processing capability and dedicated file space afforded by this cluster.

Prof Ennis is the chief investigator on the CRN portfolio approved study to recruit paediatric IBD patients and families. The Southampton Paediatric IBD research group is a part of the UK IBD genetics consortium. Laboratory facilities enabling DNA extraction, storage and targeted follow-up experiments to test analytical findings are available. DNA samples from recruited paediatric patients are extracted and stored locally for outsourced exome sequencing. Resulting raw exome sequencing data is transferred to the Genomic Informatics group for comprehensive bioinformatic analyses.

3.1.3 The Immunology Group

Immune function assays were undertaken at the immunology laboratory of the cancer sciences division based in the Somers building on the UHS site and within the Wessex Immunology Sciences Hub (WISH) laboratory, a purpose-built investigational sciences hub within the Southampton Centre for Biomedical Research. The WISH lab is a centralised interdisciplinary investigational sciences facility, undertaking research in translational immunology, cancer sciences, vaccine studies and inflammatory disorders. The immunological assays performed as part of my PhD were supervised by Professor Anthony Williams and his post-doc Dr Yifang Gao at the WISH laboratory. Professor Williams leads the Translational Immunology Laboratory, undertaking a broad portfolio of immunological analyses in cellular and humoral immunity in the context of numerous disorders of inflammation and immunity. A multidisciplinary team of scientists deliver investigational approaches for human evaluation that are compliant with Good Clinical Laboratory Practice (GCLP) across clinical cohort analyses and clinical trials. The group hosts translational research programmes in cancer immunotherapy, paediatric infectious

diseases and immunodeficiency, paediatric allergy as well as commercial partnerships to undertake immune-monitoring for international multi-centre vaccine trials.

3.2 RECRUITMENT TO THE STUDY

3.2.1 Ethics

Ethical approval for the wider study, the 'Genetics of Paediatric IBD' was obtained in August 2009 through the National Research Ethics Service (NRES), South central and Hampshire B, Regional Ethics Committee (REC Reference- 09/H0504/125). Recruitment of patients started in October 2010 and includes children and young people under eighteen years of age with a diagnosis of IBD, their biological parents and affected relatives. The main elements of the study following the initial approval included patient recruitment with detailed recording of clinical data, blood or saliva sampling to obtain DNA for exome sequencing and collection of plasma at the time of recruitment. Subsequently, four substantial amendments were drafted and submitted by myself as a clinical research fellow, which were successfully approved. These amendments enabled collection of additional specimens including blood samples for immunology assays and biochemical assessment, gastrointestinal (GI) tissue and stool specimens for microbiome analysis. Up to November 2013, patients were recruited to the study after an established diagnosis of IBD. Following a substantial amendment, we were able to recruit patients with suspected IBD before confirmation of diagnosis through endoscopy. This was crucial for my project as it enabled procurement of treatment naïve biological specimens to minimise or avoid the masking impact of drug treatments on the immune system. Additionally, we were able to use specimens for research from patients initially recruited as suspected IBD, but endoscopic investigations did not favour a diagnosis of IBD. These specimens were used as non-IBD controls. Furthermore, the amendments enabled recruitment of unaffected siblings of patients with IBD, affected relatives and adult patients with IBD, which facilitated other ongoing projects within the 'Genetics of paediatric IBD' study. Details of the amendments are presented in table **A.4** in Appendix A.

3.2.2 Patient recruitment

All patients diagnosed with IBD are entered on to a prospective clinical database by clinicians, which is a secure password protected resource accessible to paediatric gastroenterologists, gastroenterology nurse specialists and the research team. Patients are

primarily recruited through paediatric gastroenterology clinics, day ward attendances and inpatient admissions at the UHS. The recruitment of patients and their families is facilitated by a dedicated part-time research nurse or clinical research staff. During the initial screening, research study literature and the invitation to join the study is handed out to the family. During subsequent contact or following an opt-in reply, the family is approached in a clinical setting for a formal recruitment. At the time of recruitment, the study structure is explained in further detail, providing explanations and answers to any questions pertaining to the study. A study specific proforma is filled out, which includes a brief clinical history, family pedigree as appropriate, phenotypic characteristics at diagnosis, results of investigations and other clinical details. Informed written consent is obtained from parent/s and the child. On recruitment, research blood samples are obtained from the study subjects for DNA and plasma extraction. For my core project, which is primarily an evaluation of induced immune responses in peripheral blood mononuclear cells, blood samples were collected for PBMC extraction in addition to DNA and plasma samples from treatment naïve patients at or before diagnosis. For the patients included in my core project (chapter 5), recruitment and sample collection were conducted predominantly by myself with support from the research nurse.

3.2.3 Collection and storage of biological samples

3.2.3.1 Blood samples

Research blood samples were taken with routine clinical blood samples to avoid an additional venepuncture. No more than 25 mls of blood volume in children under 10 years of age, 30 mls in children over 10 years of age or 50 mls in adults were taken for research in any 6-week period. If further bloods were needed for research, a minimum period of 6 weeks was maintained between the research samples. Blood samples for DNA were collected in EDTA tubes, transported on ice, stored at -80° C pending DNA extraction and -20° C thereafter. Blood samples for plasma were collected in lithium heparin tubes, stored at room temperature up to 4 hours and at -80° C following extraction of plasma. Blood samples for peripheral blood mononuclear cells (PBMCs) extraction were collected in lithium heparin tubes, maintained at room temperature up to 4 hours from venesection until extraction of the PBMCs. Extracted PBMCs were then stored in cryovials at -160° C in liquid nitrogen tanks. For most patients, PBMCs extraction was conducted by myself, although in the third year of my project, I received support from the tissue bank technical team.

3.2.3.2 Saliva specimens for DNA

Research participants, particularly relatives who are unable to attend for blood sampling or if they preferred not to have a venepuncture were offered the option of providing a saliva specimen for DNA extraction. The saliva collection kit was delivered to participants by post with clear instructions on how to produce the sample. The saliva collection kit was stored at room temperature pending DNA extraction (usually within 3 months). As per the manufacturer's guide and published literature, saliva samples collected in the kit remain stable up to 3 years without DNA degradation[294-296]. Following DNA extraction, the samples were stored at -20° C pending further processing for DNA sequencing.

3.2.3.3 Gastro-intestinal biopsy specimens

All patients with suspected IBD undergo routine endoscopy to establish their diagnosis and further investigative biopsies subsequently depending on the course of the disease. Research biopsies were taken during routine clinical endoscopy and each biopsy specimen comprised a very small sample (ranging between 1-2 mm)[297] of the gut lining. For routine histological analysis, biopsies were taken at various levels in the gastro-intestinal tract such as the oesophagus, stomach, duodenum, ileum, colon and the rectum. Up to two additional biopsies per site were taken for research from the ileum and colon from consented patients. The samples were cryo-frozen at collection using liquid nitrogen and stored at -160 degrees at a designated storage facility.

3.3 IMMUNOLOGICAL METHODS

3.3.1 Peripheral blood mononuclear cells (PBMC) extraction

A peripheral blood mononuclear cell (PBMC) is any peripheral blood cell with a round nucleus such as lymphocytes and monocytes. PBMCs were extracted from whole blood using Ficoll gradient centrifugation method. Standard operating procedure (SOP) for PBMC extraction is provided in section **B.1** in Appendix B. Ficoll is a hydrophilic polysaccharide, which separates the various layers of blood following gradient centrifugation. Ficoll separates whole blood into two main fractions, above and below a density of 1.077g/ml. The fractions separated include a top fraction comprising plasma, followed by a layer of PBMCs and a lower column containing Ficoll with a pellet of red blood cells and polymorphonuclear cells at the bottom of the tube. The PBMC layer often referred to as the buffy coat forms a less dense layer at the upper interface layer of

Ficoll[298]. See figure **3.1.**

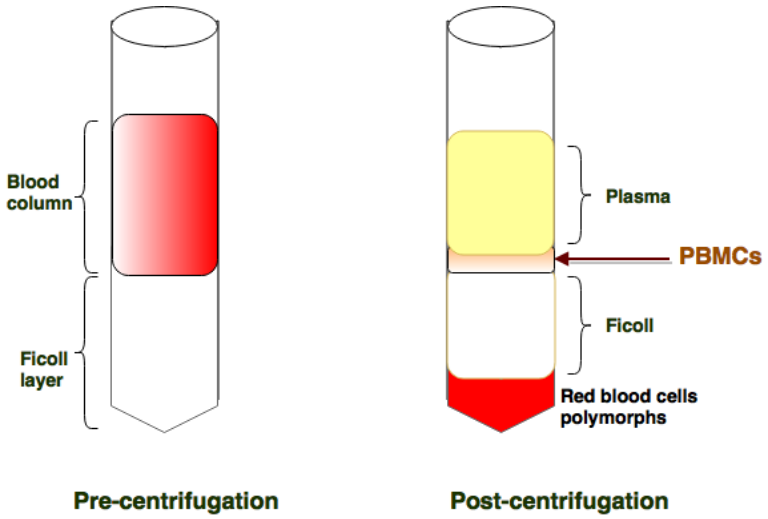


Figure 3.1 Isolation of PBMCs using Ficoll density gradient

Figure on the left shows a universal tube before centrifugation, with an overlying column of heparinized blood on a layer of Ficoll-Paque. Post-centrifugation (right), the red blood cells and the polymorphs are pelleted at the bottom of the tube. The PBMC layer forms an interphase between an underlying layer of Ficoll and an overlying column of plasma.

PBMCs include the T-cells, B-cells, natural killer (NK) cells, monocytes and dendritic cells. The frequency of distribution of these cells varies between specimens; lymphocytes are typically in the range of 60-80%, monocytes 5-10% and dendritic cells 1-2%. Within the population of lymphocytes, approximately 60-80% include CD3+ T-cells, 5-20 % B-cells and 5-20 % NK cells[299,300]. PBMC culture methods have been widely employed in immunology assays. Extracted PBMCs can be stored for further analysis and a robust level of reproducibility can be expected during experiments conducted at various time points using the same pool of extracted cells[301].

3.3.2 Isolation of monocytes from PBMCs

See section **B.2** in appendix B for further details on the method for isolation of monocytes from PBMCs.

3.3.3 Assessing induced immune function

PBMCs were used to assess immune function of innate signalling pathways by stimulating the cells with specific ligands. Cryopreserved PBMCs (frozen PBMCs) were thawed (section **B.1** in appendix B), cells suspended in 15ml Falcon tubes and the respective

ligands added. For each condition, a cell concentration of 3x 10^5 was used. Cell counts and viability assessments were performed using haemocytometer and light microscopy (section **B.1** in appendix B). Details of ligand preparation and dose optimisation are discussed in chapter 4. Activated cell suspensions were incubated at 37°C with 5% CO2 for 24 hours after adding respective ligands to each of the PBMC cell suspensions. The cell suspensions were then pelleted by centrifugation, the supernatants (conditioned media) harvested, aliquoted and stored at -20°C until quantification of cytokines through multiplex assays.

3.3.3.1 Cytokine production patterns using multiplex assays (Luminex® assay)

In this thesis, cytokines from cellular supernatants following activation of cell cultures with specific ligands were quantified using multiplex assays (Luminex®). Multiplex assays are widely used in functional genomics and high-throughput screening settings to measure multiple analytes in a single cycle. These are tools of choice for various immunological assays as they maximize efficiency by simultaneously profiling several proteins within small samples of biological fluids. Biological response modifiers such as cytokines, chemokines, immunoglobulins and other inflammatory mediators such as bioactive complement fragments are popular targets for quantification through multiplex assays[302-304].

Luminex is a bead based multiplexing assay, where spectrally encoded antibody-conjugated beads are used as a solid phase. Each bead has a unique fluorescence intensity so that they can be mixed and run simultaneously in a single tube to significantly reduce sample requirements and the time to results, when compared to traditional ELISA and Western blot techniques[303]. The sample is added to a mixture of colour-coded beads, pre-coated with analyte-specific capture antibodies. Each different group of beads is labelled with a discrete level of fluorescent dye so that it can be distinguished by its median fluorescence intensity (MFI) upon flow cytometric analysis. Beads within each group are covalently coupled with antibodies that specifically capture a particular type of molecule present within the biological fluids including sera, plasma, cell culture supernatants or cell lysates. The antibody-coupled capture beads serve as the solid capture phase for the cytometric bead assay. The immobilised high affinity antibodies function to specifically capture and localise proteins of interest that may be present in the biological fluids tested. Biotinylated detection antibodies specific to the proteins of interest are then added, which form an antigen-antibody sandwich, followed by addition of phycoerthrin- conjugated streptavidin, which binds to the biotinylated antibody. Biotinylation is a process of

covalently attaching biotin molecules to a protein. The process occurs rapidly and does not affect the natural function of the protein molecule. Biotin binds to streptavidin with high affinity, a property frequently exploited in biotechnology in order to isolate biotinylated proteins of interest. Phycoerythrin (PE)-conjugated antibodies are most frequently used since the emission wavelength of PE is easily distinguishable from the emission wavelengths of the capture beads[305,306]. The assay is performed in a 96-well plate format and the polystyrene beads are analysed on a dual laser flow-based detection instrument such as Luminex® 100™ or 200™ instrument, which monitors spectral properties of the capture beads with respect to the nature and concentration of the biological markers of interest. One laser classifies the bead and determines the protein of interest, whilst the second laser determines the magnitude of the PE-derived signal, which is in direct proportion to the amount of analyte present[307]. See figure **3.2**.

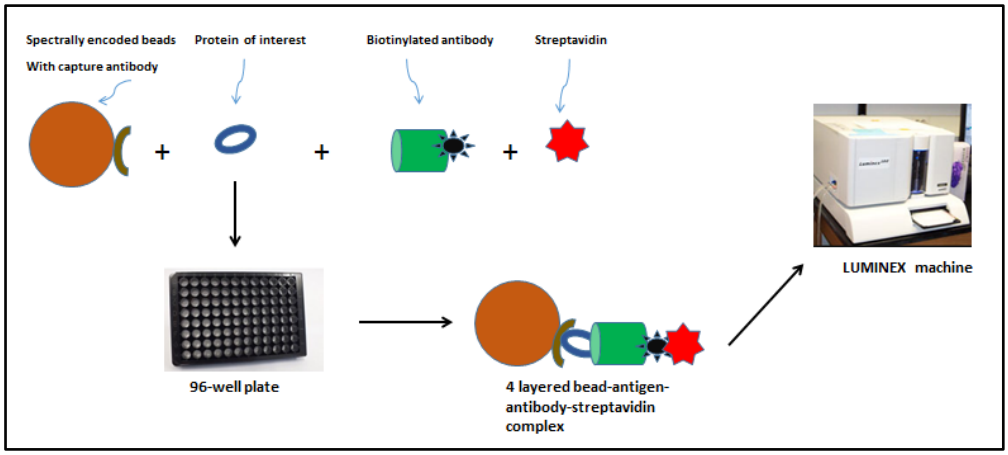


Figure 3.2 Basic principles of Luminex® assay

The protein (analyte) of interest in a given sample attaches itself to the antibody beads. The biotinylated antibody upon addition to the antigen-antibody complex will bind to the analyte. After addition of streptavidin, the biotin molecule will attach itself to the streptavidin molecule with high affinity. The multi-layered sandwich complex is then run through the analyzer for quantification.

A mixture of cytokine protein standards with known concentrations are used in serial dilutions to generate a standard curve for each analyte tested. Standard curves are used to determine the concentration of proteins in an assay. They are obtained by relating a measured quantity to the concentration of the analyte of interest in standards samples of known concentration. The concept hinges on the understanding that the measurable quantity of proteins under analysis can be compared to known concentrations of the "standards". These standards provide a reference to determine unknown concentrations. The concentrations of standards chosen need to span the range of concentrations expected to be in the "unknown" sample concentration. See figure **3.3**. The assay measurement is

graphed on y-axis and standard concentrations on x-axis. Data are analysed by fitting a line on the curve[304]. Detailed steps of the Luminex assay are presented in section **B.3** in appendix B.

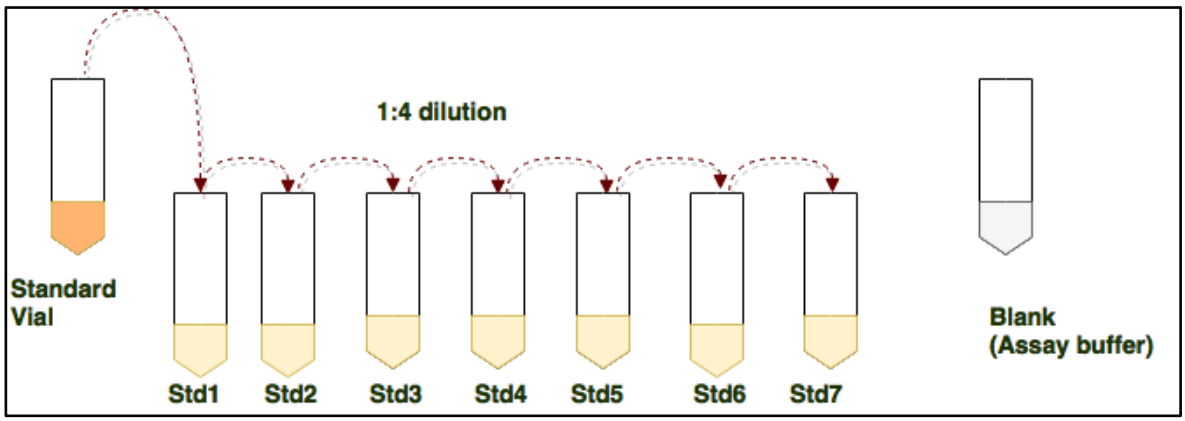


Figure 3.3 Preparation of antigen standards

Seven tubes are labelled from standard 1-standard 7. Standard 1 has the highest concentration of the analyte and the subsequent tubes have concentrations in decreasing order following 1:4 dilution. For example, if standard 1 has 200µl of the antigen standard, 50µl from standard 1 are added to standard 2. Each of the tubes from standard 2-standard 7 contain 150µl of assay buffer solution. Adding 50µl of solution from each preceding tube in succession ensures 1:4 dilution. The assay measurements from the known concentrations are then used to generate a standard curve.

3.4 WHOLE EXOME SEQUENCING (WES)

As part of the 'Genetics of paediatric IBD' study, patient samples from the Southampton paediatric IBD cohort undergo sequence analysis of all genes coding for protein. The Genomics Group has already established and continue to advance the bioinformatic pipeline necessary for processing, quality assessment and analysis of next generation sequence data. The complete NGS analysis process is complex and includes multiple steps. It involves handling vast amounts of heterogeneous data, for which several *in silico* tools have been created in order to support the specific steps of the analysis workflow. The basic steps are summarised below.

3.4.1 DNA extraction

DNA from blood or saliva specimens was isolated using the salting out method[308] and was performed by an experienced technician from the DNA laboratory, based at the Duthie building. The process involves purification of DNA from a sample using a combination of physical and chemical methods. The initial step includes the disruption of the cell and

nuclear membranes to expose the DNA. Lipids are broken down with detergents and surfactants, whilst the proteins are broken down using proteases. The solution is treated with concentrated salt which causes debris such as proteins, lipids and RNA to clump together. Centrifugation separates the clumped cellular debris from the supernatant containing the DNA. The supernatant is finally treated with ethanol to precipitate the genomic DNA[308,309]. See figure **3.4.** Details of the steps involved are included in section **B.4** in appendix B.

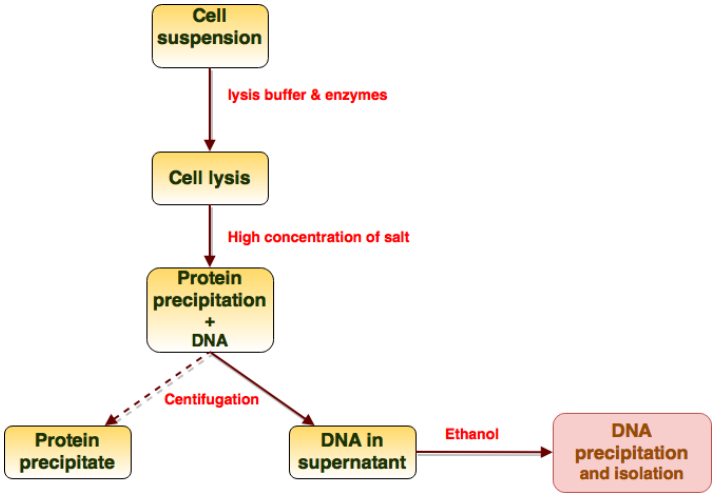


Figure 3.4 Basic steps for DNA extraction using salting out method

The broad principles include cellular lysis to expose the DNA using lysis buffers, enzymatic digestion of proteins, precipitation of proteins using concentrated salt which purifies DNA into the supernatant and finally the isolation of DNA using ethanol.

3.4.2 Exome library preparation

Genomic DNA extraction is followed by preparation of an exome enrichment library. In this project, DNA was quantified and plated for shipping to external sequencing facilities. DNA is sheared into random fragments mechanically by ultrasonication methods. Fragment ends are subsequently blunted and ligated with oligonucleotide adaptors. The DNA fragments are then hybridised to targeted baits of an exome capture kit such as Agilent SureSelect All Exon (Version 4-V4, V5 or V6). These baits are designed to be complimentary to protein coding regions and the pull down of baits provides a targeted enrichment of the exonic regions of the genome. Non-targeted sequences are washed away, hybridised fragments are eluted and the resultant enriched library is amplified[310,311].

3.4.3 Exome sequencing

Following exon enrichment, the exome library preparation is subject to a high-throughput, massively parallel sequencing to produce millions of short reads. Although several sequencing platforms are currently available, samples for this study were sequenced with Illumina's HiSeq range of sequencers (HiSeq 2000, HiSeq X10), which use sequencing by synthesis approach[310,312,313].

3.4.4 Alignment

After reads have been processed to meet a certain quality standard defined as per GATK best practice[314,315], the first step is the alignment of the raw reads against the human reference genome (hg19). The process involves alignment and mapping of the reads to the best possible match on the reference genome or conversely aligning onto positions on the reference genome with the least number of mismatches. The sequence alignment thus generated is stored in a SAM (sequence alignment/map) or BAM (binary alignment/map) file. These are standard file formats for storing sequence alignment data [316]. Various alignment algorithms or software programs are available to perform sequencing read alignment of the sample genome. Novoalign, a commercially available aligner for single-ended and paired-end reads, was a commonly used alignment tool in the version 3 Southampton pipeline[317]. In the later versions (5 & 6) of the pipeline, Burrows-Wheeler Aligner (BWA) was used. BWA applies the Burrows-Wheeler transformation (BWT) and efficiently aligns short sequencing reads against the human genome[318]. See figure **3.5**.

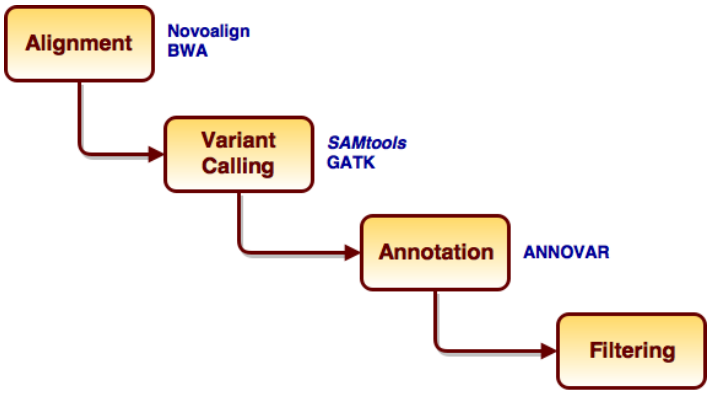


Figure 3.5 Basic workflow in computational analysis of raw sequenced data

*The initial step involves alignment of the raw reads against the reference genome, followed by variant calling and annotation, and finally applying filters. The software tools used at each step are indicated in dark blue. (**Abbreviations**: BWA- Burrows-Wheeler Aligner, GATK- Genome Analysis Toolkit)*

3.4.5 Variant calls

The next step is variant calling within the aligned reads, a process comparing the aligned sequences with the reference human genome. Positions on the aligned reads which deviate from the reference genome are referred to as 'variant calls'. These are recorded in a variant call format (VCF), which is a standardized generic format for storing sequence variants[319]. Many bioinformatic pipelines have been developed to call variants from NGS data[320]. A commonly utilized SNP calling software package is GATK (Genome Analysis Toolkit) in the current version of the Southampton bioinformatic pipeline, a structured programming framework using the functional programming philosophy of MapReduce[315]. Another SNP calling software package employed by the Genomics Group is *SAMtools*, which implements various utilities for post-processing alignments in the SAM format[316]. In this thesis, GATK haplotype caller was applied for variant calling[315].

3.4.6 Annotation

After alignment and variant calling, whole exome sequencing identifies approximately 25,000 variants in the coding regions of the genome[321]. The next step therefore is to process and condense the vast amount of information in order to determine which of these variants are likely to be of significance with a potential to explain a biological phenomenon such as disease causality. Annotation involves bioinformatic integration of information about variants that enables prioritisation of potential disease-causing mutations. This means employing software packages, which effectively perform the task of interrogating previously known information about the relevant variants through a predictive and a probabilistic framework[316,322]. The Southampton pipeline commonly utilises the ANNOVAR software tool to annotate genetic variants. The annotation system allows the user to employ several strategies, which may be based on the allele frequency in a given population, gene-based, region-based or other user preferred functionalities[323].

3.4.6.1 Allele frequencies

A key annotation type is the allele frequency, which is the relative frequency of a gene variant at a particular locus in a population and is expressed as a fraction or percentage. These data are derived from large-scale genotyping consortia such as the 1000 Genomes project (1 KG)[324], the HapMap project[89], Single Nucleotide Polymorphism Database (dbSNP)[325], the exome Variant Server (EVS) of European Americans of the NHL1-ESP project with 6500 exomes and also the allele frequency within the Southampton in-house

database of reference exomes. The Southampton database is a local repository comprising independent DNA samples from local cohorts sequenced as part of other projects. When a rare Mendelian disease with a severe phenotype is being investigated, a high allele frequency for a particular variant will support its exclusion as a causal variant. As allele frequencies vary between different ethnic groups, it is important to utilise ethnically matched data-sources for variant annotation.

3.4.6.2 Gene and variant based annotation

Gene- based annotation refers to the manner in which a variant affects known genes by inferring several types of information; chromosome location, name of the gene, whether the variant is exonic, intronic, splicing or intergenic and the functional impact it has on protein coding. A change of amino acid within a protein can have vastly differing outcomes depending on the nature of the change. Substitution by certain amino acids within the protein can result in critical distortion of protein structure through alteration of its hydrophobic properties, electrostatic interactions and interactions with other molecular complexes[326].

3.4.6.3 Deleteriousness metrics

There is a plethora of computational tools that aim to infer or predict whether observed SNVs are deleterious and impact protein function. There is no gold standard method for predicting variant deleteriousness and each of the prediction scores have their own merits and disadvantages[327,328]. Studies comparing the predictive abilities of the various prediction tools have yielded inconsistent results[327-333]. Deleteriousness is predicted based on various sources of information available on the variant such as sequence homology, evolutionary conservation and protein structure[327,330]. Table **3.1** is a synopsis of the commonly used deleteriousness prediction methods. Function prediction scores refer to scores that predict the likelihood of a given non-synonymous SNV causing deleterious functional change in the protein. Conservation scores measure the evolutionary constraint or how conserved a given nucleotide site remains across multiple species. The underpinning principle is that amino acid residues critical for protein function should be conserved among species and hence mutations in these highly ‘conserved’ or ‘naturally protected’ sites are more likely to result in a deleterious effect. Composite scores refer to measures that combine information from multiple component scores[327].

Table 3.1 Deleteriousness prediction tools

Name of the algorithm	Prediction score categories	Range	Deleterious threshold	Description
SIFT	Function prediction	0-1 0= deleterious 1= tolerated	<0.05*,†	Sorting Tolerant From Intolerant (SIFT)- Predicts whether a substitution at a particular position in a protein sequence is expected to be deleterious for protein function[329].
Polyphen-2_HVAR	Function prediction	0-1 0= tolerated 1= deleterious	>0.5*,†	Assesses the impact of an amino acid substitution on the protein structure and function[334].
LRT	Function prediction	0-1 0= tolerated 1= deleterious	Categorical prediction (deleterious, neutral or unknown)	Employs a likelihood ratio test to assess variant deleteriousness based on a comparative genomics data set of 32 vertebrate species[335].
Mutation Assessor	Function prediction	(Score-Min)/(Max-Min)	>0.65*	Assesses sequence homology of protein families and sub-families within and between species using evolutionary conservation patterns[336].
MutationTaster	Function prediction score	0-1 0= tolerated. 1= deleterious	>0.5*,†	Assesses the impact of the disease-causing potential of a sequence variant by a naive Bayes classifier using multiple resources such as evolutionary conservation, splice-site changes, loss of protein features and changes that might affect mRNA level[331].
FATHMM	Function prediction	1- (Score-Min)/(Max-Min)	≥45*,†	Uses multiple sequence alignment (MSA) built from several species to identify conserved amino acid residues[337].
PROVEAN	Function prediction	-14 to 14	<-2.5	Protein Variation Effect Analyser (PROVEAN)- Measures the change in sequence similarity of a query sequence to a protein sequence homolog before and after the introduction of an amino acid variation to the query sequence[338].
GERP++RS	Conservation	-12.0 to +6.17	>4.4*,†	Genomic Evolutionary Rate Profiling (GERP)- Estimates DNA sequence conservation by assigning conservation scores to specific nucleotides[339].
Phylop	Conservation	-20 to +10 (dbNSFP)	>1.6	Estimates the evolutional conservation at each variant from multiple alignments of placental mammal genomes to the human genome based on a phylogenetic hidden Markov model[340].
SiPhy	Conservation	0 to 37.97, (dbNSFP)	>12.17*	Based on inferred nucleotide substitution patterns per site[341].
CADD	Composite	0-35+	>15*,†	Combined Annotation Dependent Depletion (CADD)- Objectively integrates many diverse annotations into a single, quantitative score[332].
MCAP	Composite	0-1	>0.025	Combines SIFT, polyphen -2 and CADD[333].
DANN	Composite	0-1		Deleterious annotation of genetic variants using neural networks (DANN)- Uses the same feature set and training data as CADD, but uses a different 'non-linear' machine learning approach[342].

Each algorithm uses different sources of information and has its own merits and pitfalls. There is no gold standard (Dong et al, Human Mol Gen 2015[327], † Mahmood et a[328])*

3.4.7 Prioritisation and filtering

Given the magnitude of NGS data, it is important to ascertain which of the variants are functionally important against a background of neutral non-pathogenic variants. Filtering

reduces the number of variants from a multitude of thousands to a much smaller subset by excluding variants, which are convincingly less likely to be of functional significance and interpreting the remaining variants within a biological context. The commonly applied filtering strategies include variant function, population frequency and a combination of *in silico* deleteriousness metrics.

On an average, approximately 9000-11,000 of the coding variants identified through WES per individual are non-synonymous and a slightly higher number are synonymous variants[343]. In this thesis, synonymous variants were excluded on the basis that these are less likely to affect protein function and therefore less likely to be pathogenic. Insertions/deletions, commonly known as Indels, are the second most frequent type of human variation[344]. Frameshift Indels (FS-Indels) either alter the complete C-terminal region of the protein or result in premature truncation of the protein. On the other hand non-FS-Indels (NFS-Indels) involve multiples of 3 nucleotides leading to insertion/deletion of one or more amino acids[345]. In our study, NFS-Indels were excluded on the basis that these are less likely to be pathogenic compared to FS-Indels.

Most exome sequencing studies investigating rare diseases, rely on a hard-filtering approach, in which the causal mutation is assumed to be rare with a minor allele frequency (MAF) of <1% found in public databases (for example 1000 Genomes Project, ExAC, dbSNPs or in-house control datasets)[330,346]. In our study, we applied a MAF <5% rather than a more stringent threshold of <1% in order to avoid the risk of discarding variants potentially contributing to more common disease causality. Applying a frequency of over 5% would have resulted in an enormous number of common variants per individual, which are less likely to impact on disease causality and impractical to analyse in timely and effective manner[347].

Finally, filtering variants based on a combination of prediction tools often provides complementary information on the functional mechanisms and therefore enhances the accuracy of prediction of protein function[348]. See table 3.1. Details of the specific filtering strategies and deleteriousness metrics applied to my core project and the other projects will be discussed in the respective chapters.

Chapter 4 DEVELOPMENT AND OPTIMISATION OF ASSAY METHODS

This chapter describes the molecular basis of the immunological assay, the process of optimisation in healthy controls and development of a standard operating procedure (SOP) using peripheral blood mononuclear cells (PBMCs) for performing assays on patient specimens. The chapter also outlines the experiments conducted during the developmental phase of my project using monocytes both in controls and paediatric patients. These are included here as part of my learning curve, leading on to the setup of the SOP. All experiments described in this chapter were conducted wholly by myself under the supervision of Dr Yifang Gao and Prof Anthony Williams. Significant guidance was also provided by my other supervisors Professors Sarah Ennis and Mark Beattie into the concept and design of the assay.

4.1 INTRODUCTION

Over the last two decades, the findings of several genetic and immunological studies have highlighted multi-factorial mechanisms involved in the pathogenesis of IBD. Collectively, these studies support a concept of dysregulated immune checkpoints that control microbial-host interactions, including innate and adaptive immunity, as well as epithelial and mesenchymal immune responses[4,60]. Although the identification of candidate genes and potentially causal variants through genetic studies provides a robust platform to gain insight into the genetic architecture of the disease, mechanistic interpretation of the exact impact of the associated variants is essential. Therefore, in the context of immune-mediated diseases such as IBD, it is important to integrate genomics with immunological assessment, for a better understanding of the implicated mechanisms and development of bespoke therapeutics[3,349].

A number of immunological studies in IBD focus on the role of cytokines in the initiation of inflammation, propagation of inflammatory changes and also on the resolution of the inflammatory processes through anti-inflammatory mechanisms[4]. Cytokines are structurally diverse proteins that function as intercellular messenger molecules, with upstream and/or downstream regulatory roles at the various checkpoints of the immune system[350]. Hypothetically, the net effect of cytokine milieu generated through a cross fire

between the pro-inflammatory and anti-inflammatory processes can be viewed as a key operating mechanism in the pathogenesis of IBD.

Inflammatory bowel disease is a cytokine-driven disease, resulting in chronic intestinal and extra-intestinal inflammation[4]. The role of cytokines is also highlighted by the fact that blockade of tumour necrosis factor- alpha (TNF-α) is now commonly used as a standard therapy for IBD, both in children and adults[37,351]. Cytokines drive intestinal symptoms such as diarrhoea and rectal bleeding, and also regulate the extra-intestinal manifestations including arthritis, uveitis and other systemic effects of the disease. For example, interleukin-6 (IL-6) induces the release of acute phase reactants by the liver, whereas TNF-α has been implicated in the development of cachexia and arthritis[352]. Furthermore, a number of cytokines have been implicated specifically in the pathogenesis of IBD-associated complications of the bowel including stricturing disease, abscess and fistula formation, and colitis-associated bowel cancer[353,354]. Studies have identified multiple molecular components of cytokines and their receptors as potential new targets for directing treatment of IBD. These include several pro-inflammatory cytokines such as IL-6, IL-12, IL-23, as well as anti-inflammatory cytokines such as TGF-β and IL-10[3,353]. Targeting new cytokines and their immune checkpoints, optimised drug-delivery systems and personalised therapy approaches may lead to better clinical outcomes in the longer term.

A wide range of cytokines can be effectively quantified simultaneously by conducting multiplex assays. Multiplex assays can be used to assess cytokine profile in the cellular supernatant of primary cell cultures such as peripheral blood mononuclear cells (PBMCs), activated by various stimuli and the quantified effector response can be a measure of the function of the corresponding pathways. Although functional tests are frequently used in laboratory practice, the standardisation of functional assays has been very challenging given the intra/inter-individual variability of induced immune responses.

One of the key aspects of my project is to assess the molecular integrity of biological pathways of known importance in IBD including the NOD2 signalling pathway and toll-like receptor signalling, which directly or indirectly connect downstream to critical signalling cascades such as the NF-κB, mitogen-activated protein kinases (MAPK) and NLRP3-inflammasome activation. My work focuses on the integrity of these key innate signalling pathways in the context of genetic mutations, with altered cytokine production patterns as indirect signatures of functionality of these pathways.

4.2 MOLECULAR BACKGROUND OF THE ASSAY

In this assay, the four main stimulating ligands employed at the outset were synthetic tri-acylated lipoprotein (Pam3CSK4), lipopolysaccharide (LPS), L-Ala-γ-D-Glu-mDAP (Tri-DAP) and muramyl dipeptide (MDP) which are specific agonists for the TLR2, TLR4, NOD1 and NOD2 receptors respectively[355-357]. The assay hinges on the functional integrity of NF-κB, MAPK pathways and NLRP3-inflammasome activation. These are critical hubs and cross-talking portals for other innate signalling pathways involved in IBD such as the NOD and toll-like receptor signalling. Several inflammatory pathways implicated in IBD are directly involved in the upstream and downstream signaling of the NF-κB and the MAPK pathways[358,359]. Following stimulation by the respective ligands, each of the receptors undergo conformational changes triggering interactions with several adaptor proteins. Assembly of a combination of specific adaptor proteins results in the formation of multi-molecular activating complexes, which trigger further downstream specific signaling pathways. The activation of the signaling pathways results in the production of specific effector cytokines which can be quantified through multiplex assays, giving an indication of functionality of proteins involved in those specific pathways[359]. **The effector cytokines to be measured in this assay include IL-10, IL-6, IL-1β, IL-8 and TNF-α**, which are produced following transcriptional upregulation of their genes by activated NF-κB, inflammasome and the MAPK pathways. The key steps involved in the processing of the samples and the basic model of the assay are illustrated in figure **4.1** and **4.2** respectively.

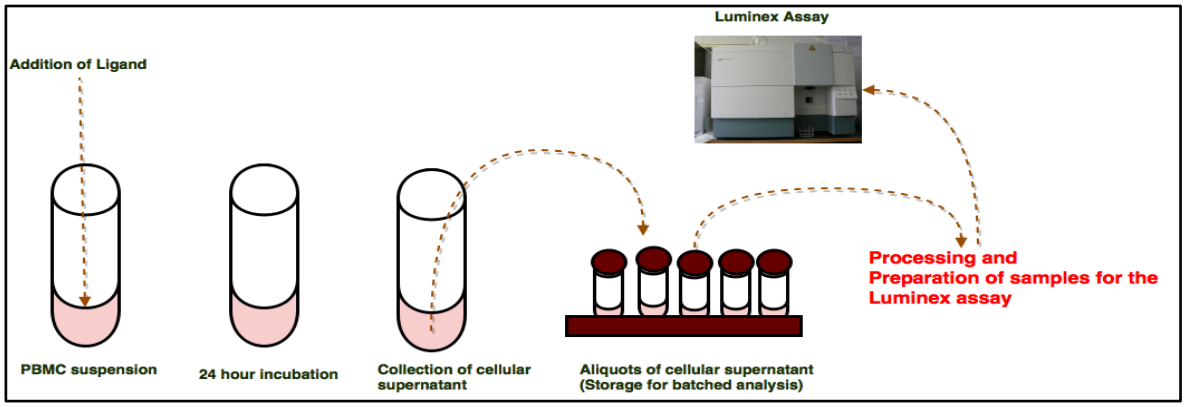


Figure 4.1 Basic steps of the assay

Specific ligands are added to the PBMCs suspended in culture media and allowed to incubate for 24 hours. The added ligands stimulate their respective cellular receptors, resulting in the activation of signaling pathways followed by production of cytokines. The cellular supernatant containing the cytokines is then harvested, aliquoted and stored at -20° C pending batched analysis. The samples are further processed with the assay-specific antibodies and reagents. The prepared

samples are then run through the Luminex® machine for quantification of the induced cytokines.

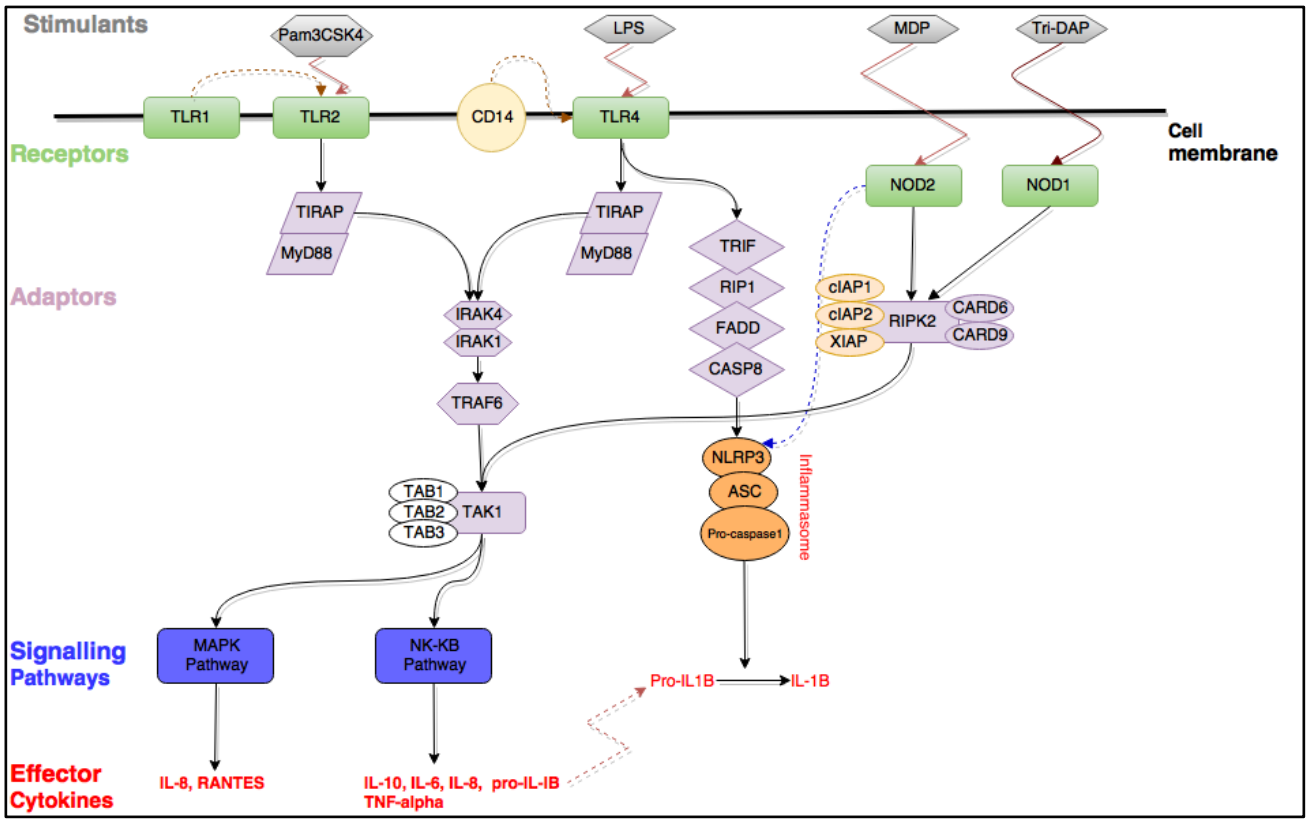


Figure 4.2 Mechanistic model of the assay

*The stimulating ligands include PAM3CSK4 binding to TLR2, LPS to TLR4, Tri-DAP to NOD1 and MDP to NOD2. TLR2 requires co-stimulation with TLR1, whilst CD14 acts as a co-receptor for TLR4. TLRs are cell surface receptors whilst the NOD receptors are intra-cellular. Binding of the ligands to the respective receptors results in the recruitment of adaptor proteins, which in turn activate the signalling pathways. At the level of adaptors, there is close interaction between various molecular complexes and cross-networking along the pathways. The figure highlights the plurality of the adaptor proteins in activating the signalling pathways and the role of the IAPs (XIAP, CIAP1 and CIAP2) in k-63 linked ubiquitination of the adaptor protein complexes, thus providing a scaffold to recruit other kinases. The ubiquitination step controls cellular signalling pathways by providing a scaffold for recruiting other effector kinases downstream. Further downstream, activation of the NF-κB, MAPK and the inflammasome result in the transcription of effector cytokines. (**abbreviations:** ASC, Adaptor protein- apoptosis associated speck-like protein containing CARD; CARD, caspase activation and recruitment domain; CD14, cluster of differentiation 14; cIAP, cellular inhibitors of apoptosis; FADD, Fas associated protein with death domain; IRAK, interleukin-1 receptor associated kinases; LPS, lipopolysaccharide; MAPK, mitogen activated protein kinases; Myd88, myeloid differentiation primary response protein 88; MDP, muramyl dipeptide; NF-κB, nuclear factor--κB; NLRP3, NOD-like receptor family pyrin domain containing 3; NOD, nucleotide-binding oligomerization domain; PAM3CSK4, synthetic triacylated lipoprotein; RIP2, receptor interacting response protein 2; TAB1, TGF-β*

activated kinase 1 binding protein; TAK1, TGF-β-activated kinase 1; TIRAP, TIR domain containing adaptor protein; TRAF6, TNF receptor associated factor 6; Tri-DAP, L-Ala-γ D-Glu mDAP; TRIF, TIR domain containing adaptor-inducing interferon-β; XIAP, X-linked inhibitor of apoptosis).

Table **4.1** provides an outline of the possible pathway defects which can be detected through the interpretation of results of the immunological assay. Further details on the specific pathways are provided in the subsequent sections.

Table 4.1 Identification of specific pathway defects through the immunological assay

Receptor	Stimulant	Cytokines					Interpretation/pathway defect
		IL-10	IL-6	IL-1β	TNF-α	IL-8	
TLR4	LPS	N	N	N	N	N	Normal immune responses
TLR1-2	Pam3CSK4	N	N	N	N	N	
NOD2	MDP	N	N	N	N	N	
NOD1	Tri-DAP	N	N	N	N	N	
TLR4	LPS	↑ or ↓	↑ or ↓	↑ or ↓	↑ or ↓	↑ or ↓	Defective TLR4-mediated signalling
TLR1-2	Pam3CSK4	N	N	N	N	N	
NOD2	MDP	N	N	N	N	N	
NOD1	Tri-DAP	N	N	N	N	N	
TLR4	LPS	N	N	N	N	N	Defective TLR1-2-mediated signalling
TLR1-2	Pam3CSK4	↑ or ↓	↑ or ↓	↑ or ↓	↑ or ↓	↑ or ↓	
NOD2	MDP	N	N	N	N	N	
NOD1	Tri-DAP	N	N	N	N	N	
TLR4	LPS	N	N	N	N	N	Defective NOD2-mediated signalling
TLR1-2	Pam3CSK4	N	N	N	N	N	
NOD2	MDP	↑ or ↓	↑ or ↓	↑ or ↓	↑ or ↓	↑ or ↓	
NOD1	Tri-DAP	N	N	N	N	N	
TLR4	LPS	N	N	N	N	N	Defective NOD1-mediated signalling
TLR1-2	Pam3CSK4	N	N	N	N	N	
NOD2	MDP	N	N	N	N	N	
NOD1	Tri-DAP	↑ or ↓	↑ or ↓	↑ or ↓	↑ or ↓	↑ or ↓	
TLR4	LPS	N	N	↑ or ↓	N	N	Defective Inflammasome activation
TLR1-2	Pam3CSK4	N	N	↑ or ↓	N	N	
NOD2	MDP	N	N	↑ or ↓	N	N	
NOD1	Tri-DAP	N	N	↑ or ↓	N	N	
TLR4	LPS	N or ↓	↑	↑	↑	↑	Defective IL-10 mediated signalling
TLR1-2	Pam3CSK4	N or ↓	↑	↑	↑	↑	
NOD2	MDP	N or ↓	↑	↑	↑	↑	
NOD1	Tri-DAP	N or ↓	↑	↑	↑	↑	
TLR4	LPS	↑ or ↓	↑ or ↓	↑ or ↓	↑ or ↓	↑ or ↓	Defective NF-Kappa B signalling
TLR1-2	Pam3CSK4	↑ or ↓	↑ or ↓	↑ or ↓	↑ or ↓	↑ or ↓	
NOD2	MDP	↑ or ↓	↑ or ↓	↑ or ↓	↑ or ↓	↑ or ↓	
NOD1	Tri-DAP	↑ or ↓	↑ or ↓	↑ or ↓	↑ or ↓	↑ or ↓	
TLR4	LPS	N	N	N	↑ or ↓	N	Defective TNF-alpha signalling
TLR1-2	Pam3CSK4	N	N	N	↑ or ↓	N	
NOD2	MDP	N	N	N	↑ or ↓	N	
NOD1	Tri-DAP	N	N	N	↑ or ↓	N	

Table 4.1 provides a synopsis of the innate immune pathway defects which can be identified through the immunological assay. The four stimulants used included LPS, Pam3CSK4, MDP and Tri-DAP, which activated the TLR4, TLR1-2, NOD2 and NOD1 receptors respectively. The table presents a simplified interpretation of the assay. It may be

noted that due to the biological variations in immune responses, various combination of normal, hyper-inflammatory (↑) or hypo-inflammatory (↓) cytokine responses as compared to reference ranges may be possible in a given individual.

4.2.1 The TLR pathway

Cell surface TLRs, including TLR2 and TLR4 recognize pathogen associated molecular patterns (PAMPs) such as lipids, lipoproteins and other microbial products. Each TLR is composed of an extra-cellular domain with leucine-rich repeats (LRR), a transmembrane domain and a cytoplasmic Toll/IL-1 receptor (TIR) domain that initiates downstream signalling. The extra-cellular domain of TLR2 along with its co-receptor (TLR1 or TLR6) recognize a wide variety of PAMPs including lipoproteins, peptidoglycans, lipotechoic acids, zymosan and others[360,361]. Pam3CSK4 is a synthetic lipoprotein that mimics the acylated amino terminus of bacterial lipoproteins and is specifically recognized by TLR2 in conjunction with TLR1[356,362]. TLR4 on the other hand, specifically recognizes bacterial lipopolysaccharide (LPS). A transmembrane protein CD14 acts as a co-receptor for TLR4. CD14 binds directly to LPS and is known to chaperone LPS molecules to the TLR4 complex, thereby regulating LPS-induced endocytosis of TLR4[363].

Following interaction with the specific ligand, TLRs recruit TIR domain-containing adaptor protein (TIRAP) to interact with downstream adaptors such as MyD88 and TIR-domain-containing adaptor-inducing interferon-β (TRIF), which activate inflammatory cascades such as the NF-κB, inflammasome and MAPK pathways. Following signal transduction from TLR via TIRAP, MyD88 forms a complex known as Myddosome with interleukin-1 associated receptor kinase 4 (IRAK4), a serine-threonine activating kinase. During the formation of the complex, IRAK4 activates IRAK1, which gets auto-phosphorylated at various sites and is then released from MyD88. IRAK1 links with TNF receptor-associated factor 6 (TRAF6), which along with ubiquitin-conjugation enzymes UBC13 and UEV1A, promotes K63-linked poly-ubiquitination of both TRAF6 itself and TGF-β-activated kinase 1 (TAK1) protein kinase complex. TAK1 forms a complex with regulatory subunits TGF-β-activated kinase 1 binding protein 1 (TAB-1), TAB-2 and TAB-3 driving forth TAK 1 activation. TAK 1 activation finally triggers transduction of the two critical inflammatory pathways, NF-κB and the MAPK pathways, resulting in the production of key effector cytokines[361,364].

In this assay, the TLR1-2 and TLR4 receptors were stimulated using Pam3CSK4 and LPS respectively. The cytokines produced as a result of the activation of downstream signalling cascades were quantified using the Luminex assay. The optimal production of the effector cytokines including IL-10, IL-6, IL-1β, IL-8 and TNF-α following stimulation of the TLRs was used as an indirect measure to assess the functional integrity of the TLR-mediated immune pathway signalling. Figure **4.2.**

Synthetic tri-acylated lipoprotein (Pam3CSK4)- TLR1-2 receptor stimulant

Pam3CSK4 is a synthetic triacylated lipoprotein that mimics the acylated amino terminus of bacterial lipoproteins. It is a potent activator of the NF-κB pathway via the cell surface toll-like receptors TLR1 and TLR2 receptors. Bacterial lipoproteins are a family of pro-inflammatory proteins found in the cell wall of both gram-positive and gram-negative bacteria. The active moiety resides in the acylated amino terminus. Pam3CSK4 is a synthetic tri-palmitoylated lipopeptide that mimics the acylated amino terminus of the bacterial lipopeptides. Recognition of Pam3CSK4 is mediated via TLR2 in conjunction with TLR1, leading to activation of the NF-κB pathway with resultant pro-inflammatory cytokine production[356,362].

Lipopolysaccharide (LPS)- TLR4 receptor stimulant

LPS is a key component of the cell wall of gram-negative bacteria and is recognised by toll-like receptor 4 (TLR4) expressed by macrophages, dendritic cells, monocytes and other cells. The molecule consists of three structural units: O-polysaccharide chain made up of repeating oligosaccharide units, the core oligosaccharide and Lipid A. Lipid A is responsible for the endotoxic activity of the entire LPS molecule. LPS contributes to the structural integrity of the bacterial cell wall and protects the bacteria against the action of bile salts and lipophilic antibiotics. LPS is a heat stable endotoxin and has long been recognised as a key factor in septic shock in humans[357].

4.2.2 The NOD- signalling pathway

Unlike TLR2 and TLR4 which are cell-surface receptors, the NOD2 receptors are intra-cellular pathogen recognition receptors (PRR) consisting of a caspase activation and recruitment domain (CARD), a nucleotide-binding oligomerization domain (NOD) and a c-terminal domain consisting of leucine-rich repeats (LRRs). NOD2 senses specific conserved fragments such as muramyl dipeptide (MDP), a peptidoglycan (PGN) motif

found in the cell wall of several gram-positive and gram-negative bacteria[365]. The mechanisms by which the bacterial PGN enters the cells is unclear although multiple routes of entry have been reported. These include phagocytosis of bacteria with subsequent bacterial degradation, transport across the host cell membrane via channels/pore-forming molecules/bacterial secretion systems and endocytosis[174]. Studies have highlighted the role of endosomes in NOD2 signalling, including the two peptide transporters SLC15A3 and SLC15A4, which transport MDP across the endosome to the NOD2 receptor[366,367]. Within the cytosol, NOD2 resides in an auto-inhibited monomeric state. Upon ligand recognition and binding at the LRR domain, the NOD domain undergoes conformational changes allowing NOD2 protein to self-oligomerize and recruit receptor-interacting serine-threonine-protein kinase 2 (RIPK2) through CARD-CARD interactions. RIPK2 then mediates the activation of TAK1, which is an essential step for downstream activation of inflammatory cascades via the NF-κB and the MAPK pathways[174].

The NOD1 receptor is an intra-cellular PRR, structurally similar to NOD2 and is activated by specific dipeptides from gram-negative bacteria. For the assay, a synthetic analog of the cell wall dipeptides (Tri-DAP) was used. Following NOD1 stimulation, conformational changes in the receptor result in downstream signaling via its specific adaptors, finally activating the transcriptional pathways[53].

In this assay, the NOD1 and NOD2 receptors were stimulated using Tri-DAP and MDP respectively as specific receptor ligands. The optimal production of the effector cytokines including IL-10, IL-6, IL-1β, IL-8 and TNF-α following the stimulation of these receptors was used as an indirect measure of the functional integrity of the NOD1 and NOD2-mediated immune signalling pathways. Figure **4.2.**

L-Ala-γ-D-Glu-mDAP (Tri-DAP)- NOD1 receptor stimulant

Tri-DAP is NOD1-receptor specific ligand comprising the iE-DAP (γ-D-Glu-mDAP) dipeptide and a L-Ala residue. Tri-DAP is found in the peptidoglycan of most common gram-negative bacteria and certain gram-positive bacteria. NOD1 acts as an intracellular pathogen recognition receptor (PRR) for a subset of bacteria through recognition of iE-DAP. Interaction with NOD1 receptor triggers a cascade of events leading to activation of the NF-KB pathway and secretion of IL-6 and TNF-α[368,369].

Muramyl dipeptide (MDP)- NOD2 receptor stimulant

MDP is a synthetic immuno-reactive peptide found in the peptidoglycan layer of the bacterial cell wall. Peptidoglycan is found in the bacterial cell wall as a thin layer in gram-negative and as a thick layer in gram-positive bacteria. Peptidoglycan serves to maintain a defined cell shape of the bacteria and also as an important scaffold for anchoring other components such as lipoproteins. Peptidoglycan is a disacharide chain consisting of an N-acetylyglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) with intercalating amino acid chains typically composed of four to five amino acids. Smaller products of the peptidoglycan complex containing MurNAc are called muropeptides. The minimum component that remains biologically potent is muramyl dipeptide (MDP), which consists of MurNAc and two amino acids, D-Ala and D-isoGln (or D-Glu)[355,370]. See figure **4.3.** As described in chapter 1, *NOD2* mutations are frequently observed in patients with Crohn's disease, suggesting the significance of the MDP-NOD2 pathway in activating immunity. Upon detection of MDP, NOD2 triggers a downstream inflammatory cascade via the NF-κβ and the MAPK pathway, resulting in the production of pro-inflammatory cytokines such as IL-6, IL-12, IL-8 and TNF-α[53].

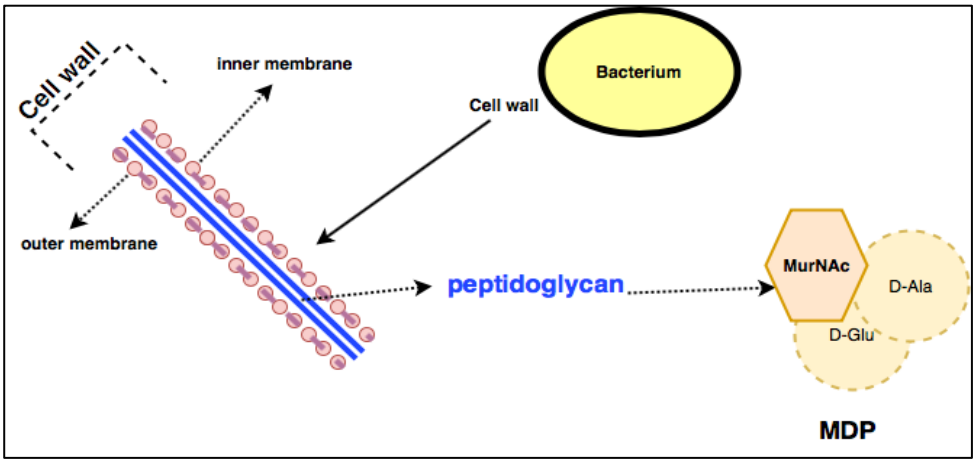


Figure 4.3 Structure of the bacterial cell wall with MDP

MDP is a component of the bacterial cell wall peptidoglycan and comprises MurNAc and two amino acids. In the host cell, MDP is detected by NOD2, an intra-cytoplasmic receptor belonging to the innate immune system.

4.2.3 The NF-κB signalling pathway

The NF-κB pathway activation results in the production of a family of transcription factors, which are critical regulators of immunity, inflammation, stress responses, cell proliferation, differentiation and apoptosis. A variety of stimuli converge on NF-κB

activation, which in turn mediates transcriptional responses for a wide spectrum of biologically diverse cellular functions. The stimuli include engagement of TLRs and other pathogen recognition receptors such as NOD2 with PAMPs, inflammatory cytokines (TNF-α, IL-1), antigen receptor engagement (T-cell and B-cell receptors), and genotoxic stimuli including UV or γ-irradiation, oxidative stress and ischemia[358,371]. Figure **4.4.**

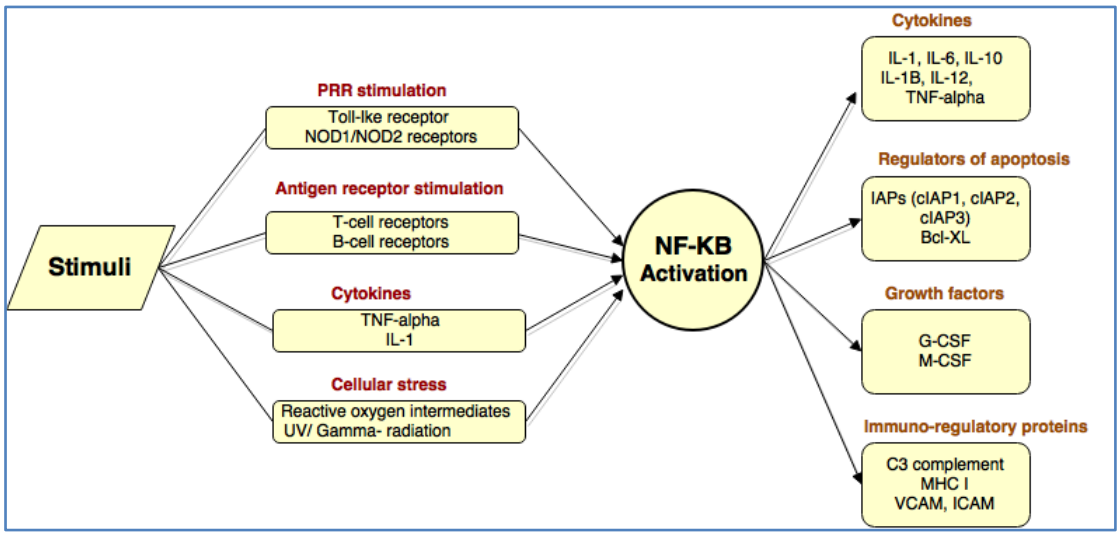


Figure 4.4 The NF-κB pathway activation stimuli and transcription of target genes

*The NF-κB pathway can be activated by a multitude of stimuli. Activation of this pathway leads to translocation of the NF-κB factors to the nucleus, leading to transcription of several proteins involved in a variety of key cellular processes. (**abbreviation**s: Bcl-XL, B-cell lymphoma extra large; cIAP, cellular inhibitors of apoptosis; G-CSF, Granulocyte- colony stimulating factor; ICAM, Intercellular adhesion molecule; M-CSF, Macrophage- colony stimulating factor; MHC, Major histocompatibility complex; VCAM-1, Vascular cell adhesion protein).*

In mammals, the NF-κB transcription factor family consists of five proteins, p65 (RelA), RelB and c-Rel, NF-κB1 (p105) and NF-κB2 (p100). NF-κB1 and NF-κB2 are precursor proteins, which are processed into p50 and p52 respectively. The transcription factors bind as dimers to the κB sites in promoters and enhancers of several target genes[359,372,373]. All the transcription factors share a Rel homology domain, responsible for DNA binding and dimerization. Under resting conditions, NF-κB dimers are bound to IκB proteins, which are inhibitory proteins keeping the transcription factors sequestered in the cytoplasm. Upon stimulation, IκB proteins are targeted by IκB kinases (IKK), which induce phosphorylation of the IκB proteins. IκB proteins undergo ubiquitination and proteasomal degradation, thereby releasing bound NF-κB dimers. These then translocate to the nucleus in order to induce or repress transcription of several target genes. The IKK complex consists of

catalytically active kinases, IKKα and/or IKKβ, and a regulatory sub-unit IKKγ (NF-κB essential modulator- NEMO). As the NF-κB activation is dependent on IκB degradation, the IKK complex plays a critical role as a gatekeeper and a converging node for interaction with other signalling pathways including TLR and NOD2 signalling[359,374].

NF-κB signalling operates via two main pathways, the canonical (the classical pathway) and the non-canonical pathway (the alternative pathway). **Figure 4.5**. For both pathways, IKK complex activation is a crucial step for the release and translocation of the NF-κB transcriptional factors to the nucleus. The canonical pathway can be activated by several stimuli including TNF-α, TLR ligands, IL-1 and other mediators such as NOD2. The activation of IKK in the canonical pathway is predominantly regulated by a multi-functional kinase, TAK1. TLR-dependent TAK1 activation has been described in the previous sections. Figure 11 represents TAK1 activation following TNF-α receptor (TNFR) stimulation with TNF-α, resulting in the recruitment of several adaptor proteins on the intra-cellular side of the TNFR such as TNF-R1-associated death domain protein (TRADD). This is followed by recruitment of other adaptors such as TRAFs, receptor interacting response protein-1 (RIP-1) as well as cellular inhibitors of apoptosis protein-1 (cIAP-1) and cIAP-2. Ubiquitination of RIP1, which is a critical protein of the TNF-α activation pathway leads to activation of TAK1 complex. Further downstream, activation of the NF-κB and MAPK ensues with transcription of pro-inflammatory genes[359,375].

The non-canonical pathway on the other hand, is induced by specific members of the cytokine family such CD40 ligand, B-cell activating factor (BAFF) and lymphotoxin-β. This pathway is largely for activation of p100/RelB complexes during B- and T-cell organ development and it proceeds through an IKK complex that contains two IKKα subunits, but without NEMO. Receptor binding with the stimulant leads to activation of the NF-κB-inducing kinase (NIK), which phosphorylates the IKK complex. This is followed by phosphorylation of the precursor protein p100 associated with RelB, with a consequent formation of a partially processed p100 and the generation of transcriptionally active p52-RelB dimers, translocating to the nucleus. This finally results in induction of genes regulating inflammatory and other cellular processes[359,376].

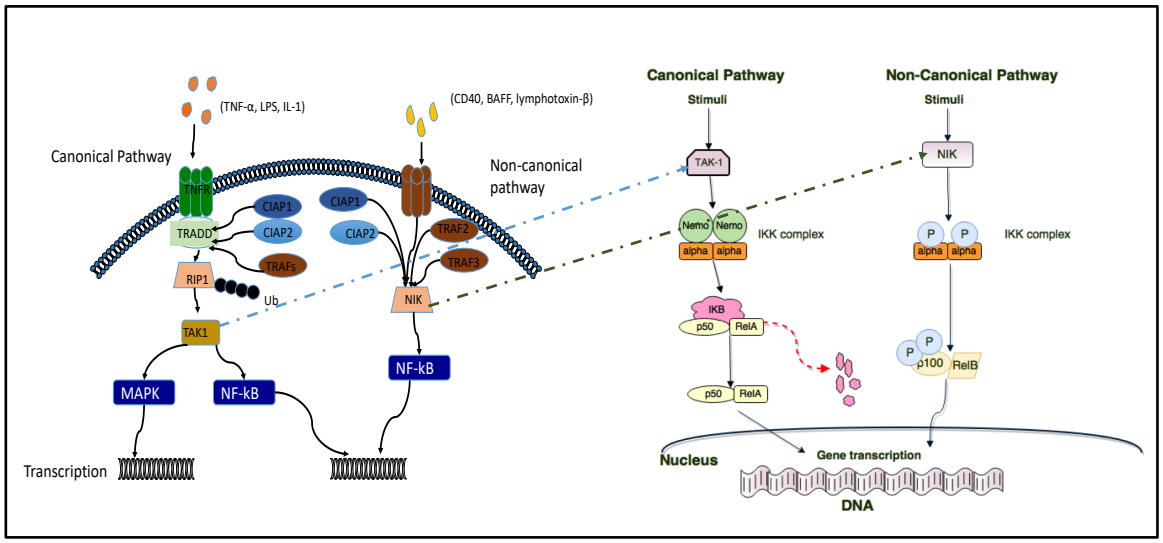


Figure 4.5 The canonical and non-canonical pathways of NF-κB activation

The figure shows the activation of NF-κB via the canonical and non-canonical pathways. ***The canonical pathway*** *is activated by TNF-α, LPS, IL-1 and other stimuli. The activation is mediated through the stimulation of TNF-R, which leads to the recruitment of several adaptor proteins. These include TRADD, which further interacts with other adaptors such as TRAFs, RIP1 and the inhibitors of apoptosis (IAPs- cIAP1 & cIAP2). Ubiquitination of RIP1 by the IAPs results in the activation of TAK1, which further downstream leads to activation of the MAPK and the NF-κB pathway. As shown on the right, TAK1 triggers activation of IκB kinases (IKK), resulting in the release of the NF-κB dimers (p50 and RelA), which translocate to the nucleus to induce transcription.* ***The non-canonical pathway*** *on the other hand, is induced by specific members of the cytokine family such CD40 ligand, BAFF and lymphotoxin-β. Stimulation of the receptor leads to the activation of NF-κB inducing kinase (NIK), which causes phosphorylation (indicated by a 'P' within a small circle) of the IKK complex. This leads to generation of transcriptionally active NF-κB dimers (RelB and p100), which translocate to the nucleus for induction of specific genes. (****Abbreviations****: BAFF, B- cell activation factor; cIAP, cellular inhibitors of apoptosis; LPS, lipopolysaccharide; MAPK, mitogen activated protein kinases; NF κB, nuclear factor--κB; NIK, NF-κB inducing kinases; RIP1, receptor interacting response protein 2; TAK1, TGF-β-activated kinase; TRADD, TNF-R1-associated death domain protein; TRAF, TNF-receptor associated factor; Ub, ubiquitin).*

In this assay, the functional integrity of the NF-κB signalling pathway was indirectly assessed through the quantification of the cytokines, IL-10, IL-6, IL-1β, IL-8 and TNF-α, which are secreted following the activation of the pathway. As the NF-κB is a critical downstream signalling hub following stimulation of the pathogen recognition receptors, an impaired functioning of the pathway, would theoretically result in an altered secretion of the cytokines following stimulation by all of the three stimulants used in this assay.

4.2.4 Inflammasome activation pathway

Inflammasomes are a group of intra-cellular multimeric protein complexes, assembled in response to pathogenic signals and sterile stressors. Inflammasomes are key signalling platforms involved in the initiation, propagation and regulation of inflammatory responses, triggering the activation of highly potent pro-inflammatory cytokines IL-1β and IL-18. Each complex consists of a ligand-sensing molecule, an adaptor protein- apoptosis associated speck-like protein containing CARD (ASC) and pro-caspase1[377]. The ligand-sensing proteins usually include members of the NOD-like receptor (NLR) family. Among the well-characterised NLR receptors that constitute the inflammasome sensors include NLRP1, NLRP3, NLRP6 and NLRC4[378]. Inflammasome sensors not containing NLR complexes include absent in melanoma 2 (AIM2) protein and IFN-γ inducible protein-16 (IFI16). The adaptor protein ASC is common to all inflammasomes and contains two death-fold domains, a pyrin domain and a caspase activation and recruitment domain (CARD). Upon stimulation, inflammasome sensors undergo self-oligomerisation and recruit pro-caspase1 via ASC. ASC interacts with the inflammasome sensor via the pyrin domain, which triggers assembly of ASC into a large multimeric protein formed of ASC dimers. Using its CARD domain, ASC brings together monomers of pro-caspase1, initiating self-cleavage and the formation of the active caspase1. Activated caspase1 then cleaves pro- IL-1β and pro-IL-18 into their mature forms, resulting in the production of highly potent pro-inflammatory cytokines[377].

Inflammasome activation is tightly regulated by two key processes; transcription of the precursor forms of these cytokines, and their maturation and release. The process starts with the activation of NF-κB pathway following upstream signalling cascades via TLRs, NOD2, TNFR and other receptors. NF-κB transcription factors upon translocation to the nucleus promote the transcription of pro- IL-1β, pro-IL-18 as well as the NLRP3 and other sensors, critical to inflammasome activation[378,379].

NLRP3 represents the most extensively studied inflammasome sensor[377,380]. It is generally accepted that NLRP3 activation leading to pro- IL-1β and pro-IL-18 maturation proceeds through a two-step activation mechanism. The first signal (priming signal) provides transcriptional upregulation pro- IL-1β and NLRP3. Pro-IL-18 on the other hand is constitutively expressed within the cells, although its expression is increased after cellular activation. Signal 1 is achieved through activation of the NF-κB pathway through various upstream stimuli as described previously. Signal 2 involves stimulation of the NLRP3

followed by activation of caspase-1 and proteolytic maturation of pro- IL-1β and pro-IL-18. Signal 2 can be triggered by diverse stimuli including PAMPs, cellular stress, TLR stimuli as well as MDP. Although the exact mechanism of NLRP3 activation remains poorly understood, a common denominator of its activation appears to be the cytosolic efflux of potassium[379]. The two-step process of NLRP3 activation is conventionally known as the canonical or classical inflammasome activation to distinguish this from the more recently reported non-canonical or 'alternative' inflammasome activation. Non-canonical activation has been described in human monocytes in response to TLR4 stimulation via LPS[380,381]. Unlike canonical inflammasome activation, alternative activation proceeds independently of K^+ efflux and pyroptosis, while it engages a series of adaptors upstream of NLRP3 following TLR4 stimulation. These adaptors include TIR-domain-containing adaptor-inducing interferon-β (TRIF), receptor-interacting protein kinase1 (RIPK1), Fas-associated protein with death domain (FADD) and caspase8 (CASP8). TLR4 stimulation therefore can induce maturation of IL-1β through both pathways and it has been observed that contribution from the non-canonical circuit can increase IL-1β secretion several-fold. The advantage of this alternative mode of activation to inflammasome biology is that unlike the classical pathway, it allows the host cell to respond to activating stimuli through a gradual release of IL-1β without committing to non-reversible cell death[381]. See figure **4.6.**

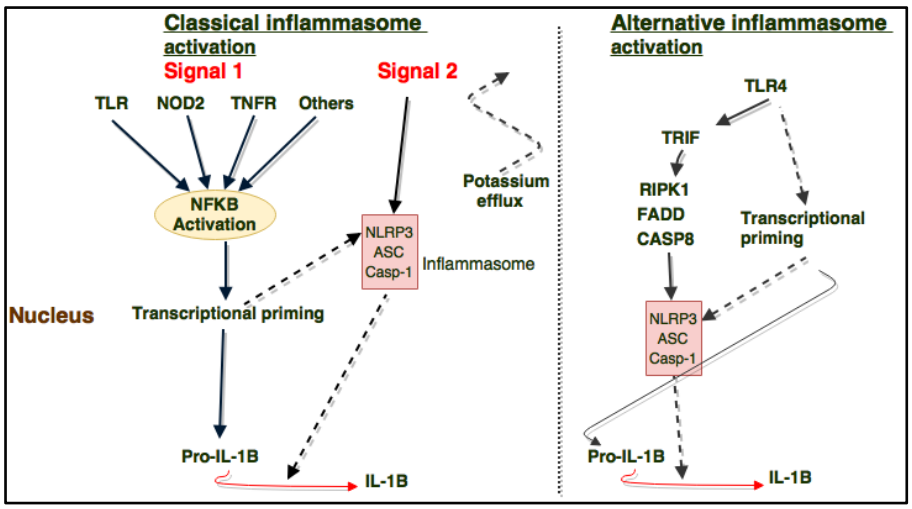


Figure 4.6 Classical and alternative inflammasome activation

Classical activation (left): *Activation of the NF-κB pathway by various upstream stimuli results in the transcription of pro-inflammatory cytokines including pro-IL-1B. This constitutes signal 1 of inflammasome activation. Signal 2 involves stimulation of the NLRP3 followed by activation of caspase-1 and proteolytic maturation of pro- IL-1β and pro-IL-18. Signal 2 can be triggered by diverse stimuli including PAMPs, cellular stress, TLR stimuli as well as MDP and is marked by*

efflux of potassium out of the cell. ***Alternative/Non-canonical activation (right):*** *This occurs in human monocytes in response to TLR4 stimulation involving various adaptor proteins including TRIF, RIPK1, FADD and CASP8. Non-canonical activation proceeds independently without K^+ efflux or signal 2. (**Abbreviation**s: ASC, adaptor protein- apoptosis associated speck-like protein containing CARD; FADD, Fas-associated protein with death domain; LPS, lipopolysaccharide; NF-κB, nuclear factor--κB; NLRP3, NOD-like receptor family pyrin domain containing 3; NOD, nucleotide-binding oligomerization domain RIP1, receptor interacting response protein 2; TRIF, TIR-domain containing adaptor-inducing interferon-β).*

As shown in figure **4.6**, the inflammasome pathway can be activated by several stimuli including the three ligands Pam3CSK4, LPS and MDP. **In this assay**, the functional integrity of the pathway was assessed through the measurement of IL-1β, which is produced as a result of cleavage of its precursor form following inflammasome activation[378]. Altered secretion of this cytokine would therefore indicate impaired functioning of the inflammasome complex.

Mitogen-activated protein kinases (MAPK) pathway

MAPKs are a highly conserved family of serine/threonine protein kinases which play a key role in transduction of extra-cellular signals to cellular responses. MAPK cascades relay, amplify and integrate signals from a diverse range of stimuli and elicit an appropriate response including cellular proliferation, differentiation, inflammatory/stress responses and cell survival. The three well-characterised MAPK families in mammalian cells include the classical MAPK (also known as extracellular signal regulated kinase- ERK), C-Jun N-terminal kinase/stress-activated protein kinase (JNK/SAPK) and p38 kinase. Each cascade comprises no fewer than three enzymes that are activated in series; a MAPK kinase kinase (MAPKKK), a MAPK kinase (MAPKK) and a MAP kinase (MAPK)[382]. TAK-1 described in the previous section is a form of MAPKKK. The events downstream to TAK-1 activation include recruitment of other MAPK family members including ERK1/2, p38 and JNK. This leads to activation of AP-1 family of transcription factors, which regulate a diverse range of cellular processes including inflammatory responses[361].

4.3 METHODS

During the early stages of my project, experiments were conducted in healthy adult controls to find the optimal set-up for the starting material (biological specimens), ligands and ligand doses, cell numbers, incubation times, effector cytokine measurements and

appropriate reference material. My initial experimental approach was guided by previously established protocols in the WISH laboratory and published literature. The experimental work following this led on to the development of an optimised assay, which became my standard operating procedure and was subsequently applied to patient specimens and paediatric controls. Figure **4.7.**

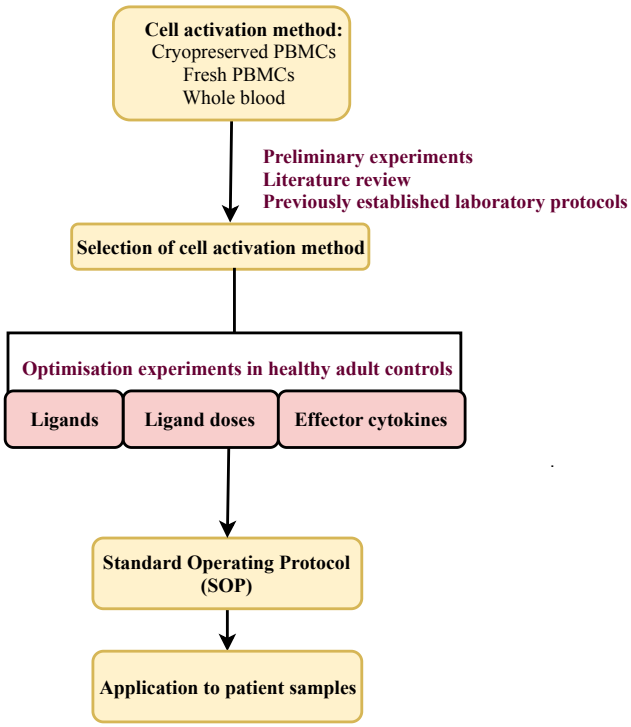


Figure 4.7 Plan of assay development and optimisation

During the developmental phase, preliminary experiments were performed to select a suitable cell activation method for assessing immune induction. This was followed by optimisation experiments in healthy controls in order to develop a standard operating protocol (SOP), which would then be applied to patient samples.

The methods section in this chapter includes **two main sub-sections:**

1. **Methodological decisions** based on previously established protocols and published literature

2. **Experimental methods**- to include the experiments conducted by myself to develop and optimise an immunological assay method

4.3.1 Methodological decisions from established protocols and published literature

The selection of ligands, ligand dose-ranges, cell concentrations and incubation times for the early experiments was based on literature review, multi-disciplinary-team discussion and results from previously conducted experiments in the WISH laboratory.

Ligands and dose-ranges

The ligands used included lipopolysaccharide (LPS), synthetic tri-acylated lipoprotein (Pam3CSK4), muramyl dipeptide (MDP) and L-Ala-γ-D-Glu-mDAP (TriDAP). For LPS, an optimised concentration of 1μg/ml, established through previous projects in the WISH laboratory was used. Previously published studies have also shown that a concentration of 1μg/ml of LPS is suitable to induce cytokine responses[263,365,369,383,384]. As the optimal dose for this ligand was already established, LPS was used as a positive control ligand to assess cytokine production induced by the other three ligands. In this assay, incremental doses of Pam3CSK4, MDP and Tri-DAP were used as shown in table **4.2.**

Table 4.2 Ligands and the concentrations used for cell stimulation

Ligand	Receptor	Conc. 1 1ng/ml	Conc. 2 10ng/ml	Conc. 3 100ng/ml	Conc. 4 1μg/ml	Conc. 5 10μg/ml
LPS	TLR4	Positive Control ligand			✔	
Pam3CSK4	TLR1-2	✔	✔	✔	✔	✔
MDP	NOD2	✔	✔	✔	✔	✔
Tri- DAP	NOD1	✔	✔	✔	✔	✔
None	Negative Control					

*LPS was used as a positive control ligand with a concentration of 1μg/ml. Incremental concentrations were used for the other ligands. Up to five different doses were used per ligand. (**Abbreviations:** LPS- lipopolysaccharide, Pam3CSK4- synthetic tri-acylated lipoprotein, MDP- muramyl dipeptide, TLR- toll-like receptor, Tri-DAP- L-Ala-γ-D-Glu-mDAP).*

Published studies varied widely in their methods using differing combinations of ligands at different cell concentrations and variable incubation times. These are summarised in table **4.3.**

Table 4.3 Published literature on induced immune responses- cell concentrations, ligand doses and incubation times

Publications	Activation method	Cell concentration (million/µl)	Ligands- Concentration (µg/ml)				Incubation (hours)
			LPS	Pam3CSK4	MDP	Tri-DAP	
Hirschfield et al. 2007[385]	PBMCs (unclear)	0.2 /200	0.1	1	-	-	24
Benyoucef et al. 1997[386]	Whole blood	25µl	25µg/ml	-	-	-	24
Li et al. 2004[124]	PBMCs (unclear)	3/1000	-	-	DR (0-10) OD (1)	-	16
Van Heel et al. 2005[251]	PBMCs (unclear)	0.2 million cells per condition	0.1	0.01	DR (0.001-1) OD (1)	DR (0.001-10) OD (1)	22
Netea et al. 2005[267]	PBMCs (fresh)	0.5 million/100	0.1	-	-	10	18
Van Heel et al. 2005[260]	PBMCs (fresh & frozen)	0.2 million cells per condition	-	-	DR (0.001-1) OD (0.1)	-	22
Orange et al. 2006[384]	PBMCs (frozen)	0.2 /100	DR (0.02-2) OD (0.1-1)	DR (4-40) OD (4-10)	-	-	24
Butler et al. 2007[264]	Dendritic cells (fresh)	0.25/500	DR (0-0.1) OD (0.1)	-	DR (0.1- 1) OD (0.1)	-	24
Canto et al. 2009[262]	PBMCs (frozen)	1/1000	0.25	-	DR (0.1-10)		LPS- 22 MDP- 16
Peeters et al. 2007[254]	Monocytes (fresh & frozen)	1/1000	0.01	-	0.1	-	20
Ferwerda et al. 2008[387]	PBMCs (unclear)	0.5 million/100	0.001	-	100 nM*	-	24
Kullberg et al. 2008[266]	PBMCs (unclear)	0.5 million/100	0.001	10	0.1-10	-	MDP- 48 LPS- 24 Tri-DAP- 24
Brosbol-Ravnborg et al. 2009[263]	PBMCs (unclear)	0.1 million cells per condition	1	-	0.02	-	18
Damsgaard et al. 2009[388]	Whole blood	500µl	1	-	-	-	24
Worthey et al. 2010[122]	PBMCs (unclear)	-	-	-	20	20	24
Wagner et al. 2013[277]	PBMCs (frozen)	0.5 million cells per condition	0.1	-	-	-	20
Zeissig et al. 2013[383]	PBMCs (unclear)	0.1 million cells	DR (0-1) OD (1)	DR (0-10) OD (10)	DR (0-10) OD (10)	DR (0-10) OD (10)	24
Ammann et al. 2014[389]	PBMCs (fresh)	1-1.5 million/1000	0.2	-	0.2	-	2.5

*Table 4.3 provides a synopsis of the cell concentrations, ligand doses and incubation times utilised by various studies. Studies varied in the number of ligands used. Only the 4 ligands pertinent to this study are shown in this table. *In this study, the dose of MDP was indicated in nM. (**Abbreviations**: DR- dose range; OD- optimal dose).*

Incubation time for cell activation

There is no clear guidance on the recommended incubation times for optimal transcriptional responses following activation of cell cultures. As evident from table **4.2,** a wide range of incubation times have been used by studies assessing cytokine responses in primary cell culture media. Orange *et al* performed a time course experiment comparing 10 mins, 1 h, 2 h, 4 h, 6 h, 24h, and 48 h of incubation with TLR ligands including Pam3CSK4, LPS, flagellin and others. Peak levels of for all ligands were achieved within 24 hours, although optimal levels were also observed for shorter incubations (>6 h). Incubation times of up to 48 hours were not significantly different compared to 24 hours for cytokine induction, although reduced amounts of TNF-α were observed with certain ligands such as TLR7 at 48 hours. Due to peak responses for all the ligands at 24 hours and the convenient assay length, the authors recommended 24 hours of incubation as adequate for assessing induced immune responses[384]. A 24 hours incubation time has also been employed for NOD2 stimulation by several studies[264,266,383,387]. Therefore, combined with evidence from published literature, practical advantages of a convenient assay length and previous experience at the WISH laboratory, a 24-h incubation time was utilised for all experiments assessing induced immune responses.

Concentration of PBMCs and whole blood

Studies using primary cell culture methods to assess induced immune responses vary widely in the concentration of cells used per stimulant condition. See Table **4.2.** The cell concentrations used by various studies range between 0.1 million up to 1.5 million cells per condition without any specific justification for the concentrations utilised. Based on the supportive evidence from published literature and local expertise at the WISH laboratory, a concentration of 0.3 million PBMCs and 0.3mls of whole blood per condition were chosen for all experiments evaluating induced immune responses. This concentration was considered appropriate bearing in mind the difficulties in obtaining adequate volumes of blood from paediatric patients for future assays and the requirement to run experiments in duplicate.

4.3.2 Experimental methods

4.3.2.1 Cytokine responses between cryopreserved (frozen) PBMCs, freshly extracted PBMCs and whole blood

During the early phases of the project, we sought to compare cytokine responses using cryopreserved PBMCs, freshly cultured PBMCs and whole blood from two healthy adult individuals. Cryopreserved PBMCs were thawed as per methods described in chapter 3 and the supplementary. Cell suspensions were stimulated with MDP, Tri-DAP, Pam3CSK4 and LPS along with a negative control (nil addition of ligand) as previously described in table **4.2**. A concentration of 0.3 million cells per condition was used. For freshly cultured PBMCs, similar stimulant conditions and cell concentrations were used. Five different concentrations of MDP, Tri-DAP and Pam3CSK4 were employed as previously described in table **4.1**. For whole blood cultures, 0.3mls of whole blood was incubated with ligands as for PBMC culture activation. After 24 hours of incubation, the samples were centrifuged, serum/culture supernatant aspirated and stored at -20°C for cytokine assessment. The samples were subsequently analysed for IL-1β, IL-6, TNF-α, IL-10 and IL-8 using the Luminex® assay method described in chapter 3.

The experiments were primarily conducted as proof of concept. The findings along with supportive evidence from published literature helped me decide on the suitability of biological specimens for conducting future assays in paediatric patients.

4.3.2.2 Dose-titration experiments using cryopreserved PBMCs

In order to optimise ligand concentrations and develop an optimal set-up for patient specimens, dose-titration experiments were conducted in a total of 10 healthy adult individuals using cryopreserved PBMCs. Given the technical nature of these assays and my developing laboratory skills, some inconsistencies were observed in the dose-response curves in the experiments conducted in the **first five individuals**. These experiments were conducted in the early stages of my project. Results of these initial experiments using cryopreserved PBMCs are presented in the appendices. The outcome of this iterative phase of development was to fine tune to the lessons learnt and improve my technical skills in the laboratory.

Assays using cryopreserved PBMCs were then **repeated in five additional individuals**. In these repeat experiments, bloods for PBMCs extraction were obtained through a one-time venesection from each individual. The extracted PBMCs were stored in cryovials, which

were used in two separate aliquots. Dose-titration experiments were conducted on the first aliquot from each individual followed a week later by experiments on the second aliquot using identical assay conditions in order to ensure repeatability. After addition of ligands, cell suspensions of PBMCs were incubated for 24 hours followed by harvesting the cellular supernatants. The supernatants were stored at -20^{o}C until a batched analysis with Luminex® assays to quantify the cytokines from the supernatants. All stimulant conditions were assayed in duplicate. Further details on the methods are described in chapter 3.

Standardisation of assay readings with monocyte counts

Monocyte counts were analysed in each of the individuals and the final quantity of cytokine read-outs assessed were standardised to a monocyte count of 10% to avoid variability in the cytokine read-outs. Studies assessing induced immune responses have shown that monocytes are the predominant cytokine producing cells within the PBMC population[390-392]. Monocyte counts within a healthy population show a wide variability, ranging from 3.43% to 14.9%[393,394]. In a study analysing the cytokine secretory profile of cells within the PBMC population upon stimulation with LPS, monocytes were observed to contribute to 90% of the cytokine levels secreted. These included among others IL-10, IL-6, IL-1β and TNF-α[391]. Studies have also shown that the other cells within the PBMCs, such as the B cells are not LPS-responsive[395] and the T cells fail to produce cytokines unless provided with co-stimulatory signals[396]. Additionally, NOD2, which is one of the key pathogen-recognition receptors investigated in our study, is primarily expressed in monocytes[397].

4.3.2.3 Induction of cytokine responses using purified monocytes

Peripheral blood monocytes were isolated from PBMCs from one healthy individual. Methods on monocytes isolation have been described in chapter 3 and the supplementary information in Appendix B. The isolated monocytes were suspended in RPMI 1640 medium and incubated for 24 hours following addition of ligands. In the optimisation experiments (section 4.3.2.2), a concentration of 0.3 million PBMCs was used per stimulant condition, which would theoretically contain monocytes ranging from 9000 to 42000 per condition (assuming monocyte counts ranging between 3-14% in a healthy population)[398,399]. Therefore, in these experiments using pure monocytes, incremental doses of monocytes per condition were selected as follows: 0, 7000, 15000, 30000 and 60000 monocytes. The absolute cell numbers used in the experiment were also based on the available number of monocytes harvested from a single individual at a one-time blood

draw to be able to conduct all the steps and conditions planned at the outset. Cellular supernatants were harvested and stored at -20°C prior to analysis. Cytokine levels were measured on Luminex®, according to the manufacturer's instructions. Cytokine induction was calculated as the concentration of the cytokine in the culture medium pre-stimulation subtracted from the culture medium post-stimulation.

4.3.2.4 Induction of cytokine response using a combined population of purified monocytes and TBNK cells

Peripheral blood monocytes were isolated from PBMCs from one healthy individual, using magnetic sorting with the Human Monocyte Isolation Kit II (Miltenyi Biotec, Amsterdam) as described in chapter 3. The isolated monocytes were re-suspended in complete RPMI 1640 medium (Invitrogen, Merelbeke, Belgium) and incubated for 24 hours following addition of ligands. Monocytes were added to a cell suspension of TBNK cells in Falcon tubes in RPMI 1640 suspension in an incremental manner for assessment of a dose-response effect. The incremental doses of monocytes per condition were as follows: 0, 7000, 15000, 30000 and 60000 monocytes as in the previous section, which were added to a fixed number of TBNK cells (240,000), generating ratios of monocytes to TBNK cells ranging from 1:40 to 1:4. Although the ratios of monocytes to TBNK cells in the peripheral blood vary widely within the population, a study conducted in 1277 children reported a ratio of monocytes to TBNK cells approximating to 1:6[400].

4.3.2.5 Application of the assay method using combination of monocytes and TBNK cells to patient specimens

The assay method described in section **4.3.2.4** was applied to 33 pIBD patients alongside 15 healthy adult controls. **The results of these experiments using a combination of monocytes and TBNK cells in patients and controls were however excluded from the final analysis.** Further details and reasons for exclusion of these experiments is discussed in the results section.

4.4 RESULTS

4.4.1 Cytokine responses between cryopreserved PBMCs, freshly extracted PBMCs and whole blood

MDP-induced cytokine responses

Cytokine response patterns for IL-1β and IL-6 were similar in both fresh and cryopreserved PBMCs, with a three-fold increase from the baseline (without a stimulant) and a plateauing effect between ligand concentrations of 1-10μg/ml. With whole blood, cytokine induction was more than two-fold higher compared to cryopreserved and fresh PBMCs, but the responses were inconsistent with incremental doses. One individual showed nil response to LPS-mediated (positive control) IL-6 with fresh PBMCs, possibly due to technical failure. There was poor or nil induction for IL-10 in both fresh and cryopreserved PBMCs. There were generally poor or nil responses for TNF-α, except in the positive controls. Dose-related induction for IL-8 was observed with whole blood in one individual. IL-8 responses were generally inconsistent with dose increments with both fresh and cryopreserved PBMCs. See figure **4.8.**

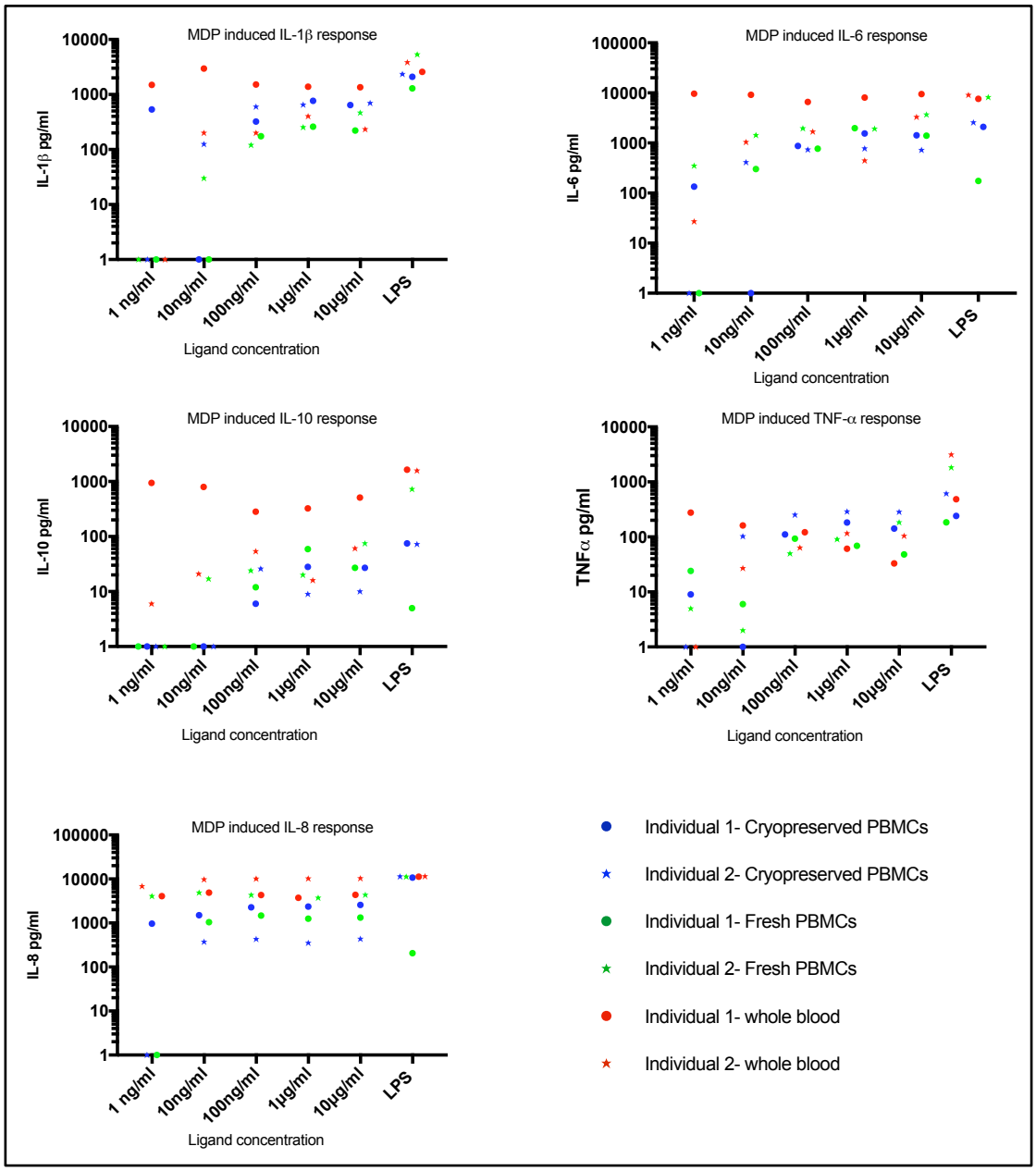


Figure 4.8 MDP-induced cytokine responses in two individuals

Incremental doses of MDP are shown along the X-axis. LPS also shown along the X-axis. MDP-induced cytokine responses are plotted along the left Y-axis and LPS-induced responses (positive control ligand) on the right Y-axis. Solid circles represent individual 1 and stars represent individual 2.

Pam3CSK4-induced cytokine responses

A three-fold increase from the baseline was observed for IL-1β in one individual plateauing at 10ng/ml and more than a five-fold response to the positive control. Similar induction was noted with whole blood in this individual. Induction with fresh PBMCs was generally poor, except at the higher dose of 10µg/ml of Pam3CSK4 and for the positive control ligand. IL-6 responses were similar for fresh PBMCs, cryopreserved PBMCs and whole blood, although some inconsistencies were observed at lower doses for fresh

PBMCs and whole blood. Poor induction was observed for IL-10 and TNF-α with all three methods, and inconsistent dose-titration responses for IL-8. Figure **4.9.**

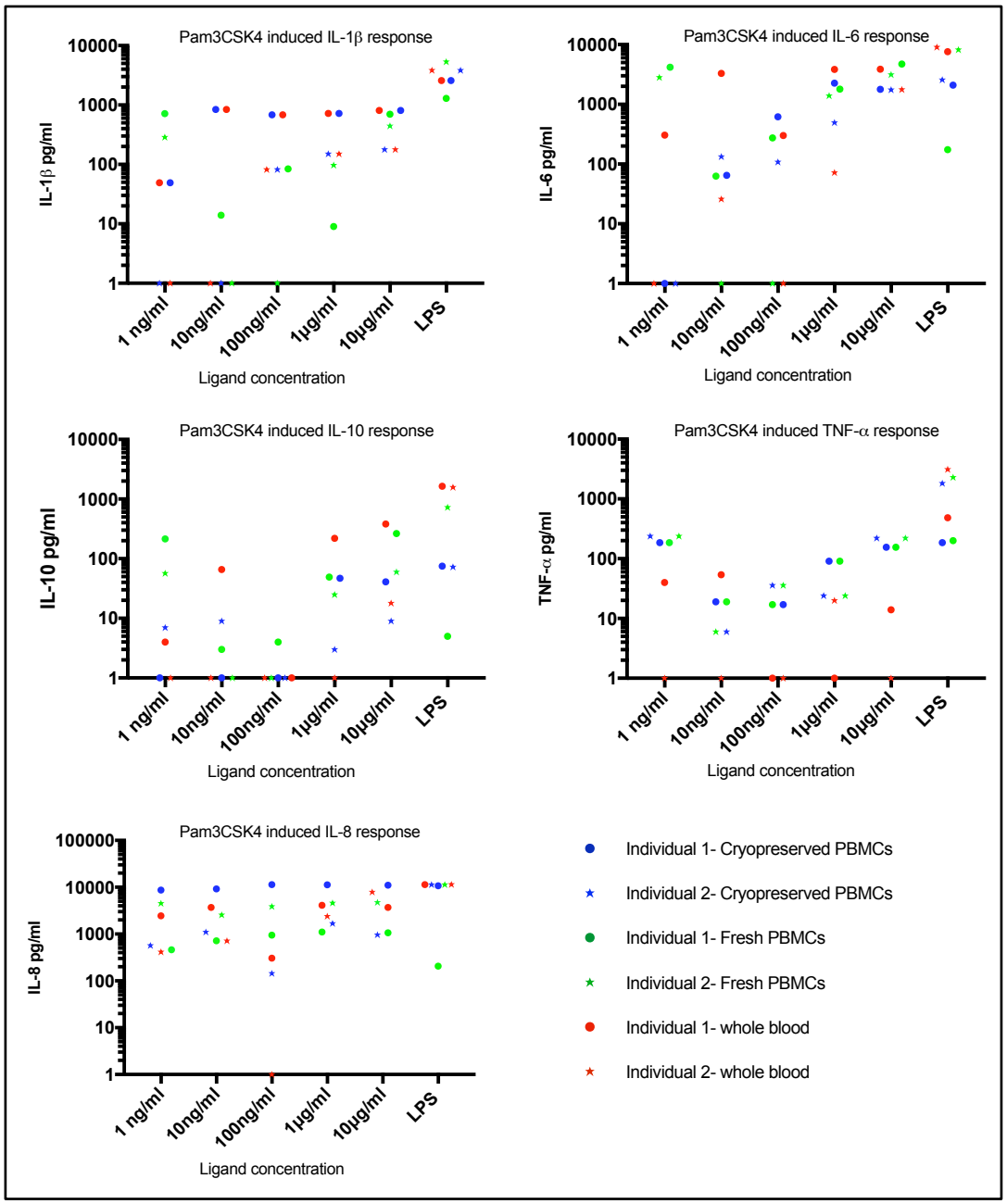


Figure 4.9 Pam3CSK4-induced cytokine responses

Incremental doses of Pam3CSK4 are shown along the X-axis. LPS also shown along the X-axis. Pam3CSK4-induced cytokine responses are plotted along the left Y-axis and LPS-induced responses (positive control ligand) on the right Y-axis. Solid circles represent individual 1 and stars represent individual 2.

Tri-DAP-induced cytokine responses

Stimulation with Tri-DAP showed transcriptional responses at a concentration of 10µg/ml, but variable and inconsistent dose-titration responses between 1ng/ml to 1µg/ml. The overall dose responses were poor and inconsistent across all three methods. Figure **4.10**.

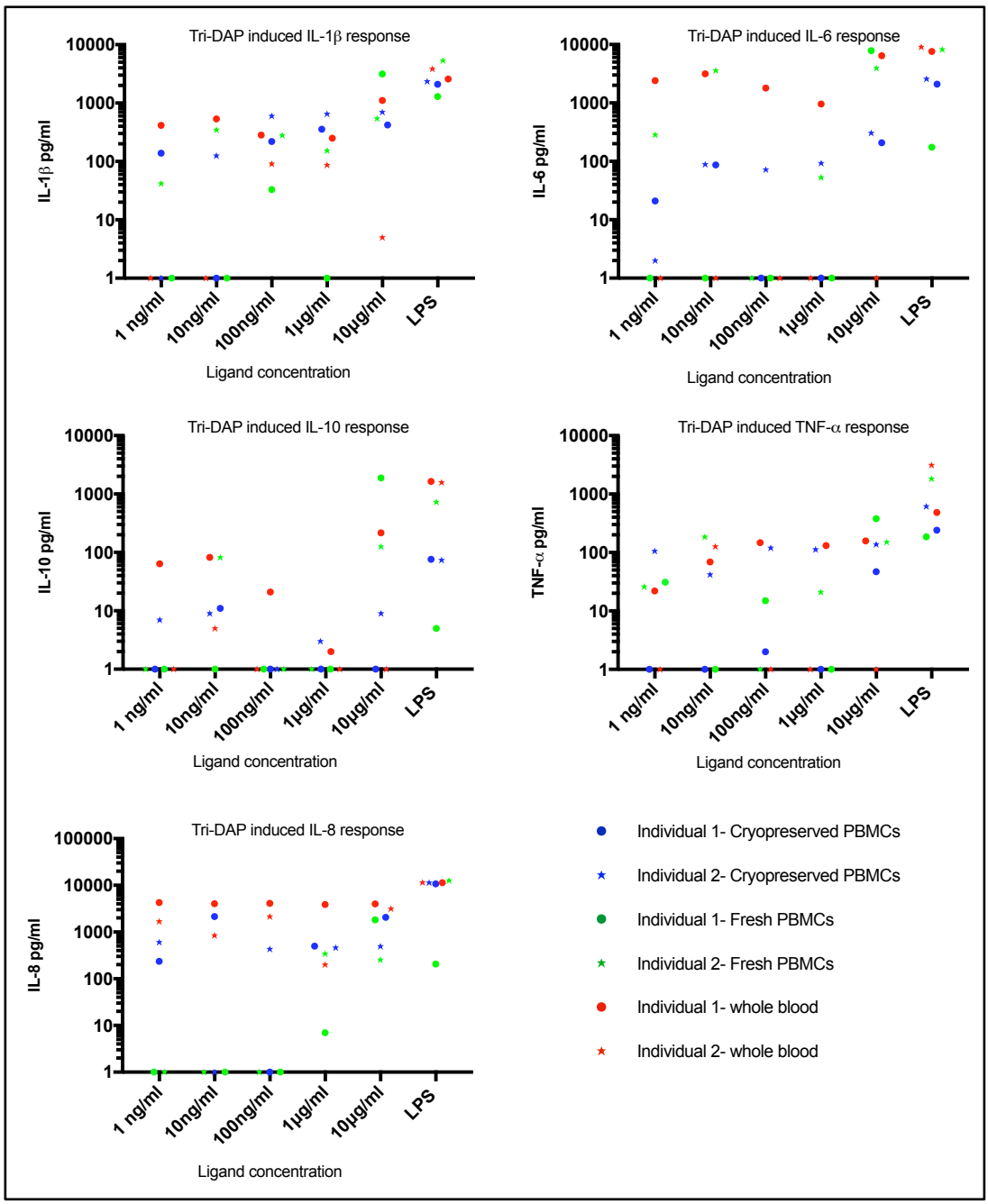


Figure 4.10 Tri-DAP induced cytokine responses in two individuals

Incremental doses of Tri-DAP are shown along the X-axis. LPS also shown along the X-axis. Tri-DAP induced cytokine responses are plotted along the left Y-axis and LPS-induced responses (positive control ligand) on the right Y-axis.

Interpretation of results

A range of variability of immune responses was observed between the two healthy controls. This was more pronounced at lower ligand concentrations, with a reduced variability at a concentration of 10µg/ml. Significant differences were observed for MDP-mediated IL-1β responses between fresh v/s cryopreserved PBMCs, and fresh v/s whole blood, using paired t-tests including cytokine values at all incremental doses for the two individuals. Also, significant differences were observed between all the three cell culture methods for MDP-mediated IL-6 responses. For the rest of ligands (Pam3CSK4, Tri-DAP and LPS), there were no significant differences between the three culture methods for IL-1β and IL-6. See table **4.4.** As cytokine responses were generally poor and inconsistent with dose-titration for IL-10, TNF-α and IL-8, t-tests were not performed for these cytokines. Overall, cytokine induction with the whole blood method was higher compared to PBMCs, and induction via cryopreserved PBMCs was higher compared to the freshly cultured PBMCs.

Table 4.4 Cytokine induction between three cell activation methods for IL-1β and IL-6

Ligand	Cytokine	Fresh V cryopreserved	Fresh V WB	cryopreserved V WB
MDP	IL-1β	0.002	0.021	0.1312
Pam3CSK4		0.161	0.459	0.671
Tri-DAP		0.661	0.464	0.767
LPS		0.664	0.947	0.3
MDP	IL-6	0.038	0.034	0.007
Pam3CSK4		0.061	0.473	0.1437
Tri-DAP		0.107	0.897	0.0638
LPS		0.708	0.425	0.053

The differences in cytokine induction between fresh PBMCs, cryopreserved PBMCs and whole blood (WB) are shown in the table. Paired t-tests for IL-β and IL-6 were performed between each of the activation methods, including cytokine values for the two individuals at all incremental doses. Significant p values are indicated in red.

4.4.2 Dose-titration experiments using cryopreserved PBMCs

MDP-induced cytokine responses

Reliable dose-response effect was observed at a concentration of 0.1µg/ml, with a peak production at the highest concentration of 10µg/ml. At a concentration of 10µg/ml, the geometric means of cytokine responses were; IL-1 β: 564pg/ml, IL-6: 574pg/ml, IL-10:

425pg/ml and TNF-α: 135pg/ml. LPS induced cytokine responses were consistently high with the strongest response for IL-10 (geometric mean for IL-10: 6780pg/ml, IL-1 β: 2836pg/ml, IL-6: 1228pg/ml and TNF-α: 2457pg/ml). As peak induction of cytokines was observed at **10µg/ml,** this concentration was chosen as the optimal dose for cell stimulation. Figure **4.11.**

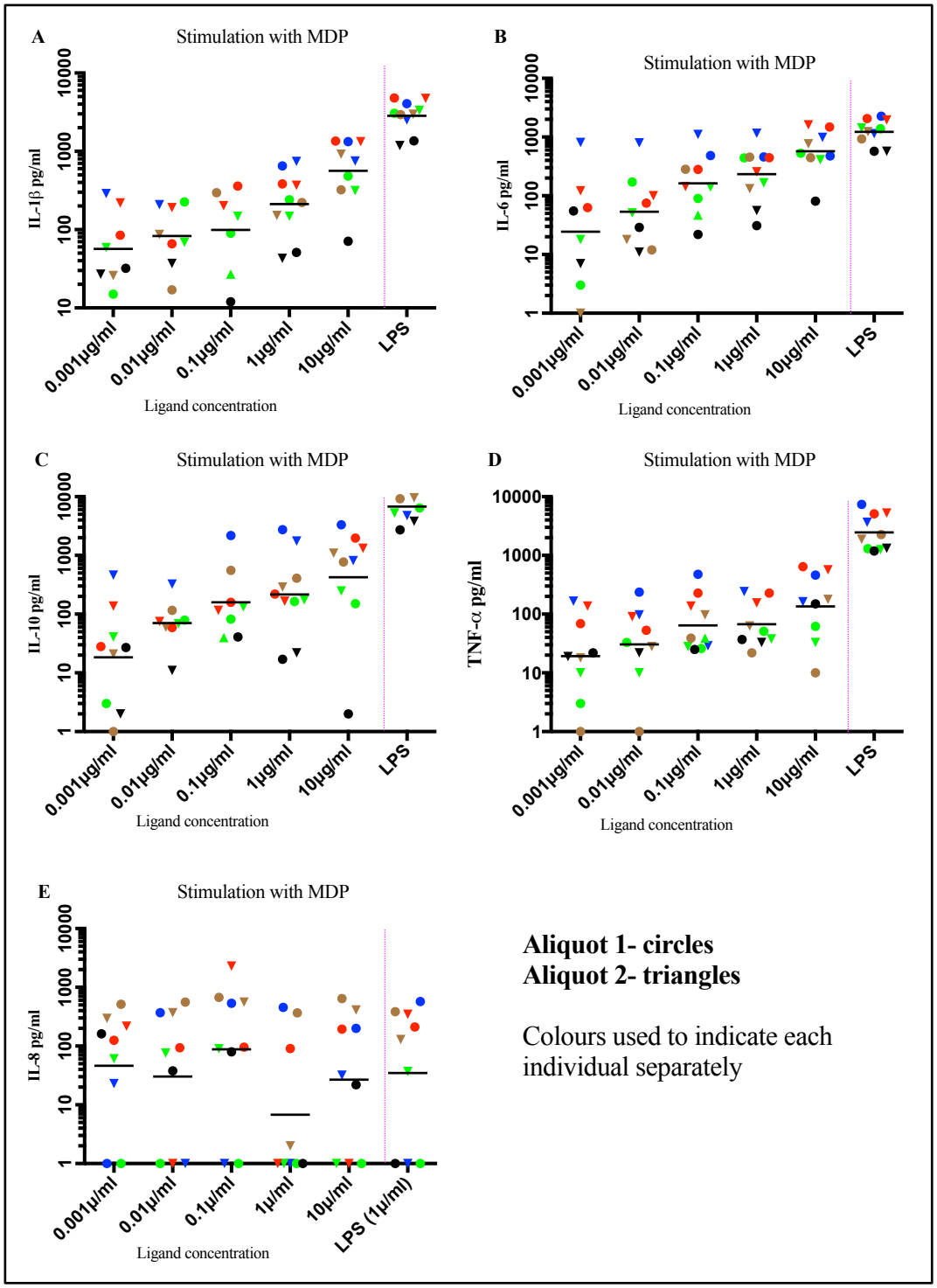


Figure 4.11 MDP-induced cytokine responses

MDP-induced cytokine responses with incremental doses of MDP from five healthy adult

individuals, performed on two separate aliquots of cryopreserved PBMCs are shown in this figure. Cytokine responses induced by MDP are plotted on the left Y-axis, LPS-induced responses (positive control) on the right Y-axis, and the incremental doses of MDP on the X-axis along with a pre determined single dose of LPS. The horizontal bars represent the geometric mean of the cytokine responses. Peak induction was observed at a concentration of 10μg/ml for MDP.

Pam3CSK4-induced cytokine responses

Dose-responses for Pam3CSK4 are shown in figure **4.12** along with responses to LPS, which was used as a positive control ligand. Robust cytokine induction was observed at concentrations of 0.1μg/ml and above. Plateauing was observed at a ligand concentration of 1μg/ml, with comparable cytokine profiles at 1μg/ml and 10μg/ml. The geometric means of cytokine responses at concentrations of 1μg/ml and 10μg/ml were; IL-10: 1565pg/ml & 1494pg/ml, IL-1 β: 476pg/ml & 529pg/ml, IL-6: 1133pg/ml & 1148pg/ml and TNF-α: 228pg/ml & 221pg/ml respectively. Based on the observed maximal induction, a concentration of **10μg/ml** of Pam3CSK4 was chosen as the optimal concentration.

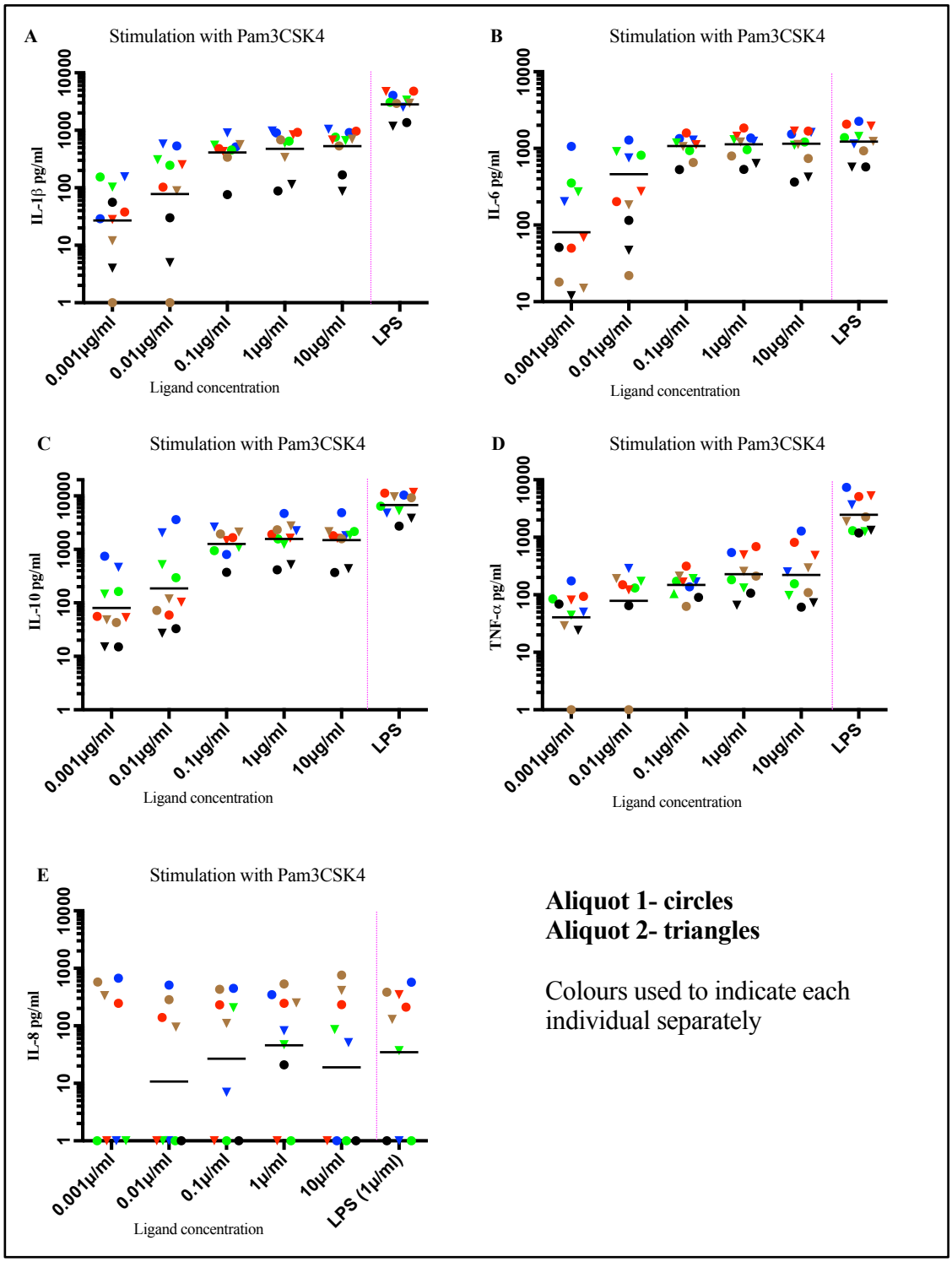


Figure 4.12 Pam3CSK4-induced cytokine responses

Pam3CSK4-induced cytokine responses with incremental doses of Pam3CSK4 from five healthy adult individuals, performed on two separate aliquots of cryopreserved PBMCs are shown in this figure. Cytokine responses induced by Pam3CSK4 are plotted on the left Y-axis, LPS-induced responses (positive control) on the right Y-axis, and the incremental doses of Pam3CSK4 on the X-axis along with a pre-determined single dose of LPS. The horizontal bars represent the geometric mean of the cytokine responses. Peak induction was observed at a concentration between 1-10µg/ml.

Tri-DAP-induced cytokine responses

Stimulation with Tri-DAP did not result in a significant and consistent induction of cytokines. The geometric means across all the cytokines were generally less than 100pg/ml. Figure **4.13.** Based on these experiments, Tri-DAP was considered unsuitable to be used as a ligand for cell stimulation. Therefore, Tri-DAP was not used as a ligand in the subsequent experiments.

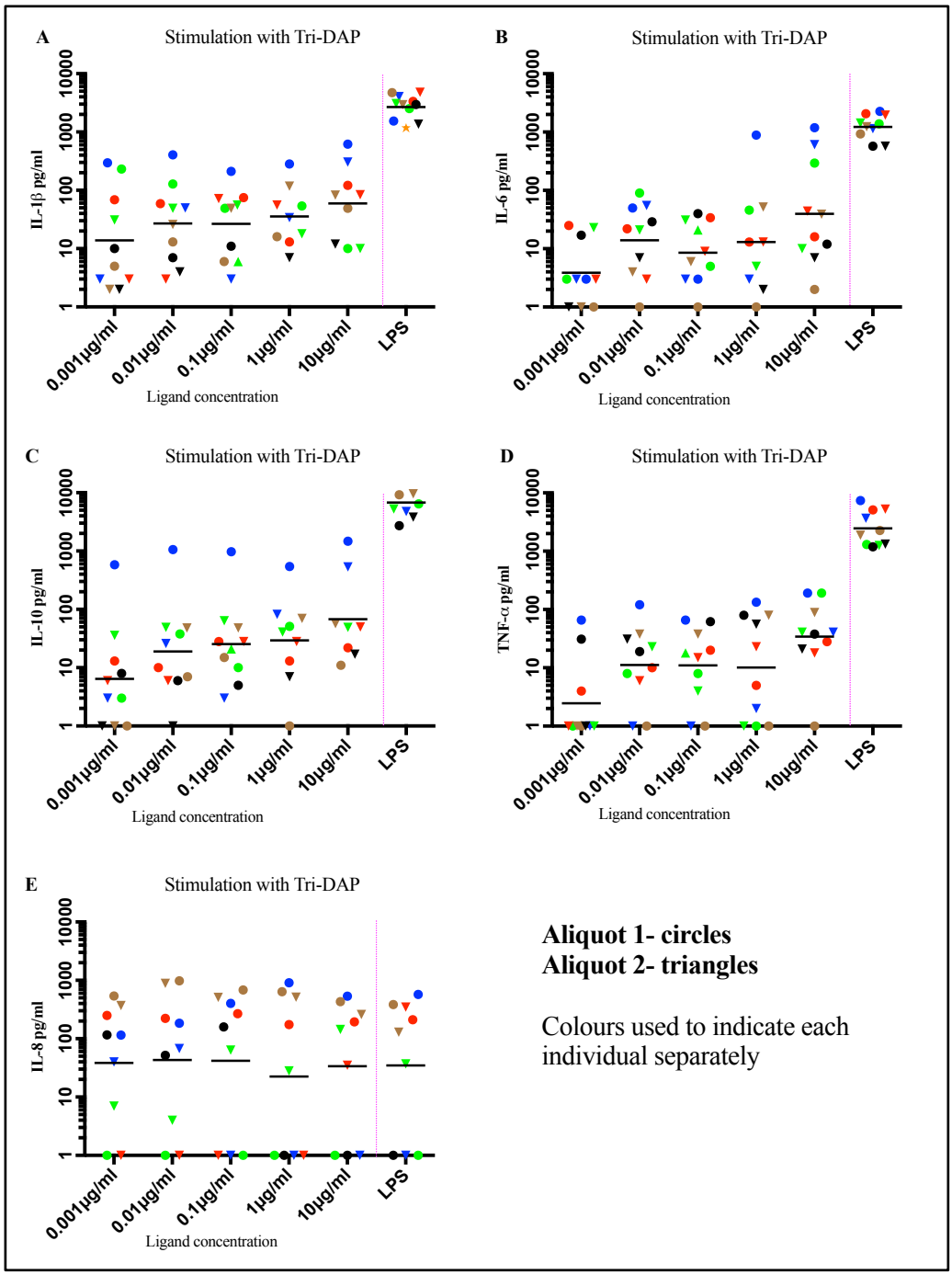


Figure 4.13 Tri-DAP induced cytokine responses

Tri-DAP-induced cytokine responses with incremental doses of Tri-DAP from five healthy adult individuals, performed on two separate aliquots of frozen PBMCs are shown in this figure.

Cytokine responses induced by Tri-DAP are plotted on the left Y-axis, LPS-induced responses (positive control) on the right Y-axis, and the incremental doses of Tri-DAP on the X-axis along with LPS. The horizontal bars represent the geometric mean of the cytokine responses. Stimulation with Tri-DAP did not result in any significant cytokine responses apart from one individual indicated in blue.

4.4.3 Experiments using purified monocytes and combination of purified monocytes with TBNK cells

This section describes experiments conducted during the assay development and optimisation, but not applied to the final assay methodology for various reasons as described in the subsequent sections. **In all the experiments using PBMCs, IL-8 responses were not consistent with increments in the ligand doses** due to high background levels of this cytokine. It was felt that one of the contributory factors was cellular stress during PBMC isolation and culture procedures, leading to IL-8 production even before stimulation with ligands. As the experiments conducted on PBMCs were blind to the cell phenotype, it was felt that if the tests were rerun on pure monocytes in sequentially increasing concentrations, it may be possible to reliably assess IL-8 responses. The presence of other cell types such as T cells, B cells and natural killer cells (TBNK cells), probably add to the background noise in terms of baseline IL-8 production. With this in mind, experiments were conducted as proof of concept on pure monocytes in a dose-dependent manner in order to determine the number of monocytes required to induce an adequate IL-8 response. The primary objective of this experiment was to assess if IL-8 responses could be better analysed with fewer cell numbers (due to a high background secretion of IL-8 by the PBMCs) and from a pure monocyte cell population.

Cytokine induction using purified monocytes

A dose-dependent IL-8 induction was observed with a cell population of pure monocytes. All three ligands (MDP, LPS, Pam3CSK4) worked better at a higher monocyte concentration. Figure **4.14**. However, overall cytokine responses (IL-1β, IL-10, IL-6 and TNF-α) were not robust using pure monocytes alone.

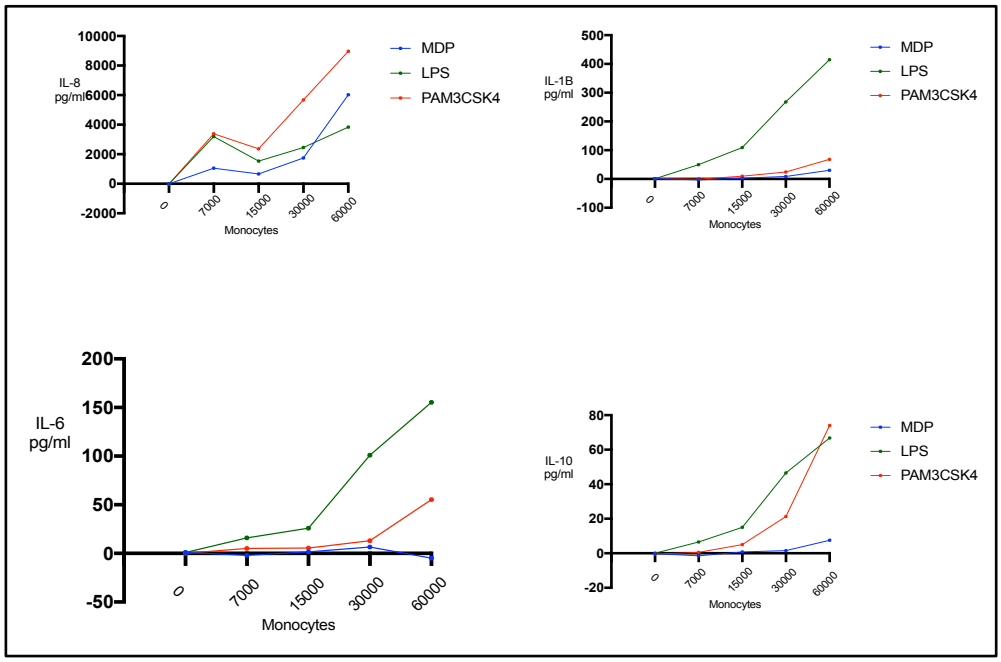


Figure 4.14 Cytokine induction with pure monocytes

Figure illustrates cytokine (IL-8, IL-1β, IL-6 and IL-10) induction with MDP, LPS and Pam3CSK4 stimulation. Optimal response was observed at a concentration of 60000 monocytes per stimulus/condition for IL-8. IL-1β induction was observed with LPS, but not Pam3CSK4 or MDP. Responses were sub-optimal for IL-6 and IL-10.

Cytokine induction using a combination of purified monocytes with TBNK cells

As the induction of IL-6, IL-10 and IL-1β were sub-optimal using pure monocytes alone, further titration experiments were performed using combinations of monocytes and TBNK cells in fixed proportions. Cytokine induction was observed to be dose-dependent with regards to the monocyte numbers. An optimal response was seen at the highest concentration used of 60,000 monocytes per condition for most cytokines except IL-8. Figure **4.15**. IL-8 plots showed inconsistent cytokine secretion profiles with dose-titration, possibly due to high background levels of IL-8 from the TBNK cells.

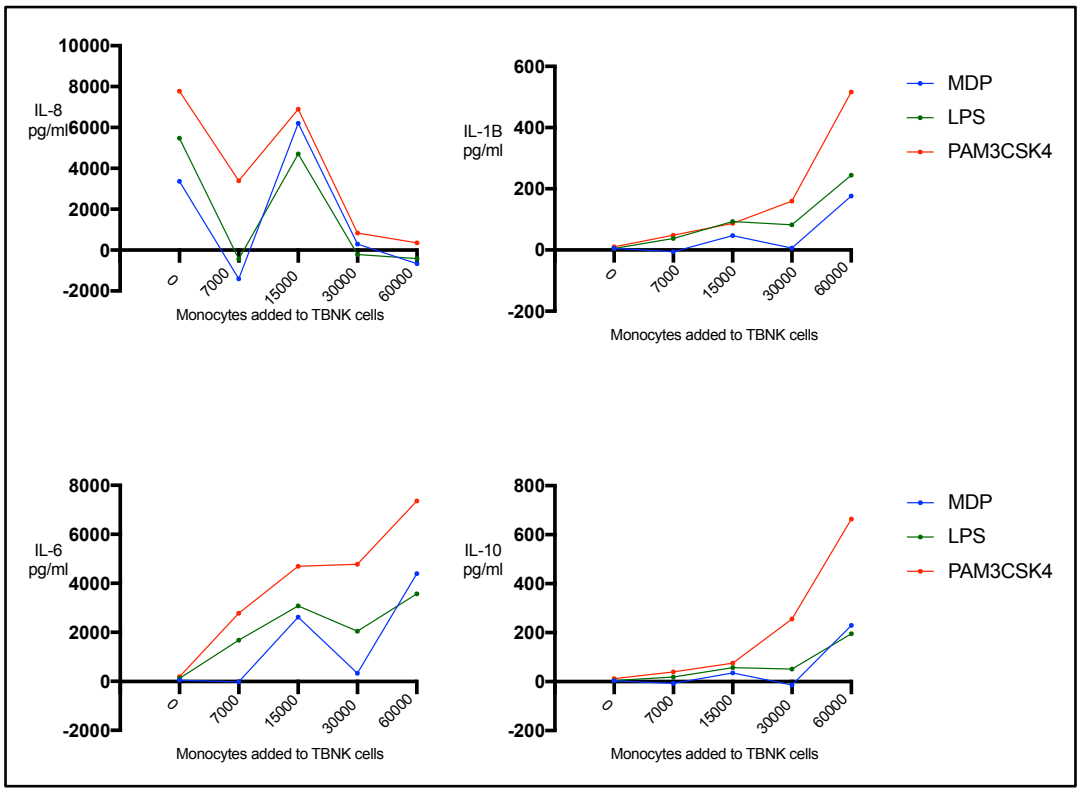


Figure 4.15 Cytokine induction with a combination of purified monocytes and TBNK cells

Figure illustrates cytokine induction with MDP, Pam3CSK4 and LPS using a combination of monocytes and TBNK cells in an enhancing ratio. An optimal response was observed at the highest concentration of monocytes to TBNK cells ratio (1:4).

Interpretation of results using purified monocytes and combination of pure monocytes with TBNK cells

Experiments using monocytes, either purified or in combination with TBNK cells, were primarily conducted to assess induction of IL-8 responses. With pure monocytes, IL-8 responses were observed to be consistent with ligand dose-titration, however the induction of other cytokines was sub-optimal. A combination of monocytes and TBNK cells in an enhanced ratio of 1:4 showed robust cytokine induction for IL-1β, IL-6, TNF-α and IL-10, but not IL-8. At our multi-disciplinary and supervisory meetings, it was felt that conducting parallel and additional experiments using isolated monocytes purely to assess IL-8 responses were significantly time-consuming, labour intensive and more expensive. The method using a combination of monocytes and TBNK cells in an enhanced ratio was felt to be sufficient to assess the integrity of the NF-κB pathway through the analysis of IL-10, IL-1β, IL-6 and TNF-α, and not Il-8. **Therefore, based on practicalities, it was decided to discontinue assessment of IL-8 for future functional assays.**

4.4.4 Application of the assay method using combination of monocytes and TBNK cells on patient specimens

This section provides a synopsis of the results of experiments conducted on patient specimens alongside healthy adult controls using a combination of monocytes and TBNK cells. Although there were important lessons learnt from these experiments, **the findings were excluded from the final analysis due to various reasons, as discussed in the next section.**

A total of fifteen healthy control samples were analysed, median age 27 years (age range 20-40 years). Five out of fifteen healthy controls were excluded based on LPS induced IL-6 response of less than 10pg/ml, which was set as lower limit of detectable induced immune response based on previous experience in the laboratory using LPS as positive control stimulant. Of the thirty-three patients (37 samples in total including 4 specimens repeated post-treatment), 61% (n= 20) were males, 75% (n= 25) had CD, 21% UC (n=7) and one patient had IBDU. Fifteen patient samples out of the thirty-seven were excluded based on LPS-induced IL-6 response of less than 10pg/ml as for the control group. In the group of patient samples with no detectable immune response (based on LPS response), approximately 80% were on drug treatment for IBD whereas only 50% of patients who were not on drugs failed to demonstrate an induced immune response. Failure rate for samples taken whilst on azathioprine or biologics was significantly higher compared to samples taken on no drugs or other drugs (65% vs 20%, p= 0.0084). Three patients on glucocorticoids showed a reduced overall cytokine response and a null response to NOD1 and NOD2 receptor stimulation.

The upper and lower limits of the 95% confidence intervals on the induced responses observed in controls were applied as estimated range of normal immune variance for the patient samples. Based on the cytokine responses falling outside the reference ranges, individuals with distinct inflammatory signatures were identified. Two individuals PR0232 and SOPR0303 showed normal responses to LPS and Pam3CSK4, with null response to MDP, indicative of a probable defect in the NOD2 signalling pathway. Proband PR0136 showed a pattern of lack of response predominantly to Pam3CSK4 with normal responses to other stimuli, suggestive of defective TLR1/2 signalling. Proband PR0208 showed poor IL-1β and TNF-α production, due to probable defects in inflammasome activation and/or TNF-α signalling. Proband SOPR0336 showed a pan-hyperinflammatory profile across most stimuli for IL-6 and IL-1β, suggestive of immune signalling defects across several

pathways or possibly a defect in IL-10 signalling due to loss of IL-10 mediated anti-inflammatory effects.

Further details of the results of these experiments are included in Appendix **C.**

4.4.5 Non-inclusion of the results of experiments using combination of monocytes: TBNK cells

The assay method was conducted and optimised using mononuclear cells from a single healthy individual. The repeatability of the method was not robustly tested prior to using patient specimens. When the assay was conducted on patient samples alongside healthy controls, there was failure of induction of immune responses to the positive control stimulant (LPS) in 33% controls (n=5) and 40% of patients (n=15), leading to exclusion of these samples from the analysis. Furthermore, the control samples used for comparison with paediatric patients were obtained from healthy adult controls. As the normal reference ranges of induced immune responses in patients were based on the responses observed in healthy controls, the results were subject to the naturally occurring age-related variability and therefore deemed not meeting the required criteria to draw robust conclusions. Furthermore, the failure rate of immune induction in controls using the monocytes: TBNK cells combination method did not conform to the required laboratory standards to continue further processing of precious patient specimens. At the supervisory meetings, revision of methodology was suggested and a re-consideration of the method using PBMCs proposed. The method using a combination of monocytes: TBNK cells had no advantage over the PBMCs method. IL-8 could not be reliably measured by either of the methods. Moreover, the monocytes: TBNK method was considerably labour-intensive, time consuming and relatively expensive. Based on all these considerations, it was felt that the optimised PBMCs method described in section **4.4.2**, which effectively assessed induced responses for IL-10, IL-6, IL-1β and TNF-α with consistent dose-response curves was sufficient to assess the integrity of the NF-κB and other signalling pathways. IL-8 was excluded from further analysis. **Finally, the cryopreserved PBMCs method was re-visited and applied as SOP for future patient specimens.**

4.5 CONCLUSIONS AND DISCUSSION

4.5.1 Cytokine responses between cryopreserved PBMCs, freshly extracted PBMCs and whole blood

Although some of the cytokine responses in these early experiments were not consistent with the dose titration, all three methods demonstrated sufficient cytokine induction following stimulation with appropriate concentration of the ligands. This meant that any of the three methods, whole blood, freshly cultured PBMCs or cryopreserved PBMCs could be used to profile cytokine responses. Within my project design, as with many immunological assays, it was necessary to perform batched analysis of cell cytokine responses on stored samples for logistic considerations and to reduce inter-assay variations, particularly when specimens are collected longitudinally over a period of time from the same or different subjects. Furthermore, aliquots of stored specimens can be used for complimentary or additional analyses at later times, as new hypotheses arise or novel techniques become available[401]. The observations from these early experiments provided supportive evidence to the use of cryopreserved PBMCs as an appropriate method for assessing induced immune responses. Therefore, based on these findings, the requirements of my project design, practical advantages and an extensively published literature favouring their use[251-253,262,383,401-407]**, cryopreserved PBMCs were chosen as a preferred method for future experiments.**

Each cell culture system has its own advantages and drawbacks. Cryopreserved PBMCs are widely used for immunological assays as this method minimises operator-dependent and inter-laboratory variability. This is particularly recommended for multi-centre clinical trials with local sample collection[408-410]. Cryopreserved PBMCs offer the advantage of performing batched analysis in a planned and controlled manner. There is also the benefit of incorporating additional analyses and tests on precious patient specimens as may be required following multi-disciplinary team input and emerging evidence from subsequently published literature. Although some studies have indicated preference for using freshly extracted PBMCs, the overall practicalities favour the use of cryopreserved PBMCs for immunological assays. One of the advantages of using freshly extracted as reported by some studies include a higher number of viable cells than cryopreserved PBMCs[411,412]. The potential impact of cryopreservation on the cell function and phenotype has been highlighted in some studies[413-415]. In a study conducted in children with type 1 diabetes mellitus, frozen PBMCs were observed to secrete higher levels of some cytokines (IL-6,

IFN-γ, IL-10, IL-12, IL-13 and monocyte chemo-attractant protein) spontaneously compared to fresh PBMCs, using a multiplex ELISA technique. This was particularly true for IFN-γ, the spontaneous secretion of which was undetectable in fresh PBMCs compared to the frozen PBMCs[402]. This observation was in line with findings from another study, where basal IFN-γ secretion was much higher in frozen PBMCs compared to fresh PBMCs, although the net antigen-specific response remained comparable[416]. The overall conclusions based on these findings were that although there are differences in the basal secretions of some cytokines between frozen and fresh PBMCs, the net antigen-induced responses are comparable. Some studies have however reported significantly increased cytokine responses with frozen PBMCs[401,402,416], which is in keeping with our findings.

Compared to whole blood assays, the potential drawbacks for PBMC culture methods include relatively low sensitivity with respect to some cytokines and a limited ability to extrapolate to the *in vivo* immune setting due to the artificially purified condition in the PBMC culture system[417 418,419]. Also there is a considerable cell loss during isolation[420]. Whole blood assays have several advantages over conventional PBMC assays; they require a much smaller volume of blood, there is no cell extraction process involved and is much quicker. The primary disadvantage of whole blood culture is that the number of cells cultured is usually unknown. On the other hand, purified PBMC cultures include a precisely known number of cytokine-producing cells[420]. The disadvantage of using freshly cultured PBMCs is that the cell culture activation should be performed immediately after PBMC extraction, which is not ideal for planned experiments. For cryopreserved PBMCs, extracted cells can be stored for future analysis and a level of reproducibility can be expected. In a study comparing the performance of PBMCs *versus* whole blood T cell assays using intracellular staining after a 6-hour challenge with cytomegalovirus (CMV), undiluted whole blood consistently yielded higher T cell frequencies than purified PBMCs[421]. However, a subsequent study using similar conditions demonstrated consistently higher cytokine responses in PBMCs than in whole blood. Responses to stimulation were observed at lower concentrations of ligand doses with PBMCs. Furthermore, responses to certain stimuli were undetectable using whole blood when compared with PBMCs[403].

4.5.2 Dose-titration experiments using cryo-preserved PBMCs

The primary objective of these experiments was to assess the optimal ligand concentration needed for cell stimulation. Optimal cytokine responses following cell activation with

MDP and Pam3CSK4 were detected at a concentration of 10μg/ml. Based on these findings, **a concentration of 10μg/ml for MDP and Pam3CSK4 was chosen as the optimal concentration to be applied for all future experiments.** This was consistent with observations from other studies[124,262,266,383]. Optimal ligand concentration for LPS (1 μg/ml) has been established through previous work in our laboratory and published literature[365,369,383,384,422]. LPS was used as a positive control ligand for all cytokine responses. Stimulation of cells with LPS showed cytokine production peaks consistently higher than the other stimuli (MDP, Tri-DAP and Pam3CSK4). Similarly, Pam3CSK4 induction was significantly higher compared to MDP or Tri-DAP induction. This may be due to the intra-cellular location of the NOD1/2 receptors as compared to the TLR receptors which are cell-surface receptors and more accessible for stimulation. Tri-DAP responses were generally poor with no significant induction from the background levels. **As stimulation with Tri-DAP showed poor induction, it was excluded as a stimulant for future experiments.**

Although robust induction levels were detected for cytokines IL-1β, IL-6, IL-10 and TNF-α, IL-8 cytokine responses could not be reliably interpreted due to a high background levels of this cytokine with the PBMC culture method. IL-8 responses were not consistent with dose-titration across all the ligands. **IL-8 was therefore deemed unsuitable for analysis using PBMCs culture method.**

4.5.3 IL-8 assessment using monocytes and application of the method to patient specimens

In order to optimise IL-8 detection, a method using incremental numbers of pure monocytes was tested, through which reliable levels of IL-8 were detected, however the assay had low sensitivity to other cytokines. To overcome the poor detection of the other cytokines including IL-10, IL-1β, IL-6 and TNF-α, an additional method using an enhanced ratio of monocytes and TBNK cells was subsequently assessed. In this method, IL-8 responses were inconsistent as with the PBMCs method, although the induction for cytokines IL-10, IL-1β, IL-6 and TNF-α was sufficiently robust. Based on this observation, patient specimens were analysed using the monocyte: TBNK combination method alongside healthy adult controls, which were used to assess reference ranges for induced immune responses. A high failure rate of cytokine induction was observed for both patient and control specimens. As the analysis of patient samples was based on the immune variance observed in healthy controls, a failure rate of 33% in the control specimens was

not consistent with the required standards of laboratory practice to continue processing precious patient samples using this method. The experiments to test this method were conducted on a single healthy individual. Therefore, the repeatability of the method was not sufficiently assessed prior to application to patient samples. Furthermore, it was argued that the findings would be reliable if age-matched controls were not used to compare immune responses observed in paediatric patient specimens. Also, the monocyte methods in general were significantly time-consuming, labour-intensive and expensive compared to unsorted PBMCs method. **Based on all these reasons, the findings on paediatric patient specimens using the monocyte: TBNK combination method were deemed unsuitable to draw robust conclusions and were excluded from the final analysis.**

4.5.4 Learning points from failed experiments using monocytes: TBNK cells

It took considerably long, up to eighteen months whole-time equivalent, to optimise assay methods in healthy controls so as to develop a standard operating procedure for conducting experiments on patient specimens. The first year of my PhD was as a full-time research fellow and the subsequent four years as a part-time PhD student, having taken up a job as a part-time consultant at the hospital, with constant pressures of clinical commitments on my time. It is possible that the time-pressures were a driving factor to conduct experiments on patient samples rather prematurely, without having robustly assessed for repeatability. The experiments using monocytes were tested on a single individual before applying the method to precious patient specimens, with poor or nil immune induction responses in a substantial proportion of patients and controls. For the reasons discussed in the previous section, we were obliged to set aside the results of these experiments from the final analysis discussed in this thesis.

There were many lessons to be learnt from experiencing failures with experiments during my research, providing opportunities for careful introspection, reflection and understanding the key tenets of scientific research. One of the most important principles that I learnt particularly during the optimisation phase of my project and following constructive feedback at the transfer thesis viva, was that experiments should be repeatable based on the methods described and reproducible when repeated, in order to ensure reliability of the study findings. Without a robust and a reproducible design, the science directing research cannot be reliably analysed with respect to the wider applicability, legacy and impact of the primary research findings[423]. Based on this understanding, when

the methods using PBMCs were revisited, further experiments were carefully planned and repeated in a timely manner under the same conditions to assess for reproducibility.

The patient samples used for the failed monocytes-based experiments were procured following diagnosis, i.e. after or during the course of treatment with medical interventions including immunosuppressive therapy. We learnt from our failed experiments that samples obtained from patients on immunosuppressive therapy are more likely to show poor or nil immune induction compared to individuals without treatment. Based on the weight of combined evidence from our own experience and previously published literature highlighting the masking effects of drugs on the functioning of immunological assays[424,425], it was subsequently decided to recruit patients before or at diagnosis, so as to be able to procure treatment naïve specimens for all future experiments. Furthermore, in our failed experiments, immune induction observed in healthy adult controls was used to derive reference ranges for immune induction in paediatric patient specimens. Although there is literature-based evidence to suggest that induced immune function in toll-like receptors remains stable over several decades (birth to 57 years), less is known about the age-related variation of other innate immune receptors[385,426]. Therefore, in order to minimise the potential variability in immune function due to these factors, it was decided to use age-matched controls to compare immune responses in paediatric patients for all subsequent experiments.

4.5.5 Developing a standard operating procedure (SOP) using the cryopreserved PBMCs method

Based on the challenges experienced in IL-8 quantification and the reasoning that IL-8 though complimentary, is not critical to the assessment of the innate inflammatory pathways, it was decided to exclude IL-8 from analysis in all subsequent experiments. Given the practical advantages in using PBMCs for cytokine quantification including IL-10, IL-1β, IL-6 and TNF-α and the consistent dose-responses observed during the optimisation experiments, **the cryopreserved PBMCs method with the optimised ligand doses was selected as the SOP** and subsequently applied to all experiments assessing innate immune induction on patient specimens.

Based on the overall results of the optimisation experiments, plan for future assay experiments using patient samples was as follows:

- **Culture Method:** Cryopreserved PBMCs method will be applied for all future assays within the framework of the research project.

- **Concentration of PBMCs:** A concentration of 0.3 million cells per condition was chosen for all experiments evaluating induced immune responses.
- **Ligand Concentration:** A ligand panel with the following concentrations will be applied for cell activation: MDP- 10μg/ml, Pam3CSK4- 10μg/ml and LPS- 1μg/ml. Tri-DAP will not be used a ligand for future assays.
- **Incubation period-** Cell suspensions will be incubated for 24 hours following activation with the specified ligands.
- **Cytokines to be analysed**: These will include IL-1β, IL-6, IL-10 and TNF-α. IL-8 analysis will not be included in the analysis for future experiments.

4.6 SUMMARY

In this chapter, the assay methods and experiments described were optimised in healthy controls to develop a standard operating procedure for future application on patient samples. The experiments were performed in order to establish the best-fit cell culture model, the optimal ligands, ligand concentration and the appropriate cytokine read-outs to assess the functional integrity of the NF-κB pathway and other cross-linking inflammatory pathways. Although the results of the experiments conducted using monocytes on patient specimens were limited in extrapolation, the productive outcome of the iterative phase of development was to fine tune to the lessons learnt and optimise an experimental set-up for future assays. The experimental work resulted in the optimisation of an assay method using frozen PBMCs for measuring four cytokines against three stimuli and utilising treatment naïve patient specimens to eliminate the masking impact of drugs for all future assays.

Chapter 5 IMMUNO-GENOMIC PROFILING IN A TREATMENT NAÏVE PAEDIATRIC COHORT

This chapter describes the application of optimised immunological assays on a treatment naïve paediatric IBD cohort alongside paediatric and adult healthy controls, followed by immune-profile based patient stratification using unsupervised machine learning and interrogation of stratified groups against whole exome sequencing data (WES). The immunological assays were carried out by myself under the supervision of Dr Yifang Gao and Prof Anthony Williams. Machine learning-based hierarchical clustering of the cytokine data and the analysis of whole exome sequencing data were carried out with the help of Dr Enrico Mossotto, co-PhD student from the Genomics group. DNA extraction for WES was carried out by Nikki Graham, an experienced technician from the DNA laboratory based at the Duthie building.

5.1 INTRODUCTION

The primary aim of this study was **to assess the functional integrity of innate signalling pathways of known importance in IBD including NOD2, toll-like receptor (TLR)-1-2 and TLR4-mediated signalling** by conducting immunological assays in a treatment naïve paediatric IBD cohort. Biological variations in patients were assessed against induced immune responses in healthy controls. Based on the understanding that individual patients may present with compromised functionality of specific signalling pathways, our secondary objective was **to stratify patients based on their immune profiles** using unsupervised machine learning and identify inflammatory signatures specific to individual patients or groups mapping on to known inflammatory pathways. Our final aim was **to interrogate immune-profile based patient stratification against genomic data**, to complement and wherever possible confirm specific pathway dysfunction. See figure **5.1.**

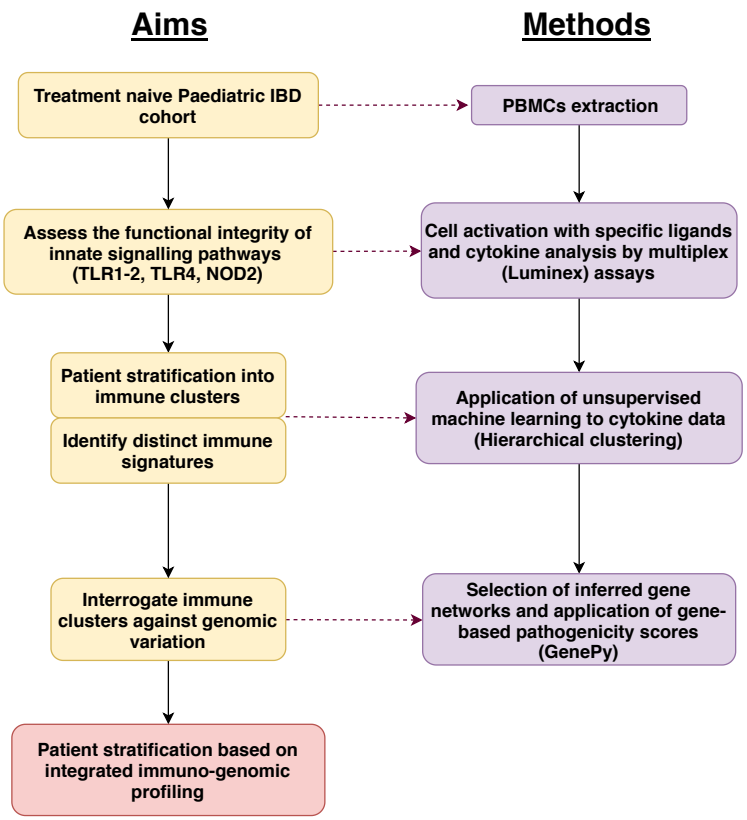


Figure 5.1 Outline of the project pathway

The figure shows the key elements of the project pathway including aims on the left and methods on the right. The functional integrity of key innate signalling pathways including toll-like receptor 1-2 (TLR1-2), TLR4 and NOD2 pathways were assessed by stimulating PBMCs from a treatment naïve paediatric IBD cohort with specific ligands and conducting immunological assays. The cytokine data generated was subjected to hierarchical clustering to stratify patients into immune clusters, followed by interrogation of the immunological findings against genomic variation, using a gene-based pathogenicity score 'GenePy'.

Immune responses depend on the functional integrity of proteins of the immune signalling pathways which mediate induced transcriptional responses. Over the last two decades, a multitude of genes have been identified as potential susceptibility genes contributing to the pathogenesis of IBD[60,98]. These include genes encoding the key immune signalling pathways, with mutations leading to aberrant responses in patients with IBD. Hyper-inflammatory responses have been described causing excessive intestinal inflammation and a predisposition to IBD[199,427]. Hypo-inflammatory immune responses following microbial breach of the intestinal epithelium have also been described with a consequent poor clearance of the pathogens and persistent inflammation leading to IBD[184]. The other mechanisms altering immune responses in IBD include the disruption of the intestinal

mucus and epithelial barrier due to genetic mutations and the complex interactions of the host immune system with the microbiota and the environment[4,65,133,134,209,211,428].

In this study, treatment naïve specimens were necessary to eliminate the potential impact of drugs on induced immune responses. Previously published studies have clearly demonstrated the effect of drug treatments on the functioning of immunological assays[424,425,429,430]. The key focus of my project was to assess the functional integrity of innate signalling pathways implicated in the pathogenesis of IBD including innate pathogen recognition receptors (NOD2, TLR1-2 and TLR4) and downstream signalling cascades such as the NF-κB, MAPK and NLRP3-inflammasome[4,60 365 174]. These cross-talking signalling pathways mediate and orchestrate inflammation through complex mechanisms, resulting in the transcription of cytokines including IL-10, IL-6, IL-1β and TNF-α, which can be quantified through multiplex assays[358,359]. The analysis of the effector cytokines induced following activation of specific receptors indicates the functionality of proteins involved along the chain of the signalling pathways, from the stimulated receptor to the specific cytokine produced[358,359]. Details of the optimised assay applied to this study are described in chapter 4.

Following the immunological assays, the cytokine data generated were subjected to unsupervised machine learning approaches to assess for patient-stratification based on their immune profiles and identify compromised functionality of signalling pathways in individuals or groups. Machine learning algorithms identify patterns within the data and use them to make predictions or infer new knowledge without being explicitly programmed to perform a defined task[293]. Machine learning methods can broadly be classified into supervised and unsupervised learning algorithms, based on the desired output and the types of input available for the training set. In supervised learning, the input consists of a set of training example with known labels. The algorithm analyses the training data and produces an inferred function, which can be used for classifying new samples. On the other hand, unsupervised algorithms utilise unlabelled samples without pre-defined classes, aiming to find patterns or clusters within the data. After unsupervised learning is used to identify classes, a supervised learning approach may be applied to classify new samples based on the programme built into the algorithm[431]. In this study, cytokine data were modelled using unsupervised machine learning approaches including principal component analysis (PCA), hierarchical clustering (HC) and data visualisation using radar plots.

PCA is an unsupervised machine learning method commonly used to simplify the complexity in high dimensional data, while retaining trends and patterns. It does this by transforming the data into fewer dimensions, which act as summaries of features without reference to prior knowledge about the sample features. High-dimensional data are very common in biological sciences and arise when multiple variables are measured for each sample. PCA reduces data by geometrically projecting them onto lower dimensions called principal components (PCs)[432,433]. The multiple number of variables is reduced to a smaller number of interpretable linear combinations of data and each linear combination is displayed as a geometrically orthogonal PC. The first PC (PC1) represents a linear combination of variables that has the maximum variance among all linear combinations, so that it accounts for as much variation in the data as possible and each succeeding component accounts for as much of the remaining variation possible[434].

Hierarchical clustering (HC) is a well-researched application of unsupervised machine learning, which seeks to identify natural groupings within a complex data set by reducing variable dimensionality and clustering data points with similar patterns. HC organises data into clusters such that there is high intra-cluster similarity and a low inter-cluster similarity. It relies on the core principle of objects being more related to nearby objects than to objects farther away. The algorithm connects objects to form ‘clusters’ based on their distance which is proportional to the degree of dissimilarity between any two given points. The bottom-up approach of hierarchical clustering, also known as agglomerative clustering, starts with each item in its own cluster finding the best pair to merge into a new cluster. This is repeated until all clusters are fused together. At different distances, different clusters will form, which can be represented using a ‘dendogram’. In a dendogram, the Y-axis marks the distance at which the clusters merge, while the objects or individuals are placed along the X-axis such that the clusters don’t overlap.

Machine learning-led stratification was followed by interrogation of whole exome sequencing (WES) data against the immune-profile based clusters of patients. This was performed to identify genomic signals contributing to the clustering pattern that could explain specific pathway defects and identify mechanisms triggering IBD in individual patients or groups. WES data were transformed into a newly developed gene-based deleteriousness metric, ‘GenePy’ score, which was used to estimate the cumulative burden of deleterious mutations across inferred gene networks based on the immunological findings. GenePy transforms WES data into whole gene-based pathogenicity scores for each individual, by combining the biological information on known deleteriousness

metrics, allele frequency and zygosity[435]. GenePy scores for genes can be summed across defined molecular pathways or networks to give a total pathogenicity score in individual patients and can also facilitate comparison between groups of individuals[435].

5.2 METHODS

5.2.1 Ethical approval

Protocols were approved by Southampton & South West Hampshire Research Ethics Committee (09/H0504/125). Informed consent was obtained from all participants and/or their parents/legal guardians before recruitment to the study.

5.2.2 Patients and controls

Children with suspected inflammatory bowel disease (IBD) were recruited prior to an established diagnosis following referral to the paediatric gastroenterology service. All children under 18 years of age, referred to the service as suspected IBD were eligible for inclusion. At referral, these patients had GI symptoms such as abdominal pain, diarrhoea, blood in stools, weight loss, lethargy, poor appetite and/or other features to suggest IBD including elevated inflammatory markers on blood tests (C-reactive protein-CRP, erythrocyte sedimentation rate- ESR) and stool tests (faecal calprotectin). Patients were recruited over a 12-month period from January 2016 to January 2017. Diagnosis of IBD was made in line with the Porto criteria[436,437]. Disease activity at diagnosis was assessed using internationally validated disease scoring tools; Paediatric Crohn's Disease Activity Index (PCDAI) for Crohn's disease (CD)[438] and Paediatric Ulcerative Colitis Activity Index (PUCAI)[439]. Blood samples for PBMC extraction were obtained prior to commencement of any treatment for IBD. Following sample procurement, all recruited children underwent endoscopic assessment and histological examination of the gastro-intestinal biopsies as part of the routine diagnostic work-up. Blood samples for DNA extraction were obtained either at diagnosis along with the samples for PBMC extraction or subsequently during the course of treatment.

Of the thirty-two children recruited as suspected IBD during the specified period, twenty-two had a subsequent diagnosis of IBD following endoscopic and histological investigations. The remaining ten children who did not have a diagnosis of IBD remained well over a six-month follow up period and were subsequently discharged from the

paediatric gastroenterology service. Biological specimens obtained from these ten children were used as paediatric controls.

Control samples for PBMC extraction were also obtained from ten healthy adult individuals, all aged over 18 years. Adult controls were included due to the practical difficulties in procurement of samples from paediatric healthy controls in a clinical setting. A further justification for the use of adult healthy controls was based on evidence from studies suggesting that the immune function remains stable across several decades between infancy and old age[385,440,441].

5.2.3 Immunological methods

5.2.3.1 PBMCs extraction

Blood samples up to 15 mls from paediatric subjects and 40 mls from adult controls were obtained for PBMCs extraction through a one-time venesection per individual. PBMCs were isolated from heparinized whole blood by density gradient centrifugation (Ficoll-Paque Plus) as described in chapter 3 and the appendices. The isolated PBMCs were stored in cryovials in liquid nitrogen until further analysis.

5.2.3.2 Cell activation

Frozen PBMCs were thawed and suspended in R10 medium, consisting of RPMI 1640 (Sigma-Aldrich) medium supplemented with 10% heat inactivated fetal calf serum, 1% L-glutamine, 1% sodium pyruvate and 1% non-essential amino acids. Cell suspensions in R10 were dispensed into 15 mls Falcon tubes at a concentration of 0.3 million cells per tube in 300µl. Cell activation was performed using muramyl- dipeptide (MDP, NOD2 agonist), Pam3CysSerLys4- a synthetic tri-palmitoylated lipopeptide (Pam3CSK4, TLR1-2 agonist) and lipopolysaccharide (LPS, TLR4 agonist) employing ligand doses at10µg/ml for MDP & Pam3CSK4, and 1µg/ml for LPS (Pam3CSK4 and MDP purchased from InvivoGen; LPS from Sigma-Aldrich). Ligand doses were based on the dose-optimisation experiments described in chapter 4. Cell activation was conducted on the first aliquot of PBMCs from each individual followed a week later by activation of the second aliquot using identical assay conditions in order to ensure repeatability. All assays were conducted in duplicate each time. The cells were incubated at 37 °C in a 5% CO_2 atmosphere for 24 hours before supernatants were harvested. The supernatants were stored at -20°C until a batched analysis with Luminex® assays to quantify the cytokines from the supernatants.

5.2.3.3 Analysis of cytokine production by multiplex (Luminex®) assay

Luminex® assays were used to simultaneously measure the concentrations of four cytokines including IL-10, IL-6, IL-1β and TNF-α. The assay was performed in a 96 well plate format and analysed with a Luminex 100^{TM} analyser. A combination of single-plex kits from Invitrogen Procartaplex was employed for detection and quantitation of the four cytokines. All stimuli were assayed in duplicate. For every individual and condition, cytokine readings from stimulated PBMCs were derived after subtraction of the baseline secretion values. **As there were three stimulating ligands employed and four effector cytokines analysed, a total of twelve conditions were assessed per individual.** Monocyte counts were determined in each of the forty-two individuals (twenty-two pIBD patients, ten paediatric controls and ten adult controls) through a flow-cytometry analyser. As described in chapter 4, based on the reported normal range of monocyte counts within a healthy population[393,394], cytokine effector responses were standardised to a monocyte count of 10% in order to avoid inconsistencies due to variable monocyte counts.

5.2.3.4 Establishing a reference range for cytokine read-outs

Cytokine responses in healthy controls were plotted against the respective ligands after subtracting from baseline secretion. The mean and standard deviation (SD) were calculated for every cytokine effector response. In order to interpret the range of normal immunological variance on patient specimens, the upper and lower limits of 2SD of the mean of control samples were applied to define a reference range. A 'hypo-inflammatory' immune response was defined if the measured level of cytokine was >2SD lower than the mean and 'hyper-inflammatory' response if >2SD above the mean.

5.2.3.5 Statistical tests for immunological assays and clinical parameters

Statistical analysis was performed using the GraphPad Prism software, version 7. Cytokine induction between the two aliquots of PBMCs per patient were compared using paired two-tailed t-tests. As all assays were conducted in duplicate and the mean of the cytokine values was used for the t-tests. Cytokine induction in paediatric patients with IBD, paediatric healthy controls and adult healthy controls were compared using unpaired t-tests. Differences were considered significant if the p value was <0.05. Regression analysis was used to assess the statistical significance between disease activity and laboratory parameters.

5.2.4 Application of machine learning methods

Normalisation of raw data is a common requirement for machine learning applications as these programmes are designed on the assumption that the data values vary on comparable scales. Presence of outliers can affect the objective and predictive performance of many machine learning algorithms. As discussed in the following sections, normalisation of raw cytokine data in this study was performed using StandardScaler and RobustScaler within the python Scikit-Learn package[442].

5.2.4.1 Principal Component Analysis (PCA) to assess for batch effects

In this study, multiplex assays on patient samples and controls were performed over a three-month period across five different batches. Principal Component Analysis (PCA) was used to assess for batch effects, by converting cytokine variables into principal components. Data were normalised for PCA using StandardScaler[442], which removes the mean and scales the data to unit variance. Each raw cytokine value (X_i) expressed in picograms/ml was scaled to a normalised value (X) by subtracting the mean (μ) from the raw value and dividing by the standard deviation (σ). Both μ and σ were calculated on paediatric controls.

$$X = X_i - \mu_{controls}/\sigma_{controls}$$

5.2.4.2 Visual representation of cytokine patterns using radar plots

To illustrate the patterns of cytokine responses in the pIBD cohort and in individual patients, radar plots were generated using python SciPy and Scikit-Learn v0.19 packages. Cytokine data were normalised using StandardScaler[442] and normalised values per patient per stimulus were plotted along the spokes or vectors of the radar plot, each spoke representing a specific cytokine response. Cytokine data from paediatric controls were used to define the boundaries of the reference range in the pIBD cohort. The upper and lower range of normality of cytokine responses were assigned as +/- 2 SD across the mean of controls. Values within 2 SD represented normal responses and, those above and below 2 SD indicated hyper-inflammatory and hypo-inflammatory responses respectively. See figure **5.2.**

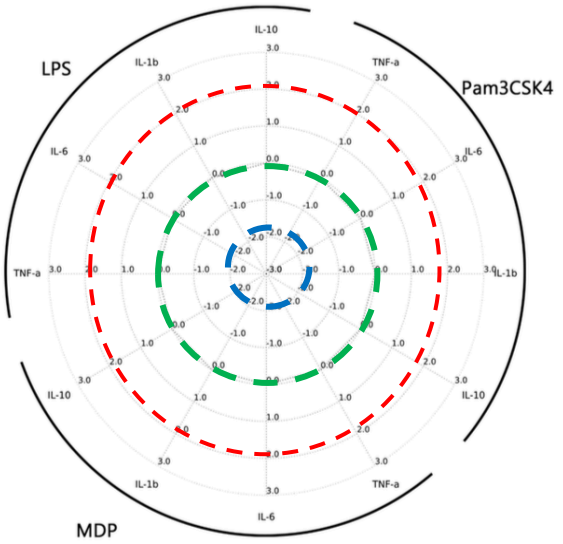


Figure 5.2 Radar plot of cytokine responses

Each sector of the radar plot represents the stimulant and the four spokes or vectors within each sector represent the cytokine responses per stimulant. The plot shows three stimulants (LPS, MDP & Pam3CSK4) and four cytokine responses (IL-10, IL-6, IL-1β and TNF-α) per stimulant. The green represents the mean generated from controls. The red and blue dotted lines represent 2 SD above the mean and 2 SD below the mean respectively.

5.2.4.3 Hierarchical clustering

Hierarchical clustering was applied to the normalised cytokine data in order to assess for patient-stratification based on the patterns of immune induction within the cohort. Raw cytokine data were normalised using RobustScaler. The data transformation and scaling statistics of this tool are based on percentiles and therefore not influenced by few numbers of large marginal outliers. RobustScaler transforms data by subtracting the median from the feature and then dividing by the interquartile range (IQR, range between the 1st and the 3rd quartile). In this study, raw cytokine data was scaled using the median and interquartile ranges calculated on non-IBD controls. Each feature is normalised as follows:

$X = X_i - Q_1 / Q_3 - Q_1$

(X- normalised feature, X_i- raw feature, Q_1 and Q_3 denote 1st and 3rd quartiles respectively calculated on paediatric controls)

5.2.5 Genomic methods

In order to understand the genomic basis of the immune clusters, exome data from the pIBD cohort were interrogated by selecting gene networks representing the innate

signalling pathways, specifically investigated through the immunological assays. WES data were transformed into a newly developed gene-based deleteriousness metric 'GenePy' score[443], which was applied to investigate the mutational burden across a defined set of genes constituting the selected immune pathways.

5.2.5.1 Selection of gene networks

Each gene network representing a specific signalling pathway, included a list of genes which were extracted by conducting a search using PathCards online database of human biological pathways. PathCards generates pathway clusters in the form of 'SuperPaths' consolidating biological information from twelve different manually curated sources, providing a unified set of genes for the interrogated pathway[444]. See figure **5.3.** PathCards is updated regularly, with a new version released two-three times per year. For our study, gene network interrogation was conducted in the first week of February 2018 to its last updated version (Version 4.6). Each SuperPath is identified by a textual name derived from one of its constituent pathways, selected as the most connected pathway or hub within the SuperPath cluster.

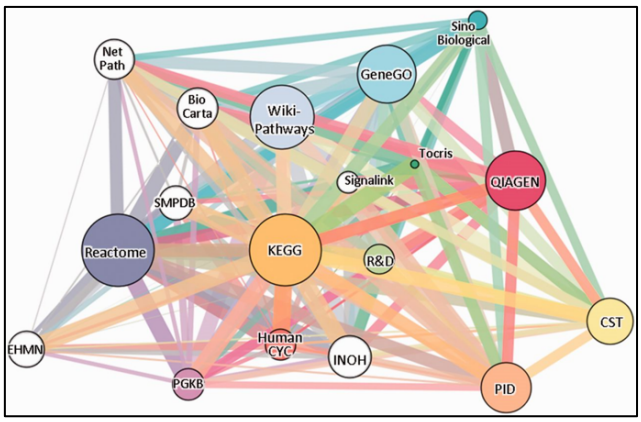


Figure 5.3 Sources of gene networks on PathCards

Eighteen sources providing gene networks for biological pathways on PathCards webserver are shown in the figure, of which 12 (coloured clusters) are included in the generation of SuperPaths. The remaining 6 (white clusters) are used as additional sources for cumulative tallying of the gene content represented in the SuperPaths generated by the 12 sources. Figure reference: Belinky et al, 2015, Database (Oxford)

Pathways were selected for gene interrogation based on the immune response profile across the twelve cytokine responses, each response representing a specific signalling pathway, beginning with the receptor activation, followed by the downstream signalling

cascade and ending with the transcription of specific cytokines. Cytokine responses with a significantly altered collective reading for the patient cohort compared to controls were selected for gene interrogation. Search terms used for gene interrogation included 'TLR1', 'TLR2', 'TLR4' & 'NOD2' for the receptors, 'NF-κB', 'inflammasome' & 'MAPK' for the downstream signalling cascades/adaptors, and 'IL-10', 'IL-6', 'IL-1β' & 'TNF' for gene networks at the cytokine level. Gene names were cross-referenced with the HUGO Gene Nomenclature Committee (HGNC) webserver to confirm the approved gene symbol.

5.2.5.2 GenePy score

GenePy score was used to estimate the cumulative deleteriousness potential within genes selected for interrogation[443]. GenePy scores are corrected for gene length. Scores may not be available for all genes as some of the genes may not be properly captured during the exome library capture process, may not harbour any missense mutations or may not be annotated by the currently available computational tools for assessing the deleteriousness potential of the detected variants. Patients harbouring multiple rare mutations in specific genes may accumulate high scores for those genes while most genes carrying few variants of neutral consequence confer very low scores. For this study, GenePy scores were generated using Combined Annotation Dependent Depletion (CADD) as the deleteriousness metric, which integrates multiple diverse annotations of single nucleotide variants into one metric to infer deleteriousness of mutations[332].

5.3 RESULTS

5.3.1 Patient cohort characteristics

Twenty-two out of the thirty-two treatment naïve children who were consecutively recruited as suspected IBD had an established diagnosis following endoscopic and histological investigations. These included 55% males (n=12); median age of 13 years (age range 5-16 years); Crohn's disease 64% (n=14) and ulcerative colitis 36% (n=8). The mean (+/- SD) PCDAI and PUCAI scores for CD and UC at the time of sample procurement were 37 (+/-12) and 54 (+/-13), reflecting a moderate-severe disease activity in most of the patients. The majority of patients with CD presented with ileo-colonic disease (57%, n=8), ileo-colonic with upper GI disease (14%, n=2) or isolated colonic disease (28%, n=4). In UC, the predominant presenting phenotype was pan-colonic disease. All patients with UC presented with rectal bleeding compared to 30% in CD. Reduced appetite, weight loss and lethargy were predominantly seen in CD compared to UC. See table **5.1** and figure **5.4.**

Table 5.1 Patient cohort characteristics

	Proband	Gender	Age at diagnosis	Diagnosis	PCDAI/ PUCAI	Severity at diagnosis	Paris Classification (disease location)	Abdo pain	Diarrhoea	Lethargy	Nocturnal stools	Reduced appetite	Rectal Bleeding	Weight loss	CRP	Faecal calprotectin(mg/g)
Crohn's Disease																
1	SOPR0304	M	10	CD	30	Mod-sev	L2L4a	✓	✓	✓	✓	✓	✓	×	4	>1800
2	SOPR0336	F	12	CD	45	Mod-sev	L3	✓	✓	✓	×	✓	×	×	67	NA
3	SOPR0339	M	15	CD	32.5	Mod-sev	L2L4a	✓	✓	×	✓	✓	×	✓	11	NA
4	SOPR0342	M	5	CD	55	Mod-sev	L2	✓	✓	×	✓	×	✓	×	78	NA
5	SOPR0345	M	14	CD	45	Mod-sev	L3	✓	✓	×	×	×	×	✓	62	NA
6	SOPR0348	M	8	CD	50	Mod-sev	L3	✓	✓	✓	×	✓	×	✓	65	NA
7	SOPR0351	F	7	CD	12.5	Mild	L2L4a	✓	×	×	×	×	×	×	NA	673
8	SOPR0352	F	16	CD	32.5	Mod-sev	L3	✓	×	✓	×	×	×	✓	25	NA
9	SOPR0359	F	9	CD	22.5	Mild	L2	✓	✓	✓	×	×	×	×	1	2664
10	SOPR0368	M	15	CD	47.5	Mod-sev	L4a	✓	✓	✓	×	×	×	✓	51	3024
11	SOPR0370	M	11	CD	35	Mod-sev	L3L4a	×	✓	×	×	✓	✓	×	47	NA
12	SOPR0377	F	13	CD	32.5	Mod-sev	L3	✓	✓	✓	×	✓	×	✓	12	360
13	SOPR0378	M	9	CD	30	Mod-sev	L2	✓	×	✓	✓	✓	✓	✓	2	NA
14	SOPR0380	F	16	CD	52.5	Mod-sev	L3L4a	✓	✓	×	×	✓	×	✓	87	NA
Ulcerative Colitis																
1	SOPR0340	F	10	UC	45	Mod-sev	E1	✓	✓	×	×	✓	✓	✓	NA	>3000
2	SOPR0346	M	16	UC	45	Mod-sev	E2	✓	✓	×	✓	×	✓	×	NA	NA
3	SOPR0355	F	13	UC	75	Severe	E4	✓	✓	✓	×	×	✓	✓	2	1690
4	SOPR0367	F	13	UC	60	Moderate	E4	✓	✓	×	✓	×	✓	×	7	NA
5	SOPR0369	M	15	UC	40	Moderate	E1	✓	✓	×	×	×	✓	×	1	166
6	SOPR0372	M	14	UC	55	Moderate	E2	✓	✓	✓	×	×	✓	×	2	NA
7	SOPR0381	F	13	UC	45	Moderate	E4	✓	✓	×	×	×	✓	✓	1	NA
8	SOPR0384	M	14	UC	70	Severe	E3	×	✓	×	✓	×	✓	×	2	NA

Patient characteristics are presented along with the PUCAI (mild 10-34, moderate 35-64, severe >65) and PCDAI (mild 10-29, moderate-severe >30) scores for UC and CD respectively. NA- not available. Disease location/s as per Paris classification for Crohn's disease include L1 (distal 1/3 of ileum), L2 (colonic), L3 (ileo-colonic) & L4 (upper GI disease; L4a- proximal to the ligament of Treitz, L4b- distal to the ligament of Treitz); UC E1(ulcerative proctitis), E2 (left-sided UC), E3 (extensive, up to hepatic flexure) & E4 (pancolitis).

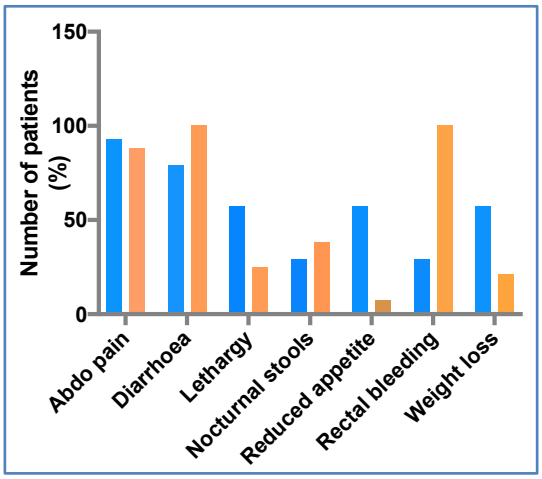


Figure 5.4 Presenting clinical features

The presenting clinical features in the cohort (n=22) are displayed along the X-axis. Patients with Crohn's disease (n=14) and ulcerative colitis (n=8) are represented with blue and orange colour respectively. The percentage of patients with the specified symptoms are indicated along the Y-axis.

Faecal calprotectin, a biomarker of bowel wall inflammation was available in 36% (n=8) of patients. Faecal calprotectin is a calcium and zinc-binding peptide, originating primarily from neutrophils, but also from monocytes. It constitutes approximately 60% of the total cytosolic protein content in neutrophils and is secreted upon activation of neutrophils. Elevated levels of faecal calprotectin indicate IBD-associated bowel wall inflammation due to neutrophilic infiltration, although other gastrointestinal conditions can also cause elevated levels[445,446]. Although there has been a sharp rise in the use of faecal calprotectin for diagnosis and management of paediatric IBD[447], it is not always routinely available to clinicians.

C-Reactive Protein (CRP) measurements were available in 19 patients (13 patients with CD, 6 with UC). Regression analysis between disease activity scores and CRP showed a linear relationship in CD ($p<0.0001$), but no correlation in UC ($p=0.598$). No correlation was observed between the age of disease onset and disease activity in either CD or UC. See Figure **5.5.**

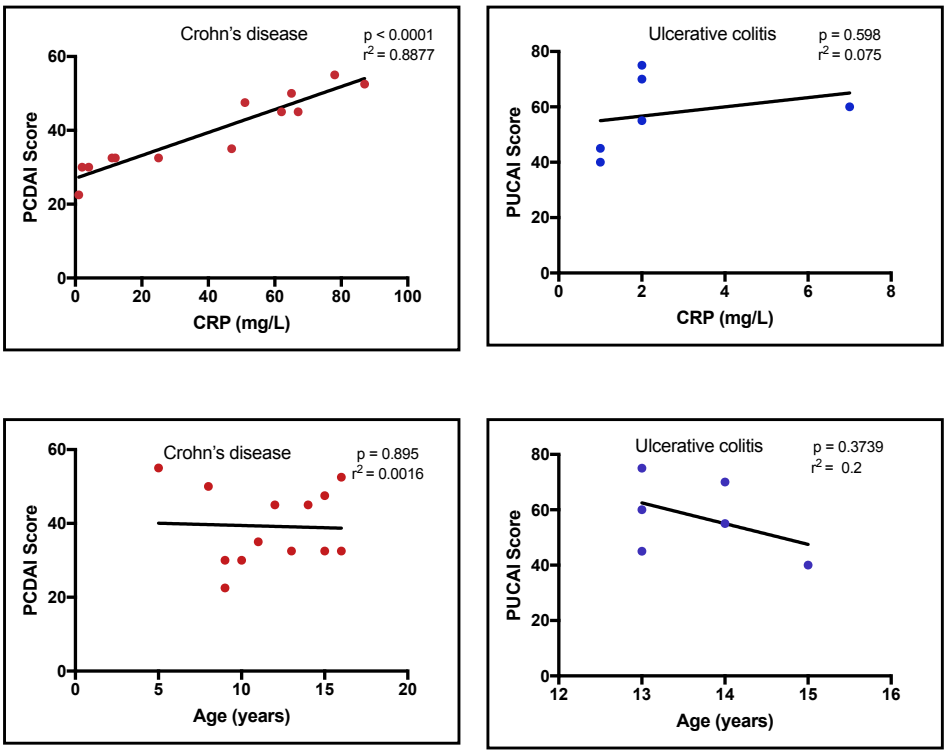


Figure 5.5 The relationship between disease activity scores, CRP and age of onset

The red dots in the diagrams on the left represent Crohn's disease and the blue dots on the right represent ulcerative colitis. Regression line is calculated using least-squares method and goodness of fit expressed as Pearson's product-moment coefficient (r). Disease activity scores are indicated along the Y-axis in all 4 diagrams. A significant association between PCDAI scores in CD at disease onset and CRP was observed.

5.3.2 Paediatric and adult healthy controls characteristics

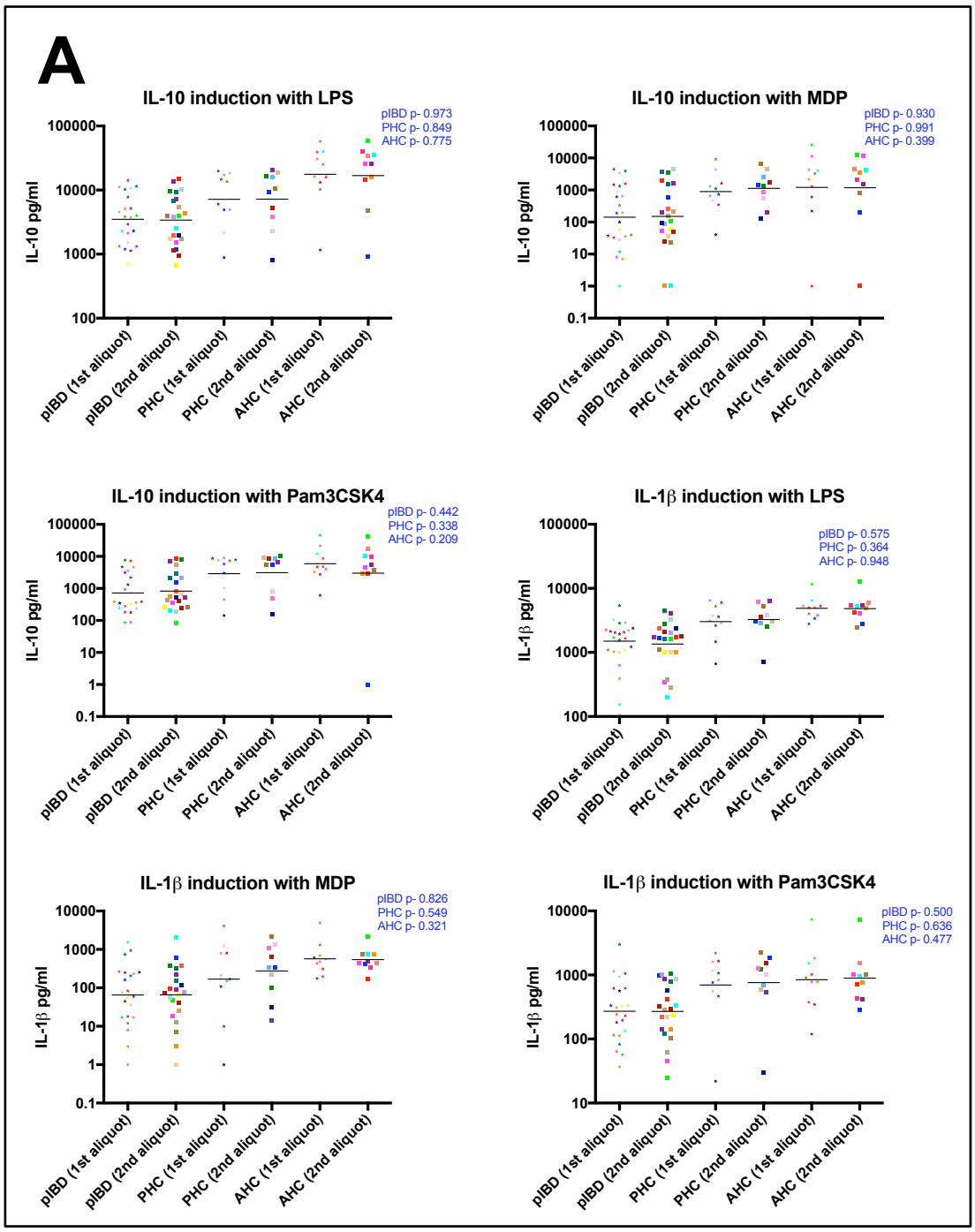
The paediatric control group included ten individuals with a median age of 14 years (age range 4-16 years), of which 60% were males (n=6). These were children who were recruited as suspected IBD, but had normal investigations, remained well over a six month follow up period and were subsequently discharged from the service. The adult control group included 10 individuals, 50% males (n=5) with a median age of 27 years (age range 20-35 years).

5.3.3 Results of the immunological assays

5.3.3.1 Assay results between the first and second aliquot of PBMCs

For all cases and controls, cytokine profiles between the two aliquots of PBMCs per individual were compared to assess for repeatability of the assay under identical conditions. Assays were conducted in duplicate each time and the mean of the duplicate

values compared between the two aliquots. Cytokine production patterns were comparable between the two aliquots across all the three groups (paediatric IBD patients, paediatric healthy controls and adult healthy controls). Paired two-tailed t-tests applied across all four cytokines per stimulus analysed showed no significant differences between the two aliquots indicating robust and repeatable experimental processes. See figure **5.6 (A & B)**. The consistency and repeatability between the aliquots provided the confidence to integrate and combine the results between the two aliquots. Therefore, the mean of the duplicate aliquots per cytokine for every individual were used for all subsequent analyses.



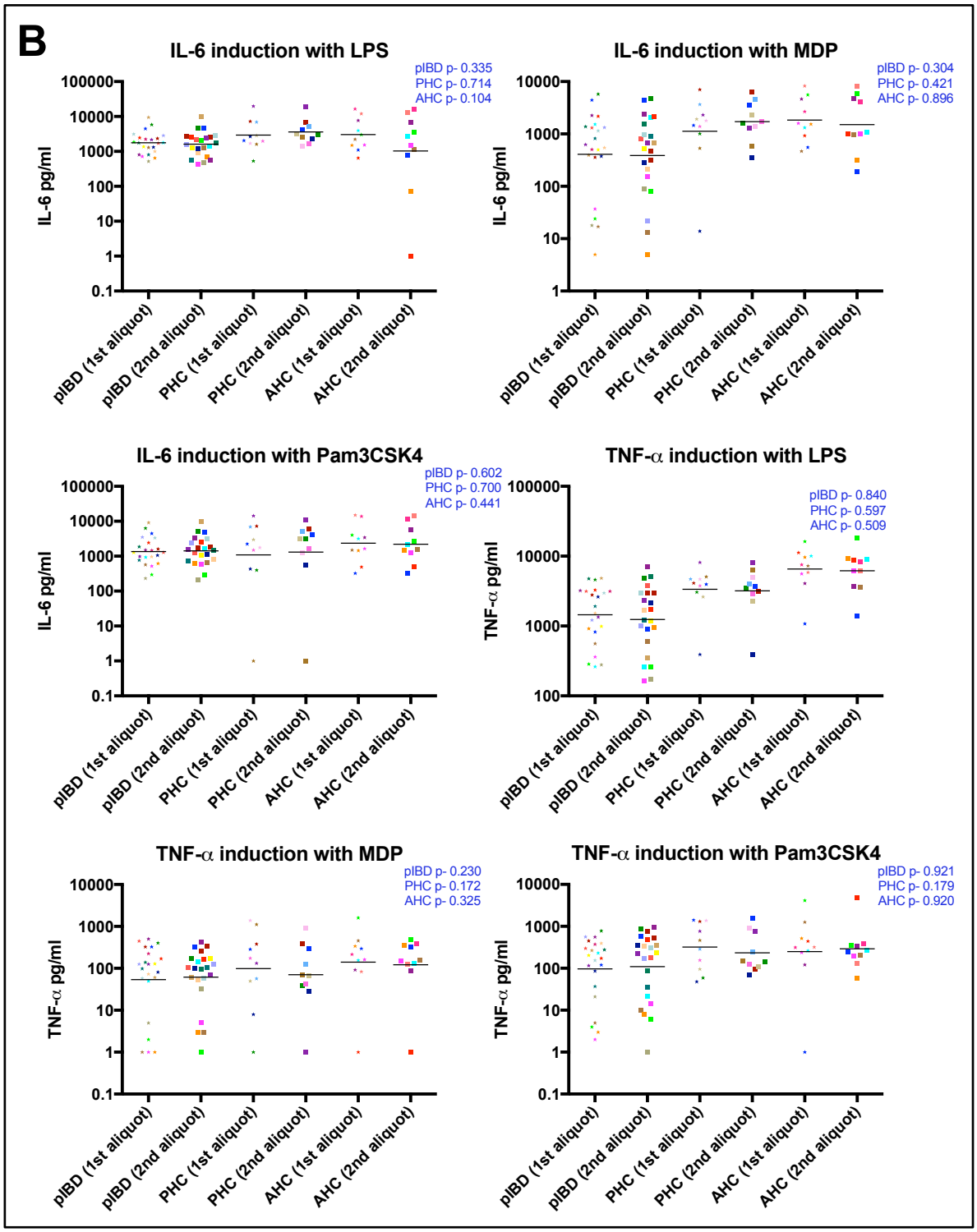


Figure 5.6 Comparing cytokine profiles between the first and second aliquots of PBMCs

*Figure **5.6A** includes IL-10 & IL-1β cytokine induction. Figure **5.6B** includes IL-6 & TNF-α induction. Multiplex assays were conducted separately on two aliquots of PBMCs per individual. PBMCs were collected at the same blood draw. Assay readings were compared between the first and the second aliquots for each of the 3 groups; paediatric IBD patients, paediatric healthy controls and adult healthy controls. Stars indicate the samples in the 1st aliquot and the solid squares the samples in the second aliquot. Each individual is indicated by a unique colour. Cytokines analysed are indicated on the Y-axis and the three groups with duplicate samples along*

*the X-axis. The p values comparing the duplicate aliquots are indicated within each of the boxes, as well as the ligands used. Paired t-tests between the two aliquots of PBMCs showed no significant differences. (**Abbreviations**: AHC- adult healthy controls; IL-6- interleukin-6; IL-10- interleukin 10; IL-1β-interleukin-1beta; LPS- lipopolysaccharide; MDP- muramyl-dipeptide; Pam3CSK4- Synthetic triacylated lipopeptide; PHC- paediatric healthy controls; pIBD- paediatric IBD patients; TNF-α- tumour necrosis factor-alpha)*

5.3.3.2 Principal component analysis (PCA) to assess for batch effects

Table **5.2** shows the distribution of samples across the five batches. The first two batches included paediatric patients and adult controls, but no paediatric controls. Paediatric controls were included in the subsequent three batches.

Table 5.2 An overview of sample distribution across batches

Batch	Patients	Adult controls	Paediatric controls
1	6	3	0
2	7	3	0
3	3	2	5
4	4	2	3
5	2	.	2
Total	22	10	10

Table 5.2 shows distribution of patients and controls across 5 different batches. Batches 1-4 included between 9-10 individuals per batch, whereas batch 5 included only 4 individuals.

The PCA plot of cytokine responses across the five batches as seen in figure **5.7** showed a varied dispersion of the samples across the five batches. Patients tended to show a distribution towards the left of the plot compared to controls, which is likely to be due to the biological nature of immune responses among patients and controls. Based on the observed distribution of the scatter points, there was no indication of batch effects.

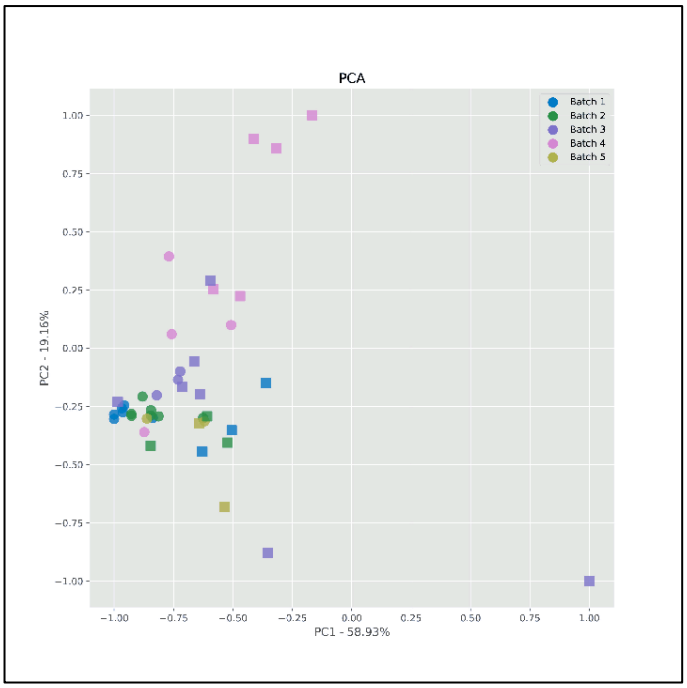


Figure 5.7 PCA on patients and controls to assess for batch effects

PCA plot was generated by converting cytokine data variables into principal components. In both figures, cases are represented as solid circles and the controls as solid squares. Each batch is represented with a unique colour. As each of the 3 stimulants (MDP, LPS & Pam3CSK4) induced 4 cytokines, a total of 12 variables were used to convert data into principal components. PCA1 accounts for the largest possible variance within the data set with the succeeding component PCA2 showing the highest variance possible under constraint with respect to the first component. The distribution of the plots was used as a surrogate tool for assessing batch effects of the immunological assay. Based on the distribution of the plot, there was no indication of batch effects.

5.3.3.3 Assay results between paediatric and adult healthy controls

Significant differences in cytokine induction for two of the twelve cytokine responses were observed between the paediatric healthy controls (PHC) and adult healthy controls (AHC). These included LPS-induced IL-10 and LPS-induced TNF-α. The p values were however not significant following multiple testing corrections. See table **5.3**. Based on first principles and MDT discussions, adult controls were excluded from all subsequent analyses comparing cytokine responses with pIBD patients.

Table 5.3 Comparing cytokine induction between paediatric and adult controls

Stimulant (Receptor)	Induced cytokine	PHC & AHC	PHC & AHC (After multiple testing correction)
LPS (TLR4)	IL-10	0.020	0.24
	IL-1b	0.120	.
	IL-6	0.984	.
	TNF-a	0.027	0.32
Pam3CSK4 (TLR1-2)	IL-10	0.252	.
	IL-1b	0.562	.
	IL-6	0.752	.
	TNF-a	0.629	.
MDP (NOD2)	IL-10	0.204	.
	IL-1b	0.772	.
	IL-6	0.624	.
	TNF-a	0.993	.

The table shows p values generated after comparing the assay readings for the 12 cytokine responses (3 stimulants × 4 cytokines) between paediatric healthy controls (PHC) and adult healthy controls (AHC) using unpaired t-tests. Significant p values (<0.05) are highlighted in red. There were significant differences between paediatric and adult controls for 2 of the 12 cytokine responses (LPS-induced IL-10 & LPS-induced TNF-α). However, the observed differences did not withstand multiple testing corrections.

5.3.3.4 Assay results between paediatric IBD patients and paediatric healthy controls

Assay results across the twelve cytokine responses were compared between paediatric IBD patients (pIBD) and paediatric healthy controls (PHC). Compared to paediatric controls, cytokine production was significantly reduced in patients for LPS-induced IL-10, IL-1β & TNF-α, and Pam3CSK4-induced IL-10 & IL-1β. MDP-induced responses for all the four cytokines did not show any significant differences. Similarly, IL-6 responses were not significantly different between patients and paediatric controls. See figure **5.8.**

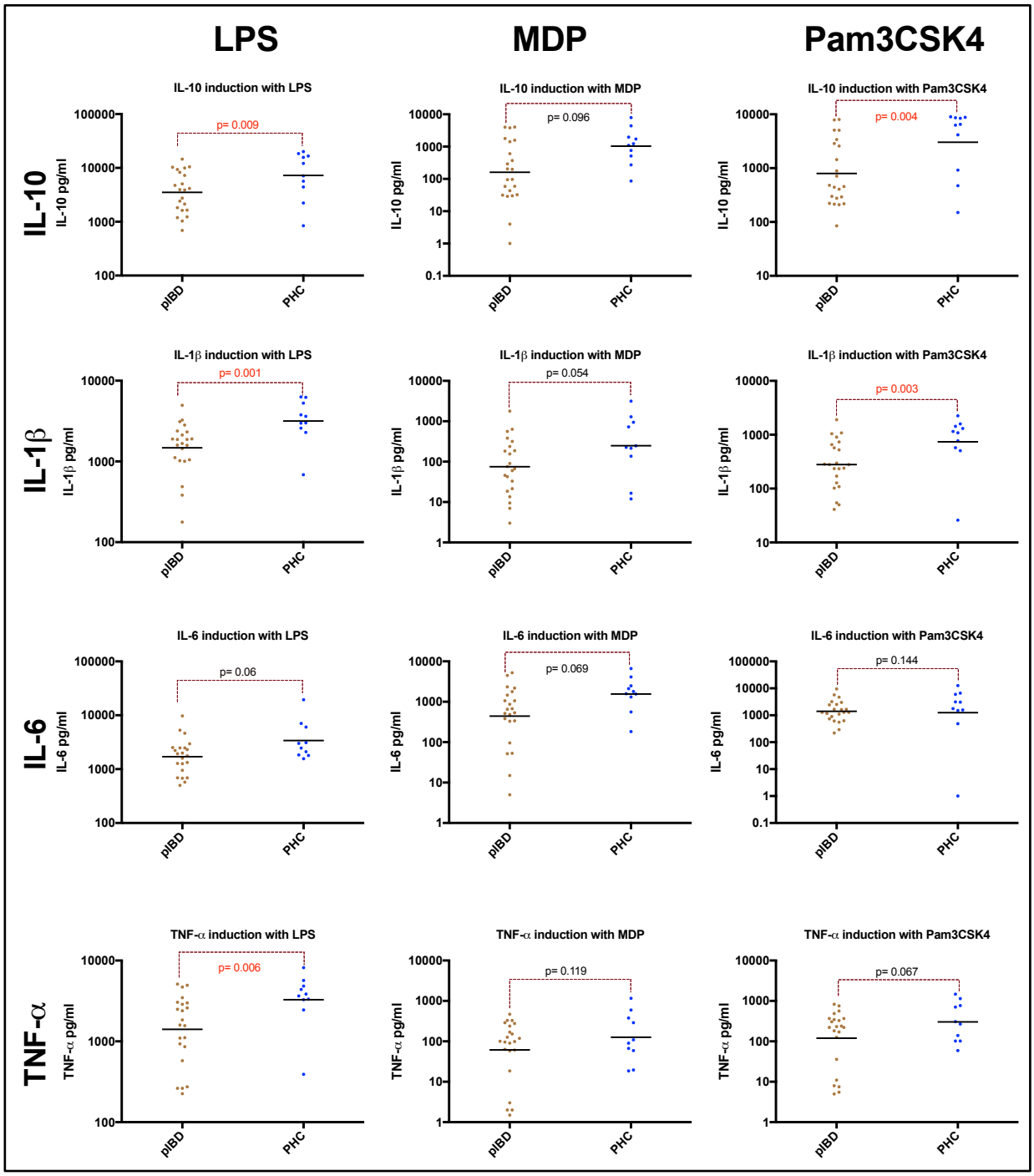


Figure 5.8 Induced immune responses between paediatric patients and paediatric controls

The figure illustrates immune induction for the twelve cytokine responses (3 stimulants × 4 cytokines), comparing paediatric IBD patients (pIBD) and paediatric healthy controls (PHC). Cytokines analysed are indicated on the Y-axis and the two groups on the X-axis. The horizontal bars across the coloured shapes are indicative of the mean cytokine values. The cytokine values between the two groups were compared using unpaired t-tests. Significant p values (<0.05) are highlighted in red.

5.3.3.5 Immune response profile of the cohort on a radar plot

Normalised cytokine values per patient per stimulus were plotted along the equi-angular spokes or radii of a radar plot, each representing a specific cytokine response (figure **5.9**). Cytokine data from paediatric controls were used to define the boundaries of normality in the patient cohort. Values within 2 standard deviations (SD) from the mean cytokine values in paediatric controls were considered within normal response limits and, those above and below 2 SD indicated hyper-inflammatory and hypo-inflammatory responses respectively. None of the patients showed responses in the hyper-inflammatory range (>2 SD). The majority of cytokine values fell between the mean and -2 SD, suggesting a trend towards hypo-inflammatory responses. One patient showed a hypo-inflammatory response along the LPS-induced IL-1β, suggestive of a signalling defect in the inflammasome activation pathway.

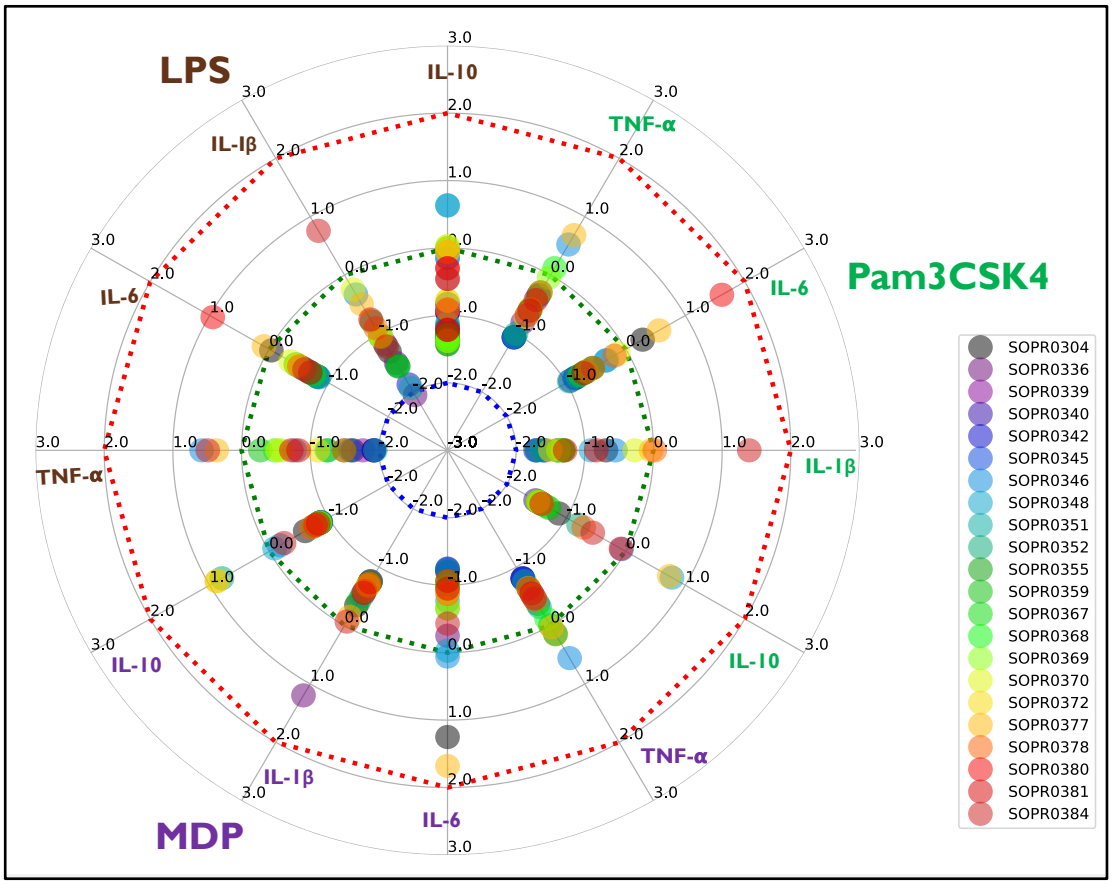


Figure 5.9 Radar plot showing cytokine responses of all patients

The ligands used for stimulation are depicted outside the boundary of the radar plot and the cytokines analysed are indicated within the boundary, with the text colour corresponding to the respective ligand. The green-dotted circle represents the mean obtained from paediatric controls. The numbers -2 and 2.0 along each vector indicate -2SD (blue-dotted circle) and 2 SD (red-dotted circle) respectively. Responses within 2 SD across the mean are within the reference range

(calculated using values from the paediatric controls). Every patient is represented by a unique colour. Individual control subjects are not represented in this figure.

5.3.3.6 Hierarchical clustering for patient stratification

In order to identify patterns or trends within the cytokine data across the twenty-two patients, hierarchical clustering of normalised cytokine profiles was used. The algorithm uses pairwise distance matrix between the observations as clustering criteria, which approximates to the value of inter-group dissimilarity based their induced immune profiles. By adopting an unsupervised machine learning approach, a complex set of twenty-two induced immune profiles across twelve cytokine responses were thus stratified into eight distinct clusters, each cluster including patients with similar immune response profiles. Figure **5.10.**

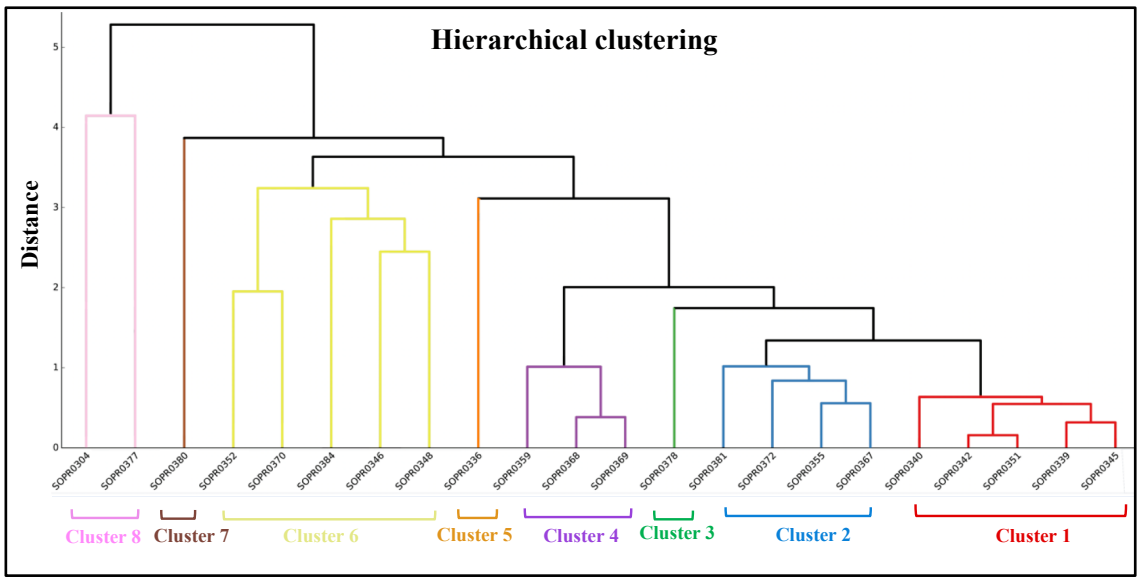
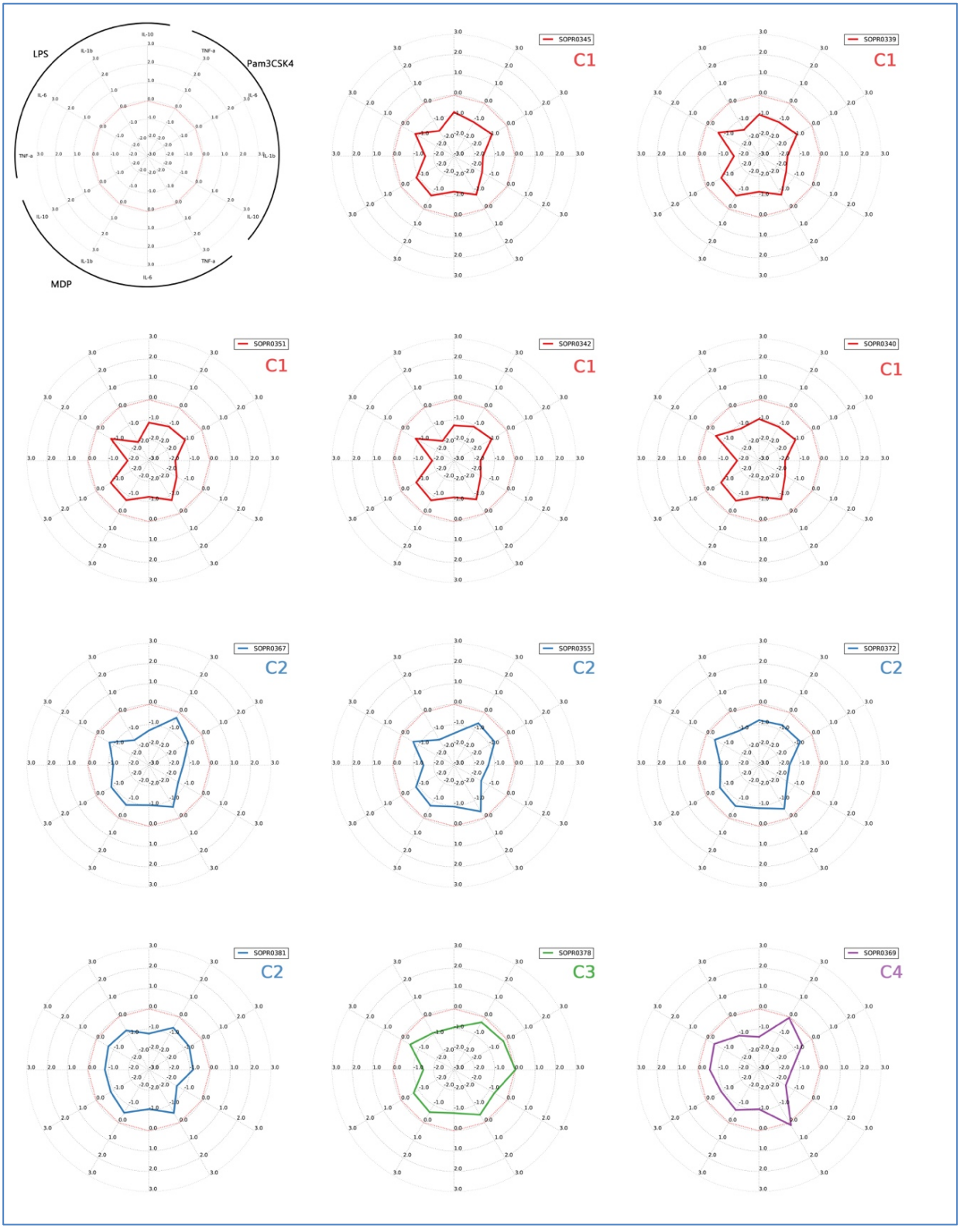


Figure 5.10 Hierarchical clustering to identify immuno-phenotypes

The figure shows eight distinct clusters of patients based on the computed similarity of their immune profiles. The tree-based representation of agglomerative clustering starts at the bottom by treating each individual as a singleton cluster and at each step of the algorithm, the two clusters that are most similar are merged into a new bigger cluster (node). The pairs of clusters are successively merged into one big cluster (root) to include the entire cohort of patients. The result is a tree-based representation of the observations per individual, which is called a dendogram. The distances were calculated using Euclidean metrics and average linkage. Each cluster is indicated in the hierarchy by a unique colour. The clusters were numbered from right to left.

5.3.3.7 Visualisation of cluster patterns on individual radar plots

In order to visualise the cytokine response patterns within individual clusters, radar plots of cytokine responses were generated per patient. Visual inspection of radar plots as groups within each cluster showed similar cytokine profiles, consistent with the stratification observed through an unsupervised machine learning approach, thereby supporting the underlying immunological basis of the clustering pattern. For example, the five patients in cluster 1 had comparatively reduced induction along the vectors for LPS-induced IL-1β & TNF-α, MDP-induced IL-6 & TNF-α and Pam3CSK4-induced IL-1β & TNF-α. See figure **5.11**.


LPS
Pam3CSK4
MDP
IL-10
TNF-a
IL-1b
IL-6
SOPR0345
C1
SOPR0339
C1
SOPR0351
C1
SOPR0342
C1
SOPR0340
C1
SOPR0367
C2
SOPR0355
C2
SOPR0372
C2
SOPR0381
C2
SOPR0378
C3
SOPR0369
C4

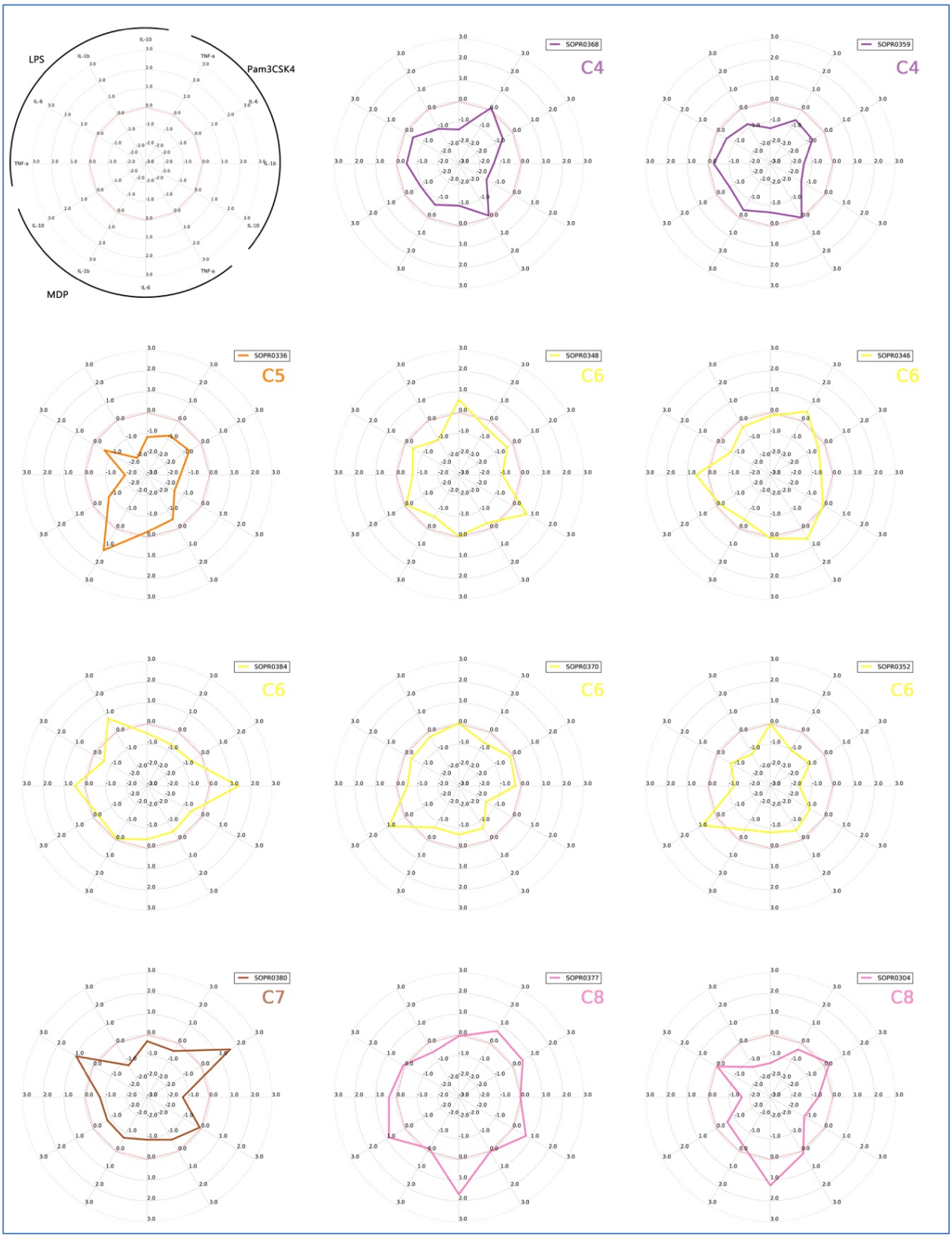


Figure 5.11 Radar plots per individual

Cytokine profiles of all patients are plotted individually on radar plots. Based on the hierarchical clustering pattern, eight different clusters or immuno-phenotypes were identified. The red-dotted line indicates the mean generated using paediatric controls. In these diagrams, patients within similar profiles are clustered and indicated by a unique colour corresponding to the hierarchical

dendogram.

5.3.3.8 Pooled cytokine data between the patient cohort and controls

Normalised cytokine values for all twenty-two patients and ten paediatric controls are presented in table **5.4.** The final column in the table represents pooled values obtained after summation of all the twelve cytokine responses per individual. There was a significant difference in the pooled responses between patients and controls (p=0.003), indicating a patient cohort signal of overall reduced immune responses. The combined reduced responses were driven predominantly by the toll-like receptor (TLR)-mediated responses [TLR4-induced IL-10 (p=0.045), IL-1β (p=0.010) & TNF-α (p=0.018) and TLR1-2-induced IL-10 (p=0.018) & IL-1β (p=0.015)], with no significant contribution from the NOD2-mediated responses. The five TLR-mediated responses with a significant difference between the patients and controls were referred to as 'dysfunctional cytokine responses' in subsequent analyses.

Table 5.4 Tabular matrix with normalised cytokine data

		TLR4 induction (with LPS)				NOD2 induction (with MDP)				TLR1-2 induction (with Pam3CSK4)				Sum
Subjects	Cluster	IL-10	IL-1B	IL-6	TNF- α	IL-10	IL-1B	IL-6	TNF- α	IL-10	IL-1B	IL-6	TNF- α	
SOPR0339	1	-0.952	-1.503	-0.690	-1.768	-0.856	-0.753	-1.242	-0.807	-1.464	-1.604	-0.846	-1.062	-13.548
SOPR0340	1	-0.945	-1.180	-0.551	-1.924	-0.845	-0.728	-1.221	-0.808	-1.519	-1.717	-0.935	-1.067	-13.440
SOPR0342	1	-1.260	-1.872	-0.825	-1.930	-0.857	-0.749	-1.197	-0.804	-1.481	-1.695	-0.863	-1.061	-14.593
SOPR0345	1	-0.829	-1.558	-0.805	-1.585	-0.868	-0.761	-1.248	-0.807	-1.405	-1.572	-0.842	-1.066	-13.345
SOPR0351	1	-1.124	-1.934	-0.840	-1.949	-0.829	-0.756	-1.221	-0.759	-1.426	-1.702	-0.955	-1.055	-14.549
SOPR0355	2	-1.427	-1.543	-0.679	-1.499	-0.845	-0.698	-0.963	-0.371	-1.459	-1.312	-0.692	-0.607	-12.095
SOPR0367	2	-1.289	-1.568	-0.752	-1.241	-0.856	-0.744	-1.037	-0.627	-1.339	-1.319	-0.768	-0.296	-11.836
SOPR0372	2	-0.780	-1.043	-0.489	-1.112	-0.784	-0.681	-0.885	-0.525	-1.412	-1.502	-0.661	-0.715	-10.589
SOPR0381	2	-1.213	-0.760	-0.693	-0.830	-0.828	-0.559	-1.066	-0.549	-1.416	-0.832	-0.710	-0.607	-10.062
SOPR0378	3	-0.913	-0.909	-0.508	-1.491	-0.708	-0.594	-0.872	-0.445	-0.710	0.018	-0.192	-0.294	-7.618
SOPR0359	4	-1.286	-0.794	-0.563	-0.274	-0.743	-0.415	-0.656	-0.004	-1.287	-1.394	-0.661	-0.569	-8.646
SOPR0368	4	-1.347	-1.057	-0.458	-0.489	-0.858	-0.713	-0.979	-0.115	-1.457	-1.319	-0.566	0.120	-9.238
SOPR0369	4	-1.377	-1.057	-0.457	-0.577	-0.851	-0.717	-1.062	0.150	-1.483	-1.386	-0.555	-0.043	-9.415
SOPR0336	5	-1.172	-2.054	-0.627	-1.929	-0.870	1.195	-0.251	-0.533	-1.480	-1.396	-0.662	-0.809	-10.587
SOPR0348	6	0.632	-1.044	-0.448	-0.776	-0.091	-0.665	-0.029	-0.320	0.775	-0.911	-0.321	-0.382	-3.580
SOPR0352	6	-0.002	-1.226	-0.803	-1.257	0.792	-0.564	-0.765	-0.519	-0.797	-1.615	-0.811	-1.002	-8.570
SOPR0370	6	0.026	-0.248	-0.362	-0.523	0.872	-0.690	-0.659	-0.643	-1.482	-0.273	-0.127	-0.585	-4.696
SOPR0346	6	-0.147	-0.331	-0.803	0.582	-0.184	-0.503	0.058	0.557	-0.071	-0.554	-0.344	0.524	-1.215
SOPR0384	6	-0.458	0.758	-0.595	0.492	-0.251	-0.068	-0.431	-0.467	-0.558	1.395	-0.555	-0.689	-1.427
SOPR0380	7	-0.300	-1.222	0.948	-0.726	-0.782	-0.740	-0.950	-0.635	-0.083	-1.288	1.614	-0.427	-4.590
SOPR0304	8	-1.353	-1.312	-0.038	-1.622	-0.606	-0.346	1.251	0.144	-1.127	-0.692	0.282	-0.343	-5.762
SOPR0377	8	-0.057	-0.497	0.087	0.361	0.881	-0.151	1.680	0.031	0.730	-0.035	0.557	0.685	4.272
SOPR0330	Controls	-0.691	-0.033	0.431	-0.183	-0.122	0.034	2.458	0.296	0.959	0.866	0.823	0.417	5.257
SOPR0364	Controls	0.264	0.945	-0.529	0.871	2.578	2.684	-0.938	0.921	0.342	1.960	-1.015	-0.423	7.660
SOPR0365	Controls	1.201	-0.409	-0.354	-0.799	1.052	-0.528	-0.067	-0.641	0.894	-0.826	-0.164	-0.861	-1.501
SOPR0371	Controls	0.792	-0.639	-0.589	-0.374	-0.333	-0.615	-0.518	-0.756	1.054	0.142	-0.530	-0.861	-3.226
SOPR0374	Controls	-1.405	-1.755	-0.459	-1.863	-0.833	-0.746	-1.149	-0.759	-1.501	-1.743	-0.881	-0.953	-14.044
SOPR0375	Controls	-0.472	-0.816	-0.337	-0.079	-0.394	-0.516	0.142	0.034	-0.338	0.400	-0.145	2.061	-0.461
SOPR0376	Controls	0.933	-0.398	0.235	0.197	-0.010	-0.493	1.061	-0.550	0.997	-0.501	0.629	-0.509	1.592
SOPR0379	Controls	1.453	1.498	2.818	2.147	-0.752	-0.751	-0.240	-0.617	0.277	-0.942	2.468	0.556	7.914
SOPR0382	Controls	-1.200	0.065	-0.634	0.427	-0.646	0.660	-0.367	2.565	-1.278	0.036	-0.603	1.353	0.378
SOPR0383	Controls	-0.875	1.542	-0.582	-0.346	-0.540	0.271	-0.384	-0.493	-1.407	0.608	-0.582	-0.780	-3.568
p values		**0.045**	**0.010**	0.178	**0.018**	0.188	0.164	0.116	0.267	**0.018**	**0.015**	0.245	0.169	**0.003**

Table 5.4 shows normalised cytokine data following stimulation with the receptor-specific ligands including LPS (TLR4 stimulant), MDP (NOD2 stimulant) and Pam3CSK4 (TLR1-2 stimulant). Induction of 4 cytokines including IL-10, IL-1β, IL-6 and TNF-α were assessed per stimulant,

thereby generating 12 assay conditions in total. Unpaired t-tests were conducted between patients and controls across the 12 conditions. These were significantly different for 5 conditions (highlighted in red).

5.3.3.9 Assessing inter-cluster differences using the 'dysfunctional cytokine responses'

In order to determine if the five significantly reduced cytokine responses observed across the twenty-two patients compared to controls were specific to individual patient clusters rather than the whole group, the within cluster values were compared against the control group using unpaired t-tests. This was conducted in order to qualitatively and statistically analyse how each cluster differed from the control group in terms of their dysfunctional cytokine responses. Cluster 1 showed significant differences for four out of five and cluster 2 for three out of five cytokine responses compared to controls (table **5.5**). As clusters 3, 5 and 7 had one individual each, t-tests were not performed. Cluster 4 showed only borderline significant differences for TLR1-2 mediated IL-10 induction (p=0.047), whilst clusters 6 and 8 did not show significant differences compared to controls. As there was stronger evidence of reduced cytokine responses in clusters 1 and 2, and to avoid multiple statistical comparisons between small clusters, clusters 3-8 were merged into one cluster, thereby leaving 3 clusters (cluster 1, cluster 2 and clusters 3-8) for subsequent analyses.

Table 5.5 Unpaired t-tests comparing cytokine data between each cluster and controls

Receptors	**TLR4-induction (with LPS)**			**TLR1-2 induction (with Pam3CSK4)**	
Cytokines	IL-10	IL-1B	TNF-a	IL-10	IL-1B
Cluster 1	0.0500	0.0060	0.0020	0.0090	0.0040
Cluster 2	0.0500	0.0400	0.0500	0.0200	0.0400
Cluster 3	.	.	.	.	.
Cluster 4	0.0500	0.1500	0.4900	0.0500	0.5600
Cluster 5	.	.	.	.	.
Cluster 6	0.9800	0.4400	0.5900	0.4400	0.5100
Cluster 7	.	.	.	.	.
Cluster 8	0.4020	0.2700	0.4700	0.8170	0.6500

Normalised cytokine data for 5 conditions were compared between each cluster and the control group using unpaired t-tests. Cluster 1 and 2 were significantly different (red) compared to the control group. As clusters 3, 5 & 7 had one individual each, t-tests were not performed.

5.3.4 Exome data analysis

5.3.4.1 Gene networks

The five significantly reduced immune responses observed in the patient cohort compared to controls were used to determine the selection of genes through PathCards for the application of GenePy scores. The search terms entered included 'TLR1', 'TLR2' and 'TLR4' for the receptors; 'NF-κB', 'MAPK' and 'inflammasome' for downstream signalling cascades, and 'IL-10', 'IL-1β' and 'TNF-α' for the cytokines. NOD2 and IL-6 were not included in the search as NOD2 receptor-mediated responses and IL-6 cytokine effector responses showed no significant differences between patients and controls.

The closest functionally connected SuperPath nucleating from the search for the receptor-associated search terms was TLR4, comprising a network of 126 genes. This was chosen as the unifying set of genes for all the three toll-like receptors, referred to as the 'TLR gene network' for subsequent analyses. As the NF-κB, MAPK and inflammasome activation pathways are inter-connected and to minimise redundancy of gene grouping, these were merged into one non-overlapping set of genes, referred to as the 'signalling cascade' genes network. Two additional genes of known importance (*GSDMD* and *TRIF*) in inflammasome activation[448], but not retrieved through PathCards search were also added to the list of signalling cascade network generating a total of 166 genes. The gene list retrieved for the cytokines IL-10, TNF-α and IL-1β included 35, 72 and 7 genes respectively. For the list of genes selected for IL-1β, there was only one individual within the cohort with a single variant gene (*NFKB2*). The IL-1β network of genes was therefore excluded from subsequent analyses. See figure **5.12** for a Venn diagram depicting the overlap in genes derived from the four selected gene networks (TLR, signalling cascades, IL-10 and TNF. See tables **D.1-D.2** in the appendix D for the list of genes nucleating from the search through PathCards and the genes with available GenePy scores.

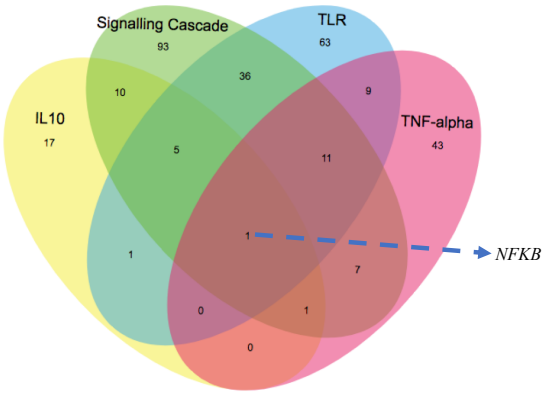


Figure 5.12 Venn diagram of selected gene networks

The four selected gene networks are shown in the diagram including TLR, signalling cascade, IL 10 and TNF-alpha. The signalling cascade was a merged network comprising NF-κB, MAPK and inflammasome SuperPath. Interestingly, the only gene that was shared by all four gene networks was NFKB, shown in the central panel overlapping all networks.

5.3.4.2 GenePy scores across the clusters

GenePy scores were generated across all selected genes for every patient. One patient of non-Caucasian ethnicity (SOPR0348) was excluded, as GenePy scores to date have been calculated using Caucasian population allele frequencies. For each of the four selected gene networks, the GenePy scores were summed for individuals within clusters 1, 2 and clusters 3-8. Unpaired t-tests performed to compare the summed GenePy scores between clusters 1, 2 and 3-8 are shown in table **5.6.** Also see tables **D.1-D.2** in Appendix D.

Table 5.6 Selected gene networks and GenePy scores across the clusters

Gene networks	Number of genes (PathCards)	Genes with GenePy scores	Genes with variations	t-tests between clusters			
				p values (uncorrected)		p values (corrected)	
				Cluster 1 V 3-8	Cluster 2 V 3-8	Cluster 1 V 3-8	Cluster 2 V 3-8
Receptors							
TLR	126	74	50	0.9383	0.4818	7.5064	3.8544
Signalling cascades							
NF-KB	70						
MAPK	65	104	64	0.2832	0.6288	2.2656	5.0304
Inflammasome	54						
Cytokines							
IL-10	35	23	11	0.3603	0.3352	2.8824	2.6816
TNF-alpha	72	32	18	0.0001	0.0959	0.0008	0.7672
IL-1B	7	5	1	.	.	.	.

The selected gene networks shown in the first column included genes involved in pathway function at the receptor level, downstream signalling level and the cytokine level. Due to overlap of genes involved in the functioning of the 3 cross-connected pathways (NF-KB, MAPK and inflammasome)

at the signalling cascade level, the list of genes nucleating from the search was combined, generating a non-overlapping list of 104 genes as extracted from 'PathCards' webserver. The list of genes retrieved through the search was further reduced as per the availability of GenePy scores for the selected genes. GenePy scores were applied across the clusters and a comparative analysis performed across three key groups of functional clusters including cluster 1, cluster 2 and clusters 3-8. IL-1β was not included in the analysis as there was only a single individual with a variant gene. The scores between cluster 1 and clusters 3-8 were significantly different for the TNF-alpha network of genes.

GenePy scores for TLR, signalling cascades and IL-10 showed no significant differences between the 3 cluster groups. A significant difference that withstands multiple testing correction, was observed between cluster 1 and clusters 3-8 for the TNF-α gene network (p= 0.0008). The mean scores were significantly higher in cluster 1 compared to cluster 3-8, suggesting statistical excess of deleterious mutations within the TNF-α signalling pathway in cluster 1. In order to further analyse variant genes within the TNF-α signalling pathway, unpaired t-tests of GenePy scores per gene were performed between cluster 1 and clusters 3-8. Out of the 16 variant genes in individuals within the two functional cluster groups, a significant difference was observed for ubiquitin-specific protease 21 gene (*USP21,* p= 0.0187), a de-ubiquitinase regulating TNF-α-mediated NF-κB signalling[449].

5.4 DISCUSSION

This study adopted a unique approach of first assessing the immunological phenotype of a treatment naïve paediatric cohort through selected innate immunity pathway profiling, followed by patient stratification through unsupervised machine learning and finally sequence interrogation of inferred gene networks. The study showed a significantly reduced cytokine induction in the patient cohort compared to controls, which was driven predominantly by hypo-inflammatory TLR-mediated responses. Application of unsupervised machine learning approaches such as hierarchical clustering to the cytokine data enabled stratification of patients into distinct clusters based on their induced immune profile. Generation of radar plots per individual for visual representation of the cytokine data demonstrated similar patterns within clusters, thereby providing a supportive conceptual evidence of the immunological basis of the hierarchical clustering. Furthermore, the clustering pattern highlighted putative genomic contributions following interrogation against the exome data. A statistical excess of mutations in the TNF-α signalling pathway were identified in one cluster, with a significant contribution from

USP21 gene, a de-ubiquitinase, which blocks downstream progression of TNF-α induced NF-κB signalling[450].

The reduced cytokine induction observed in PBMCs in this study may reflect signalling defects in the key inflammatory pathways resulting in impaired cytokine production. Reduced or absent innate immune induction has been reported previously, particularly in the context of known IBD-associated mutations in genes such as *NOD2*[251,267]. Van Heel *et al* in 2005 demonstrated absence of pro-inflammatory cytokine induction in PBMCs from CD patients with known *NOD2* double mutant genotypes, suggesting abnormal adaptive immune responses to microbial antigens in these patients[251]. In another study published in 2005, Netea *et al* showed that the PBMCs from patients homozygous for the *NOD2* mutations were totally unresponsive to stimulation by MDP as well as other muramyl peptides including NOD1 receptor stimulants (Tri-DAP). The authors proposed cross-tolerance between NOD1 and NOD2 suggesting that *NOD2* mutations could lead to down-regulation of NOD1 signalling[267]. The findings from our study, combined with the weight of evidence from previously published studies[251,267] suggest that poor or maladaptive immune response to pathogenic insults, with a consequential poor bacterial clearance and persistence of secondary inflammation may be a key driving mechanism in the pathogenesis of IBD. Other functional studies investigating the role of mutations in the autophagy gene *ATG16L1* have demonstrated increased production of pro-inflammatory cytokines such as IL-1β, upon NOD2 receptor stimulation. Genetic variations in *ATG16L1* result in impaired autophagy induction with a consequent loss of inhibition of pro-inflammatory cytokine drive, thereby favouring bacterial persistence and recurring secondary inflammation[268]. Furthermore, mechanistic studies investigating the impact of IL-10 receptor mutations in IBD have demonstrated excessive production of TNF-α and other cytokines due to loss of its functional capacity as an anti-inflammatory protein[199]. Another possible mechanism to explain the reduced cytokine induction observed in our patient cohort is down regulation of cytokine production in PBMCs due to the inhibitory effect of stromal factors secreted by the inflamed bowel tissue. It has been shown that peripheral blood cells such as monocytes when cultured with conditioned media from the intestinal stromal cells can gain features of the gut macrophages with a down regulated production of cytokines including TGF-β and IL-10[138,451,452]. The anti-inflammatory effects of IL-10 produced by actively inflamed tissue in IBD have also been well described. IL-10 has been shown to downregulate MHC class II expression and also inhibit the production of various cytokines by monocytes[452].

In this study, hypo-inflammatory immune responses were driven predominantly by TLR-mediated responses. TLRs are cell-surface or endosomal pathogen-recognition receptors, widely expressed in the GI epithelium. Alterations in TLR expression have been associated with the pathogenesis of IBD[166]. TLR2 forms a hetero-dimer with TLR1 or TLR6, which is necessary for signal transduction. Current evidence indicates a protective role for TLR1-TLR2 mediated responses[453]. Results have shown that poor expression of TLR1 during acute infection can lead to chronic bowel inflammation and dysbiosis. Experimental evidence suggests that a TLR2 focussed agonistic approach could hold promise for future therapeutic applications in IBD. On the other hand, although TLR4 signalling is protective against invading bacteria, it may trigger or aggravate mucosal inflammation through release of nitric oxide with consequent pro-inflammatory cytokine induction. Although the precise mechanistic influence of TLR4 on IBD pathogenesis is unknown, current evidence is indicative of a bi-directional role[453].

Multiple studies assessing induced immune function have been described previously, however, the vast majority have been conducted in adult IBD patients following an established diagnosis and treatment[349]. In our study, immunological assays were performed in treatment naïve paediatric IBD patients where PBMCs were obtained at or before diagnosis, thereby removing the potential impact of drug interventions on immune responses. Previously published studies have clearly demonstrated the effect of drug treatments on the functioning of immunological assays[424,425].

In this study, we initially assessed the induced immune variance using both paediatric and adult non-IBD controls. Significant differences were observed for two out of the twelve cytokine responses (TLR4-induced IL-10 and TNF-α) between paediatric and adult controls. However, the observed differences did not withstand Bonferroni correction. It has been previously reported that the immune function remains largely stable for several decades between infancy and old age[385,440,441,454-459]. Less is known about the ontogeny of innate immunity, however a study by Orange et al in 2006 showed that the production of pro-inflammatory cytokines following activation of toll-like receptors remained stable over a six-decade age range, from birth to 57 years[460,461]. Population based immunology studies have shown that age can impact on 20% of immunological traits, but overall contributes to only 5% of immune variation[440,462-467].

The pathogenesis of IBD is influenced by a complex interplay of cytokines and signalling networks within and outside the bowel micro-environment[353]. Dysregulated immunity with

aberrant cytokine responses are known drivers of inflammation in IBD[468]. Given the lack of reliable predictive markers to determine treatment responses, current treatment protocols are driven predominantly by disease severity, distribution and progression[469]. Despite the availability of newer treatment options including anti-cytokine therapies such as infliximab, adalimumab, ustekinumab and other monoclonal antibodies, primary treatment failure or inadequate responses continue to pose significant challenges in certain sub-groups of patients[45,46 44,470-472]. Treatment failures result in escalation of treatment strategies, often necessitating invasive surgical procedures, with consequential life-long morbidity. A better understanding of the molecular pathogenesis of the disease on a personalised basis would result in better outcomes. Currently, there is no guidance in place to inform a personalised approach to treatment that considers an individual's immune profile. The findings of this study indicate that patients with IBD can be stratified into sub-groups or clusters based on their immune response profiles using an unsupervised machine learning approach.

Furthermore, sequence interrogation of genes from selected molecular pathways using GenePy scores identified statistically significant genomic signals in the TNF-α signalling pathway in one cluster, with a significant contribution from the de-ubiquitinase gene *USP21* gene. This gene inhibits the downstream progression of TNF-α induced NF-κB signalling by targeting poly-ubiquitinated adaptor proteins[450]. See figure **5.13.**

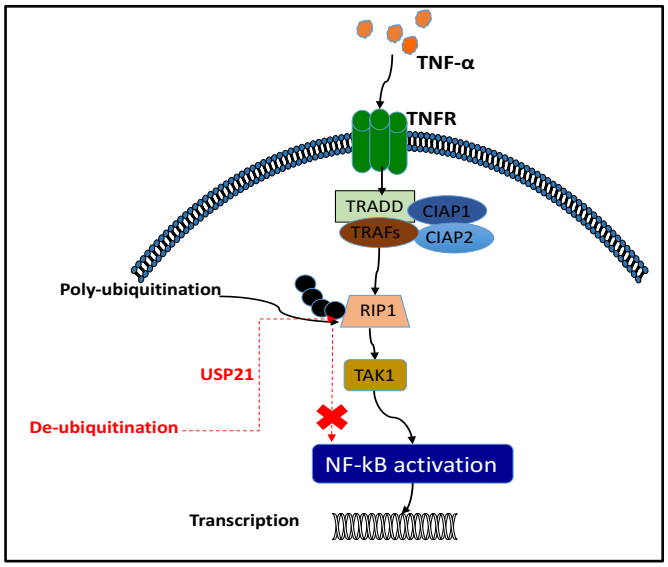


Figure 5.13 TNF-α-induced NF-κB activation pathway

Activation of the cell-surface TNF receptor (TNFR) results in the recruitment of various adaptor proteins including TRADD, cIAP1, cIAP2, TRAF and RIP1. Poly-ubiquitination of RIP1 promotes downstream progression of the cascade through activation of a multi-functional kinase TAK1,

*finally leading to activation of NF-κB pathway. The process can be reversed through de-ubiquitination of the adaptor RIP1 by de-ubiquitinase USP21. (**Abbreviations**: cIAP, cellular inhibitors of apoptosis; NF-κB, nuclear factor-κB; RIP1, receptor interacting response protein 2; TAK1, TGF-β-activated kinase; TRADD, TNF-R1-associated death domain protein; TRAF, TNF-receptor associated factor; USP21, ubiquitin specific proteases 21)*

Ubiquitination is a post-translational process which involves addition of a ubiquitin molecule to a target protein, either as a monomer or polyubiquitin chain (poly-Ub). The poly-ubiquitinated target protein may be marked for proteasomal degradation depending on the type of poly-Ub chains attached, thereby promoting downstream signalling along the cascade. Ubiquitination can be reversed by removal of the ubiquitin chains by de-ubiquitinating enzymes or de-ubiquitinases (DUBs). There are approximately 100 DUBs encoded by the human genome, falling into five different families; ubiquitin-specific proteases (USP), ubiquitin C-terminal hydrolases (UCH), ovarian tumour domain containing proteases (OTU), Machado-Josephin domain containing proteases (MJDs) and a zinc-metalloprotease group[473]. The de-ubiquitinase USP21is constitutively associated with RIP1, a key adaptor involved in NF-κB signalling. USP21 plays a crucial role in the downregulation of TNF-α-induced NF-κB activation through de-ubiquitination of RIP1[450]. In this study, PBMCs were not activated using TNF-α as a stimulant, but a TNF-α effector response was assessed as part of the assay cytokine read-outs. Therefore, we postulate an amplification pathway where TNF-α produced from initial innate stimuli activates this TNF-α mediated NF-κB pathway and the *USP21* variants impact upon this contributing to disease.

Our study was limited by modest cohort size and profiling of a limited set of innate immune cytokine responses. To confirm wider applicability of the study findings, implementation of larger cohorts and more extensive immune profiling is required. Furthermore, as all patients had active inflammation at diagnosis when the samples were collected, it is unclear if the observed hypo-functionality is an epiphenomenon of inflammation. To follow this prospectively, it is important to demonstrate a maintained 'cluster signature' by using PBMCs collected from patients during remission, whilst not on any treatment that would impact on the assay function. However, it would be practically challenging to identify a sizable cohort of patients with controlled disease, but not on immunosuppressive treatment.

This study highlights the importance of profiling the immunological genome in IBD for patient stratification in order to implement tailored therapy. The scope and clinical relevance of this approach in IBD is ever expanding with the advancements in the field of immune therapies and treatment with biologics. Despite dramatic responses to treatment with certain biologics experienced by a subset of patients, identifying biomarkers or a biological profile to determine patient responses to specific treatments remains a formidable challenge. As a futuristic model of precision medicine, patient stratification on the basis of immuno-genomic profiling will lead to development of tailored treatment strategies with optimal outcomes and minimal side-effects.

Section III

Chapter 6 BIO-CLINICAL AND GENOMIC PROFILING OF PERIOSTIN IN IBD

This chapter describes the assessment of periostin as a potential biomarker of disease activity in IBD, immuno-histochemistry of periostin expression in the GI mucosal tissue and the analysis of exome sequencing data against clinical outcomes and plasma levels of periostin. The study was conceptualised and designed by myself following discussions with Professor Dieter Riethmacher and Dr Eva Riethmacher, who have experience in tissue analysis of periostin in animal models. The analysis of plasma periostin levels through ELISA was conducted by myself under the supervision of Dr Annie Griffins and Dr Yifang Gao at the WISH laboratory. Immuno-histochemical staining of the GI specimens was performed by Dr Eva Riethmacher in her laboratory at the Duthie building on the UHS campus. Application of exome data was performed with the help of Enrico Mossotto, PhD student from the Genomics Group. The collection of GI tissue specimens, clinical data extraction and interpretation in the context of biochemical, immuno-histochemical and exome findings was performed by myself with input from supervisors.

6.1 INTRODUCTION

Periostin is an extra-cellular matrix (ECM) protein which plays a key role in wound healing, tissue re-modelling and fibrosis, in several tissues including the gastro-intestinal (GI) tract. Altered expression has been described in various inflammatory conditions including IBD[150,151]. The extra-cellular matrix is a complex micro-environment around tissues consisting of proteins which play a crucial role in regulating the behaviour of surrounding cells and maintaining tissue homeostasis. A tightly balanced ECM homeostasis is important to maintain normal functioning of tissues, wound healing and repair. The ECM proteins comprise a group of structurally unrelated proteins which include osteonectin, periostin, tenascins, thrombospondins, transforming growth factor-beta 1 (TGF-β1) and others[145,146]. Normally these proteins are expressed at very low levels within the tissues, but the expression is up regulated in response to inflammation and injury. Some of the key functions of ECM proteins include interacting with other ECM proteins to modulate their activities, re-modelling the tissue micro-environment, binding to cytokines and growth factors, and engaging

with specific cellular receptors, resulting in intra-cellular signal transduction. These proteins do not have a primary role in regulating tissue architecture, but have important biological functions during embryogenesis, tissue inflammation, tumorigenesis and tumour progression[154].

Periostin is a member of the fasciclin family based on its structural homology to fasciclin I (fas1) initially identified in *Drosophila melanogaster*. In humans it is encoded by the *POSTN* gene located on chromosome 13. It was originally identified as an adhesion protein in a mouse osteoblastic cell line and named as osteoblast-specific factor 2 (OSF-2). The name periostin was derived from its expression in the periodontal ligament and periosteum of adult mice when it was initially discovered[474,475]. Following the initial publication in 1993[474], research work has been conducted in divergent fields of biology highlighting its multi-faceted role in health and disease[155]. However a common functional denominator is the overarching role of periostin in mesenchymal remodelling, regeneration and tissue repair[155].

Although the source of plasma periostin is unclear, previous studies have shown that periostin is produced by epithelial cells and fibroblasts in the tissue, as well as monocytes and fibrocytes from the peripheral blood[156,476,477]. Its expression is upregulated in a tissue-specific context by several other proteins including TGF-β1, bone morphogenetic proteins (BMP) 2 and 4, various interleukins (IL3, 4, 6 and 13), erbB3 activation, vascular endothelial growth factor, vitamin K and others[155,478]. Periostin can influence the functioning of other signalling pathways including the NF-κB, focal adhesion kinase (FAK) signalling and phosphatidylinositol 3-kinase (PI3K)/AKT pathways through its interaction with the αv-integrins, which are trans-membrane receptors facilitating adhesion of cells to the ECM. See figure **6.1.** It also regulates the expression of several other genes within its functionally connected network including collagen, α-smooth muscle actin (α-SMA), TGF-β1 and chemokines[155,479].

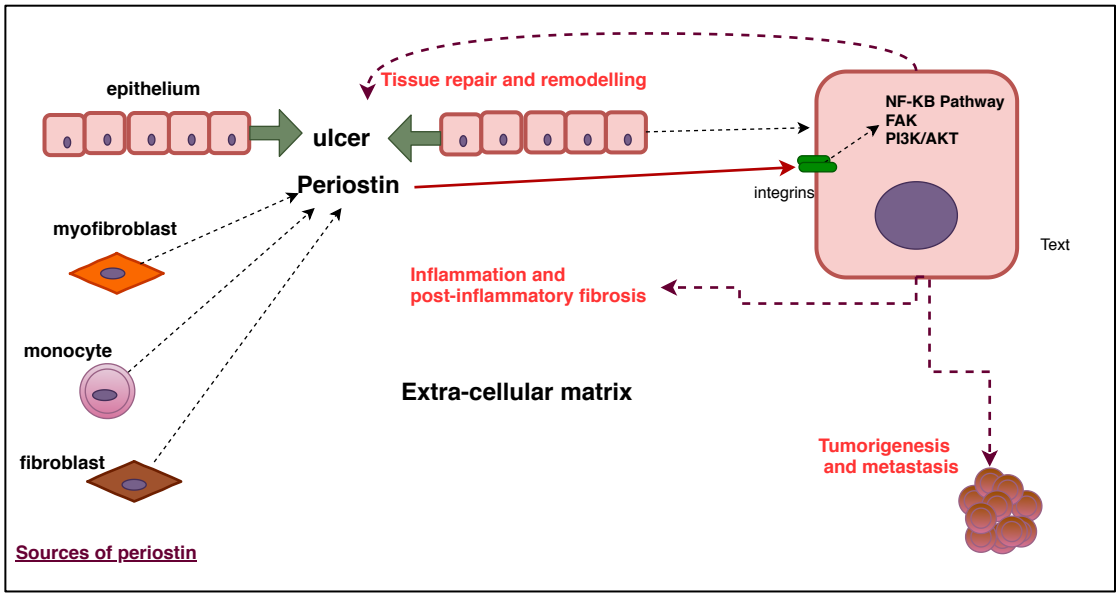


Figure 6.1 Pathophysiology of periostin

Periostin in tissue is derived from the fibroblasts, myofibroblasts and circulating monocytes. Periostin interacts with the transmembrane integrin receptors, facilitating extra-cellular matrix (ECM) adhesion and activation of signal transduction pathways, which include the NF-κB, focal adhesion kinase (FAK) signalling and phosphatidylinositol 3-kinase (PI3K)/AKT pathways.

Previous studies have demonstrated the role of periostin in the induction of inflammatory cascades in inflamed tissues through interaction with NF-κB and other pro-inflammatory pathways[147,150,480]. A study in atopic dermatitis demonstrated induction of pro-inflammatory cytokines such as IL4 and IL13 following binding of periostin to integrin molecules on the surface of keratinocytes[147]. A study in murine models and patients with UC demonstrated that periostin mediated inflammation in the lamina propria through activation of the the NF-κB pathway[150].

Studies in animal models and patients have identified the role of periostin in the biology of various malignancies, inflammatory and fibrotic conditions[146]. Overexpression of periostin has been reported in a wide range of tumours including the breast, lung, pancreas, ovaries, oesophagus, colon and liver cancers[149]. Although its precise role in cancer progression is unclear, studies have shown that cancer cells can induce periostin secretion within the cancer stem cell niche and create a supportive microenvironment to promote tumorigenesis and dissemination of the cancer cells[148,149,153,154]. Studies have shown that periostin promotes cancer cell survival, invasion and metastasis through activation of the PI3K/AKT pathway. In a

study by Bao et al, colon cancer cell lines which have a low metastatic potential, were transfected to induce overexpression of periostin, resulting in a strikingly accelerated tumour growth[152].

The role of periostin has been well described across a spectrum of inflammatory and fibrotic conditions including inflammatory conditions of the respiratory tract (asthma, chronic obstructive pulmonary disease, allergic rhinitis, idiopathic pulmonary fibrosis), systemic sclerosis & scleroderma, atherosclerosis, myocardial fibrosis, renal interstitial fibrosis, hepatic fibrosis, inflammatory bowel disease and others[146]. Fibrosis, characterised by excessive deposition of ECM components such as collagen, is the result of dysregulated repair accompanying various insults including infections, inflammation, toxins, allergens or radiations. The process involves migration of the circulating fibrocytes to the site of insult, fibroblasts differentiation, myofibroblasts accumulation, epithelial-to-mesenchymal transition and excessive collagen deposition, promoted by pro-inflammatory cytokines and chemokines[481,482]. There is a growing body of evidence linking periostin with the development of fibrosis and as a balancing mediator between adequate versus excessive tissue adaption in response to inflammation[155,156,477,481,483].

In IBD, intestinal fibrosis with consequent stricture formation, resulting from persistent inflammation and poor mucosal healing, represents a significant disease burden. Stricturing disease is a leading cause of surgical interventions in IBD, which include bowel resections, balloon dilatations and strictureplasty[484]. In UC, the inflammatory process involves the mucosa and the submucosal layer of the large bowel. A dysregulated post-inflammatory repair can cause thickening of the muscularis mucosa of the bowel wall, with fibrosis, shortening and stiffening of the involved colonic segment. In CD, the inflammatory changes show a transmural involvement and an abnormal post-inflammatory repair can lead to excessive fibrosis, stenosis and stricture formation involving any segment of the GI tract[485,486]. Although increased expression of periostin in the bowel mucosa has been described previously in patients with UC[150], the role of periostin in fibrogenesis and stricture formation in IBD has not yet been explored. A number of studies have indicated increased serum periostin levels in association with inflammatory and fibrotic conditions[487-490], but an association with IBD has not been previously described.

6.2 AIMS AND OBJECTIVES

- Explore the utility of plasma periostin as a biomarker of disease activity in IBD
- Assess plasma levels of this protein against unfavourable clinical outcomes resulting from excessive fibrosis such as stenosis and stricture formation
- Interrogate exome sequencing data to assess differences in mutational burden between patients in the context of clinical outcomes and plasma levels of periostin
- Immuno-histochemical staining of colonic mucosal tissue to investigate the expression of periostin in individuals affected with IBD and compare with healthy controls

6.3 METHODS

6.3.1 Study population

Patients

All children under 18 years of age, referred to the service as suspected IBD were eligible for inclusion. Plasma samples for periostin analysis were obtained from 144 paediatric patients. These included 132 consecutive patients from the point of recruitment to the main study (Genetics of Paediatric IBD), which started in October 2010. In twelve additional patients, recruited between January 2015 to January 2016, GI tissue was also analysed for periostin expression in addition to the plasma samples. All patients had a confirmed diagnosis of IBD as per the Porto diagnostic criteria[437].

Paediatric controls

Control samples for plasma were obtained from 13 paediatric subjects (under 18 years of age) without a diagnosis of IBD. These children were initially recruited as suspected IBD based on their symptom profile, however endoscopic and histological assessment of the gastro-intestinal biopsies performed showed no evidence of IBD. All controls remained well over 6 months follow-up period and were subsequently discharged from the paediatric gastroenterology service. In addition to plasma, GI tissue from 3 paediatric controls was used for immuno-histochemical staining.

Adult controls

Control samples for plasma extraction were also obtained from 25 healthy adult individuals with no history of any underlying immune abnormality. Individuals with a medical condition or those on immunosuppressive medications were excluded in order to ensure recruitment of controls with a minimally altered immune system. All individuals were aged over 18 years. Adult controls were not included for tissue analysis.

6.3.2 Biological specimens

Plasma samples

Whole blood was collected into commercially available heparinised tubes, transported and stored at room temperature for up to four hours pending plasma extraction. Plasma was separated after centrifugation of whole blood for 10 minutes at 2000 rpm and apportioned into 0.5ml aliquots for storage at -80 ^{o}C until analysis. Up to 0.5 ml of plasma per patient was utilised for periostin analysis.

DNA extraction

DNA extraction, sequencing and whole exome data processing was performed as described previously in chapter 3.

Colonic tissue

Up to four colonic biopsies were obtained per individual for periostin immuno-staining during routine diagnostic colonoscopies or whole thickness colonic specimens obtained during surgical resection of the bowel as part of treatment of IBD. Biopsies were taken two each from macroscopically inflamed colonic mucosa and two from areas of colonic mucosa which looked macroscopically normal.

6.3.3 Assessment of clinical parameters

Disease activity

Disease activity was assessed in all patients using internationally validated disease activity scoring tools; Paediatric Crohn's Disease Activity Index (PCDAI) for CD[438]

and Paediatric Ulcerative Colitis Activity Index (PUCAI)[439] for UC. The PCDAI score ranges between 0-100 (in remission <10, mildly active disease 10-29, moderate-severe ≥ 30) and the PUCAI between 0-85 (in remission <10, mildly active disease 10-34, moderate 35-64, severe ≥ 65)[438,439,491]. Extraction of clinical and laboratory data were performed through the hospital's electronic patient records.

Clinical outcome- surgical interventions

Patients undergoing surgical interventions due to strictures, stenotic bowel segments, severe disease unresponsive to conventional medical treatments and fistulising disease were identified and these outcomes evaluated against the plasma periostin levels and WES data. Surgical interventions including bowel resection, strictureplasty and balloon dilatation of strictures were used as proxy markers for intestinal fibrosis and impaired healing, based on the overarching role of periostin in fibrosis and tissue repair.

6.3.4 Plasma periostin measurement

Periostin concentrations in plasma were assessed in duplicate using sandwich ELISA technique (R & D DuoSet human periostin ELISA kit, R & D systems, Abingdon, UK). The assay was run according to the manufacturer's protocol. Briefly, 96-well microplate was coated with 100µl per well of the diluted capture antibody and incubated overnight at room temperature. Plates were washed three times with the wash buffer provided followed by incubation with reagent diluent for one hour. After a further wash cycle, plates were incubated with samples and standards for two hours at room temperature followed by incubation with detection antibody for 2 hours. The plates were next incubated for 20 minutes with streptavidin-HRP, followed by addition of substrate solution for a further 20 minutes. Colour development was stopped by the addition of 2N H_2SO_4 and the plates were read for optical density using a microplate reader set to 450nm. Sample readings were interpolated on standard curves using GraphPad Prism (V7) programmes, generating a logistic curve-fit and transforming values to concentration units (pg/ml).

6.3.5 Immuno-histochemical staining method

Immuno-staining for periostin was performed by Dr Eva Riethmacher in her laboratory at the Duthie building on the UHS site. Mucosal biopsies were collected in paediatric theatres during endoscopy by myself. The biopsy samples were immediately embedded in Optimal Cutting Temperature (OCT) embedding matrix (agar scientific, Essex, UK) in a labelled cryo-mold and transported in a covered foam cooler of dry ice for storage at -80 °C pending analysis. Anti-human periostin were used as the primary antibodies and cy3-labeled anti-rabbit antibody (abcam) as the secondary antibodies. Detailed steps of the immuno-staining procedure are included in the appendices.

6.3.6 Whole exome sequencing (WES) data analysis

WES data analysis was conducted using a novel gene-based score, 'GenePy' (previously described in chapter 5)[435], focusing on a list of selected genes functionally connected to the *POSTN* gene.

Gene selection

An electronic search was conducted through 'PathCards', as described previously in chapter 5, by entering the term 'periostin'. The 'SuperPath' nucleating from the search was used to select the genes for application of GenePy scores[443].

Application of GenePy scores

GenePy incorporates known deleteriousness metrics, allele frequency and individual zygosity information per gene to generate a whole gene-based pathogenicity score. The score integrates data for as many variants as are present in a given gene per individual following correction for the gene length[443]. In this study, GenePy scores were generated for a list of selected of genes functionally connected to periostin. The scores across the prioritised genes were used to assess differences in mutational burden between groups of individuals based on clinical outcomes and plasma periostin levels.

6.3.7 Statistical analysis

All numerical results were expressed as the mean +/- SEM. Statistical significance was assessed by regression analysis or student t-test, using GraphPad Prism version 7.

6.4 RESULTS

6.4.1 Cohort characteristics

The patient cohort included 65% males (n=93), median age= 14.2 years, CD= 71% (n=102) and UC= 29% (n=42). The median duration of follow up since diagnosis was 32 months. Paediatric controls included 13 individuals, 69% males (n=9), with a median age of 13.3 years (age range 4 to 16 years). The median age of adult controls was 24 years (age range 19- 40 years) and included 60% females (n=15).

6.4.2 Disease activity scores

PCDAI values were available for 69 out of the 102 CD patients at the time of sample procurement. PCDAI scores incorporate laboratory blood tests and were not available in patients who did not have routine clinical bloods tests at the time of sample procurement, or the blood tests did not include all parameters required for the generation of the scores. PUCAI scores for UC, which require only clinical symptom profile without blood tests were available in all patients. Of the 69 patients with CD where PCDAI scores were available, 70% (n=48) of patients had active disease (PCDAI ≥ 10), with a mean PCDAI score of 25. In UC, the proportion of patients with active disease (PUCAI≥ 10) was 50% (n=21) with a mean PUCAI score of 38.

6.4.3 Plasma periostin levels

Paediatric v/s adult controls: Plasma periostin levels were significantly higher in paediatric controls compared to adult controls [26824 pg/ml ± 1729 (SEM) versus 20503 pg/ml ± 1129, p= 0.003]. Due to these statistically significant differences, adult controls were not used in the subsequent analysis for interpretation of biological variations among patients.

Patients v/s paediatric controls: No significant differences were observed between pIBD patients and paediatric controls. See table **6.1.**

Table 6.1 Comparing disease activity against paediatric controls

Group	Numbers	Mean (pg/ml) +/- SEM	p value
Paediatic controls (PC)	13	26824 +/- 1729	.
CD (remission)	21	31338 +/- 2144	0.149
UC (remission)	21	28721 +/- 1379	0.399
All-remission (CD + UC)	42	30029 +/- 1276	0.204
CD (active disease)	48	25074 +/- 1108	0.452
UC (active disease)	21	25792 +/- 1570	0.672
All-active disease (CD + UC)	69	25292 +/- 901	0.49

The plasma periostin values in patients with active disease and in remission were compared against the paediatric controls using unpaired t-tests. PCDAI scores for assessing disease activity were only available in 69 CD patients (out of 102). PUCAI scores were available in all 42 patients with UC. No significant differences were observed between patients and controls.

Remission v/s active disease: Patients with IBD (both CD and UC together) showed significantly higher plasma levels during remission (n=42; CD-21, UC-21) compared to active disease (n=69; CD-48, UC-21) [remission v/s active disease: 30029 ± 1276 v/s 25292 ± 901, p=0.002]. This observation was driven by patients with CD [remission (n=21) v/s active (n=48): 31338 ± 2144 v/s 25074 ± 1108, p= 0.005], but not UC [remission (n=21) v/s active (n=21): 28721 ± 1379 v/s 25792 ± 1570, p= 0.168]. Linear regression showed an inverse correlation between plasma periostin levels and PCDAI (r^2= 0.1196, p=0.003) in CD. No association was observed between periostin levels and PUCAI scores in UC (r^2= 0.062, p=0.109). See figure **6.2.**

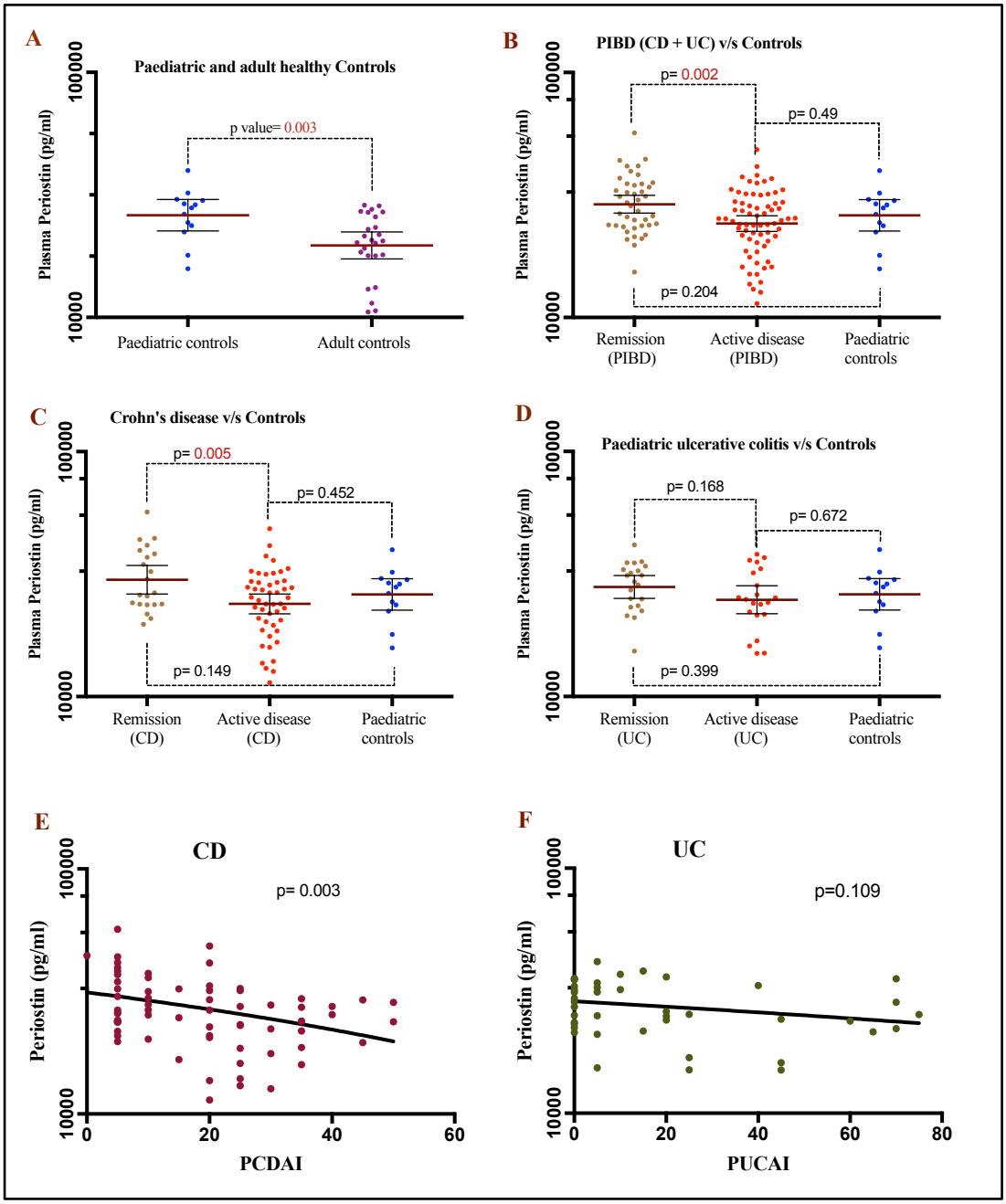


Figure 6.2 Plasma periostin levels against disease activity

Plasma periostin levels were plotted along the Y-axis. PCDAI scores for assessing disease activity in CD were available in 69 out of the 102 CD patients. PUCAI scores were available in all 42 patients with UC. A) Adult controls have significantly lower levels compared to paediatric controls. B) Patients in remission collectively show a significantly higher periostin levels compared to those with active disease. This was observed for CD (C), but not UC (D). E) Inverse relationship between periostin levels and PCDAI scores. F) No association between periostin levels and PUCAI scores.

Surgical v/s non-surgical: Periostin levels were compared between patients who had ever needed surgical interventions for their disease versus those who had not to date undergone surgical interventions. There were no significant differences in periostin levels in CD between the surgical (n=17) and non-surgical (n=85) groups (p= 0.449). Similarly, no significant differences were observed in UC patients between the surgical (n=6) and non-surgical (n=36) groups (p= 0.9338).

6.4.4 Immuno-histochemical staining in colonic biopsies

Immuno-histochemical staining for periostin was performed on colonic tissue in 15 individuals. Of these, 10 patients had CD (males=9), two patients with UC (one female) and three non-IBD controls (2 males). All specimens except one (PR0212), were obtained from treatment naïve individuals at diagnosis. Individual PR0212 was a patient with Crohn's disease who underwent surgical resection for an ileo-caecal stricture during the course of his disease. In this patient, immuno-staining was performed on the resected tissue, whereas in the rest of the patients, endoscopically biopsied mucosal tissue from the colon was used. See table **6.2.**

Table 6.2 Immuno-staining for periostin: patients and controls

	ID Number	Gender	Disease/ Control	GI Tissue	Tissue status on macroscopy	Tissue obtained at diagnosis or pre-diagnosis	Disease Scores	Disease status	Paris Classification (disease location)	Drugs
	Crohn's Disease									
1	PR0212	M	CD	Surgical*	Scarred tissue	No	10	Active	L3	AZA
2	SOPR0284	M	CD	Mucosal	Normal & inflammed	Yes	30	Active	L3	None
3	SOPR0288	F	CD	Mucosal	All inflammed	Yes	40	Active	L3L4a	None
4	SOPR0296	M	CD	Mucosal	All inflammed	Yes	25	Active	L3L4a	None
5	SOPR0303	M	CD	Mucosal	Normal & inflammed	Yes	30	Active	L3	None
6	SOPR0304	M	CD	Mucosal	All inflammed	Yes	25	Active	L2L4a	None
7	SOPR0310	M	CD	Mucosal	Normal & inflammed	Yes	25	Active	L3	None
8	SOPR0312	M	CD	Mucosal	Normal & inflammed	Yes	35	Active	L3	None
9	SOPR0313	M	CD	Mucosal	Normal & inflammed	Yes	35	Active	L3L4a	None
10	SOPR0320	M	CD	Mucosal	Normal	Yes	20	Active	.	None
	Ulcerative Colitis									
1	SOPR0311	M	UC	Mucosal	All inflammed	Yes	75	Active	E4	None
2	SOPR0314	F	UC	Mucosal	Normal & inflammed	Yes	70	Active	E4	None
	Controls									
1	SOPR0315	M	Control	Mucosal	Normal	Yes	N/A	N/A	N/A	None
2	SOPR0321	M	Control	Mucosal	Normal	Yes	N/A	N/A	N/A	None
3	SOPR0322	F	Control	Mucosal	Normal	Yes	N/A	N/A	N/A	None

Immuno-staining for periostin was performed in 15 individuals including 3 non-IBD controls. In all patients except one (PR0212), tissue was procured at diagnosis, before commencement

*of any treatment for IBD. PUCAI or PCDAI are disease activity scores for UC and CD respectively at the time of tissue procurement. AZA- azathioprine. *In individual PR0212, GI tissue was obtained during surgical resection for an ileo-caecal stricture. Disease location/s as per Paris classification for Crohn's disease include L1 (distal 1/3 of ileum), L2 (colonic), L3 (ileo-colonic) & L4 (upper GI disease; L4a- proximal to the ligament of Treitz, L4b- distal to the ligament of Treitz); UC E1(ulcerative proctitis), E2 (left-sided UC), E3 (extensive, up to hepatic flexure) & E4 (pancolitis).*

Periostin immunoreactivity was observed in the stroma surrounding the crypts in the form of 'peri-cryptal' rings. Differences in staining patterns were observed between the biopsies obtained from inflamed colonic mucosa from pIBD patients compared to non-IBD paediatric controls. The intensity of the peri-cryptal pattern was stronger in the inflamed mucosa as compared to a more diffuse pattern in the colonic mucosa obtained from controls. The change in expression pattern persisted throughout the colon of affected patients and could also be found in the non-inflammed parts of the colon. Intuitively, there were no discernible differences observed between the inflamed mucosa from patients with CD and UC. The immuno-staining on the surgically resected tissue in PR0212 showed very intense and diffuse staining, probably due to the presence of excessive fibrotic tissue in the strictured bowel segment. See figure **6.3.**

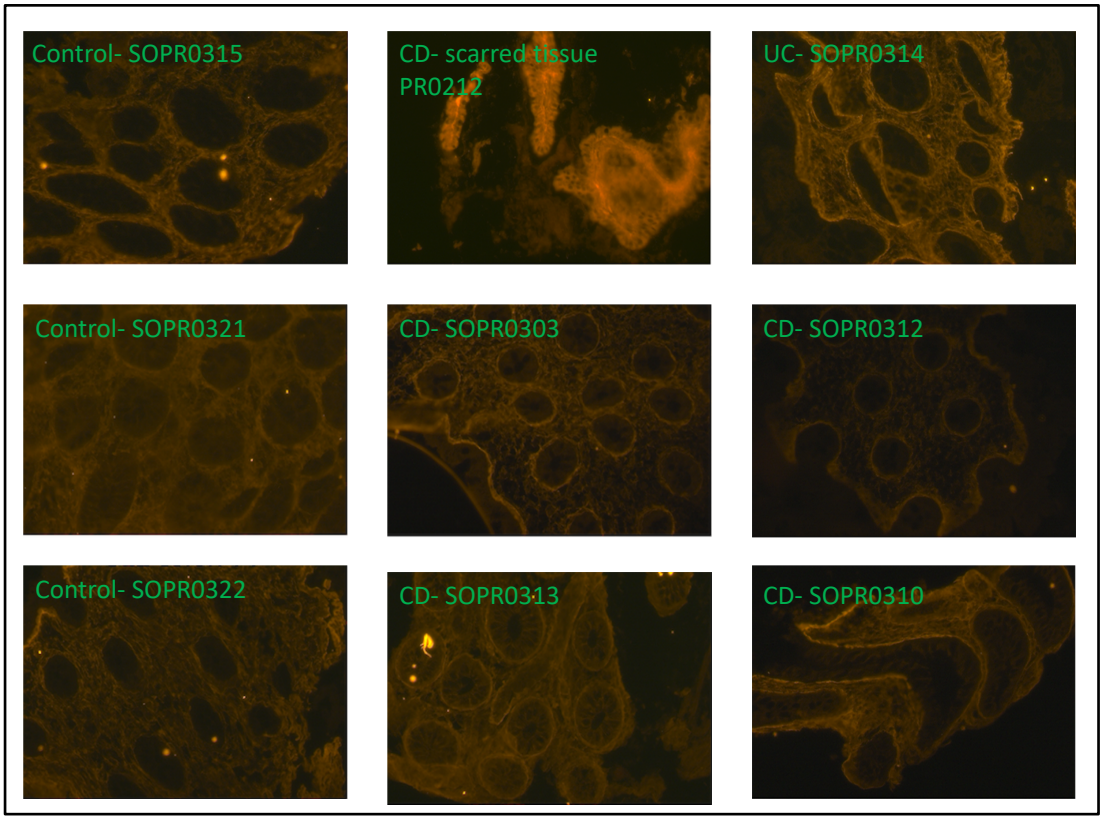


Figure 6.3 Immuno-histochemical staining for periostin in colonic tissue

The figure shows immuno-histochemical staining for periostin on colonic tissue obtained from controls and pIBD patients. The scarred and fibrotic tissue from a resected bowel specimen (PR212) shows intense staining for periostin. Patients with IBD show a prominent peri-cryptal pattern of staining compared to a more diffuse pattern in healthy controls.

6.4.5 WES data analysis

Selected genes

Interrogation through PathCards using the search term 'periostin' displayed two SuperPaths, 'Amplification and Expansion of Oncogenic Pathways as Metastatic Traits' and 'Hypothesized Pathways in Pathogenesis of Cardiovascular Disease', each comprising 17 and 25 genes respectively, including periostin. The former was selected based on its functional relevance to inflammatory pathologies, given the well-characterised roles of periostin in inflammatory and oncological diseases. Figure **6.4** shows the STRING network of the 17 genes, which were selected for further analysis and application of GenePy scores.

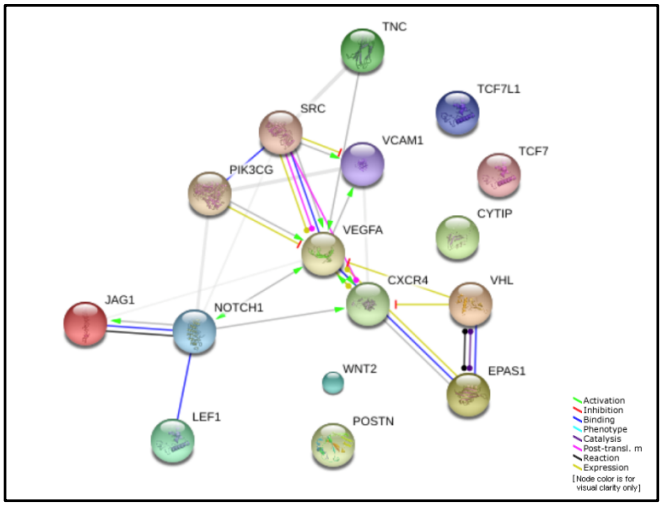


Figure 6.4 STRING network of genes for periostin

The figure shows a screenshot taken directly from PathCards page display of the SuperPath functionally connected to periostin. The SuperPath selected was 'Amplification and Expansion of Oncogenic Pathways as Metastatic Traits', which comprised 17 genes including periostin. The figure herein is a cluster representation displayed through the STRING network, a gene-network and cluster-generation tool used by PathCards for each of its SuperPaths. The protein nodes and the inter-connections within the retrieved network are automatically coloured for visual clarity and also to indicate the type of functional connectivity.

Application of GenePy scores

Exome data were available in 112 patients (CD-76, UC- 36 patients). GenePy scores were available for thirteen out of the seventeen genes selected. GenePy scores across the thirteen genes were combined per individual and the combined scores assessed between individuals needing surgical and those who did not need surgical treatments. This was performed to assess if there was a significant burden of deleterious mutations in the group needing surgical interventions, given the role of periostin in fibrosis and potentially stricture formation. Unpaired t-tests showed no significant differences between these two groups (p values; CD- 0.7639, UC- 0.9295). Furthermore, no correlation was observed between periostin levels and GenePy scores on regression analysis for each of the thirteen selected genes individually. See tables **E.1-E.2** in Appendix E for GenePy scores per individual in UC and CD respectively.

6.4.6 Summary of the results

- Significantly higher levels of plasma periostin were observed during remission compared to active disease.
- Statistically significant differences were observed in the plasma levels of periostin between paediatric and adult controls.
- Immuno-histochemical staining of the inflamed GI tissue in IBD showed an enhanced peri-cryptal ring pattern compared to a more diffuse staining pattern in GI tissue from non-IBD controls.
- No association was observed between plasma periostin levels and intestinal fibrosis, using surgical interventions as proxy markers for stricturing and fibrosing disease.
- No association was observed between plasma levels of periostin and gene-based pathogenicity scores (GenePy) for a list of prioritised genes functionally connected to the *POSTN* gene.

6.5 DISCUSSION

Although the role of periostin has been well described in a wide spectrum of inflammatory and neoplastic conditions[146,148,149,152,155], its function in the pathogenesis of IBD has not been well-characterised. To date, there have been only two publications highlighting the involvement of this protein in UC[150,151]. One of these studies[150] was published in 2016 during the generation of the results presented in this chapter. The functional role of periostin in CD has not yet been described. Furthermore, previous studies have indicated alterations in the plasma levels of periostin in several inflammatory and fibrotic conditions[487-490], but not IBD. This study presents data on a paediatric IBD cohort of 144 patients (including both CD and UC), assessing serum levels against disease activity, exome data and adverse clinical outcomes including stricturing disease. Additionally, the study describes findings of immuno-histochemical staining of colonic tissue for periostin from both CD and UC patients and comparing these with non-IBD paediatric controls.

In this study, significant differences in plasma periostin levels were observed between paediatric and adult healthy controls. It has been well-established through previous studies that although immune function can vary within a population at the extremes of

life, such early infancy or old age, immune responses remain stable across several decades in between[460,492,493]. Our study provides experimental evidence to support the use of age-matched controls for the biochemical assessment of this protein, in relevant research or clinical settings.

An inverse correlation was observed between plasma levels of periostin and disease activity for CD (PCDAI), with significantly higher levels observed during remission rather than active disease. The significance of the lower levels of plasma periostin observed during the active phase of the disease as compared to remission are unclear, but probably suggests a more prominent reparative role rather than inflammatory in IBD. Despite the seemingly disparate roles of periostin across health and disease, tissue re-modelling and repair have been shown to be the overarching functional roles of this extracellular matrix protein[155,481]. The currently used disease activity indices are based on clinical, haematological and biochemical values, and do not include endoscopic or histological parameters[438,439]. Clinical remission or resolution of symptoms often precedes mucosal healing. It is therefore possible that patients may continue to have ongoing mucosal repair-related processes in the bowel tissue even after clinical remission has been achieved. As periostin is a key protein involved in tissue repair, higher levels in circulation during remission may be secondary to ongoing mucosal repair. Tissue repair also involves an excess production of TGF-β, which upregulates periostin production, thereby contributing to the circulating pool of periostin during the repair phase or remission[156,481].

As the ECM proteins play crucial roles in tissue inflammation, immune responses and mesenchymal remodelling, altered serum levels have been observed in various disease states and used as biomarkers in monitoring disease progression[59]. For example, tenascins have been shown to be useful biomarkers in monitoring disease in malignancies, cardiovascular disease, arthritic inflammation and others[59,494-496]. Serum periostin has been described as a biomarker of disease progression in systemic sclerosis, airway inflammation in asthma and certain cancers[490,497,498]. In this study, no significant differences were observed between patients and paediatric controls. Therefore, the utility of serum periostin as a biomarker of inflammation in IBD could not be established. The control population in this study was small and included patients who were initially referred with GI symptoms, but subsequent investigations showed no evidence of IBD. Although the control subjects remained well over a 6-

months follow-up period and were discharged from the clinical service, it is possible that some these individuals may not represent 'true' healthy controls given their initial presentation with symptoms.

In our study, immuno-histochemical staining of the GI tissue showed an enhanced 'peri-cyptal' ring pattern in areas of inflamed mucosa compared to a more diffuse staining pattern in non-IBD controls. Our findings are in keeping with previously published literature[151]. *Kikuchi et al* demonstrated periostin deposition in the stroma surrounding the crypts, resulting in a peri-cryptal immunoreactivity pattern in normal colonic mucosa and an increased intensity of peri-cryptal staining in patients with UC (n=3). Findings of the study[151] indicated that in the bowel mucosa, periostin was primarily secreted by peri-cryptal fibroblasts located at the rim of the intestinal crypts. In our study, immuno-histochemical staining was conducted in both CD and UC, which showed an enhanced immuno-reactivity for periostin around the crypts. As a hypothesis generating paradigm, the intensity and staining pattern of periostin in the colonic mucosa could potentially complement histological assessment of the GI tissue as a proxy marker of inflammation. Further studies would be needed to explore the utility of this approach in a diagnostic setting.

Persistent inflammation and impaired healing in IBD lead to intestinal fibrosis, characterised by excessive deposition of matricellular proteins such as collagen, fibronectin, elastin, laminin and proteoglycans. Intestinal fibrosis leads to stricture formation, often needing surgical resection of the affected bowel. Approximately one third of patients with CD switch from an inflammatory to a fibro-stenotic phenotype over a period of ten years or longer[147,486,499]. Hypothetically, inadequate periostin expression during or post-inflammation, may result in failure of appropriate tissue remodelling and persistence of ulceration. Conversely, a sustained upregulation of periostin could drive remodelling beyond physiological adaptation, resulting in excessive ECM deposition, scar tissue formation and fibrosis[155]. In this study, surgical resection of bowel tissue due to stricturing disease was used as a surrogate index of intestinal fibrosis and compared against serum levels of periostin. No association was observed.

The contribution of mutations in the *POSTN* and other functionally connected genes in the pathogenesis of IBD remains unclear. In our study, assessment of the *POSTN*

and twelve other functionally connected genes using a novel genetic risk score 'GenePy', showed no association with plasma levels of periostin or with stricturing disease phenotype. Further research is however necessary to elucidate the underlying mechanisms involving periostin in the pathogenesis of IBD and identify a whole range of other interacting genes.

Our study represents an initial step in understanding the immunological and genomic basis of intestinal inflammation and repair in humans in the context of periostin. Although the molecular mechanisms remain obscure, the activation of signal transduction pathways such as the NF-κB via the αv-integrins receptors makes it a potentially attractive therapeutic target in future IBD management. Anti-integrin monoclonal agents such as vedolizumab have already been widely used in adult IBD treatment protocols with good outcomes[500]. Further research is needed to decipher the precise role of periostin in IBD pathogenesis, advance understanding and elucidate new targets for future therapies IBD.

Section IV

Chapter 7 *TPMT* AND OTHER GENES IMPLICATED IN THIOPURINE-INDUCED DRUG TOXICITY IN IBD

The work presented in this chapter was published in 2016 in a peer-reviewed journal. It provides an example of the application of next generation sequencing (NGS) in IBD therapeutics. The chapter examines the potential of NGS in predicting toxicity and clinical response to thiopurines, which are widely used as immunosuppressant drugs in the treatment of IBD. The study conceptualisation, design, extraction of clinical and biochemical data, and the interpretation of findings were performed by myself with input from supervisors. The application of gene-based statistical tests was performed by Gaia Andreoletti, a co-PhD student and the overall interpretation of exome sequencing data were performed with support from my primary supervisor Professor Sarah Ennis.

7.1 INTRODUCTION

Thiopurine drugs, which include azathioprine (AZA) and 6-mercaptopurine (6-MP) have been widely used in various auto-inflammatory conditions including inflammatory bowel disease for several decades[501]. As IBD is a chronic bowel condition with an unpredictable pattern of symptom flare-ups, the primary aim is to maintain prolonged remission through an effective immuno-suppressive treatment. The drug 6-MP was first synthesised in 1951 and its S-substituted precursor AZA in 1957, with the first reported use of these drugs in IBD in 1966[502-504]. Thiopurines have been used in several other conditions including different types of cancers, post-organ transplant regimens and a wide spectrum of inflammatory conditions with an underlying auto-immune basis. Although thiopurines are effective in maintaining remission in up to 70% of patients[505], concerns over adverse side-effects and toxicity continue to pose significant challenges, resulting in discontinuation of treatment in up to 40% of patients and a switch to other immunosuppressive agents[501,506-508]. Some of the known adverse effects include gastro-intestinal intolerance, liver toxicity, pancreatitis, flu-like illness, allergies, myelosuppression, increased susceptibility to infections and an increased risk of malignancies[508,509].

Thiopurines are purine analogues that interfere with DNA replication by acting as competitive antagonists, blocking proliferation of lymphocytes and other cells.

Although the precise mode of action is still not fully elucidated, much of the pharmacological activity is mediated through the incorporation of 6-thioguanine nucleotides (6-TGNs), resulting in impaired DNA replication and apoptosis of activated immune cells, particularly T-cells[508,510]. A key enzyme involved in the metabolism and degradation of these drugs is thiopurine S-methyl transferase (TPMT), which catalyses the S-methylation of aromatic and heterocyclic sulfhydryl groups within these compounds[511]. The *TPMT* gene located on chromosome 6p22, shows a wide-ranging heterogeneity among individuals and it has now been well established that variants within the gene can alter the functional activity of the enzyme, resulting in variability of therapeutic responses as well as toxicity following treatment with thiopurines[508,512,513]. Approximately 10 % of Caucasians are heterozygous for a mutant *TPMT* allele with consequent intermediate (sub-normal) enzyme activity and 0.33% either homozygous or compound heterozygous with a resultant deficient or absent enzyme activity[509,513]. In excess of 40 *TPMT* variants associated with an altered enzyme activity have been identified (http://www.imh.liu.se/tpmtalleles). As recommended by the *TPMT* nomenclature committee, in order to maintain a standardised system for all variants identified in humans, the gene and the variant are separated by an asterisk followed by an Arabic numeral. The most common mutant haplotype among Caucasians is *TPMT*3A*, which contains two nonsynonymous coding SNPs, rs1800460 (c.460G>A) and rs1142345 (c.719A>G) in exon 7 and exon 10 respectively. Among African and Asian sub-populations, *TPMT*3C (rs1142345, c.719A>G)* is the most common mutant haplotype, which contains only the exon 10 SNP[509,514]. See Figure **7.1.**

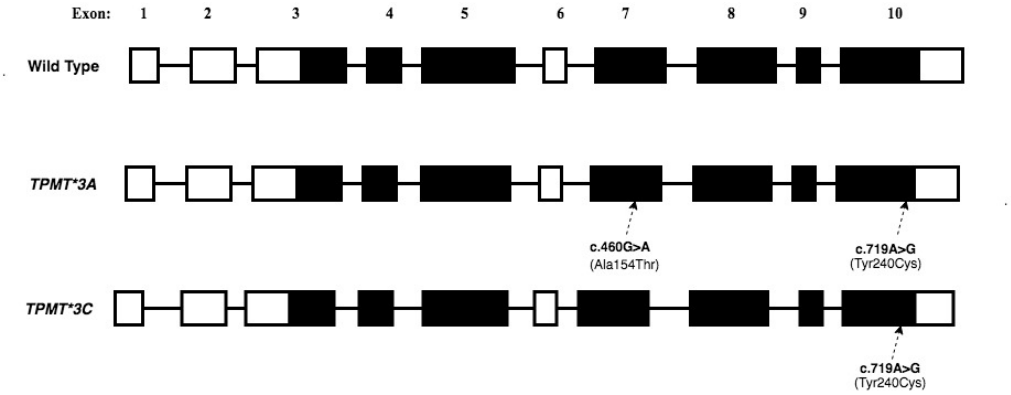


Figure 7.1 Wild-type and the most common *TPMT* mutant haplotypes

*The haplotypes TPMT*3A and TPMT*3C, which are the most common mutant haplotypes among Caucasians and African-Asian sub-populations respectively, are shown in the figure*

*along with the wild -type haplotype. The exons are displayed as black boxes, introns as connecting lines and the non-coding regions as open boxes. TPMT*3A carries two non synonymous SNPs in the exons 7 and 10, whilst TPMT*3C contains only the SNP in exon 10.*

In line with recommended treatment guidelines, the TPMT status of every patient is assessed before initiation of treatment with these drugs, in order to optimise drug dosage and minimise potential drug toxicity. This can be determined either by analysing the enzyme activity in the circulating erythrocytes or through genotyping known *TPMT* variants associated with deficient enzyme activity[35,508,515]. In an individual with deficient or very low enzyme activity, thiopurine drugs should be avoided and in those with an intermediate or low normal activity, thiopurines are prescribed at 50% or lower of standard doses[35,508,516].

Figure **7.2** provides a description of the pathway involved in the degradation of thiopurine drugs and their metabolites.

AZA and 6-MP are both inactive pro-drugs which are metabolised through three competing enzymatic pathways. Formation of the pharmacologically active metabolite, 6-thioguanine nucleotide (6-TGN) which is mediated through a series of steps which include conversion of AZA to 6-MP by the action of glutathione s-transferase, formation of 6-thioinosine monophosphate (6-TIMP) by hypoxanthine guanine phosphoribosyl-transferase (HGPRT) followed by phosphorylation to 6-thioguanine nucleotides (6-TGNs), which constitute the most crucial metabolites required to block DNA replication during cell proliferation. 6-MP is methylated to 6-methyl-MP (6-MMP) by TPMT, which also competes with inosine 5-monophosphate dehydrogenases (IMPDH 1&2) for their shared substrate 6-TIMP, converting it to 6-methyl-mercaptopurine nucleotides (6-MMPN). 6-TIMP can also be phosphorylated to 6-thioinosine triphosphate (6-TITP) through phosphokinases. The process can be reversed by ITPases, which dephosphorylate 6-TITP back to 6-TIMP.

Although a therapeutically desired profile favours 6-TGN production over 6-MMP, approximately 15% of patients preferentially metabolise thiopurines to 6-MMP which is potentially hepatotoxic. A third enzymatic pathway includes degradation of 6-MP to 6-thiouracil (6-TU) through the action of xanthine dehydrogenase (XDH), also known as xanthine oxidase (XO) or aldehyde oxidase 1 (AOX1)[508,517]. The XO inhibitor allopurinol has been shown to switch the metabolic pathway towards 6-TGN

production. This immuno-modulatory effect of allopurinol is sometimes used in clinical practice to improve the therapeutic efficacy of thiopurines by optimising 6-TGN levels in the blood[518].

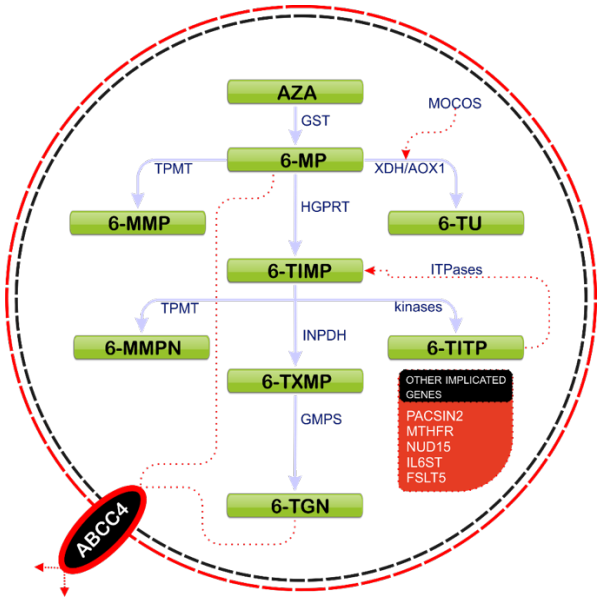


Figure 7.2 Thiopurines metabolic pathway

*Figure 7.2 shows the three main competing pathways involved in the metabolism of thiopurines following the conversion of AZA to 6-MP. These include: 1) phosphorylation to 6 thioguanine nucleotides (6-TGNs); 2) Degradation through methylation by the TPMT enzyme and; 3) Catabolism to 6-thiouracil (6-TU) by the action of xanthine oxidase (XO), XDH or AOX1 enzymes. The MOCOS enzyme sulfurates the molybdenum co-factor in XDH and AOX1. Also shown in the diagram are the other genes/proteins functionally connected to thiopurine metabolism (inset). These include PACSIN2, MTHFR, NUD15, IL6ST and FSLT5. (**Abbreviations:** ABCC4- ATP-binding cassette, sub-family C (CFTR/MRP), member 4; AOX1- aldehyde oxidase 1; AZA- azathioprine; FSLT5- follistatin-Like 5; GST- glutathione s transferase; HGPRT- hypoxanthine phosphor-ribosyltransferase;IL6ST- interleukin 6 signal transducer; 6-MP- 6- mercaptopurine; 6-MMPN- 6- methyl mercaptopurine nucleotides; MOCOS- molybdenum cofactor sulfurase; MTHFR- methyl-enetetrahydrofolate reductase ;NUDT15- nudix (nucleoside diphosphate linked moiety X)-type motif 15; PACSIN2- protein kinase C and casein kinase substrate in neurons 2; 6-TXMP- 6-thioxanthosine monophosphate); XDH- xanthine dehydrogenase.*

Although the methylation of thiopurines with TPMT is the most critical step in the detoxification pathway, several other genes have been identified which can impact on drug metabolism and alter therapeutic efficacy as well as the adverse reaction profile[519-526]. In this study, we sought to identify mutations in *TPMT* and the other

genes implicated in thiopurine drug toxicity by analysing exome sequencing data in a paediatric IBD cohort. The combined effects of common and rare variants within these genes were analysed using a gene based statistical test, sequence kernel association optimal unified test (SKAT-O) and assessed against the biochemically measured TPMT enzyme activity, therapeutic responses and drug-induced toxicity.

SKAT-O test

The phenotypic expression of complex diseases or traits can be influenced by the joint contribution of novel, rare, low-frequency as well as common variants at a given genomic locus[527]. Single variant tests are typically used to investigate associations between common variants and phenotypes. However, these are often underpowered for rare-variant analysis due to the large sample size needed to detect a significant association. Rare variant effects can be more efficiently analysed by focusing on the cumulative effects of rare variants in the genomic regions or genes. These tests for rare variant analysis can be broadly categorised into burden and non-burden tests[528]. Burden tests collapse rare variants at a given locus into a single burden variable and then regress the phenotype on the burden variable to test for the cumulative effects of rare variants in that region. Burden tests implicitly assume that all rare variants in a region are causal and affect the phenotype in the same direction with similar magnitude, resulting in a substantial loss of power when the assumptions are not true[529]. Kernel-based test methods such as the Sequence Kernel Association test (SKAT) are non-burden tests. Instead of aggregating variants, SKAT aggregates individual variant-score statistics with weights when SNP effects are modelled linearly and also allow for epistatic effects. SKAT can be particularly efficient when a genomic region has both protective and deleterious variants or many non-causal variants. However, the SKAT test can be less powerful than the burden tests if a large proportion of the variants are truly causal and affect the phenotype in the same direction. The SKAT-O test adaptively uses the data to find the 'Optimal' (O) linear combination of the burden test and SKAT to maximise power of detecting an association. The SKAT-O functions like the burden test when the burden test is more powerful than the SKAT and *vice versa*. By integrating both a burden test and a SKAT test, SKAT-O offers a powerful way of conducting association analysis on combined rare and common variants[292,528,530].

7.2 METHODS

7.2.1 Study population

Patients for this study were identified through the Wessex paediatric IBD research database, recruited as part of the wider study, 'Genetics of Paediatric IBD' and followed up through the course of their treatment. Clinical details were sourced from the electronic patient records of the University Hospital Southampton (UHS). Diagnosis of IBD was based on the modified Porto criteria for children and young adults. All patients were treated in line with the recommended national guidelines[39,437]. The study included one hundred paediatric patients (< 18 years of age at diagnosis), in whom biochemically measured TPMT enzyme activity levels and concurrent exome sequencing data were available.

7.2.2 Measurement of TPMT enzyme activity

TPMT enzyme activity levels for every individual were recorded from the electronic patient records. Biochemical measurements were performed by clinical laboratory staff using standard high-performance liquid chromatographic technique[531]. Based on the enzyme activity levels, patient sub-groups were defined as deficient (<10 mU/L), intermediate (10-67 mU/L), normal (68-150 mU/L) and high (>150 mU/L) in line with the UHS laboratory reference ranges and previously published literature[532].

7.2.3 Treatment with thiopurines and monitoring for drug toxicity

Current clinical guidelines recommend thiopurines as maintenance therapy for patients who relapse within six months of diagnosis, following 2 or more steroid courses in 12 months, extensive disease at diagnosis, steroid-dependent patients and to enhance the efficacy of anti-TNF alpha agents by reducing the immunogenicity of biological agents[37,39]. Regular blood tests are recommended as part of drug toxicity monitoring including a full blood count to assess for myelosuppression, liver function tests for hepatotoxicity and serum amylase or lipase for pancreatitis. Myelosuppression is defined as leucopoenia (WBC < 3000 mm^{-3}), neutropenia (neutrophils < 1000 mm^{-3}), lymphopenia (lymphocytes < 500 mm^{-3}) or thrombocytopenia (platelets < 100, 000 mm^{-3}); hepatotoxicity as alanine transaminase

(ALT), gamma-glutamyl transpeptidase (GGT) or alkaline phosphatase more than twice the upper limit of normal; and acute pancreatitis as severe abdominal pain within three months of commencing treatment with thiopurines, accompanied by elevation of serum amylase or lipase level to greater than twice the upper limit of normal reference values[515,516,533].

7.2.4 Response to treatment with thiopurines

The criteria defining poor clinical response or non-response included any of the following: 1) Symptom relapse within six months of therapy needing additional therapeutic interventions; 2) Inability to achieve remission within six months as assessed by global clinical assessment, laboratory tests or endoscopic evaluation; and 3) Steroid dependence for maintenance of remission. Drug intolerance was defined based on all of the following: 1) Observed adverse side effects upon commencement of therapy (myelosuppression, derangement of liver biochemistry, pancreatitis, gastro-intestinal intolerance, severe headaches, persistent flu-like illness or dermatological manifestations); 2) Observed side-effects unexplained by the natural course of the disease or other co-existing conditions; and 3) Partial or complete resolution of the adverse side effects upon cessation or reduction of treatment.

7.2.5 Search strategy for identifying genes involved in thiopurine toxicity

A literature search was conducted through ovidsp using MEDLINE and EMBASE from inception to the end of March 2015 to identify studies in humans describing *TPMT* and other genes involved in thiopurine metabolism and toxicity. The *TPMT* nomenclature committee website (http://www.imh.liu.se/tpmtalleles) that describes all clinically relevant *TPMT* variants was also applied, to cross-reference against variants identified in this study. Animal studies, review articles with no original data and conference abstracts were excluded. Search results were screened first by title and abstract, followed by full text if relevant.

7.2.6 Assessing burden of mutations

In this study, the SKAT-O test was applied to ascertain the cumulative effect of common and rare variations within the selected genes[292,530]. The test was performed to

assess the impact of mutations on TPMT enzyme activity, tolerance to thiopurine drugs and response to treatment. To conduct the test, a group file of mutations (non-synonymous, synonymous, splice-site variants, coding indels, stop gain and stop loss mutations) was created for the selected genes. SKAT-O was executed with the small sample adjustment and applying a minor allele frequency (MAF) threshold of 0.05 to define rare variants and using default weights. The EPACTS software package[534] was used to conduct this test.

7.2.7 Statistical tests

Statistical analysis was performed using GraphPad Prism version 6. The frequency distribution of the TPMT enzyme activity between the pIBD group and the Wessex paediatric population was compared using unpaired student t-test. Fisher's exact test was used to compare the biochemical method of TPMT enzyme activity assessment and the NGS approach in predicting toxicity to thiopurine drugs.

7.3 RESULTS

7.3.1 Patients

Of the hundred patients identified, seventy-eight were treated with thiopurines, while the remaining twenty-two maintained remission without thiopurine therapy. Of the patients commenced on treatment, 53% were male (n=41), CD 69% (n=54), UC 22% (n=17) and IBDU 9% (n=7). The median duration of follow up for patients commenced on thiopurines was 46 months (7-156) from diagnosis and 43 months (6-119) from starting treatment (Table **7.1**).

Table 7.1 Overview of cohort characteristics and TPMT enzyme activity

Clinical Category		Number of patients within each group	Median duration of follow-up in months	Males (%)	Disease			TPMT biochemical activity Number of patients across groups (%)			
					CD (%)	UC (%)	IBDU (%)	Low activity (<10 mU/L)	Intermediate activity (10-67 mU/L)	Normal activity (68-150 mU/L)	High activity (>150 mU/L)
Not Treated with thiopurines		22	60 (10-156)	15 (68)	13 (59)	6 (27)	3 (14)	0 (0)	12 (54)	10 (46)	0
Intolerant		14	42 (9-82)	10(71)	12 (86)	2 (14)	0	0 (0)	8 (57)	6(43)	0
Tolerant	Responders	51	52 (7-124)	24 (47)	31 (60)	13 (26)	7 (14)	0 (0)	22 (43)	28 (55)	1 (2)
	Non-responders	13	63 (13-126)	7 (54)	11 (85)	2 (15)	0	0 (0)	10 (77)	3 (23)	0
	Total	100	54 (7-156)	56 (56)	67 (67)	23 (23)	10 (10)	0 (0)	52 (52)	47 (47)	1 (1)

7.3.2 TPMT enzyme activity in erythrocytes

The biochemically measured TPMT enzyme activity across the cohort of one hundred pIBD patients showed a bimodal distribution. See figure **7.3**. The distribution was compared against an independent control group of paediatric patients (aged less than 18 years, n=524) from the Wessex region who had their TPMT enzyme activity assessed for other clinical diagnoses through the UHS laboratory. A statistically significant difference in the distribution was observed between the two groups ($P = 7.69 \times 10^{-12}$). While approximately 50% of our cohort samples had TPMT measurements within the normal range and 50% within the intermediate range, almost 80% of the control samples had their enzyme activity falling within the normal range. This observation could be due to an ascertainment bias within our study samples as exome sequencing was preferentially performed in pIBD patients with the most severe disease phenotype, resulting in enrichment for patients with inadequate response to first-line therapy.

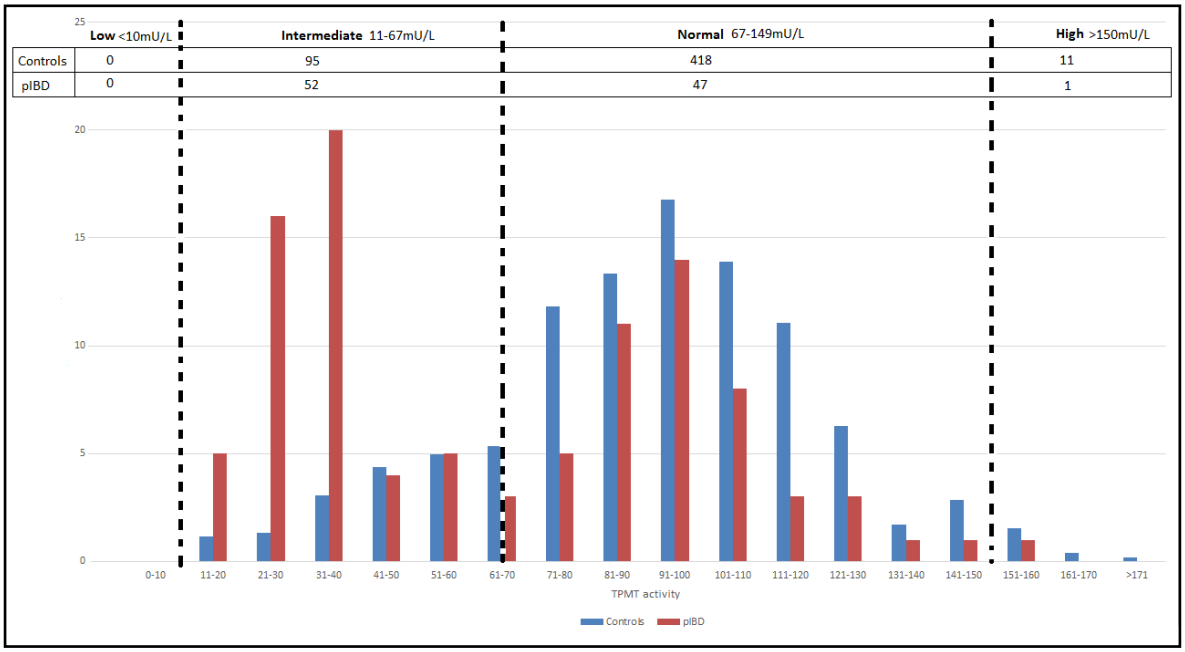


Figure 7.3 Frequency distribution of TPMT enzyme activity for the Wessex paediatric population and the study cohort

The blue bars indicate TPMT enzyme activity distribution for 524 paediatric patients with a wide range of diagnoses including IBD in the Wessex region (≤ 18 years, Dec 2010- April 2015) and the red bars indicating frequency distribution in our research cohort. A statistically significant difference between the sub-category with TPMT values between 21-40 units (p=0.001) was observed within our research cohort compared to the Wessex paediatric

population.

7.3.3 Genes implicated in thiopurine toxicity

Fifteen genes in addition to *TPMT* were identified with robust evidence from a review of over 3,000 publications (Appendix- table **F.1**). Variation in these genes was included in downstream analysis of: 1) biochemically measured TMPT enzyme activity; 2) tolerance to thiopurines and; 3) clinical response.

7.3.4 Correlation of *TPMT* gene variants with biochemical enzyme activity, thiopurine tolerance and response

Five *TPMT* variants were identified across the cohort (Table **7.2**).

Novel non-synonymous variant

One novel non-synonymous variant (p.A73V) was identified in a female patient with Crohn's disease, who had a sub-normal enzyme activity level of 55 mU/L. *In silico* analysis indicated this variant was strongly conserved (Phylop = 0.99) and likely to have a deleterious impact on the protein function (GERP = 4.98). Despite a reduced dose of AZA at 1mg/kg, this individual developed severe intractable nausea resulting in discontinuation of treatment. Our findings were reported to the *TPMT* nomenclature committee (http://www.imh.liu.se/tpmtalleles) and this variant has since been catalogued as a functionally significant allele with a unique allele number *TPMT**39.

Known *TPMT* variants associated with thiopurine toxicity

These included two non-synonymous variants (A154T and Y240C) previously known to impact TMPT function that were found to co-segregate in 9 individuals. The mean of biochemically measured TMPT enzyme activity for these 9 individuals was 36.6 mU/L, which was significantly different compared to the rest of the 91 individuals (mean = 68.1 mU/L) without these mutations (p=0.003). Although exome data are insufficient to resolve haplotypes, we postulate these two variants occur on a single haplotype in all nine individuals, consistent with the complex *TPMT* allele referred to as *TPMT*3A*, previously described as the most prevalent *TPMT* allele associated with

enzyme deficiency in Caucasians[535]. Of these nine patients, eight were commenced on treatment with thiopurines of which four subsequently developed intolerance and the treatment was discontinued. The other four patients who had the treatment were tolerant to the drug and showed a favourable clinical response.

Other *TPMT* variants

One patient with ulcerative colitis harboured a rare intronic variant (in heterozygote form) proximal to the exon 7 splice junction (c.420-4G>A). This patient concurrently had biochemical enzyme activity level of 32 mU/L, but showed tolerance as well as clinical response to thiopurine treatment. The MaxEnt score (0.95) for this splice junction variant predicted minimal impact on splicing within the gene. While this is positively consistent with drug tolerance, it does not explain the relatively low biochemical activity level observed in this patient. Finally, a common synonymous variant (I158I) at high frequency (37 heterozygotes and 57 homozygotes) was observed, with a mean enzyme activity level of 67.1 mU/L (32-99) and 66 mU/L (152-145) across the subgroup of individuals carrying the variant in heterozygous and homozygous state respectively. The very common frequency of this synonymous variant and the fact it does not alter amino acid sequence or composition is consistent with silent mutation having no impact on protein function.

Table 7.2 TPMT variants observed in the research cohort

position in hg19	variant	Coding change	Protein change	Phylop	GERP	MaxEnt	Frequency in 1000 genome	Genotypes in whole cohort (n=100)	Genotypes in intolerant (n=14)	Genotypes in tolerant (n=64)	Mean *TPMT* biochemical value (n=100)
18148069	ns	c.218C>T	p.A73V	0.99	4.98	.	.	99,1,0	13;1;0	64;0;0	55
18139272	sp	c.420-4G>A	.	.	.	0.95	0.0005	99,1,0	14;0;0	63;1;0	32
18139228	**ns**	**c.460G>A**	**p.A154T**	**0.22**	**0.93**	**.**	**0.02**	**91,9,0**	**10;4;0†**	**60;4;0†**	**36.6**
18130918	**ns**	**c.719A>G**	**p.Y240C**	**0.99**	**5.13**	**.**	**0.05**	**91,9,0**	**10;4;0†**	**60;4;0†**	**36.6**
18139214	sn	c.474C>T	p.I158I	.	.	.	0.77	6,37,57	1;5;8	3;25;36	67.4

Five TPMT variants were identified across the cohort. These included two non-synonymous variants (A154T and Y240C) previously known to impact TMPT function and were found to co-segregate in 9 individuals. A highly pathogenic novel variant was identified in an

*individual intolerant to thiopurines. The other two included a splicing variant and a common synonymous variant. For all groups, genotypes are listed as homozygous reference allele, heterozygous and homozygous alternative allele. Variants known to be associated with intermediate or low TPMT enzyme activity levels are shown in bold. Novel variant is highlighted in grey.† One out of the nine individuals harbouring p.A154T and p.Y240C was not commenced on thiopurine drugs. (**Abbreviations**: ns- non-synonymous; sn- synonymous; sp- splicing*)

7.3.5 Assessing burden of mutations within the *TPMT* and other selected genes

Exome data were analysed across fifteen additional genes implicated in thiopurine metabolism and toxicity. Mutations were identified across eleven genes (no variation in *HPRT1*, *GSTM1*, *FSLT5* and *MTHFR*). SKAT-O test was applied to assess the combined effect of rare, low frequency and common variations within these genes on TPMT enzyme activity, drug tolerance and clinical response. Significant evidence for association was observed within *MOCOS* (p=0.0015) and *TPMT* (p=0.0017) gene with TPMT enzyme activity levels. The test also detected a nominal association between *GMPS* and drug tolerance (p=0.0212) as well as variations within *IL6ST* (p=0.0084) & *ABBC4* (p=0.0452) and drug response. See Table **7.3.**

Table 7.3 SKAT-O test across TPMT and other genes involved in thiopurine toxicity

Chr	Lbp	Rbp	Gene	Total number of samples	Fraction of individuals who carry rare variants under the MAF thresholds (MAF < 0.05)	Number of all variants defined in the group file	Number of variants defined as rare (MAF < 0.05)*	p values
Biochemical activity (52 with TPMT value < 67 and 48 with TPMT value > 67)								
18	33767568	33848581	MOCOS	100	0.12000	12	5	0.0015
6	18130918	18148069	TPMT	100	0.10000	5	2	0.0017
Tolerance (14 intolerant and 64 tolerant)								
3	155588592	155654236	GMPS	78	0.12821	5	5	0.0212
Responses (51 Responders and 13 non-responders)								
5	55231311	55272085	IL6ST	64	0.10937	8	3	0.0084
13	95696540	95953517	ABCC4	64	0.31250	22	9	0.0452

*The SKAT-O test was applied to assess the joint effect of common, rare and low frequency variants on TPMT enzyme activity, tolerance and response to the drug. Only genes with significant evidence for association withstanding multiple testing correction are shown in the table. * These variants received different weights in the SKAT-O joint test. (**Abbreviations**: ABCC4, ATP-binding cassette, sub-family C (CFTR/MRP), member 4; GMPS, guanine monophosphate synthase; IL6ST, interleukin 6 signal transducer; Lbp, left base pair; MAF, minor allele frequency; MOCOS, molybdenum cofactor sulfurase; Rbp, right base pair)*

7.3.6 Drug tolerance

Thiopurine drugs were discontinued in 14 individuals due to adverse effects. Eight of these patients had TPMT enzyme activity in the intermediate category and the rest had TPMT activity in the normal range. Four of the eight individuals with TPMT enzyme activity in the intermediate range had known *TPMT* mutations associated with deficient enzyme activity and one individual had a novel *TPMT* mutation described in the previous section. In the rest of the nine individuals, there was

enrichment for potentially deleterious variants within the *MOCOS* gene and the *AOX1* gene (table **7.4**).

Gene	Chr	Position on hg19	Variant type	Coding change	Protein change	Novel	Phylop,	1-sift	Polyphen2	Mutationtaster	Gerp++	dbSNP	Frequency in 1000 genome	Frequency in EVS															
															TPMT biochemical activity	int	int	int	int	int	int	int	int	Normal	Normal	Normal	Normal	Normal	Normal
															TPMT activity	16	21	47	52	55	18	26	30	80	83	93	113	114	128
															Diagnosis	CD	CD	CD	CD	CD	CD	UC	CD	CD	CD	CD	CD	UC	CD
															Gender	M	M	M	M	F	F	M	F	F	M	M	M	M	M
															IDs	1	3	6	7	8	2	4	5	9	10	11	12	13	14
TPMT	6	18130918	ns	c.719A>G	p.Y240C	.	0.998258	0.94	0.94	0.999973	5.13	rs1142345	0.05	0.041715		1	1	1	1	0	0	0	0	0	0	0	0	0	0
TPMT	6	18148069	ns	c.218C>T	p.A73V	NOVEL	0.999474	0.95	0.898	0.999898	4.98	.	.	.		0	0	0	0	1	0	0	0	0	0	0	0	0	0
MOCOS	18	33779705	ns	c.G359A	p.S120N	.	0.961984	0.98	0.284	0.013289	2.79	rs3744900	0.06	0.047093		0	0	0	0	1	0	0	0	0	0	1	0	0	0
MOCOS	18	33831189	ns	c.C2107A	p.H703N	.	0.998524	0.58	0.706	0.010571	4.98	rs594445	0.25	0.283953		1	0	0	0	0	1	2	0	1	0	1	2	1	0
MOCOS	18	33848581	ns	c.T2600C	p.V867A	.	0.998597	1	0.063	0.006647	5.69	rs1057251	0.08	0.116279		1	0	1	0	0	1	0	1	0	0	0	0	0	0
XDH	2	31572983	ns	c.G2738A	p.R913Q	NOVEL	0.999016	1	1	1	5.38	.	.	.		0	0	0	0	0	1	0	0	0	0	0	0	0	0
XDH	2	31590917	ns	c.A2107G	p.I703V	.	0.998578	1	0.336	0.999913	4.52	rs17011368	0.05	0.034186		0	0	0	0	1	0	0	0	0	0	0	0	0	0
XDH	2	31611143	ns	c.G514A	p.G172R	.	0.998995	0.99	0.004	0.99994	5.35	rs45523133	0.04	0.026047		0	1	0	0	0	0	0	0	0	0	0	0	0	0
XDH	2	31621523	ns	c.A349T	p.T117S	.	0.998756	0.96	0.997	0.999999	5.81	.	.	.		0	0	0	0	0	0	0	0	0	0	0	2	0	0
AOX1	2	2.02E+08	ns	c.A3404G	p.N1135S	.	0.996285	0	0	2.70E-05	5.3	rs55754655	0.11	0.129767		0	0	0	1	0	1	0	0	1	1	0	0	0	1
IL6ST	5	55264153	ns	c.G442C	p.G148R	.	0.995114	0.98	0	2.70E-05	5.6	rs2228044	0.19	0.119535		0	0	1	1	0	0	0	0	0	1	0	0	0	0
NUDT15	13	48619855	ns	c.C415T	p.R139C	.	0.998597	0.92	0.057	0.937718	5.04	rs116855232	0.04	0.002442		0	0	1	0	0	0	0	0	0	0	0	0	0	0
ITPA	20	3193842	ns	c.C94A	p.P32T	.	0.998747	0.87	0.131	0.996149	5.14	rs1127354	0.08	0.067674		0	0	0	0	0	0	0	1	0	0	0	2	0	0
PACSIN2	22	43280404	ns	c.C773A	p.S258Y	.	0.999724	1	0.998	0.999962	5.16	rs200427054	0.0005	0.00314		0	0	0	0	0	0	0	1	0	0	0	0	0	0

Table 7.4 Deleterious variants occurring in individuals with intolerance to thiopurines

Table 7.4 shows patient variant profiles across TPMT and other genes implicated in thiopurine toxicity for the fourteen patients who developed intolerance. Five out of the fourteen patients had deleterious variants in the TPMT gene, including a novel non-synonymous variant. In the rest of the nine individuals, there was enrichment for deleterious variants in the MOCOS and AOX1 gene. The biochemical enzyme activity levels and the patient demographics are shown in the top right of the table.

7.3.7 Prediction of thiopurine toxicity- NGS versus biochemical test

Prediction of thiopurine toxicity showed a specificity of 93.75% through detection of *TPMT* risk variants compared to 50% for the biochemical test. Both tests had a low sensitivity of 37% and 57% respectively for predicting toxicity (Table **7.5**). However, all five patients with deleterious variants within the *TPMT* gene detected through NGS were also identified as potentially intolerant through the biochemical test. Although the biochemical test identified a higher number of individuals intolerant to thiopurines, the difference between the two approaches in predicting toxicity was statistically not significant (p value=0.45, Fisher's exact test). The clinical data of individuals intolerant to thiopurines, is shown in table **7.6**.

Table 7.5 Sensitivity and specificity for drug tolerance

		Intolerant	Tolerant	Sensitivity	Specificity
Biochemical test	+*	8	32	57.14%	50%
	−	6	32		
Deleterious variants in the *TPMT* gene	+	5	4	35.70%	93.75%
No deleterious variants in the *TPMT* gene	−	9	60		

*NGS had a specificity of 93.75% for predicting drug-induced toxicity compared to 50% for the biochemical test. The sensitivity obtained through both methods was sub-optimal. *Indicates that the biochemical test was positive for detecting intolerance (TPMT enzyme activity level<67).*

Table 7.6 Individuals intolerant to thiopurines

Patient ID	Diagnosis	Age at Diagnosis (years)	Gender	*TPMT* Gene Deleterious Variants	TPMT Activity	TPMT Value	Paris Classification (disease location)	Follow up in (months)	Thiopurine Drug	Median Dose (mg/Kg)	Adverse Effects
Crohn's Disease											
PR0011	CD	15	M	Yes	Intermediate	21	L1	14	AZA	1.5	Leuco-encephalopathy
PR0018	CD	13.5	F	No	Intermediate	18	L2L4	42	AZA	1.5	Neutropenia
PR0194	CD	11	M	No	Normal	83	NA	25	AZA	2	Persistent Nausea
PR0040	CD	14.5	F	No	Intermediate	30	L2	63	AZA	2	Persistent Nausea
PR0078	CD	5	M	No	Normal	93	L3	32	AZA	1	Persistent Nausea
PR0074	CD	10.9	M	Yes	Intermediate	16	L3	75	AZA	1.5	Pancytopenia
PR0076	CD	9.5	M	Yes	Intermediate	52	L2L4	75	6-MP	1	Persistent Nausea
PR0199	CD	12	M	Yes	Intermediate	47	L3	27	AZA	1.5	Abnormal ALT
PR0145	CD	16	F	No	Normal	80	L1	32	AZA	2.5	Persistent Nausea
PR0150	CD	14	M	No	Normal	128	L1L4	23	AZA	2	Persistent Nausea
PR0241	CD	15.5	M	No	Normal	113	L3	9	6-MP	1.5	Pancreatitis
PR0151	CD	13.5	F	Yes	Intermediate	55	L2L4	45	AZA	1	Persistent Nausea
Ulcerative Colitis											
PR0039	UC	10.5	M	No	Intermediate	26	E4	82	AZA	2	Elevated amylase
PR0077	UC	11.7	M	No	Normal	114	E4	47	AZA	1	Abnormal ALT

Fourteen patients within the study cohort were intolerant to thiopurine drugs prompting discontinuation of the drug. The predominant symptom of drug-induced toxicity included persistent nausea seen in 50% of patients. Biochemically measured TPMT enzyme activity was normal in 40% of the patients intolerant to thiopurines. Disease location/s as per Paris classification for Crohn's disease include L1 (distal 1/3 of ileum), L2 (colonic), L3 (ileo-colonic) & L4 (upper GI disease) and for UC E1(ulcerative proctitis), E2 (left-sided UC), E3 (extensive, up to hepatic flexure) & E4 (pancolitis).

7.3.8 Summary of the results

- *TPMT* variants of known functional importance in thiopurine-induced toxicity were detected in 9% of the cohort, in keeping with previously published literature[535].
- A novel and highly pathogenic variant not detectable through standard diagnostic genotyping was identified through NGS in an individual with clinical toxicity to thiopurine drugs and sub-normal TPMT enzyme activity.
- A statistically significant association between the *MOCOS* gene and TPMT enzyme activity was detected, not previously reported.
- Whilst the specificity for predicting drug toxicity through the biochemical test and NGS was 50% and 94% respectively, the sensitivity through both methods was sub-optimal.
- Although NGS has the ability to detect rare variants, not otherwise detectable through standard genotyping, there is no clear advantage over the biochemical test in predicting thiopurine-induced toxicity, observable in this small study.

7.4 DISCUSSION

Current clinical guidelines recommend assessment of TPMT status (enzyme activity or genotyping) in all patients prior to starting thiopurine treatment in order minimise the risk of adverse effects. This is performed by measuring the TPMT biochemical enzyme activity or through genotyping of known *TPMT* variants[35,508,516]. However, anticipating toxicity to thiopurine drugs through assessment of TPMT phenotype alone or in conjunction with *TPMT* genotype may be inadequate in a subset of patients. Up to 75% of patients on thiopurines can develop leukopenia despite a normal *TPMT* genotype. A normal genotype cannot exclude the possibility of TPMT enzyme deficiency or drug-induced toxicity. Conversely, drug toxicity may occur despite a normal TPMT enzyme activity potentially due to effects of other genes in the metabolism pathway[508,536,537].

The advantages of assessment of the TPMT status in a given individual through biochemical enzyme activity (phenotype) include the low long-term and circadian intra-individual variability in red blood cell TPMT activity[538]. When compared to

genotype testing, biochemical assessment is a better reflection of the current metabolic state. However, the enzyme activity can be altered by several factors such as recent blood transfusions, drug interactions and uraemia. Genotyping, unlike phenotype testing is not influenced by these factors[508,539]. However, clinically available *TPMT* genotyping analyses only a limited panel of commonly described *TPMT* variants and is likely to miss rare variants of functional significance in individual patients[540]. As a widely used practical approach, *TPMT* genotyping is considered when the TPMT phenotype suggests a deficiency or if a patient has recently been transfused[508].

Nine patients (9%) with the known *TPMT* mutations were identified in this cohort, consistent with previously published data reporting a prevalence of approximately 10% in the general population[535]. All nine patients had their biochemically measured TPMT enzyme activity levels in the sub-normal range in line with expectations, with a statistically significant difference in mean values compared to patients without these mutations. However, 43% of individuals with enzyme activity in the sub-normal range did not have the known *TPMT* mutations. Results of this study suggest that although prediction of thiopurine toxicity through NGS has a higher specificity compared to the biochemical test, the sensitivity of both methods is clinically suboptimal (35.7% and 57.14% respectively). Among the 14 individuals intolerant to thiopurines, all 5 patients who harboured deleterious *TPMT* variants would also have been predicted as potentially intolerant to thiopurines through the biochemical test. Furthermore, all the nine individuals with deleterious variants within the *TPMT* gene across the cohort of 100 patients would also have been identified as potentially intolerant as all these nine individuals had TPMT enzyme activity levels within the intermediate range. Hence based on first principles, NGS did not have a clear advantage over the biochemical test in predicting thiopurine toxicity.

We identified a highly pathogenic novel variant in *TPMT* in an individual who had TPMT enzyme activity level of 55mU/L, who developed severe gastrointestinal toxicity despite a reduced dose. This suggests that thiopurine toxicity can develop in individuals harbouring rare or yet unknown variants, not detected through standard genotyping. Exome sequencing therefore can be a powerful tool in identifying

individuals at risk of toxicity who could have been missed through standard *TPMT* genotyping.

The SKAT-O test identified a significant association between variations within the *MOCOS* gene and TPMT enzyme activity (p=0.0015). Molybdenum cofactor sulfurase (*MOCOS*) is a protein-coding gene located on 18q12, which sulfurates the molybdenum cofactor in XDH and AOX1, key enzymes involved in the degradation of thiopurines[541]. [541]. Kurzawski et al observed that patients with mutations in the *MOCOS* gene needed significantly lower doses of thiopurines for immune tolerance following kidney transplantation. A multivariate analysis showed that these mutations could influence drug dosages independent of other genetic mutations, similar to *TPMT* heterozygosity[522]. Although previous studies have suggested a role for *MOCOS* gene in thiopurine metabolism with possible impact on clinical outcomes in patients with mutations, an association between *MOCOS* and TPMT enzyme activity has not been described. Our study demonstrated that variations within this gene could significantly impact on the levels of the TPMT enzyme activity and therefore could potentially influence responses to thiopurine therapy. Further research is required to identify precise molecular mechanisms describing how *MOCOS* gene can impact on *TPMT* function, and response to thiopurine treatment.

A nominal association between *GMPS* and drug tolerance (p=0.02) was also identified in this study. *GMPS* is involved in the phosphorylation of 6-TIMP (6-thioinosine monophosphate) to 6-TGN, a key step to block DNA synthesis and replication of lymphocytes. Further research is required to identify the precise molecular mechanisms characterising the role of these genes within the thiopurine metabolic pathway.

In this study, although genotypic comparison across the individuals showed no significant evidence for association with regards to tolerance, we observed a significant association between responders and non-responders within the tolerant group for two mutations the *ABCC4* (p.K304N) and *IL6ST* (p.L8V). These were significantly enriched in the non-responsive subgroup of patients. ATP-binding cassette sub-family C member 4 (ABCC4), also known as multidrug resistance-associated protein 4 (MRP4) is encoded by the *ABCC4* gene. ABCC4 functions as an efflux pump, capable of pumping a wide variety of endogenous and xenobiotic

organic anionic compounds such as 6-MP and 6-TGN out of the cell. Single nucleotide polymorphisms in *ABCC4* reduce function of the pump and result in thiopurine toxicity through intracellular accumulation of toxic thiopurine metabolites[524,542]. A logical extrapolation of this observation would mean that mutations in ABCC4 efflux pump would result in an increased thiopurine sensitivity or toxicity. In contrast to this observation, our finding with regards to the *ABCC4* variant identified in our patients was a poor response to thiopurines. A theoretical possibility could be a gain of function, promoting rapid clearance of thiopurine nucleotides from the white blood cells resulting in a sub-optimal response.

Interleukin 6 signal transducer *(IL6ST)* encodes a signal transducer protein shared by several cytokines including IL-6 and has been identified through a genome wide association study as a myelo-suppression susceptibility candidate gene following thiopurine therapy in IBD[543]. The role of this gene in the context of response to thiopurine therapy is yet unexplored.

One of the limitations of this study is that only children who had undergone exome analysis and also had concurrent TPMT enzyme activity levels biochemically measured were prioritised for the study. While approximately half of our cohort patients had TPMT activity in the intermediate range, only 20% of control samples had TPMT activity in the intermediate category. Exome sequencing was preferentially conducted on children with the most severe disease phenotype, which would have enriched for patients with poor response to first-line treatments. Furthermore, the study cohort was relatively small and predominantly of Caucasian ethnicity, thereby limiting the general applicability of the results. Our search strategy for selecting genes known to be associated with thiopurine-induced drug toxicity identified a list of fifteen genes in addition to *TPMT*. It is possible that potentially pathogenic variants in other genes may have been missed through this approach. However, extension of the panel of genes analysed for this study would compromise power in the highest priority candidate genes.

Anticipating toxicity in patients through assessment of biochemically measured TPMT enzyme activity alone or in conjunction with a *TPMT* genotype, may be inadequate in some patients. Therapeutic response to thiopurines, as well as intolerance, can be influenced by other genes involved in thiopurine metabolism, in

addition to *TPMT*. Although there is no clear advantage of NGS over the biochemical test in predicting toxicity, our study demonstrates the strength of NGS as a powerful tool in identifying pathogenic variants in patients not detected through standard genotyping. As high throughput sequencing becomes more affordable and increasingly accessible, a more objective approach to assessing pharmaco-genomic variation in all genes comprising drug metabolism pathways may be indicated in order to capture the spectrum of individually rare but collectively common variations that influence drug response. Additional studies would be required for a comprehensive curation of candidate genes, in order to reliably predict the risk of toxicity in individual patients.

Section V

Chapter 8 FINAL CONCLUSIONS AND FUTURE DIRECTION

Over recent years, there has been a surge in the use of high throughput genomic technologies, enabling identification of candidate genes implicated in the pathogenesis of IBD. Despite these technological advancements, the functional impact of the genetic variants in disease causality remains unclear. The research presented in this thesis focuses on the development and optimisation of immunological assays in order to assess the functional integrity of key innate immune signalling pathways and incorporates analysis of whole exome sequencing data in paediatric patients with IBD. Currently, the clinical phenotypes of IBD are blind to the underlying immunological phenotype or the molecular basis of the disease. Exploring novel groupings within IBD based on the immunological phenotype was a key aim within this thesis. My research focused on the innate signalling pathways including NF-κB, MAPK and NLRP3-inflammasome pathways, assessing the functionality of several upstream activation receptors of known biological importance in IBD and the downstream cytokine effector responses.

As a basis for future investigation, a systematic review of published literature was conducted at the start of my PhD, in order to identify immunological studies with a concurrent work-up investigating genes of known biological importance in IBD. This was an important initial step for my project, exploring the wider scope of conducting immunological studies to investigate the functional impact of genetic variants within a cohort of paediatric IBD patients. The systematic review highlighted a paucity of mechanistic studies investigating the functional importance of potential candidate genes, thereby signposting a clear direction from my own research perspective for an integrated immuno-genomic approach. An update of the systematic review was conducted to include data from 2015-2019, which again showed a continued scarcity of immuno-genomic studies in IBD, despite the rapid advancements in multi-omic technologies over the recent years.

In order to develop and optimise immunological assays assessing induced immune responses, experiments were initially conducted in healthy adult controls. During this developmental phase, I optimised the set up for ligand doses, starting material, cell numbers, incubation times and effector cytokine measurements. The optimisation

experiments enabled development of the standard operating procedure for the assay and was subsequently applied to patients and paediatric controls.

Additional experiments conducted using monocytes and TBNK cells were not included in the final thesis due to inconsistent experimental observations and methodology not conforming to the required laboratory standards. However, these failed experiments provided opportunities for careful introspection, further learning, reflection and understanding the core principles underpinning robust research methodology. The results of the experiments although set aside from the final analysis did indicate individuality of immune responses among patients and comparatively poor induction in patients on immunosuppressive drug treatment, which directly influenced the next phase of the project leading to selection of a treatment naïve paediatric IBD cohort for conducting functional assays.

The optimised assays conducted in a treatment naïve cohort showed a significantly reduced immune induction driven predominantly by hypofunctional TLR-mediated responses. This may reflect signalling defects in the key inflammatory pathways resulting in impaired cytokine production. We therefore postulate that the mechanism triggering bowel inflammation in IBD in some patients could be impaired immune responses to pathogenic stimuli, with a consequential poor clearance of bacteria and persistence of secondary inflammatory responses. With the application of unsupervised machine learning, it was possible to stratify patients into subgroups or clusters based on their immune profile. Furthermore, the clustering pattern generating these immune-phenotypes highlighted putative genomic contributions when linked to next generation sequencing results. This research underscores the importance of patient stratification based on integrated immune-genomic profiling, which will facilitate future mechanistic and therapeutic evaluation, and offer better theragnostic precision.

This project was limited by a modest cohort size given the challenges in recruitment of children presenting with severe disabling symptoms prior to diagnosis. The time of diagnosis can be particularly stressful for families and young patients due to the chronic nature of the condition and the morbidity associated with it. Approaching families for research recruitment at this difficult time can therefore be very challenging. This was particularly relevant to my project as all specimens for PBMCs

were collected at diagnosis before commencement of any drug therapy. Despite the small cohort size, it was possible to identify evidence supporting immuno-genomic profile-based patient stratification. It is likely that a larger sample size would have improved power to detect significant genomic signals attributable to patient immune clusters.

This thesis also integrates the immuno-genomic profiling of periostin, a protein which has been described in a wide spectrum of inflammatory and oncological conditions. This research showed significantly higher plasma levels of the protein during remission compared to active disease, suggesting a predominantly reparative rather than inflammatory role in IBD. Although the precise molecular mechanisms of periostin remain unclear, the activation of inflammatory cascades such as the NF-κB via the αv-integrins receptors makes it a potentially attractive drug target for future treatment of IBD[150].

Pharmaco-genomic analysis included in this thesis focuses on *TPMT* and other genes implicated in thiopurine-induced drug toxicity. Results of this study indicate that the prediction of thiopurine toxicity through NGS has a higher specificity compared to the biochemical test, however the sensitivity of both methods is clinically suboptimal. A significant association between variations within the *MOCOS* gene and TPMT enzyme activity was identified, which has not been previously reported. Although there is no clear advantage of NGS over the biochemical test in predicting toxicity, our study demonstrates the strength of NGS as a powerful tool in identifying pathogenic variants not detected through standard genotyping. As high throughput sequencing becomes more affordable and increasingly accessible, a more objective approach to assessing pharmaco-genomic variation in all genes comprising drug metabolism pathways may be indicated in order to capture the spectrum of individually rare but collectively common variations that influence drug response.

Future direction

There are many interesting directions in which the projects presented within this thesis could be continued.

The immunological assays developed and optimised in this research focused on the integrity of the innate immune signalling pathways in IBD. However, abnormal

adaptive immune responses have also been well described in the pathogenesis of IBD[4,60]. Therefore, the immunological profiling of the adaptive immune system in addition to the innate immune system may provide a more comprehensive analysis and a better definition of the immune phenotype-based patient stratification.

The immunological profiling in this thesis was conducted in a treatment naïve cohort at diagnosis. As all patients had active inflammation at diagnosis when the samples were collected, it is unclear if the observed hypo-functionality in the patient cohort is an epiphenomenon of inflammation. To follow this prospectively, it is important to demonstrate a maintained 'cluster signature' by using PBMCs collected from patients during remission, whilst not on any treatment that would impact on the assay function. However, it would be practically challenging to identify a sizable cohort of patients with controlled disease, but not on immunosuppressive treatment.

In this project, exome data from the control samples were not available. For a robust interpretation of the genomic findings, it would be important to compare the mutational burden in the patient cohort with the control samples.

Genomic analysis of the immune clusters identified a statistically significant mutational burden in the TNF-α signalling pathway with a significant contribution from the *USP21* gene in one of the clusters. This gene is currently not implicated in IBD. Therefore, further functional work would be needed for a better understanding of the role of this gene in IBD.

In the periostin study, a longitudinal analysis of the plasma periostin during the course of the disease will provide better insights. Understanding the impact of variation in the *POSTN* gene and other functionally connected genes, in unaffected individuals, would enable improved understanding of this protein's role in IBD pathogenesis.

In the study investigating the genes involved in thiopurine-induced drug toxicity, a significant association was observed between the *MOCOS* gene and TPMT enzyme activity which has not been previously reported. Further research is required to identify the precise molecular mechanisms describing how *MOCOS* gene can impact on *TPMT* function, and response to thiopurine treatment.

The aetiology of this polygenic immune-mediated disease is complex with a combination of multiple factors converging to produce similar clinical phenotypes.

With the emergence of robust genomic and other omic technologies, generation of multi-omic profiles in individual patients for personalised therapy remains an ongoing research endeavour. However, despite the emergence of robust high-throughput technologies and rigorous statistical tools, the greatest challenge is to coherently integrate multiple axes of variance into biologically and clinically informative models of disease risk prediction, allowing targeted interventions, minimising treatment-related adverse effects and improving overall health outcomes by de-escalating treatments in patients at lowest risk of exacerbations. This can be achieved through co-ordinated efforts of multi-omic collaborative networks investigating data at multiple time points, multiple centres and judicious use of machine learning models. Although omics integration is still in its early phase of development, there are several exemplary studies reporting a comprehensive analysis of multi-omic data for accurately stratifying patients for personalised treatment, particularly in cancer research[544-546].

Future IBD research must include robust multi-omic based disease stratification for effective implementation of personalised therapy in a clinical setting, leading to improved quality of life in individuals affected by this condition.

Appendix A

Table A.1 Genes prioritised for the systematic review

Gene	GRAIL	DAPPLE	eSNP	eQTL Genes	Co-expression net. Genes
ATG16L1			✓	✓	
CARD11	✓	✓			
CARD9	✓		✓	✓	✓
CCDC88B				✓	✓
CCL13	✓				✓
CCL2	✓			✓	
CCR6	✓				✓
CD226	✓		✓	✓	✓
CD40	✓			✓	
CD48	✓				✓
CD6	✓		✓		
CREM	✓			✓	
CSF2	✓	✓			
CXCL1	✓	✓			
CXCL3	✓	✓			
CXCL5	✓	✓		✓	✓
CXCR1	✓	✓			
CXCR2	✓	✓			
DAP	✓			✓	
DOK3	✓				✓
EPO	✓	✓			
FASLG	✓				✓
FCGR2A	✓		✓		✓
FCGR2B	✓				✓
FCGR3A	✓				✓
FOS	✓	✓			
GPR35	✓		✓		
GPR65	✓		✓		✓
HLA-DQB1				✓	✓
HSPA6		✓		✓	
IFNAR1	✓	✓			
IFNG	✓	✓			✓
IFNGR2	✓	✓		✓	
IL10	✓	✓			✓
IL12B	✓	✓			
IL13	✓	✓			
IL15RA	✓	✓			
IL19	✓	✓			
IL2	✓	✓			
IL20	✓	✓			
IL23R	✓	✓	✓		
IL24	✓	✓			
IL2RA	✓	✓			
IL3	✓	✓			
IL4	✓	✓			
IL6ST	✓	✓			
IRF1	✓	✓		✓	✓
IRF5	✓			✓	✓
ITGAL	✓	✓			
JAK2	✓	✓			
LACC1			✓	✓	
LGALS9	✓				✓
LIF	✓	✓			
MAP3K8	✓	✓			
MST1			✓	✓	
NFKB1	✓	✓			
NOD2	✓		✓		✓
OSM	✓	✓			
PFKFB4	✓				✓
PRKCB				✓	✓
RASGRP1	✓			✓	✓
RPS6KA2	✓			✓	
SLC11A1	✓			✓	✓
SLC22A4			✓	✓	
SOCS1	✓	✓			
STAT1	✓	✓			✓
STAT3	✓	✓			
TNFRSF14	✓		✓		
TNNI2				✓	✓
TRAF3IP2	✓	✓		✓	
TYK2	✓	✓			

Table A.1 shows the list of genes prioritised using the five network connectivity tools. Genes identified by at least two of the five in silico methods were selected for the systematic search.

Table A.2 Studies selected for the systematic review

Genes	Study	Individuals Assessed	Sequencing method	Functional work
NOD2	Inohara et al. *Host recognition of bacterial muramyl dipeptide mediated through NOD2. Implications for Crohn's disease.* J Biol Chem. 278(8): 5509-12. Feb 2003	Number of patients or controls unclear. As stated in the paper, a panel of healthy and Crohn's disease individuals were genotyped for disease-associated NOD2 alleles and 2 individuals (one without clinical evidence of CD and one with CD) who were homozygous for L1007fsinsC were identified. Not clear, how many individuals had functional studies done.	DNA tested for NOD2 mutations using the SNaPshot method (Applied Biosystems, Foster City, CA)	PBMCs cultured with MDP or LPS; Cells lysed and nuclear extracts analysed for presence of NF-KB binding activity by an electrophoretic mobility shift assay (EMSA); Real-time PCR analysis for m RNA expression of IL-1β AND A1 transcripts following stimulation of PBMCs with MDP OR LPS
	Li et al. *Regulation of IL-8 and IL-1beta expression in Crohn's disease associated NOD2/CARD15 mutations. Human Molecular Genetics.* 13(16): 1715-25. Aug 2004	Total number of individuals genotyped unclear. 3 healthy controls with WT alleles and 2 CD patients homozygous for Leu1007fsinsC had functional studies done.	The Arg702Trp, Gly908Arg and Leu1007fsinsC variants were typed using allele-specific PCR. Confirmation of NOD2 genotypes performed by sequencing.	PBMCs stimulated with MDP, TNF-α alone or in combination; IL-8 and IL-1β measured by ELISA from supernatants; IL-8 and IL-1β mRNA expression measured by real-time PCR; NF-KB transcripts also assessed following treatment with MDP.
	Wehkamp et al. *NOD2 (CARD15) mutations in Crohn's disease are associated with diminished mucosal alpha-defensin expression.* Gut. 53(11): 1658-64. Nov 2004	Total cohort of 68 CD patients genotyped; 45 patients mean age 36.5 years (24 with NOD2 mutations, 21 with WT) and 12controls had functional studies performed.	Genotyping of genomic and or cDNA for NOD2 mutations (SNP8, SNP12, and SNP13) performed using TaqMan technology	Colonic biopsies assessed in 39 patients for mucosal RNA expression of alpha-defensins (HD-5, HD-6), secretory phospholipase A2, lysozyme, human hypoxanthine phosphoribosyl-transferase (house-keeping gene) TNF-α and IL-8 with real time reverse transcription-polymerase chain reaction; Immuno-histochemistry with anti-HD-5 and paneth cell staining performed performed in 10 patients
NOD2	Halme et al. *CARD15 frameshift mutation in patients with CROHN disease is associated with immune dysregulation.* Scand J Gastroenterol. 39(12): 1243-9. Dec 2004	16 CD patients (4 homozygotes, 6 heterozygotes, and 6 wild-type) and 6 healthy volunteers (one heterozygous, others homozygous for WT)	Each CARD15 variant was assayed using PCR and subsequent analysis of the PCR products by restriction enzyme cleavage and gel electrophoresis	Mononuclear cells cultured with LPS, IFN-g and GM-CSF; Immune inflammatory status evaluated by measuring monocyte HLA-DR and CD11b densities and proportion of CD14 ^{dim}CD16^{+} monocytes; TNF-α and IL-10 measured from cellular supernatants

NOD2	Van Heel et al. *Muramyl dipeptide and toll-like receptor sensitivity in NOD2-associated Crohn's disease.* Lancet. 365(9473): 1794-6.	1243 UK white individuals with CD genotyped. 19 patients with homozygous mutations and 7 controls underwent cytokine analysis.	Genotyped for 702Trp/1007fs, 702Trp/702Trp, 1007fs/1007fs, and 908Arg/1007fs	PBMCs stimulated with MDP and LPS; Cytokines (TNF-α, IL-1β and IL-8) measured by ELISA
	Braat et al. *Consequence of functional Nod2 and Tlr4 mutations on gene transcription in Crohn's disease patients.* Journal of Molecular Medicine. 83(8): 601-9. Aug 2005	411 CD patients, 226 UC patients and 137 controls genotyped. 3 CD patients homozygous for the 1007fs mutation and two CD patients homozygous for the A299G mutation in TLR4 had functional studies done.	PCR- restriction fragment-length polymorphism assay used for genotyping three SNPs within the Nod2 gene (R702W, G908R and 1007fs) and 2 SNPs within the TLR4 gene.	Dendritic cells stimulated with peptidoglycan and LPS; Gene transcription assessed with DNA micro-array between stimulated and un-stimulated cells
	Netea et al. *The frameshift mutation in Nod2 results in unresponsiveness not only to Nod2- but also Nod1-activating peptidoglycan agonists*. Journal of Biological Chemistry. 280(43): 35859-67. Oct 2005	74 patients with CD and 10 healthy volunteers genotyped. For functional studies, 5 CD patients homozygous for *NOD2* mutations, 5 patients heterozygous for the mutations, 5 patients with WT allele, 5 controls with	PCR amplification of *NOD2* gene fragments containing the polymorphic sites 3020insC (for Nod2fs) and C2104T (for R702W)	Stimulation of PBMCs with NOD1 agonists (Tri-DAP), NOD2 agonists (MDP derivatives), TLR4 agonist (LPS), TLR2 agonist (LTA) or in combinations; Cytokine analysis (IL-1β, IL-10, and TNF-α) from supernatants and IL-1a from cell lysates; Real time PCR analysis of selected genes (involved in MDP signalling) from PBMCs; Some additional experiments performed on cell lines.
	Van Heel et al. *Synergy between TLR9 and NOD2 innate immune responses is lost in genetic Crohn's disease.* Gut. 54(11): 1553-7. Nov 2005	NOD2 wild-type healthy controls (n=7) and NOD2 homozygous Crohn's disease patients (n=19), age and sex matched	PCR-restriction fragment length polymorphic genotyping for the three common NOD2 variants performed	PBMCs stimulated with CpG DNA (TLR9 ligand) and MDP; Cytokines (TNF-α and IL-8) measured by ELISA
	Kramer et al. *Impaired dendritic cell function in Crohn's disease patients with NOD2 3020insC mutation.* Journal of Leukocyte Biology. 79(4): 860-6. April 2006	150 CD patients and 10 healthy controls had their genotype assessed. 4 patients homozygous for the polymorphism, 4 patients with a WT allele and 4 controls with a WT allele selected for functional studies.	PCR amplification of NOD2 gene fragments containing the polymorphic site 3020insC performed. The 3020insC polymorphism was analyzed by Genescan analysis on an ABI PRISM 3100 genetic analyzer.	Monocyte derived DC generated from mononuclear cells; Dendritic cells stimulated with TLR ligands (LPS, Pam3cys and polyionisicpolycytidylic acid), NOD2 agonist (MDP) and interferon-g; Expression of co-stimulatory molecules CD80 and CD86 assessed using flow cytometry; Cytokine analysis (IL-12 by ELISA, TNF-α and IL-10 by commercially available kits) from supernatants

NOD2	Van Heel et al. *Normal responses to specific NOD1-activating peptidoglycan agonists in the presence of the NOD2 frameshift and other mutations in Crohn's disease.* European Journal of Immunology. 36(6): 1629-35. June 2006	18 individuals with CD and 8 healthy controls	Genotyping for the 3 common *NOD2* mutations	PBMCs stimulated with FK565/TriDAP (*NOD1* agonists), M-TriDAP (has *NOD1* and *NOD2* activity) and LPS; Cellular supernatants assessed for IL-8 and IL-1β by ELISA
	Van Heel et al. *Detection of muramyl dipeptide-sensing pathway defects in patients with Crohn's disease.* Inflammatory Bowel Diseases. 12(7): 598-605. July 2006	52 patients with CD, 1 patient with Blau syndrome and 22 with UC	Genotyping for the 3 common *NOD2* variants using polymerase chain reaction [PCR]-restriction fragment length polymorphism method	PBMCs stimulated with MDP and LPS; Cellular supernatants assessed for IL-8 and IL-1β by ELISA
	Peeters et al. *CARD15 variants determine a disturbed early response of monocytes to adherent-invasive Escherichia coli strain LF82 in Crohn's disease.* International Journal of Immunogenetics. 34 (3) (pp 181-191). June 2007	250 CD patients genotyped, of which 40 patients with a mean age of 35 years were selected for functional studies.	Genomic DNA extracted from whole blood using QIAGEN DNA kit. Genotyping for 3 NOD2 variants using restriction fragment length polymorphism-PCR. TLR4 Asp299Gly polymorphism also assessed.	Monocytes isolated from PBMCs, stimulated with MDP or LPS; Infection of monocytes with Adherent-invasive E coli (AIEC) also performed; Concentrations of IL-1β, IL-6, IL-8, IL-10, IL-12p and TNF-α measured from culture supernatants using Cytometric Bead Array and ELISA; mRNA expression for IL-1β, IL-6 and IL-10, measured by qRT-PCR (Quantitative real time-PCR); Bacterial intra-cellular survival and replication assessed using gentamicin protection assay
	Csillag et al. *CARD15 status and familial predisposition for Crohn's disease and colonic gene expression.* Dig Dis Sci. 52(8): 1783-9. August 2007	45 patients with CD (mean age 31 years) genotyped and had colonic tissue assessed for gene expression.	DNA purified from peripheral lymphocytes using QIAamp blood MiniKit (Qiagen, Hilden, Germany), PCR amplification and genotyping for the 3 *NOD2* mutations	Gene profiling analysis was performed on colonic tissue obtained from the descending colon using the Human Genome U133 Plus 2.0 GeneChip Array
	Seiderer et al. *The role of the selenoprotein S (SELS) gene - 105G>A promoter polymorphism in inflammatory bowel disease and regulation of SELS gene expression in intestinal inflammation.* Tissue Antigens. 70(3): 238-46. Sep 2007	Total of 563 individuals genotyped (205 CD patients, mean age 39.8 years and 154 patients with UC, mean age 42.7 years) and 204 healthy controls (mean age 50 years). Colonic biopsies from 12 CD patients and 10 UC patients assessed for functional studies.	*NOD2* variants detected by direct DNA sequence analysis. Patients were also assessed for *SELS* –105G>A polymorphism as part of the study.	*SELS* mRNA expression and IL-8 expression assessed from colonic biopsies of patients; TNF-α and other cytokines from serum also assessed by ELISA; Functional studies were also conducted in human intestinal cell lines and murine models (not meeting our inclusion criteria)

NOD2	Granzotto et al. *Heterozygous nucleotide-binding oligomerization domain-2 mutations affect monocyte maturation in Crohn's disease*. World Journal of Gastroenterology. 13(46): 6191-6. Dec 2007	47 individuals (mean age -16.6) with CD and 69 healthy controls genotyped for *NOD2* mutations. All CD patients and 9 controls (with wild type allele) had functional studies done.	Genotyping for R702W and G908R mutations (identified by PCR amplification and enzymatic digestion) and for the 1007fs mutation (analyzed by amplification and sequencing)	Dendritic cells stimulated with MDP or LPS; Cellular supernatants assessed for IL-12p70 (regulatory cytokine of the adaptive immune system) by ELISA; Cell surface expression (activation/maturation markers, such as CD83, HLADR and CD1a) evaluated through triple immunofluorescence staining with monoclonal antibodies through flow cytometry.
	Hedl et al. *Chronic stimulation of Nod2 mediates tolerance to bacterial products*. Proceedings of the National Academy of Sciences of the United States of America. 104(49): 19440-5. Dec 2007	? 18 individuals (including controls)	Genotyping of individuals for Leu1007insC *NOD2* mutation	Macrophages pre-treated with MDP for time periods ranging from 3 to 48 h, then retreated with synthetic lipid A (a TLR4 ligand) and Pam3Cys; TNF-α, IL-8, and IL-1β production assayed from supernatants by ELISA; IRAK-1 activity assessed through Western blot analysis.
	Butler et al. *NOD2 activity modulates the phenotype of LPS-stimulated dendritic cells to promote the development of T-helper type 2-like lymphocytes - Possible implications for NOD2-associated Crohn's disease.* Journal of Crohn's and Colitis. 1 (2) (pp 106-115). Dec 2007	Total number of individuals genotyped? Unclear. 6 CD patients homozygous for 1007fs mutation and 6 healthy controls with WT allele selected for functional studies.	Genotyped for the 3 most common *NOD2* variants.	Monocytes isolated from PBMCs, stimulated with LPS after pre-treatment with MDP or without; Supernatants analysed for cytokines (TNF-α , IL-10, IFNg, IL-12p70 and IL-13 by capture ELISA.
	Ferwerda et al. *Engagement of NOD2 has a dual effect on proIL-1beta mRNA transcription and secretion of bioactive IL-1beta.* European Journal of Immunology. 38(1): 184-91. Jan 2008	154 patients with CD and 10 controls were genotyped. Of these, 4 patients homozygous for the NOD2 variant, 4 patients with the WT	PCR amplification of *NOD2* gene fragments containing the polymorphic site 3020insC performed	MNCs stimulated with MDP or LPS or in combination at various concentrations; Cytokine profile (TNF-α, IL-1βIL-1b, proIL-1b and IL-10) studied in the supernatants by ELISA; Western blot also used for measurements of pro-IL-1b and IL-1b
	Zelinkova et al. *Muramyl dipeptide-induced differential gene expression in NOD2 mutant and wild-type crohn's disease patient-derived dendritic cells*. Inflammatory Bowel Diseases. 14(2): 186-94. Feb 2008	6 patients with CD and 3 controls (mean age= 40.5 years)	Genotyping for 3 CD-associated variants of the *NOD2* gene (R702W, G908R, and 3020Cins)	Monocyte-derived dendritic cells (DCs) stimulated with MDP; Whole-genome microarrays used to assess the differential gene expression; Clustering and transcription of various genes (those involved in the pathogen response pathways) analyzed in CD patients as well as controls, using online gene ontology mapping software.

NOD2	Lappalainen et al. *Novel CARD15/NOD2 mutations in Finnish patients with Crohn's disease and their relation to phenotypic variation in vitro and in vivo.* Inflammatory Bowel Diseases. 14(2): 176-85. Feb 2008	240 CD patients and 190 controls for DNA samples. Of these, 19 patients (6 with wild type allele and 13 with mutations) and 6 healthy volunteers (wild type allele) included for functional studies	Whole protein coding region and intron-exon boundary of the NOD2 gene sequenced	Mononuclear cells cultured alone or with MDP; IL-8 levels determined in the cellular supernatants by ELISA
	Kullberg et al. *Crohn's disease patients homozygous for the 3020insC NOD2 mutation have a defective NOD2/TLR4 cross-tolerance to intestinal stimuli.* Immunology. 123(4): 600-5. April 2008	74 patients with CD had bloods taken for genotyping. 5 patients homozygous for the polymorphism included in the final study along with 5 patients with wild type allele and 5 healthy volunteers with WT allele.	PCR amplification of *NOD2* gene fragments containing the polymorphic site 3020insC	Mononuclear cells stimulated MDP; Subsequently stimulated with LPS, Pam3cys, macrophage-activating lipopeptide 2 (MALP), heat killed Salmonella typhimurium or bacteroides fragilis; ELISA to measure TNFα , IL-6, IL-10 and IL-12 in the supernatants
	Simms et al. Reduced alpha-defensin expression is associated with inflammation and not NOD2 mutation status in ileal Crohn's disease; Gut; 2008	65 CD patients and 11 healthy controls had functional studies performed along with genotyping for *NOD2* variants.	Genotyping of the common disease-predisposing NOD2 variants (R702W, G908R and 1007FS), and a marker of the NOD2 haplotype (IVS8=158), P268S, was performed by high-resolution melt analysis on a Rotor-Gene 6000 (Corbett Life Science, Sydney, Australia).	mRNA expression levels of alpha defensins (DEFA5, DEFA6, LYZ, PLA2G2A), IL6, IL8, and VIL1 (Villin, marker of epithelial cell content) were measured using quantitative real time PCR colonic mucosa; Anti-human defensin 5 (HD-5) immune-histochemical staining performed on biopsy specimen
	Salucci et al. *Monocyte-derived dendritic cells from Crohn patients show differential NOD2/CARD15-dependent immune responses to bacteria.* Inflammatory Bowel Diseases. Issue: Volume 14(6). June 2008	53 patients with CD and 12 healthy controls	The entire coding and flanking sequence of exons 8 and 11 and the relevant segment of the large exon 4 of the CARD15 gene amplified by a standard polymerase chain reaction (PCR) for detection of the three common *NOD2* variants	Dendritic cells derived from monocytes isolated; Stimulated with a Salmonella strain; Cells collected and analysed for acquisition of activation markers (CD80 and HLA-DR) by flow-cytometry; Supernatants analysed by ELISA for IL-10, IL-12p70, IL-8, TNF-α.

NOD2	Benyon et al. *NOD2/CARD15 genotype influences MDP-induced cytokine release and basal IL-12p40 levels in primary isolated peripheral blood monocytes.* Inflammatory Bowel Diseases. 14(8): 1033-40. August 2008	40 patients with CD and 15 healthy controls	Genomic DNA isolated, genotyping for the 3 NOD2 variants (R702W, G908R, and 1007fs) performed.	Monocytes isolated from PBMCs, stimulated with MDP; Cytokine (TNF-α, IL-10, IL-12p40, and IL-1β) analysis performed on cellular supernatants by ELISA; mRNA expression from cellular lysates for IL-10, IL-12p40, IL-1β and GAPDH.
	Bodar et al. *NOD2 engagement induces proinflammatory cytokine production, but not apoptosis, in leukocytes isolated from patients with Crohn's disease.* European Cytokine Network. 19 (4) (pp 185-189). Dec 2008	4 patients with CD homozygous for the mutation, 4 patients with CD with the WT allele and 4 healthy volunteers with the WT allele selected for the study.	PCR amplification of *NOD2* gene fragments containing the polymorphic site 3020insC performed, analysed by Genescan analysis on an ABI-Prism 3100 Genetic Analyser (Applied Biosystems)	Mononuclear cells stimulated with MDP; Cellular supernatants assessed for TNF-α and IL-1β by ELISA; Cells assessed for spontaneous apoptosis, MDP-induced apoptosis and anisomycin-induced apoptosis
	Brosbol-Ravnborg et al. *Toll-like receptor-induced granulocyte-macrophage colony-stimulating factor secretion is impaired in Crohn's disease by nucleotide oligomerization domain 2-dependent and -independent pathways.* Clinical and Experimental Immunology. 155 (3) (pp 487-495). March 2009	224 patients with CD genotyped. 41 CD patients (22 with *NOD2* associated variants and 19 WT) and 12 healthy controls (WT allele) included for functional studies.	Genomic DNA extracted from cryopreserved PBMCs using the FlexiGene DNA kit (Qiagen, Hilden, Germany). Genotyping of *NOD2* variants using Taqman assays.	PBMCs stimulated with MDP (NOD2 ligand), PGN (TLR-2 ligand), polyinosine- polycytidylic acid (TLR-3 ligand), LPS (TLR-4 ligand), loxoribine (TLR-7 ligand) and CpG DNA ODN M_3 (TLR-9 ligand); Cytokine analysis from supernatants using multiplex beadskit for the presence of human cytokines (GM-CSF, IL-1β and TNF-α). Beads analysed using Luminex 100™
	Noguchi et al. *A Crohn's disease-associated NOD2 mutation suppresses transcription of human IL10 by inhibiting activity of the nuclear ribonucleoprotein hnRNP-A1.* Nature Immunology. 10 (5) (pp 471-479). May 2009	74 patients with CD and 20 healthy volunteers genotyped. 3 patients homozygous for the 3020insC mutation and 3 healthy controls were selected for functional studies.	*NOD2* gene fragments containing the 3020insC site amplified by PCR. The 3020insC polymorphism analysed by Genescan on an ABI Prism 3100 Genetic Analyser.	Primary monocytes stimulated with MDP, Pam_3Cys or a combination; m RNA expression of IL-10, IL-1β and IL-12p40 assessed ; Functional studies also carried out in mouse models and transfection in cell lines.

	Seidelin et al. *Evidence for impaired CARD15 signalling in Crohn's disease without disease linked variants.* PLoS ONE. 4 (11), Article Number: e7794. Nov 2009	18 patients with CD and 14 controls included	Both CD patients and controls screened for NOD2 mutations: SNP8, SNP12 and SNP13 by the single strand conformation polymorphism (SSCP). PCR amplification was performed using primers for the polymorphic segements.	Monocytes were isolated form PBMCs, stimulated with MDP; mRNA expression by RT quantitative PCR performed for NF-KB dependent pro-inflammatory cytokines TNF-α and IL-1β; m RNA expression of NALP3-inflammasome related molecules and expression of the p38 MAP kinase pathway also assessed.
	Canto et al. *MDP-Induced selective tolerance to TLR4 ligands: Impairment in NOD2 mutant Crohn's disease patient.* Inflamm Bowel Dis. Volume 15(11), p 1686-1696. Nov 2009	15 patients with CD and 10 healthy controls	Analysis of NOD2 variants performed using the Qiagen kit (Heiden, Germany). 8 patients identified as compound heterozygotes selected for the functional study.	PBMCs pre-treated with MDP, washed, then re-stimulated with MDP, LPS or LTA; Supernatants tested for TNF-α (ELISA), and IL-6, IL-8 and IL-10 by flowcytomix (Bender MedSystem Vienna, Austria); RNA isolation and Quantitative Real-time PCR for b-actin, HPRT1, TNF-α and NOD2.
	Cooney et al. *NOD2 stimulation induces autophagy in dendritic cells influencing bacterial handling and antigen presentation.* Nature Medicine. 16(1): 90-7. Jan 2010. (Study selected under ATG16L1*)	CD patients-number of individuals selected for the study?	Genotyping for 3 CD-associated variants of the *NOD2* gene (R702W, G908R, and 3020Cins) and ATG16L1 variants	Dendritic cells stimulated with MDP and autophagy induction assessed by measuring LC3 (modifier protein involved in autophagy) localisation; Ability of dendritic cells in bacterial trafficking and antigen presentation also examined.
	Rahman et al. *The pathogen recognition receptor NOD2 regulates human FOXP3+ T cell survival.* Journal of Immunology. 184(12): 7247-56. June 2010	30 patient with CD (mean age 45.9) and 10 healthy (mean age 56.2) individuals	Pyrosequencing performed on peripheral blood and mucosal specimens	Tregs isolated from peripheral blood as well as from lamina propria; Tregs incubated with MDP, then assessed for apoptosis; Tregs treated with MDP and NF-KB activity quantified by ELISA. Western blotting also performed
	Kosovac et al. *Association of the NOD2 genotype with bacterial translocation via altered cell-cell contacts in Crohn's disease patients.* Inflammatory Bowel Diseases. Issue: Volume 16(8). Aug 2010	36 CD patients and 30 controls	Patients with CD were genotyped for SNP 8, SNP12 and SNP13	Endotoxin stained by immunohistochemistry in 30 intestinal biopsies from patients carrying NOD2 variants (NOD2-mut) or being NOD2 wildtype (WT). Junctional proteins were visualized by immunofluorescence and quantified by Western blotting. NF-kappaB activation analyzed by immunohistochemistry in specimens from NOD2-WT and NOD2-mut CD and control patients. (All from GI tissue).

	Hamm et al. *NOD2 status and human ileal gene expression.* Inflamm Bowel Dis. 16(10): 1649-57. October 2010	18 non-smoking CD patients not treated with biologics and 9 non-smoking controls without IBD undergoing resection were genotyped. Microarray analysis performed in 4 patients with at least one risk allele, 4 patients without risk allele and 4 controls (without risk allele)	Each subject genotyped for the Leu1007fsInsC R702W, and G908R SNPs by direct sequencing or by a Taqman MGB (Applied Biosystems, Foster City, CA) genotyping platform. In addition, patients were genotyped for the ATG16L1T300A (rs2241880) and IL23R381N (rs11109026) risk alleles by the Sequenom Technology Core.	Microarray analysis performed on RNA isolated from ileal biopsies; Candidate genes selected by significance analysis of microarrays (SAM) were confirmed by quantitative reverse transcriptase PCR; IL-8 mRNA expression also assessed from the tissue.
	Glubb et al. *NOD2 and ATG16L1 polymorphisms affect monocyte responses in Crohn's disease.* World Journal of Gastroenterology. 17(23): 2829-37. June 2011. *(Study selected under ATG16L1)	12 patients	*NOD2* and *ATG16L1* SNPS genotyped (using allele-specific PCR and TaqMan SNP genotyping assays)	Monocytes challenged with mycobacterium avium paratuberculosis; Cytokine (IFNγ, IL-10, IL-12p40, IL-12p70, IL-17, IL-1b, IL-2, IL-4, IL-5, IL-6, IL-8, TNFα, TNFβ) responses assessed using 13-plex MILLIPLEX™.
NOD2	Gutierrez et al. *Antimicrobial peptide response to blood translocation of bacterial DNA in Crohn's disease is affected by NOD2/CARD15 genotype.* Inflammatory Bowel Diseases. 17(8): 1641-50. Aug 2011	50 patients with CD and 15 controls	*NOD2* SNPS genotyped	Presence of bacterial-DNA, defensin and cathelicidin gene, and protein levels in neutrophils and serum cytokine levels studied.
	Hewitt et al. *Immuno-inhibitory PD-L1 can be induced by a peptidoglycan/NOD 2 mediated pathway in primary monocytic cells and is deficient in Crohn's patients with homozygous NOD2 mutations.* Clinical Immunology. 143(2):162-9. May 2012	4 patients with Crohn's disease, 6 unaffected controls	*NOD2* SNPs genotyped	Mononuclear cells stimulated with MDP, followed by detection of TNF-α and IL-8 in the cellular supernatants by ELISA; Also mRNA expression for TNF-α, IL-8, 18S and PD-L1 assessed using real time RT-PCR.

	Vissers et al. *Respiratory syncytial virus infection augments NOD2 signaling in an IFN-beta-dependent manner in human primary cells*. European Journal of Immunology. 42 (10) (pp 2727-2735. Oct 2012	5 patients with CD homozygous for 3020insC mutation and 5 healthy volunteers participated in the functional study	Genotyping of genomic and or cDNA for NOD2 mutations	PBMCs from healthy volunteers and NOD2-deficient patients were stimulated with RSV and LTA (TLR2 ligands), LPS (for TLR4), flagellin (TLR5), CpG (TLR9) and MDP. Cytokines (TNFα, IL-10, IL-1β) from cellular supernatants analysed by ELISA
	Wagner et al. *TLR4, IL10RA, and NOD2 mutation in paediatric Crohn's disease patients: An association with Mycobacterium avium subspecies paratuberculosis and TLR4 and IL10RA expression.* Medical Microbiology and Immunology. 202 (4) (pp 267-276). Aug 2013. (*study also selected under IL-10)	62 early onset paediatric CD patients and 46 non-IBD paediatric controls	Genomic DNA extracted, multiplex PCR and extension reactions carried out using the Sequenom iPLEX Gold reaction protocol. Genotyping was performed using the matrix-assisted laser desorption ionization time-of-flight (MALDI TOF) mass spectrometry platform. 34 CD susceptibility SNPs in 18 genes genotyped including *NOD2, IL10RA and TLR4*.	MAP (Mycobacterium avium paratuberculosis) status previously assessed on patients in mucosal gut biopsies, peripheral blood mononuclear cells (PBMCs), and by culture of gut biopsies; TLR4 and IL10R expression also assessed PBMCs stimulated with MAP derivative and LPS; Cytokine analysis (TNF-α, IL-6, IL-10) from cellular supernatants;
	Lin et al. *NOD2 mutations affect muramyl dipeptide stimulation of human B lymphocytes and interact with other IBD-associated genes.* Digestive Diseases & Sciences. 58(9):2599-607. Sep 2013	294 IBD patients and 298 unrelated healthy controls	Genotype analysis of 3 *NOD2* mutations- primer-specific amplification, PCR based-RFLP, and validated with the ABI SNPlexM genotyping system	NOD2 expression in peripheral B cells and EBV-transformed B cell lines. Expression of NF-KB-p50 F following MDP treatment.
	Kuuliala et al. *Detection of muramyl dipeptide-sensing pathway defects in monocytes of patients with Crohn's disease using phospho-specific whole blood flow cytometry.* Scandinavian Journal of Clinical and Laboratory Investigation. 73 (6) (pp 494-502). Sept 2013	16 CD patients (5 wild-type, 6 heterozygous & 3 homozygous for 1007*fs* mutation, 3 rare novel mutation) and 5 healthy adult patients with wild-type allele.	Genotyped for *NOD2* variants	Monocytes supplemented with MDP and LPS; NF-kB and p38 activation detected through ELISA and phospho-specific flow cytometric analysis.

	Gutierrez et al. *Genetic susceptibility to increased bacterial translocation influences the response to biological therapy in patients with Crohn's disease.* Gut. 63 (2) (pp 272-280). Feb 2014	179 patients with CD and 25 healthy controls	3 common NOD2/CARD15 allelic variants at SNP-8 (R702W, rs2066844), SNP-12 (G908R, rs2066845) and SNP-13 (L1007finsC, rs2066847) genotyped. ATG16L1 variants also genotyped.	Phagocytic and bactericidal activities evaluated in blood neutrophils; Bacterial DNA, TNFα, IFNγ, IL-12p40, free serum infliximab/adalimumab levels and antidrug antibodies measured.
IL10	Koss et al. *Cytokine (TNF alpha, LT alpha and IL-10) polymorphisms in inflammatory bowel diseases and normal controls: differential effects on production and allele frequencies.* Genes & Immunity. 1(3): 185 -90.	Genotyping in 236 controls, 111 UC patients and 91 CD patients. Functional studies in 52 controls and 136 IBD patients (Adult cohort)	SNP genotyping	Production of the TNF alpha and IL-10 was measured by ELISA in lipopolysaccharide (LPS) stimulated whole blood.
	Gasche et al. *Novel variants of the IL-10 receptor 1 affect inhibition of monocyte TNF-alpha production.* Journal of Immunology. 170(11): 5578-82. June 2003	Blood drawn for RNA and DNA extraction from 10 control subjects and seven CD patients. Genotyping of cSNPs done in 310 DNA samples from European control subjects and 100 European CD patients.	Genotyping of sequence variants by PCR amplification with AmpliTaq Gold (Applied biosystems, Foster City, CA)	Monocytes stimulated with LPS and recombinant IL-10 and TNF-α expression analysed from cellular supernatants by ELISA; IL-R1 (receptor) expression assessed.
	Van der Linde et al. *A Gly15Arg mutation in the interleukin-10 gene reduces secretion of interleukin-10 in Crohn disease.* Scandinavian Journal of Gastroenterology. 38(6): 611-7. June 2003	17 sibling pairs with IBD and 75 healthy controls	Sanger Sequencing of IL-10 gene	PBMCs cultured with RPMI, stimulated with LPS or phorbol ester; Supernatants assessed for IL-10 protein production through ELISA or Western blotting; Activity of recombinant immature wild-type and mutated IL-10 was also tested in a proliferation assay with a human monocytic leukaemia cell line (HL60 cells).
	Glocker et al. *Inflammatory bowel disease and mutations affecting the interleukin-10 receptor.* New England Journal of Medicine. 361(21): 2033-45. Nov 2009	9 patients with IBD phenotype within the first year of life	Genetic-linkage analysis and candidate-gene sequencing (IL-10RA & IL-10RB)	PBMCs stimulated with LPS and supernatants assessed for TNF- α; TGF-β1; interleukin-1 α, -1β, -2, -6 and other cytokines. *SOCS3* (Suppressor of cytokine signaling 3 , a downstream target gene of STAT3 that is induced by IL-10) mRNA expression assessed on PBMCs following exposure to IL-10, by real-time polymerase-chain reaction. Transduction of cell lines and mutant cells with retroviral vectors also performed.

IL10	Wang et al. *The effect of IL-10 genetic variation and interleukin 10 serum levels on Crohn's disease susceptibility in a New Zealand population*. Human Immunology. 72 (5) (pp 431-435). May 2011	342 CD cases and 610 controls genotyped.	Genotyping of 3 SNPs of IL-10 and a flanking SNP	Measurement of serum IL-10 levels (in 188 CD patients and 195 controls)
	Begue et al. *Defective IL10 signaling defining a subgroup of patients with inflammatory bowel disease.* American Journal of Gastroenterology. 106(8): 1544-55. Aug 2011	75 children with IBD, including 13 infants with EO-IBD	IL10RA and B genes were sequenced	The capacities of transforming growth factor β (TGFβ) and IL10 to inhibit proinflammatory cytokine production by monocyte-derived dendritic cells (MoDC) or peripheral blood cells (PBMC) analysed; IL10 receptor-A/-B expression, STAT3 activation in response to IL6, IL10, IL21, IL22 analyzed by FACS and western blotting; The response to IL22 tested in ileal/colonic tissue cultures; Tissue gene expression analyzed by Taqman real-time polymerase chain reaction.
	Marcuzzi et al. *Inflammation profile of four early onset Crohn patients.* Gene. 493(2): 282-5. 2012 Feb	4 patients diagnosed with IBD within the first 2 years of life and 4 healthy controls	PCR amplification and DNA sequencing on the entire coding and flanking region of IL-10RA and IL-10RB genes performed. NOD2 genotyping with Taqman assays also performed.	PBMCs isolated, seeded with RPMI, then stimulated with LPS or recombinant human IL-10 alone or in combination
	Mao et al. *Exome sequencing identifies novel compound heterozygous mutations of IL-10 receptor 1 in neonatal-onset Crohn's disease*. Genes & Immunity. 13(5): 437-42.	1 patient with CD	Whole exome sequencing	IL-10R1 expression, IL-10 binding and STAT3 expression assessed. After re-constitution with wild-type IL-10R1, patient cells assessed for restoration of IL-10R binding function and STAT3 phosphorylation and expression of suppressor of cytokine signalling 3.
	Kotlarz et al. *Loss of interleukin-10 signaling and infantile inflammatory bowel disease: implications for diagnosis and therapy.* Gastroenterology. 143(2): 347-55. Aug 2012	66 patients with early onset IBD (younger than 5 years)	DNA Sanger sequencing of genes coding for IL-10R1, IL-1OR2 and IL-10	PBMCs stimulated with recombinant human IL-10, followed by Western blot analysis of STAT3; PBMCs stimulated with LPS, supernatant harvested and quantified for TNF-α upon co-stimualtion with IL-10

	Wagner et al. *TLR4, IL10RA, and NOD2 mutation in paediatric Crohn's disease patients: An association with Mycobacterium avium subspecies paratuberculosis and TLR4 and IL10RA expression*. Medical Microbiology and Immunology. 202 (4) (pp 267-276). Aug 2013. (*study also selected under NOD2)	62 early onset paediatric CD patients and 46 non-IBD paediatric controls	Genomic DNA extracted, multiplex PCR and extension reactions carried out using the Sequenom iPLEX Gold reaction protocol. Genotyping was performed using the matrix-assisted laser desorption ionization time-of-flight (MALDI TOF) mass spectrometry platform. 34 CD susceptibility SNPs in 18 genes genotyped including *NOD2, IL10RA and TLR4*	MAP (Mycobacterium avium paratuberculosis) status previously assessed on patients in mucosal gut biopsies, peripheral blood mononuclear cells (PBMCs), and by culture of gut biopsies; PBMCs stimulated with MAP derivative and LPS; Cytokine analysis (TNF-α, IL-6, IL-10) from cellular supernatants; TLR4 and IL10R expression also assessed
	Galatola et al. *Synergistic effect of interleukin-10-receptor variants in a case of early-onset ulcerative colitis.* World Journal of Gastroenterology19 (46) (pp 8659-8670). Dec 2013	One proband with early onset IBD, 3 unaffected relatives, 8 unaffected controls	Genomic PCR and sequencing of all exons of IL-10R gene	PCR quantification IL10RA and IL10RB messengers from peripheral blood cells; Beta catenin, TNF-α and TNF-α receptors analysed from peripheral blood cells and also from colorectal mucosa by Western blotting analysis.
ATG16L1	Cooney et al. *NOD2 stimulation induces autophagy in dendritic cells influencing bacterial handling and antigen presentation.* Nat Med. 16(1):90-7. Jan 2010; (Study selected under NOD2*)	Individuals with Crohn's disease. Number not specified	Genomic DNA from blood with a Flexigene DNA Extraction Kit (Qiagen) and performed PCR for specific mutations	ATG16L1 T300A–expressing DCs also showed a defect in autophagy induction and HLA-DM localization with LC3 after muramyldipeptide treatment but not after exposure to PAM3CYS4
	Glubb et al. *NOD2 and ATG16L1 polymorphisms affect monocyte responses in Crohn's disease.* World J Gastroenterol. 17(23):2829-37. June 2011; (Study selected under NOD2*)	Twelve patients with confirmed IBD	*NOD2* and *ATG16L1* SNPS genotyped (using allele-specific PCR and TaqMan SNP genotyping assays)	Monocytes were challenged with MAP and bacterial persistence assessed at subsequent time-points; Cytokine responses were assayed using a Milliplex multi-analyte profiling assay for 13 cytokines
	Plantinga et al. *Crohn's disease-associated ATG16L1 polymorphism modulates pro-inflammatory cytokine responses selectively upon activation of NOD2*. Gut. 60(9):1229-35. Sep 2011	Healthy individuals (Two independent cohorts of healthy volunteers assessed (N=46 and N=90)) and patients with Crohn's disease (74 patients).	Genotyping for the presence of the *ATG16L1* Thr300Ala polymorphism was performed by applying the TaqMan single nucleotide polymorphism assay C_9095577_20 on the 7300 ABI Real-Time PCR system	Induction of cytokine (pro-IL-1β, IL-1β, IL-6 and TNFα) production and related factors were measured at the mRNA and protein level; Protein levels of ATG16L1 were assessed by western blot.

ATG16L1	Thachil et al. *Crohn's disease-associated ATG16L1 polymorphism modulates pro-inflammatory cytokine responses selectively upon activation of NOD2.* Gastroenterology. 142(5):1097-1099.e4. May 2012	65 untreated paediatric patients, 32 CD (10 female and 22 male; median age of 11 years), 4 ulcerative colitis (1 female and 3 male; median aged 10 years), 9 celiac disease (4 female and 5 male; median age, 1.3 years), and 20 noninflammatory controls with no inflammation of the digestive tract (10 female and 10 male; median age, 10 years)	SNP genotyping using Taqman single nucleotide polymorphisms genotyping Assays	Increased LC3 accumulation in Paneth cells from IBD patients as compared to controls; Lysosomal activity was assessed by electron microscopy and staining
	Wildenberg et al. *Autophagy attenuates the adaptive immune response by destabilizing the immunologic synapse.* Gastroenterology. 142(7):1493-503.e6. June 2012	Crohn's disease patient. Number not specified	Crohn's disease patients were genotyped for rs_2241880 (ATG16L1) and rs_5743293 (NOD2) using polymerase chain reaction restriction fragment length polymorphisms	Dendritic Cell (DC) autophagy was reduced using small interfering RNAs or pharmacologic inhibitors; DC phenotype and function analyzed by confocal microscopy, time-lapse microscopy, and flow cytometry.
	Scharl et al. *Protein tyrosine phosphatase nonreceptor type 2 regulates autophagosome formation in human intestinal cells.* Inflammatory Bowel Diseases. 18(7):1287-302. July 2012	Active CD (*n* = 6, age range 31–75 years, mean: 49 ± 6 years) or from non-IBD control patients (*n* = 7, age range 37–75, mean: 54 ± 5)	Genotyped for disease-associated PTPN2 or ATG16L1 variations	Protein analysis in IEC and CLPF performed by western blotting; Autophagososme formation was assessed by LC3B immunofluorescence or immunohistochemistry
	Strisciuglio et al. *T300A Variant of Autophagy ATG16L1 Gene is Associated with Decreased Antigen Sampling and Processing by Dendritic Cells in Pediatric Crohn's Disease.* Inflammatory Bowel Diseases. 19 (11) (pp 2339-2348). Oct 2013	Paediatric patients who homozygously carry either the protective (wild type, n = 7) or risk allele (risk, n = 13) of ATG16L1, as well as heterozygous patients (het, n = 13)	Children were genotyped for SNPs *ATG16L1* rs2241880, *NOD2*rs2066844, *NOD2* rs2066845, and *NOD2* rs2066847 variants using the TaqMan system (7900HT sequence detection system; Applied Biosystems, Foster City, CA)	The monocyte-derived Dendritic Cells were analyzed for phenotype, antigen sampling, and processing by flow cytometry; The capability of DC to form transepithelial protrusions was determined by confocal microscopy
SLC22A4	Taubert et al. *Increased ergothioneine tissue concentrations in carriers of the Crohn's disease risk-associated 503F variant of the organic cation transporter OCTN1.*	24 patients with CD and 23 CD negative control subjects	The OCTN1 and OCTN2 genotypes were determined from genomic DNA obtained from whole blood samples by Taq dideoxy terminator cycle sequencing	OCTN1 m RNA expression and ergothioneine concentrations in intestinal epithelium and PBMCs analysed

	Gut. 58(2):312-4. Feb 2009			
	Repnik et al. *Haplotype in the IBD5 region is associated with refractory Crohn's disease in Slovenian patients and modulates expression of the SLC22A5 gene. Journal of Gastroenterology.* 46(9):1081-91. Sep 2011	Genotyped SNPs from the IBD5 locus (OMIM ID 606348) in 312 healthy controls and 632 IBD patients. RNA samples from 66 healthy individuals and 236 IBD analysed for gene expression.	PCR	Expression of three genes from the IBD5 locus, SLC22A4, SLC22A5, and IRF1measured
	Girardin et al. *Expression and functional analysis of intestinal organic cation/L-carnitine transporter (OCTN) in Crohn's disease.* Journal of Crohn's & colitis. 6(2):189-97. March 2012	Intestinal tissue wa s obtained from endoscopic biopsies and surgical resections from IBD patients (n=33 and 14, resp.) and controls (n=22 and 14, resp.). OCTN protein levels were measured in intestinal biopsies and carnitine transport was quantified in intestinal resections.	The genotyping of the common OCTN polymorphisms was performed using primer extension chemistry and mass spectrometric analysis (iPlex assay, Sequenom, San Diego, CA) on the Sequenom MassArray	OCTN protein levels were measured in intestinal biopsies; Carnitine transport was quantified in intestinal resections.
IL23R	Kim et al. *Genetic polymorphisms of IL-23R and IL-17A and novel insights into their associations with inflammatory bowel disease.* Gut. 60(11):1527-36. Nov 2011	201 patients with CD, 268 UC and 258 healthy controls genotyped.	Genomic DNA isolated, DNA sequencing of the promoter and exon regions of *IL-23R* and *IL-17A* performed	Transcription factor binding affinity, IL-17A m RNA expression and methylation of the promoter were evaluated in peripheral blood mononuclear cell and Jurkat cells.
	Diegelmann et al. *Intestinal DMBT1 expression is modulated by Crohn's disease-associated IL23R variants and by a DMBT1.* PLoS ONE. 8 (11). Nov 2013	818 with Crohn's disease and 972 healthy controls genotyped. Intestinal biopsies analysed in 27 CD patients for functional work.	Seven SNPs in the DMBT1 gene region analysed	The influence of *IL23R* variants on DMBT1 expression was analysed; Functional analysis included siRNA transfection, quantitative PCR, western blot, electrophoretic mobility shift and luciferase assays.
STAT3	Willson et al. *STAT3 genotypic variation and cellular STAT3 activation and colon leukocyte recruitment in pediatric Crohn disease.* Journal of Pediatric Gastroenterology &	Patients recruited not specified. ? 18 paediatric IBD patients and healthy controls (number not specified)	Genomic DNA extracted from whole blood, patients genotyped for STAT3 G>A (rs744166) SNP using the TaqMan system	Total RNA isolated from colonic biopsies for Gene array analysis, gene expression by real-time quantitative reverse transcription PCR; Expression of chemokines IL-8, CXCL2, CXCL3 and other proteins assessed

	Nutrition. 55(1):32-43. Jul 2012			
SLC11A1	Gazouli et al. *Role of functional polymorphisms of NRAMP1 gene for the development of Crohn's disease.* Inflammatory Bowel Diseases. 14(10):1323-30. Oct 2008	Blood samples from 274 patients with CD and 200 unrelated healthy individuals for genotyping. Functional work on colonic biopsies from 39 CD patients homozygous for the SNP under investigation.	The 5'(GT)n promoter polymorphism and 9 either SNPs or insertion/deletion type polymorphisms were genotyped across the *NRAMP1* gene	Reverse-transcriptase polymerase chain reaction and immunohistochemistry performed to investigate NRAMP1 m RNA levels in RNA isolated from colonic biopsies of CD patients as well as protein expression in tissues
CCL2	Herfarth et al. *Polymorphism of monocyte chemoattractant protein 1 in Crohn's disease.* International Journal of Colorectal Disease. 18(5):401-5. Sep 2003	179 patients with CD and 189 controls genotyped for MCP-1. Mucosal biopsies analysed in 31 patients with CD and 48 controls for MCP-1 concentrations.	MCP-1 genotyping carried out by polymerase chain reaction restriction fragment length polymorphism technique	MCP-1 concentration in tissue homogenates analysed in mucosal biopsies by ELISA

Table A.2 provides details of the studies identified through the systematic review. A total of sixty-seven studies were identified.

Table A.3 Systematic review update (2014-2018)

Implicated mechanisms	Genes	Number of publications
Epithelial		
Apical transport	*SLC26A3, SLC03A1*	**1.** Shao et al. Association of ulcerative colitis with solute-linked carrierf amily 26 member A3 gene polymorphisms and its expression in colonic tissues in Chinese patients. 2018. **2.** Wei et al. SLCO3A1, A novel crohn's disease-associated gene, regulates nf-kappaB activity and associates with intestinal perforation. 2014.
Junctional proteins	*MAG13* and other genes encoding tight-junction proteins	**1**. Noren et al. Genetic variation and expression levels of tight junction genes identifies association between MAGI3 and inflammatory bowel disease. 2017.
Others	*TCN2*	**1.** Zheng et al. An Analysis of Transcobalamin II Gene Polymorphisms and Serum Levels of Homocysteine, Folate and Vitamin B12 in Chinese Patients with Crohn's Disease. 2017.
Innate immunity		
NOD2 & associated proteins	*NOD2, TRIM22, XIAP*	**1.** Schafflet et al. NOD2- and disease-specific gene expression profiles of peripheral blood mononuclear cells from Crohn's disease patients. 2018. **2.** Amininejad et al. Analysis of Genes Associated With Monogenic Primary Immunodeficiency Identifies Rare Variants in XIAP in Patients With Crohn's Disease. 2018. **3.** Li et al. Variants in TRIM22 That Affect NOD2 Signaling Are Associated With Very-Early-Onset Inflammatory Bowel Disease. 2016. **4.** Dziadzio et al. Symptomatic males and female carriers in a large Caucasian kindred with XIAP deficiency. 2015.
Toll-like receptors	*TLR2, TLR5*	**1.** Meena et al. Association of TLR5 gene polymorphisms in ulcerative colitis patients of north India and their role in cytokine homeostasis. 2015. **2.** Salem. Species-specific engagement of human nucleotide oligomerization domain 2 (NOD)2 and Toll-like receptor (TLR) signalling upon intracellular bacterial infection: role of Crohn's associated NOD2 gene variants. 2015.
Complement	*MBL*	**1.** Choteau et al. Polymorphisms in the Mannose-Binding Lectin Gene are Associated with Defective Mannose-Binding Lectin Functional Activity in Crohn's Disease Patients. 2016.
Cellular processes		
Autophagy	*ATG16L1, ATG16L2*	**1.** Ma et al. A functional variant of ATG16L2 is associated with Crohn's disease in the Chinese population. 2016. **2.** Ngoh et al. The Crohn's disease-associated polymorphism in ATG16L1 (rs2241880) reduces SHIP gene expression and activity in human subjects. 2015. **3.** Wolfkamp et al. ATG16L1 and NOD2 polymorphisms enhance phagocytosis in monocytes of Crohn's disease patients. 2014. **4.** Murthy et al. A Crohn's disease variant in Atg16l1 enhances its degradation by caspase 3. 2014. **5.** Deuring et al. Genomic ATG16L1 risk allele-restricted Paneth cell ER stress in quiescent Crohn's disease. 2014.
Oxidative processes	*Genes encoding NADPH oxidases*	**1.** Denson et al. Clinical and Genomic Correlates of Neutrophil Reactive Oxygen Species Production in Pediatric Patients With Crohn's Disease. 2018.
Adaptive immunity	*PTPN22 & PTPN22, FOXP3,* genes encoding Th1 & Th17 cellular pathways	**1.** Yilmaz et al. The presence of genetic risk variants within PTPN2 and PTPN22 is associated with intestinal microbiota alterations in Swiss IBD cohort patients. 2018. **2.** Okou. Exome sequencing identifies a novel FOXP3 mutation in a 2-generation family with inflammatory bowel disease. 2014. **3.** Russo et al. Crohn's Colitis: Development of a multiplex gene expression assay comparing mRNA levels of susceptibility genes. 2017.
Innate & adaptive	*FCGR3A, ICOSLG, SHIP1*	**1.** Romero-Cara. A FCGR3A Polymorphism Predicts Anti-drug Antibodies in Chronic Inflammatory Bowel Disease Patients Treated With Anti-TNF. 2018. **2.** Hedl et al. Pattern recognition receptor signaling in human dendritic cells is enhanced by ICOS ligand and modulated by the Crohn's disease ICOSLG risk allele. 2014. **3.** Somasundaram. Analysis of SHIP1 expression and activity in Crohn's disease patients. 2017.
Cytokines	*IL6R, IL10, IFNA4 & IFNA10, TNFA, TRAIL, TNFSF15*	**1.** Parisinos et al. Variation in Interleukin 6 Receptor Gene Associates With Risk of Crohn's Disease and Ulcerative Colitis. 2018. **2.** Nemati et al. Very early onset inflammatory bowel disease: Investigation of the IL-10 signaling pathway in Iranian children. 2017. **3.** Lee et al. Novel de novo mutations of the interleukin-10 receptor gene lead to infantile onset inflammatory bowel disease. 2014. **4.** Xiao et al. Exome sequencing identifies novel compound heterozygous IFNA4 and IFNA10 mutations as a cause of impaired function in Crohn's disease patients. 2015. **5.** Kimura et al. Effects of tumor necrosis factor alpha-857C/T polymorphism on the expression of tumor necrosis factor alpha. 2016. **6.** Hu et al. Association of ulcerative colitis with TNF-related apoptosis inducing ligand (TRAIL) gene polymorphisms and plasma soluble TRAIL levels in Chinese Han population. 2015. **7.** Hedl et al. A TNFSF15 disease-risk polymorphism increases pattern-recognition receptor-induced signaling through caspase-8-induced IL-1. 2014.
Matricellular proteins	*MMP9*	**1.** Lin et al. WGCNA Reveals Key Roles of IL8 and MMP-9 in Progression of Involvement Area in Colon of Patients with Ulcerative Colitis. 2018.
Others	*SMAD3, HLA-DQB1,* other gene panels	**1.** Chiba et al. Allele-specific DNA methylation of disease susceptibility genes in Japanese patients with inflammatory bowel disease. 2018. 2. Song et al. Identification and analysis of key genes associated with ulcerative colitis based on DNA microarray data. 2018. **3.** Ranjha et al. Association of miR-196a-2 and miR-499 variants with ulcerative colitis and their correlation with expression of respective miRNAs. 2017.

Table A.3 provides a synopsis of the studies identified through the systematic search conducted since 2014 (update).

Table A.4 Amendments to 'The Genetics of Paediatric IBD' study

Amendment number	Date	Changes	Rationale	Outcome
2	Nov-13	Addition of new research sites	To enhance the power and statistical significance of the study findings	Successful
		Obtain faecal specimen for microbiome analysis	Characterise gut microbial flora and advance the understanding of disease causality	
		Obtain faecal specimen from unaffected siblings	To identify normal faecal flora from a healthy individual sharing the same environment	
		Include children under 5 years of age to the study	IBD can have very early onset including infancy.	
		Collect gastro-intestinal tissue for research	To further understand disease pathogenesis through tissue analysis	
		Collect additional bloods for immunology assays and allow repeat sampling if necessary	To assess functionality of multiple immune pathways implicated in IBD	
3	Apr-14	To include adult patients with IBD	To explore the input of genetic and immunological changes discovered in paediatric patients and assess the progression of these changes over time.	Successful
4	Nov-15	Biological samples from individuals recruited as suspected IBD, but biopsies do not favour a diagnosis of IBD to be used as control samples.	To better understand the differences between diseased and healthy tissue.	Successful
5	Jan-17	To allow collection of blood samples from healthy adult controls	To better understand the differences between immune responses in disease and health.	Successful

Appendix B ADDITIONAL METHODS

B.1 PBMC extraction

Reagents for PBMC extraction

Reagents needed for PBMC extraction include DMSO (Sigma), RPMI-1640 (Invitrogen), phosphate-buffered saline- PBS, Ficoll-Paque PLUS (GE Healthcare), Fetal calf serum- FCS and Trypan blue 0.4% (Sigma).

Table B.1 Equipment and instrumentation for PBMC extraction

Item of Equipment	Supplier
Class II laminar flow hood	HERAEUS (Kendro)
Centrifuge	SORVALL (Kendro)
Haemocytometer	(SupeRioR) Marienfeld
Adjustable pipettes	GILSON
Disposable polystyrene serological pipette 5 ml	BD Biosciences
Sterile polypropylene conical tube, 50 ml	BD Biosciences
Sterile polypropylene conical tube, 15 ml	BD Biosciences
1.8 ml cryovials	ThermoFisher Scientific
Serological pipetter	Fisher Scientific

PBMC extraction by Ficoll method (adapted from protocol for PBMC extraction and storage CSD/SOP/Template/002/UOS)

- Receive the sample in a green top lithium heparin tube. Add 6 ml Ficoll to a clean 15 ml universal tube.
- Overlay the Ficoll with up to 6 ml of blood.
- Transfer to the centrifuge without disturbing the layers and spin immediately at 2000 rpm for 25 minutes with the brake off.
- Carefully remove the sample from centrifuge without disturbing the layers and collect the interphase from each sample (cloudy layer with clear layer below and straw layer above) into 50 ml (Falcon) tubes.
- Add PBS (phosphate buffered solution) to make up to 50 mls. Spin at 1350 for 5 minutes.
- Decant the supernatant into a beaker of trigene, careful not to dislodge the pellet of cells.
- Resuspend the cells by adding PBS to make up to 1 millilitre.
- Perform a cell count with Trypan blue and improved Neubauer Haemocytometers.

- Centrifuge the remaining cell suspension at 1350 for 5 mins and aspirate the supernatant being careful not to dislodge the pellet.
- Resuspend the cell suspension in up to 5 ml (varying depending on the cell count) of freezing medium.
- Transfer to labelled cryovials.
- Snap-freeze in dry ice and then transfer to liquid nitrogen tank.

Thawing of PBMCs

- If PBMCs are not thawed properly, viability and cell recovery can be compromised; cells may not show optimal functioning in assays. Cells should be thawed quickly but diluted slowly to remove DMSO. Cells with DMSO intercalated into their membranes are very fragile and should be handled very gently.
- Warm the cryovials at 37 degrees Celsius in a water bath for 3-4 minutes
- Add the cell suspension to a 15ml tube, top it up to 15 ml with PBS
- Centrifuge: 1500 rpm, 20 degrees Celsius, 8 minutes, acceleration 5, deceleration 5
- Discard PBS, see pellet, add PBS to make up to 1 ml
- Cell count using 10µl cell suspension and 10µl trypan blue
- Wash the cells with PBS for a second cycle. Centrifuge: 1500 rpm, 20 degrees Celsius, 8 minutes, acceleration 5, deceleration 5
- Discard PBS, add RPMI to the cell pellet with a final concentration aimed at 1 million cells/ml of RPMI suspension

B.2 Monocyte fraction isolation from PBMCs

Monocytes were isolated by directly labelling PBMCs with fluorochrome-conjugated monoclonal antibodies (FITC conjugated CD14 antibodies) as primary labelling reagent and anti-FITC monoclonal antibodies conjugated as secondary labelling reagent. Magnetically labelled monocytes were isolated by retaining them on MACS column in the magnetic field of a MACS separator, while the un-labelled non-monocytes (T cells, B cells, NK cells, dendritic cells and basophils) were isolated by passing through the MACS column. The isolated monocytes were then stimulated with ligands as in the previous experiments. Incremental numbers of monocytes were used to assess cytokine response.

Sample preparation and magnetic labelling

- Peripheral blood mononuclear cells (PBMCs) were isolated by Ficoll-Paque density gradient centrifugation as described in previous sections

- Cell numbers determined
- Pipette off supernatant completely from the pellet
- Cells were kept cold using pre-cooled solutions and work fast. Higher temperatures and/or longer incubation times lead to capping of the antibodies on the cell surface and non-specific cell labelling.
- Volumes for magnetic labelling given below are for up to 10^7 total cells. With higher number of cells, reagent volumes and total volumes were scaled up accordingly. For example, for 2 x 10^7 total cells, twice the volume of the indicated reagents was employed.
- 10 µl FITC-conjugated CD14 antibody were added to the cell pellet and incubated for 10 minutes at 4-8°C.
- Washed with PBS and supernatant pipetted off completely from the pellet
- 10 µl anti-FITC antibody was added to the cell pellet and incubated for 10 minutes at 4-8°C
- Cells washed with PBS, centrifuged, supernatant pipetted off completely from the pellet

Magnetic separation

- An appropriate MACS Column and MACS separator were selected depending on the cell count. LS MACS column is appropriate for up to 2 x10^8
- The column was placed in the magnetic field of the MACS separator
- The column was prepared by rinsing the column with 3 mls of PBS
- The cell suspension was then applied to the column, allowing the cells to pass through the column. The effluent containing unlabelled non-monocytic fraction was collected into a 15 ml conical tube. Washing steps were performed by adding 1 ml PBS three times, each time once the column reservoir is empty. The effluent collected represents the unlabelled non-monocytic fraction.
- The column was then removed from the MACS separator and placed over a conical tube. Poured 5 mls of PBS in the column and swift pressure applied using the piston provided in the MACS column kit. The effluent represents the enriched labelled monocytic fraction.
- Cells from both the fractions (monocytes and non-monocytes) were then counted

B.3 Steps involved in Luminex assay (adapted from methods described in user manual: Invitrogen Human inflammatory plex panel. Catalog no. LHC0003)

1. Preparing reagents

2. Preparing standard curve
3. Prewet wells on the luminex plate
4. Add anti-body beads
5. Wash the plate
6. Add incubation buffer, standard and samples, then incubate for 2 hours
7. Wash
8. Add detection antibody (biotinylated antibody), then incubate for one hour
9. Wash
10. Add streptavidin-RPE, then incubate for 30 mins
11. Wash
12. Set up soft ware
13. Assay reading, acquire data using Luminex detection system

Guidelines for standard curve preparation

- Bring all reagents to room temperature before use
- Each kit comes with 2 complete sets of standard vials, so that 2 runs can be made with freshly prepared standards
- Protein standards should be re-constituted within an hour of performing the assay
- Confirm re-constitution volume standards from the technical data sheet
  Perform re-constitution of the standard with the assay diluent provided
- Add the suggested reconstitution of the assay diluent (in this case, add 0.5 ml of the diluent to the standard vial).
- Gentle shake, do not vortex to avoid foaming

Standard mixing and serial dilutions

- Arrange 7 Falcon glass tubes in a row for serial dilutions
- Standard curve is made by serial dilution of the reconstituted standard in assay diluent
- Do not vortex, mix by gently pipetting up and down

Wash solution

- 15 mls into a flask and dilute in 400 mls of water
- This solution can remain stable for up to 2 weeks if stored between 2-8 degrees

Prewetting wells on the luminex plate

- The luminex plate is a 96-well flat bottom plate (round bottom plate not useful as all beads will stick together)

- Add 200µl of wash solution to each well as per the plan with a multi-channel pipette
- Place the filter plate on the vacuum manifold and aspirate the liquid with gentle vacuum
- Do not exceed 5mm Hg. Excessive vacuum can cause tears in the membrane leading to loss of beads
- Stop the vacuum pressure as soon as the wells are empty
- After all wells are empty, lightly tap or press the filter plate onto clean paper towels

Preparing antibody beads

- The antibody bead concentrate is supplied as 10X concentrate and needs dilution (500µl of bead + 5 ml of wash solution)
- Vortex for 30 seconds before dispensing

Washing method

- Add 25µl of bead solution to each well
- Label Luminex plate inside and outside
- Add 200µl of wash solution to each well, place on vacuum manifold (pump), aspirate liquid, wash
- Repeat the above washing step

Adding incubation buffer

- Pipette 50µl of incubation buffer into each well
- Add 100µl of appropriate standard solution to the designated wells (a plate plan should be pre-designed before the start of the experiment. All standards and samples are performed in duplicate)
- Cover plates with an aluminium foil and incubate at room temperature on an orbital shaker @600rpm for one hour (The beads are light-sensitive. The aluminium foil prevents photo-bleaching of the embedded dye)
- Wash twice as described before

Preparing biotinylated anti-body

- BA is supplied as a 10X concentrate and is diluted with the wash solution to make up 1X concentrate
- Add 50µl of BA to all wells and 50µl of wash solution
- Incubate for 30 minutes
- Wash twice as per the steps described previously

Preparing streptavidin-RPE

- Supplied as a 10X concentrate, dilute X 10 times with wash solution
- Add 50µl of S-RPE + 50µl wash solution to the wells
- Incubate for 30 minutes (orbital shaker, 600 rpm)
- Shaking should sufficiently keep the beads suspended during incubation

Assay reading

- Wash with 200µl wash solution, aspirate with pump (vacuum manifold)
- Uncover plate and insert into XY platform of the Luminex 100 and analyse the samples

Set-up

1. Software used X-ponent, create a new protocol
2. Assign bead regions to each analyte
3. Assign settings: events/bead region, sample size, flow rate

B.4 DNA extraction

Genomic DNA is extracted from peripheral venous blood samples collected in EDTA, using the salting out method[308]. In this method, buffy coats of nucleated cells are obtained from anti-coagulated blood (EDTA) and resuspended in 15 ml polypropylene centrifugation tubes with 3 ml of nuclei lysis buffer (10 mM Tris-HCl, 400 mM NaCl and 2 Mm Na_2EDTA, pH 8.2). The cell lysates are digested overnight at 37◦ C with 0.2 ml of 10% SDS and 0.5 ml of a protease K solution (1 mg protease K in 1% SDS and 2mM Na_2EDTA). Following digestion of the cell lysates, 1 ml of saturated NaCl is added to each tube and shaken vigorously for 15 seconds, followed by centrifugation at 2500 rpm for 15 minutes. The precipitated protein pellet left at the bottom of the tube and the supernatant containing the DNA are transferred to another 15ml polypropylene tube. Exactly two parts of room temperature absolute ethanol is added and the tubes inverted several times until the DNA is precipitated. The precipitated DNA strands are removed with a plastic spatula and transferred to a 1.5 ml micro-centrifuge tube containing 100-200µl TE buffer (10:10 mM Tris-HCl, 400 mM NaCl and 2 Mm Na_2EDTA, pH 7.5). The DNA is allowed to dissolve for 2 hours at 37◦ C before quantitating. DNA concentration is estimated using the Qubit ® 2.0 Fluorometer and a 260:280 ratio calculated using a nanodrop spectrophotometer. The average DNA yield obtained is 150µg/ml and approximately 20ug of DNA is used for next generation sequencing for each patient.

Appendix C RESULTS OF EXPERIMENTS USING MONOCYTES AND TBNK CELLS IN CONTROLS AND PATIENTS

C.1 RESULTS ON CONTROLS

Characteristics of healthy controls

A total of 15 healthy control samples were analysed with multiplex assay for induced immune responses. The age range of healthy controls was between 20-40 years (median 27 years) and 80% were females.

'Pass' criterion for control samples

For a reference guideline, control samples with positive stimuli (i.e. LPS) induced IL-6 response more than 10pg/ml were marked as 'pass'. This was set based on previous experience in our laboratory through established assays where LPS was used as a positive control as well as from published literature[547]. Control samples failing this prerequisite were excluded from the analysis.

Results on control samples

Ten out of fifteen healthy controls with LPS induced IL-6 levels more than 10pg/ml were included in the analysis. The upper and lower limits of the 95% CI were subsequently applied as estimated range of normal immune variance for the patient samples. Five out of fifteen specimens were excluded from the analysis as the IL-6 levels were below the cut-off value of 10pg/ml following LPS induction. Three of the failed specimens were processed (stimulated with ligands) on the same day and the failed responses were probably due to technical failure. **Table C.1**.

Table C.1 Induced immune responses in healthy controls

Induced response	Lower limit of quantitation (LLOQ) (pg/ml)	Geo. Mean (pg/ml)	Lower 95% CI of geo. Mean (pg/ml)	Upper 95% CI of geo. Mean (pg/ml)	% of LPS response
LPS-induced IL-6	9.77	616.8	239.4	1589	100
MDP-induced IL-6		161.7	30.15	867.1	26.2
DAP-induced IL-6		61.03	9.057	411.2	10
PAM-induced IL-6		629	232	1705	102
TNF-α-induced IL-6		33.5	2.6	426.3	5.4
LPS-induced IL-1B	2.44	633.7	282.7	1421	100
MDP-induced IL-1B		251.6	94.87	667.4	39.7
DAP-induced IL-1B		57	11.29	289	9
PAM-induced IL-1B		235.7	80.18	693	37
TNF-α-induced IL-1B		42.34	12.87	139.2	6.7
LPS-induced TNF-α	8.54	103.2	24.3	437.7	100
MDP-induced TNF-α		42.52	16.4	109.9	41.2
DAP-induced TNF-α		21.91	6.4	74.8	21.3
PAM-induced TNF-α		64.79	29.18	143.9	62.8
TNF-α-induced TNF-α		116.9	43.75	312.6	113

C.2 Results on patient samples

C.2.1 Sample and data overview

A total of 37 samples (from 33 patients) were analysed; 4 samples were repeat specimens performed on the same patients at different time points. Samples were collected, processed and Luminex assays employed to measure cytokine levels as previously described. In order to assess the overall inflammatory signatures induced by the ligands, the concentrations of cytokines were plotted across all the samples and outliers per analyte identified as values 2 SD (95% CI) away from the geometric mean of the 'Pass' controls (shown in table **C.1**). Of the 33 patients, 61% (n= 20) were males, 75% had CD (n= 25), 21% UC (n=7) and one patient had IBDU. Eligibility or 'pass' criteria as applied to the control specimens were also applied to the patient specimens. 15 samples out of the 37 were excluded based on LPS-induced IL-6 response of less than 10pg/ml as for the control group. Patient sample characteristics are presented in table **C.2**. Of the 37 patient samples analysed, 40% (n=15) were obtained at the time of active disease and 38% (n=14) of samples collected from patients

who were not on any drugs at the time. Active disease was defined based on a PCDAI score >10 for Crohn's disease[438] or a PUCAI score >10 for UC[439].

Table C.2 Patient cohort characteristics

Probands	Gender	LPS Response (pg/ml)	Pass/ Fail	Diagnosis	PCDAI	PUCAI	Disease state	Drugs
SOPR0232	F	220.5	PASS	IBDU	-	0	Remission	None
SOPR0303	M	2107	PASS	CD	15	-	Active	None
PR0136	M	1373	PASS	CD	<10	-	Remission	None
PR0081	F	572	PASS	UC	-	35	Active	Aza + 5-ASA
PR0208	M	698	PASS	CD	<10	-	Remission	Sertraline
SOPR0336	F	2903	PASS	CD	40	-	Active	None
PR0047	M	12	PASS	UC	-	0	Remission	5-ASA
PR0114	M	16	PASS	CD	?	-	Remission	Inflx+ 5-ASA
PR0125	F	20	PASS	CD	0	-	Remission	None
SOPR0296	M	14	PASS	CD	0	-	Remission	None
SOPR0310-1	M	2917	PASS	CD	35	-	Active	None
SOPR0310-2	M	2670	PASS	CD	27.5	-	Active	Aza
SOPR0273	F	1174	PASS	CD	>30	-	Active	6-MP + steroids
SOPR0325	F	1810	PASS	CD	?	-	Remission	Aza + steroids
SOPR0313-1	M	2231	PASS	CD	7.5	-	Remission	None
SOPR0313-2	M	368	PASS	CD	25	-	Active	Steroids
PR0141	F	1094	PASS	UC	-	0	Remission	5-ASA
SOPR0262	M	83	PASS	CD	17.5	-	Active	Adal
SOPR0304-1	M	4087	PASS	CD	32.5	-	Active	None
SOPR0304-2	M	16	PASS	CD	32.5	-	Active	None
SOPR0311	M	155	PASS	UC	-	0	Remission	None
SOPR0327	M	2479	PASS	UC	-	0	Remission	5-ASA
PR0032	M	0	FAIL	CD	2.5		Remission	Aza
PR0121	F	0	FAIL	UC		0	Remission	None
PR0151	F	2	FAIL	CD	0	-	Remission	Inflx
PR0119-2	F	0	FAIL	CD	5	-	Remission	Aza + Adal
PR0190	M	2.75	FAIL	CD	5	-	Remission	Aza
PR0199	M	6	FAIL	CD	0	-	Remission	Aza
PR0247	M	0	FAIL	CD	5	-	Remission	Aza
PR0119-1	F	2	FAIL	CD	20	-	Active	Aza
PR0025	M	0	FAIL	CD	5	-	Remission	None
PR0148	M	1.4	FAIL	CD	0	-	Remission	Aza + inflx
SOPR0314	F	3.5	FAIL	UC	-	65	Active	5-ASA
PR0004	F	6	FAIL	CD	<10	-	Remission	Aza
PR0054	M	0	FAIL	CD	35	-	Active	Aza + inflx
SOPR0333	M	0	FAIL	CD	20	-	Active	None
PR0211	F	0	FAIL	CD	20	-	Active	Aza + inflx

Table C.2 shows the results of experiments conducted in patients using monocytes: TBNK cells in an enhanced ratio. The results were not included in the final analysis. ***(Abbreviations:*** *Adal- adalimumab; 5-ASA- 5-aminosalicyclic acid; Aza- azathioprine; inflx- infliximab; 6-MP- 6- mercaptopurine)*

C.2.2 Impact of drugs on overall response

In the group of samples failing the inclusion criteria (15 out of 37), 80% (12 out of 15) of the samples were obtained from patients on drug treatment for IBD. The drugs used to treat IBD included azathioprine, 6-mercaptopurine, steroids, 5-aminosalicylic acid drugs (5-ASA), biologics such as infliximab and adalimumab and others. In the group of specimens classed as 'pass', drugs were concurrently used at the time of sample collection in 50% of the cases. Although a higher proportion of samples failed the eligibility criteria whilst on drug therapy compared to no drugs, the difference was not statistically significant (p= 0.09). Table **C.3** shows the drugs used either as mono-therapy or in combination in the groups of samples either failing or passing the eligibility criteria. Three samples were obtained from patients whilst on oral steroids and all passed. Only 25% of samples on 5-ASA drugs monotherapy failed whilst 86% of samples on thiopurine drugs monotherapy failed the eligibility criteria. Biologic drugs are usually used in conjunction with thiopurines. The failure rate for samples taken whilst on thiopurine drugs (whether alone or in combination with other drugs) was 67% and for biologics 71% (monotherapy or combination). The overall failure rate of samples whilst on treatment with thiopurines or biologics (either, both or along with the other drugs) was 65% compared to 20% on no drugs or other drugs (steroids, 5-ASA, others). This was statistically significant (p=0.0084).

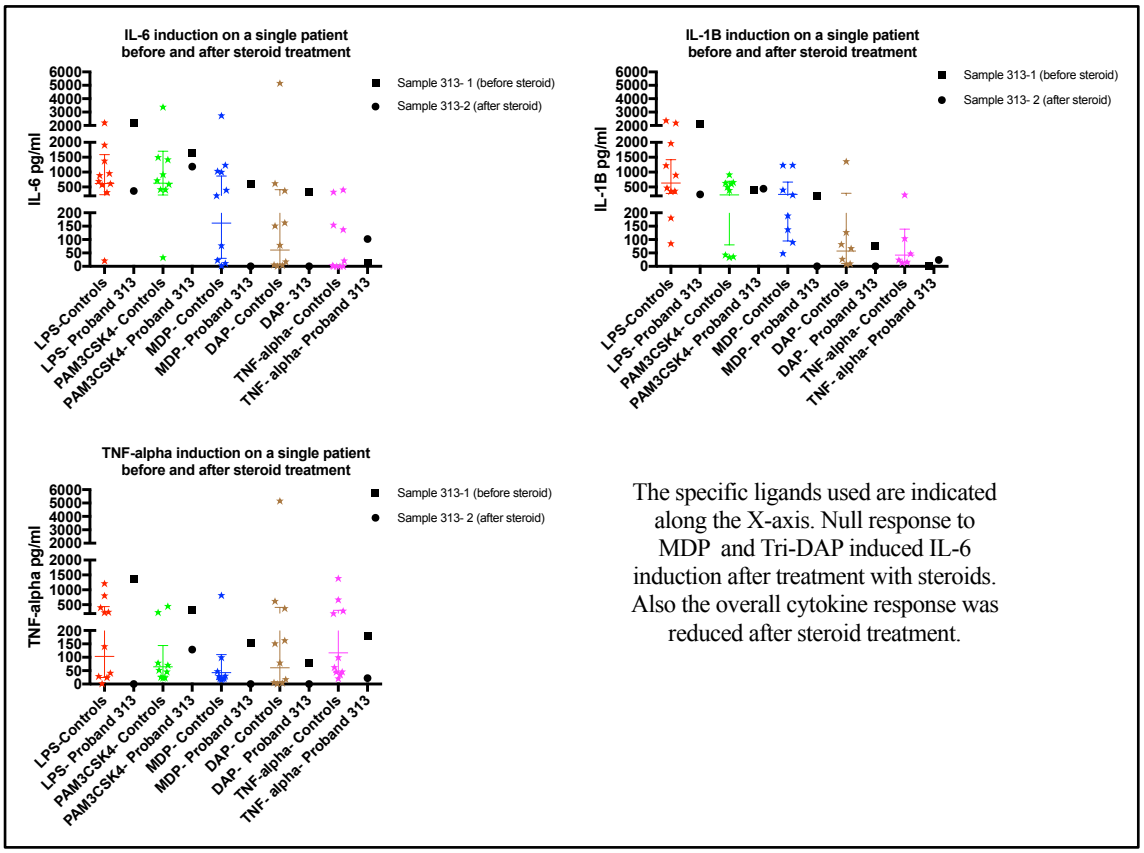
Table C.3 Impact of drug treatment on patient samples

Drugs	Pass	Fail	% Pass
5-ASA (Mesalazine/sulfasalizine)	3	1	75%
Thiopurines (AZA/6-MP)	1	6	14%
Biologics (Infliximab/Adalimumab)	1	1	50%
Steroids	1	0	100%
Thiopurine + 5-ASA	1	0	100%
Thiopurine + Biologics	1	4	20%
Steroids + Thiopurines	2	0	100%
Other drugs (example- anti-depressants)	1	0	100%
No drugs	11	3	79%
Total	22	15	

C.2.3 Impact of steroids on cytokine responses

Cryopreserved (frozen) PBMCs were analysed for induced cytokine responses from proband **0313** at diagnosis (sample **313-1**, treatment naïve, active disease) and compared with cytokine

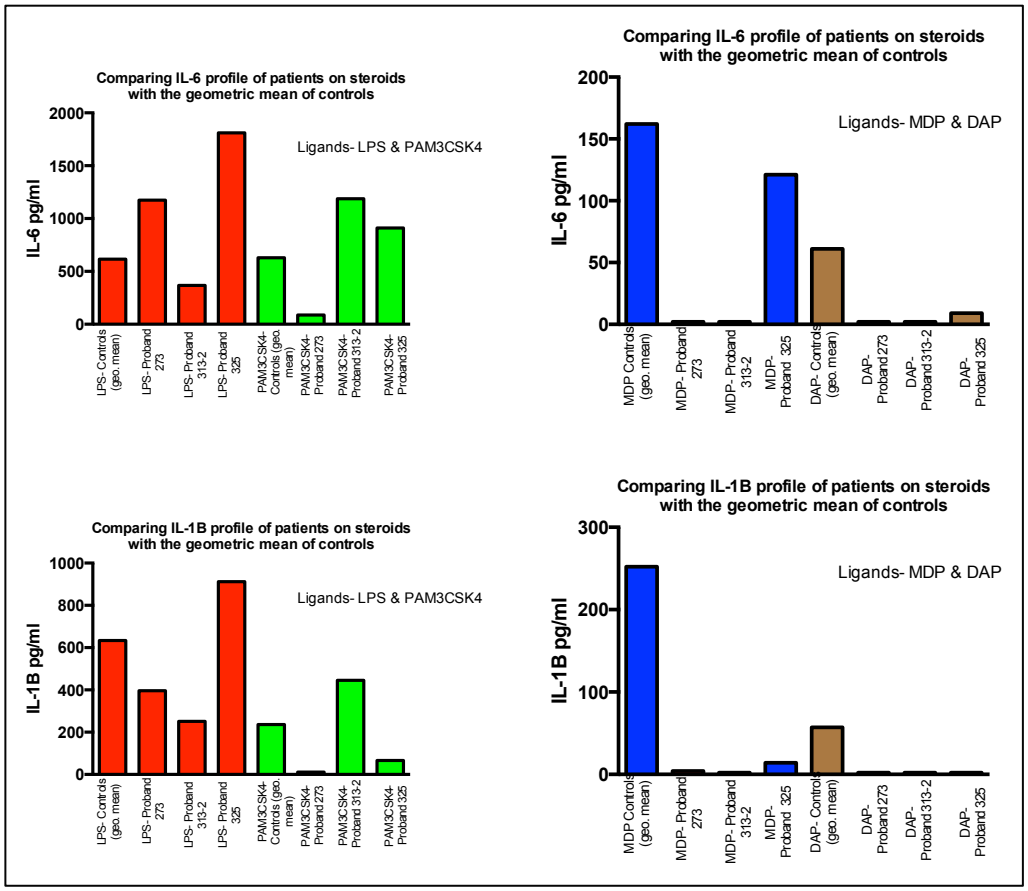
response from the same patient's PBMCs post-treatment (sample **313-2**, treated with systemic steroids, in remission). A predominantly reduced overall response was observed in this individual following treatment with steroids across all the three cytokines (IL-6, IL-1β and TNF-α) and with all the induction stimuli except TNF-α. See figure **C.1.** The LPS-induced IL-6 response in the post-treatment sample was 16.5% and PAM3CSK4 72% of the treatment-naïve sample. Similarly, the IL-1β and TNF-α responses were substantially reduced. A key observation in the post-treatment sample was a completely null response to MDP and DAP across all the three cytokines, suggesting an inhibitory effect of steroids on NOD signalling.



Appendix Fig C. 1 Cytokine induction in a single patient before and after steroids

See legend in the figure.

This effect of poor MDP and DAP response was observed in 2 additional individuals (probands **0273** and **0325**) who were treated with steroids at the time of blood sampling, although pre-treatment samples were not available in these individuals. See figure **C.2.**



Appendix Fig C. 2 Cytokine profile of patients on steroids

Marked reduction in the MDP and DAP induced cytokine induction noted in 3 patients who Were commenced on systemic steroids. The specific ligands used are indicated along the X axis for both controls as well as the proband.

C.2.4 Distinct inflammatory signatures in individual patients

Based on cytokine induction patterns, unique inflammatory signatures were identified in individual patients (see table 6.4 and figure 6.4). An overview of the distinct inflammatory signatures in selected individuals is presented in table **C.4.**

Table C. 4 Individuals with distinct inflammatory signatures

Probands	Cytokines	LPS	PAM3CSK4	MDP	DAP	TNF-α	Pathway defect
	IL6	↑	N	↓	↓	↓	NOD1/NOD2 and ?TNF-alpha signalling
	IL-1β	N	↓	↓	↓	↓	
SOPR0303	TNF-α	↑	↓	↓	↓	N	
	IL6	↓	N	↓	↓	↓	NOD1/NOD2 and ?TNF-alpha signalling
	IL-1β	N	N	↓	↓	↓	
PR0232	TNF-α	N	N	N	N	N	
	IL6	N	↓	N	N	N	TLR1/2 signalling
	IL-1β	↑	↓	N	N	N	
PR0136	TNF-α	↑	↓	N	N	N	
	IL6	N	N	N	N	↓	TNF-alpha receptor or post-TNFR canonical pathway defect
	IL-1β	N	N	↓	N	↓	
PR0081	TNF-α	N	N	↓	N	↓	
	IL6	N	N	↑	↑	↑	Defective inflammasome activation
	IL-1β	↓	↓	↓	↑	↑	
PR0208	TNF-α	↓	↓	↓	↑	N	
	IL6	↑	↑	↑	↑	↑	Pan-hyperinflammatory response
	IL-1β	N	↓	↑	↑	↑	
SOPR0336	TNF-α	N	N	↑	↑	N	
	IL6	↓	↓	↓	↓	↓	Pan-hypoinflammatory response
	IL-1β	↓	↓	↓	↓	↓	
PR0114	TNF-α	↓	↓	↓	↓	N	
	IL6	↓	↓	↓	N	N	Pan-hypoinflammatory response
	IL-1β	↓	↓	↓	↓	N	
PR0125	TNF-α	↓	↓	↓	↓	↑	
	IL6	↓	↓	↓	↓	↓	Pan-hypoinflammatory response
	IL-1β	↓	↓	↓	↓	↓	
SOPR0296	TNF-α	↓	↓	↓	↓	N	
	IL6	↓	↓	↓	↓	↓	Pan-hypoinflammatory response
	IL-1β	↓	↓	↓	↓	↓	
PR0047	TNF-α	↓	↓	↓	↓	↓	

Predominant NOD1/2 signalling defect

Although the NOD1/2 stimulated induction was weaker as compared to the other ligands, probands **0303** and **0232** showed a unique profile with a normal response to LPS and PAM3CSK4, but not to MDP and/or DAP. The IL-6, IL-1β and TNF-α were poorly induced following MDP or DAP stimulation. This would suggest an intact MAPK and NF-κB transduction via TLR signalling, but not through the NOD receptors. Proband **0303** was a male patient diagnosed with Crohn's disease at 14 years of age and treated with exclusive enteral nutrition with no drugs. The disease phenotype was ileo-colonic Crohn's disease with peri-anal lesions. PBMCs were collected 6 months following diagnosis whilst the patient was still having features of active disease (PCDAI 15), but not on any drugs apart from vitamin supplements. This patient showed a hyper-inflammatory response to LPS induced IL-6 and TNF-α, normal response to LPS induced IL-1β, normal response to PAM3CSK induced IL-1β, but a totally null response to MDP and DAP stimulation. These observations suggest a NOD1/2 signalling defect. The TNF-α induced responses were also poor which probably

suggests a signalling defect along the NF-κB canonical pathway downstream of the TNF-α receptor on the cell surface membrane. Proband **0232** was a female patient with IBD-unclassified (IBDU), aged 13 years, in remission (PUCAI score 0) and not on drugs at the time of blood sampling. This patient had a hypo-inflammatory response to LPS-induced IL-6, but a normal response to LPS-induced IL-1β and TNF-α, normal PAM3CSK4 induced responses but poor responses to MDP and DAP. As for proband 0303, the TNF-α induced cytokine induction was poor.

Poor NOD1/NOD2 stimulation responses in the context of normal TLR1/2/4 (LPS and PAM3CSK4) stimulation were also observed in other probands including **0273**, **0325** and **0313-2.** However, as described in the previous section, these 3 individuals were on steroids at the time.

Predominant TLR-1/2 signalling defect

Proband **0136**, a male patient with CD, aged 16 years, not on drugs and in remission (PCDAI <10) at the time of blood sampling showed cytokine responses consistent with a TLR-1/2 pathway signalling defect. This patient was diagnosed with CD at the age of 6 years. The LPS induced responses were normal for IL-6, hyper-inflammatory for IL-1β and TNF-α; normal MDP, DAP and TNF-α responses to all 3 cytokines. However, there was a consistent pattern of lack of response to PAM3CSK4 stimulation for all the three cytokines suggestive of a TLR1/TLR2 signalling defect.

Predominant TNF-α receptor or post-TNFR canonical pathway defect

Proband **0081** was a 16-year old female patient with UC on azathioprine and mesalazine. She had features of active disease (PUCAI 35) at the time of blood sampling. Cytokine profile analysis of this patient showed a normal response to LPS, PAM3CSK4, MDP and DAP across all or most cytokines, however there was a clear lack of response to TNF-α stimulation. As this patient was on azathioprine and mesalazine, it is difficult to speculate if there is a genetic basis for the block in the TNF-α mediated interference of the canonical pathway output or the cumulative effect of drugs.

Proband with a predominantly defective inflammasome activation

Proband **0208** was a 16-year old male patient with Crohn's disease. He was in remission (PCDAI <10) at the time of PBMC extraction and was not on any immunosuppressive treatment for IBD, although he was treated with sertraline for depression. Cytokine profile demonstrated a normal IL-6 response to LPS and PAM3CSK4, but a hyper-inflammatory response to MDP, DAP and TNF-α stimulation. A total lack of response was however noted

for LPS, PAM3CSK4 and MDP mediated induction of IL-1β and TNF-α. The defective IL-1β production may be indicative of an underlying defect in the inflammasome activation pathway.

A predominantly hyper-inflammatory or hypo-inflammatory phenotype across most stimuli

Proband **0336** is a 11-year old female patient with new diagnosis of Crohn's disease. PBMCs were collected from this patient at diagnosis before commencement of any treatment. At diagnosis, she presented with severe diarrhoea, weight loss and abdominal pain. Cytokine induction demonstrated a hyper-inflammatory response to all 5 stimuli for IL-6 read-out, a hyper-inflammatory response to MDP, DAP and TNF-α for IL-1β as well as TNF-α. Probands **0114**, **0125**, **0296** and **0047** showed an overall poor response to the inducing stimuli. Of these, only proband 0114 was on significant immunosuppression (infliximab and 6-mercaptopurine), while proband 0047 was on 5-ASA drugs. The other two individuals were not on immunosuppressive agents at the time of blood sampling.

Appendix D

D. 1 Genes and inferred gene networks for application of GenePy scores (Chapter 5)

	TLR network	TLR genes- where GenePy available	Signalling Cascade network	Signalling Cascade genes- where GenePy available	TNF-A network	TNF-alpha genes- where GenePy available	IL-10 netwo	IL-10 genes- where GenePy available
1	AGER	APP	AKT1	BRAF	ADAM17	BAG4	BCL2L1	BLVRA
2	APP	ATF1	AKT2	BTK	BAG4	BIRC2	BLVRA	CCND3
3	ATF1	ATF2	AKT3	CASP1	BIRC2	BIRC3	CCND1	CDKN2D
4	ATF2	BIRC2	ARAF	CD3E	BIRC3	CHUK	CCND2	HMOX1
5	BIRC2	BIRC3	BRAF	CD3G	CASP10	CYLD	CCND3	IL10
6	BIRC3	BPI	BTK	CD4	CASP8	IKBKB	CD14	IL1A
7	BPI	BTK	CASP1	CHUK	CAV1	MADD	CDKN2D	JAK1
8	BTK	BTRC	CASP8	ECSIT	CFLAR	NFKB1	HMOX1	MAPK12
9	BTRC	CD180	CD14	EGFR	CHUK	NSMAF	IL10	MAPK13
10	CASP8	CD36	CD3E	ERBB2	CLIP3	OTUD7B	IL10RA	MAPK14
11	CD14	CHUK	CD3G	ERBB3	CYLD	OTULIN	IL10RB	MAPKAPK2
12	CD180	CREB1	CD4	ERBB4	FADD	RBCK1	IL1A	MYC
13	CD36	DNM1	CHUK	FOS	FAS	RELA	IL1B	NFKB1
14	CHUK	DNM2	DAG1	FOSB	FASLG	RIPK1	IL6	PIK3R1
15	CREB1	DNM3	ECSIT	FYN	IKBKB	RNF31	JAK1	PIK3R2
16	CUL1	DUSP4	EGFR	GBP1	IKBKG	SHARPIN	MAP2K6	PIK3R3
17	DHX9	DUSP6	ERBB2	GBP2	MADD	SMPD2	MAPK11	PIK3R4
18	DNM1	ECSIT	ERBB3	GBP5	MAP2K3	SMPD3	MAPK12	PIK3R5
19	DNM2	FOS	ERBB4	GBP6	MAP2K7	SPPL2A	MAPK13	SOCS3
20	DNM3	IKBKB	FADD	GSK3A	MAP3K1	SPPL2B	MAPK14	TIA1
21	DUSP3	IKBKE	FOS	HSP90AA1	MAP3K3	TAB1	MAPKAPK2	TLR4
22	DUSP4	IRAK1	FOSB	HSP90AB1	MAP3K5	TAB2	MYC	TNFAIP3
23	DUSP6	IRAK2	FOSL1	HSP90B1	MAP3K7	TAX1BP1	NFKB1	TYK2
24	DUSP7	IRAK3	FYN	IKBKB	MAP4K2	TNFAIP3	PIK3R1	
25	ECSIT	IRAK4	GBP1	IKBKE	MAP4K3	TNFRSF1A	PIK3R2	
26	ELK1	IRF3	GBP2	IL1A	MAP4K4	TRADD	PIK3R3	
27	FADD	ITGAM	GBP5	IL1R1	MAP4K5	TRAF1	PIK3R4	
28	FBXW11	ITGB2	GBP6	IL1R2	NFKB1	TRAF2	PIK3R5	
29	FOS	JUN	GSDMD	IL1RAP	NRK	USP2	PIM1	
30	HMGB1	LBP	GSK3A	IL5RA	NSMAF	USP21	SOCS3	
31	IKBKB	LY86	GSK3B	IRAK4	OTUD7B	USP4	STAT3	
32	IKBKE	LY96	HLA-E	IRF6	OTULIN	XIAP	TIA1	
33	IKBKG	MAP3K1	HSP90AA1	ITGA4	PRKCI		TLR4	
34	IRAK1	MAP3K8	HSP90AA2P	ITGAV	PRKCZ		TNF	
35	IRAK2	MAPK14	HSP90AB1	JUN	RACK1		TYK2	
36	IRAK3	MAPK3	HSP90B1	LBP	RBCK1			
37	IRAK4	MAPK7	IKBKB	LY96	RELA			
38	IRF3	MAPKAPK2	IKBKE	LYN	RFFL			
39	IRF7	MEF2A	IKBKG	MALT1	RIPK1			

Appendix D

	TLR network	TLR genes-where GenePy available	Signalling Cascade network	Signalling Cascade genes-where GenePy available	TNF-A network	TNF-alpha genes-where GenePy available	IL-10 network	IL-10 genes-where GenePy available
40	ITGAM	NFKB1	IL13	MAP3K14	RNF31			
41	ITGB2	NFKB2	IL18	MAP3K3	RPS27A			
42	JUN	NFKBIB	IL1A	MAP3K8	SHARPIN			
43	LBP	NOD1	IL1B	MAPK12	SMPD1			
44	LY86	NOD2	IL1R1	MAPK13	SMPD2			
45	LY96	PELI1	IL1R2	MAPK14	SMPD3			
46	MAP2K1	PELI2	IL1RAP	MAPK3	SPPL2A			
47	MAP2K3	PELI3	IL2	MAPK7	SPPL2B			
48	MAP2K4	PLCG2	IL2RA	NFKB1	SQSTM1			
49	MAP2K6	PPP2R1B	IL5	NFKB2	STAT1			
50	MAP2K7	PPP2R5D	IL5RA	NFKBIB	TAB1			
51	MAP3K1	PTPN4	IRAK4	NFKBIE	TAB2			
52	MAP3K7	RELA	IRF6	NLRP3	TAB3			
53	MAP3K8	RIPK1	ITGA11	P2RX7	TAX1BP1			
54	MAPK1	RIPK2	ITGA4	PANX1	TNF			
55	MAPK10	RIPK3	ITGAV	PDK1	TNFAIP3			
56	MAPK11	RPS6KA1	ITGB1	PIK3C2A	TNFRSF10A			
57	MAPK14	RPS6KA2	JUN	PIK3C3	TNFRSF10B			
58	MAPK3	RPS6KA5	JUNB	PIK3CB	TNFRSF1A			
59	MAPK7	SARM1	JUND	PIK3CD	TNFRSF1B			
60	MAPK8	SIGIRR	LBP	PIK3CG	TNFSF10			
61	MAPK9	TAB1	LCK	PIK3R1	TNIK			
62	MAPKAPK2	TAB2	LY96	PIK3R2	TRADD			
63	MAPKAPK3	TANK	LYN	PIK3R3	TRAF1			
64	MEF2A	TBK1	MAA	PIK3R4	TRAF2			
65	MEF2C	TIRAP	MALT1	PIK3R5	TXN			
66	MIR6502	TLR1	MAP2K1	PRKCA	UBA52			
67	MIR718	TLR2	MAP2K2	PRKCB	UBB			
68	MYD88	TLR3	MAP2K3	PRKCD	UBC			
69	NFKB1	TLR4	MAP2K4	PRKCE	USP2			
70	NFKB2	TLR6	MAP2K6	PRKCH	USP21			
71	NFKBIA	TNIP2	MAP2K7	PRKCZ	USP4			
72	NFKBIB	TRAF3	MAP3K14	PTK2	XIAP			
73	NOD1	TRAF6	MAP3K3	PYCARD				
74	NOD2	VRK3	MAP3K7	RAF1				
75	PELI1		MAP3K8	REL				
76	PELI2		MAPK1	RELA				
77	PELI3		MAPK10	RIPK1				
78	PLCG2		MAPK11	RIPK2				
79	PPP2CA		MAPK12	RPS6KA1				
80	PPP2CB		MAPK13	RPS6KA2				
81	PPP2R1A		MAPK14	RPS6KA5				
82	PPP2R1B		MAPK15	RPS6KA6				
83	PPP2R5D		MAPK3	TAB1				
84	PTPN11		MAPK4	TANK				
85	PTPN4		MAPK6	TBK1				
86	RELA		MAPK7	TIRAP				
87	RIPK1		MAPK8	TLR1				
88	RIPK2		MAPK9	TLR10				
89	RIPK3		MYD88	TLR2				
90	RPS27A		NFKB1	TLR3				
91	RPS6KA1		NFKB2	TLR4				
92	RPS6KA2		NFKBIA	TLR5				
93	RPS6KA3		NFKBIB	TLR6				
94	RPS6KA5		NFKBIE	TLR7				
95	S100A12		NLRP3	TLR8				
96	S100B		P2RX7	TLR9				
97	SAA1		PANX1	TNFAIP3				
98	SARM1		PDK1	TNFRSF1A				
99	SIGIRR		PIK3C2A	TP53				
100								

	TLR network	TLR genes-where GenePy available	Signalling Cascade network	Signalling Cascade genes-where GenePy available	TNF-A network	TNF-alpha genes-where GenePy available	IL-10 network	IL-10 genes-where GenePy available
101	SOCS1		PIK3C2G	TRAF3				
102	TAB1		PIK3C3	TRAM1				
103	TAB2		PIK3CA	TRAM1L1				
104	TAB3		PIK3CB	ZAP70				
105	TANK		PIK3CD					
106	TBK1		PIK3CG					
107	TICAM1		PIK3R1					
108	TICAM2		PIK3R2					
109	TIRAP		PIK3R3					
110	TLR1		PIK3R4					
111	TLR2		PIK3R5					
112	TLR3		PRKCA					
113	TLR4		PRKCB					
114	TLR6		PRKCD					
115	TNIP2		PRKCE					
116	TRAF3		PRKCG					
117	TRAF6		PRKCH					
118	UBA52		PRKCI					
119	UBB		PRKCZ					
120	UBC		PTK2					
121	UBE2D1		PYCARD					
122	UBE2D2		RAF1					
123	UBE2D3		REL					
124	UBE2N		RELA					
125	UBE2V1		RELB					
126	VRK3		RIPK1					
127			RIPK2					
128			RPS6KA1					
129			RPS6KA2					
130			RPS6KA3					
131			RPS6KA5					
132			RPS6KA6					
133			SRC					
134			STAT1					
135			STAT3					
136			SUMO1					
137			SUMO2					
138			SUMO3					
139			SUMO4					
140			SYK					
141			TAB1					
142			TANK					
143			TBK1					
144			TICAM1					
145			TICAM2					
146			TIRAP					
147			TLR1					
148			TLR10					
149			TLR2					
150			TLR3					
151			TLR4					
152			TLR5					
153			TLR6					
154			TLR7					
155			TLR8					
156			TLR9					
157			TNF					
158			TNFAIP3					
159			TNFRSF1A					
160			TP53					
161			TRADD					
162			TRAF3					
163			TRAM1					
164			TRAM1L1					
165			TRIF					
166			ZAP70					

D. 2 Sum of GenePy scores across the inferred gene networks (Chapter 5)

Probands	Cluster	Sum of GenePy scores for all genes per gene network			
		TLR-gene network	Signalling cascade	TNF-α Gene network	IL-10 Gene network
SOPR0339	1	0.00312573	0.00132	0.000929	0.00092916
SOPR0340	1	0.00184174	0.002191	0.000733	0.00073278
SOPR0342	1	0.00382919	0.002659	0.000729	0.00072869
SOPR0345	1	0.00195051	0.001554	0.000944	0.00094377
SOPR0351	1	0.00141327	0.002165	0.000446	0.00044614
SOPR0355	2	0.00197723	0.004123	0.00077	0.00077023
SOPR0367	2	0.00197040	0.002614	0.000241	0.00024105
SOPR0372	2	0.00182664	0.001817	0.000594	0.00059445
SOPR0381	2	0.00166742	0.004049	0.00037	0.00037033
SOPR0378	3	0.00267811	0.002922	0.000122	0.00012151
SOPR0359	4	0.00215010	0.003591	0.000394	0.00039399
SOPR0368	4	0.00376686	0.006951	0.000635	0.00063475
SOPR0369	4	0.00155934	0.001017	0.000259	0.00025935
SOPR0336	5	0.00214067	0.002184	0.000355	0.00035518
SOPR0346	6	0.00216350	0.003193	0.00033	0.00033008
SOPR0352	6	0.00197636	0.00258	0.000261	0.00026107
SOPR0370	6	0.00212350	0.002166	0.00033	0.00032961
SOPR0384	6	0.00173234	0.002023	0.000351	0.00035099
SOPR0380	7	0.00099462	0.001842	0.000201	0.00020102
SOPR0304	8	0.00107820	0.002316	0.000164	0.00016447
SOPR0377	8	0.00616633	0.002183	0.00048	0.0004802

Table D.2 shows the sum of GenePy scores for all genes per gene network for the four selected gene networks.

Appendix E

E. 1 GenePy scores across *POSTN* and other functionally connected genes in patients with UC (Chapter 6)

Probands	Periostin Elisa (pg/ml)	PUCAI	Summation of GenePy Scores	CXCR4	CYTIP	EPAS1	JAG1	LEF1	NOTCH1	PIK3CG	POSTN	SRC	TCF7	TCF7L1	TCF7L2	TNC	VCAM1	VEGFA	VHL	WNT2
PR0173	35092	0	1.724283819	-	0	0	-	0	-	0.404300584	1.218024252	-	0	0	0	0.101958983	0	0	0	0
SOPR0311	25273	75	0.574014675	-	0	0	-	0	-	0.404300584	0	-	0	0	0	0.169714091	0	0	0	0
PR0160	31872	0	0.001320327	-	0	0	-	0	-	0	0	-	0	0	0	0.001320327	0	0	0	0
PR0167	35194	0	0.188239525	-	0	0	-	0	-	0	0	-	0	0	0	0.188239525	0	0	0	0
SOPR0314	28386	70	1.878971669	-	0	1.023147249	-	0	-	0.404300584	0	-	0	0	0	0.451523836	0	0	0	0
SOPR0316	15025	25	4.660970877	-	0	0	-	0	-	0	0	-	0.572415268	0.802458766	0	0.014885227	0	0	3.271211615	0
PR0176	16091	45	0.742129359	-	0	0	-	0	-	0	0	-	0.572415268	0	0	0.169714091	0	0	0	0
SOPR0346	15007	45	1.202188657	-	0	0	-	0	-	1.20086833	0	-	0	0	0	0.001320327	0	0	0	0
SOPR0367	23784	60	0.61879017	-	0	0	-	0	-	0.430550645	0	-	0	0	0	0.188239525	0	0	0	0
SOPR0384	35378	70	1.379872088	-	0	0	-	0	-	0	0	-	1.192311324	0	0	0.187560763	0	0	0	0
SOPR0340	24189	45	1.556360464	-	0	0	-	0	-	0.430550645	0	-	0.413880107	0	0	0.711929713	0	0	0	0
PR0096	15327	5	1.052147576	-	0	0	-	0	-	0.404300584	0	-	0.413880107	0	0	0.233966885	0	0	0	0
PR0102	32549	5	2.022721026	-	0	0	-	0	-	0	0	-	0.986295375	0.802458766	0	0.233966885	0	0	0	0
PR0110	31159	0	0.916026075	-	0	0	-	0	-	0	0	-	0	0	0	0.916026075	0	0	0	0
PR0111	36908	10	2.080418898	-	0	0	-	0	-	0.430550645	0.473894829	-	1.106897988	0	0	0.069075436	0	0	0	0
PR0112	21654	15	0.119805654	-	0	0	-	0	-	0	0	-	0	0	0	0.119805654	0	0	0	0
PR0105	26040	20	1.94888871	-	0	0	-	0	-	0.404300584	0	-	0.572415268	0.802458766	0	0.169714091	0	0	0	0
PR0115	21373	0	1.573980569	-	0	0	-	0	-	0	0.473894829	-	0.986295375	0	0	0.113790366	0	0	0	0
PR0117	25074	0	0.187560763	-	0	0	-	0	-	0	0	-	0	0	0	0.187560763	0	0	0	0
PR0121	35484	0	1.156009466	-	0	0	-	0	-	0	0	-	0.986295375	0	0	0.169714091	0	0	0	0
PR0100	41575	5	0.847296611	-	0	0	-	0	-	0	0	-	0.413880107	0	0	0.433416504	0	0	0	0
PR0124	25331	25	0.169714091	-	0	0	-	0	-	0	0	-	0	0	0	0.169714091	0	0	0	0
PR0125	27194	0	0.095264932	-	0	0	-	0	-	0	0	-	0	0	0	0.095264932	0	0	0	0
PR0128	31937	10	2.292289895	-	0	0	-	0	-	0	1.218024252	-	0.572415268	0	0	0.501850375	0	0	0	0
PR0135	21000	5	0.382975515	-	0	0	-	0	-	0	0	-	0	0	0	0.382975515	0	0	0	0
PR0134	24993	5	1.104131683	-	0	0	-	0	-	0.430550645	0	-	0.572415268	0	0	0.10116577	0	0	0	0
PR0129	31252	5	0.87951574	-	0	0	-	0	-	0.404300584	0.473894829	-	0	0	0	0.001320327	0	0	0	0
PR0138	16866	25	0.221342289	-	0	0	-	0	-	0	0	-	0	0	0	0.221342289	0	0	0	0
PR0149	24901	20	0.10116577	-	0	0	-	0	-	0	0	-	0	0	0	0.10116577	0	0	0	0
PR0179	23482	0	2.708926918	-	0	0	-	0	-	0	0	-	0	0	0	0.169714091	0	0	0	2.539212826
PR0195	33184	40	0.532509628	-	0	0	-	0	-	0.430550645	0	-	0	0	0	0.101958983	0	0	0	0
SOPR0314	22137	70	1.878971669	-	0	1.023147249	-	0	-	0.404300584	0	-	0	0	0	0.451523836	0	0	0	0
PR0180	38059	15	0.10116577	-	0	0	-	0	-	0	0	-	0	0	0	0.10116577	0	0	0	0
PR0219	21480	65	3.647889746	-	0	0	-	0	-	0.404300584	0	-	0	3.14163018	0	0.101958983	0	0	0	0
PR0081	34042	5	0.169714091	-	0	0	-	0	-	0	0	-	0	0	0	0.169714091	0	0	0	0
PR0122	35992	20	0.619742036	-	0	0	-	0	-	0.404300584	0	-	0	0	0	0.215441452	0	0	0	0

GenePy scores were available in 13 out of the 17 genes prioritised for the study. The PUCAI scores and the periostin levels also shown in the table.

E. 2 GenePy scores across POSTN and other associated genes in patients with CD (Chapter 6)

Probands	Periostin	Summation of GenePy Scores	CXCR4	CYTIP	EPAS1	JAG1	LEF1	NOTCH1	PIK3CG	POSTN	SRC	TCF7	TCF7L1	TCF7L2	TNC	VCAM1	VEGFA	VHL	WNT2
PR0186	20840	0.802513203	-	0	0.215212707	-	0	-	0	0	-	0.572415268	0	0	0.014885227	0	0	0	0
PR0182	23745	0.101958983	-	0	0	-	0	-	0	0	-	0	0	0	0.101958983	0	0	0	0
PR0163	27441	3.649602701	-	0	0.546498661	-	1.365892528	-	0.404300584	0	-	0	0	0	1.332910927	0	0	0	0
SOPR0310	23687	3.451648648	-	0	0	-	0	-	0.430550645	0	-	2.233697905	0	0	0.787400098	0	0	0	0
PR0172	27480	0.903624536	-	0	0	-	0	-	0	0	-	0	0.802458766	0	0.10116577	0	0	0	0
PR0187	12662	0.534364523	-	0	0	-	0	-	0	0	-	0.413880107	0	0	0.120484416	0	0	0	0
PR0174	43740	0.119805654	-	0	0	-	0	-	0	0	-	0	0	0	0.119805654	0	0	0	0
SOPR0312	29535	0.666304316	-	0	0.546498661	-	0	-	0	0	-	0	0	0	0.119805654	0	0	0	0
PR0031	24131	1.311603805	-	0	0	-	0	-	0	0	-	0.413880107	0.802458766	0	0.095264932	0	0	0	0
PR0210	37453	0.499626081	-	0	0	-	0	-	0.430550645	0	-	0	0	0	0.069075436	0	0	0	0
PR0213	40909	0.628875832	-	0	0	-	0	-	0.559800396	0	-	0	0	0	0.069075436	0	0	0	0
PR0215	56725	1.295190797	-	0	0	-	0	-	0.767520325	0	-	0.413880107	0	0	0.113790366	0	0	0	0
PR0093	26852	0.188239525	-	0	0	-	0	-	0	0	-	0	0	0	0.188239525	0	0	0	0
PR0097	13050	0.069075436	-	0	0	-	0	-	0	0	-	0	0	0	0.069075436	0	0	0	0
PR0099	21264	0.629138998	-	0	0.546498661	-	0	-	0	0	-	0	0	0	0.082640336	0	0	0	0
PR0101	28672	3.094162762	-	0	0	-	0	-	0.559800396	0	-	1.106897988	0.802458766	0	0.625005611	0	0	0	0
PR0103	31598	2.592668238	-	0	0.546498661	-	0	-	1.964469544	0	-	0	0	0	0.081700032	0	0	0	0
PR0104	37795	0.169714091	-	0	0	-	0	-	0	0	-	0	0	0	0.169714091	0	0	0	0
PR0106	11402	0.957314539	-	0	0	-	0	-	0	0	-	0.413880107	0	0	0.543434432	0	0	0	0
PR0107	38295	0.134268534	-	0	0	-	0	-	0	0	-	0	0	0	0.134268534	0	0	0	0
PR0108	26073	3.492605205	-	0	0	-	0	-	0	0	-	0	0	0	0.169714091	0	3.322891113	0	0
PR0109	36009	2.115819896	-	0	1.946105804	-	0	-	0	0	-	0	0	0	0.169714091	0	0	0	0
PR0113	21748	1.902430145	-	0	0	-	0	-	0.404300584	0.473894829	-	0	0	0.941594396	0.082640336	0	0	0	0
PR0114	29934	1.311603805	-	0	0	-	0	-	0	0	-	0.413880107	0.802458766	0	0.095264932	0	0	0	0
PR0116	26543	0.169714091	-	0	0	-	0	-	0	0	-	0	0	0	0.169714091	0	0	0	0
PR0118	20829	2.187643216	-	0	0	-	0	-	0	1.28401868	-	0	0.802458766	0	0.10116577	0	0	0	0
PR0120	30252	1.488778939	-	0	0	-	0	-	0.404300584	0	-	0	0	0	1.084478355	0	0	0	0
PR0119	18659	1.123450329	-	0	0	-	0	-	0.430550645	0	-	0.572415268	0	0	0.120484416	0	0	0	0
PR0122	35992	0.619742036	-	0	0	-	0	-	0.404300584	0	-	0	0	0	0.215441452	0	0	0	0
PR0127	19609	0.169714091	-	0	0	-	0	-	0	0	-	0	0	0	0.169714091	0	0	0	0
PR0126	26559	0.169714091	-	0	0	-	0	-	0	0	-	0	0	0	0.169714091	0	0	0	0
PR0130	48493	0.405620911	-	0	0	-	0	-	0.404300584	0	-	0	0	0	0.001320327	0	0	0	0
PR0131	21748	0.187560763	-	0	0	-	0	-	0	0	-	0	0	0	0.187560763	0	0	0	0
PR0136	31547	0.586639271	-	0	0	-	0	-	0.404300584	0	-	0	0	0	0.182338687	0	0	0	0
PR0146	20202	0.169714091	-	0	0	-	0	-	0	0	-	0	0	0	0.169714091	0	0	0	0
PR0132	33731	0.48694092	-	0	0	-	0	-	0.404300584	0	-	0	0	0	0.082640336	0	0	0	0
PR0133	37012	2.152389914	-	0	0	-	0	-	0.834851229	0	-	0.572415268	0	0	0.745123417	0	0	0	0
PR0143	24628	1.111308488	-	0	0	-	0	-	0	0	-	0	0	0.941594396	0.169714091	0	0	0	0
PR0144	44275	6.179294472	-	0	0	-	0	-	0	0	-	0	0	0	0.082640336	0	0	0	6.096654135
PR0145	13692	0.782405552	-	0	0	-	0	-	0.767520325	0	-	0	0	0	0.014885227	0	0	0	0
PR0150	40991	0.382975515	-	0	0	-	0	-	0	0	-	0	0	0	0.382975515	0	0	0	0
PR0151	31874	0.593552539	-	0	0	-	0	-	0.404300584	0	-	0	0	0	0.189251956	0	0	0	0
PR0153	16674	0.605825929	-	0	0	-	0	-	0	0	-	0.572415268	0	0	0.033410661	0	0	0	0
PR0141	20524	0.560449775	-	0	0	-	0	-	0.404300584	0	-	0	0	0	0.156149191	0	0	0	0
PR0188	39544	0.656114141	-	0	0	-	0	-	0.404300584	0	-	0	0	0	0.251813557	0	0	0	0
PR0190	22663	1.59604327	-	0	0	-	0	-	0.404300584	0	-	0	0	0	1.191742686	0	0	0	0
PR0192	30173	0.587685194	-	0	0	-	0	-	0	0.473894829	-	0	0	0	0.113790366	0	0	0	0
PR0193	23167	0.119805654	-	0	0	-	0	-	0	0	-	0	0	0	0.119805654	0	0	0	0
PR0194	24692	0.101958983	-	0	0	-	0	-	0	0	-	0	0	0	0.101958983	0	0	0	0
PR0196	34853	0.954349682	-	0	0	-	0	-	0	0	-	0.572415268	0	0	0.381934414	0	0	0	0
PR0197	19730	0.742139084	-	0	0	-	0	-	0.559800396	0	-	0	0	0	0.182338687	0	0	0	0
PR0199	25967	0.115743101	-	0	0	-	0	-	0	0	-	0	0	0	0.115743101	0	0	0	0
PR0208	32378	0.533685761	-	0	0	-	0	-	0	0	-	0.413880107	0	0	0.119805654	0	0	0	0
PR0198	27243	1.740393138	-	0	0	-	0	-	0	0.473894829	-	0.572415268	0	0	0.694083041	0	0	0	0
PR0218	32535	1.957753875	-	0	0	-	1.365892528	-	0.404300584	0	-	0	0	0	0.187560763	0	0	0	0
PR0203	18340	0.169714091	-	0	0	-	0	-	0	0	-	0	0	0	0.169714091	0	0	0	0
PR0204	31874	0.533685761	-	0	0	-	0	-	0	0	-	0.413880107	0	0	0.119805654	0	0	0	0
PR0191	22257	0.101958983	-	0	0	-	0	-	0	0	-	0	0	0	0.101958983	0	0	0	0
PR0207	27938	0.574014675	-	0	0	-	0	-	0.404300584	0	-	0	0	0	0.169714091	0	0	0	0
PR0217	18520	1.308718333	-	0	0	-	0	-	0.404300584	0	-	0	0.802458766	0	0.101958983	0	0	0	0
PR0216	33353	0.692220922	-	0	0	-	0	-	0	0	-	0.572415268	0	0	0.119805654	0	0	0	0
PR0220	33545	1.532310713	-	0	0	-	0	-	0.404300584	0.473894829	-	0.572415268	0	0	0.081700032	0	0	0	0
PR0184	16663	1.385977663	-	0	0	-	0	-	0	1.28401868	-	0	0	0	0.101958983	0	0	0	0
PR0185	23841	1.023801056	-	0	0	-	0	-	0	0	-	0	0.802458766	0	0.221342289	0	0	0	0
SOPR0345	23784	0.856366521	-	0	0	-	0	-	0	0	-	0	0	0	0.856366521	0	0	0	0
SOPR0351	29257	0.10116577	-	0	0	-	0	-	0	0	-	0	0	0	0.10116577	0	0	0	0
SOPR0368	35010	0.674374251	-	0	0	-	0	-	0	0	-	0.572415268	0	0	0.101958983	0	0	0	0
SOPR0359	24247	0.533685761	-	0	0	-	0	-	0	0	-	0.413880107	0	0	0.119805654	0	0	0	0
SOPR0380	21829	2.638286263	-	0	0	-	0	-	1.964705226	0	-	0.572415268	0	0	0.10116577	0	0	0	0
SOPR0342	28544	0.082640336	-	0	0	-	0	-	0	0	-	0	0	0	0.082640336	0	0	0	0
SOPR0304	15906	1.023801056	-	0	0	-	0	-	0	0	-	0	0.802458766	0	0.221342289	0	0	0	0
SOPR0222	23895	0.221342289	-	0	0	-	0	-	0	0	-	0	0	0	0.221342289	0	0	0	0
SOPR0313	23938	1.730627457	-	0	0	-	0	-	0	1.542387932	-	0	0	0	0.188239525	0	0	0	0
SOPR0284	17646	3.967909186	-	3.785570498	0	-	0	-	0	0	-	0	0	0	0.182338687	0	0	0	0
PR0091	29218	0.605825929	-	0	0	-	0	-	0	0	-	0.572415268	0	0	0.033410661	0	0	0	0
PR0137	25717	0.095264932	-	0	0	-	0	-	0	0	-	0	0	0	0.095264932	0	0	0	0

Appendix F

F. 1 Genes implicated in thiopurine toxicity (Chapter 7)

HUGO Gene	Symbol	Chromosome	Function	Percentage coverage Agilent v5	Percentage coverage Agilent v4
Aldehyde oxidase 1	*AOX1*	2	Catabolism of 6-MP to 6-thiouracil (6-TU)	100	88.87
ATP-binding cassette, sub-family C (CFTR/MRP), member 4	*ABCC4*	13	Efflux pump transporting a wide variety of endogenous and xenobiotic organic anionic compounds including 6-MP and 6-TGN out of the cell	95.61	72
Follistatin-Like 5	*FSTL5*	4	Involved in calcium ion binding and interacting with metalloproteases at the extracellular matrix level	91.4	57.4
Guanine monphosphate synthase	*GMPS*	3	Converts 6-TIMP to 6-thioguanines (6-TGN)	100	96.78
Glutathione-s-transferase	*GSTM1*	1	Detoxification enzymes catalysing the conjugation of electrophilic substrates to glutathione	100	42
Hypoxanthine phosphoribosyl transferase 1	*HPRT1*	X	Conversion of 6-MP to 6-TIMP	100	76.81
Inosine-5-Monophosphate Dehydrogenase 1	*IMPDH1*	7	Converts 6-TIMP to 6-thioguanines (6-TGN)	99.39	77.41
Inosine-5-Monophosphate Dehydrogenase 2	*IMPDH2*	X	Converts 6-TIMP to 6-thioguanines (6-TGN)	100	100
Inosine triphosphatase (nucleoside triphosphate pyrophosphatase)	*ITPA*	20	De-phosphorylation of 6-thioinosine triphosphate (6-TITP) to 6-thioinosine monophosphate (6-TIMP) in the thiopurine metabolic pathway	89.65	76.99
Interleukin 6 signal transducer	*IL6ST*	5	Signal transducer protein shared by several cytokines including IL-6	96.55	32.14
Methylenetetrahydro folate reductase	*MTHFR*	1	Intracellular folate metabolism and purine-pyrimidine synthesis	95.3	65.7
Molybdenum cofactor sulfurase	*MOCOS*	18	Sulfuration of molybdenum co-factor in the enzymes XDH and AOX1; essential for their enzymatic activities	100	100
Nudix (nucleoside diphosphate linked moiety X)-type motif 15	*NUDT15*	13	Involved in DNA repair	95.99	43.12
Protein kinase C and casein kinase substrate in neurons 2	*PACSIN2*	22	Family of proteins, involved in various biological processes including endocytosis, cell-cycle control and autophagy	89.01	49.78
Xanthine dehydrogenase (Synonyms- XO)	*XDH*	2	Catabolism of 6-MP to 6-thiouracil (6-TU)	100	75.94

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Zhou N, Chen WX, Chen SH, Xu CF, Li YM. Inflammatory bowel disease unclassified. Journal of Zhejiang University Science B 2011;12:280-6.

2. Levine A, Koletzko S, Turner D, et al. ESPGHAN revised porto criteria for the diagnosis of inflammatory bowel disease in children and adolescents. J Pediatr Gastroenterol Nutr 2014;58:795-806.

3. Uhlig HH, Powrie F. Translating Immunology into Therapeutic Concepts for Inflammatory Bowel Disease. Annu Rev Immunol 2018;36:755-81.

4. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. Nature 2011;474:307-17.

5. Olivera P, Danese S, Jay N, Natoli G, Peyrin-Biroulet L. Big data in IBD: a look into the future. Nat Rev Gastroenterol Hepatol 2019.

6. Torres J, Mehandru S, Colombel JF, Peyrin-Biroulet L. Crohn's disease. Lancet 2017;389:1741-55.

7. Torres J, Colombel JF. Genetics and phenotypes in inflammatory bowel disease. Lancet 2016;387:98-100.

8. Rialon KL, Crowley E, Seemann NM, Fahy AS, Muise A, Langer JC. Long-term outcomes for children with very early-onset colitis: Implications for surgical management. J Pediatr Surg 2018;53:964-7.

9. Ashton JJ, Mossotto E, Ennis S, Beattie RM. Personalising medicine in inflammatory bowel disease-current and future perspectives. Transl Pediatr 2019;8:56-69.

10. Cleynen I, Boucher G, Jostins L, et al. Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. Lancet 2016;387:156-67.

11. Levine A, Griffiths A, Markowitz J, et al. Pediatric modification of the Montreal classification for inflammatory bowel disease: the Paris classification. Inflamm Bowel Dis 2011;17:1314-21.

12. Eszter Muller K, Laszlo Lakatos P, Papp M, Veres G. Incidence and paris classification of pediatric inflammatory bowel disease. Gastroenterol Res Pract 2014;2014:904307.

13. Jakobsen C, Bartek J, Jr., Wewer V, et al. Differences in phenotype and disease course in adult and paediatric inflammatory bowel disease--a population-based study. Aliment Pharmacol Ther 2011;34:1217-24.

14. Uhlig HH, Schwerd T, Koletzko S, et al. The diagnostic approach to monogenic very early onset inflammatory bowel disease. Gastroenterology 2014;147:990-1007 e3.

15. Rosen MJ, Dhawan A, Saeed SA. Inflammatory Bowel Disease in Children and Adolescents. JAMA Pediatr 2015;169:1053-60.

16. Shim JO. Recent advance in very early-onset inflammatory bowel disease. Intest Res 2019;17:9-16.

17. Shivade C, Raghavan P, Fosler-Lussier E, et al. A review of approaches to identifying patient phenotype cohorts using electronic health records. J Am Med Inform Assoc 2014;21:221-30.

18. Altman RB, Ashley EA. Using "big data" to dissect clinical heterogeneity. Circulation 2015;131:232-3.

19. Ashton JJ, Beattie RM. Personalised therapy for inflammatory bowel disease. Lancet 2019.

20. Benchimol EI, Fortinsky KJ, Gozdyra P, Van den Heuvel M, Van Limbergen J, Griffiths AM. Epidemiology of pediatric inflammatory bowel disease: a systematic review of international trends. Inflammatory bowel diseases 2011;17:423-39.

21. Ng SC, Bernstein CN, Vatn MH, et al. Geographical variability and environmental risk factors in inflammatory bowel disease. Gut 2013;62:630-49.

22. Benchimol EI, Guttmann A, Griffiths AM, et al. Increasing incidence of paediatric inflammatory bowel disease in Ontario, Canada: evidence from health administrative data. Gut 2009;58:1490-7.

23. Griffiths AM. Specificities of inflammatory bowel disease in childhood. Best practice & research Clinical gastroenterology 2004;18:509-23.

24. Sykora J, Pomahacova R, Kreslova M, Cvalinova D, Stych P, Schwarz J. Current global trends in the incidence of pediatric-onset inflammatory bowel disease. World J Gastroenterol 2018;24:2741-63.

25. Henderson P, Hansen R, Cameron FL, et al. Rising incidence of pediatric inflammatory bowel disease in Scotland. Inflammatory bowel diseases 2012;18:999-1005.

26. Ashton JJ, Wiskin AE, Ennis S, Batra A, Afzal NA, Beattie RM. Rising incidence of paediatric inflammatory bowel disease (PIBD) in Wessex, Southern England. Arch Dis Child 2014;99:659-64.

27. Kevans D, Silverberg MS, Borowski K, et al. IBD Genetic Risk Profile in Healthy First-Degree Relatives of Crohn's Disease Patients. J Crohns Colitis 2016;10:209-15.

28. Childers RE, Eluri S, Vazquez C, Weise RM, Bayless TM, Hutfless S. Family history of inflammatory bowel disease among patients with ulcerative colitis: a systematic review and meta-analysis. J Crohns Colitis 2014;8:1480-97.

29. Ek WE, D'Amato M, Halfvarson J. The history of genetics in inflammatory bowel disease. Ann Gastroenterol 2014;27:294-303.

30. Santos MPC, Gomes C, Torres J. Familial and ethnic risk in inflammatory bowel disease. Ann Gastroenterol 2018;31:14-23.

31. Brant SR. Update on the heritability of inflammatory bowel disease: the importance of twin studies. Inflamm Bowel Dis 2011;17:1-5.

32. Sawczenko A, Lynn R, Sandhu BK. Variations in initial assessment and management of inflammatory bowel disease across Great Britain and Ireland. Archives of disease in childhood 2003;88:990-4.

33. Hyams JS. Extraintestinal manifestations of inflammatory bowel disease in children. Journal of pediatric gastroenterology and nutrition 1994;19:7-21.

34. Levine JS, Burakoff R. Extraintestinal manifestations of inflammatory bowel disease. Gastroenterology & hepatology 2011;7:235-41.

35. Sandhu BK FJ, Beattie RM, Mitton SG. Guidelines for the Management of Inflammatory Bowel Disease in Children in the United Kingdom. J Pediatr Gastroenterol Nutr 2010 Feb;50 Suppl 1:S1-13 2010.

36. Ashton JJ, Gavin J, Beattie RM. Exclusive enteral nutrition in Crohn's disease: Evidence and practicalities. Clin Nutr 2019;38:80-9.

37. Ruemmele FM, Veres G, Kolho KL, et al. Consensus guidelines of ECCO/ESPGHAN on the medical management of pediatric Crohn's disease. J Crohns Colitis 2014;8:1179-207.

38. Lim SZ, Chua EW. Revisiting the Role of Thiopurines in Inflammatory Bowel Disease Through Pharmacogenomics and Use of Novel Methods for Therapeutic Drug Monitoring. Front Pharmacol 2018;9:1107.

39. Sandhu BK, Fell JM, Beattie RM, et al. Guidelines for the management of inflammatory bowel disease in children in the United Kingdom. J Pediatr Gastroenterol Nutr 2010;50 Suppl 1:S1-13.

40. D'Haens GR. Top-down therapy for IBD: rationale and requisite evidence. Nat Rev Gastroenterol Hepatol 2010;7:86-92.

41. Cozijnsen MA, van Pieterson M, Samsom JN, Escher JC, de Ridder L. Top-down Infliximab Study in Kids with Crohn's disease (TISKids): an international multicentre randomised controlled trial. BMJ Open Gastroenterol 2016;3:e000123.

42. Turner D, Ruemmele FM, Orlanski-Meyer E, et al. Management of Paediatric Ulcerative Colitis, Part 1: Ambulatory Care- an Evidence-Based Guideline from ECCO and ESPGHAN. J Pediatr Gastroenterol Nutr 2018.

43. Turner D, Ruemmele FM, Orlanski-Meyer E, et al. Management of Paediatric Ulcerative Colitis, Part 2: Acute Severe Colitis; An Evidence-based Consensus Guideline from ECCO and ESPGHAN. J Pediatr Gastroenterol Nutr 2018.

44. Feagan BG, Sandborn WJ, Gasink C, et al. Ustekinumab as Induction and Maintenance Therapy for Crohn's Disease. N Engl J Med 2016;375:1946-60.

45. Buhl S, Steenholdt C, Rasmussen M, et al. Outcomes After Primary Infliximab Treatment Failure in Inflammatory Bowel Disease. Inflamm Bowel Dis 2017;23:1210-7.

46. Allez M, Karmiris K, Louis E, et al. Report of the ECCO pathogenesis workshop on anti-TNF therapy failures in inflammatory bowel diseases: definitions, frequency and pharmacological aspects. J Crohns Colitis 2010;4:355-66.

47. Mosli MH, Sandborn WJ, Kim RB, Khanna R, Al-Judaibi B, Feagan BG. Toward a personalized medicine approach to the management of inflammatory bowel disease. The American journal of gastroenterology 2014;109:994-1004.

48. Hyams JS, Davis Thomas S, Gotman N, et al. Clinical and biological predictors of response to standardised paediatric colitis therapy (PROTECT): a multicentre inception cohort study. Lancet 2019.

49. Ley RE, Peterson DA, Gordon JI. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell 2006;124:837-48.

50. Cader MZ, Kaser A. Recent advances in inflammatory bowel disease: mucosal immune cells in intestinal inflammation. Gut 2013;62:1653-64.

51. Feng T, Elson CO. Adaptive immunity in the host-microbiota dialog. Mucosal Immunol 2011;4:15-21.

52. Okumura R, Takeda K. Maintenance of gut homeostasis by the mucosal immune system. Proc Jpn Acad Ser B Phys Biol Sci 2016;92:423-35.

53. Elia PP, Tolentino YF, Bernardazzi C, de Souza HS. The role of innate immunity receptors in the pathogenesis of inflammatory bowel disease. Mediators Inflamm 2015;2015:936193.

54. Siegmund B, Zeitz M. Innate and adaptive immunity in inflammatory bowel disease. World J Gastroenterol 2011;17:3178-83.

55. Imam T, Park S, Kaplan MH, Olson MR. Effector T Helper Cell Subsets in Inflammatory Bowel Diseases. Front Immunol 2018;9:1212.

56. Manetti R, Parronchi P, Giudizi MG, et al. Natural killer cell stimulatory factor (interleukin 12 [IL-12]) induces T helper type 1 (Th1)-specific immune responses and inhibits the development of IL-4-producing Th cells. J Exp Med 1993;177:1199-204.

57. Walker JA, McKenzie ANJ. TH2 cell development and function. Nat Rev Immunol 2018;18:121-33.

58. Brand S. Crohn's disease: Th1, Th17 or both? The change of a paradigm: new immunological and genetic insights implicate Th17 cells in the pathogenesis of Crohn's disease. Gut 2009;58:1152-67.

59. Murphy-Ullrich JE, Sage EH. Revisiting the matricellular concept. Matrix Biol 2014;37:1-14.

60. Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 2012;491:119-24.

61. Goto Y, Kurashima Y, Kiyono H. The gut microbiota and inflammatory bowel disease. Current opinion in rheumatology 2015;27:388-96.

62. Uhlig HH, Muise AM. Clinical Genomics in Inflammatory Bowel Disease. Trends Genet 2017;33:629-41.

63. Rogler G, Vavricka S. Exposome in IBD: recent insights in environmental factors that influence the onset and course of IBD. Inflammatory bowel diseases 2015;21:400-8.

64. Bernstein CN, Shanahan F. Disorders of a modern lifestyle: reconciling the epidemiology of inflammatory bowel diseases. Gut 2008;57:1185-91.

65. Maaser C, Langholz E, Gordon H, et al. European Crohn's and Colitis Organisation Topical Review on Environmental Factors in IBD. Journal of Crohn's & colitis 2017;11:905-20.

66. Kaplan GG. The global burden of IBD: from 2015 to 2025. Nature reviews Gastroenterology & hepatology 2015;12:720-7.

67. Crohn BB, Ginzburg L, Oppenheimer GD. Landmark article Oct 15, 1932. Regional ileitis. A pathological and clinical entity. By Burril B. Crohn, Leon Ginzburg, and Gordon D. Oppenheimer. JAMA : the journal of the American Medical Association 1984;251:73-9.

68. Henderson P, Satsangi J. Genes in inflammatory bowel disease: lessons from complex diseases. Clinical medicine 2011;11:8-10.

69. Peeters M, Nevens H, Baert F, et al. Familial aggregation in Crohn's disease: increased age-adjusted risk and concordance in clinical characteristics. Gastroenterology 1996;111:597-603.

70. Tysk C, Lindberg E, Jarnerot G, Floderus-Myrhed B. Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. Gut 1988;29:990-6.

71. Brant SR, Shugart YY. Inflammatory bowel disease gene hunting by linkage analysis: rationale, methodology, and present status of the field. Inflammatory bowel diseases 2004;10:300-11.

72. Cardinale CJ, Kelsen JR, Baldassano RN, Hakonarson H. Impact of exome sequencing in inflammatory bowel disease. World journal of gastroenterology : WJG 2013;19:6721-9.

73. Morton NE. Sequential tests for the detection of linkage. American journal of human genetics 1955;7:277-318.

74. Altmuller J, Palmer LJ, Fischer G, Scherb H, Wjst M. Genomewide scans of complex human diseases: true linkage is hard to find. American journal of human genetics 2001;69:936-50.

75. Rice JP, Saccone NL, Corbett J. The lod score method. Advances in genetics 2001;42:99-113.

76. Hugot JP, Laurent-Puig P, Gower-Rousseau C, et al. Mapping of a susceptibility locus for Crohn's disease on chromosome 16. Nature 1996;379:821-3.

77. Satsangi J, Parkes M, Louis E, et al. Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. Nature genetics 1996;14:199-202.

78. Cho JH, Nicolae DL, Gold LH, et al. Identification of novel susceptibility loci for inflammatory bowel disease on chromosomes 1p, 3q, and 4q: evidence for epistasis between 1p and IBD1. Proceedings of the National Academy of Sciences of the United States of America 1998;95:7502-7.

79. Hampe J, Schreiber S, Shaw SH, et al. A genomewide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. American journal of human genetics 1999;64:808-16.

80. Ma Y, Ohmen JD, Li Z, et al. A genome-wide search identifies potential new susceptibility loci for Crohn's disease. Inflammatory bowel diseases 1999;5:271-8.

81. Duerr RH, Barmada MM, Zhang L, Pfutzer R, Weeks DE. High-density genome scan in Crohn disease shows confirmed linkage to chromosome 14q11-12. American journal of human genetics 2000;66:1857-62.

82. Rioux JD, Silverberg MS, Daly MJ, et al. Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. American journal of human genetics 2000;66:1863-70.

83. Williams CN, Kocher K, Lander ES, Daly MJ, Rioux JD. Using a genome-wide scan and meta-analysis to identify a novel IBD locus and confirm previously identified IBD loci. Inflammatory bowel diseases 2002;8:375-81.

84. Paavola-Sakki P, Ollikainen V, Helio T, et al. Genome-wide search in Finnish families with inflammatory bowel disease provides evidence for novel susceptibility loci. European journal of human genetics : EJHG 2003;11:112-20.

85. van Heel DA, Dechairo BM, Dawson G, et al. The IBD6 Crohn's disease locus demonstrates complex interactions with CARD15 and IBD5 disease-associated variants. Human molecular genetics 2003;12:2569-75.

86. Barmada MM, Brant SR, Nicolae DL, et al. A genome scan in 260 inflammatory bowel disease-affected relative pairs. Inflammatory bowel diseases 2004;10:513-20.

87. Hugot JP, Chamaillard M, Zouali H, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 2001;411:599-603.

88. Pearson TA, Manolio TA. How to interpret a genome-wide association study. JAMA : the journal of the American Medical Association 2008;299:1335-44.

89. International HapMap C. The International HapMap Project. Nature 2003;426:789-96.

90. Tanaka T. [International HapMap project]. Nihon rinsho Japanese journal of clinical medicine 2005;63 Suppl 12:29-34.

91. Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661-78.

92. Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. J Clin Invest 2008;118:1590-605.

93. Marian AJ. Molecular genetic studies of complex phenotypes. Transl Res 2012;159:64-79.

94. Hardy J, Singleton A. Genomewide association studies and human disease. The New England journal of medicine 2009;360:1759-68.

95. Mathew CG. New links to the pathogenesis of Crohn disease provided by genome-wide association scans. Nature reviews Genetics 2008;9:9-14.

96. Kraft P, Zeggini E, Ioannidis JP. Replication in genome-wide association studies. Stat Sci 2009;24:561-73.

97. Witte JS. Genome-wide association studies and beyond. Annual review of public health 2010;31:9-20 4 p following

98. Liu JZ, van Sommeren S, Huang H, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nature genetics 2015;47:979-86.

99. de Lange KM, Moutsianas L, Lee JC, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nat Genet 2017;49:256-61.

100. McGovern DPB. Genetic Underpinnings of IBD: Immunity, Infection, and More. wwwmedscapecom 2013.

101. Zhernakova A, van Diemen CC, Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. Nature reviews Genetics 2009;10:43-55.

102. Gregersen PK, Olsson LM. Recent advances in the genetics of autoimmune disease. Annual review of immunology 2009;27:363-91.

103. Lees CW, Barrett JC, Parkes M, Satsangi J. New IBD genetics: common pathways with other diseases. Gut 2011;60:1739-53.

104. Ellinghaus D, Jostins L, Spain SL, et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. Nat Genet 2016;48:510-8.

105. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. American journal of human genetics 2011;88:294-305.

106. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era--concepts and misconceptions. Nature reviews Genetics 2008;9:255-66.

107. Lee JC, Parkes M. Genome-wide association studies and Crohn's disease. Briefings in functional genomics 2011;10:71-6.

108. Ritchie MD, Van Steen K. The search for gene-gene interactions in genome-wide association studies: challenges in abundance of methods, practical considerations, and biological interpretation. Ann Transl Med 2018;6:157.

109. Tucker T, Marra M, Friedman JM. Massively parallel sequencing: the next big thing in genetic medicine. American journal of human genetics 2009;85:142-54.

110. Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. The New England journal of medicine 2013;369:1502-11.

111. MacArthur DG, Balasubramanian S, Frankish A, et al. A systematic survey of loss-of-function variants in human protein-coding genes. Science 2012;335:823-8.

112. Lelieveld SH, Spielmann M, Mundlos S, Veltman JA, Gilissen C. Comparison of Exome and Genome Sequencing Technologies for the Complete Capture of Protein-Coding Regions. Hum Mutat 2015;36:815-22.

113. Schwarze K, Buchanan J, Taylor JC, Wordsworth S. Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. Genet Med 2018;20:1122-30.

114. Meienberg J, Bruggmann R, Oexle K, Matyas G. Clinical sequencing: is WGS the better WES? Hum Genet 2016;135:359-62.

115. Tennessen JA, Bigham AW, O'Connor TD, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 2012;337:64-9.

116. Puckelwartz MJ, McNally EM. Genetic profiling for risk reduction in human cardiovascular disease. Genes 2014;5:214-34.

117. Choi M, Scholl UI, Ji W, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proceedings of the National Academy of Sciences of the United States of America 2009;106:19096-101.

118. Ng SB, Turner EH, Robertson PD, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature 2009;461:272-6.

119. Do R, Kathiresan S, Abecasis GR. Exome sequencing and complex disease: practical aspects of rare variant association studies. Human molecular genetics 2012;21:R1-9.

120. Need AC, Shashi V, Hitomi Y, et al. Clinical application of exome sequencing in undiagnosed genetic conditions. Journal of medical genetics 2012;49:353-61.

121. Bamshad MJ, Ng SB, Bigham AW, et al. Exome sequencing as a tool for Mendelian disease gene discovery. Nature reviews Genetics 2011;12:745-55.

122. Worthey EA, Mayer AN, Syverson GD, et al. Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. Genet Med 2011;13:255-62.

123. Girardelli M, Loganes C, Pin A, et al. Novel NOD2 Mutation in Early-Onset Inflammatory Bowel Phenotype. Inflamm Bowel Dis 2018;24:1204-12.

124. Li J, Moran T, Swanson E, et al. Regulation of IL-8 and IL-1beta expression in Crohn's disease associated NOD2/CARD15 mutations. Hum Mol Genet 2004;13:1715-25.

125. Parkhouse R, Boyle JP, Monie TP. Blau syndrome polymorphisms in NOD2 identify nucleotide hydrolysis and helical domain 1 as signalling regulators. FEBS Lett 2014;588:3382-9.

126. Li Q, Lee CH, Peters LA, et al. Variants in TRIM22 That Affect NOD2 Signaling Are Associated With Very-Early-Onset Inflammatory Bowel Disease. Gastroenterology 2016;150:1196-207.

127. Taylor JC, Martin HC, Lise S, et al. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. Nature genetics 2015;47:717-26.

128. Belkadi A, Bolze A, Itan Y, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. Proceedings of the National Academy of Sciences of the United States of America 2015;112:5473-8.

129. Laukoetter MG, Nava P, Nusrat A. Role of the intestinal barrier in inflammatory bowel disease. World journal of gastroenterology : WJG 2008;14:401-7.

130. Johansson ME, Gustafsson JK, Holmen-Larsson J, et al. Bacteria penetrate the normally impenetrable inner colon mucus layer in both murine colitis models and patients with ulcerative colitis. Gut 2014;63:281-91.

131. Sun J, Shen X, Li Y, et al. Therapeutic Potential to Modify the Mucus Barrier in Inflammatory Bowel Disease. Nutrients 2016;8.

132. McCole DF. IBD candidate genes and intestinal barrier regulation. Inflamm Bowel Dis 2014;20:1829-49.

133. Muise AM, Walters TD, Glowacka WK, et al. Polymorphisms in E-cadherin (CDH1) result in a mis-localised cytoplasmic protein that is associated with Crohn's disease. Gut 2009;58:1121-7.

134. McCole DF. Regulation of epithelial barrier function by the inflammatory bowel disease candidate gene, PTPN2. Annals of the New York Academy of Sciences 2012;1257:108-14.

135. Bevins CL, Salzman NH. Paneth cells, antimicrobial peptides and maintenance of intestinal homeostasis. Nature reviews Microbiology 2011;9:356-68.

136. Rescigno M. Intestinal dendritic cells. Advances in immunology 2010;107:109-38.

137. Zhu L, Shi T, Zhong C, Wang Y, Chang M, Liu X. IL-10 and IL-10 Receptor Mutations in Very Early Onset Inflammatory Bowel Disease. Gastroenterology Res 2017;10:65-9.

138. Smythies LE, Sellers M, Clements RH, et al. Human intestinal macrophages display profound inflammatory anergy despite avid phagocytic and bacteriocidal activity. J Clin Invest 2005;115:66-75.

139. Goldberg R, Prescott N, Lord GM, MacDonald TT, Powell N. The unusual suspects--innate lymphoid cells as novel therapeutic targets in IBD. Nat Rev Gastroenterol Hepatol 2015;12:271-83.

140. Sedda S, Marafini I, Figliuzzi MM, Pallone F, Monteleone G. An overview of the role of innate lymphoid cells in gut infections and inflammation. Mediators of inflammation 2014;2014:235460.

141. Bernink JH, Peters CP, Munneke M, et al. Human type 1 innate lymphoid cells accumulate in inflamed mucosal tissues. Nature immunology 2013;14:221-9.

142. Takayama T, Kamada N, Chinen H, et al. Imbalance of NKp44(+)NKp46(-) and NKp44(-)NKp46(+) natural killer cells in the intestinal mucosa of patients with Crohn's disease. Gastroenterology 2010;139:882-92, 92 e1-3.

143. Ciccia F, Accardo-Palumbo A, Alessandro R, et al. Interleukin-22 and interleukin-22-producing NKp44+ natural killer cells in subclinical gut inflammation in ankylosing spondylitis. Arthritis and rheumatism 2012;64:1869-78.

144. Fournier BM, Parkos CA. The role of neutrophils during intestinal inflammation. Mucosal immunology 2012;5:354-66.

145. Bornstein P. Diversity of function is inherent in matricellular proteins: an appraisal of thrombospondin 1. J Cell Biol 1995;130:503-6.

146. Liu AY, Zheng H, Ouyang G. Periostin, a multifunctional matricellular protein in inflammatory and tumor microenvironments. Matrix Biol 2014;37:150-6.

147. Masuoka M, Shiraishi H, Ohta S, et al. Periostin promotes chronic allergic inflammation in response to Th2 cytokines. J Clin Invest 2012;122:2590-600.

148. Ruan K, Bao S, Ouyang G. The multifaceted role of periostin in tumorigenesis. Cell Mol Life Sci 2009;66:2219-30.

149. Underwood TJ, Hayden AL, Derouet M, et al. Cancer-associated fibroblasts predict poor outcome and promote periostin-dependent invasion in oesophageal adenocarcinoma. J Pathol 2015;235:466-77.

150. Koh SJ, Choi Y, Kim BG, et al. Matricellular Protein Periostin Mediates Intestinal Inflammation through the Activation of Nuclear Factor kappaB Signaling. PLoS One 2016;11:e0149652.

151. Kikuchi Y, Kashima TG, Nishiyama T, et al. Periostin is expressed in pericryptal fibroblasts and cancer-associated fibroblasts in the colon. J Histochem Cytochem 2008;56:753-64.

152. Bao S, Ouyang G, Bai X, et al. Periostin potently promotes metastatic growth of colon cancer by augmenting cell survival via the Akt/PKB pathway. Cancer Cell 2004;5:329-39.

153. Malanchi I, Santamaria-Martinez A, Susanto E, et al. Interactions between cancer stem cells and their niche govern metastatic colonization. Nature 2011;481:85-9.

154. Wu T, Ouyang G. Matricellular proteins: multifaceted extracellular regulators in tumor dormancy. Protein Cell 2014;5:249-52.

155. Conway SJ, Izuhara K, Kudo Y, et al. The role of periostin in tissue remodeling across health and disease. Cell Mol Life Sci 2014;71:1279-88.

156. Naik PK, Bozyk PD, Bentley JK, et al. Periostin promotes fibrosis and predicts progression in patients with idiopathic pulmonary fibrosis. Am J Physiol Lung Cell Mol Physiol 2012;303:L1046-56.

157. Karin M, Clevers H. Reparative inflammation takes charge of tissue regeneration. Nature 2016;529:307-15.

158. Algeri M, Conforti A, Pitisci A, et al. Mesenchymal stromal cells and chronic inflammatory bowel disease. Immunol Lett 2015;168:191-200.

159. Lazebnik LB, Konopliannikov AG, Kniazev OV, et al. [Use of allogeneic mesenchymal stem cells in the treatment of intestinal inflammatory diseases]. Ter Arkh 2010;82:38-43.

160. Forbes GM, Sturm MJ, Leong RW, et al. A phase 2 study of allogeneic mesenchymal stromal cells for luminal Crohn's disease refractory to biologic therapy. Clin Gastroenterol Hepatol 2014;12:64-71.

161. Regmi S, Pathak S, Kim JO, Yong CS, Jeong JH. Mesenchymal stem cell therapy for the treatment of inflammatory diseases: Challenges, opportunities, and future perspectives. Eur J Cell Biol 2019.

162. Delves PJ. Essential Immunology. Singapore: Wiley-Blackwell; 2011.

163. Akira S, Uematsu S, Takeuchi O. Pathogen recognition and innate immunity. Cell 2006;124:783-801.

164. Abreu MT. Toll-like receptor signalling in the intestinal epithelium: how bacterial recognition shapes intestinal function. Nature reviews Immunology 2010;10:131-44.

165. Lee J, Mo JH, Katakura K, et al. Maintenance of colonic homeostasis by distinctive apical TLR9 signalling in intestinal epithelial cells. Nature cell biology 2006;8:1327-36.

166. Cario E, Podolsky DK. Differential alteration in intestinal epithelial cell expression of toll-like receptor 3 (TLR3) and TLR4 in inflammatory bowel disease. Infect Immun 2000;68:7010-7.

167. Wang F, Tahara T, Arisawa T, et al. Genetic polymorphisms of CD14 and Toll-like receptor-2 (TLR2) in patients with ulcerative colitis. Journal of gastroenterology and hepatology 2007;22:925-9.

168. Shen X, Shi R, Zhang H, Li K, Zhao Y, Zhang R. The Toll-like receptor 4 D299G and T399I polymorphisms are associated with Crohn's disease and ulcerative colitis: a meta-analysis. Digestion 2010;81:69-77.

169. Hume GE, Fowler EV, Doecke J, et al. Novel NOD2 haplotype strengthens the association between TLR4 Asp299gly and Crohn's disease in an Australian population. Inflammatory bowel diseases 2008;14:585-90.

170. Arnott ID, Nimmo ER, Drummond HE, et al. NOD2/CARD15, TLR4 and CD14 mutations in Scottish and Irish Crohn's disease patients: evidence for genetic heterogeneity within Europe? Genes and immunity 2004;5:417-25.

171. Williams A, Flavell RA, Eisenbarth SC. The role of NOD-like Receptors in shaping adaptive immunity. Current opinion in immunology 2010;22:34-40.

172. Robertson SJ, Rubino SJ, Geddes K, Philpott DJ. Examining host-microbial interactions through the lens of NOD: From plants to mammals. Seminars in immunology 2012;24:9-16.

173. Fritz T, Niederreiter L, Adolph T, Blumberg RS, Kaser A. Crohn's disease: NOD2, autophagy and ER stress converge. Gut 2011;60:1580-8.

174. Caruso R, Warner N, Inohara N, Nunez G. NOD1 and NOD2: signaling, host defense, and inflammatory disease. Immunity 2014;41:898-908.

175. Bertrand MJ, Doiron K, Labbe K, Korneluk RG, Barker PA, Saleh M. Cellular inhibitors of apoptosis cIAP1 and cIAP2 are required for innate immunity signaling by the pattern recognition receptors NOD1 and NOD2. Immunity 2009;30:789-801.

176. Goncharov T, Hedayati S, Mulvihill MM, et al. Disruption of XIAP-RIP2 Association Blocks NOD2-Mediated Inflammatory Signaling. Mol Cell 2018;69:551-65 e7.

177. Van Limbergen J, Radford-Smith G, Satsangi J. Advances in IBD genetics. Nature reviews Gastroenterology & hepatology 2014;11:372-85.

178. Lesage S, Zouali H, Cezard JP, et al. CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. American journal of human genetics 2002;70:845-57.

179. Adler J, Rangwalla SC, Dwamena BA, Higgins PD. The prognostic power of the NOD2 genotype for complicated Crohn's disease: a meta-analysis. The American journal of gastroenterology 2011;106:699-712.

180. Van Limbergen J, Nimmo ER, Russell RK, et al. Investigation of NOD1/CARD4 variation in inflammatory bowel disease using a haplotype-tagging strategy. Human molecular genetics 2007;16:2175-86.

181. Kim JG, Lee SJ, Kagnoff MF. Nod1 is an essential signal transducer in intestinal epithelial cells infected with bacteria that avoid recognition by toll-like receptors. Infection and immunity 2004;72:1487-95.

182. McGovern DP, Hysi P, Ahmad T, et al. Association between a complex insertion/deletion polymorphism in NOD1 (CARD4) and susceptibility to inflammatory bowel disease. Human molecular genetics 2005;14:1245-50.

183. Brain O, Allan P, Simmons A. NOD2-mediated autophagy and Crohn disease. Autophagy 2010;6:412-4.

184. Parkes M, Barrett JC, Prescott NJ, et al. Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. Nat Genet 2007;39:830-2.

185. Hampe J, Franke A, Rosenstiel P, et al. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. Nat Genet 2007;39:207-11.

186. Muzes G, Tulassay Z, Sipos F. Interplay of autophagy and innate immunity in Crohn's disease: a key immunobiologic feature. World journal of gastroenterology : WJG 2013;19:4447-54.

187. Rufini S, Ciccacci C, Di Fusco D, et al. Autophagy and inflammatory bowel disease: Association between variants of the autophagy-related IRGM gene and susceptibility to Crohn's disease. Dig Liver Dis 2015;47:744-50.

188. Bogaert S, De Vos M, Olievier K, et al. Involvement of endoplasmic reticulum stress in inflammatory bowel disease: a different implication for colonic and ileal disease? PloS one 2011;6:e25589.

189. Kaser A, Lee AH, Franke A, et al. XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease. Cell 2008;134:743-56.

190. Fujino S, Andoh A, Bamba S, et al. Increased expression of interleukin 17 in inflammatory bowel disease. Gut 2003;52:65-70.

191. Rovedatti L, Kudo T, Biancheri P, et al. Differential regulation of interleukin 17 and interferon gamma production in inflammatory bowel disease. Gut 2009;58:1629-36.

192. Basu R, Hatton RD, Weaver CT. The Th17 family: flexibility follows function. Immunol Rev 2013;252:89-103.

193. Dambacher J, Beigel F, Zitzmann K, et al. The role of the novel Th17 cytokine IL-26 in intestinal inflammation. Gut 2009;58:1207-17.

194. Kleinschek MA, Boniface K, Sadekova S, et al. Circulating and gut-resident human Th17 cells express CD161 and promote intestinal inflammation. The Journal of experimental medicine 2009;206:525-34.

195. Jin W, Dong C. IL-17 cytokines in immunity and inflammation. Emerging microbes & infections 2013;2:e60.

196. Hundorfean G, Neurath MF, Mudter J. Functional relevance of T helper 17 (Th17) cells and the IL-17 cytokine family in inflammatory bowel disease. Inflammatory bowel diseases 2012;18:180-6.

197. Franke A, McGovern DP, Barrett JC, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nature genetics 2010;42:1118-25.

198. Sandborn WJ, Feagan BG, Fedorak RN, et al. A randomized trial of Ustekinumab, a human interleukin-12/23 monoclonal antibody, in patients with moderate-to-severe Crohn's disease. Gastroenterology 2008;135:1130-41.

199. Glocker EO, Kotlarz D, Boztug K, et al. Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. N Engl J Med 2009;361:2033-45.

200. Chaudhry A, Samstein RM, Treuting P, et al. Interleukin-10 signaling in regulatory T cells is required for suppression of Th17 cell-mediated inflammation. Immunity 2011;34:566-78.

201. Moran CJ, Walters TD, Guo CH, et al. IL-10R polymorphisms are associated with very-early-onset ulcerative colitis. Inflammatory bowel diseases 2013;19:115-23.

202. Engelhardt KR, Shah N, Faizura-Yeop I, et al. Clinical outcome in IL-10- and IL-10 receptor-deficient patients with or without hematopoietic stem cell transplantation. The Journal of allergy and clinical immunology 2013;131:825-30.

203. Kotlarz D, Beier R, Murugan D, et al. Loss of interleukin-10 signaling and infantile inflammatory bowel disease: implications for diagnosis and therapy. Gastroenterology 2012;143:347-55.

204. O'Shea JJ, Holland SM, Staudt LM. JAKs and STATs in immunity, immunodeficiency, and cancer. N Engl J Med 2013;368:161-70.

205. Coskun M, Salem M, Pedersen J, Nielsen OH. Involvement of JAK/STAT signaling in the pathogenesis of inflammatory bowel disease. Pharmacological research : the official journal of the Italian Pharmacological Society 2013;76:1-8.

206. De Vries LCS, Wildenberg ME, De Jonge WJ, D'Haens GR. The Future of Janus Kinase Inhibitors in Inflammatory Bowel Disease. J Crohns Colitis 2017;11:885-93.

207. Kisseleva T, Bhattacharya S, Braunstein J, Schindler CW. Signaling through the JAK/STAT pathway, recent advances and future challenges. Gene 2002;285:1-24.

208. Wang X, Lupardus P, Laporte SL, Garcia KC. Structural biology of shared cytokine receptors. Annual review of immunology 2009;27:29-60.

209. Hammer T, Nielsen KR, Munkholm P, Burisch J, Lynge E. The Faroese IBD Study: Incidence of Inflammatory Bowel Diseases Across 54 Years of Population-based Data. Journal of Crohn's & colitis 2016;10:934-42.

210. de Lange KM, Barrett JC. Understanding inflammatory bowel disease via immunogenetics. J Autoimmun 2015;64:91-100.

211. Kostic AD, Xavier RJ, Gevers D. The microbiome in inflammatory bowel disease: current status and the future ahead. Gastroenterology 2014;146:1489-99.

212. Knox NC, Forbes JD, Van Domselaar G, Bernstein CN. The Gut Microbiome as a Target for IBD Treatment: Are We There Yet? Curr Treat Options Gastroenterol 2019;17:115-26.

213. Sheehan D, Moran C, Shanahan F. The microbiota in inflammatory bowel disease. Journal of gastroenterology 2015;50:495-507.

214. Suskind DL, Brittnacher MJ, Wahbeh G, et al. Fecal microbial transplant effect on clinical outcomes and fecal microbiome in active Crohn's disease. Inflammatory bowel diseases 2015;21:556-63.

215. Lopez J, Grinspan A. Fecal Microbiota Transplantation for Inflammatory Bowel Disease. Gastroenterol Hepatol (N Y) 2016;12:374-9.

216. Paramsothy S, Paramsothy R, Rubin DT, et al. Faecal Microbiota Transplantation for Inflammatory Bowel Disease: A Systematic Review and Meta-analysis. J Crohns Colitis 2017;11:1180-99.

217. Cole DS, Dayal AK, Chazotte C. Elective primary cesarean delivery. The New England journal of medicine 2003;348:2364-5; author reply -5.

218. Li Y, Tian Y, Zhu W, et al. Cesarean delivery and risk of inflammatory bowel disease: a systematic review and meta-analysis. Scandinavian journal of gastroenterology 2014;49:834-44.

219. Bager P, Simonsen J, Nielsen NM, Frisch M. Cesarean section and offspring's risk of inflammatory bowel disease: a national cohort study. Inflammatory bowel diseases 2012;18:857-62.

220. Gronlund MM, Lehtonen OP, Eerola E, Kero P. Fecal microflora in healthy infants born by different methods of delivery: permanent changes in intestinal flora after cesarean delivery. Journal of pediatric gastroenterology and nutrition 1999;28:19-25.

221. Salminen S, Gibson GR, McCartney AL, Isolauri E. Influence of mode of delivery on gut microbiota composition in seven year old children. Gut 2004;53:1388-9.

222. Jin S, Zhao D, Cai C, et al. Low-dose penicillin exposure in early life decreases Th17 and the susceptibility to DSS colitis in mice through gut microbiota modification. Sci Rep 2017;7:43662.

223. Ko Y, Kariyawasam V, Karnib M, et al. Inflammatory Bowel Disease Environmental Risk Factors: A Population-Based Case-Control Study of Middle Eastern Migration to Australia. Clin Gastroenterol Hepatol 2015;13:1453-63 e1.

224. Ng SC, Tang W, Leong RW, et al. Environmental risk factors in inflammatory bowel disease: a population-based case-control study in Asia-Pacific. Gut 2015;64:1063-71.

225. Theochari NA, Stefanopoulos A, Mylonas KS, Economopoulos KP. Antibiotics exposure and risk of inflammatory bowel disease: a systematic review. Scand J Gastroenterol 2018;53:1-7.

226. Bloomfield SF, Stanwell-Smith R, Crevel RW, Pickup J. Too clean, or not too clean: the hygiene hypothesis and home hygiene. Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology 2006;36:402-25.

227. Tsironi E, Feakins RM, Probert CS, Rampton DS, Phil D. Incidence of inflammatory bowel disease is rising and abdominal tuberculosis is falling in Bangladeshis in East London, United Kingdom. The American journal of gastroenterology 2004;99:1749-55.

228. Koloski NA, Bret L, Radford-Smith G. Hygiene hypothesis in inflammatory bowel disease: a critical review of the literature. World journal of gastroenterology : WJG 2008;14:165-73.

229. Feeney MA, Murphy F, Clegg AJ, Trebble TM, Sharer NM, Snook JA. A case-control study of childhood environmental risk factors for the development of inflammatory bowel disease. European journal of gastroenterology & hepatology 2002;14:529-34.

230. Luther J, Dave M, Higgins PD, Kao JY. Association between Helicobacter pylori infection and inflammatory bowel disease: a meta-analysis and systematic review of the literature. Inflammatory bowel diseases 2010;16:1077-84.

231. Bernstein CN, Rawsthorne P, Cheang M, Blanchard JF. A population-based case control study of potential risk factors for IBD. The American journal of gastroenterology 2006;101:993-1002.

232. Molodecky NA, Kaplan GG. Environmental risk factors for inflammatory bowel disease. Gastroenterology & hepatology 2010;6:339-46.

233. Lee D, Albenberg L, Compher C, et al. Diet in the pathogenesis and treatment of inflammatory bowel diseases. Gastroenterology 2015;148:1087-106.

234. Martinez-Medina M, Denizot J, Dreux N, et al. Western diet induces dysbiosis with increased E coli in CEABAC10 mice, alters host barrier function favouring AIEC colonisation. Gut 2014;63:116-24.

235. Wagner SJ, Schmidt A, Effenberger MJ, Gruber L, Danier J, Haller D. Semisynthetic diet ameliorates Crohn's disease-like ileitis in TNFDeltaARE/WT mice through antigen-independent mechanisms of gluten. Inflammatory bowel diseases 2013;19:1285-94.

236. Roberts CL, Keita AV, Duncan SH, et al. Translocation of Crohn's disease Escherichia coli across M-cells: contrasting effects of soluble plant fibres and emulsifiers. Gut 2010;59:1331-9.

237. Nickerson KP, McDonald C. Crohn's disease-associated adherent-invasive Escherichia coli adhesion is enhanced by exposure to the ubiquitous dietary polysaccharide maltodextrin. PloS one 2012;7:e52132.

238. Tilg H, Moschen AR. Food, immunity, and the microbiome. Gastroenterology 2015;148:1107-19.

239. Teahon K, Smethurst P, Pearson M, Levi AJ, Bjarnason I. The effect of elemental diet on intestinal permeability and inflammation in Crohn's disease. Gastroenterology 1991;101:84-9.

240. Triantafillidis JK, Vagianos C, Papalois AE. The role of enteral nutrition in patients with inflammatory bowel disease: current aspects. BioMed research international 2015;2015:197167.

241. Khalili H, Huang ES, Ananthakrishnan AN, et al. Geographical variation and incidence of inflammatory bowel disease among US women. Gut 2012;61:1686-92.

242. Nerich V, Jantchou P, Boutron-Ruault MC, et al. Low exposure to sunlight is a risk factor for Crohn's disease. Alimentary pharmacology & therapeutics 2011;33:940-5.

243. Shivananda S, Lennard-Jones J, Logan R, et al. Incidence of inflammatory bowel disease across Europe: is there a difference between north and south? Results of the European Collaborative Study on Inflammatory Bowel Disease (EC-IBD). Gut 1996;39:690-7.

244. Ye Y, Pang Z, Chen W, Ju S, Zhou C. The epidemiology and risk factors of inflammatory bowel disease. Int J Clin Exp Med 2015;8:22529-42.

245. Probert CS, Jayanthi V, Pinder D, Wicks AC, Mayberry JF. Epidemiological study of ulcerative proctocolitis in Indian migrants and the indigenous population of Leicestershire. Gut 1992;33:687-93.

246. Carr I, Mayberry JF. The effects of migration on ulcerative colitis: a three-year prospective study among Europeans and first- and second- generation South Asians in Leicester (1991-1994). Am J Gastroenterol 1999;94:2918-22.

247. Benchimol EI, Mack DR, Guttmann A, et al. Inflammatory bowel disease in immigrants to Canada and their children: a population-based cohort study. Am J Gastroenterol 2015;110:553-63.

248. Graham DB, Xavier RJ. From genetics of inflammatory bowel disease towards mechanistic insights. Trends in immunology 2013;34:371-8.

249. Ogura Y, Bonen DK, Inohara N, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature 2001;411:603-6.

250. Hampe J, Cuthbert A, Croucher PJ, et al. Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. Lancet 2001;357:1925-8.

251. van Heel DA, Ghosh S, Butler M, et al. Muramyl dipeptide and toll-like receptor sensitivity in NOD2-associated Crohn's disease. Lancet 2005;365:1794-6.

252. van Heel DA, Hunt KA, Ghosh S, Herve M, Playford RJ. Normal responses to specific NOD1-activating peptidoglycan agonists in the presence of the NOD2 frameshift and other mutations in Crohn's disease. European journal of immunology 2006;36:1629-35.

253. van Heel DA, Hunt KA, King K, et al. Detection of muramyl dipeptide-sensing pathway defects in patients with Crohn's disease. Inflammatory bowel diseases 2006;12:598-605.

254. Peeters H, Bogaert S, Laukens D, et al. CARD15 variants determine a disturbed early response of monocytes to adherent-invasive Escherichia coli strain LF82 in Crohn's disease. International journal of immunogenetics 2007;34:181-91.

255. Hedl M, Li J, Cho JH, Abraham C. Chronic stimulation of Nod2 mediates tolerance to bacterial products. Proceedings of the National Academy of Sciences of the United States of America 2007;104:19440-5.

256. Lappalainen M, Paavola-Sakki P, Halme L, et al. Novel CARD15/NOD2 mutations in Finnish patients with Crohn's disease and their relation to phenotypic variation in vitro and in vivo. Inflammatory bowel diseases 2008;14:176-85.

257. Seidelin JB, Broom OJ, Olsen J, Nielsen OH. Evidence for impaired CARD15 signalling in Crohn's disease without disease linked variants. PloS one 2009;4:e7794.

258. Kuuliala K, Lappalainen M, Turunen U, et al. Detection of muramyl dipeptide-sensing pathway defects in monocytes of patients with Crohn's disease using phospho-specific whole blood flow cytometry. Scandinavian journal of clinical and laboratory investigation 2013;73:494-502.

259. Rahman MK, Midtling EH, Svingen PA, et al. The pathogen recognition receptor NOD2 regulates human FOXP3+ T cell survival. Journal of immunology 2010;184:7247-56.

260. van Heel DA, Ghosh S, Hunt KA, et al. Synergy between TLR9 and NOD2 innate immune responses is lost in genetic Crohn's disease. Gut 2005;54:1553-7.

261. Braat H, Stokkers P, Hommes T, et al. Consequence of functional Nod2 and Tlr4 mutations on gene transcription in Crohn's disease patients. Journal of molecular medicine 2005;83:601-9.

262. Canto E, Moga E, Ricart E, et al. MDP-Induced selective tolerance to TLR4 ligands: impairment in NOD2 mutant Crohn's disease patients. Inflammatory bowel diseases 2009;15:1686-96.

263. Brosbol-Ravnborg A, Hvas CL, Agnholt J, et al. Toll-like receptor-induced granulocyte-macrophage colony-stimulating factor secretion is impaired in Crohn's disease by nucleotide oligomerization domain 2-dependent and -independent pathways. Clin Exp Immunol 2009;155:487-95.

264. Butler M, Chaudhary R, van Heel DA, Playford RJ, Ghosh S. NOD2 activity modulates the phenotype of LPS-stimulated dendritic cells to promote the development of T-helper type 2-like lymphocytes - Possible implications for NOD2-associated Crohn's disease. Journal of Crohn's & colitis 2007;1:106-15.

265. Vissers M, Remijn T, Oosting M, et al. Respiratory syncytial virus infection augments NOD2 signaling in an IFN-beta-dependent manner in human primary cells. European journal of immunology 2012;42:2727-35.

266. Kullberg BJ, Ferwerda G, de Jong DJ, et al. Crohn's disease patients homozygous for the 3020insC NOD2 mutation have a defective NOD2/TLR4 cross-tolerance to intestinal stimuli. Immunology 2008;123:600-5.

267. Netea MG, Ferwerda G, de Jong DJ, et al. The frameshift mutation in Nod2 results in unresponsiveness not only to Nod2- but also Nod1-activating peptidoglycan agonists. J Biol Chem 2005;280:35859-67.

268. Cooney R, Baker J, Brain O, et al. NOD2 stimulation induces autophagy in dendritic cells influencing bacterial handling and antigen presentation. Nat Med 2010;16:90-7.

269. Koss K, Satsangi J, Fanning GC, Welsh KI, Jewell DP. Cytokine (TNF alpha, LT alpha and IL-10) polymorphisms in inflammatory bowel diseases and normal controls: differential effects on production and allele frequencies. Genes and immunity 2000;1:185-90.

270. Gasche C, Grundtner P, Zwirn P, et al. Novel variants of the IL-10 receptor 1 affect inhibition of monocyte TNF-alpha production. J Immunol 2003;170:5578-82.

271. van der Linde K, Boor PP, Sandkuijl LA, et al. A Gly15Arg mutation in the interleukin-10 gene reduces secretion of interleukin-10 in Crohn disease. Scandinavian journal of gastroenterology 2003;38:611-7.

272. Begue B, Verdier J, Rieux-Laucat F, et al. Defective IL10 signaling defining a subgroup of patients with inflammatory bowel disease. The American journal of gastroenterology 2011;106:1544-55.

273. Mao H, Yang W, Lee PP, et al. Exome sequencing identifies novel compound heterozygous mutations of IL-10 receptor 1 in neonatal-onset Crohn's disease. Genes and immunity 2012;13:437-42.

274. Marcuzzi A, Girardelli M, Bianco AM, et al. Inflammation profile of four early onset Crohn patients. Gene 2012;493:282-5.

275. Galatola M, Miele E, Strisciuglio C, et al. Synergistic effect of interleukin-10-receptor variants in a case of early-onset ulcerative colitis. World journal of gastroenterology : WJG 2013;19:8659-70.

276. Wang AH, Lam WJ, Han DY, et al. The effect of IL-10 genetic variation and interleukin 10 serum levels on Crohn's disease susceptibility in a New Zealand population. Human immunology 2011;72:431-5.

277. Wagner J, Skinner NA, Catto-Smith AG, et al. TLR4, IL10RA, and NOD2 mutation in paediatric Crohn's disease patients: an association with Mycobacterium avium subspecies paratuberculosis and TLR4 and IL10RA expression. Med Microbiol Immunol 2013;202:267-76.

278. He C, Klionsky DJ. Regulation mechanisms and signaling pathways of autophagy. Annu Rev Genet 2009;43:67-93.

279. Wolfkamp SC, Verseyden C, Vogels EW, et al. ATG16L1 and NOD2 polymorphisms enhance phagocytosis in monocytes of Crohn's disease patients. World J Gastroenterol 2014;20:2664-72.

280. Ma T, Wu S, Yan W, Xie R, Zhou C. A functional variant of ATG16L2 is associated with Crohn's disease in the Chinese population. Colorectal Dis 2016;18:O420-O6.

281. Murthy A, Li Y, Peng I, et al. A Crohn's disease variant in Atg16l1 enhances its degradation by caspase 3. Nature 2014;506:456-62.

282. Parisinos CA, Serghiou S, Katsoulis M, et al. Variation in Interleukin 6 Receptor Gene Associates With Risk of Crohn's Disease and Ulcerative Colitis. Gastroenterology 2018;155:303-6 e2.

283. Xiao CX, Xiao JJ, Xu HZ, et al. Exome sequencing identifies novel compound heterozygous IFNA4 and IFNA10 mutations as a cause of impaired function in Crohn's disease patients. Sci Rep 2015;5:10514.

284. Hedl M, Abraham C. A TNFSF15 disease-risk polymorphism increases pattern-recognition receptor-induced signaling through caspase-8-induced IL-1. Proc Natl Acad Sci U S A 2014;111:13451-6.

285. Cavanaugh J, Consortium IBDIG. International collaboration provides convincing linkage replication in complex disease through analysis of a large pooled data set: Crohn disease and chromosome 16. Am J Hum Genet 2001;68:1165-71.

286. Capes-Davis A, Theodosopoulos G, Atkin I, et al. Check your cultures! A list of cross-contaminated or misidentified cell lines. Int J Cancer 2010;127:1-8.

287. Momozawa Y, Dmitrieva J, Theatre E, et al. IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. Nat Commun 2018;9:2427.

288. Liu JZ, van Sommeren S, Huang H, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet 2015;47:979-86.

289. Glubb DM, Gearry RB, Barclay ML, et al. NOD2 and ATG16L1 polymorphisms affect monocyte responses in Crohn's disease. World J Gastroenterol 2011;17:2829-37.

290. Christodoulou K, Wiskin AE, Gibson J, et al. Next generation exome sequencing of paediatric inflammatory bowel disease patients identifies rare and novel variants in candidate genes. Gut 2013;62:977-84.

291. Coelho T, Andreoletti G, Ashton JJ, et al. Genes implicated in thiopurine-induced toxicity: Comparing TPMT enzyme activity with clinical phenotype and exome data in a paediatric IBD cohort. Sci Rep 2016;6:34658.

292. Andreoletti G, Ashton JJ, Coelho T, et al. Exome analysis of patients with concurrent pediatric inflammatory bowel disease and autoimmune disease. Inflamm Bowel Dis 2015;21:1229-36.

293. Mossotto E, Ashton JJ, Coelho T, Beattie RM, MacArthur BD, Ennis S. Classification of Paediatric Inflammatory Bowel Disease using Machine Learning. Sci Rep 2017;7:2427.

294. Goode MR, Cheong SY, Li N, Ray WC, Bartlett CW. Collection and extraction of saliva DNA for next generation sequencing. Journal of visualized experiments : JoVE 2014.

295. Garcia-Closas M, Egan KM, Abruzzo J, et al. Collection of genomic DNA from adults in epidemiological studies by buccal cytobrush and mouthwash. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2001;10:687-96.

296. Rylander-Rudqvist T, Hakansson N, Tybring G, Wolk A. Quality and quantity of saliva DNA obtained from the self-administrated oragene method--a pilot study on the cohort of Swedish men. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2006;15:1742-5.

297. Turk DJ, Kozarek RA, Botoman VA, Patterson DJ, Ball TJ. Disposable endoscopic biopsy forceps: comparison with standard forceps of sample size and adequacy of specimen. Journal of clinical gastroenterology 1991;13:76-8.

298. Bigaud M, Maurer C, Vedrine C, Puissant B, Blancher A. A simple method to optimize peripheral blood mononuclear cell preparation from cynomolgus monkeys and improve mixed lymphocyte reactions. J Pharmacol Toxicol Methods 2004;50:153-9.

299. Autissier P, Soulas C, Burdo TH, Williams KC. Evaluation of a 12-color flow cytometry panel to study lymphocyte, monocyte, and dendritic cell subsets in humans. Cytometry Part A : the journal of the International Society for Analytical Cytology 2010;77:410-9.

300. Corkum CP, Ings DP, Burgess C, Karwowska S, Kroll W, Michalak TI. Immune cell subsets and their gene expression profiles from human PBMC isolated by Vacutainer Cell Preparation Tube (CPT) and standard density gradient. BMC immunology 2015;16:48.

301. Ramachandran H, Laux J, Moldovan I, Caspell R, Lehmann PV, Subbramanian RA. Optimal thawing of cryopreserved peripheral blood mononuclear cells for use in high-throughput human immune monitoring studies. Cells 2012;1:313-24.

302. Cook EB, Stahl JL, Lowe L, et al. Simultaneous measurement of six cytokines in a single sample of human tears using microparticle-based flow cytometry: allergics vs. non-allergics. Journal of immunological methods 2001;254:109-18.

303. Tarnok A, Hambsch J, Chen R, Varro R. Cytometric bead array to measure six cytokines in twenty-five microliters of serum. Clinical chemistry 2003;49:1000-2.

304. Cox KL, Devanarayan V, Kriauciunas A, Manetta J, Montrose C, Sittampalam S. Immunoassay Methods. In: Sittampalam GS, Coussens NP, Brimacombe K, et al., eds. Assay Guidance Manual. Bethesda (MD)2004.

305. Elshal MF, McCoy JP. Multiplex bead array assays: performance evaluation and comparison of sensitivity to ELISA. Methods 2006;38:317-23.

306. Zhou X, Fragala MS, McElhaney JE, Kuchel GA. Conceptual and methodological issues relevant to cytokine and inflammatory marker measurements in clinical research. Current opinion in clinical nutrition and metabolic care 2010;13:541-7.

307. Invitrogen. User Manual: Human Cytokine 10-Plex Panel. Catalog no LHC0001 2010.

308. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 1988;16:1215.

309. Carpi FM, Di Pietro F, Vincenzetti S, Mignini F, Napolioni V. Human DNA extraction methods: patents and applications. Recent Pat DNA Gene Seq 2011;5:1-7.

310. Seaby EG, Pengelly RJ, Ennis S. Exome sequencing explained: a practical guide to its clinical application. Briefings in functional genomics 2016;15:374-84.

311. Samorodnitsky E, Jewell BM, Hagopian R, et al. Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing. Human mutation 2015;36:903-14.

312. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 2008;456:53-9.

313. Ju J, Kim DH, Bi L, et al. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. Proceedings of the National Academy of Sciences of the United States of America 2006;103:19635-40.

314. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43:491-8.

315. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297-303.

316. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078-9.

317. Ruffalo M. Accurate estimation of short read mapping quality for next-generation genome sequencing. Bioinformatics 2012.

318. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754-60.

319. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. Bioinformatics 2011;27:2156-8.

320. Liu X, Han S, Wang Z, Gelernter J, Yang BZ. Variant callers for next-generation sequencing data: a comparison study. PLoS One 2013;8:e75619.

321. Chilamakuri CS, Lorenz S, Madoui MA, et al. Performance comparison of four exome capture systems for deep sequencing. BMC genomics 2014;15:449.

322. Dolled-Filhart MP, Lee M, Jr., Ou-Yang CW, Haraksingh RR, Lin JC. Computational and bioinformatics frameworks for next-generation whole exome and genome sequencing. TheScientificWorldJournal 2013;2013:730210.

323. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research 2010;38:e164.

324. Genomes Project C, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56-65.

325. Bhagwat M. Searching NCBI's dbSNP database. Current protocols in bioinformatics 2010;Chapter 1:Unit 1 19.

326. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nature protocols 2015;10:1556-66.

327. Dong C, Wei P, Jian X, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Human molecular genetics 2015;24:2125-37.

328. Mahmood K, Jung CH, Philip G, et al. Variant effect prediction tools assessed using independent, functional assay-based datasets: implications for discovery and diagnostics. Hum Genomics 2017;11:10.

329. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4:1073-81.

330. Li MX, Kwan JS, Bao SY, et al. Predicting mendelian disease-causing non-synonymous single nucleotide variants in exome sequencing studies. PLoS genetics 2013;9:e1003143.

331. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 2010;7:575-6.

332. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014;46:310-5.

333. Jagadeesh KA, Wenger AM, Berger MJ, et al. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. Nat Genet 2016;48:1581-6.

334. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248-9.

335. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. Genome Res 2009;19:1553-61.

336. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res 2011;39:e118.

337. Hassan MS, Shaalan AA, Dessouky MI, Abdelnaiem AE, ElHefnawi M. Evaluation of computational techniques for predicting non-synonymous single nucleotide variants pathogenicity. Genomics 2018.

338. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. PLoS One 2012;7:e46688.

339. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput Biol 2010;6:e1001025.

340. Sadri J, Diallo AB, Blanchette M. Predicting site-specific human selective pressure using evolutionary signatures. Bioinformatics 2011;27:i266-74.

341. Garber M, Guttman M, Clamp M, Zody MC, Friedman N, Xie X. Identifying novel constrained elements by exploiting biased substitution patterns. Bioinformatics 2009;25:i54-62.

342. Quang D, Chen Y, Xie X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. Bioinformatics 2015;31:761-3.

343. Gonzaga-Jauregui C, Lupski JR, Gibbs RA. Human genome sequencing in health and disease. Annual review of medicine 2012;63:35-61.

344. Wheeler DL, Barrett T, Benson DA, et al. Database resources of the National Center for Biotechnology Information. Nucleic acids research 2007;35:D5-12.

345. Mullaney JM, Mills RE, Pittard WS, Devine SE. Small insertions and deletions (INDELs) in human genomes. Human molecular genetics 2010;19:R131-6.

346. Li MX, Gui HS, Kwan JS, Bao SY, Sham PC. A comprehensive framework for prioritizing variants in exome sequencing studies of Mendelian diseases. Nucleic acids research 2012;40:e53.

347. Park JH, Gail MH, Weinberg CR, et al. Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. Proceedings of the National Academy of Sciences of the United States of America 2011;108:18026-31.

348. Ritchie GR, Flicek P. Computational approaches to interpreting genomic sequence variation. Genome medicine 2014;6:87.

349. Coelho T, Andreoletti G, Ashton JJ, et al. Immuno-genomic profiling of patients with inflammatory bowel disease: a systematic review of genetic and functional in vivo studies of implicated genes. Inflamm Bowel Dis 2014;20:1813-9.

350. Steinke JW, Borish L. 3. Cytokines and chemokines. The Journal of allergy and clinical immunology 2006;117:S441-5.

351. Mowat C, Cole A, Windsor A, et al. Guidelines for the management of inflammatory bowel disease in adults. Gut 2011;60:571-607.

352. Masuko K. Rheumatoid cachexia revisited: a metabolic co-morbidity in rheumatoid arthritis. Front Nutr 2014;1:20.

353. Neurath MF. Cytokines in inflammatory bowel disease. Nat Rev Immunol 2014;14:329-42.

354. Becker C, Fantini MC, Schramm C, et al. TGF-beta suppresses tumor progression in colon cancer by inhibition of IL-6 trans-signaling. Immunity 2004;21:491-501.

355. Ogawa C, Liu YJ, Kobayashi KS. Muramyl dipeptide and its derivatives: peptide adjuvant in immunological disorders and cancer therapy. Current bioactive compounds 2011;7:180-97.

356. Aliprantis AO, Yang RB, Mark MR, et al. Cell activation and apoptosis by bacterial lipoproteins through toll-like receptor-2. Science 1999;285:736-9.

357. Rietschel ET, Kirikae T, Schade FU, et al. Bacterial endotoxin: molecular relationships of structure to activity and function. FASEB journal : official publication of the Federation of American Societies for Experimental Biology 1994;8:217-25.

358. Oeckinghaus A, Ghosh S. The NF-kappaB family of transcription factors and its regulation. Cold Spring Harbor perspectives in biology 2009;1:a000034.

359. Oeckinghaus A, Hayden MS, Ghosh S. Crosstalk in NF-kappaB signaling pathways. Nature immunology 2011;12:695-708.

360. Kawai T, Akira S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. Nature immunology 2010;11:373-84.

361. Kawasaki T, Kawai T. Toll-like receptor signaling pathways. Front Immunol 2014;5:461.

362. Ozinsky A, Underhill DM, Fontenot JD, et al. The repertoire for pattern recognition of pathogens by the innate immune system is defined by cooperation between toll-like receptors. Proceedings of the National Academy of Sciences of the United States of America 2000;97:13766-71.

363. Zanoni I, Ostuni R, Marek LR, et al. CD14 controls the LPS-induced endocytosis of Toll-like receptor 4. Cell 2011;147:868-80.

364. Kagan JC, Medzhitov R. Phosphoinositide-mediated adaptor recruitment controls Toll-like receptor signaling. Cell 2006;125:943-55.

365. Inohara N, Ogura Y, Fontalba A, et al. Host recognition of bacterial muramyl dipeptide mediated through NOD2. Implications for Crohn's disease. J Biol Chem 2003;278:5509-12.

366. Irving AT, Mimuro H, Kufer TA, et al. The immune receptor NOD1 and kinase RIP2 interact with bacterial peptidoglycan on early endosomes to promote autophagy and inflammatory signaling. Cell Host Microbe 2014;15:623-35.

367. Nakamura N, Lill JR, Phung Q, et al. Endosomes are specialized platforms for bacterial sensing and NOD2 signalling. Nature 2014;509:240-4.

368. Kitaura Y, Nakaguchi O, Takeno H, et al. N2-(gamma-D-Glutamyl)-meso-2(L),2'(D)-diaminopimelic acid as the minimal prerequisite structure of FK-156: its acyl derivatives with potent immunostimulating activity. Journal of medicinal chemistry 1982;25:335-7.

369. Chamaillard M, Hashimoto M, Horie Y, et al. An essential role for NOD1 in host recognition of bacterial peptidoglycan containing diaminopimelic acid. Nature immunology 2003;4:702-7.

370. Vollmer W, Blanot D, de Pedro MA. Peptidoglycan structure and architecture. FEMS microbiology reviews 2008;32:149-67.

371. Baeuerle PA, Henkel T. Function and activation of NF-kappa B in the immune system. Annual review of immunology 1994;12:141-79.

372. Hayden MS, Ghosh S. Signaling to NF-kappaB. Genes Dev 2004;18:2195-224.

373. Pahl HL. Activators and target genes of Rel/NF-kappaB transcription factors. Oncogene 1999;18:6853-66.

374. Chen ZJ, Parent L, Maniatis T. Site-specific phosphorylation of IkappaBalpha by a novel ubiquitination-dependent protein kinase activity. Cell 1996;84:853-62.

375. Pedersen J, LaCasse EC, Seidelin JB, Coskun M, Nielsen OH. Inhibitors of apoptosis (IAPs) regulate intestinal immunity and inflammatory bowel disease (IBD) inflammation. Trends in molecular medicine 2014;20:652-65.

376. Gilmore TD. Introduction to NF-kappaB: players, pathways, perspectives. Oncogene 2006;25:6680-4.

377. Liu T, Zhang L, Joo D, Sun SC. NF-kappaB signaling in inflammation. Signal Transduct Target Ther 2017;2.

378. Latz E, Xiao TS, Stutz A. Activation and regulation of the inflammasomes. Nat Rev Immunol 2013;13:397-411.

379. Guo H, Callaway JB, Ting JP. Inflammasomes: mechanism of action, role in disease, and therapeutics. Nature medicine 2015;21:677-87.

380. He Y, Hara H, Nunez G. Mechanism and Regulation of NLRP3 Inflammasome Activation. Trends Biochem Sci 2016;41:1012-21.

381. Gaidt MM, Ebert TS, Chauhan D, et al. Human Monocytes Engage an Alternative Inflammasome Pathway. Immunity 2016;44:833-46.

382. Zhang W, Liu HT. MAPK signal pathways in the regulation of cell proliferation in mammalian cells. Cell Res 2002;12:9-18.

383. Zeissig Y, Petersen BS, Milutinovic S, et al. XIAP variants in male Crohn's disease. Gut 2015;64:66-76.

384. Deering RP, Orange JS. Development of a clinical assay to evaluate toll-like receptor function. Clinical and vaccine immunology : CVI 2006;13:68-76.

385. Hirschfeld AF, Bettinger JA, Victor RE, et al. Prevalence of Toll-like receptor signalling defects in apparently healthy children who developed invasive pneumococcal infection. Clin Immunol 2007;122:271-8.

386. Benyoucef S, Hober D, Shen L, et al. Production of TNFalpha and IL-6 by activated whole blood from HIV-1 infected patients detected by a one-stage procedure: relationship with the phenotype of HIV-1 isolates. Microbiol Immunol 1997;41:939-46.

387. Ferwerda G, Kramer M, de Jong D, et al. Engagement of NOD2 has a dual effect on proIL-1beta mRNA transcription and secretion of bioactive IL-1beta. Eur J Immunol 2008;38:184-91.

388. Damsgaard CT, Lauritzen L, Calder PC, Kjaer TM, Frokiaer H. Whole-blood culture is a valid low-cost method to measure monocytic cytokines - a comparison of cytokine production in cultures of human whole-blood, mononuclear cells and monocytes. J Immunol Methods 2009;340:95-101.

389. Ammann S, Elling R, Gyrd-Hansen M, et al. A new functional assay for the diagnosis of X-linked inhibitor of apoptosis (XIAP) deficiency. Clinical and experimental immunology 2014;176:394-400.

390. Netea MG, Nold-Petry CA, Nold MF, et al. Differential requirement for the activation of the inflammasome for processing and release of IL-1beta in monocytes and macrophages. Blood 2009;113:2324-35.

391. Smedman C, Gardlund B, Nihlmark K, Gille-Johnson P, Andersson J, Paulie S. ELISpot analysis of LPS-stimulated leukocytes: human granulocytes selectively secrete IL-8, MIP-1beta and TNF-alpha. Journal of immunological methods 2009;346:1-8.

392. Daigneault M, De Silva TI, Bewley MA, et al. Monocytes regulate the mechanism of T-cell death by inducing Fas-mediated apoptosis during bacterial infection. PLoS pathogens 2012;8:e1002814.

393. Hudig D, Hunter KW, Diamond WJ, Redelman D. Properties of human blood monocytes. I. CD91 expression and log orthogonal light scatter provide a robust method to identify monocytes that is more accurate than CD14 expression. Cytometry B Clin Cytom 2014;86:111-20.

394. Hudig D, Hunter KW, Diamond WJ, Redelman D. Properties of human blood monocytes. II. Monocytes from healthy adults are highly heterogeneous within and among individuals. Cytometry B Clin Cytom 2014;86:121-34.

395. Wagner M, Poeck H, Jahrsdoerfer B, et al. IL-12p70-dependent Th1 induction by human B cells requires combined activation with CD40 ligand and CpG DNA. Journal of immunology 2004;172:954-63.

396. Ulmer AJ, Flad H, Rietschel T, Mattern T. Induction of proliferation and cytokine production in human T lymphocytes by lipopolysaccharide (LPS). Toxicology 2000;152:37-45.

397. Ogura Y, Inohara N, Benito A, Chen FF, Yamaoka S, Nunez G. Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF-kappaB. The Journal of biological chemistry 2001;276:4812-8.

398. Hudig D, Hunter KW, Diamond WJ, Redelman D. Properties of human blood monocytes. I. CD91 expression and log orthogonal light scatter provide a robust method to identify monocytes that is more accurate than CD14 expression. Cytom Part B-Clin Cy 2014;86:111-20.

399. Hudig D, Hunter KW, Diamond WJ, Redelman D. Properties of human blood monocytes. II. Monocytes from healthy adults are highly heterogeneous within and among individuals. Cytom Part B-Clin Cy 2014;86:121-34.

400. Warimwe GM, Murungi LM, Kamuyu G, et al. The ratio of monocytes to lymphocytes in peripheral blood correlates with increased susceptibility to clinical malaria in Kenyan children. PloS one 2013;8:e57320.

401. Mallone R, Mannering SI, Brooks-Worrell BM, et al. Isolation and preservation of peripheral blood mononuclear cells for analysis of islet antigen-reactive T cell responses: position statement of the T-Cell Workshop Committee of the Immunology of Diabetes Society. Clin Exp Immunol 2011;163:33-49.

402. Axelsson S, Faresjo M, Hedman M, Ludvigsson J, Casas R. Cryopreserved peripheral blood mononuclear cells are suitable for the assessment of immunological markers in type 1 diabetic children. Cryobiology 2008;57:201-8.

403. Hoffmeister B, Bunde T, Rudawsky IM, Volk HD, Kern F. Detection of antigen-specific T cells by cytokine flow cytometry: the use of whole blood may underestimate frequencies. European journal of immunology 2003;33:3484-92.

404. Kreher CR, Dittrich MT, Guerkov R, Boehm BO, Tary-Lehmann M. CD4+ and CD8+ cells in cryopreserved human PBMC maintain full functionality in cytokine ELISPOT assays. Journal of immunological methods 2003;278:79-93.

405. Mazur P. Freezing of living cells: mechanisms and implications. The American journal of physiology 1984;247:C125-42.

406. Damgaard RB, Fiil BK, Speckmann C, et al. Disease-causing mutations in the XIAP BIR2 domain impair NOD2-dependent immune signalling. EMBO molecular medicine 2013;5:1278-95.

407. Brooks-Worrell B, Tree T, Mannering SI, et al. Comparison of cryopreservation methods on T-cell responses to islet and control antigens from type 1 diabetic patients and controls. Diabetes/metabolism research and reviews 2011;27:737-45.

408. Hensley-McBain T, Heit A, De Rosa SC, McElrath MJ, Andersen-Nissen E. Optimization of a whole blood phenotyping assay for enumeration of peripheral blood leukocyte populations in multicenter clinical trials. Journal of immunological methods 2014;411:23-36.

409. Stevens VL, Patel AV, Feigelson HS, Rodriguez C, Thun MJ, Calle EE. Cryopreservation of whole blood samples collected in the field for a large epidemiologic study. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2007;16:2160-3.

410. Honge BL, Petersen MS, Olesen R, Moller BK, Erikstrup C. Optimizing recovery of frozen human peripheral blood mononuclear cells for flow cytometry. PloS one 2017;12:e0187440.

411. Bull M, Lee D, Stucky J, et al. Defining blood processing parameters for optimal detection of cryopreserved antigen-specific responses for HIV vaccine trials. Journal of immunological methods 2007;322:57-69.

412. Jeurink PV, Vissers YM, Rappard B, Savelkoul HF. T cell responses in fresh and cryopreserved peripheral blood mononuclear cells: kinetics of cell viability, cellular subsets, proliferation, and cytokine production. Cryobiology 2008;57:91-103.

413. Seale AC, de Jong BC, Zaidi I, et al. Effects of cryopreservation on CD4+ CD25+ T cells of HIV-1 infected individuals. J Clin Lab Anal 2008;22:153-8.

414. Ferreira Vda S, Mesquita FV, Mesquita DJ, Andrade LE. The effects of freeze/thawing on the function and phenotype of CD4(+) lymphocyte subsets in normal individuals and patients with systemic lupus erythematosus. Cryobiology 2015;71:507-10.

415. Shete A, Jayawant P, Thakar M, Kurle S, Singh DP, Paranjape RS. Differential modulation of phenotypic composition of HIV-infected and -uninfected PBMCs during cryopreservation. J Immunoassay Immunochem 2013;34:333-45.

416. Mallone R, Martinuzzi E, Blancou P, et al. CD8+ T-cell responses identify beta-cell autoimmunity in human type 1 diabetes. Diabetes 2007;56:613-21.

417. Jason J, Larned J. Single-cell cytokine profiles in normal humans: comparison of flow cytometric reagents and stimulation protocols. Journal of immunological methods 1997;207:13-22.

418. Katial RK, Sachanandani D, Pinney C, Lieberman MM. Cytokine production in cell culture by peripheral blood mononuclear cells from immunocompetent hosts. Clinical and diagnostic laboratory immunology 1998;5:78-81.

419. Ferry B, Antrobus P, Huzicka I, Farrell A, Lane A, Chapel H. Intracellular cytokine expression in whole blood preparations from normals and patients with atopic dermatitis. Clinical and experimental immunology 1997;110:410-7.

420. Deenadayalan A, Maddineni P, Raja A. Comparison of whole blood and PBMC assays for T-cell functional analysis. BMC research notes 2013;6:120.

421. Suni MA, Picker LJ, Maino VC. Detection of antigen-specific T cell cytokine expression in whole blood by flow cytometry. Journal of immunological methods 1998;212:89-98.

422. Yaqoob P, Newsholme EA, Calder PC. Comparison of cytokine production in cultures of whole human blood and purified mononuclear cells. Cytokine 1999;11:600-5.

423. Haddaway NR, Verhoeven JT. Poor methodological detail precludes experimental repeatability and hampers synthesis in ecology. Ecol Evol 2015;5:4451-4.

424. Wong SH, Gao Q, Tsoi KK, et al. Effect of immunosuppressive therapy on interferon gamma release assay for latent tuberculosis screening in patients with autoimmune diseases: a systematic review and meta-analysis. Thorax 2016;71:64-72.

425. Edwards A, Gao Y, Allan RN, et al. Corticosteroids and infliximab impair the performance of interferon-gamma release assays used for diagnosis of latent tuberculosis. Thorax 2017;72:946-9.

426. Gomez CR, Boehmer ED, Kovacs EJ. The aging innate immune system. Curr Opin Immunol 2005;17:457-62.

427. Maeda S, Hsu LC, Liu H, et al. Nod2 mutation in Crohn's disease potentiates NF-kappaB activity and IL-1beta processing. Science 2005;307:734-8.

428. Lin Z, Poritz L, Franke A, et al. Genetic association of DLG5 R30Q with familial and sporadic inflammatory bowel disease in men. Disease markers 2009;27:193-201.

429. Brinkmann V, Kristofic C. Regulation by corticosteroids of Th1 and Th2 cytokine production in human CD4+ effector T cells generated from CD45RO- and CD45RO+ subsets. Journal of immunology 1995;155:3322-8.

430. Hu X, Li WP, Meng C, Ivashkiv LB. Inhibition of IFN-gamma signaling by glucocorticoids. Journal of immunology 2003;170:4833-9.

431. Greene CS, Tan J, Ung M, Moore JH, Cheng C. Big data bioinformatics. J Cell Physiol 2014;229:1896-900.

432. Bishop CM. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag, New York 2006.

433. Lever J, Krzywinski, M., Altman, N. Principal Component Analysis. Nature methods 2017;14:641-2.

434. Abdi H, Williams, L.J. Principal component analysis. Wiley Interdisc Rev Comput Stat 2010;2:433-59.

435. Mossotto EA, JJ. Pengelly, R.J. GenePy - a score for estimating gene pathogenicity in individuals using next-generation sequencing data. bioRxiv (Pre-print) 2018.

436. Group IW. Inflammatory Bowel Disease in Children and Adolescents: Recommendations for Diagnosis-The Porto Criteria

. J Paediatr Gastroenterol Nutr 2005;41:1-7.

437. Ibd Working Group of the European Society for Paediatric Gastroenterology H, Nutrition. Inflammatory bowel disease in children and adolescents: recommendations for diagnosis--the Porto criteria. J Pediatr Gastroenterol Nutr 2005;41:1-7.

438. Hyams J, Markowitz J, Otley A, et al. Evaluation of the pediatric crohn disease activity index: a prospective multicenter experience. J Pediatr Gastroenterol Nutr 2005;41:416-21.

439. Turner D, Hyams J, Markowitz J, et al. Appraisal of the pediatric ulcerative colitis activity index (PUCAI). Inflamm Bowel Dis 2009;15:1218-23.

440. Stiehm ER, Winter HS, Bryson YJ. Cellular (T cell) immunity in the human newborn. Pediatrics 1979;64:814-21.

441. Pawelec G, Barnett Y, Forsey R, et al. T cells and aging, January 2002 update. Frontiers in bioscience : a journal and virtual library 2002;7:d1056-183.

442. Oliphant TE. SciPy: Open source scientific tools for Python. Comput Sci Eng 2007;9:10-20.

443. Mossotto E, Ashton JJ, O'Gorman L, et al. GenePy - a score for estimating gene pathogenicity in individuals using next-generation sequencing data. BMC Bioinformatics 2019;20:254.

444. Belinky F, Nativ N, Stelzer G, et al. PathCards: multi-source consolidation of human biological pathways. Database (Oxford) 2015;2015.

445. Angriman I, Scarpa M, D'Inca R, et al. Enzymes in feces: useful markers of chronic inflammatory bowel disease. Clin Chim Acta 2007;381:63-8.

446. Kostas A, Siakavellas SI, Kosmidis C, et al. Fecal calprotectin measurement is a marker of short-term clinical outcome and presence of mucosal healing in patients with inflammatory bowel disease. World journal of gastroenterology : WJG 2017;23:7387-96.

447. Ashton JJB, R.M. Faecal Calprotectin: What Does This Mean for the Paediatric Inflammatory Bowel Disease Phenotype? JPGN 2017.

448. Gaidt MM, Hornung V. Alternative inflammasome activation enables IL-1beta release from living cells. Curr Opin Immunol 2017;44:7-13.

449. Harhaj EW, Dixit VM. Regulation of NF-kappaB by deubiquitinases. Immunol Rev 2012;246:107-24.

450. Xu G, Tan X, Wang H, et al. Ubiquitin-specific peptidase 21 inhibits tumor necrosis factor alpha-induced nuclear factor kappaB activation via binding to and deubiquitinating receptor-interacting protein 1. J Biol Chem 2010;285:969-78.

451. Babyatsky MW, Rossiter G, Podolsky DK. Expression of transforming growth factors alpha and beta in colonic mucosa in inflammatory bowel disease. Gastroenterology 1996;110:975-84.

452. de Waal Malefyt R, Abrams J, Bennett B, Figdor CG, de Vries JE. Interleukin 10(IL-10) inhibits cytokine synthesis by human monocytes: an autoregulatory role of IL-10 produced by monocytes. J Exp Med 1991;174:1209-20.

453. Lu Y, Li X, Liu S, Zhang Y, Zhang D. Toll-like Receptors and Inflammatory Bowel Disease. Front Immunol 2018;9:72.

454. Tang A, Sharma A, Jen R, et al. Inflammasome-mediated IL-1beta production in humans with cystic fibrosis. PloS one 2012;7:e37689.

455. Liston A, Carr EJ, Linterman MA. Shaping Variation in the Human Immune System. Trends in immunology 2016.

456. Brodin P, Jojic V, Gao T, et al. Variation in the human immune system is largely driven by non-heritable influences. Cell 2015;160:37-47.

457. Orru V, Steri M, Sole G, et al. Genetic variants regulating immune cell levels in health and disease. Cell 2013;155:242-56.

458. Carr EJ, Dooley J, Garcia-Perez JE, et al. The cellular composition of the human immune system is shaped by age and cohabitation. Nature immunology 2016;17:461-8.

459. Goronzy JJ, Weyand CM. Understanding immunosenescence to improve responses to vaccines. Nature immunology 2013;14:428-36.

460. Aaron F. Hirschfeld a JABa, Rachel E. Victor a,, Donald J. Davidson b AJCc, J. Mark Ansermino d,, David W. Scheifele a JSOe, Stuart E. Turvey a. Prevalence of Toll-like receptor signalling defects in

apparently healthy children who developed invasive

pneumococcal infection

. Clinical Immunology (2007) 122, 271–278 2006;122:271–8.

461. Anthony Tang AS, Roger Jen, Aaron F. Hirschfeld, Mark A. Chilvers, Pascal M. Lavoie, and Stuart E. Turvey. Inflammasome-Mediated IL-1β Production in Humans with Cystic Fibrosis. PloS one 2012;PMC3359311.

462. Adrian Liston EJC, Micelle A. Linterman. Shaping Variation in the Human Immune System. Trends in immunology 2016;37:p637–46.

463. P. Brodin ea. Variation in the human immune system is largely driven by non-heritable influences. Cell 2015;160:pp. 37–47.

464. V. Orru ea. Genetic variants regulating immune cell levels in health and disease

. Cell 2013;155:pp. 242–56.

465. E.J. Carr ea. The cellular composition of the human immune system is shaped by age and cohabitation. Nat Immunol 2016:461–8.

466. C.C. Whiting ea. Large-scale and comprehensive immune profiling and functional analysis of normal human aging. PloS one 2015.

467. J.J. Goronzy CMW. Understanding immunosenescence to improve responses to vaccines. Nat Immunol 2013:428–36.

468. Kamada N, Hisamatsu T, Okamoto S, et al. Unique CD14 intestinal macrophages contribute to the pathogenesis of Crohn disease via IL-23/IFN-gamma axis. The Journal of clinical investigation 2008;118:2269-80.

469. Viennois E, Zhao Y, Merlin D. Biomarkers of Inflammatory Bowel Disease: From Classical Laboratory Tools to Personalized Medicine. Inflamm Bowel Dis 2015;21:2467-74.

470. Tanaka T, Narazaki M, Kishimoto T. Anti-interleukin-6 receptor antibody, tocilizumab, for the treatment of autoimmune diseases. FEBS letters 2011;585:3699-709.

471. Atreya R, Mudter J, Finotto S, et al. Blockade of interleukin 6 trans signaling suppresses T-cell resistance against apoptosis in chronic intestinal inflammation: evidence in crohn disease and experimental colitis in vivo. Nature medicine 2000;6:583-8.

472. Kai Y, Takahashi I, Ishikawa H, et al. Colitis in mice lacking the common cytokine receptor gamma chain is mediated by IL-6-producing CD4+ T cells. Gastroenterology 2005;128:922-34.

473. Lopez-Castejon G, Edelmann MJ. Deubiquitinases: Novel Therapeutic Targets in Immune Surveillance? Mediators Inflamm 2016;2016:3481371.

474. Takeshita S, Kikuno R, Tezuka K, Amann E. Osteoblast-specific factor 2: cloning of a putative bone adhesion protein with homology with the insect protein fasciclin I. Biochem J 1993;294 (Pt 1):271-8.

475. Horiuchi K, Amizuka N, Takeshita S, et al. Identification and characterization of a novel protein, periostin, with restricted expression to periosteum and periodontal ligament and increased expression by transforming growth factor beta. J Bone Miner Res 1999;14:1239-49.

476. Sidhu SS, Yuan S, Innes AL, et al. Roles of epithelial cell-derived periostin in TGF-beta activation, collagen production, and collagen gel elasticity in asthma. Proceedings of the National Academy of Sciences of the United States of America 2010;107:14170-5.

477. Takayama G, Arima K, Kanaji T, et al. Periostin: a novel component of subepithelial fibrosis of bronchial asthma downstream of IL-4 and IL-13 signals. The Journal of allergy and clinical immunology 2006;118:98-104.

478. Sonnenberg-Riethmacher E, Miehe M, Riethmacher D. Promotion of periostin expression contributes to the migration of Schwann cells. J Cell Sci 2015;128:3345-55.

479. Norris RA, Moreno-Rodriguez R, Hoffman S, Markwald RR. The many facets of the matricelluar protein periostin during cardiac development, remodeling, and pathophysiology. J Cell Commun Signal 2009;3:275-86.

480. Jackson-Boeters L, Wen W, Hamilton DW. Periostin localizes to cells in normal skin, but is associated with the extracellular matrix during wound repair. J Cell Commun Signal 2009;3:125-33.

481. Ashley SL, Wilke CA, Kim KK, Moore BB. Periostin regulates fibrocyte function to promote myofibroblast differentiation and lung fibrosis. Mucosal Immunol 2017;10:341-51.

482. Rogler G, Hausmann M. Factors Promoting Development of Fibrosis in Crohn's Disease. Front Med (Lausanne) 2017;4:96.

483. Oka T, Xu J, Kaiser RA, et al. Genetic manipulation of periostin expression reveals a role in cardiac hypertrophy and ventricular remodeling. Circ Res 2007;101:313-21.

484. Legnani PE, Kornbluth A. Therapeutic options in the management of strictures in Crohn's disease. Gastrointest Endosc Clin N Am 2002;12:589-603.

485. Burke JP, Mulsow JJ, O'Keane C, Docherty NG, Watson RW, O'Connell PR. Fibrogenesis in Crohn's disease. Am J Gastroenterol 2007;102:439-48.

486. Rieder F, Fiocchi C. Intestinal fibrosis in inflammatory bowel disease - Current knowledge and future perspectives. J Crohns Colitis 2008;2:279-90.

487. Honsawek S, Udomsinprasert W, Vejchapipat P, Chongsrisawat V, Phavichitr N, Poovorawan Y. Elevated serum periostin is associated with liver stiffness and clinical outcome in biliary atresia. Biomarkers 2015;20:157-61.

488. Ohta S, Okamoto M, Fujimoto K, et al. The usefulness of monomeric periostin as a biomarker for idiopathic pulmonary fibrosis. PLoS One 2017;12:e0174547.

489. Takahashi K, Meguro K, Kawashima H, et al. Serum periostin levels serve as a biomarker for both eosinophilic airway inflammation and fixed airflow limitation in well-controlled asthmatics. J Asthma 2018:1-8.

490. Yamaguchi Y, Ono J, Masuoka M, et al. Serum periostin levels are correlated with progressive skin sclerosis in patients with systemic sclerosis. Br J Dermatol 2013;168:717-25.

491. Turner D, Griffiths AM, Walters TD, et al. Appraisal of the pediatric Crohn's disease activity index on four prospectively collected datasets: recommended cutoff values and clinimetric properties. Am J Gastroenterol 2010;105:2085-92.

492. E.R. Stiehm HSW, Y.J. Bryson.

Cellular (T cell) immunity in the human newborn

. Pediatrics, 64 1979:814–21.

493. G. Pawelec YB, R. Forsey, D. Frasca, A. Globerson, J. McLeod, C. Caruso, C. Franceschi, T. Fulop, S. Gupta, E. Mariani, E. Mocchegiani, R. Solana.

T cells and aging,

. Front Biosci, 2002:1056–d183.

494. Orend G, Chiquet-Ehrismann R. Tenascin-C induced signaling in cancer. Cancer Lett 2006;244:143-63.

495. Karatas Z, Baysal T, Alp H, Toker A. Serum tenascin-C: a novel biomarker for diagnosis and predicting prognosis of rheumatic carditis? J Trop Pediatr 2013;59:476-82.

496. Midwood K, Sacre S, Piccinini AM, et al. Tenascin-C is an endogenous activator of Toll-like receptor 4 that is essential for maintaining inflammation in arthritic joint disease. Nat Med 2009;15:774-80.

497. Lv Y, Wang W, Jia WD, et al. High preoparative levels of serum periostin are associated with poor prognosis in patients with hepatocellular carcinoma after hepatectomy. Eur J Surg Oncol 2013;39:1129-35.

498. Jia G, Erickson RW, Choy DF, et al. Periostin is a systemic biomarker of eosinophilic airway inflammation in asthmatic patients. J Allergy Clin Immunol 2012;130:647-54 e10.

499. Louis E, Collard A, Oger AF, Degroote E, Aboul Nasr El Yafi FA, Belaiche J. Behaviour of Crohn's disease according to the Vienna classification: changing pattern over the course of the disease. Gut 2001;49:777-82.

500. Engel T, Ungar B, Yung DE, Ben-Horin S, Eliakim R, Kopylov U. Vedolizumab in IBD-Lessons From Real-world Experience; A Systematic Review and Pooled Analysis. J Crohns Colitis 2018;12:245-57.

501. Jharap B, Seinen ML, de Boer NK, et al. Thiopurine therapy in inflammatory bowel disease patients: analyses of two 8-year intercept cohorts. Inflamm Bowel Dis 2010;16:1541-9.

502. Bowen GE, Irons GV, Jr., Rhodes JB, Kirsner JB. Early experiences with azathioprine in ulcerative colitis; a note of caution. JAMA 1966;195:460-4.

503. Dubinsky MC. Azathioprine, 6-mercaptopurine in inflammatory bowel disease: pharmacology, efficacy, and safety. Clin Gastroenterol Hepatol 2004;2:731-43.

504. Elion GB, Hitchings GH. Antagonists of nucleic acid derivatives. V. Pteridines. J Biol Chem 1951;188:611-21.

505. Coenen MJ, de Jong DJ, van Marrewijk CJ, et al. Identification of Patients With Variants in TPMT and Dose Reduction Reduces Hematologic Events During Thiopurine Treatment of Inflammatory Bowel Disease. Gastroenterology 2015;149:907-17 e7.

506. de Boer NK, Wong DR, Jharap B, et al. Dose-dependent influence of 5-aminosalicylates on thiopurine metabolism. Am J Gastroenterol 2007;102:2747-53.

507. Present DH, Korelitz BI, Wisch N, Glass JL, Sachar DB, Pasternack BS. Treatment of Crohn's disease with 6-mercaptopurine. A long-term, randomized, double-blind study. The New England journal of medicine 1980;302:981-7.

508. Benkov K, Lu Y, Patel A, et al. Role of thiopurine metabolite testing and thiopurine methyltransferase determination in pediatric IBD. J Pediatr Gastroenterol Nutr 2013;56:333-40.

509. Liu YP, Wu HY, Yang X, et al. Association between thiopurine S-methyltransferase polymorphisms and thiopurine-induced adverse drug reactions in patients with inflammatory bowel disease: a meta-analysis. PLoS One 2015;10:e0121745.

510. Haglund S. Interindividual differences in thiopurine metabolism. Linkoping University Medical Dissertations 2011.

511. Tai HL, Krynetski EY, Yates CR, et al. Thiopurine S-methyltransferase deficiency: two nucleotide transitions define the most prevalent mutant allele associated with loss of catalytic activity in Caucasians. American journal of human genetics 1996;58:694-702.

512. Hamdan-Khalil R, Gala JL, Allorge D, et al. Identification and functional analysis of two rare allelic variants of the thiopurine S-methyltransferase gene, TPMT*16 and TPMT*19. Biochemical pharmacology 2005;69:525-9.

513. Weinshilboum RM, Sladek SL. Mercaptopurine pharmacogenetics: monogenic inheritance of erythrocyte thiopurine methyltransferase activity. American journal of human genetics 1980;32:651-62.

514. von Ahsen N, Armstrong VW, Oellerich M. Rapid, long-range molecular haplotyping of thiopurine S-methyltransferase (TPMT) *3A, *3B, and *3C. Clin Chem 2004;50:1528-34.

515. Relling MV, Gardner EE, Sandborn WJ, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for thiopurine methyltransferase genotype and thiopurine dosing. Clin Pharmacol Ther 2011;89:387-91.

516. Relling MV, Gardner EE, Sandborn WJ, et al. Clinical pharmacogenetics implementation consortium guidelines for thiopurine methyltransferase genotype and thiopurine dosing: 2013 update. Clin Pharmacol Ther 2013;93:324-5.

517. Al Hadithy AF, de Boer NK, Derijks LJ, Escher JC, Mulder CJ, Brouwers JR. Thiopurines in inflammatory bowel disease: pharmacogenetics, therapeutic drug monitoring and clinical recommendations. Dig Liver Dis 2005;37:282-97.

518. Sparrow MP, Hande SA, Friedman S, Cao D, Hanauer SB. Effect of allopurinol on clinical outcomes in inflammatory bowel disease nonresponders to azathioprine or 6-mercaptopurine. Clin Gastroenterol Hepatol 2007;5:209-14.

519. Palmieri O, Latiano A, Bossa F, et al. Sequential evaluation of thiopurine methyltransferase, inosine triphosphate pyrophosphatase, and HPRT1 genes polymorphisms to explain thiopurines' toxicity and efficacy. Alimentary pharmacology & therapeutics 2007;26:737-45.

520. Farfan MJ, Salas C, Canales C, et al. Prevalence of TPMT and ITPA gene polymorphisms and effect on mercaptopurine dosage in Chilean children with acute lymphoblastic leukemia. BMC cancer 2014;14:299.

521. Kudo M, Saito Y, Sasaki T, et al. Genetic variations in the HGPRT, ITPA, IMPDH1, IMPDH2, and GMPS genes in Japanese individuals. Drug metabolism and pharmacokinetics 2009;24:557-64.

522. Kurzawski M, Dziewanowski K, Safranow K, Drozdzik M. Polymorphism of genes involved in purine metabolism (XDH, AOX1, MOCOS) in kidney transplant recipients receiving azathioprine. Therapeutic drug monitoring 2012;34:266-74.

523. Stocco G, Yang W, Crews KR, et al. PACSIN2 polymorphism influences TPMT activity and mercaptopurine-related gastrointestinal toxicity. Human molecular genetics 2012;21:4793-804.

524. Ban H, Andoh A, Imaeda H, et al. The multidrug-resistance protein 4 polymorphism is a new factor accounting for thiopurine sensitivity in Japanese patients with inflammatory bowel disease. Journal of gastroenterology 2010;45:1014-21.

525. Yang SK, Hong M, Baek J, et al. A common missense variant in NUDT15 confers susceptibility to thiopurine-induced leukopenia. Nature genetics 2014;46:1017-20.

526. Karas-Kuzelicki N, Jazbec J, Milek M, Mlinaric-Rascan I. Heterozygosity at the TPMT gene locus, augmented by mutated MTHFR gene, predisposes to 6-MP related toxicities in childhood ALL patients. Leukemia 2009;23:971-4.

527. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet 2008;40:695-701.

528. Lee S, Emond MJ, Bamshad MJ, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet 2012;91:224-37.

529. Basu S, Pan W. Comparison of statistical tests for disease association with rare variants. Genet Epidemiol 2011;35:606-19.

530. Ionita-Laza I, Lee S, Makarov V, Buxbaum JD, Lin X. Sequence kernel association tests for the combined effect of rare and common variants. Am J Hum Genet 2013;92:841-53.

531. Ford LT, Berg JD. Determination of thiopurine S-methyltransferase activity in erythrocytes using 6-thioguanine as substrate and a non-extraction liquid chromatographic technique. J Chromatogr B Analyt Technol Biomed Life Sci 2003;798:111-5.

532. Ford L, Graham V, Berg J. Whole-blood thiopurine S-methyltransferase activity with genotype concordance: a new, simplified phenotyping assay. Ann Clin Biochem 2006;43:354-60.

533. Heap GA, Weedon MN, Bewshea CM, et al. HLA-DQA1-HLA-DRB1 variants confer susceptibility to pancreatitis induced by thiopurine immunosuppressants. Nature genetics 2014;46:1131-4.

534. H. M. Kang XZ, X. Sim, C. Ma Biostatistics Dept, Univ Michigan, Ann Arbor, Ann Arbor, MI. “EPACTS (Efficient and Parallelizable Association Container Toolbox).” N. p., n.d.

535. Schaeffeler E, Fischer C, Brockmeier D, et al. Comprehensive analysis of thiopurine S-methyltransferase phenotype-genotype correlation in a large population of German-Caucasians and identification of novel TPMT variants. Pharmacogenetics 2004;14:407-17.

536. Chisick L, Oleschuk C, Bernstein CN. The utility of thiopurine methyltransferase enzyme testing in inflammatory bowel disease. Can J Gastroenterol 2013;27:39-43.

537. Warner B, Johnston E, Arenas-Hernandez M, Marinaki A, Irving P, Sanderson J. A practical guide to thiopurine prescribing and monitoring in IBD. Frontline Gastroenterol 2018;9:10-5.

538. Giverhaug T, Klemetsdal B, Lysaa R, Aarbakke J. Intraindividual variability in red blood cell thiopurine methyltransferase activity. European journal of clinical pharmacology 1996;50:217-20.

539. Lewis LD, Benin A, Szumlanski CL, et al. Olsalazine and 6-mercaptopurine-related bone marrow suppression: a possible drug-drug interaction. Clinical pharmacology and therapeutics 1997;62:464-75.

540. Colombel JF, Ferrari N, Debuysere H, et al. Genotypic analysis of thiopurine S-methyltransferase in patients with Crohn's disease and severe myelosuppression during azathioprine therapy. Gastroenterology 2000;118:1025-30.

541. Ichida K, Matsumura T, Sakuma R, Hosoya T, Nishino T. Mutation of human molybdenum cofactor sulfurase gene is responsible for classical xanthinuria type II. Biochemical and biophysical research communications 2001;282:1194-200.

542. Russel FG, Koenderink JB, Masereeuw R. Multidrug resistance protein 4 (MRP4/ABCC4): a versatile efflux transporter for drugs and signalling molecules. Trends in pharmacological sciences 2008;29:200-7.

543. Zabala W, Cruz R, Barreiro-de Acosta M, et al. New genetic associations in thiopurine-related bone marrow toxicity among inflammatory bowel disease patients. Pharmacogenomics 2013;14:631-40.

544. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. Nature 2012;490:61-70.

545. Akbani R, Ng PK, Werner HM, et al. Corrigendum: A pan-cancer proteomic perspective on The Cancer Genome Atlas. Nat Commun 2015;6:4852.

546. Hoadley KA, Yau C, Wolf DM, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. Cell 2014;158:929-44.

547. Duffy D, Rouilly V, Libri V, et al. Functional analysis via standardized whole-blood stimulation systems defines the boundaries of a healthy immune response to complex stimuli. Immunity 2014;40:436-50.

Publications

Immuno-Genomic Profiling of Patients with Inflammatory Bowel Disease: A Systematic Review of Genetic and Functional In Vivo Studies of Implicated Genes

Tracy Coelho, MRCPCH,[*,†] Gaia Andreoletti, MSc,*[*] James J. Ashton, BM, BS,*[†] Reuben J. Pengelly, MBiol,*[*] Yifang Gao, PhD,*[‡] Ananth RamaKrishnan, MD,*[‡] Akshay Batra, MD,*[†] Robert M. Beattie, MRCP, FRCPCH,*[†] Anthony P. Williams, PhD,*[‡] and Sarah Ennis, PhD**

Background: Over the last 2 decades, there has been an ever-expanding catalog of genetic variants implicated in inflammatory bowel disease (IBD) through genome-wide association studies and next generation sequencing. In this article, we highlight the remarkable developments in understanding the genetic and immunological basis of IBD. The main objective of the study was to perform a systematic review of published literature detailing functional/immunological studies in patients known to harbor genetic variations in the implicated genes.

Methods: A panel of 71 candidate genes implicated in IBD was prioritized using 5 network connectivity in silico methods. An electronic search using MEDLINE and EMBASE from 1996 to February 2014 for each of the selected genes was conducted. Only studies describing genotyped IBD cohorts with concurrent in vivo functional studies were included.

Results: Between the reviewers, a total of 35,142 potentially eligible publications were identified. Only 8 genes had publications meeting the inclusion criteria. A total of 67 studies were identified across the selected genes. The *NOD2* gene had the most number with 41 studies followed by *IL-10* with 11 eligible studies. A meta-analysis was not practical given the heterogeneity of the study design and the number of implicated genes with diverse immunological and physiological functions.

Conclusions: There is a clear lack of functional studies in humans to assess the in vivo impact of the various genetic variants implicated. A collaborative approach merging genomics and functional studies will help to unravel the obscure mechanisms involved in IBD.

(Inflamm Bowel Dis 2014;20:1813–1819)

Key Words: inflammatory bowel disease, Crohn's disease, ulcerative colitis, functional studies, genetic studies

Inflammatory bowel disease (IBD), like most other common diseases, has a complex pathobiology involving multiple genetic, immunological, and environmental factors. Crohn's disease (CD) and ulcerative colitis (UC) are the 2 main phenotypes of IBD, which can present with a diverse but quite often overlapping symptomatology. Genetics plays a major role in IBD. Over the last decade or so, with the advent of genome-wide association studies (GWAS) and next generation sequencing, a more defined albeit complex interplay between genes, host immunity, and the resident microbiota has emerged.[1,2] This has prompted a proliferation of research studies refining genetic loci to identify causal variants, functional studies to assess the complexities of the immune networks, and more recently a hugely topical subject of the role of microbiome in the pathogenesis of IBD.[3,4] The current hypothesis is that in health, there is a well-balanced homeostasis between the gut immune system and the resident microbiota of the gut. Any break down of this homeostasis which can occur in genetically susceptible individuals due to inherent "weaknesses" in their immune check points can lead to inflammatory changes in the gut wall as seen in IBD.[3] In this article, we present an overview of the genetic milestones that underpin the substantial contribution of the immune sensing and response in the pathogenesis of IBD. We then provide a systematic review of available literature highlighting studies with a combined approach of in vivo assessment of "aberrant" pathways suspected on the basis of genetic variations in human cohorts with IBD.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.ibdjournal.org).

Received for publication June 10, 2014; Accepted July 3, 2014.

From the *Human Genetics and Genomic Medicine, Faculty of Medicine, University of Southampton, Southampton, United Kingdom; †Pediatric Gastrointestinal Unit, University Hospital Southampton, Southampton, United Kingdom; and ‡Cancer Sciences Division, Faculty of Medicine, University of Southampton, Southampton, United Kingdom.

T. Coelho is funded by CICRA (Crohn's in Childhood Research Association, United Kingdom). The remaining authors have no conflicts of interest to disclose.

Reprints: Tracy Coelho, MRCPCH, Human Genetics and Genomic Medicine, Duthie Building (Mp 808), University Hospital Southampton Foundation Trust, Southampton SO16 6YD, United Kingdom (e-mail: t.f.coelho@soton.ac.uk).

DOI 10.1097/MIB.0000000000000174

Published online 28 August 2014.

A HISTORICAL JOURNEY THROUGH GENETICS OF IBD

Since the original description of CD by Crohn in 1932,[5] there have been several lines of epidemiological pointers implicating genetic predisposition in the pathogenesis of IBD. In the 1980s, several studies confirmed the early findings of familial clustering of CD and suggested a positive family history ranging between 8% and 25%.[6] Large studies were carried out in the late 1980s in the United Kingdom and Scandinavia, which showed an increased concordance in twin studies.[7,8] The search for a long-known or suspected heritability in IBD, prompted by these early studies was then driven toward family-based linkage analysis in the 1990s to identify specific susceptibility genes.[9] At least 6 chromosomal regions were identified through linkage studies as linked to IBD and were named from IBD locus 1 to 6 according to their date of reporting and independent replication. The discovery of IBD-1 locus on chromosome 16 was a major breakthrough and till date is regarded as a single largest genetic risk factor for CD.[9,10]

Over the last 2 decades, gene discovery in complex diseases such as IBD has advanced rapidly through genome-wide scans. These studies have made a substantial impact in providing insights into the understanding of the disease and its complex biology. Until now, there have been at least 15 GWAS of IBD and 3 meta-analyses, which have successfully identified a total of 163 loci for IBD.[11] However, a major disadvantage of GWAS has been their intrinsic limitation to detection of common disease variation and so at best, this approach can only account for a portion of the heritability of the disease. Rare variants missed by GWAS may contribute significantly toward the missing heritability of IBD.[1,12] Technological advancements in recent years such as next generation sequencing now offer a feasible modality for studying rare and novel variation in disease causality.[13]

GENES AND GASTROINTESTINAL HOMEOSTASIS

Candidate genes implicated in IBD highlight the interplay between several cellular mechanisms and immune pathways that are crucial for maintaining gastrointestinal homeostasis. These mechanisms broadly include the barrier function of the epithelium, innate immune regulation of microbial invasion, and the various effectors and regulators of adaptive immune response. Mutations in the key regulatory genes result in perturbations in the carefully balanced homeostasis that exists between the gastrointestinal immune system and the complex microbial milieu of the gut. The adverse outcome of this imbalance is inflammation of the gut resulting in IBD. We briefly present an overview of the complex levels of defense constantly on the role in the gut and their role in the pathogenesis of IBD.

EPITHELIAL BARRIER AND OTHER LUMINAL ELEMENTS OF DEFENSE

The goblet cells of the intestinal epithelium secrete glycosylated mucins that form a mucus matrix over the epithelium, forming the first level of defense against any microbial invasion. The colon has a dual mucus layer; the inner layer has properties that restrict bacterial motility and adhesion to the epithelium.[14,15] The inner mucus layer is formed from sheets of *Muc2* mucin, which physically separates the epithelium from the bacteria. *Muc2*-null (*Muc2*$^{-/-}$) mice do not have any protective mucus and develop spontaneous colitis. Interestingly, it has also been observed that patients with active UC can have defective and penetrable mucus overlying the epithelium.[16] Epithelial cells associate with each other through a series of intercellular junctions, the most important being apical junction complex, which consists of tight junctions and adherens junctions. Epithelial barrier integrity is crucial for intestinal homeostasis in the context of IBD, and several genes associated with epithelial integrity are now implicated in IBD. Some of these genes include *CDH1*, *GNA12*, *HNF4A*, *ERRRFI1*, *MUC19*, *ITLN1*, and *PTPN2*.[1]

NOD2 AND AUTOPHAGY

NOD2, also called CARD15 is an intracellular pathogen recognition receptor that recognizes N-acetyl muramyl dipeptide (MDP), derived from bacterial cell wall degradation.[17] One of the most remarkable achievements in the genetics of IBD came in 2001 when fine mapping of the IBD-1 locus on chromosome 16 by a French group identified the leucine-rich repeat variants of the *NOD2* gene (Nucleotide-binding oligomerization domain-containing-2) as conferring susceptibility to CD. The variants commonly found in the *NOD2* gene as described in the original study include single-nucleotide polymorphism (SNP) 8 (R702W), SNP 12 (G908R), and SNP 13 (1007 fs).[10] A number of rare genetic variants have been described subsequently, all of which almost exclusively localize to the leucine-rich repeat region. NOD2 plays a critical role in the induction of autophagy, which is a process whereby the cell tends to autodigest damaged intracellular organelles or intracellular bacteria by formation of an "isolation membrane" which is sequestered and marked for degradation. NOD2 recruits the autophagy protein ATG16L1 to the plasma membrane at the bacterial entry site. Cells with mutant NOD2 are incapable of this directive ATG16L1 recruitment and consequently fail to entrap pathogens through autophagy.[18]

TH17 CELL SIGNALING

Until recently, it was believed that intestinal inflammation in CD is mediated by a Th1 response (tumor necrosis factor-α [TNF-α], interleukin [IL]-12, interferon-γ) and in UC by Th2 cytokine pathways (e.g., IL-4, IL-5, IL-13). Evidently, with the emerging role of other T-cell lineages such as regulatory T cells (Tregs) and Th17 cells, the concept of Th1 response and Th2 response in CD and UC, respectively, as the primary pathways of inflammation has been largely superseded. In health, a finely tuned homeostasis exists between the Tregs and the proinflammatory T-helper cells. An overzealous Th17 response or an inadequate Treg response can tip the balance toward an undesired inflammatory response. Th17 cells, which produce highly potent proinflammatory cytokines such as IL-17, are abundantly found in the mucosa of patients with active IBD.[19,20] Several genes in the Th17 pathway have been linked with IBD susceptibility, including *IL-23R*, *TNFSF15*, *STAT3*, *IL-12B*, *CCR6*, and *JAK2*.[21]

IL-10 SIGNALING PATHWAY IN IBD

IL-10 is secreted by a wide variety of cells and over the last many years, it has been identified as crucial anti-inflammatory cytokine essential for maintaining gut homeostasis. IL-10 restricts the secretion of proinflammatory cytokines such as TNF-α and IL-12.[22] IL-10 signaling is required for limiting the expansion of Th17 cells, which are proinflammatory mercenaries.[23] The IL-10 receptor (IL-10R) has 2 alpha subunits and 2 beta subunits, which are encoded by *IL-10RA* and *IL-10RB*, respectively. Homozygous or compound heterozygous mutations in IL-10 or its receptor subunits have been well described in literature. Patients usually present with a very severe form of IBD at a very young age (<1 yr), with poor response to conventional therapy. These studies have highlighted the successful role of hematopoietic stem cell transplant in bringing about a sustained remission in these patients.[22,24,25]

JAK-STAT PORTAL IN IBD

The Janus kinase/signal transducer and activator of transcription (JAK/STAT) pathway constitute a major portal for vital cellular processes including cell growth, differentiation, proliferation, and several immune mechanisms. Various cytokines and effectors communicate through this pathway to orchestrate an appropriate cellular response through target gene expression.[26] GWAS have implicated several genes in JAK-STAT pathway as candidate genes for IBD. Some of these include *JAK2*, *TYK2*, *STAT3*, and *STAT4*, with genetic variants associated with an increased risk of developing IBD.[11,21]

SYSTEMATIC REVIEW OF IMMUNE FUNCTION STUDIES IN GENOTYPED INDIVIDUALS

Clear interpretation of the functional relevance of the many newly discovered genomic loci in IBD and associated variants is limited by the palpable lack of functional studies to characterize the molecular aberrations caused by these variants. Although many candidate genes implicated by genetic studies can be mapped to known pathways, a substantial fraction of these genes (>40%) are poorly understood at the functional level.[27] Establishing the causal effect of a genetic variant requires a mechanistic insight by the way of functional studies. With this in mind, we performed a systematic search of a selected panel of genes implicated in IBD through GWAS and other genetic studies. We specifically searched for studies in patients with IBD, where functional/immunological assessment of genotyped individuals was carried out to assess the in vivo functional impact of the genetic variants. This is a unique systematic review of literature aimed at exploring the strengths of integrating genetics/genomics with mechanistic studies in establishing causality of the multitude of variants implicated in IBD.

METHODS

Selection of Genes

Based on the largest and the most recent meta-analysis of IBD genome-wide association scans, the overall number of IBD loci is estimated as 163.[11] In this study, causal genes within the IBD loci were prioritized using 5 network connectivity tools such as Gene Relationships Across Implicated Loci, Disease Association Protein–Protein Link Evaluator, 3 different sources of expression quantitative trait locus, coding SNPs, and Coexpression network analysis (Fig. 1). Given the individual limitations of each of the tools used for functional clustering of genes, we adopted a consensus approach, requiring genes to be simultaneously implicated by at least 2 of the in silico methods. This approach focused our analysis on 71 candidate genes (see Table, Supplemental Digital Content 1, http://links.lww.com/IBD/A555).

Literature Search

We conducted an electronic search through OvidSP using MEDLINE and EMBASE from 1996 to February 2014 for each of the 71 selected genes, looking specifically for genetic studies in IBD. A uniform search strategy was developed using a structured approach for every selected gene combining key words, gene symbol, and name as approved on "HUGO Gene Nomenclature Committee," using Boolean terms enabling multiple combinations of terms to be searched at once. We used the approved gene symbol, approved name, previous names, and synonyms as on HGNC (HUGO Gene Nomenclature Committee) by standardizing syntax as and where necessary, during the conduct of the search. An example of the search on *ATG16L1* gene is shown in Table 1.

The initial search intentionally focused on the genetic aspects and not the functional component, so as to be able to retrieve all the relevant articles. A study was eligible if it reported in vivo immunological/functional assessment of patients with genotyped IBD, relevant to the genetic variants detected. No language restrictions were applied. The titles and abstracts of articles obtained from electronic searches were screened and assessed for their relevance. Selected articles identified after the initial screen were retrieved in full text, and the reference lists were scanned to look for literature that had not been obtained by searches. At each stage of the study selection, eligibility of publications was assessed independently by 2 reviewers, and discrepancies if any, were resolved through discussion.

Inclusion Criteria

Only studies describing genotyped cohorts with concurrent in vivo functional/immunological studies to assess the functional impact of genetic variants in individuals with IBD were included.

Exclusion Criteria

1. Murine models and other animal studies
2. Functional studies carried out on cell lines
3. Review articles with no original data
4. Functional studies in healthy volunteers (and not IBD patients), bearing genetic variants implicated in IBD
5. Studies on same/overlapping sets of patients published by same groups in different journals
6. Conference abstracts.

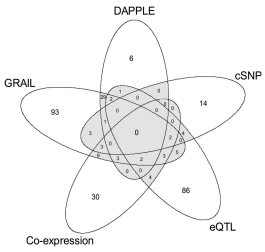


FIGURE 1. Shows the method used for prioritizing the genes implicated in IBD. The 5 network connectivity tools used in the third meta-analysis[11] to identify candidate genomic loci shown in the Venn diagram, namely Gene Relationships Across Implicated Loci (GRAIL), Disease Association Protein-Protein Link Evaluator, cSNP, Coexpression, and expression quantitative trait locus (eQTL). GRAIL uses text searching of abstracts in the scientific literature to identify linked genes; Disease Association Protein-Protein Link Evaluator (DAPPLE) interrogates known protein–protein interactions to identify proteins which are likely to physically interact; cSNP indicates where associated SNPs directly alter the protein, or are in strong linkage disequilibrium with known SNPs which alter the protein; eQTL denotes genes where associated SNPs are correlated with an alteration in protein expression levels and coexpression refers to genes for which expression patterns are linked to known inflammatory processes. For further details, see Jostins et al.[11] Genes identified by at least 2 connectivity tools were selected for the systematic search (71 genes out of 300; shaded region).

TABLE 1. Example of a Systematic Search on *ATG16L1*

1	IBD*.tw.
2	IBD*.tw.
3	Crohn*.tw.
4	CD.tw.
5	UC.tw.
6	UC.tw.
7	1 or 2 or 3 or 4 or 5 or 6
8	ATG16L1.tw.
9	Autophagy related 16-like 1.tw.
10	"APG16 autophagy 16-like (*Saccharomyces cerevisiae*)".tw.
11	APG16L.tw.
12	"ATG16 autophagy related 16-like (*S. cerevisiae*)".tw.
13	"ATG16 autophagy related 16-like 1 (*S. cerevisiae*)".tw.
14	ATG16L.tw.
15	ATG16A.tw.
16	FLJ10035.tw.
17	WDR30.tw.
18	8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17
19	7 and 18

Numbers 1 to 7 to include all publications with IBD, number 8 includes approved gene symbol, number 9 approved gene name, 10 to 14 previous names of the gene, and 15 to 17 include synonyms for the gene.

RESULTS

A structured search across the 71 genes yielded a large number of studies involving genetic and functional studies on individuals with IBD. Between the reviewers, we identified 35,142 potentially eligible publications, which were assessed through titles, abstracts, and full text where appropriate. Of the 71 genes interrogated, only 8 genes had publications meeting our inclusion criteria. A total of 67 studies were identified across the selected genes, which met the criteria for inclusion. The *NOD2* gene had the most number of immuno-genomic studies, 41 studies meeting the eligibility criteria, followed by *IL-10* with 11 eligible studies. We also included *IL-10RA* and *IL-10RB* (genes encoding the receptor subunits), as most genetic studies on *IL-10* invariably included the receptor genes as well. Other genes investigated using functional studies on genotyped cohorts of IBD included *ATG16L1* (7 studies), *SLC22A4* (3 studies), *IL-23R* (2 studies), and *SLC11A1*, *CCL2*, and *STAT3* had one each (see Table, Supplemental Digital Content 2, http://links.lww.com/IBD/A556).

For brevity, here, we present an overview of the studies retrieved through our search strategy on *NOD2* and *IL-10* genes (including *IL-10RA* and *IL-10RB*).

Review of Selected Studies on the *NOD2* Gene

A structured search on the *NOD2* gene retrieved a total of 4994 publications between the 2 reviewers. Details of genotyping and functional methods on the 41 selected articles are given in Table, Supplemental Digital Content 2, http://links.lww.com/IBD/A556.

It is understood that NOD2 after the recognition of MDP triggers a defense response through NF-κB pathway. Hence, most of the functional studies focus on the assessment of the read-outs of the NF-κB pathway such as IL-8, TNF-α, IL-1β, and other cytokines for mechanistic assessment of the NOD2 protein.[28–34] Some studies have shown impaired NF-κB activation in response to MDP by detecting reduced phosphorylation levels in nuclear extracts of cells stimulated by MDP.[35,36] Rahman et al performed studies on Treg cells isolated from patients with CD with *NOD2* mutations, patients with wild-type allele, and healthy volunteers. They demonstrated that MDP-stimulated Tregs were normally protected from apoptosis; however, this protection was not evident in patients with *NOD2* polymorphisms.[37]

Hedl et al in 2007 found that pretreatment with MDP, significantly decreased production of the proinflammatory cytokines TNF-α, IL-8, and IL-1β on NOD2, TLR4, and TLR2 restimulation in primary human monocyte-derived macrophages. MDP-stimulated macrophages from CD-relevant Leu1007insC

NOD2 homozygote individuals were deficient in their ability to cross-tolerize to subsequent treatment with TLR2 and TLR4 ligands.[33] On the same lines, NOD2/TLR cross talk has also been implicated by work from other groups.[29,38–44] Some studies proposed a cross talk and synergy between NOD1 AND NOD2 pathways, suggesting that *NOD2* mutations could lead to down regulation of NOD1 signaling.[45]

More recently, studies have also concentrated on the role of NOD2 in autophagy induction and the impaired bacterial handling consequent to *NOD2* mutations. Cooney et al showed that in cells homozygous for the CD-associated *NOD2* frame-shift mutation, mutant NOD2 fails to recruit ATG16L1 to the plasma membrane, resulting in an impaired engulfment of invading bacteria by autophagosomes. Their findings revealed that dendritic cells from patients with CD bearing *NOD2* mutations were defective in autophagy induction, bacterial trafficking, and antigen presentation.[46]

In terms of the methods used for genotyping of individuals, 38 of the selected studies performed genotyping for *NOD2* variants, 3 studies performed candidate gene sequencing, and none employed exome sequencing. Blood specimens were analyzed for functional work in 34 studies, ileal/colonic tissue in 5 studies, and a combination of peripheral blood cells and gut tissue in 2 studies. The synopsis of excluded studies on *NOD2* gene is given in Table 2.

Review of Studies on the *IL-10*, *IL-10RA*, and *IL-10RB* Genes

Our search retrieved a total of 7300 publications, 11 of which met the selection criteria (see Table, Supplemental Digital Content 2, http://links.lww.com/IBD/A556). Immunological studies have focused on the functional capacity of IL-10 to limit the secretion of TNF-α and other proinflammatory cytokines. The IL-10 receptor consists of 2 alpha molecules (IL-10R1 encoded by *IL-10RA*) and 2 beta molecules (IL-10R2 encoded by *IL-10RB*). Mutations abrogate IL-10–induced signaling, causing disruption of the IL-10 dependent "negative feedback" regulation. Koss et al studied the influence of biallelic polymorphisms in TNF-α, lymphotoxin-α, and IL-10 genes on TNF-α and IL-10 production. Three haplotypes of IL-10 were identified. The functional effect of IL-10 haplotypes on IL-10 protein production in whole blood samples stimulated with lipopolysaccharide (LPS) was analyzed.[47] Van der Linde et al identified a point mutation (Gly15Arg) in the leader sequence of IL-10 following genotyping of IL-10 alleles in 17 sibling pairs with CD and 75 healthy controls. The functional consequences of this genetic variation were tested by stimulating peripheral blood mononuclear cells of patients bearing this mutation with LPS or phorbol ester, and then assessing the cellular supernatants for IL-10 production by enzyme-linked immunosorbent assay or Western blotting. The activity of recombinant immature wild-type or mutated IL-10 was also tested in vitro in a proliferation assay with LPS-stimulated human monocyte cell line (HL60 cells).[48] In keeping with previous reports, Glocker et al, in 2009, proposed that the pathophysiology of a deficiency in the IL-10 receptor involves undue and prolonged activation of mononuclear cells on exposure to bacterial particles, resulting in an exaggerated efflux of inflammatory cytokines such as TNF-α causing mucosal damage. To test this hypothesis, TNF-α secretion of monocytes and monocyte-derived macrophages was analyzed on exposure to LPS or LPS plus IL-10 in patients with the receptor *IL-10RA* mutations and in healthy controls. Patients in this group had a very early onset IBD (in the first year of life) with a very severe disease refractory to conventional treatment. Mutations were identified through genetic-linkage analysis and candidate gene sequencing on samples from 2 unrelated consanguineous families with children who were affected by very early onset IBD and from 6 additional patients with very early onset IBD. IL-10 substantially reduced the release of TNF-α in cells from control subjects; however, this inhibitory effect was absent in cells from patients with the receptor mutations. The impairment in the capacities of mononuclear cells with IL-10 or IL-10 receptor deficiencies to inhibit TNF-α production have been reproduced through studies by other groups subsequently.[49–52]

Of the 11 selected studies, whole exome sequencing was used in 1 study,[51] candidate gene sequencing in 6 studies[22,48–50,52,53] and SNP genotyping in 4 studies.[47,54–56] Specimens analyzed for functional studies were blood samples in 8 studies, and in 3 studies, ileal/colonic tissue was assessed in addition to blood samples. The synopsis of excluded studies on *IL-10* is given in Table 2.

TABLE 2. A Synopsis of All Excluded Studies on *NOD2* Gene and *IL-10/RA/RB* Gene Retrieved Through Initial Search (Relevant Studies but not Meeting Eligibility Criteria)

Studies	*NOD2* Gene	*IL-10/IL-10RA/ IL-10RB* Gene
Total no. publications screened	1670	2354
Functional studies in humans (no genotyping), n (%)	62 (3.7)	289 (12.3)
Human genetic studies including GWAS and meta-analysis (no functional studies), n (%)	315 (18.8)	55 (2.3)
Studies on animal models (genetic or functional), n (%)	98 (5.8)	845 (35.9)
Studies on cell lines (genetic or functional), n (%)	22 (1.3)	22 (1)
Reviews (no original data), n (%)	81 (4.8)	59 (2.5)
Selected studies, n (%)[a]	41 (2.5)	11 (0.46)
Publications not pertinent to our review and duplicate studies, n (%)	1051 (63)	1073 (45.5)

Assessed by a Single Reviewer TC on EMBASE.
[a]Details of selected articles included in Supplemental Digital Content.

DISCUSSION

During our structured search for immuno-genomic studies in IBD, we found a large number of functional studies conducted in humans, focusing on the implicated immunological pathways. A meta-analysis was not practical given the heterogeneity of the study design and the number of implicated genes with diverse immunological and physiological functions. For the *NOD2* gene, 3.7% of the studies (66 studies out of 1670 publications retrieved) were functional studies carried out in humans, and for the *IL-10* gene (including *IL-10/IL-10RA/IL-10RB*), human functional studies were 12.3% (289 studies out of 2354 publications). However, these studies were conducted on cohorts without a genotyped profile. Conversely, we found an ever-expanding number of studies with genotyped cohorts, but without the functional element (19% for *NOD2* and 2.5% for *IL-10*). There could be several reasons for this discrepant observation. One of the key factors may be due to the fact that IBD in a given individual may be due to multiple hits at the implicated loci. Several inflammatory mechanisms and pathways may be involved with varying degrees of contributory impact, thereby making it extremely difficult to design functional studies bespoke to the genetic variants identified in a given individual with IBD. Yet, another reason for the paucity of functional studies to backup implicated genetic variants is the relative lack of enthusiasm to invest in conducting functional work in a GWAS variant where there is no proven causality.

An extensive number of murine models have contributed significantly toward understanding the mechanistic basis with both induced and spontaneous mutations in a diversity of genes. This was clearly obvious during our search, which generated a vast number of experimental studies in knockout models. We identified 98 (6%) of 1670 publications, conducted as preclinical experiments in animal models for the *NOD2* gene and 845 (36%) of 2354 publications for the *IL-10/IL-10RA/IL-10RB* gene. As is obvious, the search retrieved a significantly larger number of animal model studies for *IL-10* as compared with *NOD2* because of the extensive use of *IL-10* knockout models for experimental colitis. These experimental models have enhanced our understanding of the functional impact of specific genes in a defined biological process; however, it does not effectively map out how genetic variants in human population will impact immune function. Animal models, although crucial for identifying pathways of susceptibility, are not an ideal platform to establish and define the pathogenetic associations in humans. Similarly, transformed cell lines are good experimental models for transfecting in a gene when the behavior of the cell itself is not of interest. However, this can be a major pitfall whilst trying to establish the actual functional impact caused by genetic variants in human disease. Also, transfection of cell lines can lead to overexpression of genes resulting in an unpredictable outcome. Therefore, in this context, to determine and decipher the role of candidate genes in the causation of disease, in vivo studies in human cohorts may be more meaningful.

Our study was limited to only 71 genes based on our inclusion strategy. It is possible that the immuno-genomic landscape is different for the other genes identified in IBD pathogenesis. Some of the commonly implicated genes such as *IRGM*, *ERAP2*, *MUC19*, *CDH1*, and others were not identified for our review based on the nature of our methodology. Given the heterogeneity of the functional methods used, selection of publications was subject to reviewer bias. This was however kept to the minimum through a standardized assessment, discussion, and consensus at all stages of the review. A substantial number of studies evaluated a panel of genes for genetic profiling, rather than a single gene resulting in an unavoidable overlap between studies across the genes under assessment. For example, some of the selected studies were included for both *ATG16L1* and *NOD2* genes, as the studies evaluated both the genes from an immunogenomic angle given that the 2 genes work closely together.[46,57]

Human genomics has provided key insights into the complexities of the biology of IBD. To consolidate and take it to the next level of understanding, there is now a clear pressing need for more collaborative approach between human genomics and immunology. Immuno-genomic profiling can possibly inform a risk-prediction model for complicated IBD progression. This will enhance the prospects of a more refined tailor-made diagnostic and therapeutic approach in IBD for the foreseeable future.

REFERENCES

1. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. *Nature*. 2011;474:307–317.
2. Lees CW, Barrett JC, Parkes M, et al. New IBD genetics: common pathways with other diseases. *Gut*. 2011;60:1739–1753.
3. Hansen R, Berry SH, Mukhopadhya I, et al. The microaerophilic microbiota of de-novo paediatric inflammatory bowel disease: the BISCUIT study. *PLoS One*. 2013;8:e58825.
4. Qin J, Li R, Raes J, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. 2010;464:59–65.
5. Crohn BB, Ginzburg L, Oppenheimer GD. Landmark article Oct 15, 1932. Regional ileitis. A pathological and clinical entity. By Burril B. Crohn, Leon Ginzburg, and Gordon D. Oppenheimer. *JAMA*. 1984;251:73–79.
6. Peeters M, et al. Familial aggregation in Crohn's disease: increased age-adjusted risk and concordance in clinical characteristics. *Gastroenterology*. 1996;111:597–603.
7. Henderson P, Satsangi J. Genes in inflammatory bowel disease: lessons from complex diseases. *Clin Med*. 2011;11:8–10.
8. Tysk C, Lindberg E, Järnerot G, et al. Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. *Gut*. 1988;29:990–996.
9. Cardinale CJ, Kelsen JR, Baldassano RN, et al. Impact of exome sequencing in inflammatory bowel disease. *World J Gastroenterol*. 2013;19:6721–6729.
10. Hugot JP, Chamaillard M, Zouali H, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature*. 2001;411:599–603.
11. Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012;491:119–124.
12. Van Limbergen J, Radford-Smith G, Satsangi J. Advances in IBD genetics. *Nat Rev Gastroenterol Hepatol*. 2014;11:372–385.
13. Rivas MA, Beaudoin M, Gardet A, et al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat Genet*. 2011;43:1066–1073.
14. Cader MZ, Kaser A. Recent advances in inflammatory bowel disease: mucosal immune cells in intestinal inflammation. *Gut*. 2013;62:1653–1664.
15. Laukoetter MG, Nava P, Nusrat A. Role of the intestinal barrier in inflammatory bowel disease. *World J Gastroenterol*. 2008;14:401–407.
16. Johansson ME, Gustafsson JK, Holmén-Larsson J, et al. Bacteria penetrate the normally impenetrable inner colon mucus layer in both murine colitis models and patients with ulcerative colitis. *Gut*. 2014;63:281–291.
17. Fritz T, Niederreiter L, Adolph T, et al. Crohn's disease: NOD2, autophagy and ER stress converge. *Gut*. 2011;60:1580–1588.

Copyright © 2014 Crohn's & Colitis Foundation of America, Inc. Unauthorized reproduction of this article is prohibited.

Downloaded from https://academic.oup.com/ibdjournal/article-abstract/20/10/1813/4578872 by University of Southampton user on 03 July 2019

18. Muzes G, Tulassay Z, Sipos F. Interplay of autophagy and innate immunity in Crohn's disease: a key immunobiologic feature. *World J Gastroenterol*. 2013;19:4447–4454.
19. Fujino S, Andoh A, Bamba S, et al. Increased expression of interleukin 17 in inflammatory bowel disease. *Gut*. 2003;52:65–70.
20. Rovedatti L, Kudo T, Biancheri P, et al. Differential regulation of interleukin 17 and interferon gamma production in inflammatory bowel disease. *Gut*. 2009;58:1629–1636.
21. Franke A, McGovern DP, Barrett JC, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet*. 2010;42:1118–1125.
22. Glocker EO, Kotlarz D, Boztug K, et al. Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. *N Engl J Med*. 2009;361:2033–2045.
23. Chaudhry A, Samstein RM, Treuting P, et al. Interleukin-10 signaling in regulatory T cells is required for suppression of Th17 cell-mediated inflammation. *Immunity*. 2011;34:566–578.
24. Engelhardt KR, Shah N, Faizura-Yeop I, et al. Clinical outcome in IL-10- and IL-10 receptor-deficient patients with or without hematopoietic stem cell transplantation. *J Allergy Clin Immunol*. 2013;131:825–830.
25. Moran CJ, Walters TD, Guo CH, et al. IL-10R polymorphisms are associated with very-early-onset ulcerative colitis. *Inflamm Bowel Dis*. 2013;19:115–123.
26. Coskun M, Salem M, Pedersen J, et al. Involvement of JAK/STAT signaling in the pathogenesis of inflammatory bowel disease. *Pharmacol Res*. 2013;76:1–8.
27. Graham DB, Xavier RJ. From genetics of inflammatory bowel disease towards mechanistic insights. *Trends Immunol*. 2013;34:371–378.
28. Li J, Moran T, Swanson E, et al. Regulation of IL-8 and IL-1beta expression in Crohn's disease associated NOD2/CARD15 mutations. *Hum Mol Genet*. 2004;13:1715–1725.
29. van Heel DA, Ghosh S, Butler M, et al. Muramyl dipeptide and toll-like receptor sensitivity in NOD2-associated Crohn's disease. *Lancet*. 2005;365:1794–1796.
30. van Heel DA, Hunt KA, Ghosh S, et al. Normal responses to specific NOD1-activating peptidoglycan agonists in the presence of the NOD2 frameshift and other mutations in Crohn's disease. *Eur J Immunol*. 2006;36:1629–1635.
31. van Heel DA, Hunt KA, King K, et al. Detection of muramyl dipeptide-sensing pathway defects in patients with Crohn's disease. *Inflamm Bowel Dis*. 2006;12:598–605.
32. Peeters H, Bogaert S, Laukens D, et al. CARD15 variants determine a disturbed early response of monocytes to adherent-invasive Escherichia coli strain LF82 in Crohn's disease. *Int J Immunogenet*. 2007;34:181–191.
33. Hedl M, Li J, Cho JH, et al. Chronic stimulation of Nod2 mediates tolerance to bacterial products. *Proc Natl Acad Sci U S A*. 2007;104:19440–19445.
34. Lappalainen M, Paavola-Sakki P, Halme L, et al. Novel CARD15/NOD2 mutations in Finnish patients with Crohn's disease and their relation to phenotypic variation in vitro and in vivo. *Inflamm Bowel Dis*. 2008;14:176–185.
35. Seidelin JB, Broom OJ, Olsen J, et al. Evidence for impaired CARD15 signalling in Crohn's disease without disease linked variants. *PLoS One*. 2009;4:e7794.
36. Kuuliala K, Lappalainen M, Turunen U, et al. Detection of muramyl dipeptide-sensing pathway defects in monocytes of patients with Crohn's disease using phospho-specific whole blood flow cytometry. *Scand J Clin Lab Invest*. 2013;73:494–502.
37. Rahman MK, Midlting EH, Svingen PA, et al. The pathogen recognition receptor NOD2 regulates human FOXP3+ T cell survival. *J Immunol*. 2010;184:7247–7256.
38. van Heel DA, Ghosh S, Hunt KA, et al. Synergy between TLR9 and NOD2 innate immune responses is lost in genetic Crohn's disease. *Gut*. 2005;54:1553–1557.
39. Braat H, Stokkers P, Hommes T, et al. Consequence of functional Nod2 and Tlr4 mutations on gene transcription in Crohn's disease patients. *J Mol Med (Berl)*. 2005;83:601–609.
40. Canto E, Moga E, Ricart E, et al. MDP-Induced selective tolerance to TLR4 ligands: impairment in NOD2 mutant Crohn's disease patients. *Inflamm Bowel Dis*. 2009;15:1686–1696.
41. Brosbol-Ravnborg A, Hvas CL, Agnholt J, et al. Toll-like receptor-induced granulocyte-macrophage colony-stimulating factor secretion is impaired in Crohn's disease by nucleotide oligomerization domain 2-dependent and -independent pathways. *Clin Exp Immunol*. 2009;155:487–495.
42. Butler M, Chaudhary R, van Heel DA, et al. NOD2 activity modulates the phenotype of LPS-stimulated dendritic cells to promote the development of T-helper type 2-like lymphocytes—possible implications for NOD2-associated Crohn's disease. *J Crohns Colitis*. 2007;1:106–115.
43. Vissers M, Remijn T, Oosting M, et al. Respiratory syncytial virus infection augments NOD2 signaling in an IFN-beta-dependent manner in human primary cells. *Eur J Immunol*. 2012;42:2727–2735.
44. Kullberg BJ, Ferwerda G, de Jong DJ, et al. Crohn's disease patients homozygous for the 3020insC NOD2 mutation have a defective NOD2/TLR4 cross-tolerance to intestinal stimuli. *Immunology*. 2008;123:600–605.
45. Netea MG, Ferwerda G, de Jong DJ, et al. The frameshift mutation in Nod2 results in unresponsiveness not only to Nod2- but also Nod1-activating peptidoglycan agonists. *J Biol Chem*. 2005;280:35859–35867.
46. Cooney R, Baker J, Brain O, et al. NOD2 stimulation induces autophagy in dendritic cells influencing bacterial handling and antigen presentation. *Nat Med*. 2010;16:90–97.
47. Koss K, Satsangi J, Fanning GC, et al. Cytokine (TNF alpha, LT alpha and IL-10) polymorphisms in inflammatory bowel diseases and normal controls: differential effects on production and allele frequencies. *Genes Immun*. 2000;1:185–190.
48. van der Linde K, Boor PP, Sandkuijl LA, et al. A Gly15Arg mutation in the interleukin-10 gene reduces secretion of interleukin-10 in Crohn disease. *Scand J Gastroenterol*. 2003;38:611–617.
49. Begue B, Verdier J, Rieux-Laucat F, et al. Defective IL-10 signaling defining a subgroup of patients with inflammatory bowel disease. *Am J Gastroenterol*. 2011;106:1544–1555.
50. Kotlarz D, Beier R, Murugan D, et al. Loss of interleukin-10 signaling and infantile inflammatory bowel disease: implications for diagnosis and therapy. *Gastroenterology*. 2012;143:347–355.
51. Mao H, Yang W, Lee PP, et al. Exome sequencing identifies novel compound heterozygous mutations of IL-10 receptor 1 in neonatal-onset Crohn's disease. *Genes Immun*. 2012;13:437–442.
52. Marcuzzi A, Girardelli M, Bianco AM, et al. Inflammation profile of four early onset Crohn patients. *Gene*. 2012;493:282–285.
53. Galatola M, Miele E, Strisciuglio C, et al. Synergistic effect of interleukin-10-receptor variants in a case of early-onset ulcerative colitis. *World J Gastroenterol*. 2013;19:8659–8670.
54. Gasche C, Grundtner P, Zwirn P, et al. Novel variants of the IL-10 receptor 1 affect inhibition of monocyte TNF-alpha production. *J Immunol*. 2003;170:5578–5582.
55. Wang AH, Lam WJ, Han DY, et al. The effect of IL-10 genetic variation and interleukin 10 serum levels on Crohn's disease susceptibility in a New Zealand population. *Hum Immunol*. 2011;72:431–435.
56. Wagner J, Skinner NA, Catto-Smith AG, et al. TLR4, IL10RA, and NOD2 mutation in paediatric Crohn's disease patients: an association with Mycobacterium avium subspecies paratuberculosis and TLR4 and IL10-RA expression. *Med Microbiol Immunol*. 2013;202:267–276.
57. Glubb DM, Gearry RB, Barclay ML, et al. NOD2 and ATG16L1 polymorphisms affect monocyte responses in Crohn's disease. *World J Gastroenterol*. 2011;17:2829–2837.

Downloaded from https://academic.oup.com/ibdjournal/article-abstract/20/10/1813/4578872 by University of Southampton user on 03 July 2019

SCIENTIFIC REPORTS

OPEN

Received: 06 June 2016
Accepted: 13 September 2016
Published: 05 October 2016

Genes implicated in thiopurine-induced toxicity: Comparing TPMT enzyme activity with clinical phenotype and exome data in a paediatric IBD cohort

Tracy Coelho[1,2,*], Gaia Andreoletti[1,*], James J. Ashton[2], Akshay Batra[2], Nadeem Ahmad Afzal[2], Yifang Gao[3], Anthony P. Williams[3], Robert M. Beattie[2] & Sarah Ennis[1]

The aim of our study was to assess the utility of next generation sequencing (NGS) for predicting toxicity and clinical response to thiopurine drugs in paediatric patients with inflammatory bowel disease. Exome data for 100 patients were assessed against biochemically measured TPMT enzyme activity, clinical response and adverse effects. The *TPMT* gene and a panel of 15 other genes implicated in thiopurine toxicity were analysed using a gene based statistical test (SKAT-O test). Nine patients out of 100 (Crohn's disease- 67, ulcerative colitis- 23 and IBDU-10) had known *TPMT* mutations associated with deficient enzyme activity. A novel and a highly pathogenic *TPMT* variant not detectable through standard genotyping, was identified through NGS in an individual intolerant to thiopurines. Of the 16 patients intolerant to thiopurines, NGS identified deleterious *TPMT* variants in 5 individuals whereas the biochemical test identified 8 individuals as intolerant (sensitivity 35.7% and 57.14%; specificity 93.75% and 50% respectively). SKAT-O test identified a significant association between *MOCOS* gene and TPMT activity ($p = 0.0015$), not previously reported. Although NGS has the ability to detect rare or novel variants not otherwise identified through standard genotyping, it demonstrates no clear advantage over the biochemical test in predicting toxicity in our modest cohort.

Thiopurine drugs, which include azathioprine and 6-mercaptopurine (6-MP) have been effectively used in inflammatory bowel disease for more than 30 years. They are also widely used in the treatment of patients with neoplastic conditions, post-organ transplantation and a wide range of autoimmune and inflammatory conditions. However concerns over toxicity and adverse reactions frequently result in discontinuation of treatment and a switch to alternative therapies[1,2]. Known adverse reactions include bone marrow suppression, severe gastric intolerance, pancreatitis, hepatotoxicity, skin reactions, susceptibility to infections, risk of malignancy and flu-like symptoms[2,3].

Thiopurine S-methyl transferase (TPMT) is a key enzyme involved in the metabolism of thiopurine drugs and functions by catalysing the S-methylation of aromatic and heterocyclic sulfhydryl groups[4]. TPMT is encoded by a gene located on chromosome 6p22, consisting of 10 exons, encoding one protein domain[5]. Although the precise mode of action of thiopurines is still unclear, the most important mechanism is thought to be the incorporation of 6-TGNs (thioguanine nucleotides) into the cell DNA, resulting in an impaired DNA synthesis and cell death[2,6] (Fig. 1). The *TPMT* gene is known to exhibit genetic heterogeneity resulting in wide inter-individual differences, both in terms of clinical efficacy and toxicity profiles based on the enzyme activity[7,7]. Approximately 4–11% of individuals of Caucasian origin are heterozygous for a mutant *TPMT* allele with intermediate enzyme activity, whereas approximately 1 in 300 individuals are homozygous or compound heterozygous with a consequent low or absent *TPMT* activity[4,7]. Current clinical guidelines recommend determining *TPMT* status in a given patient

[1]Human Genetics and Genomic medicine, University of Southampton, Southampton, UK. [2]Department of Paediatric Gastroenterology, University Hospital Southampton, Southampton, UK. [3]Cancer Sciences Division, Faculty of Medicine, University Hospital Southampton, Southampton, UK. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to T.C (email: T.F.Coelho@soton.ac.uk)

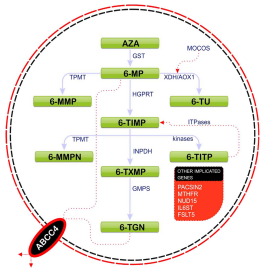


Figure 1. Schematic diagram showing thiopurine drug metabolism and genes implicated in thiopurine-induced toxicity. There are three main catabolic pathways for thiopurine drugs, following conversion of AZA to 6-MP: (1) Phosphorylation to 6-thioguanines (6-TGN) which are active metabolites; 6-MP is first converted by hypoxanthine guanine phosphoribosyl-transferase (HGPRT) to 6-thioinosine monophosphate (6-TIMP), which is then phosphorylated to 6-TGN with inosine 5-monophosphate dehydrogenase (IMPDH1 and IMPDH2) and guanosine monophosphate synthetase (GMPS) (2) Methylation by of 6-MP by TPMT to form 6-methyl-MP (6-MMP) which is an inactive metabolite and not a substrate for IMPDH) (3) Catabolism of 6-MP to 6-thiouracil (6-TU) via xanthine dehydrogenase (XDH, synonym- Xanthine oxidase) or aldehyde oxidase 1 (AOX1). TPMT competes with IMPDH for their common substrate 6-TIMP to form 6- methylmercaptopurine nucleotides (6-MMPN). 6-TIMP can be phosphorylated by kinases to 6-thioinosine triphosphate (6-TITP), which can get dephosphorylated by inosine triphosphatase (ITPase) to form 6-TIMP again. Although the precise mode of action of thiopurines is still unclear, the most important mechanism is thought to be the incorporation of 6-TGNs into the cell DNA, resulting in an impaired DNA synthesis and cell death[2,6]. **Abbreviations:** ABCC4- ATP-binding cassette, sub-family C (CFTR/MRP), member 4; AZA- azathioprine; FSLT5- Follistatin-Like 5; GST- glutathione s-transferase; HGPRT-hypoxanthine phosphoribosyltransferase;IL6ST- Interleukin 6 signal transducer; 6-MP- 6- mercaptopurine; 6-MMPN- 6- methyl mercaptopurine nucleotides; MOCOS- Molybdenum cofactor sulfurase; MTHFR-Methyl-enetetrahydrofolate reductase ;NUDT15- Nudix (nucleoside diphosphate linked moiety X)-type motif 15; PACSIN2- Protein kinase C and casein kinase substrate in neurons 2; 6-TXMP- 6-thioxanthosine monophosphate)

before commencement of thiopurine therapy; this is achieved by measuring *TPMT* enzyme activity in the circulating red blood cells *or* through genotyping known *TPMT* variants associated with enzyme deficiency[2,8,9].

The rationale for assessing the TPMT status before commencement of thiopurine therapy is to minimise the risk of adverse effects whilst aiming for an optimal clinical response. Thiopurines are best avoided in individuals who are deficient or have extremely low TPMT activity and administered at a reduced dose if TPMT activity is intermediate or low normal[2,9,10].

Although *TPMT* is the most crucial pharmaco-gene involved in the metabolism of thiopurines, previous studies have highlighted the role of other genes, whose products substantially alter drug metabolism and consequently impact clinical efficacy or toxicity[11–16]. It is plausible that some proportion of adverse effects observed whilst on treatment in the context of a normal *TPMT* status (genotype or phenotype), could be explained by variation in genes encoding the other enzymes involved in thiopurine metabolism.

In this study, we identify genes known to be involved in thiopurine metabolism and determine all coding mutations in these genes in a cohort of 100 children with IBD.

We assess the joint effect of rare and common variants within the *TPMT* gene and other genes implicated in thiopurine toxicity on TPMT enzyme activity through the application of a gene based statistical test (SKAT-O). The test was also conducted to investigate the association of these variants with thiopurine tolerance and clinical response.

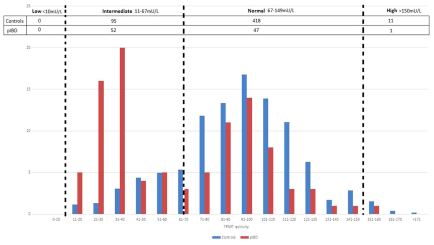


Figure 2. TPMT phenotype frequency distribution for the Wessex paediatric population and the research cohort. Figure shows the TPMT phenotype frequency distribution for 524 paediatric patients, in the Wessex region (524 patients ≤ 18 years, between Dec 2010–April 2015, includes patients with/without IBD) and the TPMT phenotype distribution in our research cohort. Within our cohort we observe a statistically significant difference between the sub-category with TPMT values between 21–40 units ($p = 0.001$) compared to the Wessex paediatric population. This difference is not observed between the sub-categories with TPMT values between 41 to161 ($p = 0.90$)

Materials and Methods

Study Population. Patients were identified through the paediatric gastroenterology service database based at the University Hospital Southampton (UHS), recruited from outpatient clinics and followed through their treatment. All children were diagnosed using the Porto diagnostic criteria[17] and treated according to British Society of Paediatric Gastroenterology, Hepatology and Nutrition (BSPGHAN) published guidelines[9]. Data for one hundred paediatric patients with TPMT phenotype defined as red blood cell enzyme activity and concurrent exome data were analysed.

Ethical Approval. The study was ethically approved by Southampton and South West Hampshire Research Ethics Committee (09/H0504/125). Informed consent was obtained from all participants before recruitment to the study. All methods were carried out in accordance with the approved and published guidelines.

TPMT Phenotype Determination. TPMT enzyme activity was measured using standard high performance liquid chromatographic technique[20]. TPMT enzyme activity level groups were defined as previously described (Fig. 2)[20,21].

Use of Thiopurines and Monitoring for Adverse Effects. British Society of Paediatric Gastroenterology, Hepatology and Nutrition (BSPGHAN) recommend initiation of treatment with thiopurines for maintenance of remission in individuals who relapse in less than 6 months, have 2 or more relapses per year following initial successful therapy and in all steroid-dependent patients. Practice with regards to initiation of treatment varies among clinicians, but is usually commenced and monitored as per the BSPGHAN guidelines[9,19], with regular blood tests to monitor adverse effects such as bone marrow suppression, pancreatitis, hepatotoxicity and patients are clinically followed up to assess progress through treatment. For this study, bone marrow suppression was defined as leucopoenia (WBC < 3000 mm^{-3}) and/or thrombocytopenia (platelets < 100,000 mm^{-3}); liver toxicity was alanine transaminase (ALT), gamma-glutanyl transpeptidase (GGT) or alkaline phosphatase more than twice their normal levels; acute pancreatitis was defined as significant abdominal pain within 3 months of starting thiopurines, accompanied by a serum amylase or lipase level of greater than twice their normal levels as per our local laboratory values[22].

Evaluation of response to therapy. Assessment of clinical poor response or non-response was based on one of the following: (1) Inability to achieve clinical improvement as assessed by global clinical assessment after at least 6 months of thiopurine therapy; (2) Corticosteroid dependence after at least 3 months of thiopurine therapy; (3) Relapse within 6 months of therapy; (4) Use of biologics within 6 months of therapy with thiopurines and; (5) Disease progression needing surgery within 6 months of thiopurine therapy commencement.

Intolerance to treatment was based on all of the following: (1) Occurrence of adverse side effects (including bone marrow suppression, severe gastric intolerance, pancreatitis, hepatotoxicity, skin reactions, flu-like symptoms); (2) Adverse effects unexplained by disease course or other concomitant co-morbidities; and (3) Partial or complete resolution of the observed adverse effects following discontinuation of therapy.

Clinical Category		Number of patients within each group	Median duration of follow-up in months	Males (%)	Disease			TPMT biochemical activity Number of patients across groups (%)			
					CD (%)	UC (%)	IBDU (%)	Low activity (<10 nU/L)	Intermediate activity (10–67 nU/L)	Normal activity (68–150 nU/L)	High activity (>150 nU/L)
Not Treated with thiopurines		22	60 (10–156)	15 (68)	13 (59)	6 (27)	3 (14)	0 (0)	12 (54)	10 (46)	0
Intolerant		14	42 (9–82)	10(71)	12 (86)	2 (14)	0	0 (0)	8 (57)	6(43)	0
Tolerant	Responders	51	52 (7–124)	24 (47)	31 (60)	13 (26)	7 (14)	0 (0)	22 (43)	28 (55)	1 (2)
	Non-responders	13	63 (13–126)	7 (54)	11 (85)	2 (15)	0	0 (0)	10 (77)	3 (23)	0
	Total	100	54 (7–156)	56 (56)	67 (67)	23 (23)	10 (10)	0 (0)	52 (52)	47 (47)	1 (1)

Table 1. A summary of the key clinical and biochemical (TPMT) features of the cohort. Of the 100 patients within the cohort, 67 individuals had Crohn's disease (CD), 23 had ulcerative colitis (UC) and 10 had inflammatory bowel disease unclassified (IBDU). The proportion of males was 56% and the median duration of follow up was 54 months.

Genes Implicated in Thiopurine Toxicity. A systematic search was conducted through ovidsp using MEDLINE and EMBASE from inception to the end of March 2015. Only studies in humans describing *TPMT* genetic variants and other genes involved in thiopurine metabolism and toxicity were included. We applied the *TPMT* nomenclature committee website (http://www.imh.liu.se/tpmtalleles) that outlines all reported *TPMT* variants[23], to cross-reference against variants identified in this study.

Whole Exome Sequencing and Data Analysis

DNA Extraction. Genomic DNA was extracted from peripheral venous blood samples collected in EDTA, using the salting out method as previously described[24,25].

Exome Sequencing. Whole-exome capture was performed using Agilent SureSelect Human all Exon 51 Mb (versions 4 and 5) capture kit as previously described[25]. A bespoke script was used to assign individual variants as "novel" if they were not previously reported in the dbSNP137 databases[26], 1000 Genomes Project (1 KG)[27], the Exome Variant Server (EVS) of European Americans of the NHLI-ESP project with 6500 exomes [http://evs.gs.washington.edu/EVS/], in 46 unrelated human subjects sequenced by Complete Genomics[28] or in the Southampton database of reference exomes.

Burden of Mutation Testing. The burden of genetic variations within the genes was conducted using a gene-based statistical test (the sequence kernel association optimal unified test- SKAT-O)[29]. To conduct the test, a group file of non-synonymous, synonymous, splicing, frameshifts and non-frameshift, stop gain and stop loss mutations was created for each of the genes analysed. SKAT-O was executed with the small sample adjustment, by applying MAF threshold of 0.05 to define rare variations within the whole cohort, and using default weights. The EPACTS software package[30] was use to perform this test. The test was conducted to assess the impact of variations on TPMT enzyme activity, drug tolerance and clinical response.

Results

Clinical Data. Of the 100 patients, 78 initiated thiopurines as part of their clinical management while 22 maintained remission without recourse to this therapy. The median duration of follow up for the 78 patients commenced on thiopurines was 46 months (7–156) from diagnosis and 43 months (6–119) from starting therapy with thiopurines (Table 1).

Genes identified in TMPT metabolism and toxicity. A systematic search of genes implicated in TPMT metabolism and toxicity identified 15 genes with robust evidence from a review of over 3,000 articles (Fig. 1 and Supplementary Table S1) for which we could ascertain variation from exome data. Variation in *TMPT* and these genes was included in downstream analysis of: (1) biochemically assessed TMPT activity; (2) tolerance to thiopurines and; (3) response to treatment.

Biochemically measured TPMT activity in red blood cells. The TPMT enzyme activity was unexpectedly bimodal in distribution across the 100 patients examined in this study. We compared the distribution in our pIBD cohort against the same measure for an independent control group of individuals with various clinical diagnoses, aged ≤18 years from the Wessex region for whom TPMT activity was assessed by the same laboratory over the same period and observed a statistically significant distribution between the groups ($P = 7.69 \times 10^{-12}$) (Fig. 2). While approximately half of our cohort samples had biochemical activity levels within the normal range and half in the intermediate range, almost 80% of control samples had biochemical activity levels falling within the normal range. We believe this may reflect an ascertainment bias in our study cohort due to the fact that exome sequencing was preferentially conducted on children with most severe disease at earliest onset. This will have enriched for children with poor response to first-line treatments.

***TPMT* gene variants - their correlation with biochemically measured TPMT enzyme activity, thiopurine tolerance and response.** We identified five *TPMT* variants across our cohort (Table 2). These included two non-synonymous variants (A154T and Y240C) previously known to impact TMPT function that were found to co-segregate in 9 individuals. The mean biochemical measurement of TMPT activity for this group was 36 mU/L (range 16–56), which is significantly different ($p = 0.003$) from that found in the 91 individuals

Clinical Category		Number of patients within each group	Median duration of follow-up in months	Males (%)	Disease			TPMT biochemical activity Number of patients across groups (%)			
					CD (%)	UC (%)	IBDU (%)	Low activity (<10 mU/L)	Intermediate activity (10–67 mU/L)	Normal activity (68–150 mU/L)	High activity (>150 mU/L)
Not Treated with thiopurines		22	60 (10–156)	15 (68)	13 (59)	6 (27)	3 (14)	0 (0)	12 (54)	10 (46)	0
Intolerant		14	42 (9–82)	10(71)	12 (86)	2 (14)	0	0 (0)	8 (57)	6(43)	0
Tolerant	Responders	51	52 (7–124)	24 (47)	31 (60)	13 (26)	7 (14)	0 (0)	22 (43)	28 (55)	1 (2)
	Non-responders	13	63 (13–126)	7 (54)	11 (85)	2 (15)	0	0 (0)	10 (77)	3 (23)	0
	Total	100	54 (7–156)	56 (56)	67 (67)	23 (23)	10 (10)	0 (0)	52 (52)	47 (47)	1 (1)

Table 1. A summary of the key clinical and biochemical (TPMT) features of the cohort. Of the 100 patients within the cohort, 67 individuals had Crohn's disease (CD), 23 had ulcerative colitis (UC) and 10 had inflammatory bowel disease unclassified (IBDU). The proportion of males was 56% and the median duration of follow up was 54 months.

Genes Implicated in Thiopurine Toxicity. A systematic search was conducted through ovidsp using MEDLINE and EMBASE from inception to the end of March 2015. Only studies in humans describing *TPMT* genetic variants and other genes involved in thiopurine metabolism and toxicity were included. We applied the *TPMT* nomenclature committee website (http://www.imh.liu.se/tpmtalleles) that outlines all reported *TPMT* variants[23], to cross-reference against variants identified in this study.

Whole Exome Sequencing and Data Analysis

DNA Extraction. Genomic DNA was extracted from peripheral venous blood samples collected in EDTA, using the salting out method as previously described[24,25].

Exome Sequencing. Whole-exome capture was performed using Agilent SureSelect Human all Exon 51 Mb (versions 4 and 5) capture kit as previously described[25]. A bespoke script was used to assign individual variants as "novel" if they were not previously reported in the dbSNP137 databases[26], 1000 Genomes Project (1 KG)[27], the Exome Variant Server (EVS) of European Americans of the NHLI-ESP project with 6500 exomes [http://evs.gs.washington.edu/EVS/], in 46 unrelated human subjects sequenced by Complete Genomics[28] or in the Southampton database of reference exomes.

Burden of Mutation Testing. The burden of genetic variations within the genes was conducted using a gene-based statistical test (the sequence kernel association optimal unified test- SKAT-O)[29]. To conduct the test, a group file of non-synonymous, synonymous, splicing, frameshifts and non-frameshift, stop gain and stop loss mutations was created for each of the genes analysed. SKAT-O was executed with the small sample adjustment, by applying MAF threshold of 0.05 to define rare variations within the whole cohort, and using default weights. The EPACTS software package[30] was use to perform this test. The test was conducted to assess the impact of variations on TPMT enzyme activity, drug tolerance and clinical response.

Results

Clinical Data. Of the 100 patients, 78 initiated thiopurines as part of their clinical management while 22 maintained remission without recourse to this therapy. The median duration of follow up for the 78 patients commenced on thiopurines was 46 months (7–156) from diagnosis and 43 months (6–119) from starting therapy with thiopurines (Table 1).

Genes identified in TMPT metabolism and toxicity. A systematic search of genes implicated in TPMT metabolism and toxicity identified 15 genes with robust evidence from a review of over 3,000 articles (Fig. 1 and Supplementary Table S1) for which we could ascertain variation from exome data. Variation in *TMPT* and these genes was included in downstream analysis of: (1) biochemically assessed TMPT activity; (2) tolerance to thiopurines and; (3) response to treatment.

Biochemically measured TPMT activity in red blood cells. The TPMT enzyme activity was unexpectedly bimodal in distribution across the 100 patients examined in this study. We compared the distribution in our pIBD cohort against the same measure for an independent control group of individuals with various clinical diagnoses, aged ≤18 years from the Wessex region for whom TPMT activity was assessed by the same laboratory over the same period and observed a statistically significant distribution between the groups ($P = 7.69 \times 10^{-12}$) (Fig. 2). While approximately half of our cohort samples had biochemical activity levels within the normal range and half in the intermediate range, almost 80% of control samples had biochemical activity levels falling within the normal range. We believe this may reflect an ascertainment bias in our study cohort due to the fact that exome sequencing was preferentially conducted on children with most severe disease at earliest onset. This will have enriched for children with poor response to first-line treatments.

***TPMT* gene variants - their correlation with biochemically measured TPMT enzyme activity, thiopurine tolerance and response.** We identified five *TPMT* variants across our cohort (Table 2). These included two non-synonymous variants (A154T and Y240C) previously known to impact TMPT function that were found to co-segregate in 9 individuals. The mean biochemical measurement of TMPT activity for this group was 36 mU/L (range 16–56), which is significantly different ($p = 0.003$) from that found in the 91 individuals

Chr	Lbp	Rbp	Gene	Total number of samples	Fraction of individuals who carry rare variants under the MAF thresholds (MAF < 0.05)	Number of all variants defined in the group file	Number of variants defined as rare (MAF < 0.05)*
Biochemical activity (52 with TPMT value <67 and 48 with TPMT value >67)							
18	33767568	33848581	MOCOS	100	0.12000	12	5
6	18130918	18148069	TPMT	100	0.10000	5	2
Tolerance (14 intolerant and 64 tolerant)							
3	155588592	155654236	GMPS	78	0.12821	5	5
6	18130918	18148069	TPMT	78	0.01282	5	1
Responses (51 Responders and 13 non-responders)							
5	55231311	55272085	IL6ST	64	0.10937	8	3
13	95696540	95953517	ABCC4	64	0.31250	22	9

Table 3. SKAT-O test association analysis across TPMPT and other genes involved in thiopurine toxicity. The SKAT-O test was applied to assess the joint effect of common, rare and low frequency variants within the genes implicated in thiopurine toxicity (only significant genes are shown) on TPMT enzyme activity, tolerance and response to the drug. *These variants received different weights in the SKAT-O joint test. Genes are ordered by p-value.

Gene	Chr	Position on hg19	Variant type	Coding change	Protein change	Novel	Phylop,	1- sift	Polyphen 2	Mutationtaster	Gerp++	dbSNP	Frequency in 1000 genome	Frequency in EVS	TPMT biochemical activity	int	int	int	int	int	int	int	int	Normal	Normal	Normal	Normal	Normal	Normal
															TPMT activity	16	21	47	52	55	18	26	30	80	83	93	113	114	128
															Diagnosis	CD	CD	CD	CD	CD	CD	UC	CD	CD	CD	CD	CD	UC	CD
															Gender	M	M	M	M	F	F	M	F	F	M	M	M	M	M
															IDs	1	3	6	7	8	2	4	5	9	10	11	12	13	14
TPMT	6	18130918	ns	c.719A>G	p.Y240C	.	0.998258	0.94	0.94	0.999973	5.13	rs1142345	0.05	0.041715		1	1	1	1	0	0	0	0	0	0	0	0	0	0
TPMT	6	18148069	ns	c.218C>T	p.A73V	NOVEL	0.999474	0.95	0.898	0.999898	4.98	.	.	.		0	0	0	0	1	0	0	0	0	0	0	0	0	0
MOCOS	18	33779705	ns	c.G359A	p.S120N	.	0.961984	0.98	0.284	0.013289	2.79	rs3744900	0.06	0.047093		0	0	0	0	1	0	0	0	0	0	1	0	0	0
MOCOS	18	33831189	ns	c.C2107A	p.H703N	.	0.998524	0.58	0.706	0.010571	4.98	rs594445	0.25	0.283953		1	0	0	0	0	1	2	0	1	0	1	2	1	0
MOCOS	18	33848581	ns	c.T2600C	p.V867A	.	0.998597	1	0.063	0.006647	5.69	rs1057251	0.08	0.116279		1	0	1	0	0	1	0	1	0	0	0	0	0	0
XDH	2	31572983	ns	c.G2738A	p.R913Q	NOVEL	0.999016	1	1	1	5.38	.	.	.		0	0	0	0	0	1	0	0	0	0	0	0	0	0
XDH	2	31590917	ns	c.A2107G	p.I703V	.	0.998578	1	0.336	0.999913	4.52	rs17011368	0.05	0.034186		0	0	0	0	1	0	0	0	0	0	0	0	0	0
XDH	2	31611143	ns	c.G514A	p.G172R	.	0.998995	0.99	0.004	0.99994	5.35	rs45523133	0.04	0.026047		0	1	0	0	0	0	0	0	0	0	0	0	0	0
XDH	2	31621523	ns	c.A349T	p.T117S	.	0.998756	0.96	0.997	0.999999	5.81	.	.	.		0	0	0	0	0	0	0	0	0	0	0	2	0	0
AOX1	2	2.02E+08	ns	c.A3404G	p.N1135S	.	0.996285	0	0	2.70E-05	5.3	rs55754655	0.11	0.129767		0	0	0	1	0	1	0	0	1	1	0	0	0	1
IL6ST	5	55264153	ns	c.G442C	p.G148R	.	0.995114	0.98	0	2.70E-05	5.6	rs2228044	0.19	0.119535		0	0	1	1	0	0	0	0	0	1	0	0	0	0
NUDT15	13	48619855	ns	c.C415T	p.R139C	.	0.998597	0.92	0.057	0.937718	5.04	rs116855232	0.04	0.002442		0	0	1	0	0	0	0	0	0	0	0	0	0	0
ITPA	20	3193842	ns	c.C94A	p.P32T	.	0.998747	0.87	0.131	0.996149	5.14	rs1127354	0.08	0.067674		0	0	0	0	0	0	0	1	0	0	0	2	0	0
PACSIN2	22	43280404	ns	c.C773A	p.S258Y	.	0.999724	1	0.998	0.999962	5.16	rs200427054	0.0005	0.00314		0	0	0	0	0	0	0	1	0	0	0	0	0	0

Table 4. Deleterious variants occurring within the group of individuals with intolerance to thiopurines. Fourteen out of the 100 patients were intolerant to thiopurines. Five of the fourteen individuals had deleterious *TPMT* variants; there was enrichment for deleterious variants within the *MOCOS* gene and the *AOX1* gene in the other 9 individuals. Deleterious variants included: frameshift indels, stopgain/loss, splicing with Maxent score >3 and nonsynonymous variants with a gerp score >2. (ns- non-synonymous; 1 and 2 indicate heterozygous and homozygous genotype respectively).

Discussion

It is well established that in a small percentage of patients, *TPMT* genotyping alone or in combination with TPMT phenotype is insufficient to predict tolerance to thiopurine drugs[2]. Several *TPMT* variants associated with deficient enzyme activity have been described[23], however an appreciable subset of patients with intermediate to low activity do not harbour known risk alleles[32]. Our study identified nine individuals with the known *TPMT* mutations (9% compared to 11% reported in previous studies)[31]. All nine individuals had TPMT enzyme activity levels in the intermediate range in line with expectations; mean TPMT value across this group was significantly different compared to individuals without these mutations (36.6 mU/L compared to 68.1 mU/L, $p = 0.003$). However 43% of individuals with TPMT activity in the intermediate range did not have the known *TPMT* mutations. Our results indicate that although prediction of thiopurine toxicity through NGS has a higher specificity compared to the biochemical test, the sensitivity of both methods is clinically suboptimal (35.7% and 57.14% respectively). Among the 14 individuals intolerant to thiopurines, all the 5 patients who harboured deleterious *TPMT*

		Intolerant	Tolerant	Sensitivity	Specificity
Biochemical test	+	8	32	57.14%	50%
	−	6	32		
Deleterious *TPMT* variants	+	5	4	35.7	93.75%
Non-deleterious variants	−	9	60		

Table 5. Specificity and sensitivity for drug intolerance and tolerance. The specificity for predicting toxicity through the biochemical test and through application of TPMT genetic variants was 50% and 93.75% respectively. The sensitivity obtained through both methods was sub-optimal.

Patient ID	Diagnosis	Age at Diagnosis (years)	Gender	*TPMT* Gene Deleterious Variants	TPMT Biochemical Activity	TPMT Value (mU/L)	Undergoing treatment with 5-aminosalicylic drugs at time of test	Follow up (months)	Thiopurine Drug	Median Dose (mg/Kg)	Adverse Effects
6	CD	15	M	Yes	Intermediate	21	No	14	AZA	1.5	Leuco-encephalopathy
10	CD	13.5	F	No	Intermediate	18	No	42	AZA	1.5	Neutropenia
22	UC	10.5	M	No	Intermediate	26	Yes	82	AZA	2	Elevated amylase
23	CD	14.5	F	No	Intermediate	30	Yes	63	AZA	2	Persistent Nausea
28	CD	5	M	No	Normal	93	Yes	32	AZA	1	Persistent Nausea
39	CD	10.9	M	Yes	Intermediate	16	No	75	AZA	1.5	Pancytopenia
40	CD	9.5	M	Yes	Intermediate	52	No	75	6-MP	1	Persistent Nausea
41	UC	11.7	M	No	Normal	114	Yes	47	AZA	1	Abnormal ALT
69	CD	16	F	No	Normal	80	No	32	AZA	2.5	Persistent Nausea
73	CD	14	M	No	Normal	128	No	23	AZA	2	Persistent Nausea
74	CD	13.5	F	Yes	Intermediate	55	No	45	AZA	1	Persistent Nausea
88	CD	11	M	No	Normal	83	No	25	AZA	2	Persistent Nausea
92	CD	12	M	Yes	Intermediate	47	No	27	AZA	1.5	Abnormal ALT
95	CD	15.5	M	No	Normal	113	No	9	6-MP	1.5	Pancreatitis

Table 6. The group of individuals with intolerance to thiopurines. The drugs used were azathioprine (AZA) in 12 patients (dose range 1–2.5 mg/kg/day) and 6-mercaptopurine (6-MP) in 2 patients (dose range 1–1.5 mg/kg/day). The choice of drug was based on clinician preference.

variants would also have been predicted as potentially intolerant to thiopurines through the biochemical test. Furthermore, all the nine individuals with deleterious variants within the *TPMT* gene across the cohort of 100 patients would also have been identified as potentially intolerant as all these nine individuals had TPMT enzyme activity levels within the intermediate range. Hence based on first principles, NGS did not have a clear advantage over the biochemical test in predicting thiopurine toxicity.

We identified a highly pathogenic novel variant in *TPMT* in an individual who had TPMT enzyme activity level of 55 mU/L, but developed severe gastrointestinal toxicity despite a reduced dose. This suggests that thiopurine toxicity can develop in individuals harbouring rare or yet unknown variants, not detected through standard genotyping. As a widely used practical approach, *TPMT* genotyping of known variants is considered only when biochemical tests suggest a deficiency or if a patient has recently been transfused. However, a normal genotype for known variants cannot exclude the possibility of rare variation causing TPMT deficiency and development of adverse effects[2]. Exome sequencing therefore can be a powerful tool in identifying individuals at risk of toxicity who could have been missed through standard *TPMT* genotyping.

The SKAT-O test identified a significant association between variations within the *MOCOS* gene and TPMT enzyme activity ($p = 0.0015$). Molybdenum cofactor sulfurase (*MOCOS*) is a protein-coding gene located on 18q12, which sulfurates the molybdenum cofactor in XDH and AOX1, key enzymes involved in the degradation of thiopurines[33]. Previous studies have suggested a role for *MOCOS* gene in thiopurine metabolism with possible impact on clinical outcomes in patients with mutations, however an association between *MOCOS* and TPMT enzyme activity has not been explored. This is the first study to identify a significant role for this gene with variations causing alterations in biochemical enzyme activity. Further work is required to determine how *MOCOS* influences *TPMT* function.

We also detected a nominal association between *GMPS* and drug tolerance ($p = 0.0212$). *GMPS* is involved in the phosphorylation of 6-TIMP (6-thioinosine monophosphate) to thioguanine nucleotides, which is an important step for thiopurines to exert their cytotoxic effects. Further mechanistic studies will be required to elucidate the molecular mechanisms and clearly define the role of these genes involved in the thiopurine metabolic pathway.

A limitation of this study is that only children who had undergone exome analysis with concurrent TPMT values were selected for the study. While approximately half of our cohort patients had TPMT activity in the intermediate range, only 20% of control samples had TPMT activity in the intermediate category. Exome sequencing was preferentially conducted on children with the most severe disease phenotype, which would have enriched for

patients with poor response to first-line treatments. Secondly, the selected population is predominantly Caucasian and a relatively small cohort, thereby limiting the general applicability of the results. Following a systematic search of literature to identify genes implicated in thiopurine toxicity, a panel of 15 genes was prioritised for assessment. It is it is possible that potentially pathogenic variants in other genes may have been missed through this approach. However extension of the list of genes examined would compromise power in the highest priority candidate genes.

Although there is no clear advantage of NGS over the biochemical test in predicting toxicity, our study demonstrates the strength of NGS as a powerful tool in identifying pathogenic variants in patients not detected through standard genotyping. As high throughput sequencing becomes more accessible and affordable, a more objective approach to assessing pharmaco-genomic variation in all genes involved in the drug metabolism pathway may be indicated in selected patients to guide treatment strategies. Replication in larger studies would be required for a comprehensive curation of candidate genes, possibly paving the way for a targeted gene panel as a reliable predictor of toxicity.

References

1. Present, D. H. *et al.* Treatment of Crohn's disease with 6-mercaptopurine. A long-term, randomized, double-blind study. *The New England journal of medicine* **302,** 981–987, doi: 10.1056/NEJM198005013021801 (1980).
2. Benkov, K. *et al.* Role of thiopurine metabolite testing and thiopurine methyltransferase determination in pediatric IBD. *Journal of pediatric gastroenterology and nutrition* **56,** 333–340, doi: 10.1097/MPG.0b013e3182844705 (2013).
3. Liu, Y. P. *et al.* Association between Thiopurine S-methyltransferase Polymorphisms and Thiopurine-Induced Adverse Drug Reactions in Patients with Inflammatory Bowel Disease: A Meta-Analysis. *PloS one* **10,** e0121745, doi: 10.1371/journal.pone.0121745 (2015).
4. Tai, H. L. *et al.* Thiopurine S-methyltransferase deficiency: two nucleotide transitions define the most prevalent mutant allele associated with loss of catalytic activity in Caucasians. *American journal of human genetics* **58,** 694–702 (1996).
5. Hamdan-Khalil, R. *et al.* Identification and functional analysis of two rare allelic variants of the thiopurine S-methyltransferase gene, TPMT*16 and TPMT*19. *Biochemical pharmacology* **69,** 525–529, doi: 10.1016/j.bcp.2004.10.011 (2005).
6. Haglund, S. Interindividual differences in thiopurine metabolism. *Linkoping University Medical Dissertations* (2011).
7. Weinshilboum, R. M. & Sladek, S. L. Mercaptopurine pharmacogenetics: monogenic inheritance of erythrocyte thiopurine methyltransferase activity. *American journal of human genetics* **32,** 651–662 (1980).
8. Relling, M. V. *et al.* Clinical Pharmacogenetics Implementation Consortium guidelines for thiopurine methyltransferase genotype and thiopurine dosing. *Clinical pharmacology and therapeutics* **89,** 387–391, doi: 10.1038/clpt.2010.320 (2011).
9. Sandhu B. K., Fell, J., Beattie, R. M. & Mitton, S. G. Guidelines for the Management of Inflammatory Bowel Disease in Children in the United Kingdom. *J Pediatr Gastroenterol Nutr. 2010 Feb;50 Suppl 1*, S1–13 (2010).
10. Relling, M. V. *et al.* Clinical pharmacogenetics implementation consortium guidelines for thiopurine methyltransferase genotype and thiopurine dosing: 2013 update. *Clinical pharmacology and therapeutics* **93,** 324–325, doi: 10.1038/clpt.2013.4 (2013).
11. Palmieri, O. *et al.* Sequential evaluation of thiopurine methyltransferase, inosine triphosphate pyrophosphatase, and HPRT1 genes polymorphisms to explain thiopurines' toxicity and efficacy. *Alimentary pharmacology & therapeutics* **26,** 737–745, doi: 10.1111/j.1365-2036.2007.03421.x (2007).
12. Farfan, M. J. *et al.* Prevalence of TPMT and ITPA gene polymorphisms and effect on mercaptopurine dosage in Chilean children with acute lymphoblastic leukemia. *BMC cancer* **14,** 299, doi: 10.1186/1471-2407-14-299 (2014).
13. Kudo, M. *et al.* Genetic variations in the HGPRT, ITPA, IMPDH1, IMPDH2, and GMPS genes in Japanese individuals. *Drug metabolism and pharmacokinetics* **24,** 557–564 (2009).
14. Kurzawski, M., Dziewanowski, K., Safranow, K. & Drozdzik, M. Polymorphism of genes involved in purine metabolism (XDH, AOX1, MOCOS) in kidney transplant recipients receiving azathioprine. *Therapeutic drug monitoring* **34,** 266–274, doi: 10.1097/FTD.0b013e31824aa681 (2012).
15. Stocco, G. *et al.* PACSIN2 polymorphism influences TPMT activity and mercaptopurine-related gastrointestinal toxicity. *Human molecular genetics* **21,** 4793–4804, doi: 10.1093/hmg/dds302 (2012).
16. Ban, H. *et al.* The multidrug-resistance protein 4 polymorphism is a new factor accounting for thiopurine sensitivity in Japanese patients with inflammatory bowel disease. *Journal of gastroenterology* **45,** 1014–1021, doi: 10.1007/s00535-010-0248-y (2010).
17. Yang, S. K. *et al.* A common missense variant in NUDT15 confers susceptibility to thiopurine-induced leukopenia. *Nature genetics* **46,** 1017–1020, doi: 10.1038/ng.3060 (2014).
18. Karas-Kuzelicki, N., Jazbec, J., Milek, M. & Mlinaric-Rascan, I. Heterozygosity at the TPMT gene locus, augmented by mutated MTHFR gene, predisposes to 6-MP related toxicities in childhood ALL patients. *Leukemia* **23,** 971–974, doi: 10.1038/leu.2008.317 (2009).
19. Ibd Working Group of the European Society for Paediatric Gastroenterology, H. & Nutrition. Inflammatory bowel disease in children and adolescents: recommendations for diagnosis–the Porto criteria. *Journal of pediatric gastroenterology and nutrition* **41,** 1–7 (2005).
20. Ford, L. T. & Berg, J. D. Determination of thiopurine S-methyltransferase activity in erythrocytes using 6-thioguanine as substrate and a non-extraction liquid chromatographic technique. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **798,** 111–115 (2003).
21. Ford, L., Graham, V. & Berg, J. Whole-blood thiopurine S-methyltransferase activity with genotype concordance: a new, simplified phenotyping assay. *Annals of clinical biochemistry* **43,** 354–360 (2006).
22. Heap, G. A. *et al.* HLA-DQA1-HLA-DRB1 variants confer susceptibility to pancreatitis induced by thiopurine immunosuppressants. *Nature genetics* **46,** 1131–1134, doi: 10.1038/ng.3093 (2014).
23. Appell, M. L. *et al.* Nomenclature for alleles of the thiopurine methyltransferase gene. *Pharmacogenetics and genomics* **23,** 242–248, doi: 10.1097/FPC.0b013e32835f1cc0 (2013).
24. Miller, S. A., Dykes, D. D. & Polesky, H. F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic acids research* **16,** 1215 (1988).
25. Andreoletti, G. *et al.* Exome analysis of patients with concurrent pediatric inflammatory bowel disease and autoimmune disease. *Inflammatory bowel diseases* **21,** 1229–1236, doi: 10.1097/MIB.0000000000000381 (2015).
26. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic acids research* **29,** 308–311 (2001).
27. Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65, doi: 10.1038/nature11632 (2012).
28. Drmanac, R. *et al.* Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* **327,** 78–81, doi: 10.1126/science.1181498 (2010).
29. Ionita-Laza, I. e. a. "Sequence Kernel Association Tests for the Combined Effect of Rare and Common Variants." American journal of human genetics 92.6 (2013): 841–53. Web. 13 Jan. 2014.

Inflamm Bowel Dis. 2015 Jun;21(6):1229-36. doi: 10.1097/MIB.0000000000000381.

Exome analysis of patients with concurrent pediatric inflammatory bowel disease and autoimmune disease.

Andreoletti G[1], Ashton JJ, Coelho T, Willis C, Haggarty R, Gibson J, Holloway J, Batra A, Afzal NA, Beattie RM, Ennis S.

Author information

Abstract

BACKGROUND: Pediatric Inflammatory Bowel Disease (PIBD) is a chronic condition seen in genetically predisposed individuals. Genome-wide association studies have implicated >160 genomic loci in IBD with many genes coding for proteins in key immune pathways. This study looks at autoimmune disease burden in patients diagnosed with PIBD and interrogates exome data of a subset of patients.

METHODS: Patients were recruited from the Southampton Genetics of PIBD cohort. Clinical diagnosis of autoimmune disease in these individuals was ascertained from medical records. For a subset of patients with PIBD and concurrent asthma, exome data was interrogated to ascertain the burden of pathogenic variants within genes implicated in asthma. Association testing was conducted between cases and population controls using the SKAT-O test.

RESULTS: Forty-nine (28.3%) PIBD children (18.49% CD, 8.6% UC, and 21.15% IBDU patients) had a concurrent clinical diagnosis of at least one other autoimmune disorder; asthma was the most prevalent, affecting 16.2% of the PIBD cohort. Rare and common variant association testing revealed 6 significant genes ($P < 0.05$) before Bonferroni adjustment. Three of these genes were previously implicated in both asthma and IBD (ZPBP2 IL1R1, and IL18R1) and 3 in asthma only (PYHIN1, IL2RB, and GSTP1).

CONCLUSIONS: One-third of our cohort had a concurrent autoimmune condition. We observed higher incidence of asthma compared with the overall pediatric prevalence. Despite a small sample size, SKAT-O evaluated a significant burden of rare and common mutations in 6 genes. Variant burden suggests that a systemic immune dysregulation rather than organ-specific could underpin immune dysfunction for a subset of patients.

Acta Paediatr. 2015 Aug;104(8):831-7. doi: 10.1111/apa.13017. Epub 2015 May 20.

Presenting phenotype of paediatric inflammatory bowel disease in Wessex, Southern England 2010-2013.

Ashton JJ[1], Coelho T[1], Ennis S[2], Batra A[1], Afzal NA[1], Beattie RM[1].

Author information

Abstract

AIM: There has been at least a twofold increase in the incidence of paediatric inflammatory bowel disease (PIBD) over the last 20 years; we report the presenting features from 2010 to 2013 and compare with previous data.

METHODS: All patients diagnosed with PIBD at University Hospitals Southampton from 2010 to 2013 were identified from an in-house database. Data were obtained from paper and electronic notes. Height, weight and BMI SDS are presented as median values (95% CI).

RESULTS: One hundred and seventy-two patients were included (median age at diagnosis 13.5, 115 male); Crohn's disease (CD) - 107, UC - 50, inflammatory bowel disease unclassified (IBDU) - 15. The most common presenting features of CD were abdominal pain (86%), diarrhoea (78.5%) and weight loss (56.1%); 42.1% of patients had all three. In UC blood in stool (92%), diarrhoea (92%) and abdominal pain (88%) were the most common; all three in 76% of patients. CD presented with ileocolonic disease in 52.5%. UC presented with pancolitis in 64%. There was growth delay in CD: height -0.37 (-0.60 to -0.14); weight -1.09 (-1.35 to -0.83). Growth was maintained in UC: height 0.53 (0.19 to 0.87); weight 0.14 (-0.20 to 0.48).

CONCLUSION: Paediatric inflammatory bowel disease phenotype remains as extensive despite increasing incidence. Although the classical phenotype is common, a reasonable proportion present with atypical features, normal growth and normal blood markers.

J Pediatr Gastroenterol Nutr. 2016 Feb;62(2):246-51. doi: 10.1097/MPG.0000000000001032.

Endoscopic Versus Histological Disease Extent at Presentation of Paediatric Inflammatory Bowel Disease.

Ashton JJ[1], Coelho T, Ennis S, Vadgama B, Batra A, Afzal NA, Beattie RM.

Author information

Abstract

OBJECTIVES: The Paris classification (PC) of paediatric inflammatory bowel disease categorises disease extent and therefore affects treatment decisions. Histological (microscopic) disease extent is not incorporated, and endoscopic (macroscopic) findings may underrepresent disease extent when compared with histological findings; this study compares disease extent at presentation.

METHODS: Data were obtained of patients <17 years of age diagnosed with inflammatory bowel disease from 2010 to 2013 at University Hospital Southampton. Data are presented as percentage of patients undergoing endoscopy. PC was performed alongside a modified PC by histological disease location.

RESULTS: A total of 172 patients were identified (median age at diagnosis 13.5 years, 115 boys); Crohn disease (CD) 107, ulcerative colitis (UC) 50, inflammatory bowel disease unclassified (IBDU) 15; 159 had undergone upper gastrointestinal (GI) endoscopy, 163 had undergone lower GI endoscopy. Histological disease was more extensive at all points for CD, UC, and IBDU. CD--endoscopic ileal disease in 49% of patients compared with histological disease in 71.3%. Comparing PC--a 10% increase in L3 disease (ileocolonic), a 24% increase in L3+L4a disease (ileocolonic plus upper GI), and a 27% increase in all of the upper GI involvement if histological disease extent was used. UC--the most common disease location was the rectum (endoscopic 91.5% vs histological 93.6%) and descending colon (endoscopic 89.4% vs histological 95.7%). Comparing PC--a 19% increase in E4 disease (pancolitis) if histological disease extent was used.

CONCLUSIONS: These data confirm that histological disease extent is greater than endoscopic disease extent. This should be considered when the PC is used. Further study is needed to elucidate which classification would better predict disease outcome.

Inflamm Bowel Dis. 2016 Oct;22(10):2317-27. doi: 10.1097/MIB.0000000000000890.

Identification of Variants in Genes Associated with Single-gene Inflammatory Bowel Disease by Whole-exome Sequencing.

Ashton JJ[1], Andreoletti G, Coelho T, Haggarty R, Batra A, Afzal NA, Beattie RM, Ennis S.

Author information

Abstract

BACKGROUND: Most cases of inflammatory bowel disease (IBD) are caused by complex host-environment interaction. There are a number of conditions associated with a single-gene mutation, most cases are very early onset (aged < 6 yr), present with a unique form of disease and often have atypical features.

METHODS: Whole-exome data for 147 pediatric patients with IBD were interrogated for a panel of 51 genes associated with monogenic IBD. Observed variation was categorized according to the American College of Medical Genetics (ACMG) guidelines to identify rare, novel, and known variants that might contribute to IBD.

RESULTS: Five hundred seventy-four variants were identified across 51 genes. These were categorized in line with ACMG guidance to remove benign variants and to identify "pathogenic" and "likely pathogenic" variants. In 6 patients, we observed 6 pathogenic variants of which CYBA(c.287+2T>C), COL7A1(c.6501+1G>C), LIG4(p.R814X), and XIAP(p.T470S) were known causative mutations, and FERMT1(p.R271Q) and SKIV2L(c.354+5G>A) were novel. In the 3 patients with XIAP, SKIV2L, and FERMT1 variants, individuals' disease features resembled the monogenic phenotype. This was despite apparent heterozygous carriage of pathogenic variation for the latter 2 genes. The XIAP variant was observed in a hemizygous male.

CONCLUSIONS: Whole-exome sequencing allows for identification of known and de novo potentially causative mutations in genes associated with monogenic IBD. Although these are rare conditions, it is vital to identify causative mutations early to improve prognosis. We postulate that in a subset of IBD, heterozygous mutations (in genes believed to manifest IBD through autosomal recessive inheritance) may contribute to clinical presentation.

Medicine (Baltimore). 2017 Jun;96(26):e7347. doi: 10.1097/MD.0000000000007347.

16S sequencing and functional analysis of the fecal microbiome during treatment of newly diagnosed pediatric inflammatory bowel disease.

Ashton JJ[1], Colquhoun CM, Cleary DW, Coelho T, Haggarty R, Mulder I, Batra A, Afzal NA, Beattie RM, Scott KP, Ennis S.

Author information

Abstract

The human microbiome is of considerable interest to pediatric inflammatory bowel disease (PIBD) researchers with 1 potential mechanism for disease development being aberrant immune handling of the intestinal bacteria. This study analyses the fecal microbiome through treatment in newly diagnosed PIBD patients and compares to cohabiting siblings where possible. Patients were recruited on clinical suspicion of PIBD before diagnosis. Treatment-naïve fecal samples were collected, with further samples at 2 and 6 weeks into treatment. Samples underwent 16S ribosomal ribonucleic acid (RNA) gene sequencing and short-chain fatty acids (SCFAs) analysis, results were analyzed using quantitative-insights-into-microbial-ecology. Six PIBD patients were included in the cohort: 4 Crohn disease (CD), 1 ulcerative colitis (UC), 1 inflammatory bowel disease (IBD) unclassified, and median age 12.6 (range 10-15.1 years); 3 patients had an unaffected healthy sibling recruited. Microbial diversity (observed species/Chao1/Shannon diversity) was reduced in treatment-naïve patients compared to siblings and patients in remission. Principal coordinate analysis using Bray-Curtis dissimilarity and UniFrac revealed microbial shifts in CD over the treatment course. In treatment-naïve PIBD, there was reduction in functional ability for amino acid metabolism and carbohydrate handling compared to controls ($P = .038$) and patients in remission ($P = .027$). Metabolic function returned to normal after remission was achieved. SCFA revealed consistent detection of lactate in treatment-naïve samples. This study adds in-depth 16S rRNA sequencing analysis on a small longitudinal cohort to the literature and includes sibling controls and patients with UC/IBD unclassified. It highlights the initial dysbiosis, reduced diversity, altered functional potential, and subsequent shifts in bacteria from diagnosis over time to remission.

Sci Rep. 2017 Apr 19;7:46454. doi: 10.1038/srep46454.

Exome Analysis of Rare and Common Variants within the NOD Signaling Pathway.

Andreoletti G[1], Shakhnovich V[2,3], Christenson K[2], Coelho T[1,4], Haggarty R[5], Afzal NA[4], Batra A[4], Petersen BS[6], Mort M[7], Beattie RM[4], Ennis S[1].

Author information

Abstract

Pediatric inflammatory bowel disease (pIBD) is a chronic heterogeneous disorder. This study looks at the burden of common and rare coding mutations within 41 genes comprising the NOD signaling pathway in pIBD patients. 136 pIBD and 106 control samples underwent whole-exome sequencing. We compared the burden of common, rare and private mutation between these two groups using the SKAT-O test. An independent replication cohort of 33 cases and 111 controls was used to validate significant findings. We observed variation in 40 of 41 genes comprising the NOD signaling pathway. Four genes were significantly associated with disease in the discovery cohort (BIRC2 $p = 0.004$, NFKB1 $p = 0.005$, NOD2 $p = 0.029$ and SUGT1 $p = 0.047$). Statistical significance was replicated for BIRC2 ($p = 0.041$) and NOD2 ($p = 0.045$) in an independent validation cohort. A gene based test on the combined discovery and replication cohort confirmed association for BIRC2 ($p = 0.030$). We successfully applied burden of mutation testing that jointly assesses common and rare variants, identifying two previously implicated genes (NFKB1 and NOD2) and confirmed a possible role in disease risk in a previously unreported gene (BIRC2). The identification of this novel gene provides a wider role for the inhibitor of apoptosis gene family in IBD pathogenesis.

Thorax. 2017 Oct;72(10):946-949. doi: 10.1136/thoraxjnl-2016-209397. Epub 2017 Feb 3.

Corticosteroids and infliximab impair the performance of interferon-γ release assays used for diagnosis of latent tuberculosis.

Edwards A[1], Gao Y[2,3], Allan RN[1,4], Ball D[4], de Graaf H[1,4,5], Coelho T[1,6], Clifford V[7], Curtis N[7], Williams A[2,3,8,9], Faust SN[1,4,10,11], Mansour S[1], Marshall B[1,10,12], Elkington P[1,10,12], Tebruegge M[1,7,10,11].

Author information

Abstract

The impact of immunosuppression on interferon-γ release assays and novel cytokine biomarkers of TB infection, mycobacteria-specific IL-2, IP-10 and TNF-α responses was investigated in an ex vivo model. Cytokine responses in standard QuantiFERON-TB Gold in-Tube (QFT-GIT) assays were compared with duplicate assays containing dexamethasone or infliximab. Dexamethasone converted QFT-GIT results from positive to negative in 30% of participants. Antigen-stimulated interferon-γ, IL-2 and TNF-α responses were markedly reduced, but IP-10 responses were preserved. Infliximab caused QFT-GIT result conversion in up to 30% of participants and substantial reductions in all cytokine responses. Therefore, corticosteroids and anti-TNF-α agents significantly impair interferon-γ release assay performance. IP-10 may be a more robust TB biomarker than interferon-γ in patients receiving corticosteroids.

PostScript

Letter

Is the incidence of paediatric inflammatory bowel disease still increasing?

James John Ashton[1, 2], Mick Cullen[1], Nadeem A Afzal[1], Tracy Coelho[1], Akshay Batra[1], R Mark Beattie[1]

Author affiliations +

View Full Text

http://dx.doi.org/10.1136/archdischild-2018-315038