Exploring Genetic Susceptibility: Using a combined systems biology, *in vitro* and *ex vivo* approach to understand the pathology of ulcerative colitis

Dr Johanne Brooks MBChB, Bsc (Hons)

Submitted for the degree of Doctor of Philosophy (PhD)

University of East Anglia

The Quadram Institute

Abstract

The overall aim of this PhD is to use a multidisciplinary approach to determine the function of Ulcerative Colitis (UC) associated SNPs, to help understand the role of SNPs in the pathogenesis of UC in general and in a patient specific context.

UC is a chronic, relapsing inflammatory disease of the large bowel for which the aetiology is thought to be a trifecta of 1) dysregulation of the immune system in response to 2) an environmental trigger in a 3) genetically susceptible host. Genetic susceptibility or susceptibility loci for UC have been identified by Genome Wide Associations Scanning and subsequent fine mapping and deep sequencing.

This work intended to further characterise these susceptibility loci at a global level and a patient specific level using both a systems biology approach and experimental validation of the *in-silico* work. Using publicly available datasets non exonic UC associated SNPs were functionally annotated to regulatory regions within the genome. Exonic SNPs were also analysed looking at impacts in protein linear motifs and splice enhancement motifs. Bioinformatics was used to identify interacting proteins and create a UC-interactome network. This suggested that UC was a disease of fine regulators as opposed to a disease of specific target proteins.

Analysis of the UC-interactome identified the focal adhesion complex (FAC) that is involved in regulating wound healing as major component of the network. One member of the FAC, Leupaxin (LPXN), was identified as a potential target for validation. Using CRISPR-Cas9 technology, *LPXN* overexpressing cell lines and knock out cell lines were created. Wound healing assays and cytokine analysis identified that overexpression of LPXN impaired wound healing and reduced the secretion of MCP-1. In addition, using genotyped colonic biopsies from UC patients and control patients in a polarised *in vitro* organ culture (pIVOC) system we show that the *LPXN* risk allele may impact on cytokine production.

Finally, UKIBD genetics consortium data was used to access a pilot dataset of 58 patients' SNP profiles from Immunochip data who were patients at the Norfolk and Norwich University Hospital to create patient-specific UC-interactomes. Analysis of these footprints identified convergent interacting proteins affected by multiple SNPs and novel pathogenic pathways.

Declaration

I certify that the work contained in this thesis submitted by me for the degree of Doctor of Philosophy is my original work, except where due reference is made to other authors, and has not been previously submitted by me at this or any other university.

In line with regulations for the degree of Doctor of Philosophy, I have submitted a thesis that has a word count, including footnotes and bibliography but excluding appendices of 66,399 words

Acknowledgments

I am very grateful to the Norfolk and Norwich University Hospital for providing funding for the first year of this PhD and the Wellcome Trust for funding the two years to complete the PhD as a Wellcome Trust Clinical Training Fellow.

My heart and soul went into this work, trying to shed a little more light on an answer to my patient's question of 'why do I have this disease?' I would not have been able to do this without the initial blind faith of Dr Mark Tremelling and Professor Simon Carding that I would be PhD material and would be able to pull in funding. For taking that risk and for guiding and supporting me in this endeavour, I am eternally grateful to you both.

To Professor Alastair Watson, many thanks for the life lessons you gave me during your time as my supervisor. They will stand me in good stead.

To Dr Lizzy Bassity, for giving me the well-deserved ear-bashing whenever my plans were too extreme, and for suggesting I talk to the new research leader in Bioinformatics when I had come up with a madcap idea at the end of year 1; thank you.

To Dr Duncan Gaskin. I will never be able to emulate your magic with cloning, but I will continue to give it a darn good go. Thank you for your time, your advice and your reagents that you allowed me to 'borrow'.

To Dr Isabelle Hauteforte, Dr Emily Jones, Dr Ana Carvalho and Dr Sarah Clements for your friendship and providing the endless protocols, guidance, time, and patience during this PhD (let alone the coffee); thank you.

To Dr Arnoud Van Vliet and Dr Mark Fernandes for providing cake, waffles, strong coffee, a water filtration system and some good sound common sense; thank you.

To Gary, Tim, Isa, Oli and Liz for providing the soundtrack and letting me play along, thank you.

To Dr Tamas Korcsmaros, you saw the potential and you made it happen. Koszonom baratom.

I wholeheartedly dedicate this PhD with love to Richard, Ben and George Warburton. You provided the light in the darkness, the sound of laughter in the silence and the hugs when 'a little more trouble shooting' was required. My love, may you never have to proof read this again.

Contents

List of Figures

List of Tables

Abbreviations

5ASA	= 5 - Aminosalicylic Acid
AMP	= Adenosine Monophosphate
ANCA	= Antineutrophil Cytoplasmic Antibody
Anti-TNF	= Anti - Tumour Necrosis Factor
ARN	= Autophagy Regulatory Network
CAP-D3	= Chromosome Associated Protein D3
CARD	= Caspase Recruitment Domain Families
Cas	= CRISPR associated nucleases
CD	= Crohn's Disease
CD6	= Cluster of Differentiation 6
CI	= Confidence Intervals
DAMPs	= Damage Associated Molecular Patterns
DCs	= Dendritic Cells
DSS	= Dextran Sodium Sulphate
DMEM	= Dulbeccos Modified Eagles Media
DNA	= Deoxyribonucleic Acid
EBV	= Epstein Barr Virus
ECM	= Extracellular Matrix
EIM	= Extra-intestinal Manifestations
EPC	= Enhanced Parent Cohort
ESS	= Exonic Splicing Silencer
FAC	= Focal Adhesion Complex
GALT	= Gut Associated Lymphoid Tissue
GI	= Gastrointestinal
GIANT	= Genome scale Integrated Analysis of gene Networks in Tissues
GIT	= Gastrointestinal Tract
GPR35	= G-protein Complex Receptor 35
GWA	= Genome Wide Association
HBD	= Human beta-defensin 1

HDR = Homology Directed Repair

HLA = Human Leukocyte Antigen

HSF = Human Splicing Factor

ICAM1 = Intercellular Adhesion Molecule 1

IBD = Inflammatory Bowel Disease

IC = Immunochip

IEC = Intestinal Epithelial Cells

IFN = Interferon

IgA = Immunoglobulin A

IL = Interleukin

IL17REL = Interleukin 17 Receptor E Like

ILC = Innate Lymphcytoid Cells

LPXN = Leupaxin

JNK = Jun N-terminal Kinase

LncRNA = Long non-coding Ribonucleic Acid

LPS = Lipopolysaccharide

MAF = Minor Allelle Frequency

MAMP = Microbe Associated Molecular Patterns

MAPKS = Mitogen- Activated Protein Kinases

MES = Max Entrophy Scan

MHC = Major Histocompatibility Complex

MiRNA = Micro Ribonucleic Acid

MiRNA-BS = Micro Ribonucleic Acid Binding Sites

MMP = Matrix Metalloproteinase

mSNP = Missense Single Nucleotide Polymorphisms

MST1 = Macrophage Stimulating 1

NFKB = Nuclear Factor Kappa light chain enhancer of activated B cells

NHEJ = Non-homologous End Joining

NICE = National Institiute of Clinical Excellence

NOD = Nucleotide Oligomerisation Domain

NRAMP2 = Natural Resistance Associated Macrophage Protein 2

NXPE1 = Neurexophilin and PC-esterase domain Family Member 1

PAMP = Pathogen Associated Molecular Patterns

PCSK7 = Pro-protein Convertase 7

PRKCB = Protein Kinase C beta

PRR = Pattern Recognition Receptors

PTPN2 = Protein Tyrosine Phosphatase N2

PSC = Primary Sclerosing Cholangitis

PUC = Preceding Ulcerative Colitis

RSAT = Regulatory Sequence Analysis Tool

RTEL1 = Regulator of Telomere Elongation Helicase 1

sgRNA = Single Guide Ribonucleic Acid

SNP = Single Nucleotide Polymorphism

sSNP = Synonymous Single Nuceotide Polymorphisms

TB = Tuberculosis

TDT = Transmission Disequilibrium Testing

TFBS = Transcription factor Binding Sites

TLR = Toll Like Receptor

UC = Ulcerative Colitis

UCI = Inflammed Ulcerative Colitis

UC-Ome = Ulcerative Colitis Interactome

UCQ = Quiescent Ulcerative Colitis

UTR = Untranslated region

WHO = World Health Organisation

1. Introduction

1.1 Anatomy of the Gastrointestinal Tract

The gastrointestinal (GI) tract compromises a number of hollow organs from the mouth to the anus whose function includes digestion of food, absorption of nutrients, water and electrolyte balance and excretion of waste. The GI tract can be divided into two regions, based on embryonic origins of the foregut or midgut. The upper GI tract consists of the mouth, pharynx, stomach and duodenum. The lower GI tract consists of the remainder of the small intestine (jejunum and ileum), the colon and the anus (Figure 1-1). In order for digestion to occur, other visceral organs are required e.g. the liver, pancreas , gallbladder.

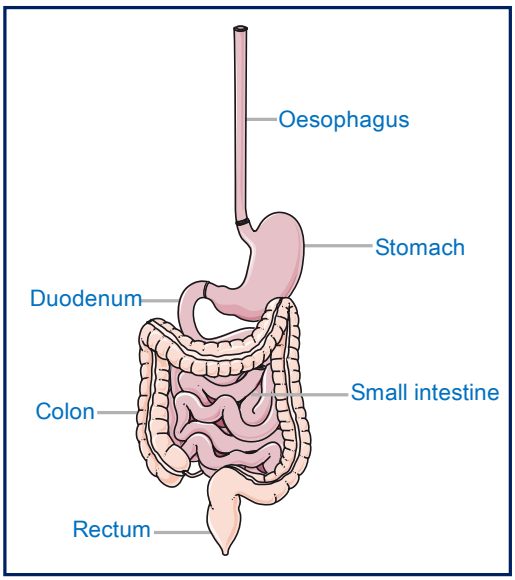


Figure 1-1 Anatomy of the human GI tract. Schematic representation of GI tract from oesophagus to rectum. (Adapted from Servier Medical Art, 2016)

The GI tract along its length is essentially a hollow tube comprising four layers (Figure 1-2), the mucosa, submucosa, muscularis and serosa.

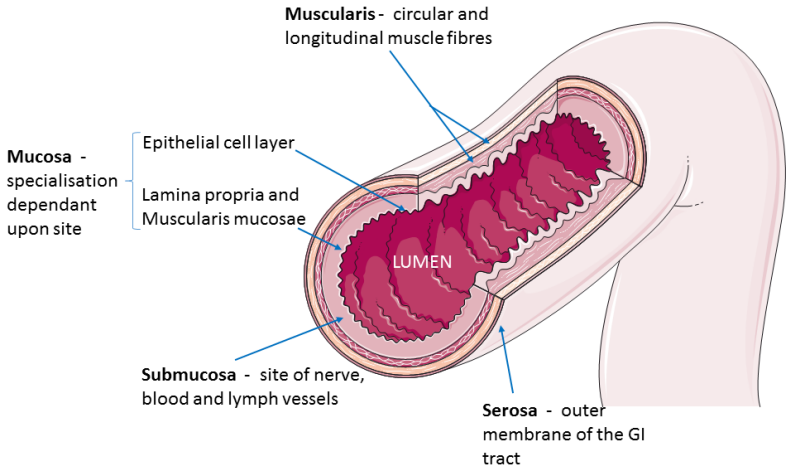


Figure 1-2 Layers of the GI tract. Schematic of the four layers of the GI tract; mucosa, submucosa, muscularis and serosa.(Adapted from Servier Medical Art).

Although each layer of the GI tract wall is present, each part of the GI tract has anatomical features and functions unique to that section. An example of this is the stomach wall epithelium containing parietal cells to produce the hydrochloric acid required for the chemical breakdown of food bolus or the small intestine having villi to increase the surface area available for nutrient absorption.

The function of the healthy colon is to reabsorb fluid and electrolytes from the stool and propulsion of intestinal contents towards the anal canal. The colon also plays a significant function in the induction of mucosal tolerance to gut microbiota. The human intestine contains over 100 trillion microorganisms, which live in symbiosis and homeostasis with the human host, providing a role in extracting energy, mineral and bioactive compounds from food. The gut associated lymphoid tissue (GALT) is the largest immune system in the body, which is continuously stimulated by gut microorganisms who provide natural antigens to induce mucosal immune tolerance (local and systemic immune unresponsiveness) to innocuous antigens (1). A functioning GALT is necessary to prevent acute proinflammatory immune responses against commensal microbiota (2). Tolerance to colonic microbiota attenuates local immune responses but not systemic immune responses, this is appropriate as an *E. coli* which is part of the common microbiota, may produce rapid fatality if allowed to populate the blood stream and other body organs.

1.2 Ulcerative Colitis

Inflammatory bowel disease is an umbrella term used to describe chronic or recurring inflammation of the gastrointestinal tract. It is a comprised of two major clinical entities. Crohn's disease (CD) – characterised by patchy inflammation of the gastrointestinal tract (GIT) which can occur anywhere between the mouth and the anus and has multiple phenotypes associated with the advent of stricturing of the bowel, perianal disease, or the formation of fistulae (abnormal connections between two hollow viscous or the bowel and the skin). This is opposed to Ulcerative Colitis (UC), which is characterised by continuous superficial ulceration of the large bowel from the rectum proximally, which can extend to involve the entire large bowel.

Classical histological features characterising CD are chronic inflammation comprising increased lamina propria, plamacytosis in association with chronic architectural distortion and patchy neutrophilic inflammation, cryptitis, crypt abscesses or erosions; skip lesions of focal, patchy erosions, vertical fissures and fistulas; transmural inflammation with multiple lymphoid aggregates; granulomas and submucosal fibrosis and neuromuscular hyperplasia of submucosa. This is in comparison to features of UC which also comprise alteration of crypt architecture, basal plasmacystosis, neutrophilic cryptitis, crypt abscesses; inflammation is typically limited to the mucosa and submucosa. The use of immunosuppressive medication can cloud the histological picture, making initial diagnosis and distinguishing between CD and UC more difficult.

UC has a bimodal age distribution with the first peak of diagnosis in the first and second decade of life and a second incidence peak occurring after the sixth decade of life. The annual incidence of UC in Europe is 24.3 per 100,000 person years (prevalance 505 per 100,000 person years), 6.3 per 100,000 person years in Asia and Middle East (prevalance 249 per 100,000 person years), 19.2 per 100,000 person years in North America.

Within Europe, extensive review of incidence figures indicate and estimate of 178,000 new cases of UC across Europe per year, with the prevalence expected to increase due to the early age of onset and low mortality of UC patients.

UC can affect the large bowel from the rectum extending proximally. The site, extent and severity of the disease dictate therapeutic management strategies. The site can be that which affects only the rectum (proctitis), that affecting the rectosigmoid (proctosigmoiditis also known as distal colitis), that affecting up to the splenic flexure (left

sided colitis), and intermediary of that affecting around to the transverse colon ('extensive') and that affecting the whole large bowel (pancolitis).

In the IBSEN cohort (3), the distribution of disease extent was 32% patients had proctitis, 35% had left sided colitis (above the rectum to the splenic flexure) and 33% had extensive (past the splenic flexure to pan colitis). Interestingly 28% of the patients with proctitis had disease that progressed proximally with 10% extending to extensive colitis over the 5 years follow up period.

There are additional phenotypes which include those patients who development extraintestinal manifestations (EIMs) (Table 1-1). The classification of EIMs is wide and can be broadly characterised into complications of the disease and or treatment e.g. osteoporosis or dermatological manifestations of nutritional deficiencies, or diseases that are associated with a patient's HLA status e.g. ankylosing spondylosis, psoriasis, vitiligo, or "inflammatory processes" that follow the inflammatory course of the IBD e.g. seronegative arthritis, reactive dermatological lesions, and ocular manifestations. Extraintestinal manifestations are seen in 25-40% of patients with inflammatory bowel disease, why they occur in some patients but not others are not always clear. There is also a subset of patient with UC who develop Primary Sclerosing Cholangitis(PSC) – inflammation and scarring of the bile ducts which can progress to liver cirrhosis requiring the patient to potentially undergo a liver transplant and predisposes to the development of bile-duct carcinoma. The severity of PSC is well documented to be inversely proportional to the UC inflammation severity.

Site	Manifestation	Chronological Status	Comment
Musculoskeletal	'Seronegative' arthritis	Preceding UC (PUC), synchronous(S), post diagnosis(PD)	Inflammation associated
	Ankylosing spondylitis	PUC	HLA
	Osteoporosis	PD	Vitamin D deficiency (as opposed to steroid use)
Dermatologic	Erythema nodosum, Pyoderma gangrenosum	S, PD	'Reactive'/Inflammation associated
	Acrodermatitis	PD	Nutritional deficiencies
	Vitiligo, psoriasis, amyloidosis	PUC, S, PD	HLA
Hepatobiliary	Primary sclerosing cholangitis	PUC (PSC patients will get colonoscopy if no known diagnosis of UC), S, PD	
	Autoimmune chronic active hepatitis, granulomatous disease	PD	Inflammatory
Ocular	Uveitis/iritis, episcleritis, scleromalacia, corneal ulcers, retinal vascular disease	PUC, S PD	Inflammatory
Metabolic	Growth retardation in children	PD	Disease and treatment associated

Table 1-1 Extraintestinal manifestations of UC

Treatment of UC is dependent on severity as determined by the Truelove and Witt criteria(4) (Table 1-2) and to some degree the extent of the disease in the milder cases (Figure 1-3). As shown by the flow charts below, mild to moderate flares are treated depending on the site of the disease to bring the acute symptoms under control. Patient tolerability and side effects of drugs also play a significant part in medical management.

Feature	Mild	Moderate	Severe
Stool frequency per day	<4	4-6	>6 plus at least one other feature of systemic upset (*)
Blood in stools	Small amounts	Between mild and severe	Visible blood
Pulse >90bpm (*)	No	No	yes
Pyrexia >37.8 ℃(*)	No	No	Yes
ESR (mm/hr) (*)	30 or below	30 or below	Above 30
Anaemia (*)	No	No	Yes

Table 1-2 Truelove and Witt Criteria for disease severity

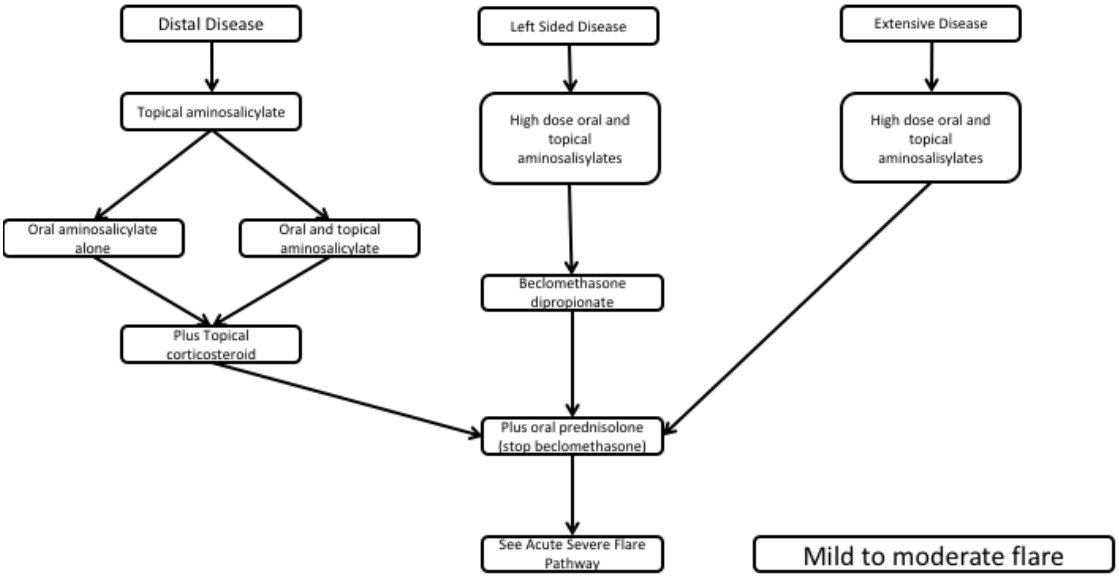


Figure 1-3 Flow chart of treatment for mild to moderate flares (adapted from NICE Guidelines for the Management of UC)

For acute severe flares, a different strategy is required to rapidly bring the inflammation under control in an inpatient setting, this includes intravenous drug administration and close monitoring (Figure 1-4). The font size is reference to the most frequent administrated therapy or approach.

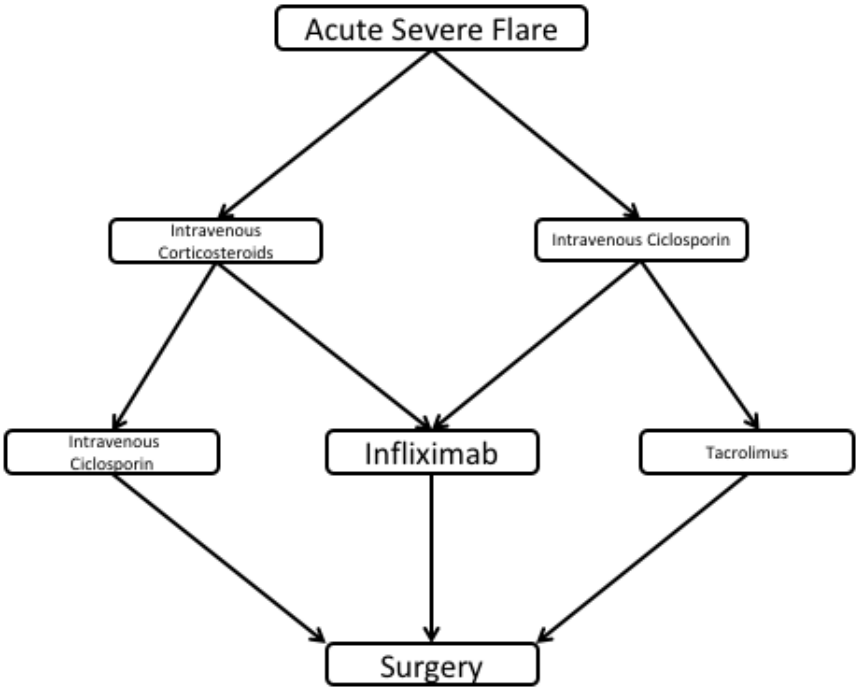


Figure 1-4 Flow chart for the treatment of acute severe UC flare (adapted from NICE guidelines for the management of UC)

Once the flare is under control, the aim of ongoing management is to prevent further flares or maintain the patient in remission. Maintenance therapy, like treatment of flares, is site specific and functions, like the WHO Analgesia ladder, in terms of starting with the 'gentler' therapies first and escalating though increasing immunosuppression to monoclonal antibody use (Figure 1-5).

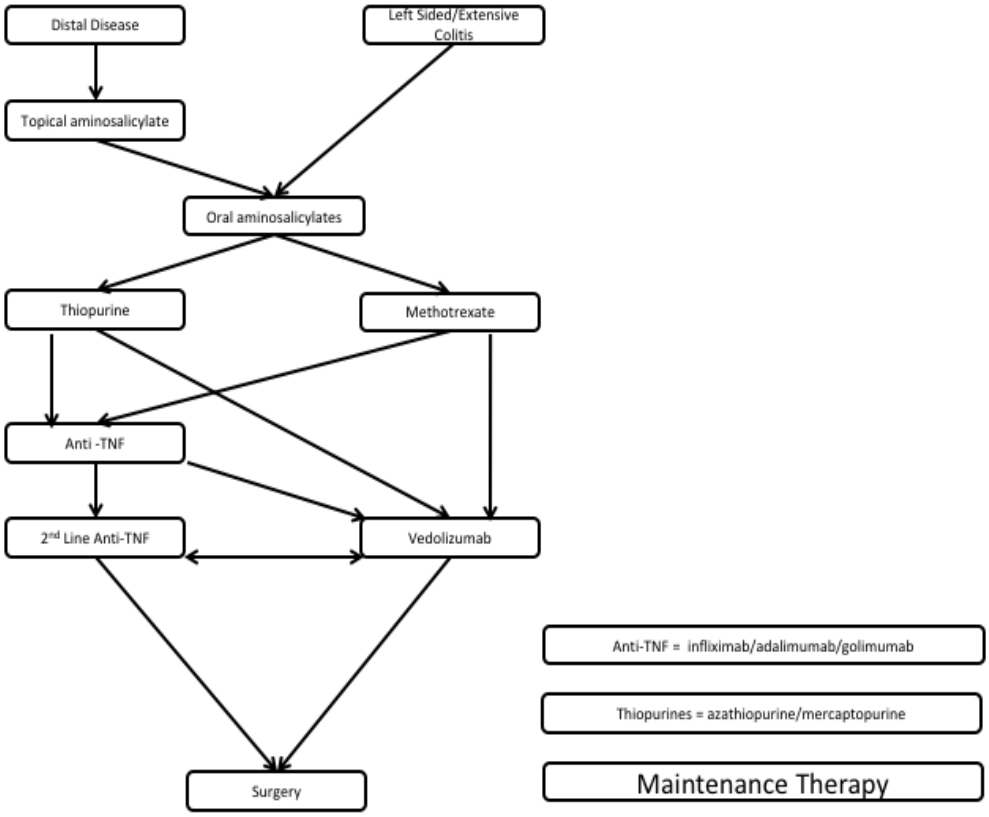


Figure 1-5 The ladder of maintenance therapy for UC

The side effect profile of 5Amino salicylic acids (5ASAs) are not insignificant, but affect a minority of patient with 11.1 reports per million prescriptions of interstitial nephritis (5, 6) and 7.5 reports per million prescriptions of pancreatitis (5). 50% of UC patients treated with 5ASAs require escalation of treatment. The first rung of maintenance therapy escalation is to the thiopurines; purine analogs which halt DNA replication in actively replicating cells (such as T and B cells). Consequently, bone marrow suppression and hepatoxicity are dose-dependent side effects of thiopurines. Other side effects that have been reported by multivariate pharmacosurveillance include pancreatitis, anaemia, cell lysis and nephrotoxicity (7). Thiopurines are associated with an increased risk of lymphoma and non-melanoma skin cancer (8). Patients requiring further escalation of therapy to monoclonal antibody use, are at risk of serious and opportunistic infections , such as pneumonia, sepsis, fungal infections and tuberculosis (TB) (9) (10). A variety of infections have been reported in patients treated with Anti-TNF monoclonal antibodies including candidiasis, varicella zoster, herpes zoster, EBV, CMV, herpes simplex, *Listeria* infections and *Pneumocystis jirovecii.* Reactivation of infectious diseases that are dormant is also a risk, including TB and Hepatitis B (11). Other adverse events for Anti-TNFs include infusion reactions which are characterised by arthralgia, myalgia, urticarial rash, fever and or malaise, injection site reactions, skin lesions, lupus like syndrome, demyelinating disease, heart failure, melanoma and cervical dysplasia. Combination therapies such an anti-TNFs and thiopurines, may be associated with marginally higher rates of TB, candidiasis and herpes zoster, although this is not consistent across the literature field, as some studies suggest there is no difference between risk of infection between single agent and combination therapy (SONIC trial(12)). The risk of lymphoma is increased in combination therapy and specifically the risk of hepatosplenic T cell lymphoma is increased in young men (13).

1.2.1 The Genetics of UC

Although genetics can by no means explain the increasing incidence of Inflammatory Bowel Disease, there is a genetic predisposition. In 1987, Monsen et al(14) published an observational study of the familial occurrence of inflammatory bowel disease in patient with UC. They showed that the prevalence of UC in first-degree relatives was 15 times higher than in non-relatives and there was a general prevalence of 7.9% for IBD (regardless of the type) among relatives.

By 1989 the first genetic associations for UC were being identified by candidate gene analysis in the major histocompatibility complex; HLA genes (15), as well as T cell receptor and Immunoglobulin heavy chains (16). The HLA region continued to be of significant interest, with the HLA-B locus associated with UC susceptibility in Japanese populations (17), and HLA-DR2, HLD-DR4 being associated with UC in the context of ANCA antibodies (18). The HLA regions continued to have both positive and negative associations with UC depending on the size, the clinical heterogeneity and ethnicity of the cohort (19-21). The impact of disease heterogeneity, ethnicity and differing methodologies on genetic marker studies was highlighted by Satsangi *et al* (22) in their investigation of the relationship between ANCA status, HLA genotype and clinical patterns on IBD where ANCA positive status was not associated with HAL-DR2 or DR4 in the UC population.

Also within the MHC complex lie the tumour necrosis factor genes. In 1996, the distribution of 4 polymorphisms in TNF genes were analysed in IBD (20), which began to identify some of the complexities associated with susceptibility polymorphisms. The authors highlighted that some of the polymorphisms were more present in UC compared to controls, some were less present, and there were tendencies as opposed to statistically significant associations. The authors concluded that although the TNF genes were not susceptibility markers, they may be markers for subsets of patients with UC.

Cytokines and their role in inflammation provided a good source of genes for candidate gene analysis. The IL-1 receptor antagonist was associated with UC susceptibility in 1994 (23)with allele 2 of interleukin-1 receptor antagonist (IL-1RA) being significantly over-represented in UC patients (35% vs 24% in controls), the authors also noted that this was more prevalent in UC patients with total colitis. Further examination of allele 2 of IL-1RA identified that a synergistic effect between IL-1beta/IL-1RA allelic cluster participated in the susceptibility to UC (24). This cluster was reanalysed in a different cohort and no significant difference in genotype distribution was found, however when the cohort was

stratified by disease severity, there was a higher frequency of UC patients who required surgery who had the genotype of interest (25).

In a further candidate gene study, IL-2 microsatellite markers (polymorphic dinucleotide repeats) were modestly linked with UC, but due to small numbers the TDT failed to reach significance (26).

ICAM1 (intercellular adhesion molecule-1) was the next candidate gene to be analysed in UC associated with ANCA status (27), which again on a population wide scale identified no association with UC or CD, but when stratified for ANCA-status there was a borderline statistically significant association with ANCA negative UC patients. How ANCA status and ICAM-1 polymorphisms functioned synergistically was not identified.

The hint of genetic involvement of epithelial barrier dysfunction in UC, came with the Kyo et al paper (28) which identified rare polymorphisms of variable number of tandem repeats in the intestinal mucin gene MUC3 that was associated with UC in both Japanese and Caucasian populations

With the advent of improving gene mapping technology and GENEHUNTER (a linkage analysis program (29), IBD2, a pericentromeric region of chromosome 12 was identified as associated with UC. This was confirmed with strong evidence for linkage of both CD and UC (30).

A further IBD locus, IBD1, on chromosome 16 was associated with UC. By typing eight microsatellite markers from the IBD1 locus in 70 kindreds Mirza et al (31) identified that the locus D16S419 was associated with an estimated relative risk of 1.46. The authors concluded that IBD1 may contribute to the susceptibility of UC.

GENEHUNTER again, with transmission disequilibrium testing (TDT) was used to confirm IBD2 as being linked to UC in 122 North American Caucasian families (32). Chromosome 12 and chromosome 16 continued to provide strong linkages to IBD in sibling genotyped cohorts but this technique requires large cohorts of sibling related disease and non-disease sufferers.

In a step away from candidate gene analysis, genome wide screening using 377 autosomal markers in sibling pairs identified IBD susceptibility loci in 1p, 3q and 4q with potential epistasis between 1p and IBD1 (33). Genome wide screening on larger sibling pair cohorts using autosomal markers began to link further chromosomal loci to UC, including loci on

chromosome 1, 6 (6p particularly – the site of leukocyte antigen and TNF genes), 10 and 22.

Genome wide scanning provides the basis for further candidate gene analysis, an example of this is the gene for natural resistance associated macrophage protein 2 (NRAMP2). This was examined on the basis of the genome wide scanning and a plausible hypothesis that NRAMP2 had a role in innate immunity. Unfortunately, nonparametric linkage analysis and TDT did not provide evidence of linkage of NRAMP2 to IBD, nor UC or CD specifically. Further sequence analysis, although identified that the signal from the genome wide scanning of chromosome 12 was not NRAMP2 and that the signal was due to linkage disequilibrium with the disease causative gene (34). Higher density genome wide scans aimed to overcome this difficulty, but locus heterogeneity and sample size were limiting factors in the fine localisation of disease susceptibility loci (35, 36) in the early 2000s.

With the completion of the Human Genome Project in 2003(37) which identified 3 billion bases of human DNA and the International HapMap Project 2005(38) it was possible to catalogue the wealth of single nucleotide genetic variants within humans. The HapMap described the SNP position within the DNA and their distribution amongst populations. As described above, allelic variations in candidate genes had already been associated with UC, however now, scientists had the availability to undertake hypothesis free analysis of the genetics of disease cohorts, thereby potentially identifying novel susceptibility genes associated with disease susceptibility by genome wide association scanning.

Genome wide association (GWA) is an approach that involves rapid scanning of markers across genomes of specific populations to highlight genetic variations associated with a particular disease. GWA involves identifying SNPs that are present in the target population e.g. patients with UC, significantly more or less frequently than in the control group. The SNPs, however, may not be disease causative, and like the NRAMP2 story, the identified SNP may be in linkage disequilibrium with the causative variant.

Fine mapping for possible causal alleles may alleviate this problem, but again markers that are strongly correlated could still be relatively distant, but statistically close. Deep sequencing where whole genomes are sequenced multiple times allows for validation of previously identified SNPs, may identify causal variants as well as highlighting rare variants.

Over 163 SNPs(39) have been associated with Inflammatory Bowel Disease, many involving the adaptive immune system overlapping between both CD and UC, suggesting a common pathway(40). Key pathways(41) that have been identified for CD include autophagy pathways, the innate and adaptive immunity and bacterial recognition. For UC, the mucosal epithelial barrier has been highlighted as an area of interest (41)(Table 1-3).

Gene	Protein	Function
CDH1	E-cadherin	Adherens junction protein
LAMB1	Laminin B1	Involved in adhesion and differentiation
GNA12	Guanine nucleotide binding protein a12	Inhibition of tight junction assembly
PTPN2	TCPTP	Inhibits IFN-y induction of claudin 2
HNF4A	Hepatocyte nuclear factor 4a	Regulates differentiation along crypt-villus axis
ECM1	Extracellular matrix protein 1	Cell proliferation
ITLN1	Intelectin-1	Brush border protection
PTGER4	EP4 Receptor	Epithelial restitution
CARD15	NOD2	Recognition of PAMPS

Table 1-3 UC risk associated genes that have roles in the intestinal epithelial barrier

Whole genome sequencing at low coverage has identified a further missense variant at *ADCY7* that doubles the risk of UC (42). Further high resolution fine mapping, identified 18 associations to a single causal variant with >95% certainty and a further 27 associations with a single variant with >50% certainty. The variants were enriched for protein coding changes, direct disruption of transcription factor binding sites and tissue specific epigenetic marks, with gut mucosa associations stronger in UC (43). Translational impacts of IBD SNPs have been seen in the IL23/IL17 axis, as well as the JAK/STAT pathway. SNPs in the IL23 receptor have been associated with both CD and UC, as well as psoriasis and ankylosing spondylitis suggesting a shared inflammatory pathway. The IL23/IL17 axis has been heavily implicated in inflammatory bowel disease with risk variants being identified

in the *IL23R*, *STAT3*, *JAK2*(signal transduction), *IL12b* (common subunit of IL12 and IL23) and *CCR6* (chemokine expressed on IL17 producing cells)(39, 44, 45). In terms of translational aspects of this, the JAK pathway has been targeted via Tofacitinib, on oral JAK inhibitor, which has been trialled for induction of remission in patients with moderate to severe UC(46); Ustekinumab (Stelara), which targets the p40 subunit of IL12/23 is another novel therapeutic agent used in CD(47).

Many of the loci identified within the GWA share association with other inflammatory disorders such as coeliac disease, multiple sclerosis, primary biliary cirrhosis (which is clinically associated with UC) highlighting that these susceptibility loci create a predisposition to a chronic inflammatory state(41), but the phenotype is determined by other factors, such as tissue specific epigenetics (as highlighted by differential genome wide transcriptome(48) and methylome analysis of colonic biopsies in UC patients compared to healthy controls(49) or environmental factors such as smoking, the role of the gut microbiota(50), enteric infections and antibiotic exposure in childhood. To add weight to this, twin studies have demonstrated a change in microbiota between patients with CD and their healthy twin(50). The relapsing/remitting nature of UC and the onset of disease in the 2nd the 3rd decade of life also (51)implicates environmental processes in the instigation of the inflammatory response. These environmental factors and epistasis(52) may explain the gap between the explained heritability and the true heritability of UC.

IBD related genes have been shown to organise into regulatory networks enriched for inflammatory and immune networks (53). Using integration of large scale DNA, RNA variation data in the context of active IBD using immune networks, a conserved inflammatory component was highlighted which was enriched for genes associated with known CD and UC susceptibility loci. Key driver genes including *CD53, RHOH, DOCK2, FGR, NCKAP1L, CXCL10, FCER1G, SLAMF8, NFAM1, P1K3CD, DOK3, GBP5, AIF1, GPSM3* have been identified and validated in macrophage cell culture models. None of the individual mutations gave rise to spontaneous colitis in the models, highlighting that none were causal, but result in subtle modulations of regulatory states. The networks demonstrated a high degree of connectivity suggesting that the key driver genes are linked (54).

The Colonic Mucosal Barrier

The human GI tract is exposed from birth to food, environmental toxins and microorganisms. Colonisation of the human GI tract from birth by microbes is essential for normal function of the GI tract and development of the GI associated immune system (51). This complex and dynamic ecosystem containing a diverse intestinal microbiota that includes Archae, Bacteria and Eukarya have been shown to impact on human health parameters including metabolic (55), nutritional (56), physiological (57, 58) and immunological processes(59) within the human body. Disturbance of the microbiota in childhood, for example, by multiple exposures to antibiotics has been associated with development of allergies (60) and CD (61, 62). Across IBD a reduction in bacterial diversity of the microbiota has been observed consistently and in UC a reduction in Bacteriodetes has been documented (63). The role of microbiota dysbiosis in the aetiology of intestinal inflammation remains unclear (64), but we know there is a fine homeostatic balance between the beneficial effect of the commensals and the risk of commensals causing a systemic inflammatory response should they cross the gut wall. The human body defends against this at the point of contact with the formation of a selectively semi-permeable epithelial barrier.

This intestinal barrier has three major components composed of physical, biochemical and immune elements.

1.2.2 The physical barrier of the mucosal epithelium

The physical barrier consists of a stratified mucous layer and mucosal epithelium comprising enterocytes, goblet cells, crypt epithelial cells, and intraepithelial lymphocytes which form and are part of a semi-permeable epithelial barrier. Beneath the mucosal epithelium is connective tissue and supportive tissue of the lamina propria. Within the lamina propria are the immunocompetent cells including dendritic cells, macrophages and lymphocytes which form a functional unit with the epithelial cells (Figure 1-6).

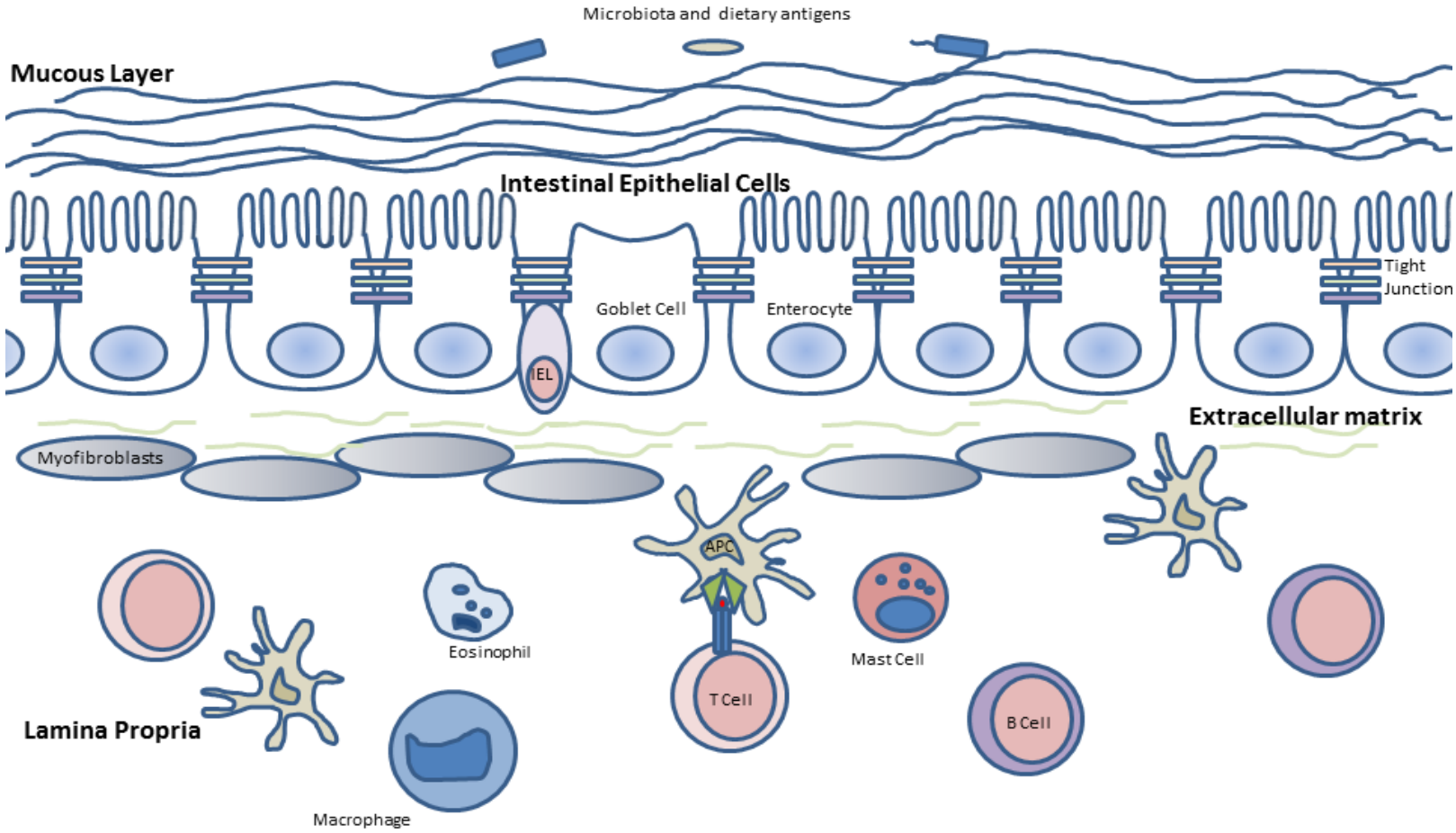


Figure 1-6 Graphic representation of the cells of the mucosal epithelium and within the lamina propria of the intestine.

IEL = Intra - Epithelial Lymphocyte
APC = Antigen Presenting Cell

The mucous layer is mainly comprised of densely O-glycosylated MUC2 (65) secreted by goblet cells. *Muc2-/-* mice develop colitis and have an increased risk of colorectal cancer associated with the presence of bacteria in direct contact with the mucosal epithelium(66). Examination of colonic biopsies from UC highlight a reduction in the number of goblet cells and corresponding reduction in expression of MUC2 and 3, with reduction correlating with severity and extent of disease(67). During inflammation, the mucous layer in UC is thinner or absent (68), probably due to an increase in mucolytic gut microbiota in UC patients (69).

Enterocytes, which form the majority of the mucosal epithelium, are columnar cells approximately 25um in height and 8um in width. Their apical surface is covered with microvilli which is covered in a mucous layer. The enterocytes are connected with adjacent cells by junctional complexes; tight junctions, adherens junctions, desmosomes, and gap junctions. These junctions help maintain barrier integrity. The basal membrane is anchored to the extracellular matrix by hemidesmasomes and focal adhesion complexes (Figure 1-7). The extracellular matrix is composed of collagen, laminin, fibronectin and glycosaminoglycans. Defects in the epithelial barrier and increased intestinal permeability has been demonstrated in UC and CD as well as non-affected family members and spouses (70-72).

Tight junctions regulate epithelial permeability (73). The junctional complexes are disordered in UC, with desmoglein-2 (desmosome), occludin, E-cadherin and Beta catenin expression reduced in inflamed UC colonic samples compared to non-inflamed UC colonic samples and non IBD controls (74). Tight junction structure is known to be altered in CD in both inflamed and non-inflamed mucosa leading to an impaired barrier function (75). Inflammatory cytokines TNFa and IFN-y trigger intestinal barrier defects acutely by cytoskeletal contraction (actin reorganisation) (76) or chronically by modulation of tight junction protein expression(77-80).

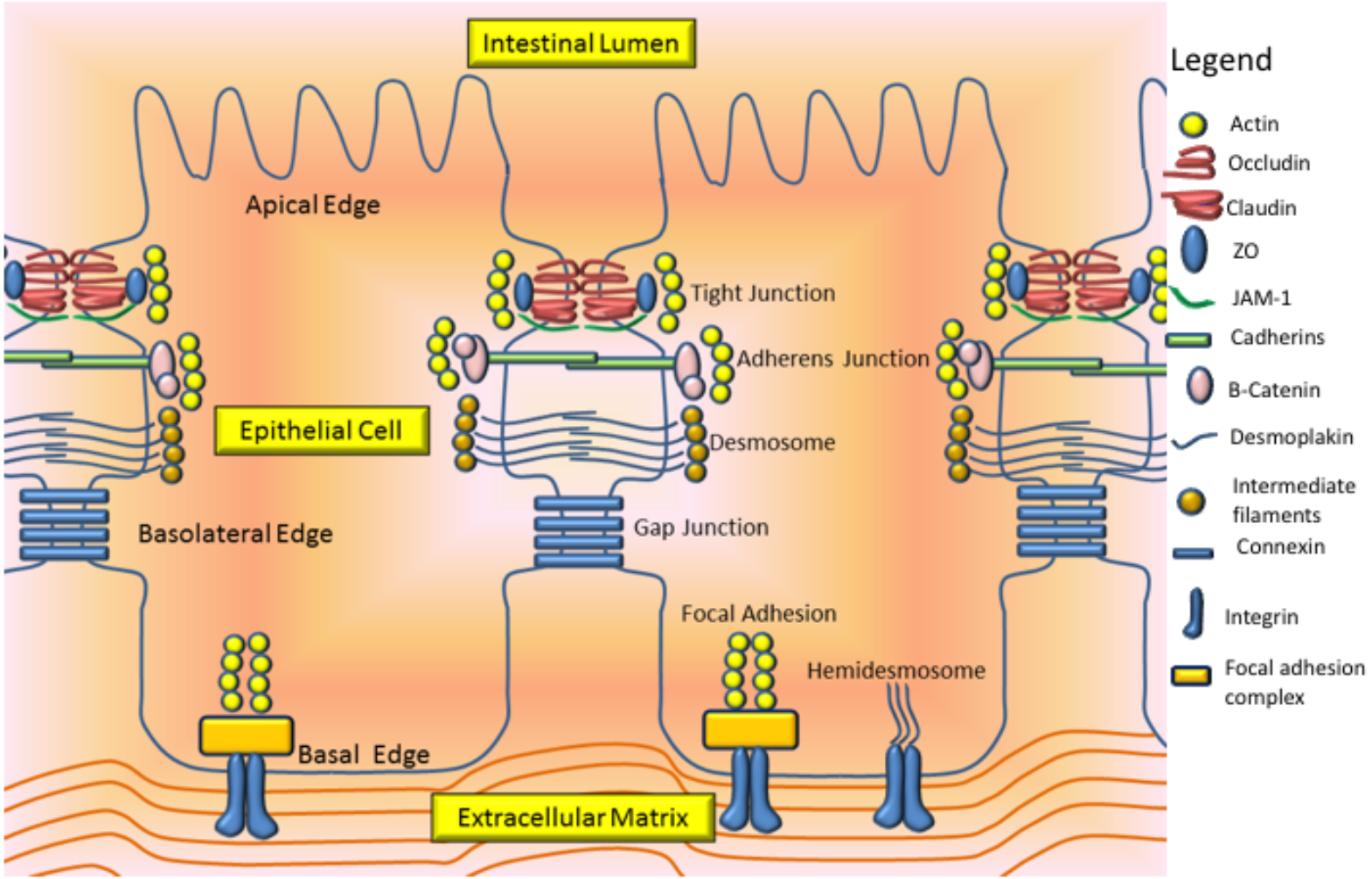


Figure 1-7 Epithelial junctional complexes. A graphical representation of the cell-cell and cell to matrix junctional complexes.

The junctional complexes are also regulated by microRNAs. miRNAs are small, endogenous RNA molecules that can negatively regulate target gene expression at the post transcriptional level. Table 1-4 identifies the miRNAs involved in intestinal epithelial permeability, several of which have been shown to be overexpressed in UC. MicroRNAs are aberrantly expressed in both the mucosa and peripheral blood of UC patients (Table 1-5)(81), however their functions have yet to be fully elucidated.

	miRNA	Function	↑or ↓in UC
miRNAs affecting epithelial permeability	Mir-21	Degradation of RhoB mRNA, leading to an increase in epithelial permeability due to loss of TJ proteins	Over expressed
	Mir-150	Intestinal epithelial disruption due to repressing c-Myb	Over expressed
	Mir-874	Repressed AQA3, reduced MUC2, suppression of occluding and claudin 1	Over expressed
	Mir-9 and Mir-374	Claudin-14 translational repression	unclear
	Mir-145	Repressing junctional adhesion molecule -1 expression	unclear
	Mir-212	Repression of ZO-1 expression	unclear

Table 1-4 MicroRNAs affecting epithelial permeability in UC patients.

	miRNA	Function	↑or ↓in UC
miRNA in UC mucosal tissue	Mir-192	Regulation of chemokine production in colonic epithelial cells	Down regulated
	Mir-375 Mir 422b	Unknown	
	Mir-16 Mir-23a Mir-24 Mir-126 Mir-195 Let-7f Mir-29a/b Mir-7 Mir- 127-3p Mir-135b Mir-223 Mir324-3p Mir-31 Mir-155 Mir-146a Mir-206 Mir-424 Mir-20b Mir-125b-1		Over expressed
	Mir-188-5p Mir-215 Mir-320a Mir-346 Mir-200b		Down regulated
miRNA in peripheral blood	Mir-16 Mir-21 Mir-28-5p Mir-151-5p Mir-155 Mir-199a-5p Mir-188-5p Mir-378 Mir-422a Mir-500 Mir-501=5p Mir-769-5p Mir-874		Over expressed
	Mir-505		Down regulated

Table 1-5 MicroRNAs aberrantly expressed in UC patients, both in the mucosa and peripheral blood.

1.2.3 The biochemical barrier of the mucosal epithelium

As well as being a physical barrier, enterocytes recognise gut microbiota with pattern recognition receptors (PRR) such Toll Like Receptors (TLRs), nucleotide oligomerisation domain (NOD) or caspase recruitment domain families (CARD). These receptors recognize microbes by essential and highly conserved 'pathogen-associated molecular patterns' (PAMPs) or 'microbe-associated molecular patterns' (MAMPs). When stimulated with a ligand, the PRRs induce a rapid first line of defence. There is, however, tolerance towards the presence of commensal microbiota by limiting the expression of PRRs in the human colon e.g. TLR3 and TLR5 which recognise viral infections and flagellin are abundantly expressed, but TLR2 and TLR4 (which recognise lipoptide agonists and lipopolysaccharide) expression is low in a healthy gut but can be induced (82). Toll like receptor signalling compounds include MyD88, which, when knocked out in mice leads to reduced levels of mucin-2, impaired antimicrobial activity, a greater number of mucus associated bacteria, translocation events and colitis susceptibility (83). PRR signalling, therefore plays an important role in intestinal homeostasis as well as the mucosal innate immune response.

Stimulation of PRR in colonic enterocytes leads to production and secretion of antimicrobial peptides called Beta-defensins. Beta-defensins are dependent on PRRs for transcription induction or for secretion. Human Beta –defensin 1 (HBD1) is constitutively expressed in colonic enterocytes, HBD2 is inducible and upregulation is dependent on inflammation or pathogens present (84, 85). Patients with colonic CD (CD) have a diminished ability to upregulate HBD2 and secrete it (85). In CD, a single nucleotide polymorphism (SNP) in the HBD1 gene promoter was annotated to affect the HBD1 mRNA level (86), leading to low HBD1 mRNA levels in colonic CD (87). In active UC defensins are upregulated (88) indicating stimulation of PRRs in the inflammatory process.

Once the mucosa is injured AMPs from immune cells within the lamina propria are released, this includes calprotectin which is now utilised in a clinical setting by physicians as a marker of inflammatory in IBD patients (36) (89). Calprotectin is a calcium binding S-100 protein family member that is released from dead and dying neutrophils. It inhibits growth of *Staphylococcus aureus, E. coli* and *Candida albicans*. Lactoferrin, a glycoprotein AMP that is found in the secretory granules of neutrophils, was also investigated as a biomarker of inflammation in IBD (90), but was found to be less sensitive than faecal calprotectin (91) (92). Lactoferrin sequesters iron in the mucous layer, thereby inhibiting microbial growth. Immune cells within the lamina propria also release TNFalpha and IFNy

which induces epithelia cell apoptosis, leading to increased permeability of the intestinal mucosa (93).

1.2.4 The colonic innate and adaptive immunity cross over

The interplay between the rapid, non-specific response of the innate immune system and the highly specific, long lasting immunity of the adaptive immune system plays a key role in intestinal homeostasis and when dysregulated, the pathogenesis of IBD. The triggers of the innate immune system, the PRRs are controlled via expression and localisation. For example, intestinal epithelial cells lack expression of TLR4, or MD2/CD14 (required proteins for TLR4 recognition and signalling), or TLR4 is located in intracellular compartments to avoid responding to LPS from commensal microbiota. Upregulation of MD2 and TLR4 is seen in CD due to high local levels of IFN-y or TNFa (94). Low levels of TLR2 protein are found at subapical locations and TLR5 (flagellin recognition) is found on the basolateral side, thereby reacting only when microorganisms cross the epithelial barrier (95). If TLR9 (intracellular DNA recognition) is activated at the basolateral side of intestinal epithelial cells, it causes secretion of cytokines via degradation of the NFKB inhibitor IkBa, thereby activating NFKB. If TLR9 is activated at the apical side of IECs, it causes IkBa accumulation, thereby preventing NFKB signalling (96).

Once a TLR response is triggered, as opposed to the oral and airway epithelium which secrete predominantly antimicrobial peptides, the intestinal epithelium secretes inflammatory mediators. Pro-inflammatory cytokines such as IL-1b, IL-7, IL-8, IL-15, IL-12, IL-33 and IL-18 drive recruitment and activation of granulocytes (predominantly neutrophils), macrophages, dendritic cells (DC), Natural Killer cells and innate lymphocytoid cells (ILC) (97-99). Activation of dendritic cells leads to increased production of the cytokines described above, causing proliferation and activation of NK and ILCs. ILCs regulate CD4+ T cell responses to commensal bacteria(100), NK cells modulate CD8+T cells (101, 102). IL-18 secretion also increases IL-2 and INF-y production, thereby changing mucus production and composition. IECs also secrete anti-inflammatory cytokines IL-10 and TGF-b, thereby reducing an excessive inflammatory response and aiding the tissue repair process (103). If the intracellular TLRs are triggered, it produces a rapid interferon (IFN) response which has two major effects, firstly the expression of viral restriction factors is upregulated by IFNs, secondly IFNS can modulate the functions and activation of DCs, NK Cells, T cells and B cells (104).

Intestinal DCs have functional plasticity in their ability to generate either inflammatory or tolerogenic immune responses. In the homeostatic state, DCs are hyporesponsive (105).In IBD, activated DC accumulate at the sites of inflammation (106). They function as a hub for multiple multicellular immune cascades, bridging the innate and the adaptive immune systems. They contribute to T reg generation (107), imprint their homing properties on T cells and B cells in order to localise immune responses to particular tissues (108), for example by increasing expression of $\alpha 4\beta 7$ integrin (gut homing) on T cells and induce B cell class twitching to IgA producing B cells with tolerogenic properties (109).

There are multiple different T cell populations that can be present in the intestinal lamina propria, but two specific types show distinct intestinal trophism; $\gamma\delta$T cells and Th17 cells.

$\gamma\delta$T cells, are unconventional CD3+ T cells which have both innate and adaptive immune functions. They have TLR, Notch and NKG2D activating receptors which they use in conjunction with the T cell receptor to rapidly respond to stress-induced ligands and infection – irrespective of their molecular nature. In the intraepithelial lymphocyte compartment, they are CD8aa+, when circulating they are MHC unrestricted.

$\gamma\delta$T cells demonstrate cytotoxic responses in the same manner as conventional T cells through perforin/granzyme , Fas/FaL pathways and production of TNF alpha(110) (111). $\gamma\delta$T cells have also been shown to modulate goblet cell numbers and thereby mucus production (112); $\gamma\delta$T deficient mice have an increased susceptibility to DSS colitis due, in part, to a reduction in the number of goblet cells.

Th17 cells are CD4+ T cells which are enriched in the lamina propria under homeostatic conditions. Although in IBD they have been shown to be microbiota reactive (113), they have a dynamic response which can be both pro-inflammatory and anti-inflammatory, as evidenced by the lack of response or exacerbation of symptoms in trials of anti-IL-17A or anti-IL17RA antibodies in CD patients (114) (115).

Th17 cells produce signature cytokines in excess in both CD and UC (116); IL17, IL-21 and IL22. IL17 stimulates innate immune cells and epithelial cells to produce IL-1, IL-6, G-MSF and IL-8 which in turn leads to increase neutrophil production and recruitment leading to increased pro-inflammatory signalling (117) (118). IL22 is not only produced by Th17 cells, but also by NK cells and ILCs (119). Detection of IL22 by intestinal epithelial cells leads to increased epithelial barrier integrity (120) and production of REGIII antimicrobial proteins (121, 122).

Intestinal B Cells are specialised in IgA production as differentiated plasma cells. Secretory IgA in the intestinal lumen acts as a barrier by opsonising pathogens to protect the epithelium from invasion. The sIgA system is tightly integrated with both innate and adaptive immune mechanisms, influencing adaptive T cell responses (123) and contribute to immune homeostasis. B cells also function as antigen presenting cells, polarising effector T cell responses (124). IL10 produced by B cells with the cognate B cell/T cell interaction is required for the generation of mucosal Tregs (125) (126). Regulatory B cells have a inflammatory suppressive function, producing IL-10 and TGFβ, thereby playing a role in suppressing inappropriate responses to intestinal microbiota and innocuous antigens.

Several SNPs correlating to genes encoding proteins involved in the maintenance of the epithelial barrier have been demonstrated in association with UC. A selection of the candidate genes from Jostins et al(39, 127) and their potential site of protein impact on the intestinal barrier are diagrammatically represented in Figure 1-5.

The genes highlighted by eQTL in Jostins et al as UC risk associated are hypothesised to have roles in maintenance of the epithelial cell function via gene regulation or downstream signalling of PPR recognition (*SLC9A3, CALM3, RNF186, PRKCD, CDH1, HVELM, HNF4A, NFKB1, CARD11, IRF5, GNA12*) , roles in wound healing in the extracellular matrix (*MMP24, ECM1, ITIH1, LAMB1*) or roles in the innate-adaptive cross talk (*IRF5, NFKB1, HLAs, CARD11*). The functional annotation of the SNPs to drive the hypothesis, however is lacking.

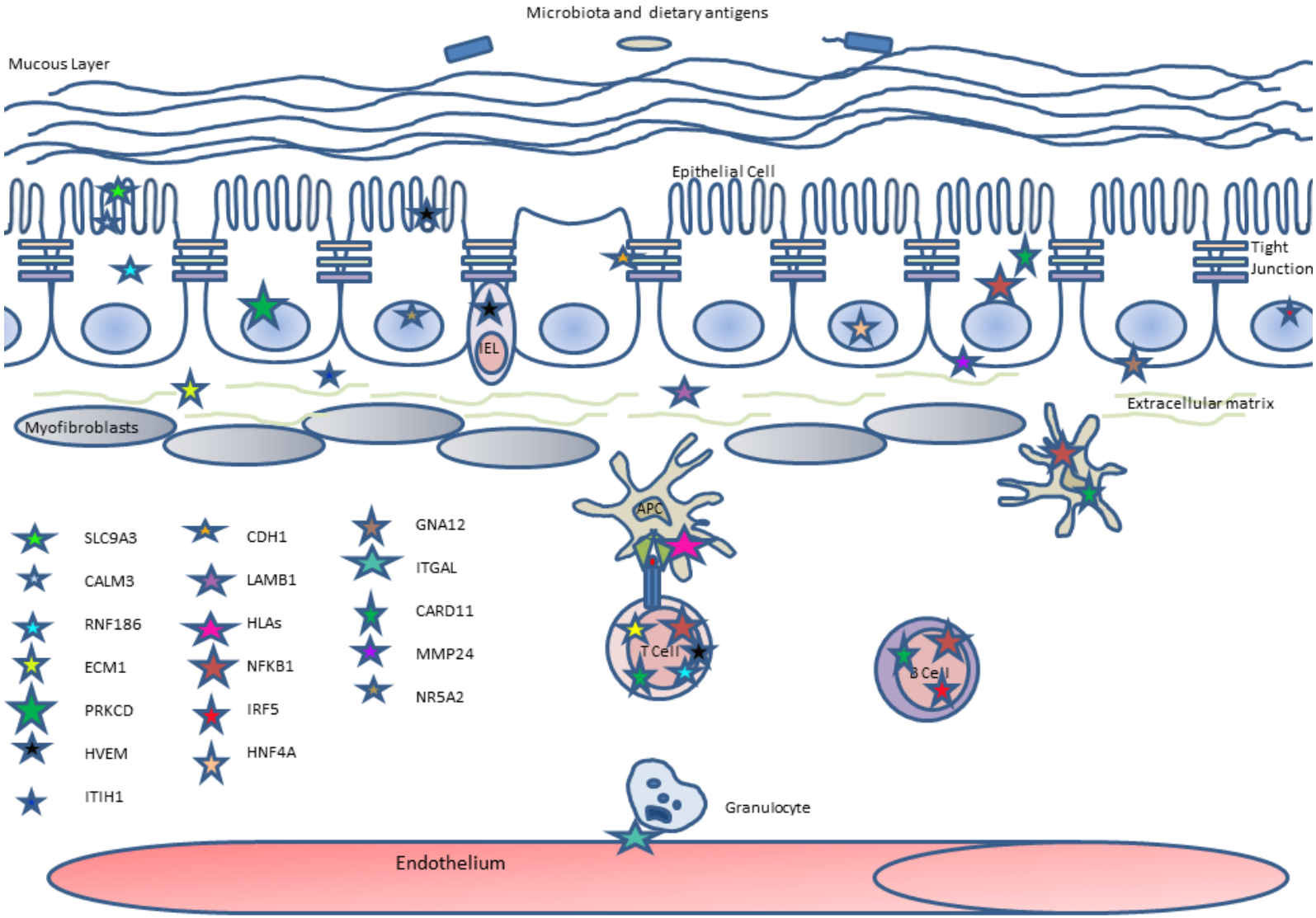


Figure 1-8 Graphical representation of UC risk associated genes and their potential site of effect

1.2.5 The role of autophagy in UC

Autophagy is the regulated self-degradative process by which there is controlled digestion of damaged and unnecessary cellular components. Under homeostatic conditions, damaged cellular components are targeted by protein adaptors which engage autophagy machinery to engulf and deliver the components to the lysosome for degradation. In conditions of nutrient starvation, autophagy is utilised to liberate energy stores, thereby promoting cellular survival. Antibacterial autophagy can be triggered by recognition of bacterial pathogens, or indicators of pathogen invasion such as membrane damage, membrane remnants, amino acid starvation, protein aggregate formation or the presence of bacterial DNA (128). Functional autophagy is required in the intestinal epithelium to maintain the secretion of mucins as well as maintenance of normal gut microflora (129). Autophagy is required in goblet cells that secrete mucins, Fc-gamma binding protein (crosslinks MUC2 to provide stabilisation) and trefoil factors (for mucosal defence). Three core autophagy proteins ATG5, ATG7 and LC3B have all been demonstrated to regulate goblet mucin secretion (130). In enterocytes, autophagy functions as part of the cellular innate immune system to restrict bacterial replication and dissemination. Functioning autophagy is also required for the maintenance of microvilli, the microscopic membrane protrusions that increase the surface area of the apical side of enterocytes.

CD has well defined autophagy risk susceptibility genes (ATG16L1, IRGM) as well as extensive examination of the role of autophagy in macrophages and dendritic cells in CD mouse models and AG16L1 models (131, 132). Conversely UC does not have the genetic markers of autophagy involvement, however deficient autophagy processes have been identified. Schuster et al (133) showed that patients with active UC have reduced chromosome associated protein D3 (CAP-D3) expression. In colonic cell lines, they showed the reduced CAP-D3 decreased autophagy and impaired intracellular bacterial clearance of intracellular *Salmonella*. Hao et al (134) have shown that a potent regulator of autophagy, Beclin 1 is expressed at a higher levels in UC patients' colonic mucosa as compared to patents with irritable bowel syndrome. Whether this is due to the inflammatory process, or whether this is an instigator of the inflammatory processes seen in UC is unclear.

Human viruses are well documented to subvert and in some cases, regulate the autophagy process; Epstein-Barr virus (EBV), hepatitis C viruses, Influenza A, Herpes Simplex Viruses and Measles virus all require autophagy to replicate and shed their viral particles (135-138). The documented human gut virome (139) is predominated by

bacteriophages, but the role of enteropathogenic viruses has been mooted previously as an instigating factor in the pathogenesis of mucosal inflammation(140). In the CD ATG16L1 mouse model, Paneth cell abnormalities are triggered by infection with a mouse norovirus, which alters the transcriptional signature of Paneth cells and the inflammatory response when the mice were treated with dextran sodium sulphate (DSS) (141). In both small paediatric and adult cohorts, enteropathogenic viruses such as rotavirus, norovirus, astrovirus or adenovirus have not been detected during follow up or relapses (140, 142) suggesting that the ongoing inflammation of UC is not due to enteropathogenic viruses. EBV and to a lesser extent CMV have been implicated in the instigation of IBD, by Lopes et al (143). In a study of 95 IBD patients with active disease and 50 healthy controls and EBV were more prevalent in the colonic mucosa of patients with IBD compared to the controls, with no difference between the inflamed and non-inflamed mucosa. EBV utilises both TLR and autophagy dependant pathways in dendritic cells to evade the immune system (144) as well as to enhance viral replication (145). EBV has been shown to infect the basolateral surface of epithelial cells via integrins and activation of adhesion molecules (146), giving credence to the presence of EBV not just in mature B cells and dendritic in the lamina propria of UC patients, but in the enterocytes also.

It is plausible, that even though core autophagy genes do not form a part of the UC risk susceptibility cohort, the core autophagy process or autophagy regulation is involved in instigation or maintenance of mucosal inflammation that characterises UC. Potential pathways for dysregulation of autophagy include downstream effects of a genetic susceptibility gene and/or via external agents such as invasive bacteria or viruses.

1.2.1 The role of the focal adhesion complex and inflammasomes in intestinal inflammation

Focal adhesion complexes (FACs) are large protein assemblies that physically link and transduce signals from the external environment to the intracellular environment via integrins (147). The function of the focal adhesion is both mechanical and responsive. It functions as a cellular anchor via integrins binding to their extracellular ligands and to the actin cytoskeleton to modify the physical features of the cell. Depending on the initiating signal, FAC can be involved in regulating inflammatory gene expression via signal transduction pathways such as interleukin 1 (IL-1) signalling (148, 149) or regulating calcium fluxes via phosphatidyl inositol signalling (150) which impact in inflammatory

cascades. Many components of the FAC are involved in downstream signalling cascades such as the MAPK/ERK pathway (151), AKT1 (152) and Wnt signalling (153). In this way, pathways affected by the FAC range from apoptosis (154), production of cellular protrusions (155) to cell cycle progression (156)and cellular proliferation (157).

A major component of the FAK is the focal adhesion kinase (FAK). Activation of FAK is necessary to maintaining and repairing the epithelial barrier in cell culture via tight junctions (158). Lipopolysaccharide induces tight junction permeability via FAK/MyD88/IL1 receptor pathways (159, 160). In addition to this, FAC GTPases such as RAC1 (161) and tyrosine phosphatase members of the FAC have a role in the regulation of the NACHT, LRR and PYD domains-containing protein 3 (NLRP3) inflammasome (162), which mediates the release of IL-1 and IL-18 from cells. IL-18 signalling drives the breakdown of barrier integrity in murine models of UC. Intestinal cell inflammasome activation occurs in pathogen recognition via integrin signalling though the FAC. Impaired inflammasome signalling leads to increased sensitivity of murine intestinal mucosa to mild chemical disruption in DSS models of colitis (163).

Although neither the focal adhesion complex, nor the NLRP3 inflammasome have been implicated in UC in the GWAS, integrin genes have been (164). GWAS and meta-analysis identified variants at *ITGA4,* ITGB8 and PLCG2 and *SLAMF8* (as identified by Peters et al 2017). It is plausible that by using a network and systems biology approach, non-trivial regulatory pathways affected by UC associated SNPs that may impact on the function of the FAC or inflammasomes can be identified.

1.3 The evolution of this PhD project

This body of work has evolved significantly from the initial PhD project which was to evaluate the role of UC associated SNPs in Extracellular Matrix Protein 1 in the maintenance of the epithelial barrier via a dysregulation of matrix metalloproteinase 9 function. The first 12 months of the PhD included work identifying co-localisation of ECM1 and MMP9 in epithelial cell lines and early work in genome editing to create ECM1 knock out cell lines. At this point, new deep sequencing results were released to myself and Dr Mark Tremelling by the UKIBD Genetics Consortium that identified that the ECM1 SNPs I was investigating were not UC associated. I made a decision not to pursue the ECM1-MMP9 interaction any further. The field of SNP epistasis and non-coding SNP function annotation was nascent but very interesting, so I approached Dr Tamas Korcsmaros with a concept of utilising SNP-SNP interactions to stratify SNPs for experimental validation. Under his guidance, I manually created a SNP analysis workflow to identify the function of UC associated SNPs and interactions between them, which underwent multiple iterations and updates as the databases were updated or changed. This is outlined in Chapter 2. Over the next year, having identified SNPs for experimental validation utilising the techniques I outline in Chapter 3 it was clear we could take the bioinformatics pipeline further to advance the goal of having personalising medicine based on a patient genotype. This final step is outlined in Chapter 4 and would not have been possible without the collaboration of Dr Dezso Modos from Professor Andreas Bender's cheminformatics group (Department of Chemistry, Cambridge), Dr Miles Parkes (IBD Research Group, Cambridge), Dr Jeff Barrett and Dan Rice at the Wellcome Trust Sanger Institute.

1.4 Aims and Objectives of the PhD

UC is a chronic debilitating disease characterised by ulceration of the colonic mucosa leading to symptoms of profuse bloody diarrhoea, abdominal pain, and fatigue and weight loss. The aetiology of UC is thought to be due to a disordered immune response to a microbiota signal in genetically susceptible hosts. Advances in genetics have led to the identification of single nucleotide polymorphisms (SNPs) associated with chronic diseases such as inflammatory bowel disease (IBD). Further advances in fine mapping and deep sequencing have narrowed down these associations; however, the assessment of the physiological function of these SNPS is still in its infancy.

The overall aim of this PhD is to use a multidisciplinary approach to determine the function of UC associated SNPs, to help understand the role of SNPs in the pathogenesis of UC in general and in a patient-specific context.

The specific objectives of this project are as follows:

1. To annotate UC associated SNPs and identify network interactions to create a UC interactome (UC-Ome) (Chapter 2);
2. To experimentally validate a stratified SNP identified as relevant within the UC-Ome (Chapter 3);
3. To identify patient-specific pathogenic pathways to disease from their genotype (Chapter 4).

2. From Genes to Disease: A network medicine approach to UC

Overarching Aim :

To annotate UC associated SNPs and identify network interactions to create a UC interactome (UC-Ome)

2.1 Introduction

2.1.1 Cellular connectivity

The human cell is complex, containing hundreds of thousands of interacting components from large complex proteins, DNA, RNA, to small molecules, lipid and carbohydrate messengers, and elements such as calcium. One facet of this, the protein-protein interactions, is described as the human interactome, which is estimated to contain more than 100,000 individual proteins(165), with each of these interacting with each other. The human interactome contains experimentally validated protein-protein interactions for the open reading frames for 17,500 unique genes (HuRI-CCSB – unpublished). A further facet of the complexity of the human cell is the extensive regulatory interactions between RNA molecules e.g. miRNAs regulating messenger-RNA, and protein-DNA interactions e.g. transcription factor activation and repression of genes. Given this level of complex interconnectivity it is no wonder that specific gene abnormalities can cause phenotypic perturbations in downstream molecular pathways. There is, however, significant redundancy in pathways such as cell cycle and regulatory signalling pathways. Damaging mutations that affect expression of key members of these pathways such as p53 or members of the Wnt pathway are seen in colon cancer but the phenotypic impact of individual single allele genetic variations in the same pathways are difficult to identify and even more difficult to experimentally validate.

A global view of how multiple single nucleotide polymorphisms (SNPs) associated with disease fit within the human interactome is required to identify common pathological pathways associated with disease. However, this requires an in-depth analysis of the function of individual disease associated SNPs using functional genomics and network biology.

Functional genomics is a term used to describe the use of genome-wide assays such as genome sequencing, and transcriptomics to study gene and protein function which focuses on the dynamic aspects of transcription, translation, regulation of gene expression and protein-protein interactions. The use of bioinformatics allows insight into complex dynamic cellular activities enabling stratification of proteins of interest. Integrating functional genomics with bioinformatics, and network biology allows identification and visualisation of each individual SNP impact within the context of a larger cellular interactome. By this mechanism, specific combinations of SNPs can be identified that in combination impact on strategic pathways for the disease of interest. This is known as network medicine.

A network at its simplest is defined as elements, or nodes, that are connected by links, or edges. We utilise networks in every-day life, from networks of work colleagues, to social networks accessed by social media. In these examples we, as individuals, are nodes. The edges are our link with others. The number of edges (links) to a node or individual is called a degree, so the more 'social network friends' we have, the higher our degree. The edges can be directed or undirected e.g. the network can show that there is an interaction present (two people are social media friends) or the direction of the interaction (person A sends a message to person B, but not the other way round).

Biological networks are not random, they will have a large number of nodes with very few connected neighbours, and a small number of nodes that have a huge number of connecting neighbours. These highly 'sociable' nodes are called hubs. An analogy of this would be UK airports. The UK has multiple airports with just a few connecting destinations, and a few airports e.g. those in London, with a huge number of connecting destinations. The London airports would be considered hubs.

To identify the most important nodes within a network, indicators of centrality (e.g nodes that are central to the working of the network) can be used. Degree centrality is an indicator of the number of edges or connections a node has; therefore, the highly social nodes or hubs will have the highest degree centrality, like the person with the greatest number of social media friends described above. In a biological context, an abundant transcriptional regulator such as STAT3 or NFKB1 will have the highest degree centrality as they interact with a large number of proteins and genes. Degree centrality only gives information regarding the number of edges a particular node has, it will not give information regarding the importance of that node to the network. For example, an individual node with six connections that sits separately from several connected nodes

will have a higher degree centrality than each of the connected nodes which only have two or three edges, however from a biological pathways perspective, the connected nodes are more important. To use the airport analogy again, if a traveller needs to get from Scotland to New York, then the airport in Scotland which connects to 10 other local destinations but nothing further afield will have a high degree centrality but is not important to the traveller (and therefore not important to the network). The airport with 1 destination that will connect him to a New York airport is more important to the traveller and therefore the network.

For an indicator of the node with the most amount of information going through it, e.g. the node that created cohesiveness within a network by connecting different parts of the network together, or joins the network together, a measurement called betweeness centrality is used. In graph theory terms betweeness centrality is used to indicate the node that has the largest number of shortest paths running through it. The higher the betweeness centrality, the more 'influential' that node is to the network. An example of this would be a party planner who controls each aspect of a wedding from the flower arrangers, to the musicians and the caterers, all of which are different groups, but the planner provides the cohesiveness to create the wedding.

A biological example of this would be an adapter protein that gets phosphorylated by a MAPK pathway which then activates the apoptosis pathway and inhibits autophagy. This adapter protein links three distinct components of the network and in order to get from one part of the network to the next, the information has to go through this node. The node described is Stat3, therefore a node can have both high degree centrality and betweeness centrality, but they are mutually exclusive and the measurement confers import depending on the question asked.

The building and interpretation of robust biological interaction networks is dependent upon three major factors;

1. High quality multi-omics data
2. Creation and curation of interaction databases
3. A multidisciplinary team with experimental and computational skillsets

With the advent of technologies allowing cost effective and rapid genotyping, transcriptomics, proteomics, lipidomics and DNA methylomics there is a wealth of data available associated with specific disease states. Cancer is a model for the use of network medicine to identify novel biomarkers, underlying molecular mechanisms of disease or

genetic 'fingerprints' of disease (166) and drug repurposing. However, identification of disease pathways of interest in autoimmune diseases such as Diabetes (167) and Juvenile Idiopathic Arthritis (168) using multi-omics datasets has also been successful. Using multi-omics techniques and visualising the results within a network has the potential to identify the interaction between SNP affected proteins, and SNP- SNP interaction, or SNP epistasis.

The limitations of using bioinformatics include the dependence on experimentally validated datasets for example the lower the quality of the data, the increasing risk of false positive or false negative results and random network generations.

2.1.2 SNP epistasis

Epistasis is defined as the interaction between non-allelic genes such that the phenotypic effect of genes can change depending on their combination. SNP epistasis explores the phenotypic impact between SNPs. A common approach is to undertake pairwise analyses of SNPs but this requires immense statistical power in terms of cohort numbers and numbers of SNP pairs (127, 169). Other approaches have used logic regression to identify higher order interactions, which still uses a pairwise approach, but identifies groups or trees of SNPs with interactions. This has been utilised within the UC cohort and identified potential epistatic mechanisms between 4 SNP groups: 1. HLA.DQA1.B1/DRA.B1 and UBQLN4/RIT1 or IFNG/IL26/IL22; 2. FNG/IL26/IL22 or REXO2; 3. IL23R/IL12RB or REXO2 or GPR35 and 4. MST1 or near KIF11 and near USP25 (52). The phenotypic effect of the epistasis has not been identified. Phenotypic analysis requires a gene by gene approach, as undertaken by Diegelmann et al (170) who identified that IL23R variants influence DMBT1 expression and that DMBT1 variants have altered transcription factor binding. The gene by gene approach requires a stratification strategy for candidate SNPs.

2.1.3 SNP functional annotation:

Two major questions in SNP research are which are the causative SNPs and what do they do? Fahr et al (171) identified putative causative SNPs in coding and non-coding disease variants in 21 autoimmune diseases, using high density genotyping and epigenomic data mapping causal variants to create an tissue enriched predictive resource. Immunochip was designed to finemap previously identified risk associated SNPs, and meta-analysis has been used to further hone potential causative SNPs in IBD (53, 172). This research aims to build on the wealth of data accumulated in the two ground-breaking papers from Fahr et al(171) and Jostins et al (39) in an attempt to provide evidence for the question of what SNPs do.

Approximately 10% of IBD associated SNPs (UC/CD and IBD overlap) are within protein coding regions (non-synonymous/missense or synonymous SNPs), the majority (90%) are within introns, intergenic regions or regulatory regions.

Tools such as SIFT (173) or SNPEff (174) utilise structural analyses to determine the predicted outcome of missense SNPs either tolerated or deleterious and SNP nexus (175), which functionally annotates individual SNPs into site and predicted effect on transcriptome/proteome based on cancer data. In IBD, there has been a lot of interest in exonic SNPs in *ATG16L1* (176-201), *MST1* (195, 207), *IL23R* (176, 178, 181, 208-213), *IRGM* (177, 179, 180, 183, 190, 191, 194, 195, 214-218), *NOD2* (184, 185, 189, 196, 201, 219-229), *CARD9* (229-235), *RNF186* and *PRDM1* (231, 236-240).

As missense and synonymous disease associated SNPs account for approximately 10% of GWAS SNPs, analysis of SNPs in non-coding regions is important. Enrichment of risk SNPs in active regulatory elements in non-coding DNA such as promoters and enhancers have been used to create refined hypothesis for the genetic predisposition of disparate diseases such as Parkinson's disease (241), schizophrenia (242), systemic sclerosis (243) and prostate cancer (244).

Non-coding SNPs are potentially within sites of splice sites, microRNA (miRNA), miRNA binding sites (miRNA-BS), transcription factor binding sites (TFBS) or long non-coding RNAs (lncRNAs).

MiRNAs are small non-coding RNAs, which bind to *cis*-elements in the 3'untranslated region (UTR) of target mRNAs to fine-tune target gene expression. MiRNAS are generated by a two-step process: pre-miRNA (hairpin like partially duplexed) from pri-miRNA by the drosha/DGCR8 complex in the nucleus and mature miRNA from pre-miRNA by the DICER/TRBP complex in the cytoplasm (245). For translational suppression, base pairing between the 'seed' sequence of miRNA (nucleotides 2-7 or 2-8 at the miRNA 5'end) is required with the target mRNA. The miRNA/mRNA interaction guides RISC for translation inhibition (246). MiRNA signatures in colon, blood and saliva are able to differentiate between CD and UC indicating a differing miRNA regulation or genetic impact between the two IBDs (Schaefer 2015).

SNPs could impact on the hairpin structure guided miRNA processing, have thermodynamic effects on strand loading as well as causing a change in the seed sequence

or shift in the processing sites that could all result in a change in production of mature miRNAs, change the mRNA targets that the seed miRNA can bind to, change the affinity of the binding or create novel miRNA with different targets (247).

There is evidence that SNPs within conserved miRNA binding sites are deleterious. This is due to the specific Watson and Crick pairing of miRNA to mRNA, a SNP within the 3'UTR of mRNA would impact on miRNA binding (248). This has been shown in schizophrenia associated SNPs (249) and Coronary Heart Disease associated SNP rs4846049 (250). There are 458 papers identifying miRNA binding site changes caused by SNPs, however there is only one paper identifying a change in a miRNA binding site for IBD SNPs – a SNP in IRGM which alters the binding site to mir-196 in CD (251).

MiRNAs are involved in intestinal epithelial homeostasis as demonstrated in the Dicer1ΔIEC mice where the deletion of Dicer abolishes all miRNA function in intestinal epithelial cells. These mice display a phenotype with a reduction in goblet cells, and increase in inflammatory immune cells, disorganised intestinal architecture with associated increased intestinal permeability, decreased mucus production and decreased Th2 cytokines with IBD symptoms and inappropriate Th1 responses during infection. With regard to the epithelial barrier the Dicer1ΔIEC mice exhibit a weakened epithelial barrier associated with a decrease in claudin expression (252).

As with colonic cancer, there is differential expression of several miRNAs in IBD as compared to controls and this has been assessed both by colonic biopsy and in peripherally circulating miRNAs (253).

Transcription factors are proteins that bind to DNA usually in the promotor region, upstream and close to the transcription start site of a target gene. They regulate the expression of the gene by activating or inhibiting the transcription machinery. These promotor sequences containing TFBS have some conserved structural properties including stability, bendability and nucleosome position preference and curvature. Changes in TFBS by SNPs have been assessed for *ADRBK1, AKT3, ATF3, DIO2, TBXA2R* and *VEGFR* genes and hypothesised to be associated with a variety of disease phenotypes such as diabetes, high altitude sickness, asthma and hypospadias (254). From a pubmed search in 2017, there are 582 papers identifying alterations in TFBSs by SNPs across a wide variety of disease and across a variety of species. In terms of IBD; in CD DMBT1 variants have been shown to have altered transcription factor binding sites (170). In both CD and UC

PTPN2 SNP rs2241879 has been predicted to alter transcription factor binding sites to multiple inflammatory transcription factors including NFKB and GATA-3(255). The IBD associated NKX2-3 variant rs11190140 has been shown to have altered NFAT binding to its promoter compared to the non IBD associated allele (256). *In Silico* analysis of IBD SNPs to undertake functional annotation have identified 18 noncoding regulatory SNPs in known transcriptional factor bindings sites which are hypothesised to dysregulate expression of nearby genes including *PLC1, ANKRD55, BACH2, CCDC26, CREM, FADS1, FOSL1, SMAD3, PRKCB, IKZF3, DNMT3B, CD40* and *UBE2L3 (257)*.

Long non-coding RNAs (lncRNAs) are non-protein coding RNAs >200 nucleotides in length which are typically transcribed by RNA polymerase II (258). They participate in the regulation of gene expression via transcriptional and post transcriptional mechanisms, regulation of proteins post translationally, organisation of protein complexes and cellular signalling (259, 260). The GENCODE project (258) collected over 10,000 human LncRNA genes. The role of these is still being characterised, however certain conserved lncRNAs have been shown to have a role in immune regulation. lncRNA secondary structure disruptive SNPs have been identified within or in close proximity of IBD loci associated candidate genes with tissue specific expression patterns (261).

The generation of mature RNA for protein translation requires the removal of non-coding intronic regions from precursor mRNAs and the ligation of exons. This removal of intronic regions and ligation of exonic regions is called splicing and is an essential step in gene expression in eukaryotes. Splicing is performed by large protein-RNA complexes called spliceosomes. Splicing occurs when the splicesomes recognise specific splice sites. The splicing process is regulated by proteins within the splicesome, splicing factors and RNA sequence elements including the core splicing motifs and splicing enhancer or silencer sequences.

Splice site consensus sequences are located at the ends of introns. The transition of exon to intron is known as the splice donor site, and the transition from intron to exon is known as the splice acceptor sequences. The highly conserves splice sites are characterised by having GT (donor) or AG (acceptor) dinucleotides at the intron ends.

The splicing process can be disrupted by mutations in *trans*-acting splicing factors or *cis*-acting sequences in introns and exons. *Cis* acting mutations include those that disrupt the constitutive splice sites or disrupt the regulatory sequences (enhancers or silencers).

Splice site destruction can result in deletion of the adjacent exon or retention of the adjacent intron. A SNP in BTNL2 donor splice site (rs2076530) has been identified in sarcoidosis leading to a frameshift and a premature stop codon truncating the protein, however it is not of the same import in other granulomatous diseases(262). In IBD, the *CARD9* SNP rs4077515 (230) affects a splice site and is protective for IBD.

There are a variety of tools which offer SNP analysis. Var2GO uses a SNPeff platform and applies gene ontology to the genes local to the SNP. FASTSNP analyses SNPs to identify protein changes, transcription factor binding sites in promoter sequences or enhancer regions and alternative splicing regulation, preferentially identifying phenotypic candidates with a low minor allele frequency given the strong selective pressure against strong phenotypes. SNP Function Portal analysis includes analysis of genomic elements, transcription regulation, protein function, pathway, disease and population genetics. However, given that it is over 10 years old, it utilises HapMap Phase II so is now obsolete.

In recent years, there has been an upsurge in mathematical modelling based and experimentally validated tools for identifying changes individually in protein binding motifs (Eukaryotic Linear Motifs (263)), miRNA seed sequences (miRBASe, mirDB, Tarbase), miRNA binding sites (Tarbase, miRANDA), transcription factor binding sites (JASPAR), splice sites (Human Splicing Finder 3.0 (HSF) (264), Max Entropy Scan (MES) (265) Alternative Splicer Site Predictor (ASSP) (266)), and long non-coding RNA sequences as well as curated protein-protein interaction databases ((267)) which together allow for in-depth analysis of multiple disease associated SNPs.

At the time of writing there is no single tool which brings all of these modalities together with a network visualisation tool such as cytoscape and downstream gene ontology which allows the ability to assess the network for novel disease pathways downstream of functional SNP sites.

2.2 Hypothesis, Aims and Objectives

I hypothesised that utilising the wealth and breadth of bioinformatics tools available we could functionally annotate the UC associated SNPs and using a systems biology approach we could identify disease associated pathways. Given that UC is thought to be a disease of the intestinal epithelial barrier, I further hypothesised that the UC associated SNPs would be involved in integral epithelial barrier pathways including tight junction maintenance, autophagy or the focal adhesion complex.

The aim of this project, therefore, was to identify UC associated SNPs that had a role in tight junction maintenance or regulation, autophagy, and the focal adhesion complex, with a view to identifying candidates for experimental validation and shedding light of pathogenic processes associated with UC.

The objectives were:

1. Identify the predicted and experimentally validated regulatory (splicing, miRNA, transcription factor, long non-coding RNA) effects and protein-protein interaction effects of UC associated SNPs
2. Create a UC interactome network
3. Identify components of the network associated with tight junction maintenance and regulation, autophagy and the focal adhesion complex to enable SNP prioritisation for experimental validation.

2.3 Methods

2.3.1 Identification of UC associated SNPs

A glossary of databases and technical terms can be found before the appendix.

I identified UC associated index SNPs from either Jostins et al(39)supplementary data 2 which passed the GWAS threshold for significance ($p= 5 \times 10^{-8}$), or Farh et al 2014 (166) dataset pulled directed from the Broad Institute repository, which were enhancing to the colonic mucosa. Using the Fahr et al dataset, I identified finemapped SNPs with the highest 'PICS' value to index SNPs to all the index SNPs from the Broad and Jostin's datasets. The finemapped SNPs were included if they enhanced to the colonic mucosa. This SNP list is called the 'parent cohort' (PC).

I also identified index SNPs and finemapped SNPs not enhancing to any tissue from the Fahr et al dataset, as well as IBD associated SNPs from Jostins et al dataset, to be used to broaden the network. This SNP list in addition to the parent cohort is called the 'enhanced parent cohort' (EPC).

If there was no finemapping available for an index SNP (e.g. the Immunochip finemapped SNP had an $R2<0.8$) then the highest proxy partners (based on tightest linkage disequlibirium and distance) were assessed using a SNAP proxy search (268) and were included in the analysis.

Each SNP was annotated using Ensembl (269) from the rsID using GRCh38.p7.

Risk alleles were taken from the original data source. For the broad institute finemapped SNPs, the risk allele was annotated as the non-ancestral allele from dbSNP(263), and the non-risk allele as the ancestral allele.

The SNPs and alleles used as well as their source can be found in Appendix 2a.

The workflow from SNP data to network can be seen in overview in Figure 2-1. Each individual method for each segment of the workflow is described below.

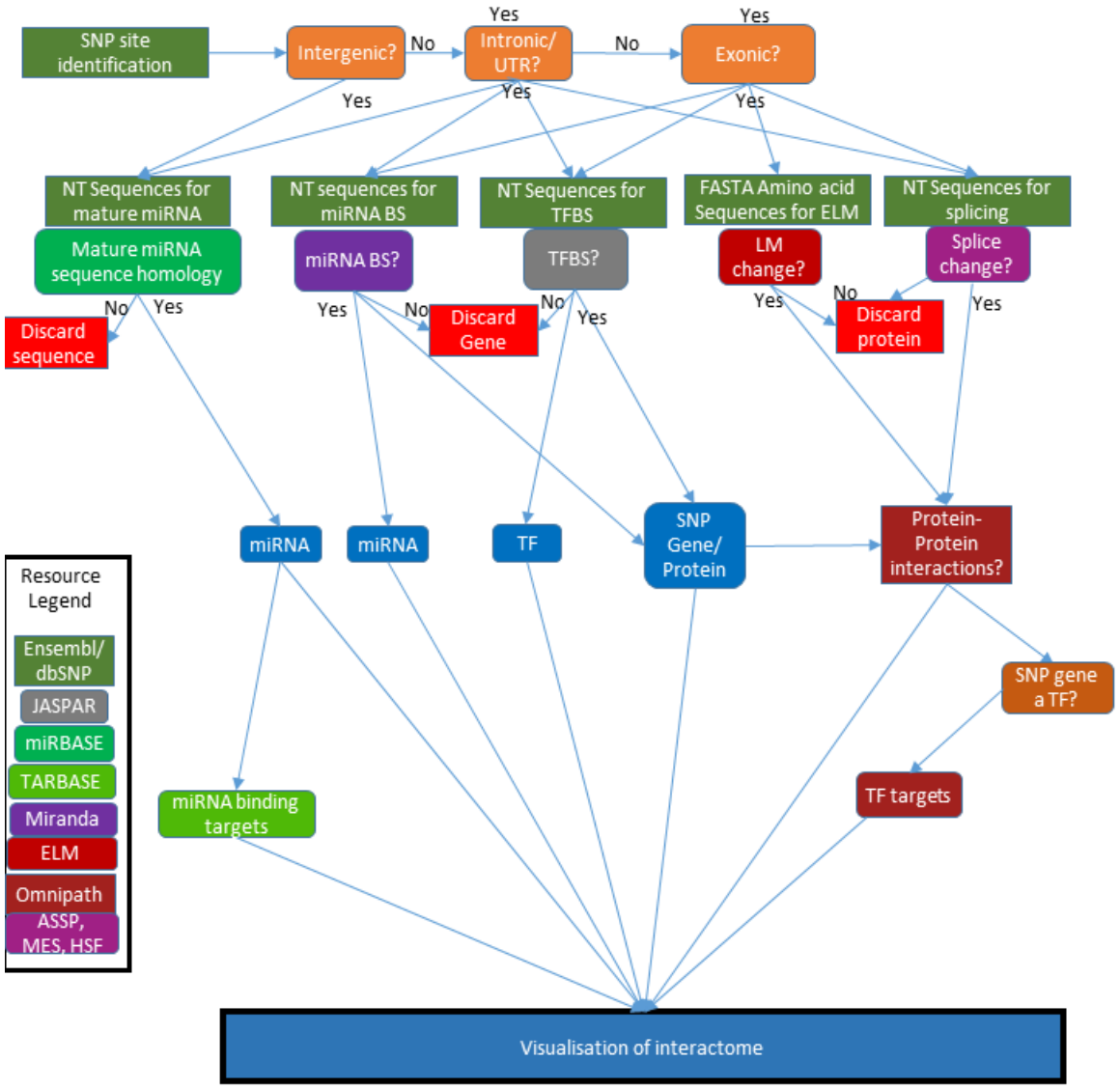


Figure 2-1 SNP workflow logic flow chart from SNP site identification, through functional annotation using multiple databases to visualisation of the interactome.

2.3.2 Identification of missense SNP effects

I extracted SNPs annotated to be missense in Ensembl(269) from the extended parent data set. Each rsID was input into PolyPhen2 (270, 271) to ascertain a prediction of how deleterious the change was.

Using linkouts (NP or XP identifier) from Ensembl (269) and dbSNP (263) I extracted the FASTA amino acid sequence for both the risk allele and the non-risk allele. The full amino acid sequences for the proteins were input manually into the stand alone web-based service ELM (263). The motif cutoff was 100. The output for risk and non-risk allele was compared by eye in excel and any differences identified at the amino acid site. Changes to binding sites or enzymatic functions were highlighted and the corresponding protein target was identified. Each change was annotated as a loss or gain with reference to the non-risk allele. For effects associated with autophagy 'LIR' motifs, these were further analysed in iLIR (272), a web based service for identifying and scoring LIR motifs.

Data mining to confirm literature evidence of identified protein-protein interactions was undertaking using STRING v10 (273), with interaction score set to high confidence (0.7) and no more than 50 interactors in the first shell. I looked for text mining, experiments, databases, co-expression and co-occurrence. I excluded neighbourhood and gene fusion functions.

Deleteriously affected protein names and their binding partners were converted to Uniprot (274) identifiers and input into the UC Network in Cytoscape (275) v 3.3.0. The first neighbour proteins (downstream binding partners) were not identified as this was already a protein-protein interaction.

2.3.3 Identification of splice sites, splice enhancing and silencing motifs affected by SNPs

The nucleotide sequence containing the SNP risk or non-risk allele +/- five nucleotides from intronic, synonymous and missense SNPs were compared against 606 curated hexamer exonic splice enhancing (ESE) motifs (annotated as CHESEL) collated and provided by Dr Wilfried Haerty, Earlham Institute (Appendix 1). This initial run identified there were sites of interest therefore this avenue was examined further.

It is best practice for splice site analysis to use at least three different programmes, therefore I used a combination of Human Splicing Finder 3.0 (HSF) (264), Max Entropy Scan (MES) (265) – based on the HSF web based server (http://www.umd.be/HSF3/HSF.html) with Alternative Splicer Site Predictor (ASSP) (266), another web based server (http://wangcomputing.com/assp/index.html). The combination of HSF/MES/ASSP has been shown to have the highest performance based on receiver operative curve analysis (276). For exonic splice enhancers, exonic splice silencers and branch points; HSF was used.

The HSF analysis was done using the corresponding Ensembl transcript ID for the rsID. If there was more than one transcript available, the transcript corresponding to the parent gene was used first. If multiple transcripts from different genes were present corresponding to one SNP site – all transcripts were run if available.

There can be a difference between the ensemble ancestral allele and minor allele, and the UC denoted risk and non-risk alleles. If there was a difference, each allele was run against the ancestral allele in HSF. If the risk /non risk alleles were concordant with the Ensembl alleles then just these alleles were run. Default parameters were used for the HSF server utilising the mutation analysis function.

Transcripts that were identified as having an alternative splice site were analysed using ASSP. The same Fasta sequence for transcripts used in HSF were used in ASSP with the 'wild type'(ancestral), the risk or non-risk alleles being present and analysed. The results were analysed with regard to concordance or discordance with the HSF and MES results. The translated transcript (e.g. protein) was included into the downstream network as 'splice site affected' if there was 2/3 concordance between the three programmes, even if the confidence value or variation value differed.

Transcripts were converted to their protein name Uniprot ID.

lncRNAs were excluded from the analysis as were any transcripts which were unavailable in cDNA format (e.g. were in n. format). Statistics were undertaken in GraphPad Prism.

OmniPath (267) interactions were downloaded on 25/01/2017 and imported into Cytoscape v3.3.0. The transcripts affected by splicing sites or splicing regulation were converted to uniprot identifiers and were used as search terms in the OmniPath Cytoscape Network to identify downstream directed protein-protein interactions. By using a directed network upstream interactions such as transcription factors for SNP affected genes were specifically excluded, as these would not be biologically relevant to the SNP function and impact.

2.3.4 Identification of mature miRNAs affected by SNPs

For all the intronic, intergenic, downstream gene variants (DGV), upstream gene variants (UGV), 5'untranslated region(UTR) and 3'UTR SNPs, flanking sequences of 10 nucleotides either side of the SNP risk allele or non-risk allele (or ancestral allele if no non-risk allele was identified) were retrieved from dbSNP. These were then manually input into miRBase (277) standalone web browser. Using the BLASTN facility (E cut off of 10) and SSEARCH facility (E cut off 10), the 21 bp sequences were analysed for sequence homology with known human mature miRNAs. Only homology sequences where the presence of the risk allele caused a loss of sequence homology, as compared to the non-risk allele sequence were considered biologically relevant. A literature review of the identified miRNA was undertaken in 2014 and repeated in 2017 using PubMed and searching for the mir term e.g. mir-1266 filtered to human species.

miRNAs would only be included if 2 out of 3 of the methods used were concordant or suggestive. When further stringency was required 3/3 concordance was utilised.

The human mature miRNAs identified were then input into miRDB(278) to identify predicted miRNA target genes, initially only those with a target score of >80 were put forward for inclusion into the UC network. On creation of the network, further stringency was required, therefore a higher cut off point of target score >95 was chosen.

A miRNA homology network was created in Cytoscape 3.3.0 using, for ease of reading, the gene in which the SNP resides (or intergenic or lncRNA) as the source node and the miRNA predicted targets from miRDB as the interaction nodes. The first layout was perfuse force directed to give an overview of any networks that were formed. On identifying two separate networks, these were separated and gene ontology (biological function in homo sapiens) analysed via BinGO (279) (gene ontology app in Cytoscape) and Genome scale

Integrated Analysis of Gene Networks in Tissues (GIANT) (280). The larger network was further analysed identifying nodes which had 2 or more source or interacting nodes interacting with them. This network was then visualised in an attributable circle format. Network statistics were performed within Cytoscape on the larger network.

The SNP sites were then compared against the known genomic site in Ensembl for the identified mature miRNA.

2.3.5 Identification of miRNA binding sites affected by SNPs

The 22bp sequences of mature miRNAs were retrieved from miRBase(277). The 10 base pair flanking sequences of all the risk alleles and non-risk alleles were assessed for the presence of miRNA binding sites using the web based tool miRanda (281-283). Hits predicted to occur in the seed region (2'-8') of the miRNAs and with pairing scores >= 150 and energy threshold <= -7 kcal/mol were considered as significant. Further stringency was required therefore the threshold of >155 pairing score and energy threshold of<-20 kcal/mol were used. Other parameters were set to default settings. The outcomes were denoted as a loss , a gain of a putative binding site or a neutral results relative to the non-risk allele e.g. the risk allele caused a loss of binding site that was present with the non-risk allele sequence. A neutral result was found when the binding site remained regardless of risk or non-risk allele.

Analysis of conserved miRNA binding sites in target mRNA was undertaken in miRanda, using the target mRNA search function identifying conserved miRNAs with good mirSVR scores, then non conserved miRNA with mirSVR scores.

Genes affected by changes in miRNA binding sites were converted to their protein Uniprot identifiers and used as search terms in the OmniPath Cytoscape programme as described above. Direct outgoing first neighbours of the affected genes were input into the UC network in Cytoscape with the translated proteins.

2.3.6 Identification of transcription factor binding sites affected by SNPs

Each risk or non-risk allele plus 50 base pair flanking sequences were compared individually against binding profiles represented by Position Specific Scoring Matrices (PSSMs) corresponding to 140 human transcription factors from the JASPAR database (284) using the nucleotide ambiguity code to identify transcription factor binding motifs which covered the SNP site individually. Coverage of the Jaspar Database increased from 140 to 396 human transcription factors in 2016 (283), so one by one analysis was unfeasible. With the assistance of Dr P Sudhakar the PSSMs downloaded in JASPAR format were converted to the TRANSFAC format to enable easier handling of results. The Regulatory Sequence Analysis Tool (RSAT) called matrix-scan (285) was used to search for potential transcription factor binding sites in both the risk and non-risk allele sequences The background model estimation was determined by using residue probabilities from the input sequences with a Markov order of 1. Both forward and reverse strands of the sequences were searched. Hits with a P-value <= 1e-05 were considered as putative binding sites. Other parameters were set at default values.

The SNP sites were visualised in UCSC genome browser on Human Dec 2013 (GRCh38/hg38) Assembly to identify regulatory elements for the associated genes (from ORegAnno) and to identify if the SNP occurs within an enhancer area. If the SNP occurred within an enhancer, the enhancer score was identified from GeneCards and the known transcription factor binding sites within the enhancer compared against the SNP prediction.

2.3.7 Creation of the UC interactome

The data was imported into Cytoscape, using ID of effecting gene as source node and ID of SNP effected gene as the target node, with the type of interaction being the interaction edge. The name of effecting gene and effected gene as well as loss/gain data was imported as edge attributes. Each subtype of interaction e.g. miRNA BS, TFBS was demarcated by a change in shape, colour or border colour as per the legend. The UC network was demarcated as UC in node table column to separate it easily from the UC interactome, the 'affected genes' were demarcated as affecting genes in a separate column to ensure we only identified first neighbours of the affected proteins, not any effecting proteins in the network.

Omnipath was imported into a new network (with direction as edge attribute) and this network was merged in union format with the UC SNP network, keeping the same style.

First nodes from UC column and their first neighbours were selected from the network, then effected proteins and their first neighbours were selected from the network, then effected proteins and directed outgoing nodes were selected from the network. Nodes that were not downstream first neighbours of the effected gene (not effecting gene) were removed to create the UC interactome. First neighbours were demarcated by a change in label colour.

2.3.8 Subnetwork identification

The nodes for the Autophagy Regulatory Network(ARN) were obtained from http://autophagy-regulation.org. These were compared against the UC Interactome network and the Autophagy_UC subnetwork was created. Nodes were only included if they were a UC node or first neighbour of a UC node.

The nodes for the focal adhesion complex were downloaded from the Adhesome (286) in November 2015.

Tight junction proteins, their regulators and maintenance nodes were obtained directly from Dr Emily Jones, collated as part of her PhD.

2.4 Results:

2.4.1 SNP Effectors

In the Parent Cohort (PC) analysis of 60 SNPs associated with UC (UC), 52 (88%) had fine mapping within the Broad Institute dataset and were enhancing to the colonic mucosa. 7 were finemapped on immunochip only and therefore we did not have tissue enhancing information for them. 1 SNP had no finemapping and no proxy (rs6927022). For the extended analysis, a further 66 finemapped SNPs which had no tissue enhancement but were UC associated and 36 IBD associated SNPs were included, this added a further 103 SNPs to complete the total of 163. The breakdown of annotation can be found in Table 2-1.

Annotated Consequence	Numbers in Parent Cohort (number of index SNPs)	Numbers added by Extended Parent Cohort (number of index SNPs)	% Total in PC /%Total in EPC	Overall Totals PC/EPC
Intronic	21 (16)	63 (37)	35% /51%	
Downstream Gene Variant	2 (2)	5(4)	3.33%/4%	
Upstream Gene Variant	16 (7)	5(4)	26.66%/12%	
Intergenic	4 (4)	15 (10)	6.66%/12%	**90% 'non coding'/91.6% non coding**
3'untranslated region	2(2)	3(3)	3.33%/3%	
5'untranslated region	1(1)	1(1)	1.66%/1.2%	
Regulatory region variant	5(2)	2(2)	8.33%/4%	
Non-coding transcript variant	2(2)	0	3.33%/1.8%	
Synonymous	3(3)	3(3)	5%/3.6%	**10% coding /8.4% coding**
Missense	3(3)	5(5)	5%/4.8%	
Within lincRNA	21 (8)	8(6)		**35%**/17%

Table 2-1 Breakdown of annotations within the Parent Cohort and added by the extended parent cohort.

2.4.2 UC associated missense SNPs affect linear motifs within proteins

There were three missense SNPs in the Parent Cohort (Table 2-2);

- Thr139Met in G-protein Coupled Receptor 35 (GPR35)
- Leu333Pro in Interleukin17 Receptor E Like (IL17REL)
- Gln1042His in Regulator of Telomere Elongation Helicase 1 (RTEL1)

Only Thr139Met in GPR35 was possibly damaging in polyphen2. In ELM, there were no ligand changes, but amino acid 139 resides within a transmembrane domain. Methionine is a very hydrophobic amino acid which is fairly non reactive, but may play a role in binding/recognition of hydrophobic ligands such as lipids. Threonine is less hydrophobic than methionine but is polar and usually found at the surface of proteins(287). There is no specific protein identified so no datamining was undertaken.

Leu333Pro in IL17REL was identified as a benign change in PolyPhen2. Leucine and proline are both very hydrophic amino acids. Proline is a unique amino acid, being an imino acid as it contains an NH2+, not an NH3+. This means that it is unable to conform to main chain structures adopted by other amino acids. It is used to form tight turns in protein structures, or introduce kinks into alpha helices. Due to this it is usually found at the protein surface and forms part of WW and SH3 motifs that are key to intracellular signalling cascades (287). This is represented in the ELM results, with a gain in WW2 motif, but losing MYND1 and SUMO motifs at the same site. MYND1 is a domain that binds proline rich motifs and has been shown to be mainly involved in protein-protein interactors in the context of transcriptional regulation (288). The SUMO ligand binds small ubiquitin related modifiers which regulate extensive protein-protein and protein-DNA interactions (289). Of interest is that there are no other MYND ligand sites on IL17REL. There is one other SUMO ligand site which remains intact in both the SNP affected and 'wild type' protein. There are no other WW2 motifs on the wildtype protein. There is no specific protein identified so no datamining was undertaken.

Gln1042His in RTEL1 was also predicted as benign by PolyPhen2. Glutamine is a polar amino acid, found on the surface of proteins. Histidine is also a polar amino acid, but it has a pKa near physiological pH , thereby is able to switch from a neutral to a positive charge, altering its preference for being in the protein core or exposed at the surface (287). RTEL1 regulates homologous recombination (290, 291), limits excessive crossing over during meiotic recombination (292) and maintains integrity of telomeres (293). Only 1 ligand change was identified by ELM, with a loss of motif which is phosphorylated by

phosphoinositide-3-OH-kinase related kinases (PIKK) family members. PIKK members,; mTOR, ATM, ATR, PRKDC, SMG1 and TRRAP are proteins with serine/threonine kinase activity with roles in DNA repair and DNA damage checkpoints (294). There are seven PIKK domains in the wild type protein, with the loss of only one it is likely to be a mild phenotype at best. None of the PIKK members were identified during data mining as associated with RTEL1.

In the extended cohort there were a further five missense SNPS (Table 2-2);

- Arg381Gln in Interleukin 23 receptor (IL-23R)
- His167Arg in the Fc Fragment of IgG Receptor IIa (FCGR2A)
- Arg689Cys in Macrophage Stimulating 1 (MST1)
- Arg225Trp in Cluster of Differentiation 6 (CD6)
- Gly353Arg in Neurexophilin and PC-esterase domain family member 1 (NXPE1)

Polyphen2 was unable to compute results on multiple occasions for rs3197999 Arg689Cys in MST1, and there was no difference seen in ELM. Arginine is a very positively charge polar amino acid, whereas cysteine can be either charged or hydrophobic, suggesting the substitution would have an impact on protein function. Hauser et al (207) have seen a gain of function in MST1 Cys689 cell systems, leading to an increased stimulatory effect of MSP on chemotaxis and proliferation by THP-1 cells. There is no specific protein-protein interaction identified so no datamining was undertaken.

Arg381Gln in the IL-23 Receptor was identified as probably damaging by PolyPhen2, with a loss of a pro-protein convertase 7 cleavage (PCSK7) site identified by ELM. Both arginine and glutamine are polar amino acids; arginine has a predominately positive charge. PCSK7 is a membrane bound calcium dependent endoprotease in the trans-Golgi network (295), therefore hypothetically it could be involved in post translational modification of the IL-23 Receptor. There are no other PCSK7 binding domains on the wild type protein, however there are 3 other PCSK family domains that remain intact in both the wild type and SNP affected protein. Due to the large IL-23R datamining result, PCSK7 was used as a search term instead and only 7 interactors were returned, none were IL-23R, reflecting a global paucity of data on PCSK7.

His167Arg in FCGR2A was identified as benign by PolyPhen2. Both histidine and arginine are positively charged polar amino acids. In ELM there was a loss of FHA domain and SH3 binding domain which was not contiguous with the amino acid site (both motifs 202-208). There was a gain of BRCA1, and GSK3 just adjacent to the amino acid site (161-165 and

159-166, respectively). Neither BRCA1 nor GSK3 were found on datamining. At the amino acid site was a gain of a LIR motif that binds ATG8 protein family members (QKFSRL 163-168). There are already four other independent LIR motifs in the wild type protein (EPPWINV 46-52, SEWLVL 121-126, DPTFSIP 169-175, DGGYMTL 285-291). iLIR(272) identified the motifs in the wild type (position 121-126) in a disordered region and with a high position specific scoring matrix (PSSM) making it a functional candidate. QKFSRL 163-168 was not in a disordered region and has a low PSSM, therefore has a low prediction to be functional. None of the ATG8 family was found associated with FCGR2A on datamining.

Arg225Trp in Cluster of Differentiation 6 (CD6) was predicted to be benign in PolyPhen2. Whilst arginine is positively charged, tryptophan is an aromatic hydrophobic amino acid. As it is aromatic amino acid it can interact with other aromatic groups that are not protein ligands. There were no ligand differences seen in ELM, but the amino acid occurs within a scavenger receptor domain. This contains the activated leucocyte adhesion molecular binding site. Scavenger receptor areas also function to recognise and remove unwanted entities e.g non-self molecules such as lipopolysaccharide (296). There is no specific protein-protein interaction identified so no datamining was undertaken.

Gly353Arg in Neurexophilin and PC-esterase domain family member 1 (NXPE1) is predicted to be probably damaging by PolyPhen2. Glycine is a very flexible small, hydrophobic amino acid, and can often be found in tight turns in structures that can give it a functional role, therefore a change in a conserved glycine could have an impact. Given that arginine is a large positively charged amino acid, one can see why PolyPhen2 identified this substitution as probably damaging. ELM identified a single gain of the WH2 motif which binds to the hydrophobic cleft in actin subdomains 1 and 3. Upon actin binding , the area forms an alpha helix, followed by a flexible loop with is stabilised by actin binding (297). STRING identified that at the high stringency cut offs for datamining, no associations were found for NXPE1, and even at the lowest stringency cut offs no association with actin was seen.

Gene	SNP identifier	Amino acid change	Polyphen2 result	Functional site	Loss/Gain	Site function	Probability
IL-23R	Rs11209026	Arg381gln	Probably damaging Score 1	CLV PCSK PC7 1	Loss	Proprotein convertase 7 cleavage site	5.09E-04
FCGR2A	Rs1801274	His167arg	Benign Score 0 Sensitivity 1, Specificity	Within in immunoglobulin domain			
				LIG FHA 1 202-208	loss	FHA domain is a signal transduction module, prevalent in nuclear proteins	8.66e-03
				LIG BRCT BRCA1_1 161-165	gain	Phosphopeptide motif directly interacts with the carboxy terminal of BRCA1	1.91e-03
				LIG LIR Gen 1 122-126 162-168	gain	Canonical LIR motif that binds to atg8 protein family members to mediate autophagy	5.20E-03
				LIG SH3 3 202-208	loss	Motif recognized by SH3 domains with a non canonical class 1 recognition specificity	1.32E-02
				MOD GSK3 1 159-166	gain	GSK3 phosphorylation recognition site	2.68e-03
GPR35	Rs3749171	Thr139met	Possibly damaging Score 0.956	Within transmembrane domain, recognition unchanged			
MST1	Rs3197999	Arg689cys	Unable to compute	None	NA	NA	NA
CD6	Rs11230563	Arg225trp	Benign Score 0.371	Occurs within scavenger receptor domain. Recognition unchanged			
IL17REL	Rs5771069	Leu333Pro	Benign Score 0	Lig MYND1 331-335	Loss	PxLxP motif is recognized by a subset of MYND domain containing proteins.	6.50E-04
				Lig SUMO Sim Par 1 330-336	Loss	Motif for the parallel beta augmentation mode of non-covalent binding to SUMO protein.	4.55E-03
				LIG WW2 330-333	gain	PPLP is the motif recognized by WW domains of Group II	6.13E-05
				TrG LysEnd ApsAcLL 1 329-334	Loss	Endocytic vesicle (Nb 1 or 4 ligands)	2.75E-03
RTEL1	Rs3208008	Gln1042his	Benign Score 0	MOD PIKK1 1038-1044	Loss	(ST)Q motif which is phosphorylated by PIKK family members.	9.23E-03
NXPE1	Rs10891692	Gly353arg	Probably damaging Score 1	LIG_Actin_WH2_2 195-213	gain	The WH2 motif is of variable length (16-19 amino acids) binding to the hydrophobic cleft formed by actin's subdomains 1 and 3. At the N-terminus it forms an alpha-helix followed by a flexible loop stabilised upon actin binding.	6.6E-04

Table 2-2Eukaryotic linear motif results for all the Missense SNPs. ELM downloaded 28/11/16, rechecked 30/11/16

2.4.3 Intronic and Exonic SNPs affect splicing sites, splice enhancement and silencing

All intronic, missense and synonymous SNPs from the extended parent cohort were analysed in the splicing workflow(n=96). 65 SNPs were intronic, 17 were in long non coding RNAs, 8 were missense SNPs, 6 were synonymous SNPs. One of the synonymous SNPs (rs3742130 GPR18) also encoded an intronic SNP (UBAC2).

31 SNPs were excluded for having no cDNA HGVS transcript to be able to run in HSF/MES. A further 11 SNPs had HGVS annotation that brought up permanent error messages in HSF/MES (c.- numbers or c.* numbers); these were also excluded.

A total of 65 SNPs were therefore analysed in this component, 21 from the parent cohort and 44 additional in the extended parent cohort.

Within the parent cohort, three cryptic splice sites were identified by Human Splice Finder (HSF).

These were in *CARD9* (Rs10781499 - synonymous SNP), *IL17REL* (rs5771069 - missense SNP) and *MST1* (rs13085791 - synonymous SNP). *CARD9* rs10781499 risk allele A is predicted to activate a cryptic acceptor site which would mean cleaving the exon at the 3' site, reducing this exon by 161 nucleotides. Identification of this site as a splice site was also found by MES and ASSP. Although of note – compared to the wild type allele C the non-risk allele in ASSP was also potentially a splice site, but with a lower confidence; 0.55 compared to 0.62 with the risk allele. The non-risk allele was identified as altering a exonic splice enhancer, however the 'wild type'/ancestral allele did not form part of an ESE from the curated hexamer ESE list that could then be altered by other alleles. All three non-wild type alleles at this site were predicted to either alter ESE sites (G,A) or create an exonic splicing silencer (ESS) indicative that this site is a hotspot for splicing or splicing regulation. *CARD9* is predominantly expressed in peripheral blood mononuclear cells and B lymphocytes(232, 235) as an integral part of innate immune signalling by intracellular and extracellular pathogens(229).

IL17REL rs5771069 risk allele G is predicted to activate a cryptic splice donor site, cleaving 32 nucleotides from the exon. There was agreement from both ASSP and MES, but with low confidence values (0.3). Alleles A and C are both predicted to have no effect on splicing in HSF (splice sites, ESE, ESS or branch points). In CHESEL, I found two hexamers ESEs for the wild type sequence and three hexamer ESEs for the non-risk allele. There

were no hexamers for the risk allele sequence. By RNAseq data, IL17REL is predominantly expressed in the terminal ileum and upper GI tract(298).

MST1 rs13085791 risk allele A is predicted to activate an exonic cryptic acceptor splice site, cleaving 90 nucleotides from the exon in HSF. However, there is no concurrence between HSF, MES or ASSP, therefore this result been discounted. The non-risk allele is predicted to alter an exonic ESE site, however neither the wild type allele (G), nor the non-risk allele (C) or the risk allele (A) encode an ESE motif in CHESEL.

With the extended parent cohort 3 further cryptic splice sites from missense SNPs were identified in HSF; FCGR2A c.500G<A (G allele - non-risk, A allele - risk), however there was no concordance with MES or ASSP; MST1 c.515C<A (C allele – wild type, A allele – non-risk), again with no concordance with MES or ASSP. These two results were therefore discounted. The final cryptic splice site was via a missense SNP in NXPE1 c.631G<T (G allele wild type, T allele – risk). HSF predicted a cryptic donor site with 46.6% variability. MES was unable to identify the wildtype, therefore no results from that matrix, but ASSP was in concordance with 0.8 confidence value, leading to a cleavage of 169 base pairs from 213 base pair long exon 5. NXPE1 is predominantly over expressed in the colon (299)

Name	Non-risk allele	Risk allele	Mutation	HSF Intepretation	HSF Variation Values	Max Ent Variation Values	ASSP	Length of exon
CARD9 ENST00000371732	G	A	c.126C>A	Activation of an exonic cryptic acceptor site, with presence of one or more cryptic branch point(s). Potential alteration of splicing.	3' 51% new site	Pos 110%	Agreement. Confidence 0.62 but non-risk 0.55	-166
			c.126C>G	Alteration of an exonic ESE site. Potential alteration of splicing.	NA	NA		
IL17REL ENST00000341280	A	G	c.998T>G	Activation of an exonic cryptic donor site. Potential alteration of splicing.	3' 68.28% new site.	Pos 118. 3' 1.04%	Agreement. Confidence 0.3	-32
			c.998T>A	No significant splicing motif alteration detected. This mutation has probably no impact on splicing.	NA	NA		
MST1 ENST00000448220	C	A	c.373G>A	Activation of exonic cryptic acceptor site, with presence of one or more cryptic branch points. Potential alteration of splicing. Alteration of exonic ESE site. Potential alteration of splicing.	3' 74.54% New site. 5' 15.76% site broken.	No result in this matrix	Disagreement. Non significant difference between wild type, risk or non-risk allele	-90
			c.373G>C	Alteration of exonic ESE site. Potential alteration of splicing	NA	NA		

Table 2-3 Parent Cohort Cryptic splicing site analysis in Human Splice Finder (HSF), Maximal Entropy Scan (Max Ent) and Alternative Splice Site Predictor (ASSP). ESE = Exonic Splice Enhancer. Confidence values calculated from ASSP 0= unsure 1 = full confidence.

Of the 606 hexamer motifs in CHESEL, 30% (n= 184) were found to be affected by SNPs. 82% (n= 79) of the SNP sequences contained hexamer motifs. In order to understand this further, they need to be broken down into the individual sites/effects e.g. intronic, synonymous and missense SNPs. By considering the UC associated SNPs as biallelic (risk/non-risk), there can be one of three outcomes for the SNPs: Type 0 – neither of the SNP sequences encode an ESE; Type 1 – either SNP sequences encode an ESE but not both; or Type 2 – both sequences encode an ESE. Currently there is no *in silico* mechanism to determine the strength of the ESE motif in comparison to another out of the context of the site of the ESE e.g. adjacent to a branch point, at an exon terminus. Therefore, with the current tools available, only Type 1 outcomes are considered deleterious.

The literature is mixed with regard to synonymous SNPs (sSNPs) and ESEs. There is evidence for both positive selection of sSNPs in ESEs (300), but also that ESEs have the lowest frequency of SNP affected ESEs due to removal of deleterious variants by purifying and/or natural selection (301, 302). In this cohort; of the 6 sSNPs; two were Type 0 (MST1, CPSF3L), three were type 1 and one was type 2 (Figure 2-2). For the Type 1 SNPs, sSNPs in APEH is predicted to cause a loss of ESE motif, whereas sSNPs in CARD9 and RTEL1 are predicted to cause a gain of ESE.

Of note, the Type 2 SNP was rs3742130 which is both a synonymous SNP for GPR18 and an intronic SNP for UBAC2. UBAC2 is primarily overexpressed in neutrophils and peripheral blood mononuclear cells, although is also expressed in the colon(298), whereas GPR18 is overexpressed in B lymphocytes and is predominately expressed in immune cells as opposed to the colon(298). This is indicative of cell specificity of gene expression, for which a Type 2 SNP may have a differential effect depending on the strength of the ESE in different cell types.

By comparison to the sSNPs, the missense SNPs analysis (mSNPs) painted a very different picture; none of the eight missense SNPs were type 0, four were type 1 and four were type 2 (Figure 2-2)The mSNPs in GPR35, FCGR2A and IL17REL are predicted to cause a loss of ESE motif with the risk allele. The mSNP in CD6 is predicted to cause a gain of ESE motif. mSNPS in NXPE1, MST1, RTEL1 and IL23R all have ESE hexamer motifs in both the risk and non-risk allele sequences.

Transcript	WT	NR	R	Mutation	HSF Splicing enhancer/silencer	CHESEL	WT
CARD9 ENST00000371732	C	G	A	c.126C>G	Alteration of an exonic ESE .	Nil	nil
				c.126C>A	nil	TCATC**A** CATC**A**G	
CD6 ENST00000313421	C	C	T	c.673C>T	Creation of an exonic ESS, alteration of an exonic ESE.	CAC**T**GG	nil
IL23R ENST00000347310	G	A	G	c.1142G>A	Creation of an exonic ESS and alteration of an ESE.	**G**AACTG	TCC**A**AA
NXPE1 ENST00000251921	G	C	T	c.631G>C	Alteration of an exonic ESE.	C**C**CAGG **C**CAGGA	nil
				c.631G>T	Creation of an exonic ESS.	**T**CAGGA	
RTEL1 ENST00000369996	C	C	T	c.147C>T	Creation of an exonic ESS. Alteration of an exonic ESE.	TGTG**T**G TG**T**GCC	nil
CPSF3L ENST00000411962	A	A	G	c.1347A>G	Alteration of exonic ESE .	Nil	nil
MST1 ENST00000448220	G	C	A	c.373G>C	Alteration of exonic ESE .	Nil	nil
				c.373G>A	Alteration of exonic ESE .	Nil	

Table 2-4 Predicted exonic splicing enhancer (ESE) motifs corresponding to changes in ESE or Exonic Splicing Silencer (ESS) motifs identified in HSF . Allele of interest in bold for CHESEL results. Nil= no corresponding motif found within the sequence. WT = wild type allele. NR = non-risk allele. R = risk allele. Grey shading indicates source from extended parent cohort.

As previously noted in the literature(302-304) functioning ESE motifs are also found in the intronic sequences. Of the fifty-four intronic (non lncRNA) SNPs, twenty-two had type 1 outcomes, twenty-two had type 2. Eight intronic SNP sequences had type 0 outcome (Figure 2-2).

To analyse whether there was a significant difference between the numbers of ESEs in intronic sequences and the exonic sequences, I used a Two-way ANOVA and Tukey's multiple comparison's test, there was no significant difference between the 95% confidence intervals (CI) of the difference of means between the numbers of sequences found without any ESE hexamers in any of the SNP sites (e.g. intronic, synonymous etc). There was a significant difference as described above between the hexamers identified in either risk or non-risk allele sequences between intronic sites and the rest of the potential sites ($p = 0.033$, $p=0.03$, $p=0.045$ for synonymous, missense and lncRNAs respectively). There was carried over to a significant difference between the 95% CI of the difference of means between the numbers of intronic sequences and the numbers of synonymous sequences that had a type 2 outcome. The rest of the comparisons were non-significant (

Table 2-6).

The intronic ESE results from CHESEL (Table 2-5) indicate that where there is more than one fine mapped SNP in an intronic area within a gene e.g PUS10, C1orf106, CCNY, C5orf66, NFKB1, GNA12, IL17R, SFMBT1, each of those SNPs encode an (different) ESE hexamer motif, suggesting co-localisation of those motifs with enrichment of disease associated SNPs within them. There are two exceptions, APEH has a finemapped intronic SNP with no ESE hexamer motif and IL23R has two finemapped SNPs with no ESE hexamer motifs, where both have one SNP sequence each which does contain a motif. Although co-localisation of exonic splice enhancers has been seen in exonic regions in oncogenes such as BRCA2, with polymorphisms localising to these sites associated with splice variants (305), this has not been documented in intronic regions. A common feature of intronic splice enhancer motifs is C triplets or G triplets. These are not commonly found at exonic splice sites, and mutations in exonic C tripletswhen they do occue does not change their function. Co-occurrence of C and G triplets within an intron is thought to create a functional synergy to create a larger ribonuclearprotein complex through transacting sites(306).

NR5A2 rs2816958 and C5orf66 have C triplets in their hexamer motif, APEH has a G triplet in its hexamer motif. Looking at the wider sequences for the putative co-localising intronic ESE motifs; neither PUS10 SNPs, CCNY nor NFKB1 have C or G triplets within their 11 bp sequences, C1orf106 had a G triplet in one SNP sequence (rs59655222), C5orf 66 had a triplet C in one SNP sequence (rs254562), GNA12 had a C triplet in rs1182188 and a G triplet in rs798502, IL17R (rs11567701) had a G triplet and SFMBT1 (rs2564956) had a G triplet. There was no co-occurrence of C and G triplets within one sequence, further analysis is required to assess if the disease associated intronic SNPs that cluster with putative enhancing motifs, be it CHESEL motifs or triplet motifs, are close enough to function synergistically to cause a splicing variation or aberration. The presence or absence of the triplets within the SNP sequences does not indicate that these are definitely are or are not enhancing sites, but it does add weight to the CHESEL intronic results. In terms of the downstream effect, the heterogenous ribonucleoprotein (hnRNP) family bind intronic sequences to have an enhancer effect. Thus, a gain of an intronic ESE motif with the risk allele was denoted as a gain of splicing enhancement, whereas a loss of intronic ESE motif with the risk allele sequence was a loss of splicing enhancement.

Non risk	Risk	SNP/ID	Site	Gene Name
TATGGCAGTGA	TATGGTAGTGA	rs6911490	intronic	ATG5
AATCAGAACTA	AATCATAACTA	rs11229555	intronic	GLYAT
TCTCCTTCCAC	TCTCCCTCCAC	rs1182188	intronic	GNA12
GCTAATGTACA	GCTAACGTACA	rs8005161	intronic	GPR65
TGTGTGTAGGA	TGTGTATAGGA	rs11041476	intronic	LSP1
TACCTGTGTCC	TACCTATGTCC	rs1893217	intronic	PTPN2
ATTTTTGTACT	ATTTTCGTACT	rs7596362	intronic	PUS10
TTGATATTTGA	TTGATCTTTGA	rs7608697	intronic	PUS10
CTCTTGTAGAC	CTCTTATAGAC	rs7608910	intronic	PUS10
TGTGGAACTTG	TGTGGGACTTG	rs2564956	intronic	SFMBT1
TTTAAATTTAA	TTTAAGTTTAA	rs12942547	intronic	STAT3
TCACATTGGGA	TCACACTGGGA	rs59655222	intronic	C1orf106
ACAGTACTGTC	ACAGTGCTGTC	rs254560	intronic	C5orf66
ATTCCGTCTCA	ATTCCATCTCA	rs28671712	intronic	CHP1
CTGTTCGCTCT	CTGTTTGCTCT	rs7495132	intronic	CRTC3
gtcagTtgtca	gtcagAtgtca	rs267984	intronic	DAP
CTCTTTCAGTT	CTCTTCCAGTT	rs10896794	intronic	LPXN
GTTTAGATAAC	GTTTAAATAAC	rs543104	intronic	MAML2
ATGCCGTGTAT	ATGCCATGTAT	rs2425019	intronic	MMP24
GATACAATGTT	GATACGATGTT	rs3774937	intronic	NFKB1
AGCAGGTCAAC	AGCAGTTCAAC	rs3766606	intronic	PARK7
GTGAGTAACAC	GTGAGGAACAC	rs2266959	intronic	UBE2L3
actgcGttcca	actgcAttcca	rs6062504	intronic	ZGPAT
TCAAACCTGGG	TCAAATCTGGG	rs11130213	intronic	APEH
CAAGTGGTTTT	CAAGTAGTTTT	rs12131796	intronic	C1orf106
ctttgTtgaat	ctttgCtgaat	rs12132298	intronic	C1orf106
GATAACTGCAG	GATAAATGCAG	rs7554511	intronic	C1orf106
TTTACTCCTGC	TTTACGCCTGC	rs41299637	intronic	C1orf106
CTTCCTTACCC	CTTCCCTACCC	rs254562	intronic	C5orf66
TTGCTCTGCAC	TTGCTGTGCAC	rs12254167	intronic	CCNY
GATGATAGCAA	GATGAGAGCAA	rs12261843	intronic	CCNY
gtcacCgtact	gtcacTgtact	rs6481950	intronic	CCNY
TAGCAGGAGGT	TAGCAAGAGGT	rs11879191	intronic	CDC37
TGCAAGTGCTA	TGCAACTGCTA	rs267939	intronic	DAP
tctgaAgggtc	tctgaCgggtc	rs798502	intronic	GNA12
GTGGCCTTGAT	GTGGCTTTGAT	rs11168249	intronic	HDAC7
ACGCACATCTG	ACGCAGATCTG	rs11567699	intronic	IL17R
CAACTGGGATT	CAACTTGGATT	rs11567701	intronic	IL17R
TATGATGTTAG	TATGACGTTAG	rs113935720	intronic	IL23R
GAGAGAGACTT	GAGAGGGACTT	rs7657746	intronic	KIAA1109
ATAGCAAGAAA	ATAGCGAGAAA	rs3774959	intronic	NFKB1
TCCCAGGCAGT	TCCCAAGCAGT	rs2816958	intronic	NR5A2
CAGGCTCTGCT	CAGGCGCTGCT	rs4560096	intronic	PUS10
TCCAACCTGGT	TCCAAGCTGGT	rs2581817	intronic	SFMBT1
GGAAACACCAT	GGAAAAACCAT	rs1517352	intronic	STAT4

Table 2-5 CHESEL Intronic sequences with hexamer motif matches, type 1 and type 2 outcomes in grey. EPC.

Comparing the ESE hexamers found in CHESEL with the HSF results for the intronic samples, although may be of interest to compare the two modalities is an untenable comparison given the large number of exclusions in HSF due to inabilities to run a number of intronic transcripts, therefore the intronic SNPs found in CHESEL were put forward to the network, with the caveat that we are unable to assess the false positive nature of these results.

HSF also identified multiple exonic splicing silencer motifs in the synonymous and missense SNP sequences identified in Table 2-4.

Attempts to confirm these with other ESS prediction tools was unsuccessful as published or validated prediction tools such as FAS-ESS (http://genes.mit.edu/fas-ess/) or MutDB (www.mutdb.org) are either no longer working or do not have the rsIDs we are interested in within their databases. The ESS motifs (genes affected) were put forward to the network, again, with the caveat that we are currently unable to assess the false positive nature of these results.

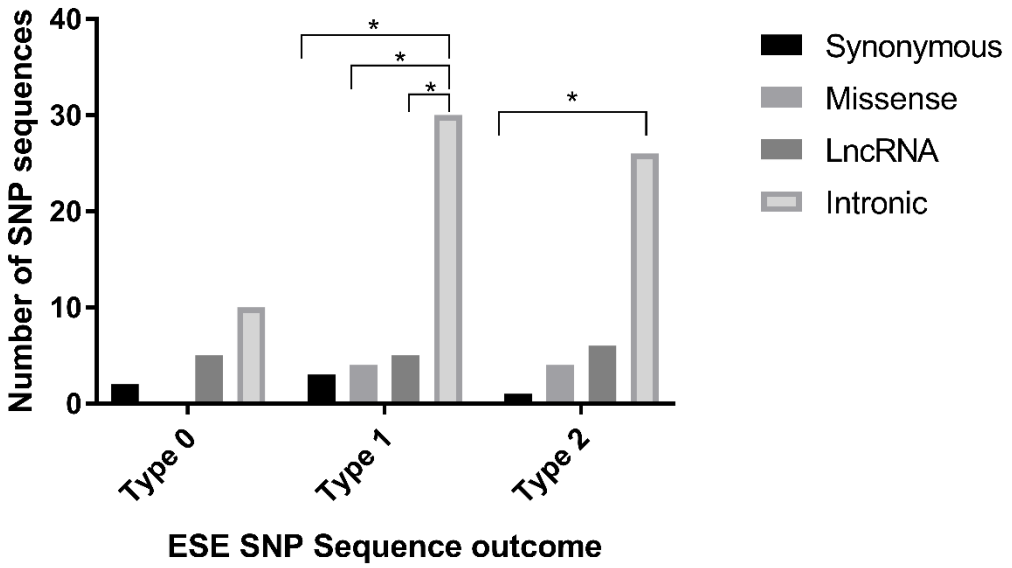


*Figure 2-2Comparison of number of SNP sequences for each ESE outcome. Type 0 outcome = no ESE hexamer in either risk or non-risk allele sequence. Type 1 =ESE hexamer found in either risk or non-risk SNP allele sequence. Type 2 outcome = ESE hexamer found in both risk and non-risk allele sequences. * = p=<0.05, actual values within the text. The rest was not significant. Based on Tukey's Multiple Comparison.*

	Mean Diff.	95.00% CI of diff.	Summary	Adjusted P Value
Type 0				
Synonymous vs. Missense	2	-22.46 to 26.46	ns	0.9913
Synonymous vs. LncRNA	-3	-27.46 to 21.46	ns	0.9721
Synonymous vs. Intronic	-8	-32.46 to 16.46	ns	0.6854
Missense vs. LncRNA	-5	-29.46 to 19.46	ns	0.8906
Missense vs. Intronic	-10	-34.46 to 14.46	ns	0.5347
LncRNA vs. Intronic	-5	-29.46 to 19.46	ns	0.8906
Type 1				
Synonymous vs. Missense	-1	-25.46 to 23.46	ns	0.9989
Synonymous vs. LncRNA	-2	-26.46 to 22.46	ns	0.9913
Synonymous vs. Intronic	-27	-51.46 to -2.536	*	0.0333
Missense vs. LncRNA	-1	-25.46 to 23.46	ns	0.9989
Missense vs. Intronic	-26	-50.46 to -1.536	*	0.039
LncRNA vs. Intronic	-25	-49.46 to -0.5356	*	0.0458
Type 2				
Synonymous vs. Missense	-3	-27.46 to 21.46	ns	0.9721
Synonymous vs. LncRNA	-5	-29.46 to 19.46	ns	0.8906
Synonymous vs. Intronic	-25	-49.46 to -0.5356	*	0.0458
Missense vs. LncRNA	-2	-26.46 to 22.46	ns	0.9913
Missense vs. Intronic	-22	-46.46 to 2.464	ns	0.0752
LncRNA vs. Intronic	-20	-44.46 to 4.464	ns	0.1055

Table 2-6 Tukey's Multiple Comparison of three groups, Type 0 outcome, Type 1 outcome and Type 2 outcome.

2.4.4 SNPs predicted impact on miRNA function

2.4.4.1 SNP effects on mature miRNAs

Using a BLASTN search, sixteen human mature microRNA homologous sequences were suggested to be 'lost' with the presence of a SNP risk allele as compared to the non-risk allele (Table 2-7). The scores for homology are based on the statistics for local alignments lacking gaps, where the ideal is to find two segment pairs which cannot be improved by extension or trimming called high scoring pairs – the higher the score the better the alignment. To ensure this is not by random alignment, the E value tells you how likely the alignment is to have occurred by chance. As shown by in Table 2-7, the BLASTN scores are only moderately high (which could be a function of the short sequence analysed) and the E values are less than ten, but an E value of 10 gives a probability of random chance of 0.999. Given these 'grey' results, we searched the MiRBase database miRNA for miRNA motifs using SSEARCH to add confidence to the findings. The SSEARCH facility is better placed to search short sequences, hence the higher scores but the payoff is higher E values in some cases, such as mir-6755-3p, to the extent of not finding the previously identified homology sequence.

Human MiRNA	Gene	rs ID	Site	BLASTN Score	E-value	SSEARCH Score	E-value
miR-1266-3p		rs4380874	intergenic	62	5.7	110.9	8.1
miR-6775-3p				60	8.4	not found	
miR-548at-3p	*NXPE2P1*	rs561722	UGV	61	6.9	109.7	9.9
miR-6777-3p	*TMBIM1*	rs2382817	5'UTR	67	2.2	not found	
miR-5692a	*IKZF3*	rs12946510	DGV	61	6.9	not found	
miR-519c-3p	lncRNA	rs6920220	intronic	61	5.3	116.3	4.1
miR-526b-5p	*KIAA1109*	rs7657746	intronic	63	4.3	not found	
miR-552-5p	*CDC37*	rs11879191	intronic	61	6.9	116.6	4.1
miR-6835-5p				not found		112.6	6.5
miR-5701	lncRNA	rs11742570	UGV	63	4.7	116	4.9
miR-365a-3p	*ITLN1*	rs4656958	UGV	61	6.9	114.5	5.1
miR-365b-3p				61	6.9	114.5	5.1
miR-4699-3p				63	4.7	116	4.9
miR-5010-5p	*CCNY*	rs12254167	intronic	62	5.7	110.9	8.1
miR-511-5p				66	2.6	not found	
miR-6082	*CHP1*	rs28374715	intronic	68	1.8	121	3.2
miR-6875-5p	lncRNA	rs941823	intronic	65	3.2	118	4.1

Table 2-7 Human Mature miRNA homology sequences lost with the presence of the risk allele in the 21bp flanking sequence, when compared to the non-risk allele 21bp sequence. Data summarised from MirBase. Extended parent cohort

Analysis of the miRNA targets in Cytoscape 3.3.0 identified two subnetworks, a smaller subnetwork consisting of two star clusters around a lncRNA SNP site and an intergenic SNP site. This suggested that there were interlinking (or first neighbour) interaction nodes between the source nodes. First neighbour nodes are of interest as these can link our SNPs of interest to a downstream impact not otherwise identified from looking directly at the SNP affected site. Unfortunately, despite visually looking connected, the clustering coefficient was zero, BinGO did not identify any biological process enrichment, and confirmation of no biological enrichment in all tissues analysed (intestine, colon, B lymphocyte, T lymphocyte, Whole blood) in Genome-scale Integrated Analysis of Gene Networks in Tissues (GIANT) all suggested that the network created was not a indictor of a biological process, but by random chance.

Genome analysis of the annotated sites for the miRNA in Ensembl identified no correlation between the SNP sites and the true miRNA annotated sites. The homology miRNA and their binding targets were therefore removed from further network creation.

2.4.4.2 SNPs effect miRNA binding sites

449 putative miRNA binding sites (MBS) were identified in 117 SNP sites using the initial cut-off threshold of total score >150, energy score<-7. Given the high rate of false positive motifs we had identified in previous sections, and the literature quoting false positive rates between 20-40% (Baek 2008, Krek 2005, Selbach 2008, Cloonan 2008) for algorithm based predictive miRNA binding site prediction tools; a more stringent threshold of total score >155, energy score <-20 was applied.

This resulted in 104 putative MBS in 56 SNP sites. A further 6 MBS in 3 SNP sites were excluded from network creation for being neutral (loss and gain of the same miRNA binding site), 18 were excluded as being within intergenic regions, 1 was excluded for being within a non-coding transcript variant, 5 were excluded for being in antisense RNAs, 2 were excluded in uncharacterised RNA (RP11-415K). After exclusions, there were 71 putative MBS in 36 SNP sites (Table 2-8).

2.4.4.3 miRNA-mRNA sites affected by SNPs

Seed sequences of miRNAs are at positions 2'-8'of the miRNA, however there can be mismatch or wobble at positions 5,6 and 7 (Loeb 2012) which can create variation in binding. Perfect complementarity of 11 nucleotides starting at position 3, 4 or 5 can repress mRNA translation (Shin 2010). Unfortunately, these only accounts for <10% of miRNA/mRNA interactions found in the human transcriptome (Bartel 2009).

The miRanda algorithm used creates an optimal local alignment of the miRNA with the mRNA using a weighted dynamic programme algorithm – this gives preferential weight to the '2-8' pairing, and is a sum of match/mismatch scores with gap penalties. The level of perfect complementarity is given as a percentage alignment score, which takes into account the miRNA-mRNA alignment length and the gaps within that alignment. The alignment lengths for the SNP sequences were set to >11, the longest alignments were 20nt long. The best alignment was 100%, in 11 nt in rs72703058 which is an intergenic SNP. The lowest alignment score was 57.89% in a 20nt alignment sequence for rs41299637 (C1orf106). The intergenic and lncRNAs miRNA binding sites were interspersed equally across the range of percentage alignments.

Each SNP site could have multiple miRNA binding sites within in it (mean = 2.2, median = 2, mode = 1, minimum-maximum =1-6). The Pearson correlation was significant with r = -0.85 (-0.98-0.1 95%CI); r^2=0.73, p=0.036 which was not unsurprising given the low

probability of 1 SNP site having 6 miRNA binding sites associated with it as opposed to 1SNP site having 1 miRNA binding site associated with it.

There was no overlap in the specific miRNAs that would bind, within the included set. Two genes (KIR3DL2-rs1654644 and ITGAL-rs12716977) shared miRNA (hsa-miR-4716-3p, hsa-miR-1268b respectively) with lncRNA sites (rs200073939 and rs11757201)(see below). These SNPs are not finemapped or in linkage disequilibrium with their shared miRNA partner. From the intergenic sites (excluded) 1 miRNA (hsa-miR-6812-5p) was shared between two intergenic sites (rs72703050, rs2413583). Again, these SNPs are in different sites and are not in linkage disequilibrium with each other or fine mapped together.

The main role of miRNAs is to negatively regulate gene expression. Binding of the miRNA to mRNA leads to splitting and deadenylation of the mRNA, thereby causing translation repression. A loss of a miRNA binding site would suggest that negative regulation could not occur by that particular miRNA and vice versa; a gain of a miRNA binding site may indicate gene silencing. 54% of the putative MBS were lost with the risk allele, 46% were gained, loss affecting 17 genes; gain affecting 18 genes, with an overlap of 8 genes between loss and gain (MST1, NXPE1, CD6,ITGAL, CCNY, C1orf106,KIAA1109, NR52A) via different miRNAs. The overlap gene SNP sites were synonymous (MST1), missense (NXPE1, CD6), upstream gene variants (ITGA) and intronic (C1orf106, CCNY,KIAA1109, NR52A). Although 3'UTRs are the primary site of miRNA binding to mRNAs, miRNA binding sites have been shown to be active in open reading frames (exonic and intronic regions), and 5'UTRs (307-310). 53% (n=29)of the MBS (regardless of loss or gain) were intronic, 16% were missense (n=9), 13% were synonymous (n=7), 11% were upstream gene variants (n=6), 4% were down stream gene variant (n=2), 2% were 3'UTR and 5'UTR (n= 1 each).

Similarly to the splice site motifs, there are six finemapped genes that have MBS motifs in multiple SNP sequences; TNFRSF14, C1orf106, CCNY, RTEL, MST1 and ITGAL. Although there is evidence that multiple binding sites for one miRNA within a target gene optimises the repression of said target gene (281, 311, 312), this is not represented as a SNP effect in this dataset but there is increasing evidence that combinations of different miRNAs work in conjunction to repress target genes(313-315).

Analysis of the conserved miRNAs in miRanda with good mirSVR scores(316) for each of the gene mRNAs did not identify any other binding sites for any of the miRNAs we have identified (no redundancy of conserved sites). Analysis of non-conserved sites in miRanda

did confirm two SNP predictions; mir-661 in RTEL1 (rs2257440) and in FCGR2A miR-204-5p (rs1801274) as being where we expected it to be in the mRNA (it is lost with the risk allele). The RTEL MBS also had redundancy – another binding site approximately 100 base pairs part. For the broad institute finemapped SNP rs59655222 (that used the alleles denoted in Ensembl), miRanda identified the miRNA binding site (which should have been a gain in binding site with the risk allele). This emphasises the importance of having the right allele for understanding the disease process. In ITGAL mRNA there were eight miR-548 family sites, but none were specifically miR-548ay, miR-548aa or miR-548at.

miRNA	SNP	Site	Gene	Total Score	Energy (kcal/mol)	Loss/Gain
miR-2392	rs727088	3'UTR	*CD226*	155	-25.3	loss
miR-1291	rs2382817	5'UTR	*TMBIM1*	164	-22.59	gain
miR-8073	rs10797432	DGV	*TNFRSF14*	155	-25.01	loss
miR-7157-3p	rs10910092	DGV	*TNFRSF14*	157	-20.19	loss
miR-3941	rs11041476	Intronic	*LSP1*	156	-20.61	loss
miR-6511b-5p	rs41299637	Intronic	*C1orf 106*	156	-24.63	gain
miR-7159-5p	rs12132298	Intronic	*C1orf106*	162	-21.55	loss
miR-199a-5p	rs59655222	Intronic	*C1orf106*	171	-30.61	gain
miR-199b-5p	rs59655222	Intronic	*C1orf106*	167	-28.54	gain
miR-4733-5p	rs59655222	Intronic	*C1orf106*	169	-22.79	loss
miR-1229-5p	rs254562	Intronic	*C5orf66*	162	-31.66	gain
miR-433-5p	rs6481950	Intronic	*CCNY*	161	-24.35	loss
miR-4430	rs6481950	Intronic	*CCNY*	155	-23.82	gain
miR-6870-3p	rs12261843	Intronic	*CCNY*	165	-27.05	gain
miR-5589-5p	rs28374715	Intronic	*CHP1*	162	-25.64	gain
miR-204-5p	rs1801274	Intronic	*FCGR2A*	158	-21.98	loss
miR-6867-3p	rs1801274	Intronic	*FCGR2A*	157	-21.06	loss
miR-4433a-3p	rs1182188	Intronic	*GNA12*	161	-28.4	loss
miR-6880-5p	rs1182188	Intronic	*GNA12*	163	-21.27	loss
miR-4510	rs1182188	Intronic	*GNA12*	159	-21.06	loss
miR-6760-5p	rs1182188	Intronic	*GNA12*	156	-20.62	loss
miR-7847-3p	rs1182188	Intronic	*GNA12*	158	-20.44	loss
miR-4647	rs3024495	Intronic	*IL10*	159	-21.88	gain
miR-3183	rs7657746	Intronic	*KIAA1109*	161	-23.84	loss
miR-2114-5p	rs7657746	Intronic	*KIAA1109*	160	-22.53	loss
miR-642a-5p	rs7657746	Intronic	*KIAA1109*	158	-21.4	loss
miR-3184-3p	rs7657746	Intronic	*KIAA1109*	164	-20.91	gain
miR-625-5p	rs1654644	Intronic	*KIR3DL2*	165	-23.44	loss
miR-4716-3p	rs1654644	Intronic	*KIR3DL2*	155	-20.19	loss
miR-6839-5p	rs483905	Intronic	*MAML2*	162	-20.79	Loss
miR-619-5p	rs2816958	Intronic	*NR5A2*	164	-27.61	gain
miR-6513-5p	rs2816958	Intronic	*NR5A2*	156	-21.29	loss
miR-4538	rs2581817	Intronic	*SFMBT1*	165	-22.06	gain
miR-3653-5p	rs10891692	Miss	*NXPE1*	156	-23.87	gain
miR-1200	rs10891692	Miss	*NXPE1*	159	-23.24	gain
miR-3192-5p	rs10891692	Miss	*NXPE1*	158	-23.11	loss
miR-6761-5p	rs10891692	Miss	*NXPE1*	155	-20.67	gain
miR-4281	rs11230563	Miss	*CD6*	159	-29.88	gain
miR-6849-5p	rs11230563	Miss	*CD6*	157	-25.48	loss
miR-6759-5p	rs3208008	Miss	*RTEL1*	159	-26.71	loss

Table 2-8 Summary of miRNA binding site affinities to SNP sequences in the extended UC cohort.

miRNA	SNP	Site	Gene	Total Score	Energy (kcal/mol)	Loss/Gain
miR-6747-3p	rs5771069	Miss	*IL17REL*	157	-25.13	gain
miR-7113-3p	rs12103	Syn	*CPSF3L*	162	-28	gain
miR-4502	rs10781499	Syn	*CARD9*	165	-22.14	gain
miR-369-3p	rs1131095	Syn	*APEH*	179	-21.76	loss
miR-6746-5p	rs13085791	Syn	*MST1*	162	-32.94	loss
miR-8085	rs13085791	Syn	*MST1*	157	-23.59	loss
miR-661	rs2257440	Syn	*RTEL1*	155	-31.88	loss
miR-6769a-3p	rs9822268	Syn	*MST1*	156	-22.9	gain
miR-548aa	rs11150589	UGV	*ITGAL*	169	-24.04	gain
miR-548t-3p	rs11150589	UGV	*ITGAL*	169	-24.04	gain
miR-548ay-3p	rs11150589	UGV	*ITGAL*	158	-23.32	loss
miR-548at-3p	rs11150589	UGV	*ITGAL*	162	-21.13	loss
miR-1268b	rs12716977	UGV	*ITGAL*	166	-35.99	loss
miR-1268a	rs12716977	UGV	*ITGAL*	166	-35.49	loss

Table2-8 continued Summary of miRNA binding site affinities to SNP sequence. Highlighted genes are SNPs in extended parent cohort.

2.4.4.4 miRNA-long non-coding RNA sites affected by SNPs.

Long non-coding RNA with miRNA binding sites within them have been shown to affect post-transcriptional regulation by miRNAs by competing with the mRNA targets of miRNAs, thereby reducing the availability of the miRNA, reducing repression of the target mRNA(317). Of the 71 putative binding sites, 17 (24%) were predicted to occur due to a combination of 8 SNPs in 6 long non coding RNAs and 2 SNPs in upstream gene variants of long non coding RNA RP11-386E5.1(Table 2-9). In order for this competition to occur, the lncRNA and mRNA pair that is targeted by a common miRNA must be expressed in the same tissues. Using lnCeDB, the miRNA identified were not found within the only 2 long non-coding RNA the database held data for, from within our dataset; RP11-95M15.1 and LINC00484. Predictions for the mRNA pairs that share the miRNA with the lncRNA have not been undertaken, as this would make the network fraught with false positive results from non-validated 'predictions on predictions'. Accordingly, the SNPs have not been separated into parent and extended parent cohorts but visualised as a whole cohort in the table below.

miRNA	SNP	Site	LncRNA	Total Score	Energy (kcal/mol)	Alignment length	miRNA % alignment	Loss/Gain
miR-1268b	rs11757201	intronic	RP11-95M15.1	159	-27.46	17	81.25%	gain
miR-499a-3p	rs11757201	intronic	RP11-95M15.1	157	-21.53	19	71.43%	loss
miR-557	rs13277237	intronic	CCDC26-001	158	-20.41	19	72.22%	loss
miR-608	rs2396087	intronic	RP11-344J7.3	166	-32.43	21	71.43%	loss
miR-6782-5p	rs2396087	intronic	RP11-344J7.3	156	-28.38	21	75.00%	loss
miR-3936	rs2396087	intronic	RP11-344J7.3	155	-25.19	20	73.68%	loss
miR-1913	rs2396088	intronic	RP11-344J7.3	168	-28.04	20	78.95%	loss
miR-3137	rs2396088	intronic	RP11-344J7.3	157	-26.52	22	66.67%	gain
miR-1301-3p	rs2396088	intronic	RP11-344J7.3	165	-24.59	17	81.25%	loss
miR-1254	rs4743820	intronic	LINC00484	160	-33.17	22	71.43%	loss
miR-323b-3p	rs6584283	intronic	LINC01475	162	-27.27	20	73.91%	gain
miR-8089	rs6940798	intronic	RP11-344J7.3	155	-24.31	22	65.22%	gain
miR-2110	rs6940798	intronic	RP11-344J7.3	163	-22.27	19	72.22%	gain
miR-6799-3p	rs9770544	intronic	RP1-170O19.14	156	-21.17	16	73.33%	loss
miR-6809-3p	rs6883964	UGV	RP11-386E5.1	168	-23.68	18	88.24%	gain
miR-6830-3p	rs6883964	UGV	RP11-386E5.1	163	-20.02	19	72.22%	loss
miR-3934-5p	rs6888952	UGV	RP11-386E5.1	163	-22.36	16	87.50%	loss

Table 2-9Putatuve miRNA binding sites within long non-coding RNAs

2.4.5 SNPs predicted effect on Transcription Factor Binding Sites

Thirteen SNPs were predicted by JASPAR to occur within transcription factor binding sites.

RTEL1 encodes a DNA helicase crucial for telomere maintenance and DNA repair. It is known to be regulated by NRSF1, NRSF2, AP1, NFKB1, ARP1, E47, Ik1. Rs2297441 is predicted to gain the homologous sequence for transcription factors HOXB13 and HOXD13. HOXB13 and HOXD13 are sequence specific transcription factors which form part of the developmental regulatory system which provides cells with specific positional identities on the anterior- posterior axis. There was no ORegAnno annotation for this site, the SNP does occur within an enhancer region which is an elite enhancer, however the enhancer region does not regulate RTEL1.

TNFRSF14 encodes a member of the tumour necrosis factor superfamily. It functions in signal transduction pathways to activate inflammation and inhibit T cell immune respose. It also binds herpes simplex virus envelope proteins to mediate the viral entry into the cell. It is regulated by GATA1, NKx2-5, Nkx5-1, RP58, STAT3, ARE6, E2F. rs10797432 is predicted to encode a loss of binding to INSM1 is a transcriptional repressor involved in beta cell development. There was no ORegAnno annotation, nor does the SNP occur within an enhancer region.

HDAC7 encodes a histone deacytylase which plays a critical role in transcriptional regulation by altering chromosomal structure to prevent transcription factor binding, thereby repressing transcription. It has transcription factor binding sites to *PAX4A, E47, C-ETS-1, SEF1, NKX2-5, ZIC3, FAC1, PPAR-GAMMA 1 and 2*. Rs11168249 is predicted to cause a loss of transcription factor binding to *LEF1* and *TC7L2*. The SNP occurs both within an ORegAnno annotation site for SMARCA4, as well as an enhancer region. This elite enhancer region does contain a binding site to *TC7L2* but not *LEF1*.

IL17R encodes a membrane bound glycoprotein which binds with low affinity to interleukin 17A. It is regulated by AP1, c-JUN, ATF-2, GATA3, NF-AT1,2,3,4, NF-AT, C/EBPalpha. Rs11567699 is predicted to cause of loss of binding by ZBTB18. There is no ORegAnno annotation, but it does occur within an elite enhancer region which does regulate *IL17R*, but a ZBTB18 binding site is not found within this enhancer.

FCGR2A encodes a protein that is found on the cell surface of phagocytic immune cells, including macrophages and neutrophils and is directly involved in the phagocytic process and initiates cellular responses against pathogens and soluble antigens. It is regulated by STAT3, AML1a, P53, FOXD1, STAT1, HLF, HAND1, E47 and EVI-1. rs4657041 is predicted to

cause the loss of binding sites for DUXA and RARA transcription factors. There was no ORegAnno nor enhancer annotation for this site.

CRTC3 is a member of the CREB regulated transcription coactivator family. It is regulated by deltaCREB, CREB, NKX6-1, MSX-1, AML1a, CRE-BP1, ATF-2, CP2 and ZIC3. Rs7495132 is predicted to cause a loss of binding site for FOXA1 and a gain of TBX19. There was no ORegAnno nor Enhancer annotation for this site.

LPXN is a member of the focal adhesion associated adapter protein family. It is involved in the regulation of cell adhesion, cell migration and is a negative regulator in integrin mediated cell adhesion events. It is regulated by C/EBPalpha, c-ETS1, HOAX3 and FOXC1. Rs10896794 is predicted to cause a gain of binding site for IRF1, STAT1 and STAT2. IRF1 transcriptional regulation occurs with IRF1 as a homodimer, or as a heterodimer with unphosphorylated STAT1 which bind partially overlapping interferon consensus sequences or as part of an interferon enhancesome with NFKB, ATF-2/c-jun, and IRF3, IRF7. There are no other *cis* or *trans* acting transcription factor binding sites to create an interferon enhancesome 2kB upstream or downstream of *LPXN*. The SNP does not occur within an enhancer region, however, it is immediately downstream of two ORegAnno annotations for CEBPA which raises the possibility of transcription factor cooperativity which is a previously documented phenomena between STATs and C/EBPs(318-320). The SNP is also immediately upstream of a microsatellite region.

CHP1 encodes a calcium binding phosphoprotein that binds to the Na/H exchanger NHE1. It mediates the association between microtubules and membrane bound organelles of the endoplasmic reticulum and Golgi apparatus. It is also an endogenous inhibitor of calcineurin activity. rs28671712 is predicted to cause a gain of transcription factor binding site to IRF1, however the p value is significantly smaller than the rest of the predictions. There was no ORegAnno nor enhancer annotations for this site.

IL10 encodes Interleukin 10, which is primarily produced by monocytes. It is a cytokine that has pleiotropic effects on inflammation and immunoregulation. It downregulates the expression TH1 cytokines, MHC class II antigens, and costimulatory molecules on macrophages. Conversely it enhances B cell survival, proliferation and antibody production. It is considered to be an essential immunoregulator in the intestinal tract. It is regulated by STAT3, SP1, deltaCREB, CREB, GR-beta, GR-alpha, GR and PBX1a. rs3024495 occurs within an elite enhancer region that does regulate IL10. It is predicted to cause a gain in binding site to ZNF263. ZNF263 is a KRAB domain containing zinc finger

protein which has both transcription activation and repression properties. *IL10* is not known to be regulated by ZNF263.

ITGAL encodes integrin subunit alpha L which combines with the beta2 subunit (ITGB2) to form the lymphocyte function associated antigen 1 (LFA1). LFA1 has a role in leucocyte intercellular adhesion and as a lymphocyte costimulatory molecule. It is known to be regulated by AML1a. rs11150589 is within an elite enhancer site which is known to regulate ITGAL, and is predicted to cause a gain of binding site to *IRF1*.

SNP	Site	Gene	Transcription Factor	strand	sequence	P Value	Loss/Gain
rs2297441	3'UTR	*RTEL1*	HOXB13	D	CTAATAAAAC	9.10E-06	Gain
rs2297441	3'UTR	*RTEL1*	HOXD13	D	CTAATAAAAC	6.20E-06	Gain
rs10797432	DGV	*TNFRSF14*	INSM1	D	TGCCAGGGGGAG	5.20E-06	Loss
rs12946510	DGV	*IKZF3*	MEF2C	D	GAGTTAAAAATAAAA	7.10E-06	Gain
rs11168249	intronic	*HDAC7*	LEF1	R	CAAGATCAAAGCCAC	6.40E-06	Loss
rs11168249	intronic	*HDAC7*	TCF7L2	R	CAAGATCAAAGCCA	3.90E-06	loss
rs11567699	intronic	*IL17R*	ZBTB18	R	TGGACAGATGTGC	8.20E-06	loss
rs1654644	intronic	*KIR3DL2*	ZNF263	R	GGGGGATTTGGGTGAGGGGGA	7.30E-06	loss
rs4657041	intronic	*FCGR2A*	DUXA	R	ATGACCTAATCAC	9.00E-06	loss
rs4657041	intronic	*FCGR2A*	RARA	D	TAGGTGATTAGGTCATGA	5.30E-06	loss
rs7495132	intronic	*CRTC3*	FOXA1	D	AGTCTGTTTGCTCTT	6.50E-06	Loss
rs10896794	intronic	*LPXN*	IRF1	D	GATTTGTTACTCTTTCAGTTT	6.60E-06	gain
rs10896794	intronic	*LPXN*	STAT1::STAT2	D	ACTCTTTCAGTTTTT	5.10E-06	gain
rs28671712	intronic	*CHP1*	IRF1	R	TTTTTTTTTTTTTTTGAGAT	1.00E-05	gain
rs3024495	intronic	*IL10*	ZNF263	R	GATGGTGAGAGGAGAGGAGGG	3.70E-06	gain
rs7495132	intronic	*CRTC3*	TBX19	D	GTTCGCTCTTATATGTGAAA	8.80E-06	gain
rs11150589	UGV	*ITGAL*	IRF1	D	CTGTGGTTTTGATTTGCATTT	9.50E-06	gain

Table 2-10 Putative transcription factor binding sites (lncRNA sites excluded)

2.4.6 The UC SNP effects – what and how makes and difference.

A total of 43 SNP affected proteins identified from the previous tables were input into Cytoscape with their annotation, totalling 117 nodes and 79 edges (Figure 2-3). Although the UC nodes had 30 connected components, with 13 isolated nodes, as can be seen in the demonstrative figure 2-3, the UC nodes do not form a network.

The commonest modality (e.g type of effect) was a splice effect (not separating the splicing effect types). 65% of the affected proteins/genes were affected by splicing changes; 56% were affected by miRNA binding site changes; 25% were affected by transcription factor binding site changes, and 14% were affected by changes in the protein linear motif structure.

51% of SNP genes/proteins were affected by 1 modality alone (not taking into account the frequency that the modality was identified), 37% were affected by 2 modalities, and 12% were affected by 3 or more modalities. Interestingly, only one had all 4 modalities (RTEL1). We noted a trend between the number of modalities/frequency of hits and the minor allele frequency (Figure 2-4) but not with PIC values (Figure 2-5). We saw disease associated SNPs with a potentially significant phenotype e.g. multiple predicted effects, had higher minor allele frequencies. There was a natural break at a minor allele frequency (MAF) of 0.5, with 12 SNPs having a MAF <0.5 (group A) and 11 SNPs having a MAF >0.5 (group B). The groups were significantly different for both modality and frequency of modalities using a Mann-Whitney two tailed T-test (Figure 2-6) ($p = 0.0071$ and $P=0.0029$ respectively), covariance (using Two-way ANOVA) was significant $p=0.0013$. Minor allele frequency in combination with the number and frequency of modalities predicted at a SNP site could therefore be used to stratify SNPs for further investigation.

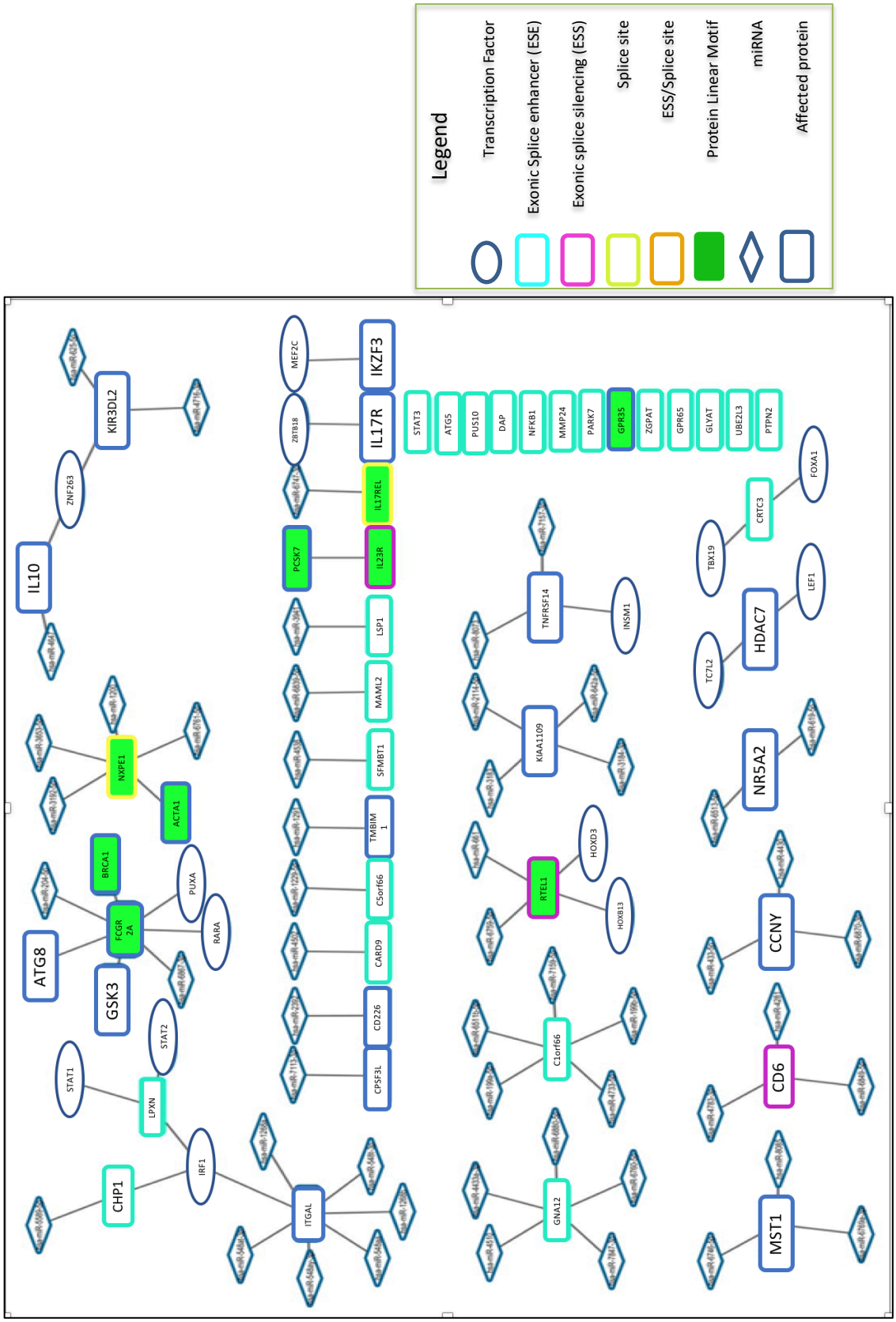


Figure 2-3 UC SNP Nodes denoted by their Uniprot IDs with the miRNA whose binding site is affected (diamond) or transcription factor whose binding site is affected (oval) by SNPs. Some proteins are affected by protein linear motifs (ELM) (green blocks), SNPs affecting slice sites in mRNA have a coloured outline to the node. mRNA with splice enhancer sites affected by SNPs have a turquoise border, mRNA with splice silencing sites affected by SNPs have a pink border and mRNA with splice motifs affected by SNPs have an olive border. mRNA with both a splice motif and a splice silencing site affected by a SNP have a light brown border.

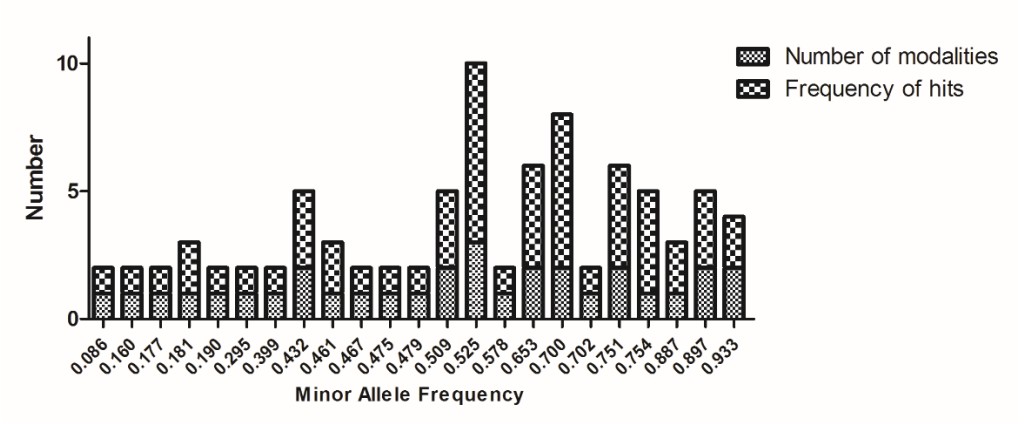


Figure 2-4 The relationship between single nucleotide polymorphism minor allele frequency and number of modalities predicted to affect the SNP site and the frequency of the modality at that site. Kolmogorov-Smirnov Test significant; not a normal distribution.

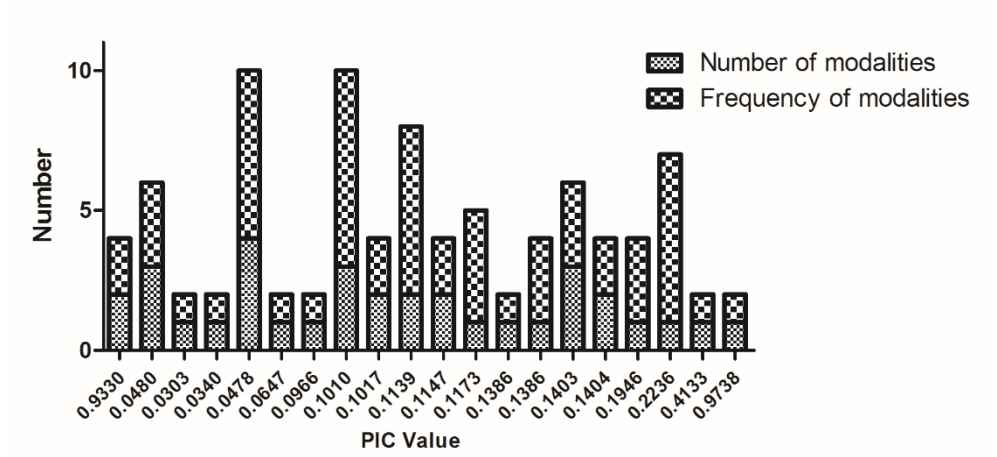


Figure 2-5 The relationship between single nucleotide polymorphism PIC value and number of modalities predicted to affect the SNP site and frequency of the modality at that site. Kolmogorov Smirnov Test not significant; normal distribution

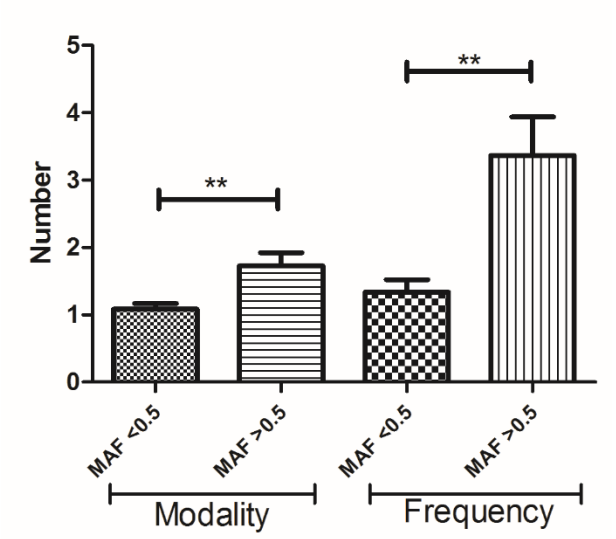


*Figure 2-6. Comparing number of modalities or frequency of hits affecting SNP sites with a minor allele frequency < or >0.5, with standard error bars and significance levels **P<0.01*

The 5 genes that were affected by 3 modalities or more were *RTEL1, IL17REL, NXPE1, CHP1* and *FCGR2A*. All are UC specific except *FCGR2A*, which is IBD specific (*NXPE1* and *CHP1* have no tissue localisation in the Broad dataset). These 5 proteins, when analysed in GIANT are found predominantly in mononuclear phagocytes and neutrophils, with the gene enrichment from mononuclear phagocytes highlighting intracellular pathogen pathways (Leishmaniasis pathways (p = 2.25 e-5), human tuberculosis pathways (p=9.6 e -4)), and immune function pathways (Phagosome pathways (p=5.61 e-4), regulation of monocyte chemotaxis (p=8.28 e-4), regulation of smooth muscle cell migration (p=9.07 e-4), muscle cell migration (p=1.01 e-3), monocyte chemotaxis (p=1.20 e-3), myeloid leukocyte migration (p=1.23 e-3), smooth muscle cell migration (p=1.35 e-3), leukocyte chemotaxis (p=1.54 e-3), and regulation of cell adhesion mediated by integrin (p=1.58 e-3)). The gene enrichment for neutrophils was equally interesting highlighting intracellular pathogens again (human tuberculosis p=9.88 e-4), *Staphylococcus aureus* infection (p=1.16 e-2)) and immune function (phagocyte bactericidal dysfunction (p=2.27 e-3), phagosome pathway (p=3.26 e-3), leucocyte cell-cell adhesion (p=3.28 e-3), granulocyte migration and chemotaxis (p=3.62 e-3), Natural killer cell mediated cytotoxicity (p=1.18 e-2), cytokine-cytokine receptor interaction (p=1.19 e-2 and myeloid leukocyte migration (p=1.26 e-2)).

There are 15 SNP affected genes with 2 or more modalities (11 UC specific genes; *IL23R, ITGAL, IL10, KIR3DL2, CARD9, C5orf66, MAML2, LSP1, GNA12, C1orf106, SFMBT1* and 4 IBD specific genes; *LPXN, TNFRSF14, CRTC3, CD6*). The GIANT analysis of these 15 genes identified tissue specificity to dendritic cells, natural killer cells, the caecum and the appendix. In the caecum, the only gene enrichment present was for Staphylococcus aureus infection (p=4.85 e-2).

For dendritic cells, infection pathways formed a major part of the gene enrichment (Tuberculosis p=8.81 e-3, Trypanosomiasis p=3.06 e-2, Epstein-Barr Virus infection p=3.29 e-2, Toxoplasmosis p=4.01 e-2, *Staphylocuccus aureus* infection p=4.27 e-2, Measles virus infection p=4.45 e-2, HTLV-1 infection p=4.92 e-2), as well as JAK-STAT pathways p=2.88 e-2, cytokine to cytokine receptor interaction p=2.81 e-2, Inflammatory Bowel Disease from 2 different sources (OMIM and KEGG) p=3.53 e-2 and p=4.82 e-2 and intestinal disease p=4.87 e-2. The NK cells all enriched for immune functions.

Infectious agents have been proposed as triggers for inflammatory bowel disease, including Cytomegalovirus (321), alpha hemolysin secreting *Escherichia coli*(322), Salmonella(323), *Mycobacterium avium* ssp *paratuberculosis*(324), Epstein-Barr Virus

(325), *Helicobacter pylori* (326) , enterohepatic helicobacter species (327), Varicella zoster virus (328), enteropathogenic viruses (140) and intestinal parasites (329). Pathobiont accumulation in the gut microbiota has also been proposed as an aetiological agent including expansion of Prevotellaceae, *Escherichia coli*, *Enterococcus faecalis*, *Bilophila wadsworthia*, *Desulfovibrio* spp. A reduction in abundance of Firmicutes, and *Faecalibacterium prausnitzii* have been identified as associated with IBD and IBD outcomes. What is not clear with all these proposed agents is whether these are triggers to, or a consequence of, the intestinal disease.

2.4.7 Immune responses, apoptosis and host-microbe interactions: The UC interactome takes shape.

The results from the gene enrichment indicate that the UC associated SNPs affect infection handling pathways, not necessarily indicating causative infectious agents, but could point towards a commonality and a synchronicity of the SNPs to particular pathogenic pathways. To examine this further and clarify phenotypic effect we created the UC interactome, identifying the first neighbours of the UC nodes e.g binding targets, enzymatic targets, signalling proteins and pathways downstream of the UC nodes themselves. Omnipath is the largest, curated protein-protein interaction and transcription factor-protein database available, combining a large number of smaller databases. The merged Omnipath-UC_SNP network contained 8058 nodes and 49149 edges. Once the non-first neighbour nodes were removed the UC interactome contained 338 nodes with 2023 edges (Figure2-7). It had 143 connected components, with a clustering co-efficient of 0.113 (highly clustered = 1.0, un-clustered =0.0). There were 5 isolated nodes, which were removed. There were no self-loops. As is normal with a large network, the gene ontology (using BiNGO in Cytoscape) identified large number (n=1344) of biological processes which were enriched within the network. Most of these were non-specific e.g biological, cellular or metabolic processes, however several themes were over represented in the enrichment, particularly in the first 500 enrichments (adjusted p ranged from p=6.6526E-47 to p=6.1878E-5). The themes were signalling (12%; 4.7% kinases, 0.1% kinases involved in NF kappa B regulation), immune response (8.9%), induction and regulation of apoptosis/programmed cell death (4%), handling of infectious agents (3.8%), response to wounding (2.2%), the WNT pathway (0.74%), and autophagy (0.2%).

Honing the network to just the genes/proteins with >2 modalities predicted at the SNP site and their first neighbour (82 nodes, 289 edges); there were 768 gene ontology enrichments (correct p=2.8561E-26 to p=4.9178E-2). The most prevalent theme was immune response (19.5%) which, unlike the larger network, included gamma delta T cell differentiation and regulation, alpha beta T cell differentiation and regulation, production and response to interleukins 1,2,3,6,8,10,12,17 and 18 as well as mucosal immunity. Signalling themes accounted for 10.5%, including small GTPase pathways, RHO and RAS pathways, as well as Toll Like Receptor 2 and 4 signalling pathways, which could equally be counted in the response to an infective agent which accounted for 7.6% of the ontology. 1% of the 7.6% was the production or regulation of antimicrobial peptides.

Apoptosis or cell death accounted for 3% of the Gene Ontology. The last cohesive group was the cell matrix, adhesion or integrin (not signalling) themed group accounting for 1.5%. Autophagy did not feature in this subgroup.

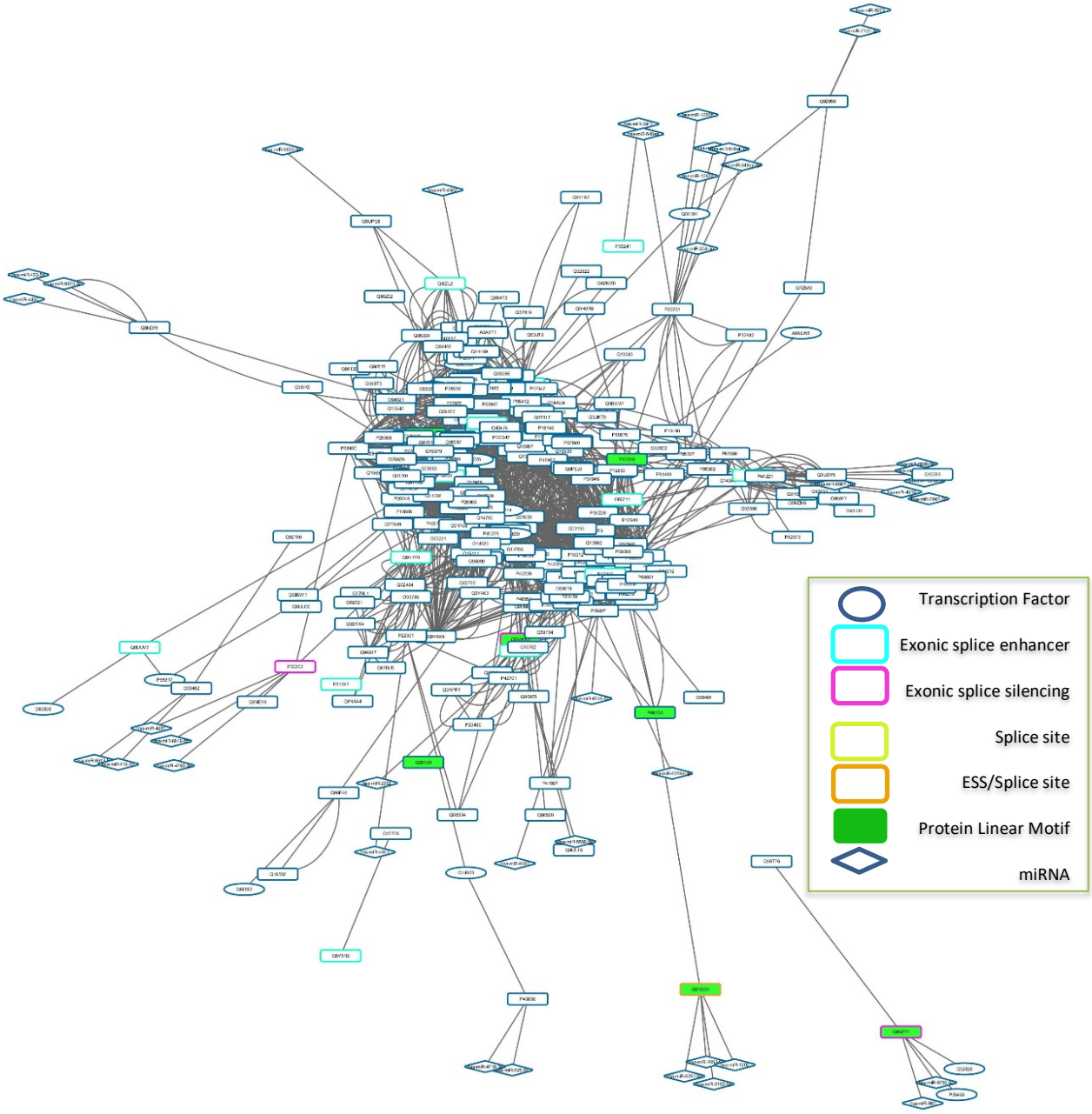


Figure2-7 The UC Interactome – identified from the merged Omnipath-UC network, by identifying the downstream first neighbours of the UC network in a perfuse force directed layout. Although not readily readable, the complexity of the network can be readily appreciated. UC SNP Nodes and Omnipath first neighbours, are denoted by their Uniprot IDs. The node shape denotes either a protein (rectangle), a miRNA whose binding site is affected (diamond) or transcription factor whose binding site is affected (oval) by SNPs. Some proteins are affected by protein linear motifs (ELM) (green blocks), SNPs affecting slice sites in mRNA have a coloured outline to the node. mRNA with splice enhancer sites affected by SNPs have a turquoise border, mRNA with splice silencing sites affected by SNPs have a pink border and mRNA with splice motifs affected by SNPs have an olive border. mRNA with both a splice motif and a splice silencing site affected by a SNP have a light brown border.

2.4.8 The UC interactome – Tight Junction Analysis

TLR4 expression is low in colonic intestinal epithelial cells, the receptors are found in the basolateral surface of the enterocytes, however in active CD they are found apically. TLR2 expression in the colon is also low – it is found mainly in colonic crypts. The role of TLR2 in the colonic epithelium is preservation of tight junction structure. The tight junction did not feature as a gene ontology theme, however regulation of wound healing and signalling cascade associated with regulation of actin filaments did feature. The tight junction subgroup network (figure 7) identified 2 UC associated nodes which were tight junction nodes: GNA12, MST1. These are both signalling proteins involved in tight junction regulation, GNA12 had 8 binding partners (high degree centrality, low betweenness centrality). Interestingly MST1 was a stand alone node, this could be explained by the directionality of the network as we were attempting to identify downstream binding partners only. There were ten additional UC associated nodes which were first neighbours to tight junction nodes (table 11). There were 33 nodes based around 8 components. More importantly there were four tight junction nodes which are putatively affected by 2 different SNP affected proteins, 3 kinases (protein kinase C alpha, -SRC and FYN) and one transcriptional regulator, Jun. The c-SRC kinase node is affected by both protein tyrosine phosphatase N2 (PTPN2) node and Stat3 node, which are both predicted to have higher expression compared to the non-risk allele via loss of splice enhancers. The Fyn kinase node is affected by both the PTPN2 node and the CD226 node, both of which are predicted to have higher expression secondary to the loss of a splice enhancer and the loss of a miRNA binding site, respectively. The protein kinase C node is affected by the CD226 node as described above and the GNA12 node which is also predicted to be over expressed due to a loss of a miRNA binding site. The Jun node is affected by STAT3 and NFKB1 in opposite directions (NFKB1 is predicted to be under expressed due to the gain of an ESE and STAT3 is due to be overexpressed due to the loss of an ESE). STAT3 and JUN, and NFKB1 and Jun are known to be transcriptionally co-operative in different pathways; the JAK/STAT pathway and the Activated TLR4 pathway respectively and together in the response to IL6.

Two proteins were identified as UC associated binding partners which were a scaffolding protein Actin, an integral protein of cell cytoskeletons, and an adaptor Par1 or Coagulation factor II receptor (F2R) which is involved in the maintenance and disruption of the endothelial barrier.

In terms of numbers of modalities affecting the UC nodes in the tight junction subnetwork, half of the UC nodes had >2 modalities. The four tight junction nodes with 2 UC binding partners were from predominantly singleton modalities (splicing or miRNA binding sites), except for GNA12 which has 2 modalities. The majority of findings are very non-specific as the kinases, transcriptional regulators and signalling proteins are ubiquitous. When analysed for UC specificity, only half of the nodes are UC specific; GNA12, NFKB1, MST1, ITGAL, NXPE1 and CHP1. The other half were IBD specific; CD226, IKZF3, CD6, PTPN2, LPXN and STAT3. The network is, therefore, made up of a small number of UC specific associated proteins, and as suggested by the prior gene ontology work, it is not clear from this methodology how much of the pathological driving force behind UC specific disease is dysregulation of tight junctions.

Stratification using minor allele frequency (where available), number of modalities effecting the SNP site and UC specificity of the most important nodes, identifies that GNA12 (over expressed) would be a potential candidate for experimental validation, as would LPXN (over expressed) from the IBD side.

GNA12 is a membrane bound GTPase, which binds to the tight junction protein ZO-1 and activates SRC to increase paracellular permeability. Tight junction integrity is disrupted by GNA12 stimulated SRC phosphorylation of ZO-1 and ZO-2, which leads to dissociation of two important tight junction proteins claudin 1 and occluding from the ZO-1 complex (330). GNA12 also has a role in the adaptive immune system in mouse models of T cell mediated pathology where inactivation of GNA12 leads to an increased activity of integrin leukocyte function antigen 1 in murine CD4+ T Cells and lymphadenopathy sue to increased lymph node entry and enhance T cell proliferation (331). GNA12 also has a role in LPXN translocation to focal adhesions (332).

LPXN is a member of the paxillin protein family, is a thrysoine phosphoprotein which forms a part of the focal adhesion complex, serving as a scaffold to focus and regulate specific effector molecules (kinases) to a subcellular location. Via this mechanism it mediates the phosphorylation of the actin binding protein CaD via the extracellular signal regulated kinase(ERK) 1 and 2 pathway (333). When phosphorylated by Lyn, LPXN also activates the JNK pathway which leads to phosphorylation of ZO-1 and dissociation of the tight junction (334).

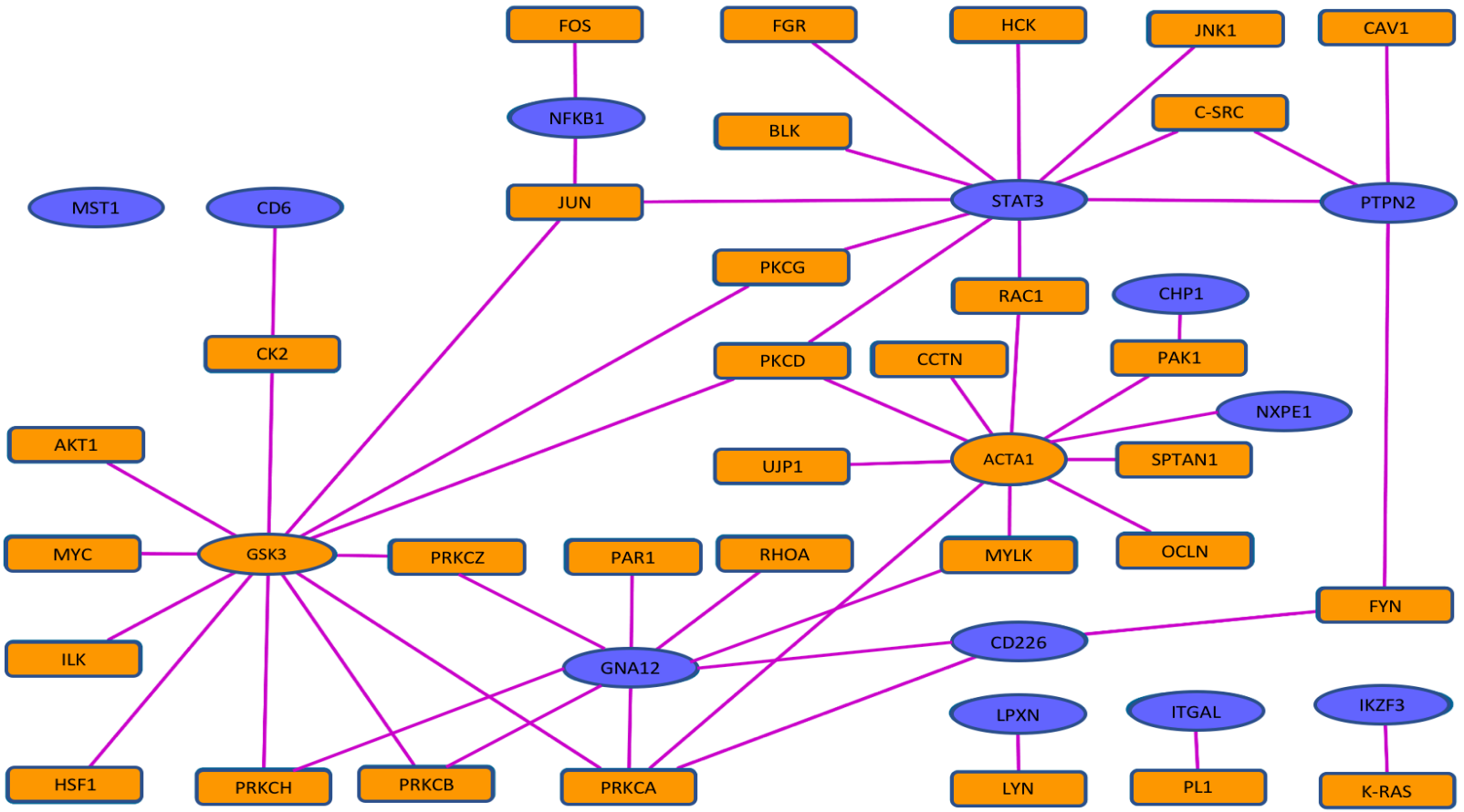


Figure 2-8 UC interactome-Tight Junction associations. The blue ovals are the nodes from the UC interactome. The orange ovals are the first neighbours of UC affected proteins. The orange rectangles are tight junction associated proteins. More details regarding the tight junction associated protein functions can be found in the table below.

CHP1

UC Gene Name	Uniprot ID	Modality	Effect	TJ Uniprot	Name	TJ function
CD226	Q15762		Loss_up	P17252	PKCA	Kinase
				P06241	FYN	Kinase
IKZF3	Q9UKT9	TFBS	Gain_up	P01112	GTPase H-RAS	Signalling
NFKB1	P19838	ESE	Gain_down	P01100	Fos	Transcriptional regulators.
				P05412	Jun	
CHP1	Q99653	TFBS	gain_up	Q13153	Pak1	Kinase
CD6	P30203	ESS	gain_up	P68400	CK2	Kinase
PTPN2	P17706	ESE	Loss_up	P06241	FYN	Kinase
				Q03135	CAV1	Signalling
				P12931	c-SRC	Kinase
LPXN	O60711	TFBS	Gain_up	P07948	Lyn	Kinase
STAT3	P40763	ESE	Loss_up	P09769	Fgr	Kinase
				P51451	Blk	Kinase
				P08631	Hck	Kinase
				P45983	Jnk1	Kinase
				P12931	c-SRC	Kinase
				P63000	Rac1	Signalling
				P05129	PKCgamma	Kinase
				Q05655	PKCD	Kinase
				P05412	Jun	Transcriptional regulator
NXPE1	Q8N323	multiple	up and down	P68133	Actinin	Scaffolding/Adaptor
ITGAL	P20701	multiple	down and up	P25116	Par1	Adaptor

Table 2-11 UC-Tight Junction proteins and general functions

2.4.9 The UC Interactome: Apoptosis-Autophagy cross talk.

Programmed cell death/apoptosis featured highly within the gene ontology with the network pathways including positive regulation of caspase activity, positive and negative regulation of the Jun N-terminal Kinase (JNK) pathway, RAS and GTPase pathways, positive and negative regulation of apoptosis, cell death and regulation of programmed cell death. Autophagy and apoptosis often occur in the same cell, with multiple common upstream signals including JNK which phosphorylates BCL-2, promotes apoptosis, as well as decreasing BCL-2 inhibition of Beclin 1, leading to an active Beclin 1-VPS34 complex and autophagy. There is also shared downstream proteins in both networks e.g ATG5 is an autophagy protein, which also generates an amino-terminal fragment that can sensitise cells to apoptosis (via mitochondria) and precursors of proteins that, with caspase activation can accelerate the apoptotic process.

The UC Interactome-Autophagy subnetwork (Figure 2-9), when visualised as a directed network identified 12 UC associated genes which interact with major regulators of autophagy e.g. BCL-2 and BCL-3 (Table 2-12), only ATG5 was the only autophagy and UC overlapping node. The modalities are all represented in the network Figure 2-8, with splicing changes forming the majority. The nodes highest betweeness centrality (the most important nodes for the network) are two UC nodes STAT3 and STAT1, as shown by figure 2-10. The larger the circle, the greater the betweeness centrality and the more important that node is to the network. Both Stat1 and Stat3 activation are regulated by PTPN2. Stat3 has been implicated in multiple steps in autophagy depending on the cellular site of stat3. It is the main transcriptional enhancer of several autophagy related genes including BCL2, BECN1, and PIK3C3, all of which are anti autophagy. STAT3 within the UC interactome is predicted to have increased expression. STAT3 is known to be overexpression in UC colonic biopsies, as compared to normal colons and even more so in inflamed UC colonic specimens. It is a significant transcriptional regulator interleukins IL6, IL22 and IL23, and is the signal transductor from the IL10, IL27 and IL6 receptors. The STAT3 SNP has a minor allele frequency of 57.8% (0.578), indicating that a significant percentage of UC patients tested had this allele, whereas the regulator PTPN2 (which is predicted to be over expressed) has a MAF of 16%, therefore globally would not be having such a large impact on the UC population, and indeed the regulation of stat3 will not be the only effect. Stat3 is also significantly implicated in the apoptosis pathway.

The apoptosis_UC interactome subnetwork (Figure 2-11) is made up of 2 groups; the NKFB1/MAML2/LPXN/CRTC3 group and a GNA12/CARD9/STAT3/IL23R group. There is

significant overlap in the regulators of autophagy and apoptosis from within the UC interactomes (

Figure 2-12). Once again, to get an overarching view of the pathogenesis of UC, looking at the nodes with the highest minor allele frequency e.g STAT3 57.8% (increased), IL23R 93% (increased), GNA12 70% (increased), CRTC3 89% (increased) , these overlap as regulators in both autophagy and apoptosis, therefore depending on the incoming signal the SNP may affect the cellular decision to undergo autophagy or apoptosis and dysregulation of either has been shown to have a negative impact on intestinal barrier function (335, 336).

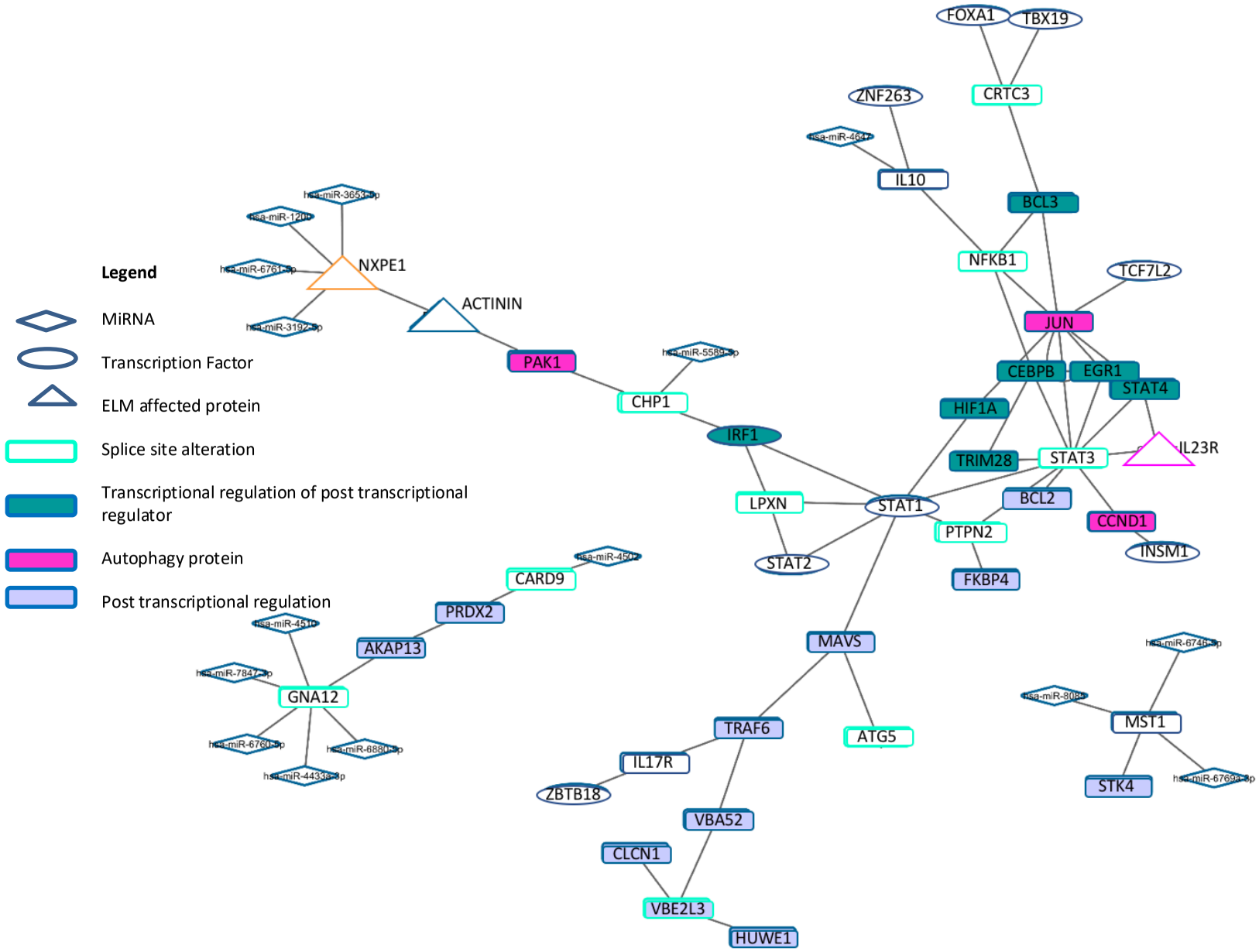


Figure 2-9 UC-interactome-autophagy undirected network. The nodes are from the UC interactome overlapping with autophagy regulatory network(ARN). The UC nodes with their miRNA or transcription factor which has a binding site altered by a SNP, and the first neighbours from Omnipath which appear in the ARN. Each protein node present in the ARN is coloured to the function of the protein.

UC gene Name	Uniprot ID	Effect	Autophagy Uniprot	Name	Function
STAT3	P40763	loss_up	Q13263	TRIM28	Transcriptional regulation of post transcriptional regulator
			P18146	EGR1	Transcriptional regulation of post transcriptional regulator
			P17676	CEBPB	Transcriptional regulation of post transcriptional regulator
			Q14765	STAT4	Transcriptional regulation of post transcriptional regulator
			P24385	CCND1	Autophagy protein
			P10415	BCL2	Post transcriptional regulation
			Q16665	HIF1A	Transcriptional regulation of post transcriptional regulator
			P05412	Jun	Autophagy protein
IL23R	Q5VWK5	GAIN_UP	Q14765	STAT4	Transcriptional regulation of post transcriptional regulator
PTPN2	P17706	LOSS_UP	Q02790	FKBP4	Post transcriptional regulation
ATG5	Q9H1Y0	LOSS_UP	Q7Z434	MAVS	Post transcriptional regulation
			Q7Z6Z7	HUWE1	Post transcriptional regulation
MST1	P26927	LOSS_UP	Q13043	STK4	Post transcriptional regulation
IL17R	Q96F46	GAIN_ DOWN	Q9Y4K3	TRAF6	Post transcriptional regulation
UBE2L3	P68035	GAIN_ DOWN	O75592	CLCN1	Post transcriptional regulation
			P62987	UBA52	Post transcriptional regulation
			Q7Z6Z7	HUWE1	Post transcriptional regulation
CARD9	Q9H257	GAIN_ DOWN	P31946	PRDX2	Post transcriptional regulation
GNA12	Q03113	LOSS_UP	Q12802	AKAP13	Post transcriptional regulation
CRTC3	Q6UUV7	GAIN_UP	Q13153	Pak1	Autophagy protein
			P20749	BCL3	Transcriptional regulation of post transcriptional regulator
NFKB1	P19838	GAIN_ DOWN	P20749	BCL3	Transcriptional regulation of post transcriptional regulator
			P17676	CEBPB	Transcriptional regulation of post transcriptional regulator
			P05412	Jun	Autophagy protein

Table 2-12 Directional autophagy network binding targets from UC SNP associated genes

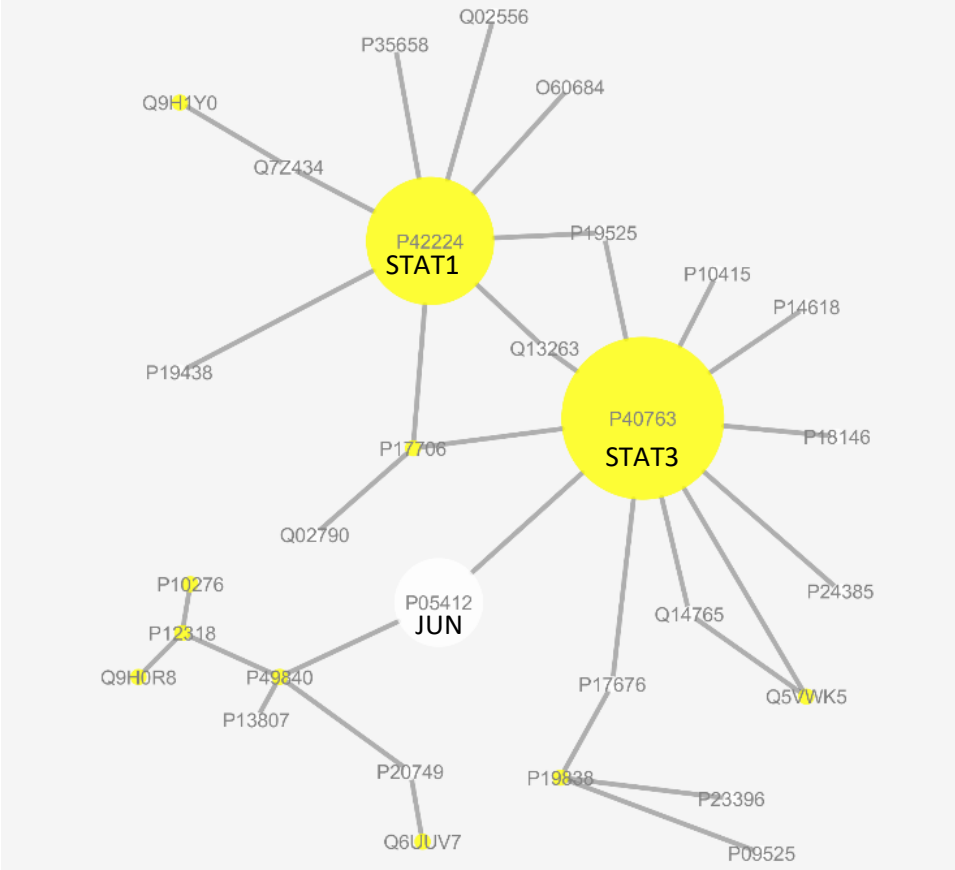


Figure 2-10 Betweenness centrality of the major autophagy UC cluster in a directed network. UC nodes are yellow. The larger the circle, the higher the between-ness centrality, therefore the more important the nodes are to the cohesiveness of the network.

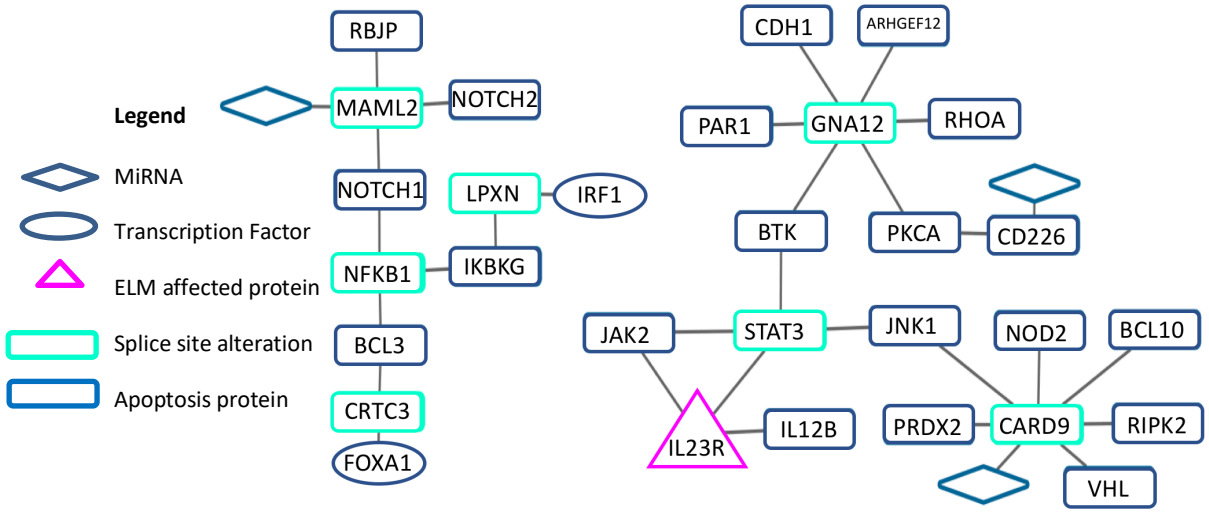


Figure 2-11 UC-interactome-apoptosis undirected network. The nodes are from the UC interactome overlapping with apoptosis network. The UC nodes with their miRNA or transcription factor which has a binding site altered by a SNP, and the first neighbours from Omnipath which are also apoptosis proteins.

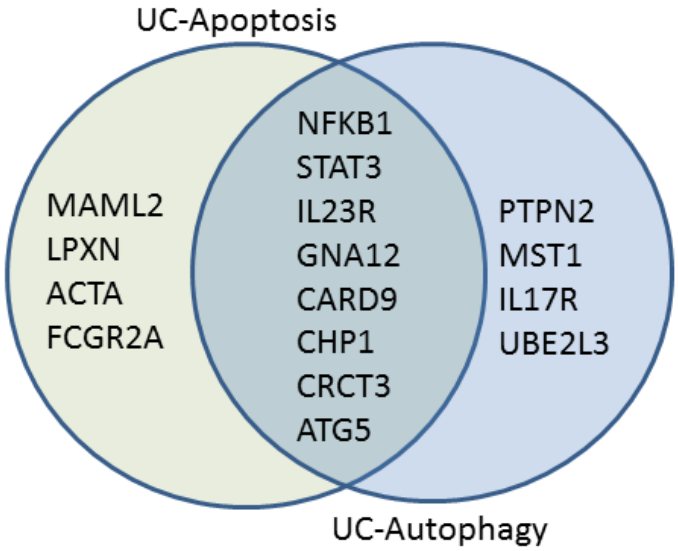


Figure 2-12 Venn diagram identifying the overlap of UC proteins with first neighbours in both autophagy and apoptosis pathways (using protein names for ease of reference).

2.4.10 The Focal Adhesion Complex within the UC interactome

Integrin receptor based adhesion and adhesion to the cell matrix have been identified consistently within the UC SNP gene ontology analysis. The focal adhesion complex (FAC) is a potential signalling and scaffolding link between tight junctions, apoptosis and autophagy. Unlike all the other networks, which had isolated nodes, or multiple separate smaller networks, the FAC_UC network (Figure 2-13) is a small but cohesive network. Four nodes are both UC and FAC nodes: LPXN, Stat3, NFKB1 and PTPN2. Like the other networks, signalling regulation forms the backbone of the network as seen with the kinase and phosphatase nodes been predominant within the network. We can see a direct impact on protein-protein interaction with integrins and important cytoskeletal proteins via PTPN2 which phosphorylates integrin alpha 5, and caveolin1. Caveolin 1 is major protein component of caveolae, which are small invaginations in cell plasma membranes which are a type of lipid raft. Caveolae undertake clathrin independent-raft dependant endocytosis which is exploited by Echovirus, Coronovirus, Rotavirus and by some strains of *E.coli*. Another structural protein affected by protein-protein interactions with UC risk protein is vinculin. LPXN binds vinculin, an integral cystoskeletal protein of the focal adhesion, via it's LIM domain. Vinculin links the talin-integrin complex with F-actin thereby controlling the transmission of extracellular mechanical cues to the actin cytoskeleton, which includes cues to polarise and response to mechanical shear force including wound healing.

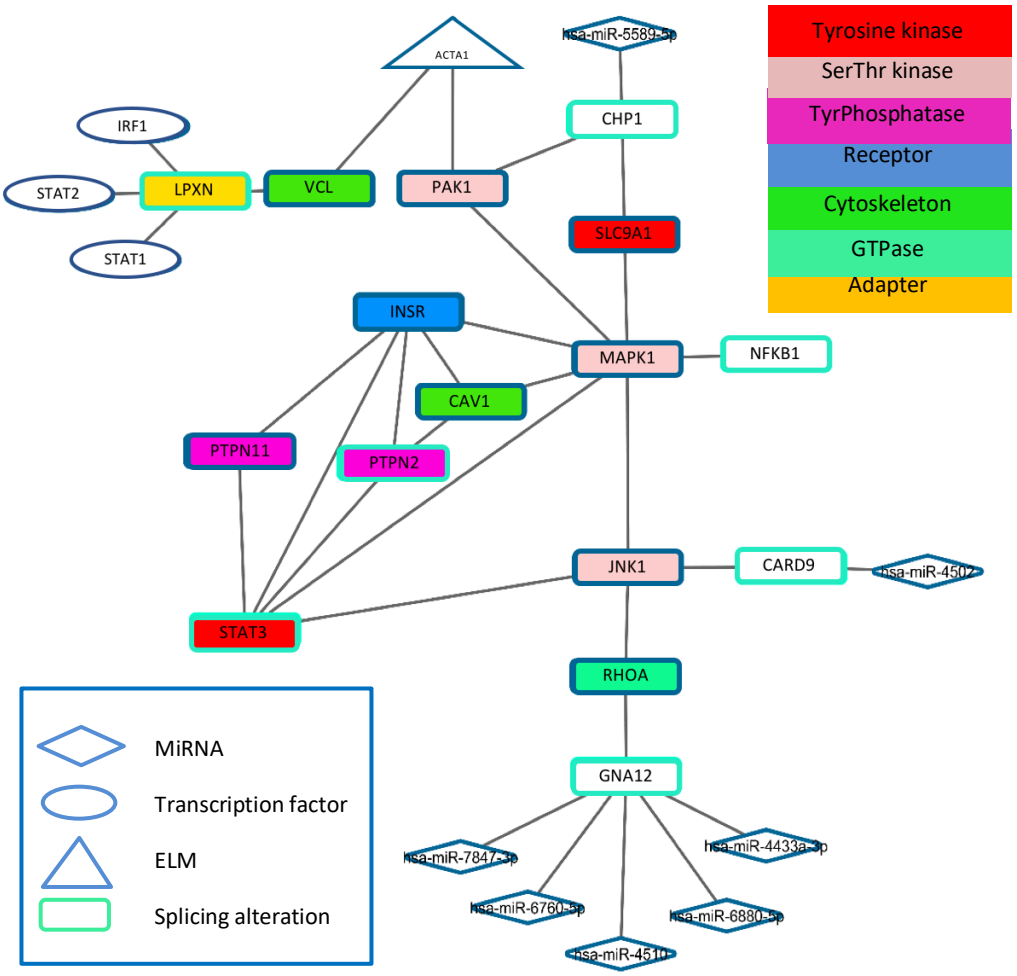


Figure 2-13 UC-interactome-focal adhesion complex undirected network. The nodes are from the UC interactome overlapping with the adhesome network. The UC nodes with their miRNA or transcription factor which has a binding site altered by a SNP, and the first neighbours from Omnipath which appear in the adhesome. Each protein node present in the adhesome is coloured to the function of the protein.

2.5 Discussion

UC is a heterogenous disease in terms of clinical phenotype, response to medication and outcome as described in Chapter 1. Despite the heterogeneity, there are a handful of SNPs that re-occur across multiple networks, have more than two predicted modalities for their SNP effect and have a high minor allele frequency. This is almost counter-intuitive, as genetic selection would suggest that a SNP that has a strong phenotype would be selected against and therefore within the population would have a lower minor allele frequency. I would argue that the modalities we have identified produce weaker phenotypes *via* transcriptional, post transcriptional and translational regulation and that the cumulative effect is key, not the individual SNP affect. A case in point are the rare variants in CARD9, and IL23R. Using different techniques, we confirmed the splice site alteration in CARD9 which has been shown to produce a truncated protein, however it is in a very small number of UC patients. Accepting the significant bias of the hypothesis driven networks, neither CARD9 nor IL23R were the most key components of any of the networks we analysed. Taken out of context the SNP in PTPN2 would be an antithesis to this argument, it is a relatively rare variant, has a low modality yet it is a component of each of the subnetworks we have analysed. It also has roles in haematopoiesis, T cell receptor signalling, cytokine mediated signalling among multiple other signalling cascades. As such there is redundancy as other phosphatases can 'cross cover' PTPN2. Unfortunately, like STAT3, it would appear to be so ubiquitous it would feature in the networks regardless of the hypothesis. Interestingly PTPN2 knock out mice display severe impairment of T cell development and exhibit severe inflammation and survive only 5-6 weeks after birth associated with a loss of the epithelial barrier. Our results indicated that the UC associated SNP leads to the loss of an exonic splice enhancer (consequence could be a retained intron at worst) which is difficult to equate with a loss of PTPN2 function.

The results from the network analysis confirm in broad strokes what has been postulated previously in terms of the host-microbe interactions, interestingly highlighting viral-host interactions as a potential avenue of exploration, and the role of immune regulation in UC. It has also suggested further evidence for the role of the UC associated SNPs in the 3 components of the maintenance of the epithelial barrier; tight junctions, autophagy, and potentially joining it all together, the focal adhesion complex.

In terms of stratification for experimental validation, we need to apply three key concepts:

1. They have to be a significant component of the network

2. The hypothesis has to be clear and therefore they have to produce a clear downstream validateable phenotype

3. The modality has to be experimentally testable.

There were 4 SNP affected proteins which were in all the subnetworks; NFKB1, STAT3, PTPN2 and LPXN. NFBK1, STAT3 and PTPN2 are all regulatory with multiple complicated downstream signalling effects as well as being difficult to test modalities (splicing enhancement). LPXN, although an IBD node, had both a putative splicing enhancement as well as strong evidence for transcription factor binding sites for IRF1, STAT1 and STAT2 and there are clear documented roles of LPXN within the focal adhesion complex.

There are limitations of this workflow. The workflow is dependent on high quality GWAS, finemap or deep sequenced SNPs. The workflow itself is not biased and cannot differentiate between poorly associated and causal SNPs, therefore care must be taken as to the data input into the workflow. Consideration was applied to this when separating the data from the parent cohort and the extended parent cohort, however given the broad overview of the interactome utilising the IBD SNPs as well as the UC associated SNPs not just localised to the colonic mucosa gave more cohesion to the networks.

For each modality, we used more than one *in* silico technique to functionally annotate the SNPs. Each technique has its own stringency and statistical validation and where possible we used a maximal concordance rule e.g. the two techniques had to agree and if more than two techniques were used, then the majority result went forward. Where there was not a comparable technique available, then the stringency of the network was based upon the fallibility of one technique e.g. the splicing silencing motifs. This can lead to a high false positive hit rate. Within each of the techniques, where available we used the most stringent settings, so it is possible that novel, weaker binding sites for miRNA and transcription factors have been missed, but the pay-off was lower false positive rates. Each technique has its own limitation for example ELM did not identify a specific ligand binding site change in GPR65, despite Polyphen identifying thr139met as possibly damaging, but the transmembrane domain was impacted. Emtage et al (337)have shown that standard computational docking tools for ligands in GPCRs are limited as ligand binding pockets are dynamic and using a pressurisation modelling method they identified that minor changes in amino acid side chain orientations can open a fissure between transmembrane helices, which can then accommodate ligands.

The mature mirNA seed sites were a significant limitation as multiple mature miRNA sequences were identified but could not be used as we were unable to confirm that the full stem-loop sequence was present, or that the mature miRNA was at a documented locus for that miRNA, therefore given the significant number of targets each miRNA has, these were removed from network analysis. This produces a potential gap within the network, as we've already seen that regulation is a key component of the UC interactome function.

We were also unable to incorporate lncRNAs within the network due to the lack of experimentally validated datasets or accessible sequence databases which would allow us to identify impact of SNPs on the lncRNA sequence.

There is clearly a bias to the subnetwork analysis, we were using a network to create hypotheses within the broad hypothesis that UC associated SNPs have an impact on the intestinal epithelial barrier. This bias was necessary to create focus for the research, as the interactome is broad enough to create multiple avenues of research.

Peters (54) et al 2017 created an integrative approach for constructing a predictive network model of IBD. They utilised a paediatric cohort of CD patients (n=322), as well as an TNF refractory CD cohort (n=118), and an advanced IBD population (n=134), and utilising transcriptomics and probabilistic causal gene networks (Bayesian networks) to identify co-expression networks and enriched risk genes within those networks. They then utilised those networks to stratify a key driver genes list to validate in macrophages, as a hypothesis driven approach. Our approach globally takes a similar route of hypothesis free network creation, then hypothesis driving to stratify a list of genes to validate, however we differ significantly in our methodology. We utilised a hypothesis free approach to create the UC interactome, integrating proteomic data post functional annotation of SNPs as opposed to transcriptomic data with gene enrichment. Unlike Peters et al, who looked specifically at macrophage specific signatures in a specific subset of CD patients with severe disease, we looked for enrichments and pathways involved in maintenance of the colonic mucosa across the whole UC disease severity spectrum. Utilising transcriptomics from colonic mucosa from quiescent UC patients and integrating it with the UC interactome would provide a significant wealth of data and although for further stratification and hypothesis driving from the UC interactome and will be done in the future work.

2.6 Conclusion

UC appears to be a disease predominantly of regulation, and such may be difficult to identify phenotypic effects of the SNPs. By utilising a network medicine technique, we have added further evidence to the current literature on SNP effects and created a tool by which pathological hypotheses can be obtained and stratification of SNPs to experimentally validate can be undertaken. Further work to tease apart the SNP effects in clinical cohorts are described in the follow chapters.

3. Validating the UC network *in vitro* and *ex vivo*.

Overarching aim:

To experimentally validate a stratified SNP identified as relevant within the UC-Ome

3.1 Introduction

3.1.1 Techniques of validation: *in vitro and ex vivo techniques*

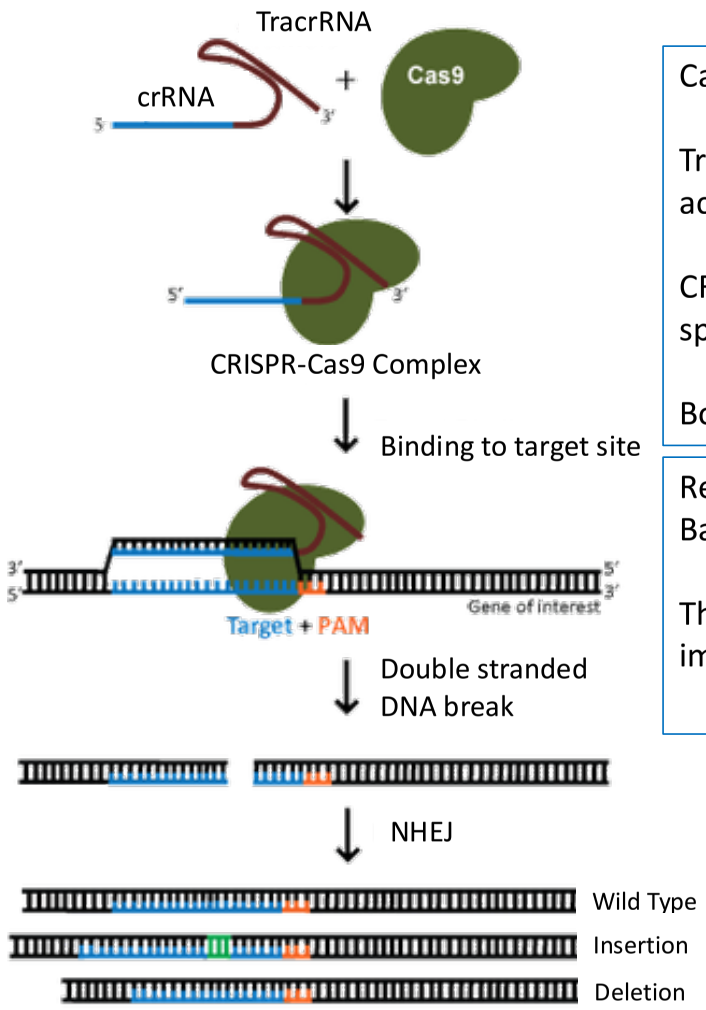
3.1.1.1 In vitro techniques

Leupaxin (*LPXN*) SNP rs10896794 was identified as a candidate for evaluation from the bioinformatics workflow described in chapter 2, where the hypothesis was that SNPs had an impact on the intestinal epithelial barrier function. To validate the *LPXN* SNP as playing a role in the pathogenesis of UC and thereby show that the global UC-ome technique produced relevant results we had to evaluate whether LPXN has a role in epithelial cell function. The commonest techniques to do this in isolation are utilising intestinal epithelial cell models such as Caco2, HT29 and T84. Immortalised cell lines have been used to characterise animal models of colitis such as using DSS on caco2 cells (338)Specific cells such as dendritic cells and cell lines such as THP1 have been used to investigate pathogenesis pathways of colitis such as enteric cell education of dendritic cells (339) and identifying how carrageenan works on THP1 cells to cause the phenotype seen in carrageenan induced colitis mouse models(340).

The hypothesis is that the *LPXN* SNP affects *LPXN* expression. Techniques to assess the impact of changes in gene expression include strategies such as siRNA knock downs vs transient ectopic overexpression plasmids or using Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) technology to create over expression and knock out (as opposed to knock down) cell lines. The more complex strategy utilises CRISPR to create 'SNP' cell lines that contain the risk allele of interest.

CRISPR and CRISPR associated nucleases (Cas) in bacteria provide adaptive immunity against viruses and plasmids with their RNAs are used to guide the Cas cleavage of foreign nucleic acids (ref). The Cas nuclease Cas 9 (originally found in *Streptococcus pyogenes*) can cleave a strand of double stranded target when directed by single guide RNAs that consists

of the CRISPR RNA and a transactivating RNA (Figure 3-1). The double stranded DNA breaks can be repaired by cellular repair machinery either by non-homologous end joining (NHEJ) or homology directed repair(HDR), which depends on the presence of a repair template. NHEJ requires less energy expenditure (taking less than 30 mins as opposed to approximately 7 hours like HDR), and can introduce unpredictable insertions or deletions creating knock out cell lines. HDR uses a template with desired changes to make mutations, such as changing alleles to SNP risk alleles in the genomic loci. HDR occurs less frequently than NHEJ as it occurs only in S and G2 phases, but can be encouraged by the use of exogenous HDR promoters such as brefeldin (341). To encourage HDR further CRISPR Cas9 nickases can be used with sense specific sgRNAs to cause single DNA strand nicks in opposing DNA strands thereby “dropping out” a segment of DNA which have to be repaired by HDR, as recently published by Zhang et al (342).



Cas 9 endonuclease (*S. pyogenes)*

Transactivating RNA (tracrRNA) – activates Cas9 enzymatic activity (scaffold sequence)

CRISPR RNA (crRNA or spacer) – determines substrate specificity

Both RNA together = sgRNA

Requirements:
Base pairing between 5' end of sgRNA and target DNA

The presence of a protospacer adjacent motif (PAM) NGG, immediately down stream of the target sequence.

Repair by Non-Homologous End Joining (NHEJ) is Intrinsically mutagenic
Takes approx. 30 mins cf 7 hours for Homologous Recombination (HR)

Figure 3-1 The CRISPR-Cas9 system with non-homologous end joining. Adapted from addgene.org this diagram shows how the basic CRISPR Cas-9 gene editing system works using a fully active Cas9 to cause double stranded DNA breaks.

The CRISPR system has been further engineered for activation of endogenous gene expression. It has been engineered in two ways, firstly, the Cas9 nuclease has been deactivated(dCas9), and is fused to a transcription activation domain (VP64). Using sgRNAs to direct the dCas9-VP64 complex to a -200bp region from the transcriptional start site of the desired gene to upregulate gene expression. Secondly, the sgRNA hairpin has had an aptamer added to it which binds MS2 bacteriophage coat proteins. On the 'activation plasmid' is a MS2-P65-HSF1 fusion protein. P65 and HSF1 bound to the sgRNA via the MS2 protein enhance recruitment of transcription factors, thereby causing further gene activation.

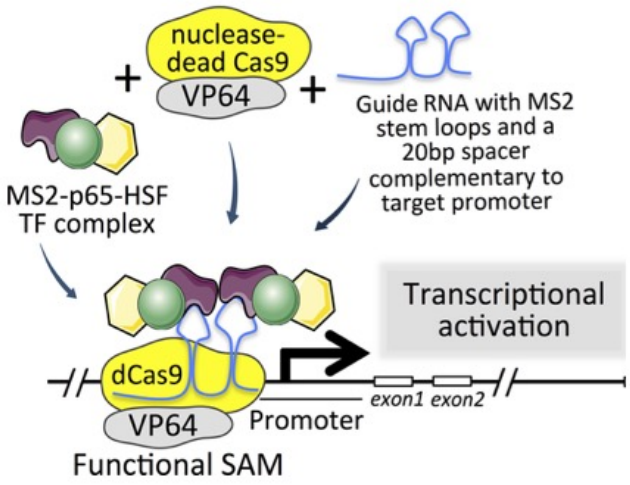


Figure 3-2 Diagram of the CRISPR-Cas9 activation mechanism, whereby a transactivation domain is fused to a deactivated (dead) Cas9. The guide RNA directs the Cas9 to 200bp upstream from the transcriptional start site of the gene of interest. The guide RNA has been altered to contain an aptamer that binds to MS2 proteins. On the commercial plasmid is a MS2-p65-HSF fusion protein which forms a large transcription factor complex and when the MS2 binds to the aptamer on the guide RNA, the enhanced recruitment of transcription factors leads to increased transcription of the gene of interest. (Image adapted from Kaczmarczyk et al PLos ONE 2016).

3.1.2 *Ex vivo* techniques

SNP analyses have been undertaken extensively in peripheral blood cells taken from both healthy human subject and those with diseases of interest. In CD, genotyped peripheral blood mononuclear cells have been studied for the role of a non-coding SNP within FOXOP3 in T cells(343), but analysis of SNP effect on intestinal mucosa remains challenging. For analysis the effect of SNPs on the function of the colonic mucosa, it

requires more than the analysis of intestinal epithelial cells as these cells work in concert with goblet cells, immune cells and the extracellular matrix as well as the milieu of cytokines and chemokines produced in response to microbial stress. In vitro organ culture utilises colonic biopsies from human patients as a model of the functioning human mucosa which can be kept alive for up to 8 hours. This provides a snapshot of the function of the intestinal epithelial cells and any immune cells present in the mucosa at the time of biopsy. Cytokine analysis of the mucosal biopsy has been shown to be a useful tool for identifying pathogenic responses when analysing lamina propria lymphocyte cytokine responses in IBD (344), polarised organ culture of mucosal biopsies has been used to create a secretory cytokine readout for treatment efficacy analysis(345). We plan to use polarised in vitro organ culture to characterise the role of the UC associated *LPXN* risk allele on cytokine production.

3.1.3 The role of LPXN within the cell

LPXN is a 43kDa phosphotyrosine protein member of the focal adhesion complex functioning as an adapter molecule. It has multiple different roles depending on the cell type it is expressed in. In T cells it binds to PYK2 (a focal adhesion kinase) which modulates integrin dependent adhesion in response to integrin engagement, T cell receptor engagement or chemokine stimulation (346).In B cells LPXN inhibits B cell receptor signalling and IL2 secretion (334). In osteoblasts LPXN is a critical component of the podosomal signalling complex by binding to Pyk2 and pp125FAK and protein tyrosine phosphorylation-PEST (PTP-PEST)(347, 348). Overexpression of LPXN in prostate cancer cells resulted in an increased associated with Pyk2 with increased cellular migration(347). In macrophages Pyk2 has roles in migration, F-actin localisation an Rho and PI-3 kinase activation in response to integrin mediated cell adhesion (349) TNF-alpha stimulated phosphorylation of LPXN leads to the production of cytoplasmic projections at the leading edge of the cell (350). LPXN has also been shown to modulate migration of cancer cells by phosphorylation of actin associated proteins, including caldesmon (333). In prostate cells, LPXN has also been shown to be a transcriptional regulator of the androgen receptor (351).

LPXN is expressed in multiple cell types and in multiple cancer cell lines, including colorectal cancer cell lines. siRNA knock down of *LPXN* in MDA-MB-231 cell lines inhibited adhesion to both collagen I and fibronectin, but only inhibited spreading of cells on collagen I not fibronectin (332) However, in other cell lines *LPXN* expression reduced

spreading of NIH3T3 on fibronectin and in K562 cells LPXN supressed a5b1 mediated cell adhesion to fibronectin (352).

LPXN is known to share structural characteristics with paxillin, including LIM domains, and functional modules including leucine and aspartate (LD) motifs and like paxillin is known to localise to focal adhesions upon cell adhesion to fibronectin. The molecular motifs that LPXN and paxillin have in common form functional modules which bind to Pyk2, FAK, SRC, LYN, and PTP-PEST; all of which have roles in integrin signalling. Integrin engagement with its receptor triggers tyrosine phosphorylation of paxillin which generates SH2-binding sites for other SH2 domain containing focal adhesion proteins and Rac1 signalling, potentiating the integrin signal mediated by mitogen-activated protein kinases (MAPKs). LPXN lacks the homologous tyrosine residues that are phosphorylated in paxillin and its expression reduces tyrosine phosphorylation of paxillin thereby supressing the integrin signalling. Specifically LPXN has been shown to suppress integrin a5b1 function including cell adhesion to fibronectin and cell spreading (352).

3.1.4 The focal adhesion complex and the NLRP3 inflammasome

Although there is a paucity of information regarding LPXN function in multiple cells types, the role of the focal adhesion complex (FAC) that LPXN functions within is better understood. FACs are large dynamic protein assemblies and scaffolds that mechanically link and transduce signals from the extracellular matrix to the internal cell via receptor modules. The complex participants are most easily characterised by their function; structural proteins such as paxillin, talin, actinin, vinculin, or dynamic signalling protein including protein kinases such as PYK2, phosphatases, small guanosine triphosphatases (GTPases) as regulatory molecules, or adapter molecules that mediate core protein-protein interactions.

Within intestinal epithelial cells (IECs), the FAC anchors the intestinal epithelial cell via integrins to extracellular matrix ligands(154). Integrins are heterodimers containing an alpha and a beta subunits. A wide variety of integrins are expressed on epithelia including a such as a1b1, a2b1, a3b1 and a6b4 which are collagen/laminin receptors. Fibronectin receptors on epithelia include a5b1, a8b1 and aV containing integrins (353). Signalling through the integrins in response to mechanical signals such sheer forces or compression allows for transduction of the signal to the actin cytoskeleton (354). This has direct implications on wound healing, as well as the invasive and metastatic nature of cancer cells, in which this function is best described. FACs have also been implicated in the

activation of the NLRP3 inflammasome (355). The inflammasome is another multiprotein complex composed of pattern recognition receptors (PRRs) including NOD-like receptors and RIG receptors, with the adaptor proteins, apoptosis-associated speck-like protein containing CARD (ASC) and pro-caspase-1. To form an inflammasome, ASC oligomerises, recruits and activates caspase 1. Caspase 1 cleaves pro-IL-1b, and pro-IL-18 to proinflammatory cytokines IL-1b and IL-18. The NLRP3 inflammasome contains the NLRP3 protein which is held in an inactive state by a heart shock protein (HSP90) and a ubiquitin ligase (SGT1). NLRP3 is activated by microbe associated molecular patterns(MAMPs), microbial toxins and damage-associated molecular patterns (DAMPs). On activation NLRP3 binds to and activates ASC. Both FAK and Pyk2 regulate this inflammasome activation; Pyk2 activates ASC directly by phosphorylation and in clinical trials a dual inhibitor of Pyk2/FAK significantly reduced monosodium urate-mediated peritonitis, a disease model used to study NLRP3 activation (355, 356).

Integrin-mediated signalling via b1 subunits activates the NLRP3 inflammasome in intestinal epithelial cells, with production of IL-18 (355) and in macrophages via integrin heterodimer (fibronectin receptor) a5b1 (357) with IL-1b secretion, and caspase 1 activation. In UC there is evidence for both protection (in an oxazolone-induced colitis model) (358) and involvement in chronic UC (in peripheral blood mononuclear cells) (359). There is evidence that adaptor members of the FAC, such as integrin-linked kinase(ILK), when mutated in mice, or knocked down in cell lines then there is a reduction in chemokine production, specifically MCP-1 and that ILK mutant mice are protected from DSS colitis (360).

In summary; LPXN abrogates integrin signalling in prostate cancer cells, BCR signalling in B cells and modulates migration in macrophages and cancer cells. There is no data on the normal role of LPXN in IECs, therefore it is difficult to predict the effect any changes in *LPXN* expression will have on IECs or functional colonic mucosa.

3.2 Hypothesis

The bioinformatics workflow in Chapter 1 indicates that cells containing UC susceptibility SNP rs10896794 have increased transcriptional regulation from IRF1/STAT1:STAT2 binding. As these transcription factors are predominantly activators we hypothesise that *LPXN* will be over expressed in cells containing this SNP.

Given the hypothesis that a significant part of the UC pathogenesis is disruption of the intestinal epithelial barrier, we hypothesise that IECs containing the SNP rs10896794 will have suppressed integrin-mediated functions of adhesion and spreading, thereby impaired epithelial wound healing, and decreased NLRP3 activation in response to MAMPS. In colonic mucosa, it is difficult to hypothesise the role of SNP rs10896794 given the multiple different cell types present, however if integrin signalling is supressed we hypothesise that the NLRP3 inflammasome will be affected with a reduction in IL-1B and IL-18 production in response to MAMPS.

3.3 Aims and Objectives

To identify the phenotypic effect of *LPXN* overexpression in epithelial cell lines and the *LPXN* SNP within colonic biopsies.

The objectives are:

1. To create an *LPXN* overexpressing epithelial cell line using immortalised cell lines HT29, Caco2 or T84 if possible, or other epithelial cells lines such as Hela if unsuccessful in IECs;
2. To assess the response to inflammatory stimuli in *LPXN* overexpressing cell lines on a) wound healing and b) cytokine production;
3. To investigate the effect of *LPXN* risk allele rs10896794 homozygosity in genotyped colonic biopsies on cytokine production, and in response to bacterial ligand stimulation.

3.4 Ethics and patient recruitment

The investigating genetics of UC (iGUC) study obtained ethical approval from both the University of East Anglia Faculty of Medicine and Health Sciences Ethics Committee and Human Tissue act subcommittee (ref 20152016-39HT) and the Norfolk and Norwich University Hospital Research and Development Committee (ref 02-01-16)

Patients with known UC undergoing colonoscopy for surveillance or disease assessment and those undergoing colonoscopy for polyp surveillance were identified from the colonoscopy lists by a member of the medical team. On the day of the procedure the patients were fully briefed and asked for written informed consent *via* the Norwich Biorepository Information sheet and consent form.

3.5 Methods

3.5.1 Cell Culture

Complete cell media for HT29 and Hela cells was Dulbeccos Modified Eagle's Media-DMEM (Lonza BE12-604F) 500ml, Foetal Bovine Serum, FBS (LabTech FCS-SA) 50ml and

L-glutamine (Lonza BE17-161E) 5ml.

HT29 (ATCC HTB38) and Hela (ATCC CCL2) adherent cell lines were maintained in either T25 or T75 flasks and passaged following trypsinisation and incubating at 37^0C for 5-10 minutes until the cells are rounded, detached single cells. Complete media was then added at the required volume for seeding or for cell counting (10mls total for T25, 30mls total for T75). If directly seeding to T25 or T75 then a 1:10 passage rate was used. Cell media was changed every 3-4 days. HT29 were used from passage 10- 25. Hela were used from passage 7-20. Cells are counted using a haemocytometer and standard protocols.

Hela (ATCC CCL2) adherent cell lines were a gift from Dr I Hautefort, Carding Laboratory, Institute of Food Research.

3.5.2 Cell transfection toxicity assessments

3.5.2.1 Plasmid and transfection reagent toxicity

The cells were grown to 60% confluency in a 24 well plate. 50ul Jet Prime buffer (Polyplus #114-01) was mixed with the required amount of trial plasmid DNA by vortexing. The required amount of jet prime (see figure 3.1) was added to appropriate DNA/buffer mix on a 1:2w/v basis or just buffer and incubated at room temperature for 10 mins. 50ul of transfection mix was added per well and the plate was rocked from side to side gently to get maximal coverage then incubated at 37^0C for 24 hours. It was trypsinised, and trypan blue stained for alive/dead counts at 24 hours. The toxicity assay extended from 0.25ug plasmid to 0.75ul plasmid for the 24 well plates.

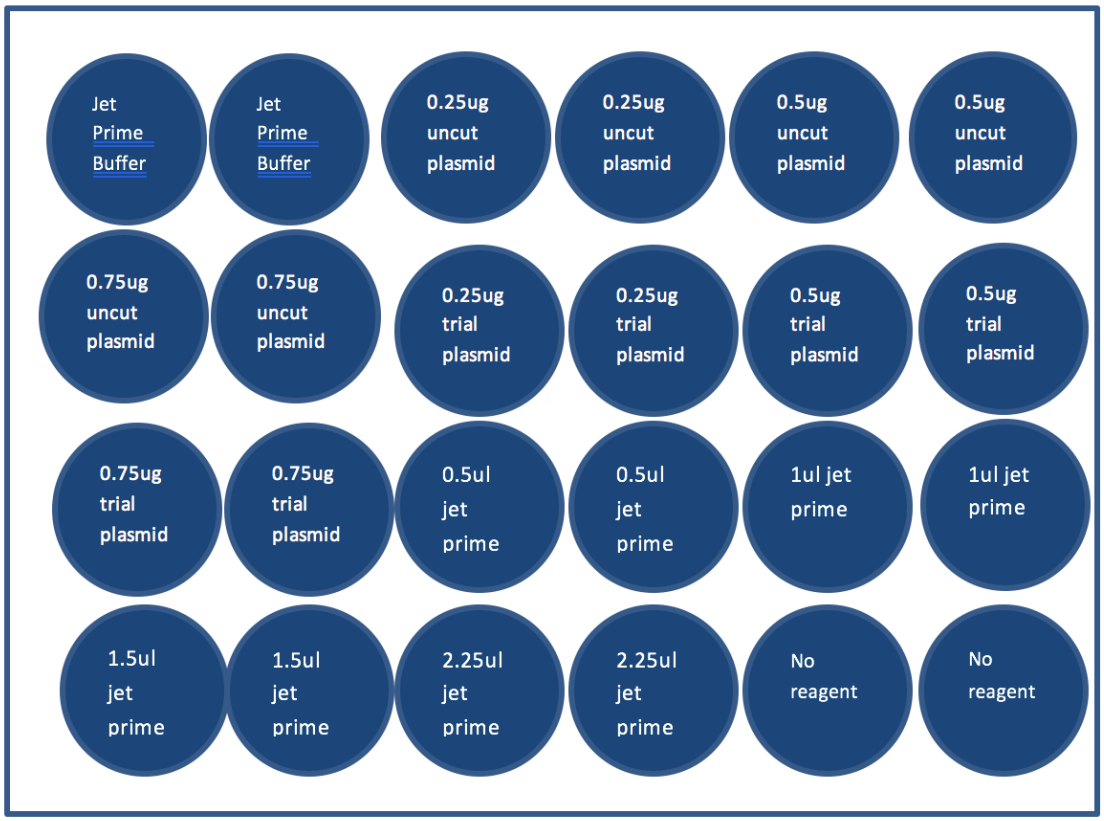


Figure 3-3 Plasmid and Jet Prime Reagent toxicity assay layout of 24 well plate.

3.5.1 Plasmids used

Table 3-1 identifies the plasmids assessed during the process of optimisation, their source and the amount of DNA required for transfection based on the toxicity assays in each cell type for each plasmid.

DNA	Source	DNA for transfection in 6 wells (Hela)	DNA for transfection in 6 wells (HT29)
***LPXN* Double Nickase Plasmid and Control**	Santa Cruz sc-405101-NIC Santa Cruz sc-437281 (control)	2ug	1ug
***LPXN* Activation Plasmid and Control**	Santa Cruz – Sc 405101-ACT Santa Cruz sc-437275 (control)	2ug	1ug
pX334-U6-DR-BB-DR-Cbh-NLS-hSpCas9n(D10A)-NLS-H1-shorttracr-PGK-puro	Addgene: Gift from the Zhang Lab	3ug	1.5ug
Floxed Neo resistence Homology Cassette	Integrated DNA Technology	2ug	

Table 3-1 Plasmids used to modify the expression of LPXN in Hela and HT29 cells. The source of the plasmid and amount of DNA required for the transfection reaction for HT29 and Hela cells is given.

3.5.1.1 Antibiotic Toxicity

Puromycin was used for positive selection of plasmid transfected cells. This causes premature chain termination during translation at the ribosome and was toxic to eukaryotic cells. Geneticin (G148) toxicity was also assessed as it was used to selected for cells which incorporated a floxed Neo Resistence gene by homologous recombination.

Hela and HT29 cells were plated onto 24 well plates and grown to 80 % confluence. Duplicate wells were then subjected to puromycin from 0.25ug/ml to 10ug/ml. Unlike normal kill curves where the optimal concentration was reached at which all cells are dead at 7 days, because the transfection protocol required selection pressure for 3 days maximum the concentration of puromycin used was that that caused total cell death at 3 days to prevent false positive transfection rates.

3.5.2 Transfection and Positive Selection

Cells are seeded at a density of 2 x 10^5 per 2ml into 6 well plates and grown to 60% confluency. 200ul of JetPrime (Polyplus, Cat no 114-15) buffer was mixed by vortexing with the optimised amount of plasmid DNA determined by plasmid toxicity assay. JetPrime transfection reagent was added to the appropriate DNA/buffer mix on a 1:2w/v basis and incubated for 10 minutes. 200ul was removed from the 6 well media and replaced by 200ul of JetPrime/DNA mix. The 6 well plates were rocked gently from side to side to ensure good coverage and incubated at 37°C, 5% CO_2.

For single cell colonies:

After 48 -72 hours (48 for Hela, 72 for HT29) the cells were trypsinised using 500ul of trypsin and incubated at 37°C for 10 minutes. The trypsin was deactivated by the addition of complete media. The detached cells were processed through a limiting dilution in a 96 well plate using 50% fresh culture medium and 50% conditioned cell media from a T25 pre-confluent respective cell line with puromycin (2ug/ml for HT29, 3ug/ml for Hela). The 96 well plate was incubated at 37°C, 5% CO_2. After three days, the media was changed to a 1:1 conditioned cell media/fresh media mix. This was changed every three days. During the second week, once the surviving cells form colonies covering >50% of the 96 well base, they were trypsinised and passaged into a 24 well, then 6 well, then T25 flask. The whole passaging process took 6-8 weeks. Once the cells are in a T25 they no longer need conditioned cell media, so fresh complete media was used instead.

A divergence from protocol was required for CRISPR activation plasmids, as these required a continuous selection pressure, puromycin 2 or 3ug/ml was added to all media used depending on the cell type.

For mixed colonies:After 48 -72 hours (48 for Hela, 72 for HT29) the cells were trypsinised as above using 500ul of trypsin and incubated at 37°C for 10 minutes. The trypsin was deactivated by the addition of complete media. The detached cells were seeded in a 1:2 dilution to 24 well plates with puromycin/conditioned/fresh media. After three days the media was changed to conditioned/fresh media. After two weeks, conditioned media did not need to be added.

3.5.3 Immunocytochemistry

The cells were seeded at a density of 5 x 10^4 into single wells of an IBIDI u-slide 12 well chamber with a removable gasket (Thistle Scientific IB-81201). Once confluent (usually within 36 hours depending on phenotype), 250ul apical media was removed and flash

frozen for cytokine analysis. The cells were gently washed in 200ul sterile 1 x Dulbecco's phosphate buffer solution (DPBS)(Sigma D837) three times and fixed in 4% paraformaldehyde (PFA)(made in house) for 15 minutes. The PFA was removed and discarded and the cells washed in 200ul DPBS three times. The DPBS wash was discarded and 200ul permeabilisation buffer (0.5% Triton x-100 in DPBS) applied for 15 minutes. The permeabilisation buffer was then discarded and the cells washed in 200ul DPBS three times, before blocking buffer was applied for 1 hour (0.2% Triton X-100, 3% Bovine Serum Albumin in DPBS). This was then discarded. The cells were incubated with primary antibodies or isotype at the optimised concentration (1:100) in blocking buffer at 37°C for an hour. The antibodies used were a mouse anti-LPXN monoclonal antibody (Sigma SAB4200301) with mouse IgG1 isotype control (BD Biosciences BUV661) or a mouse anti-LPXN F12 (IgG3 K) (Santa Cruz)sc-376903 with mouse IgG3 K isotype control (Crown Bio#C0009). After an hour, the solution was discarded and the cells washed gently in 200ul DBPS three times. The cells were incubated with optimised secondary antibody (1:1000) goat anti mouse IgG Texas Red (Abcam ab6787) in blocking buffer at 37°C for thirty minutes. The gasket was then removed and the slides washed in DPBS baths three times. Vectashield Hardset AntifadeMounting Medium with DAPI (Vector Laboratories, Cat No: H-1500) was then applied with the. The slides were cured 15 mins in the dark, then placed at 4°C.

3.5.4 Fibronectin coating of glass slides

12 well IBIDI glass slides with a growth surface area of 0.56cm^2 per well were coated with 5mg/cm^2 human fibronectin (Merck Millipore FC010-100mg) as per manufacturers instructions and stored at 4°C until used (life span 1 month).

3.5.5 Wound Healing Assay

Cells were seeded into either a fibronectin coated 12 well glass slide (F^+) or non fibronectin coated 12 well glass slide (F^-) grown to confluency. Once confluent the cells were scratched in a diagonal using a 20ul pipette tip. Half of the slides (F^+, F^-) were exposed for 24 hours or 12 hours to a bacterial ligand cocktail of 1ug/ml Lipopolysaccharide from *E.coli* (Sigma L2880), Muramyl dipeptide (Bachem G-1060.0005) , and *Micrococcus lutues* Peptidoglycan (Sigma 53243). After the allotted time period the slides were fixed in 4% PFA and stained as per the immunocytochemistry protocol. The samples were imaged on a Zeiss Imager M2 microscope. Wound distance at 5 points along the wound using 10x

magnification and phase contrast was determined and then calculated using imageJ (Fiji) using the closest distance edge to edge averaged across the 5 points.

3.5.6 Immunoblotting

3.5.6.1 Whole cell lysate

Cell culture bottles were placed on ice and the cells washed with ice-cold sterile PBS. 400ul Ice-cold NP-40 lysis buffer (Thermofisher FNN0021), 2ul PMSF (Sigma PMSF-RO) and protease/phosphatase inhibitor cocktail 4ul (Sigma 11836170001 ROCHE) was added to the cells and incubated on ice for 5 minutes before being scraped into a cooled microcentrifuge tube. The mixture was then sonicated for 3 x 30 seconds in ice cold water and centrifugation at 10,000 RCF for 10 mins at 4°C. The supernatant was carefully removed and the pellet discarded. The supernatant was either kept on ice if it was going to be used immediately or flash frozen in liquid nitrogen and stored at -80°C.

3.5.6.2 Biopsy protein lysate

Biopsies were thawed on ice in screwtop eppendorfs to which 122ul CelLytic MT (Sigma, C3228) and 3ul protease inhibitor cocktail (Sigma P2714-1BTL) were added with 5 acid washed glass beads (3mm diameter). Samples were bead beaten on MP FastPrep at 4m/s for 30 seconds, before centrifugation at 10,000RCF at 4°C for 2 mins. The lysate was transferred to 1.5ml pre-cooled eppendorfs for centrifugation at 10,000 rom for 10 mins at 4°C. The supernatant was carefully removed and stored at -80°C The pellet was discarded.

3.5.6.3 Protein quantification

Protein content from whole cell or biopsy lysate were quantified using a Pierce 660nm Colorimetric Assay, with fresh BSA as a reference, following the manufacturers protocol. In short, a standard curve of BSA was created by serial dilution, in lysis buffer. The test samples were assayed neat, and at 1:2 and 1:10 dilutions. All samples were run in duplicate adding 1ul sample to 15ul of Pierce Assay Reagent in flat bottomed 96 well plate for 5 mins before being read on a spectrophotometer (Bioscreen C plate reader) at 660nm. The result was averaged across the duplicate samples. In the event of significant discrepancy between duplicates, the samples were vortexed and re-analysed. The protein content was quantified against the standard curve.

3.5.7 Western Blotting

Samples normalised to protein content in 13ul volume were mixed with 5ul 4X NuPage buffer and 2ul 10X NuPage Reducing Agent in Eppendorf tubes and heated to 70°C for 10

mins. 20ul reduced sample was added to NuPage Novex Bis-Tris gels 4-12% 1.0mm 12 well gels with 5ul Page Rule Prestained Protein Ladder (Fermentas) also run on the gel for size determination. Gels were run in NuPage mini tank in NuPage MOPS SDS Running Buffer at 200V for 50 minutes.

Blotting pads were soaked in transfer buffer (NuPage Transfer Buffer plus 10% methanol) and PVDF membrane was soaked in 10ml methanol. The precast electrophoresed gels were levered open using a gel knife and the wells and foot were removed. Soaked blotting paper was placed on the gel and air bubbles removed. The gel was turned over and pre-soaked PVDF membrane was placed on top of the gel then blotting paper on top of the PVDF. The blotting paper/PVDF/Gel sandwich was placed in the transfer block (PVDF side uppermost) and clamped shut. This was then soaked in transfer buffer and electrophoresed for 1 hour at 30V. The PVDF membrane was washed in NATT buffer (0.24% Tris Base, 0.8% NaCl, 0.05% Tween20 in 1000ml MilliQ), then blocked in 5% BSA for 1 hour (shaking). The PVDF membrane was incubated with the primary antibody mouse anti-LPXN F12 (IgG3 K) (Santa Cruz sc-376903) 1:200 in antibody solution (1% BSA in NATT buffer), and anti-mouse GAPD loading control primary antibody for 16 hours at 4°C. The membrane was then washed extensively (in NATT buffer) then probed with donkey anti mouse HRP conjugate 1:5000 dilution for 30 mins at 37°C. The PVDF membrane was then washed copiously again before incubation with 1:1 dilution of enhancer and buffer PicoRabbit IgG detection Kit (Supersignal West) for 5 mins. After this the PVDF was imaged using a Protein Imagine machine (Protein Simple).

3.5.7.1 Creation of the LPXN sgRNA plasmids

CHOPCHOP(https://chopchop.rc.fas.harvard.edu/), an online single guide RNA(sgRNAs) design tool, was used to design and assess the LPXN sgRNAs using FASTA sequences corresponding to the rsID from dbSNP (http://www.ncbi.nlm.nih.gov/SNP/).

The potential sgRNAs were analysed and stratified depending on their proximity to the SNP site, their predicted off target effects and primer off target binding.

The two target sequences were based on their proximity to the SNP, efficiency, and minimising off target effects which was essential for the insertion of the homologous recombination template. The two sgRNA sequences were (forward) gatgttgcaatttaaatagttgg and (reverse) tggtcctgaacttagacccgataga. The underlined sequences did not form part of the sgRNA insert as these are the protospacer adjacent motifs (PAMs).

Each target sequence with complimentary reverse strand was designed with *BbSI1* complimentary overhangs and manufactured by IDT. The DNA oligos were duplexed in DNAse free water with 50nM NaCl, by heating to 95°C for five minutes in a thermocycler, then allowing to cool on the benchtop. Duplexes were confirmed by agarose gel electrophoresis. The duplex was ligated into the Cas9 plasmid backbone pX334-U6-DR-BB-DR-Cbh-NLS-hSpCas9n(D10a)-NLS-H1-shottracr-PGK-puro following the protocol in the New England Biolab Quick Ligation Kit (Cat M2200S). The ligated plasmid was transduced into One Shot™ Top10 Chemically competent *E.coli* (Thermofisher C404010) using a heatshock protocol and plated onto 100ug/ml ampicillin-LB plates and grown overnight at 37°C. Colonies were picked and underwent colony PCR using the sgRNA oligo (forward) as an internal primer and a Cas9 (rev) primer as the plasmid primer. The colonies were also grown in a 1ml deep well block containing terrific broth/phosphate buffer/LB /Ampicillin mix. Cultures grown from corresponding positive colony PCRs had the plasmid extracted using Qiagen Mini-Prep Kit as per the manufacturer's instructions and sequenced by Eurofins using the Mix2Seq kit. Once positive identification of the sgRNA sequence in the correct orientation in the plasmid by sequencing had been confirmed in SnapGene, the culture was grown for 12 hours in 100ml LB media supplemented with ampicillin. The plasmid was extracted as per the manufacturers protocol with QiaPrep Spin Mini Prep Kit (Qiagen 27104) The extracted plasmid was then analysed using NanoDrop to ascertain the quality and quantity of the DNA.

3.5.8 DNA extraction

DNA was extracted from cell cultures and colonic biopsies using a Sigma GenElute Mini Prep Kit (G1N10) according to manufacturer's instructions. In short, cells in culture were trypsinised, resuspended, RNase treated and incubated with proteinase K before the lysis solution was added. The samples were then vortexed and heated to 70 °C for 10 minutes. The DNA was extracted from the lysate via ethanol extraction and the use of spin columns. The extracted DNA was washed and eluted into 200ul of Tris-EDTA. The DNA quantity was confirmed using NanoDrop. DNA samples were stored at -20°C. The protocol for DNA extraction from colonic biopsies varied from cell culture in the lysis/digestion step. Biopsies were individually resuspended in 180ul Lysis T solution and 20ul proteinase K was added. The biopsies were then incubated at 55°C in a water bath for 4 hours until the tissue was completely lysed. The mixture then underwent RNase treatment and protocol followed as per cultured cells.

3.5.9 RNA extraction

Total RNA extraction from the cell lines was undertaken using the ISOLATE II RNA mini kit (Bioline BIO-52072), according to the manufacturer instructions.

3.5.10 SNP sequencing

Human genomic DNA in TRIS-EDTA solution underwent custom sequencing at Eurofins Medigenomix GmbH to identify the alleles at SNP sites rs1598859, rs 2227551, rs37774937, rs10896794 and rs12254167.

3.5.11 PCR

The primers to detect the *LPXN* SNP site (Chr 11 Position 58571651) in both parental cells (Hela, HT29, Cac02, T84) and in transfected cells were:

Forward: CCTGTCTTTTAGGGTGTGGAGA (pos 58571546) TM 59°C

Reverse: GCCCAGATTCAAGTCCTGGT (pos58571901) TM 59°C

HotStarTaq Master mix kit (Qiagen 203443) was used according to manufacturer's instructions (changing annealing temp to 54°C) to produce a 356bp amplicon visualised on 1.5% agarose gels. The PCR amplicon was purified using QIAquick PCR purification kit (Qiagen 28104) following the manufacturer's instructions before sequencing at Eurofins using the forward primer with the Mix2Seq service.

3.5.12 qPCR for *LPXN* gene expression

After RNA extraction, each sample was tested for quality and quantity using a Nanodrop. cDNA reverse transcription was undertaken using a HighCapacity cDNA Reverse Transcription Kit (Applied Biosystems 4374966), as per the manufacturers instruction, with the RNAse inhibitor. For each reaction 0.5ug of RNA was used. After cDNA reverse transcription, the samples were diluted 1:5 in PCR grade water. Each of the samples were qPCR in triplicate for *LPXN* expression and beta actin expression, used as the internal control. 'No cDNA' controls were also used to assess for pipetting error and contamination of reagents. TaqMan Gene Expression Master Mix (ThermoFisher Scientific 439016) was used. The reactions were prepared in a 384-well plate with a reaction volume of 10ul (2.5ul template, 7.5ul mastermix). The qPCR was run on a Roche LC48011 Light Cycler.

3.5.13 Polarised *in vitro organ* culture of colonic biopsies

6 colonic biopsies taken from the rectosigmoid junction (approx. 18cm from rectum) via 5mm pinch biopsy forceps were taken from consented patients. Only those that were macroscopically normal at the time of endoscopy were used for the 'normal' colon

patients. 1 sample per patient was flash frozen on dry ice immediately (for cytokine analysis or DNA extraction), 5 x samples were placed in IVOC media for the 10min transport to the laboratory. As per the protocol of Schuller et al 2009 (361) IVOC media consisted of (per 100ml): NCTC-135 (Sigma-aldritch) 0.47g, sodium bicarbonate (Sigma-aldritch) 0.11g, Dulbeccos modified Eagle's Medium (Sigma-Aldritch) 45ml, and newborn calf serum (NCS)(Sigma-Aldritch) 10ml, distilled water 45ml. Each media aliquot was filtered (0.22um) and stored at 4°C, but warmed to 37°C prior to sample collection.

Each sample was orientated with the mucosal side uppermost and halved across the shortest access and prepared for polarised IVOC (pIVOC) (Figure 3-4) and mounted on a 6 well plate containing 3ml of IVOC media in each well base.

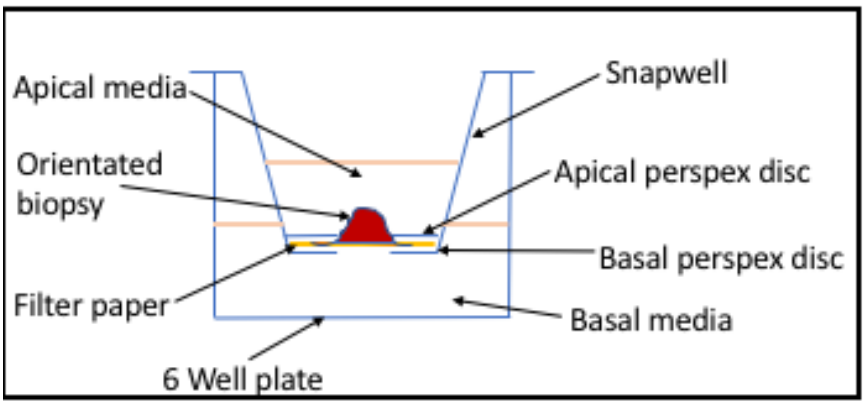


Figure 3-4 Polarised in vitro organ culture (pIVOC). The biopsy orientated with the mucosal side uppermost and sealed in between two 'O' shaped perspex discs, held in place using a snapwell. The apical and basal media are kept separate by the biopsy.

Once mounted, 200ul of media was added apically and for those undergoing bacterial stimulation, 1ul/ml or 10ul/ml lipopolysaccharide, peptidoglycan and muryldipeptide were added and the plate was incubated on a rotor (12 RPM) at 37°C for 2 hours. The apical media was then removed and 2 drops of fresh media placed onto the apical surface of the biopsy to prevent the sample drying out. The plate was then re-incubated for a further 6 hours after which the biopsy was removed intact from each snapwell, flash-frozen in liquid nitrogen and stored at -80°C for later RNA or protein extraction. The basal media was flash-frozen and stored at -80°C for cytokine analysis.

3.5.14 Cytokine Analysis

The LEGENDplex Human Inflammation Panel (BioLegend 740118) was used to simultaneously quantify 13 human inflammatory cytokines/chemokines (IL-1B, IFN-a, IFN-y, TNF-a, MCP-1, IL-6, IL-8, IL-10, IL-12p70, IL-17A, IL-18, IL-23, and IL-33) according to the manufacturers protocol, and run on the BD LSRFortessa using the PE and APC channels.

3.6 Results

3.6.1 Characterisation of LPXN in epithelial cell lines

Using western blot the presence of Leupaxin protein was confirmed in HT29, Caco2 and Hela cell lines (Figure 3-5) compared against a positive control, Ramos whole cell lysate (Santa-cruz sc-2216)

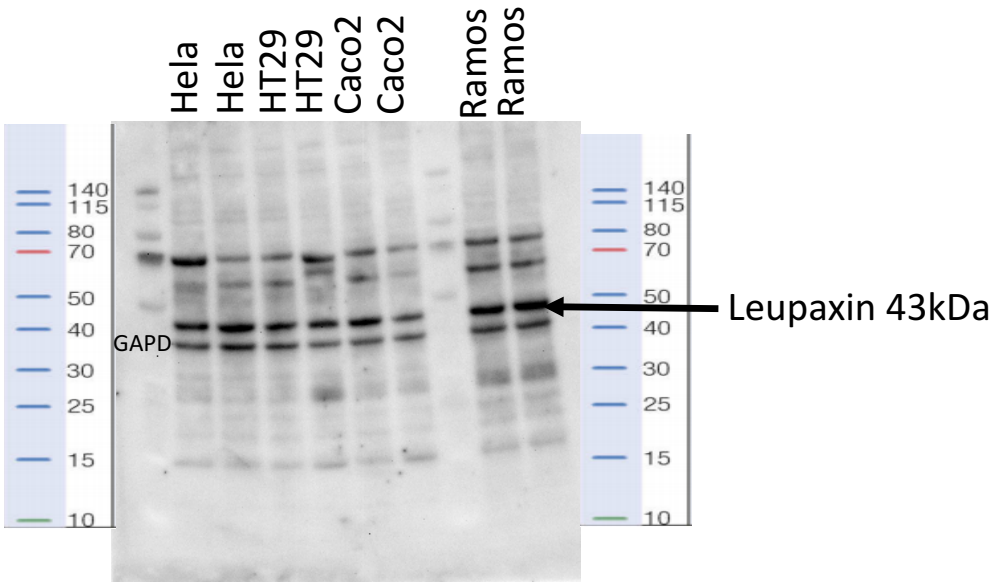


Figure 3-5 Western blot images of HT29, Hela and Caco2 whole cell lysate probed for LPXN expression. Blots were incubated with primary Mouse anti-LPXN F12 (IgG3 K) sc-376903 1:200 and anti-mouse GAPD (as loading control and donkey anti-mouse-HRP conjugated secondary antibody. The Ramos cell line whole cell lysate was used as a positive control LPXN is a 43kDa protein, GAPD is 37kDA.

HT29, T84, Caco2 and Hela cell lines were sequenced across the *LPXN* SNP allele site to ensure they did not already contain the risk allele using the PCR primers as described above, cleaning the PCR product and sequencing using the forward PCR primer with a Mix2Seq kit (Eurofins). Caco2 was not used after this stage as the amplicons from the primers which should have been specific to the *LPXN* site reproducibly produced multiple different products when sequenced, indicating genetic instability and potentially polyploidy of chromosome 11.

We experienced poor transfection rates in HT29s (<0.01 %), using multiple transfection and selection techniques, therefore genetic manipulation using CRISPR plasmids was therefore undertaken in Hela cells as a surrogate epithelial cell line, notwithstanding the obvious limitations to using a non-intestinal cell.

3.6.1.1 Creation of LPXN overexpression and knock out cell lines

Transfection rates of each of the plasmids into Hela cells were assessed by cell comparing alive/dead cell counts from parental HeLa cells at day 3 of growth in 6 well plates vs cells growing under positive antibiotic selection at day 3 (Table 3.2).

Plasmid	**Average Transfection Rate (n= 6 transfections)**
D10A-puro with forward sgRNA D10A-puro with reverse sgRNA	62% cells alive under puromycin selection
LPXN Activation Plasmids	60% cells alive under puromycin selection
LPXN Activation plasmids control	59% cells alive under puromycin selection
Double nickase *LPXN* knock out plasmids	38% cells alive under puromycin selection
Double nickase *LPXN* knock out control plasmids	46% cells alive under puromycin selection
Homologous recombination template gBlock with Neo resistence (1:3 w/v ratio)	0.1% cells alive under dual puromycin and G148 selection

Table 3-2 Plasmid transfection rates in Hela cells using JetPrime transfection reagent. Cells were trypsinised at day 3 of puromycin selection, stained with trypan blue and counted using a haemocytometer. Cells were considered to have been successfully transfected if they were still alive under puromycin selection at 3 days.

Both the knock out plasmids and knock out plasmid controls had lower transfection rates than the D10A and activation plasmids, despite this 5 hela cell lines were created. 1. Hela cells with a customised *LPXN* knock out using a cloned double nickase system with the CHOPCHOP designed sgRNAs ligated onto the cassette (D10A-puro) 2. *LPXN* over expressing Hela cell line (puromycin selection) 3. Control hela cells with the activation plasmid control transfected (puromycin selection) 4. *LPXN* knock out via a commercial double nickase system (puromycin selection) and 5. *LPXN* knock out control (using sham sgRNAs on the knock out plasmids with the same selection cassette).

3.6.1.1.1 Creation of the *LPXN* SNP risk allele cell line

The customised 3119bp homologous recombination (HR) template with the SNP risk allele, a floxed EF1A promoter sequence- Neo resistence selection cassette with 430bp homology arms either side was designed using SnapGene and manufactured as a gBlock by Integrated DNA technologies (IDT). The gBlock was transfected with the D10A-puro plasmids and optimisation steps were undertaken. Optimisation of the plasmid: gblock ratio from 1:1w/v to 1:3 w/v, increasing the number of cells transfected from 6 wells to 10cm plates, using brefeldin to force the cells to undergo a higher rate of homologous recombination (209), using of conditioned media to keep single cells alive, and reducing

the dose of antibiotic for selection did not work. After 72 hrs of dual antibiotic selection the cells would not survive due to low transfection rates and presumed low HR rates, therefore it was not possible to progress further to undertake the final steps of confirmatory sequencing and floxing the EF1A promoter- Neo resistance cassette out of the genome.

3.6.1.2 Characterisation of the modified LPXN expression cell lines

The comparative growth between the cell lines was documented in fibronectin coated 6 wells at 4 days and 7 days post puromycin selection. On visual inspection (with representative areas shown in Figure 3-6), the *LPXN* (Santa Cruz) knock out cell lines produced long elongated spindle shaped cells by day 7 with an average confluency of 18% (n= 6, analysed with image J) which was significantly lower than the parental hela cells ($p=0.0077$). The double nickase (D10A) knockout (SNP specific) produced circular colonies of cells by day 7, with an average confluency of 19.5% (n=6) which was also significantly lower than the parental HeLa cells ($p=0.0039$). The circular colonies had large areas of no growth between each colony hub suggesting that the colonies had arisen from individual surviving cells, but had not yet grown to confluency with other hubs. This pattern was unlike the other cell lines which had an even spread of cells growing across the plate. The HeLa cell line transfected with the sham control knock out plasmid (n=6) showed normal HeLa morphology, but slightly delayed growth (80% confluency by day 7, NS) compared to the parental HeLa cell line. This is expected with puromycin selection as the number of transfected cells would be less than the total number seeded, and would therefore take longer to grow. The control cell lines were confluent by 10 days (data not shown), whilst the commercial knock out cell line never achieved confluency (cell death at <50% confluency). The D10A cell line reached confluency by day 21 via convergence of multiple single colony hubs.

The *LPXN* over expression cell lines grew rapidly and formed clumps as opposed to the flat layer of the normal HeLa cell (Figure 3.5, B). The *LPXN* overexpression cell line was at 100% confluency in <4 days under puromycin selection, significantly faster than the parental HeLa cells under no puromycin selection ($p=0.0058$). The control plasmids for the overexpression containing the sham sgRNA responded in the same way as the knock out sham plasmids; with 80% confluency at 7 days (NS).

qPCR to confirm *LPXN* gene expression of the 6 cell lines was undertaken in duplicate on two separate occasions with freshly transfected and cultured cell lines at day 10 (Figure

3.6,Figure 3-9). This confirmed overexpressing cell lines had (on average) 78 fold higher levels of *LPXN* expression than the parental HeLa cells. It was also confirmed the commercial *LPXN* knock out cell line had 1.98-fold less expression of *LPXN* than the parental cell lines (with Ct values of 34-35). The assay indicated that the D10A system was only a 'partial' knock out with 0.9-fold less expression of *LPXN* than the parental cell line (1.08-fold more expression than the commercial knockout). One explanation for this is D10A knock out required 2 large plasmids to be transfected simultaneously. This reduces the probability of successful transfection and increases the probability of a mixed genetic population. If only one plasmid was transfected and worked at maximal efficacy it would cause a single stranded nick which could have deleterious effects, but is more likely to be accurately repaired by the cellular DNA repair mechanisms, but the puromycin resistance containing plasmid would still be retained, allowing the cell to survive. A further explanation is that the sgRNAs cloned into the D10A plasmids were less efficient than the sgRNAs in the commercial knock out plasmids. The sgRNAs used for the D10A plasmids had the highest efficacy as designed by CHOPCHOP, but we were limited by the number of sgRNAs available in the SNP vicinity.

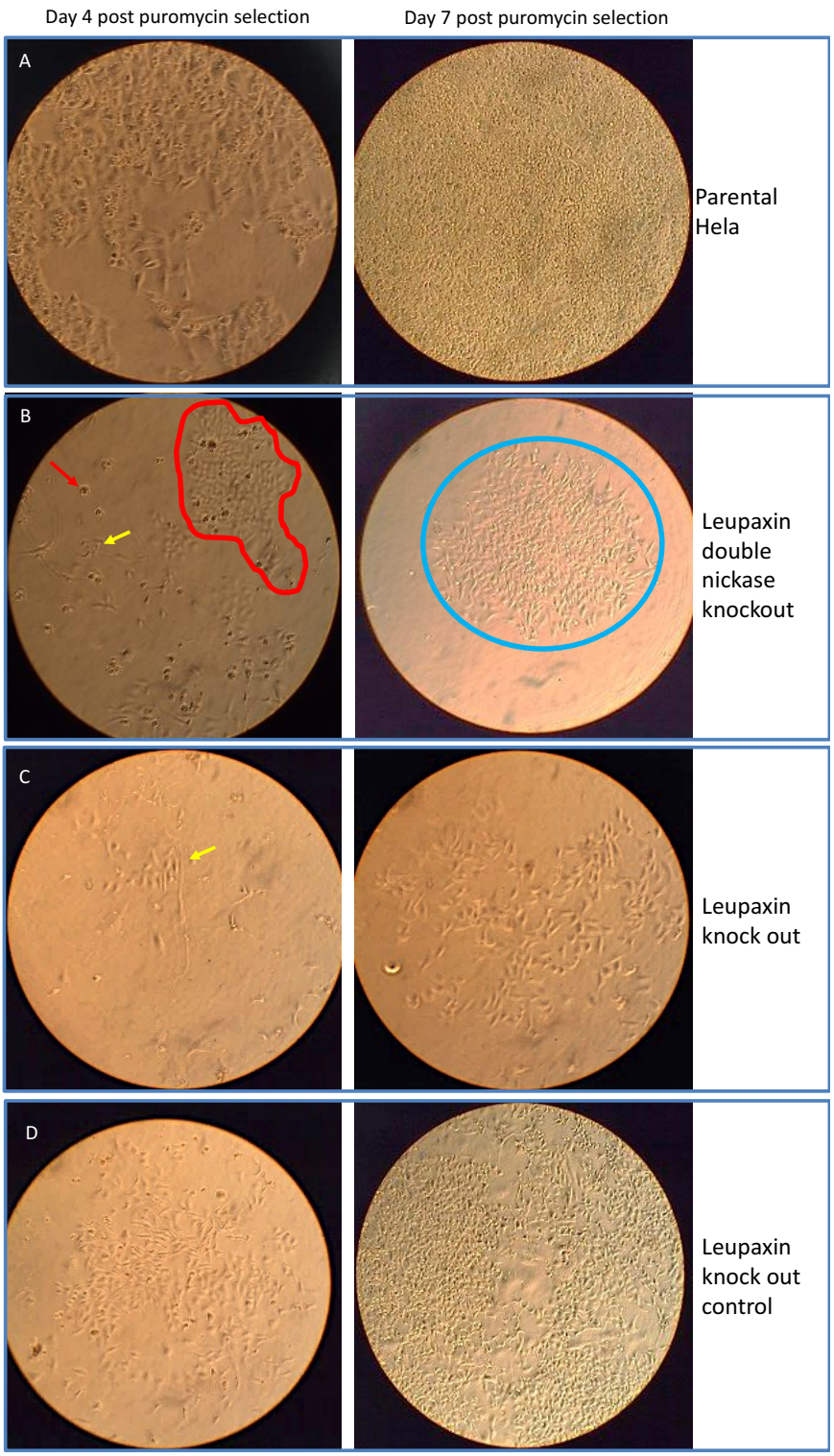


Figure 3-6 Characterisation of growth of Hela cells and Leupaxin knock out Hela cell lines over 7 days. ***A****: Parental Hela cells images at x 40(day 4) and x 10 (day 7) magnification showing normal growth and 100% confluency at day 7.* ***B****: Leupaxin double nickase knock out hela cell lines post puromycin selection both at x 40 magnifications showing at day 4 a mix of scattered individual cells (yellow arrow), dead cells (red arrow) and expanding colony growth (outlined in red); at day 7 single colony growth expansion only was present (blue circle) with <20% confluency.* ***C*** *Leupaxin (Santa Cruz) knock out cell lines post puromycin selection both at x 40 magnification showing individual cells with elongated morphology at day 4 (yellow arrow) and cellular expansion at day 7, <20% confluency.* ***D*** *Leupaxin (Santa Cruz) knock out control with sham plasmids post puromycin selection images at x 40 (day 4) and x 10 (day 7) showing delayed growth compared to the parental cells line post puromycin selection; 80% confluency at day 7.*

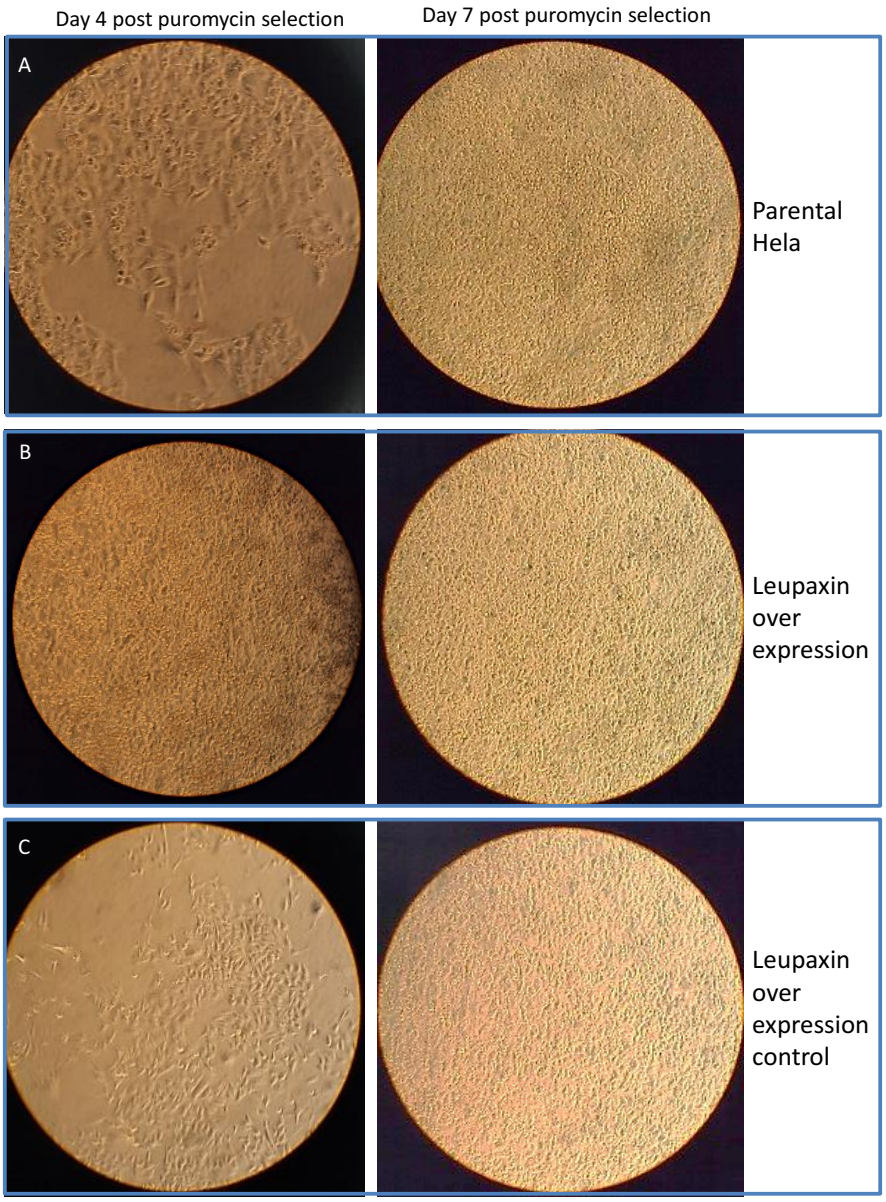


Figure 3-7 Characterisation of growth of Hela cells and Leupaxin overexpression Hela cell lines over 7 days. ***A****: Parental Hela cells images at x 40(day 4) and x 10 (day 7) magnification showing normal growth and 100% confluency at day 7.* ***B****: Leupaxin (Santa Cruz) overexpression hela cell lines post puromycin selection both at x 10 magnifications, showing at day 4 100% confluency.* ***C*** *Leupaxin (Santa Cruz) overexpression control with sham plasmids post puromycin selection images at x 40 (day 4) and x 10 (day 7) showing the same growth pattern compared to the parental cells line post puromycin selection with 100% confluency at day 7*

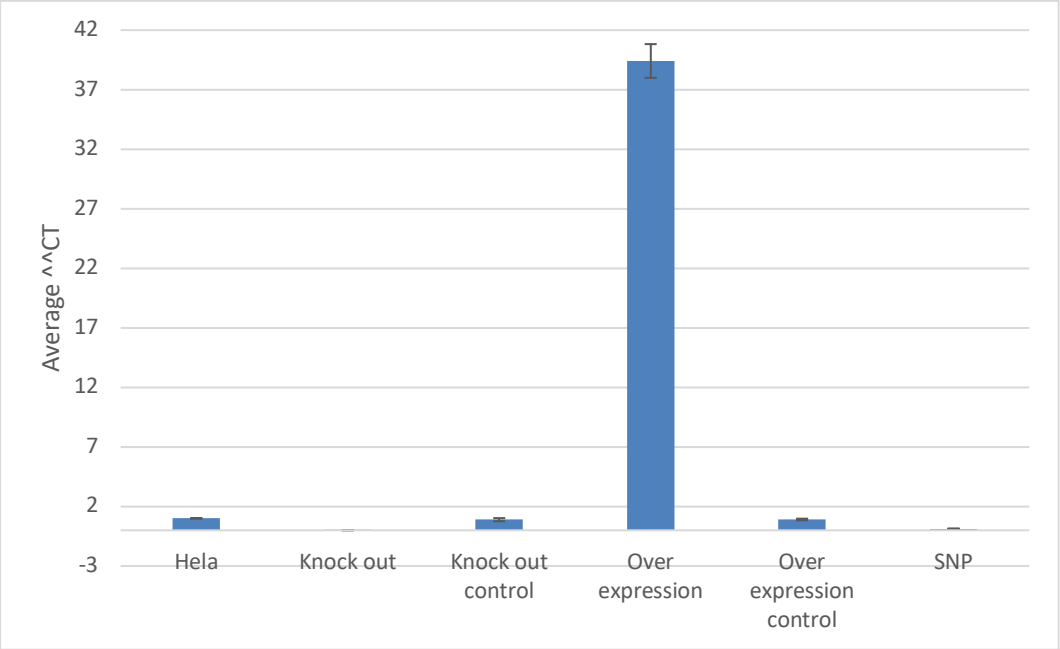


Figure 3-8 Fold changes (^^CT) of leupaxin expression in Hela cells. Each of the cell lines are controlled with beta actin with two biological replicates and two technical replicates for each biological replicate with standard deviations. Hela = Parental Hela cells, Knock out =commercial double nickase LPXN knock out, knock out control = double nickase sham plasmids, Overexpression = HeLa cell line overexpressing LPXN via LPXN activation plasmids, Over expression control =activation sham plasmids, SNP= D10A double nickase cell line.

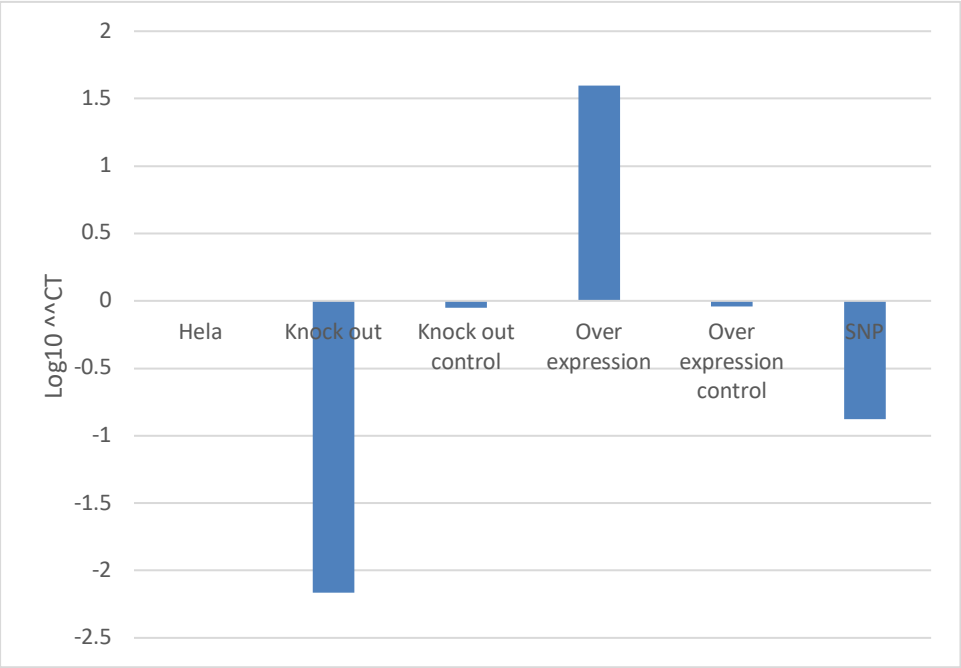


Figure 3-9 Log10 fold changes of leupaxin expression in Hela cells indicating down and up regulation of leupaxin expression. Results confirming knock out and overexpression of LPXN compared to the parental hela cells. Hela = Parental Hela cells, Knock out =commercial double nickase LPXN knock out, knock out control = double nickase sham plasmids, Overexpression = HeLa cell line overexpressing LPXN via LPXN activation plasmids, Over expression control =activation sham plasmids, SNP= D10A double nickase cell line.

The cellular location of LPXN in the epithelial cell lines, and confirmation of mRNA translation in the cells over expressing LPXN was visualised by immunocytochemistry. (Figure 3-10). The two antibodies used show slightly different localisations but the isotype control antibody produced no staining. At 24 hours, the LPXN staining in the overexpression cell line was visualised as granular cytoplasmic staining, with significantly less visible staining in the Parental Hela cell line. A composite image was required to ascertain the localisation. At 12 hours, the overexpressing cell line displayed clearly identified intense fluorescent signal at the plasma membrane (white arrows), indicative of focal adhesions, in addition to patchy staining across the cytoplasm (yellow arrow) and potentially in the nucleus. Focal adhesion sites were also seen in the parental Hela cells at 12 hours (white arrows) but there was no evidence of nuclear localisation in these cells.

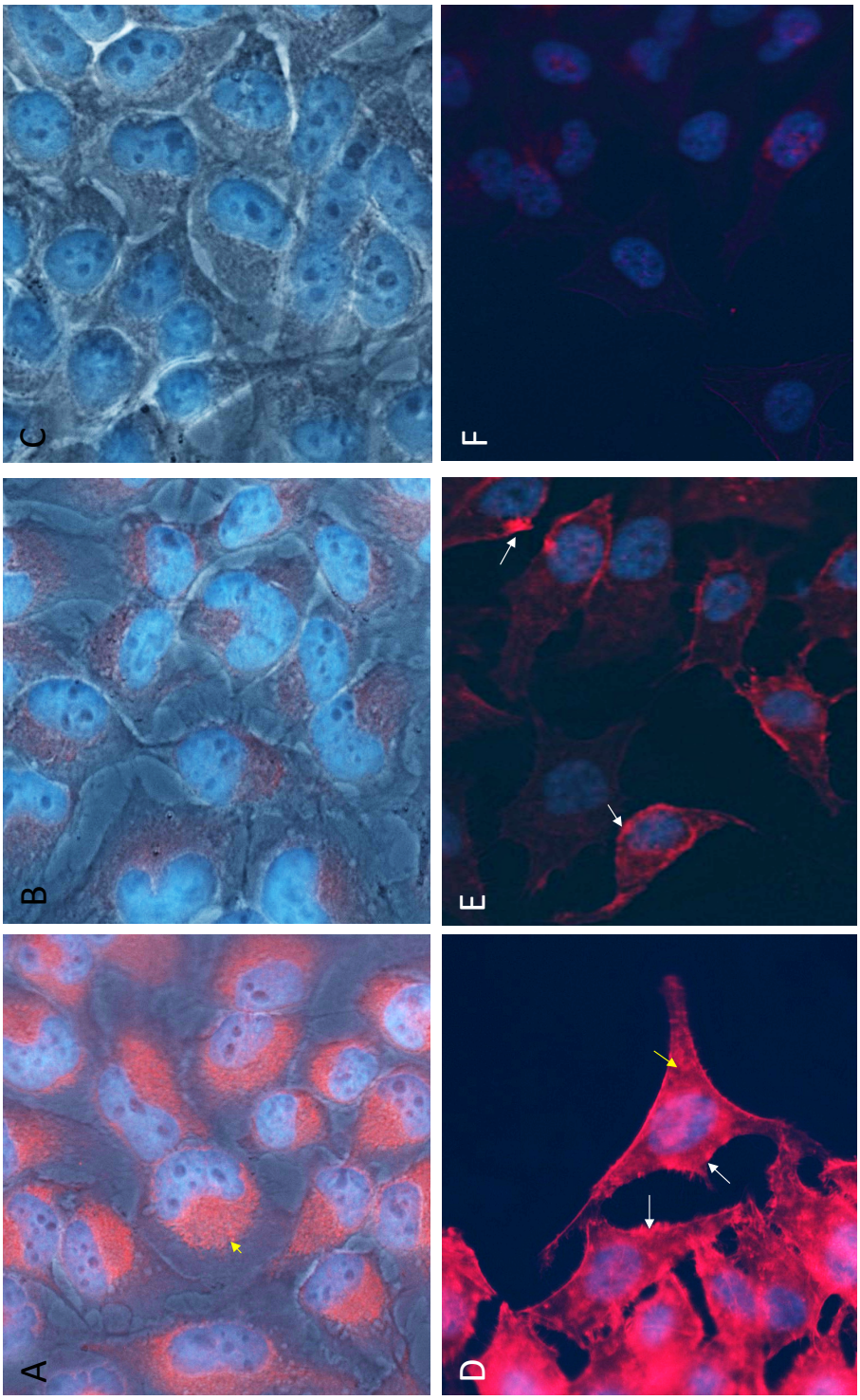


Figure 3-10 Immunocytochemistry staining for LPXN (texas red). A-B: Composite image of transmitted light image and fluorescence imaging x 20 at 24 hours Sigma Mouse anti LPXN Monoclonal antibody (1:100) and goat anti-mouse-Texas red (1:1000), and DAPI (blue). A. LPXN over expression hela cell line B Hela cell line C: Composite image of transmitted light image and fluorescence imaging x 20 Mouse IgG1 isotype control (1:100) with goat anti mouse texas red (1:1000, and DAPI (blue). D-F immunofluorescence imaging at 12 hours x 20. D-E Santa Cruz Mouse LPXN antibody (F12) (1:100) and goat antiimouse texas-red (1:1000), and DAPI (blue). F: Mouse IgG3 isotype control(1:100) withgoat anti-mouse texas red secondary antibody (1:1000) and DDAPI (blue). All images visualised with Zeiss Fluorescence microscope.

3.6.2 LPXN over expression and wound healing

The wound healing assay was undertaken on cell lines grown on fibronectin of which only the parental HeLa and *LPXN* over expressing cell lines grew to confluency on the glass coverslips and therefore could be used for the wound healing assay. The wound healing assay was done in the presence and absence of bacterial ligands (1ug/ml LPS, peptidoglycan and MDP).

The experiment was undertaken twice, with freshly transfected cell lines grown under puromycin control for 72 hours and then rested for 24 hours prior to causing the wound. The distance between the opposing epithelial edges was measured at the same 5 points for each replicate using a template in ImageJ (Figure 3-11)

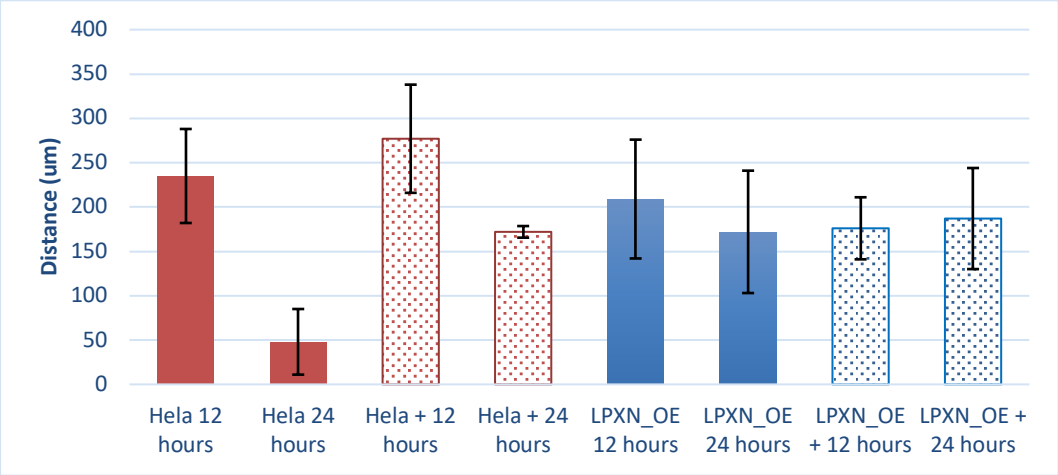


Figure 3-11 Wound healing assay. The wound defect was measured at 12 hours and 24 hours in parental HeLa cells (red solid columns) and LPXN over expressing(LPXN_OE) HeLa cells (blue solid columns) in the absence or presence (+/dotted columns) of bacterial ligands. There were two biological replicates for each cell line, with two technical replicates for each time point. Cells were fixed and stained and images viewed with an ImageJ template measuring the same 5 points along the wound defect for each replicate.

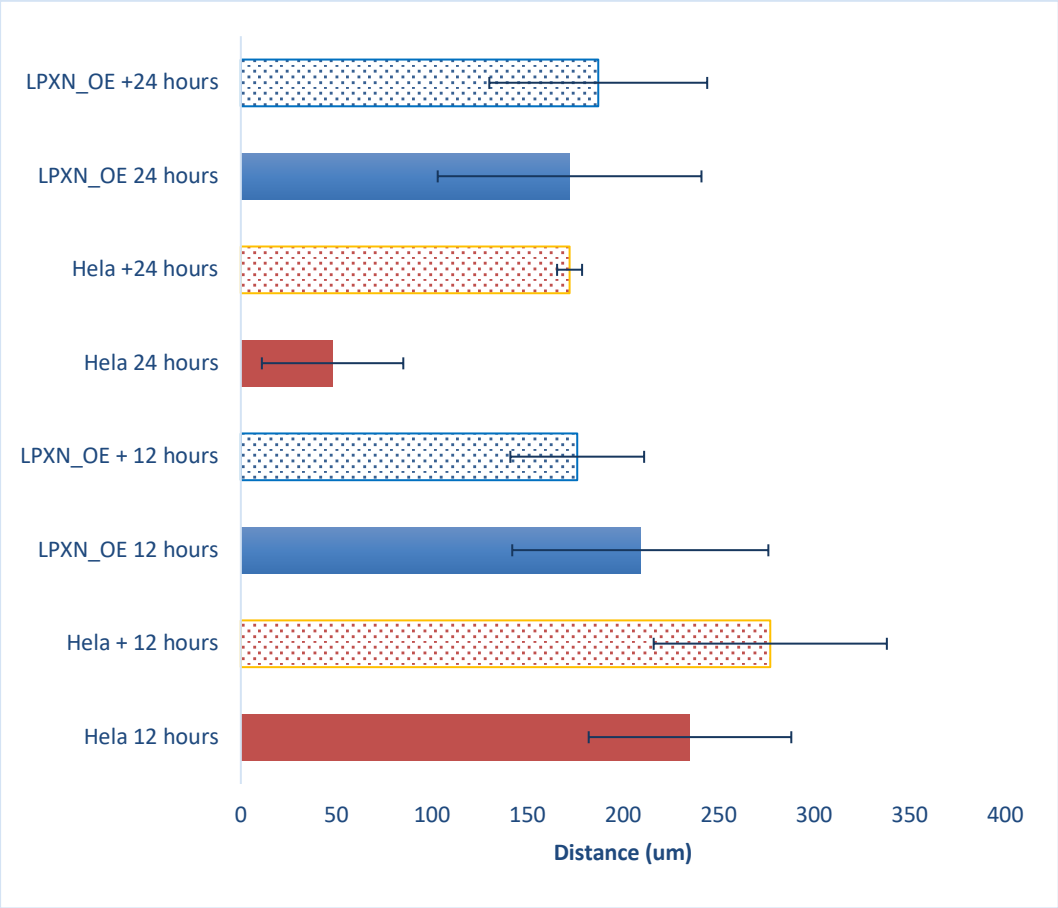


*Figure 3-12 Comparison of wound defect at 12 hours and 24 hours time points between parental hela cells (red solid columns) and LPXN over expressing(LPXN_OE) hela cells (blue solid columns) in the absence or presence (+/dotted columns) of bacterial ligands. Statistical significance assessed with a two tailed T test. * p=0.0198, **p=0.0001*

The *LPXN* over expression cells lines were able to heal significantly faster in the first 12 hours than the HeLa controls (p=0.0198) even in the presence of bacterial ligand stress (p=0.0001), however they were unable to maintain that rate of healing and a significant (p=0.0001) defect remained compared to the parental HeLa cells. The parental HeLa cells stimulated with bacterial ligands also had a significant defect remaining compared to the unstimulated HeLa cells (p=0.0001). The *LPXN* overexpressing cell line was difficult to assess as there were single cells within the defect tract which, in part, accounts for the wide SD bars. In summary, the wound healing assay indicates that although *LPXN* overexpression initially leads to a faster healing response wound healing over time is significantly impaired.

3.6.3 *LPXN* expression alters the cytokine profile excreted from epithelial cells

To quantify cytokine responses to changes in LPXN expression the BioLegend LEGENDplex Inflammation Panel was used to quantify cytokines secreted by Hela cells, either containing the *LPXN* overexpression plasmids, the *LPXN* knock out plasmids or parental Hela cells grown to confluency on fibronectin coated glass coverslips and on plain glass coverslips at 12 hours and 24 hours. Fibronectin was used to assess the requirement of the integrin signal to LPXN function, using a bacterial ligand cocktail of LPS to activate TLR2 and TLR4, peptidoglycan to activate TLR2, and MDP to activate NOD2. Hela cells did not grow on plain glass coverslips and the knock out cell lines grew poorly regardless of the growth environment (as described above) therefore these groups could not be analysed.

3.6.3.1 MCP-1

The pattern of MCP1 secretion by parental HeLa cells was a significant, but expected, rise in MCP1 secretion from 12 hours to 24 hours (p=0.0005). They show increased MCP-1 secretion in response to bacterial antigens at 12 hrs post stimulation (p=0.0151), but not at 24 hours (Figure 3.12).and with a trend of more MCP1 secretion with bacterial ligand stimulation (p=0.0151 at 12 hours, no significant difference at 24 hours).

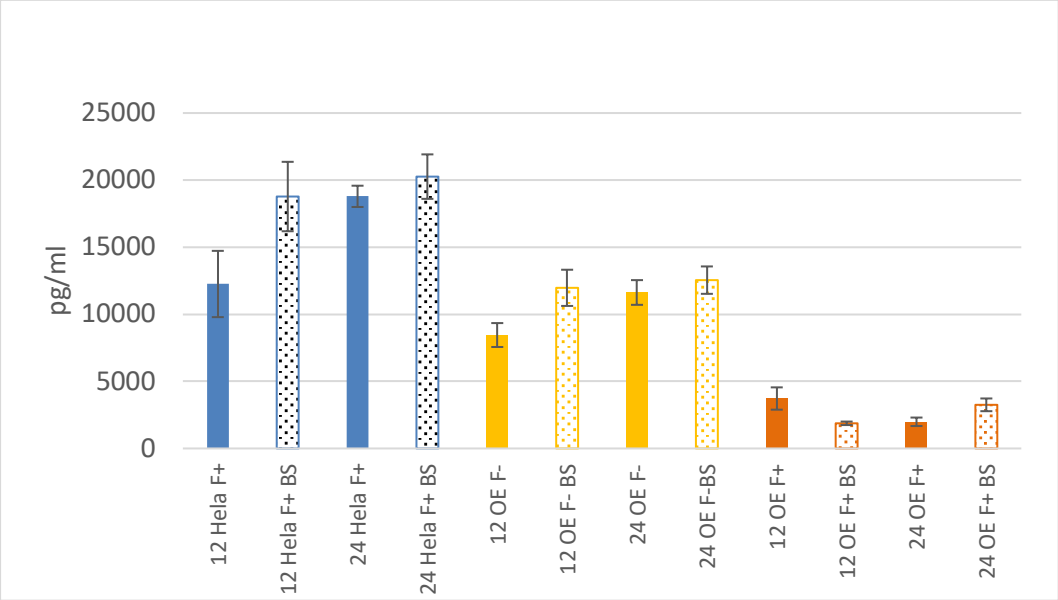


*Figure 3-13 Column graph showing secreted MCP1 in pg/ml of conditioned cell media averaged from quadruple replicates, each with duplicate technical replicates. F+ = fibronectin coated glass slide, OE = LPXN overexpressing Hela cell BS = bacterial antigen stimulation. Solid bars are non stimulated cells, dotted bars are stimulated with bacterial antigens. * p=<0.05 compared to 12 hour parental non stimulated hela sample, ** p= <0.001 compared to 12 hour sample,** p=<0.001 compared to the equivalent parental hela sample, **p=<0.001 compared to the equivalent fibronectin negative LPXN overexpressing sample.*

Levels of MCP1 in the media was significantly lower from the LPXN over expressing cells grown without fibronectin at both 12 hours (p=0.0034) and 24 hours (p=0.0006) compared to the media from parental HeLa cell grown on fibronectin, the same was seen with media from bacterial antigen stimulated LPXN overexpressing cells grown without fibronectin compared to the media from parental HeLa cells; p = 0.0021, p=0.0006 respectively.

With fibronectin coating, MCP1 secretion into the media by *LPXN* overexpressing cells was significantly lower than both the parental HeLa cell lines and the *LPXN* overexpressing cells grown without fibronectin. At 12 hours, both the unstimulated and bacterial antigen stimulated *LPXN* overexpressing cells secreted <60% of the MCP1 than was expected (compared to parental HeLa cells) (p=0.0003 and p=0.0001). The bacterial antigen stimulation led to more MCP1 secretion at 24 hrs than 12 hours on the fibronectin coated slides, however it was still significantly lower than the level secreted by parental hela and the *LPXN* overexpressing cells lines without fibronectin (p=0.0001 for both). The data from the cells not grown on fibronectin indicate an integrin independent effect of the *LPXN* overexpression, whilst data from cells grown on fibronectin indicate an integrin dependant effect. Together this indicated that LPXN may have an effect on MCP1 secretion by epithelial cells, both in an integrin-independent and integrin-dependant fashion.

It is unclear from this experiment whether *LPXN* overexpression affects the signal cascade for secretion or whether it downregulates *MCP1* expression. We know that cell adhesion to fibronectin via integrins induces phosphorylation of the focal adhesion kinase (FAK) (362). Over expression of FAK increases expression of MCP-1 with a subsequent increase in protein levels, and knock out of FAC expression abolishes adhesion induced MCP-1 expression as well as IL6 expression. FAK and NFKB inhibitors genistein and tosyl phenylalanyl chloromethylketone also inhibit MCP1 expression (363). LPXN binds to FAK and other phospho-tyrosine kinases such as PYK2 and in doing sequesters both LPXN and the kinases to the focal adhesion plaques thereby negatively regulating FAK functions. Knock out of FAK allows for LPXN shuttling to the nucleus to function as a transcription cofactor (351), knock out of FAK has also been shown to alter expression of other cytokines including IL6 (364). Therefore, it is possible that *LPXN* overexpression negatively regulates members of the FA, altering cytokine production as well as behaving as a transcriptional cofactor. Given that MCP1 is upregulated by IL6, to confirm that the MCP1

response seen was not in response to IL6, the level of IL6 in the conditioned media was quantified.

3.6.3.2 IL-6

As shown in Figure 3-14, IL-6 was constitutively secreted by the parental HeLa cells with a trend (not significant) of decrease in secretion after 24 hours culture. Stimulation of parental cell lines with bacterial antigens did not significantly change constitutive IL6 secretion at 12 or 24 hours. Interestingly in the absence of integrin signal (fibronectin -), more IL6 was secreted in the *LPXN* overexpressing cell lines (F- LPXN OE), with a trend of more secretion with bacterial antigen stimulation. The decrease in measurable IL6 at 24 hours was again noted within the F- *LPXN* OE cell lines (p=0.0379), although this was not significant in the bacterial antigen stimulated F- *LPXN* OE cell lines. At 24 hours, the F- OE *LPXN* cell line secreted significantly more IL6 than the parental HeLa cells (p=0.0354).

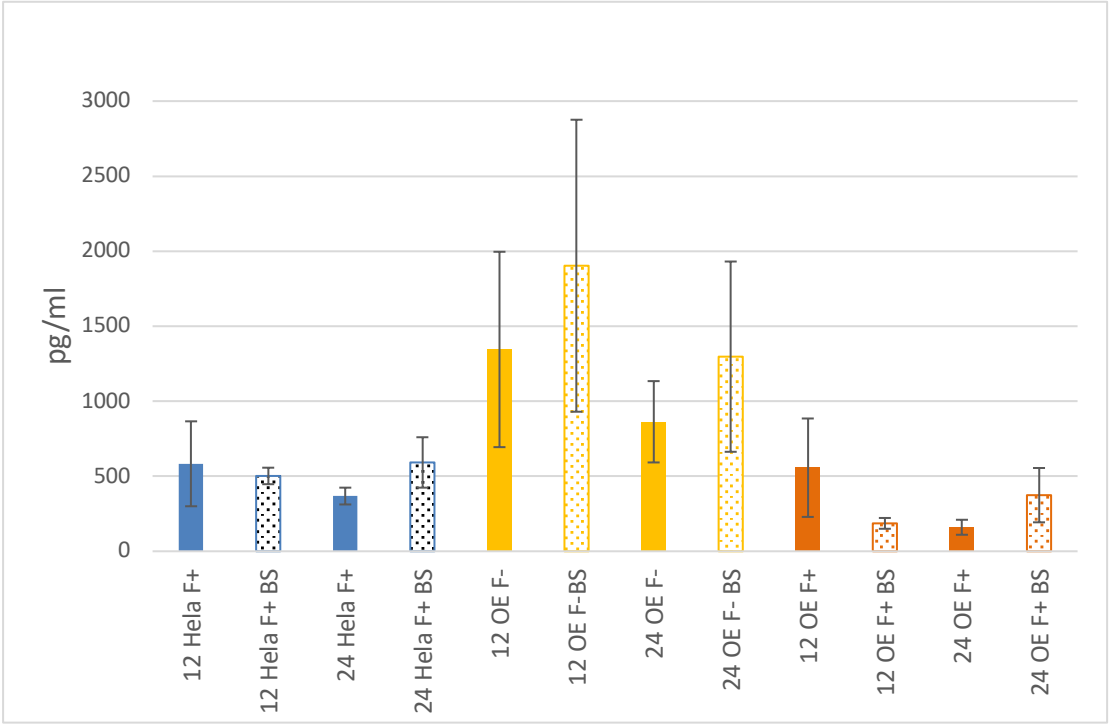


*Figure 3-14 Column graph showing secreted IL6 in pg/ml of conditioned cell media averaged from quadruple replicates, each with duplicate technical replicates. F+ = fibronectin coated glass slide, OE = LPXN overexpressing Hela cell BS = bacterial antigen stimulation. Solid bars are non stimulated cells, dotted bars are stimulated with bacterial antigens. * p=<0.05 compared to 24 hour parental non stimulated hela sample, *p=<0.05 compared to the 12 hr F- LPXN overexpressing sample, **p=<0.001 compared to the 12 hr F+ LPXN overexpressing sample.*

Analysis of the IL6 in the conditioned media of *LPXN* over expressing cell lines grown on fibronectin (F+ *LPXN* OE) gave a very different picture to the non integrin stimulated over expressing cell lines. Whilst at 12 hours the IL6 concentration in the media was

comparable to constitutive levels, by 24 hours it had decreased to a barely detectable (160pg/ml) level (p=0.0046). With bacterial ligand stimulation, the measured IL6 was significantly lower at 12 hours than the non stimulated cultures (p=0.0002), although this increased and was comparable to constitutive levels by 24 hours suggesting a delayed response.

Using Pearson correlation (Figure 3-15) we identified that MCP1 concentration measured across all the cell lines did not correlate with the measured IL6 concentration (r=0.2156, p= 0.1412). Therefore, the changes seen with MCP1 may not be in response to IL6, although more detailed analysis was required. When analysing the individual data points for each cell line the MCP1 concentration in the media from the *LPXN* OE F+ cell line correlated well with IL6 concentration (r=0.934 p= 0.0001), the LPCN OE F- cell line less well (r=0.57 p = 0.019) and the parental HeLa cells did not correlate at all (r= -0.206 p = 0.443). This indicates IL6 involvement in the MCP1 response, but other factors may contribute. To identify if the IL6 response was anomalous, other cytokines were measures including IL8.

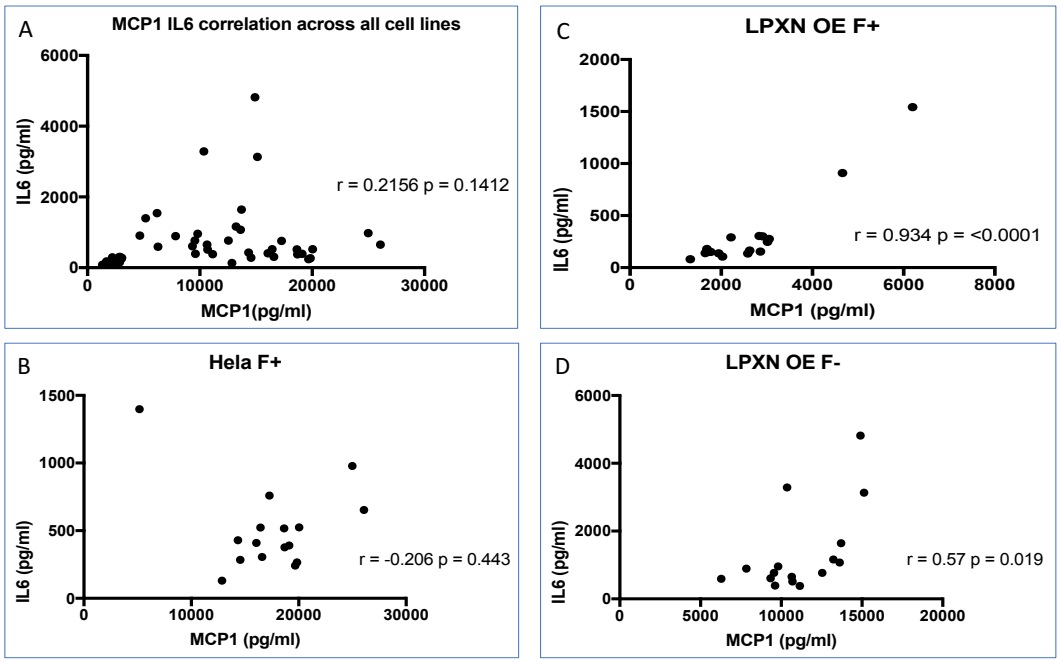


Figure 3-15 Pearson Correlation Graphs assessing the correlation between MCP1 and IL6. A: Analysing all time point values for IL6 and MCP1 in parental Hela cells, LPXN over expressing cell lines grown without and without fibronectin. B Individual time point results for MCP/IL6 secreted from parental HeLa cells grown on fibronectin. C. Individual time point results for MCP/IL6secreted from LPXN overexpressing cell lines grown on fibronectin. D Individual time point results for MCP/IL6secreted from LPXN overexpressing cell lines not grown on fibronectin

3.6.3.3 IL-8

As shown in Figure 3-16, IL8 was constitutively secreted at low level by the parental HeLa cell lines grown on fibronectin, with no significant changes in response to bacterial antigen stimulation. The *LPXN* overexpressing lines grown on fibronectin produced lower levels of IL8 than the parental HeLa controls, but it was only significant in the bacterial antigen stimulated lines at 12 hours (p=0.0003). The cell lines grown without fibronectin (without integrin activation), had higher levels of IL8 in the media compared to both the parental Hela cell lines and the *LPXN* overexpressing cell lines grown on fibronectin, however, due to variability these changes were not significant. The reduction in measurable IL8 at 24 hours is expected, as post stimulation epithelial cells have maximal expression of IL8 at 4-6 hours (365, 366). As expected, changes in IL6 and IL8 correlate r=0.92 p=<0.0001 (Figure 3-17).

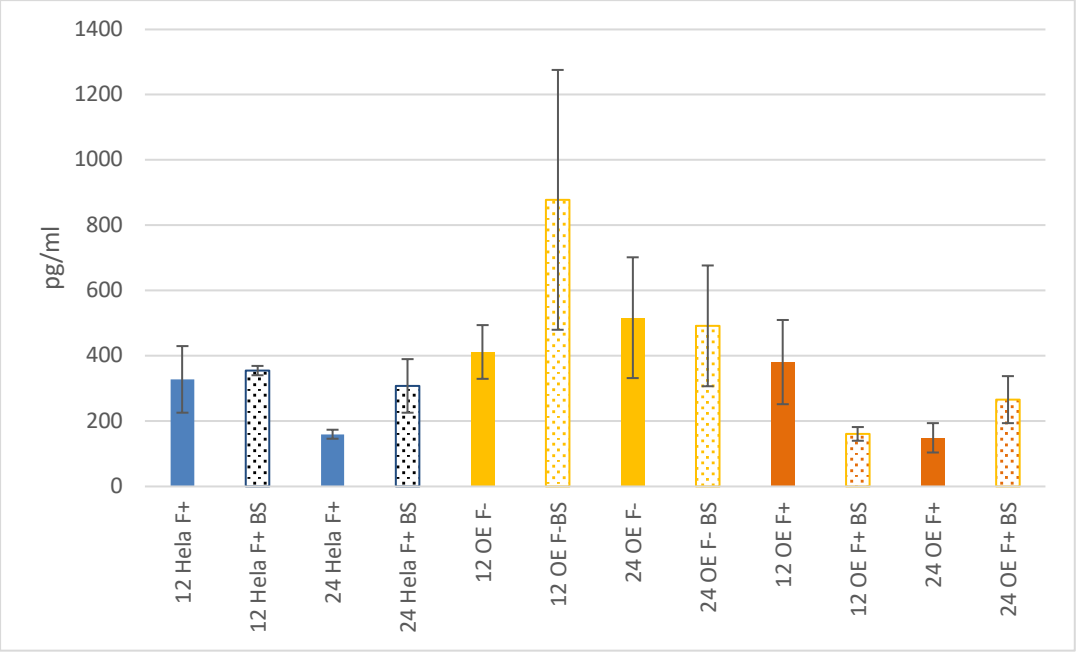


*Figure 3-16 Column graph showing secreted IL8 in pg/ml of conditioned cell media averaged from quadruple replicates, each with duplicate technical replicates. F+ = fibronectin coated glass slide, OE = LPXN overexpressing Hela cell BS = bacterial antigen stimulation. Solid bars are non stimulated cells, dotted bars are stimulated with bacterial antigens. ** p=<0.001 compared to 12 hour parental stimulated hela sample, **p=<0.001 compared to the 12 hr F+ LPXN overexpressing sample.*

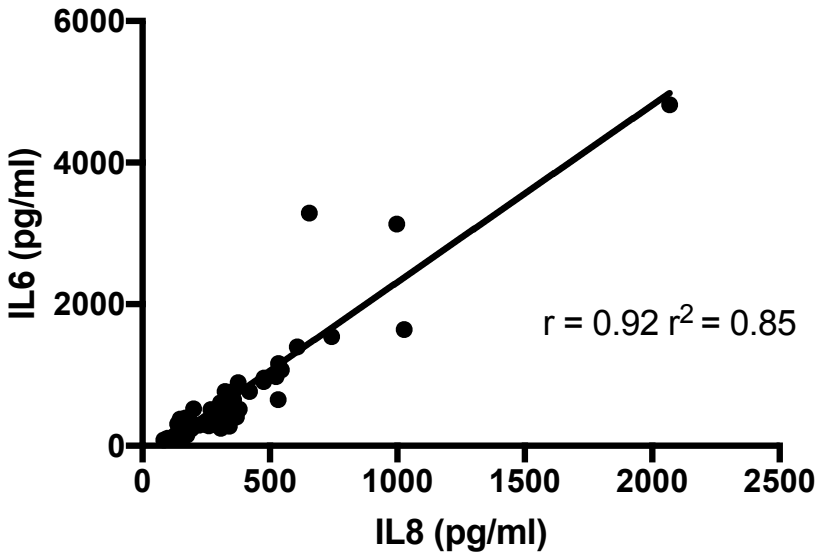


Figure 3-17 Pearson Correlation Graph assessing the correlation between levels of IL-8 and IL-6 secreted into media from parental HeLa cell lines and LPXN overexpressing cell lines grown with and without fibronectin at 12 hours and 24 hour time points.

In summary, *LPXN* overexpression reduces MCP1 secretion in epithelial cells, more so when the integrin signalling cascade is activated by fibrinogen. Part of this response may be due to a reduction in IL6 production. The trend of lower IL6 and IL8 production in *LPXN* over expressing cells with integrin activation may indicate that the cells are less able to respond to TLR and NOD2 stimulation, although more work is required to examine this further.

3.6.4 Colon biopsy cytokine responses

Of the samples collected over 6 months, we utilised 9 sets of biopsies (a biopsy set is 3 biopsies halved into IVOC system and 1 biopsy for baseline measurements) from normal colons, 10 sets of biopsies from quiescent UC patients, and 3 sets of biopsies from inflamed UC colons for cytokine analysis.

3.6.4.1 Cell type by cytokine profile

The minimum detectable concentration (MDC) of the Legendplex assay for each analyte was noted as per the Biolegend MDC in serum which was the closest validated equivalent to the biopsy samples containing cellular material (Table 3-3). There were undetectable levels of IL17A, IL12p70 and IFNa in any of the colonic biopsies therefore the presence of activated T cells was not detectable by their cytokine profile in these biopsy samples. The detectable cytokine profile at baseline would be consistent with cells of the innate immune system including epithelial cells, macrophages, and dendritic cells (Figure 3-17).

Analyte	MDC in Serum (pg/ml)
Human IL1-B	0.9
Human IFN-a	1.5
Human IFN-y	1.1
Human TNF-a	1.0
Human MCP-1	1.1
Human IL-6	1.0
Human IL-8	1.0
Human IL-10	0.8
Human IL-12p70	0.6
Human IL-17A	1.9
Human IL-18	1.1
Human IL-23	1.2
Human IL-33	1.2

Table 3-3 Minimum detectable concentration(MDC) of analytes in serum (pg/ml)based on results from Biolegend LegendPlex protocol.

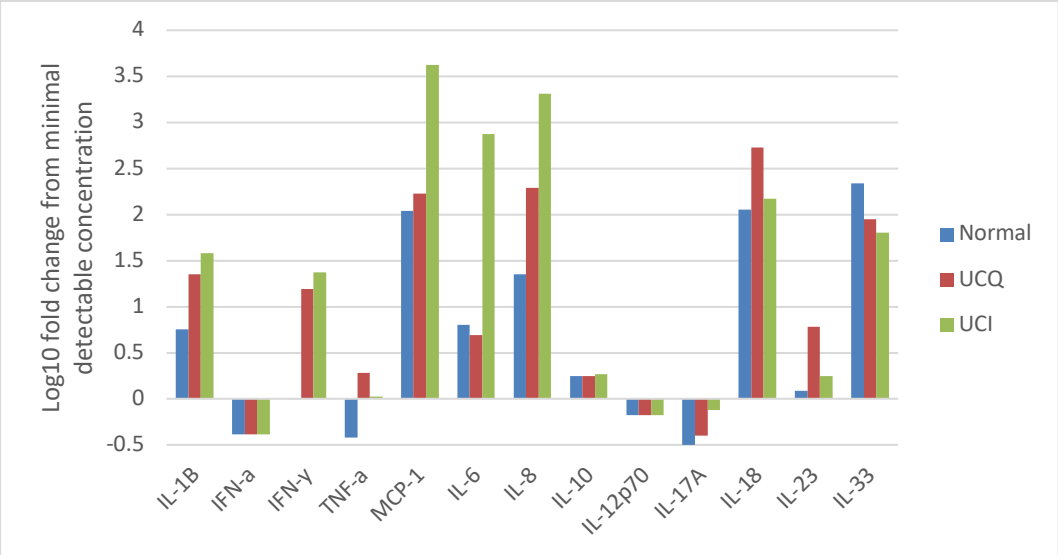


Figure 3-18 Logarithmic fold change from minimal detectable concentration of analytes at baseline (time zero) samples in normal (blue), quiescent UC (UCQ - red) and inflamed UC (UCI- green) colonic biopsies.

Using the polarised IVOC system to assess cytokines response, we were able to quantify the mucosal response to the mechanical stress of pIVOC in terms of cytokine production over 8 hours. The method and the manipulation of the biopsies to place them into the pIVOC system and maintain them for up to 8 hours produced a change in cytokine response over time. Apart from the static cytokines (IL18, IL23), the cytokine response increased over time in the IVOC system

When analysed together as a group, the UC samples which were macroscopically quiescent did not differ significantly from the normal samples in the patterns of cytokine response. The inflamed samples differed at baseline in interferon gamma production compared to the quiescent UC samples ($p=0.0004$) and the normal samples ($p=0.0002$)(Figure 3-19). The UC inflamed samples had significantly altered TNF production compared to normal colons ($p=0.04$) and MCP1 production compared to quiescent UC samples ($p=0.03$). This data confirms that manipulation of biopsies produces a cytokine response, but other than in the expected and documented changes in IFN gamma, TNF and MCP1 production, there is no other significant difference seen in the cytokine response to pIVOC between the diseased and non-diseased samples. This provided a baseline from which to determine if any additional stress or factors can affect cytokine production.

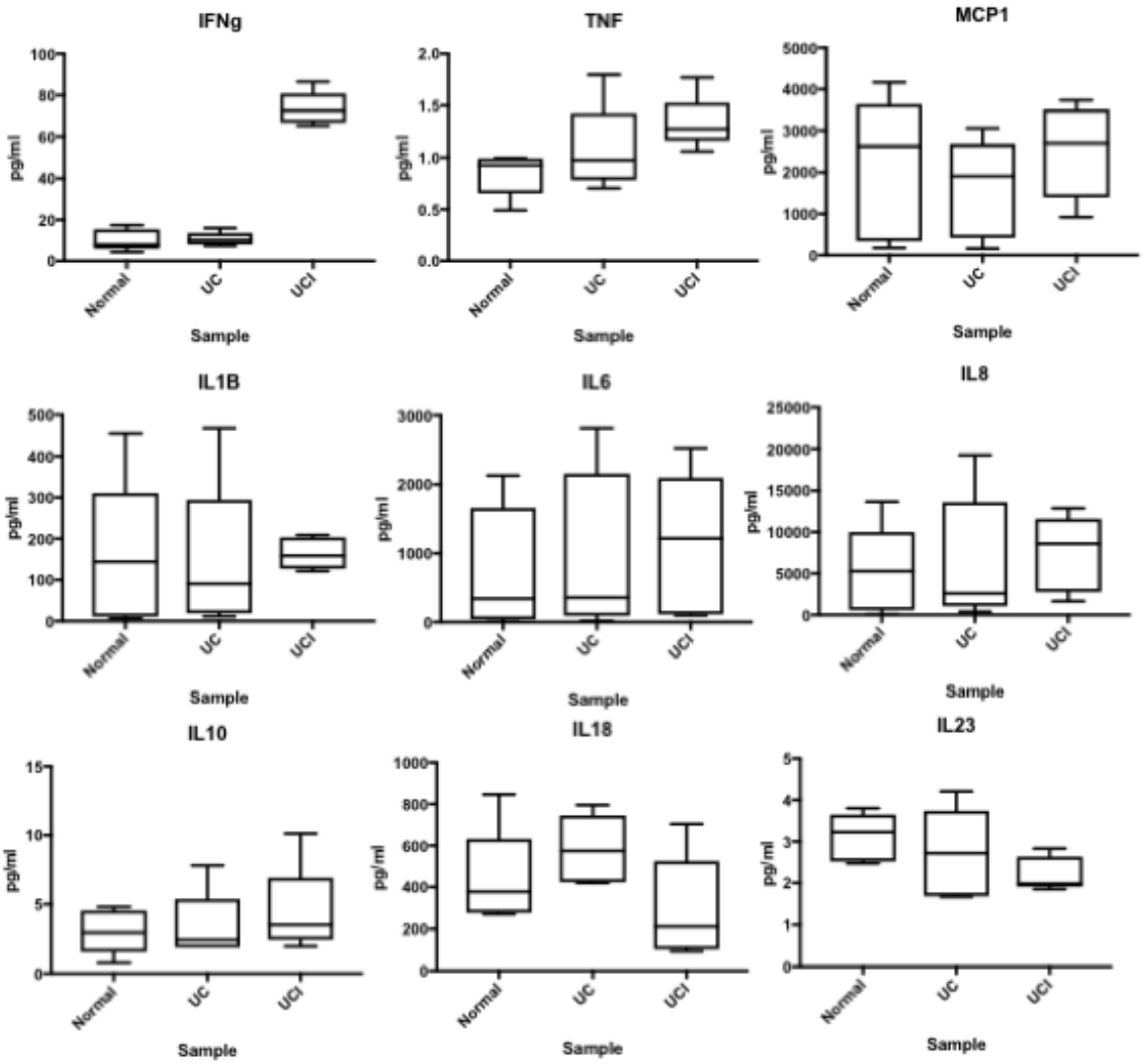


Figure 3-19 Box Whisker plots of measured cytokines (normalised to protein content) in colonic biopsies in pIVOC from normal colons (n=63 individual samples analysed in duplicate from 9 patients), quiescent UC colons (UC)(n= 69 individual samples analysed in duplicate from 10 patients) and inflamed UC colons (UCI) (n= 27 individual samples analysed in duplicate from 3 patients) over 8 hours. The cytokines were quantified used LegendPlex immunobeads.

3.6.4.2 *Cytokine response in genotyped colonic biopsies*

To identify if the *LPXN* SNP rs10896794 had any effect on cytokine production, 11 patients' samples (2 normal, 6 quiescent UC and 3 inflamed UC) were genotyped for the *LPXN* SNP. The samples were considered risk allele positive if they were homozygotes for the *LPXN* SNP risk allele (C<T). Heterozygotes are denoted (Y), and for those samples we were unable to ascertain a phenotype for that SNP in that sample (Table 3-4). All the genotyped samples had technical cytokine analysis duplicates which were reproduced on the legendplex analysis. To test the hypothesis that the *LPXN* SNP may have a role in altered response to bacterial stress, half of the samples underwent bacterial antigen stimulation. These results are only included in the analysis if the test was able to be reproduced with a technical replicate. Each of the analytes were normalized to the total protein in each sample to normalise for differing biopsy sizes.

ID	Disease Group	Age	Sex	IBD Medication	rs10896794 Allele	Bacterial Antigen stimulation
16TB0392	Normal (polyp surveillance)	55	M	Nil	T	No
16TB0398	Normal (IBS)	57	F	Nil	Y	No
16TB0405	UCQ	37	F	5ASA	T	No
16TB0413	UCQ	63	F	5ASA	Y	No
16TB0410	UCQ	45	M	nil	T	1ug + null controls
16TB0409	UCQ	42	M	5ASA	Y	1ug + null controls
16TB0455	UCQ	58	M	Azathioprine	T	10ug + null controls
16TB0457	UCQ	60	F	5ASA	Y	10ug + null controls
16TB0274	UCI	51	M	Nil	Y	No
16TB0389	UCI	34	F	5ASA	T	No
16TB0495	UCI	28	F	5ASA	T	No

Table 3-4 Patients who had colonic biopsies genotyped for rs10896794.Patient samples are anonymised via the tissue bank (TB) ID code. T allele = risk allele, Y = T/C heterozygote at the SNP site. Age, gender and IBD medication are also highlight. (IBS = irritable bowel syndrome).

Proinflammatory cytokines and chemokines

Given the hypothesis that the *LPXN* risk allele has an effect on the NLRP3 inflammasome, IL-1b and IL-18 were analysed as downstream products of NLRP3 activation. Based on the *LPXN* over expressing cell line results; MCP-1, IL-6 and IL-8 were also analysed.

3.6.4.2.1 IL-1Beta

The *LPXN* risk allele in the normal colon had no significant effect on IL-1B production when compared to heterozygote controls (Figure 3-19). IL-1b production was significantly higher in *LPXN* heterozygote (*LPXN* T/C) quiescent UC samples (UCQ) than in *LPXN* risk allele homozygotes (*LPXN* T/T) at 6 and 8 hours. When comparing the normal samples to the UCQ samples; the T/T UCQ samples produced <50% of the IL1-B that the *LPXN* T/T normal samples did at the same time points (4,6,8 hours), with the reverse being true for the *LPXN* T/C samples. This indicates that the *LPXN* SNP risk allele may be affecting the production of IL1-B in the UCQ samples. For the inflamed UC samples, the *LPXN* T/T samples contained more IL-1b at 6 hours and 8 hours (p=0.0006, p=0.0003) as compared to the *LPXN* T/C time matched pairs. The *LPXN* T/C samples produced lower levels of IL1-b than the normal colonic samples.

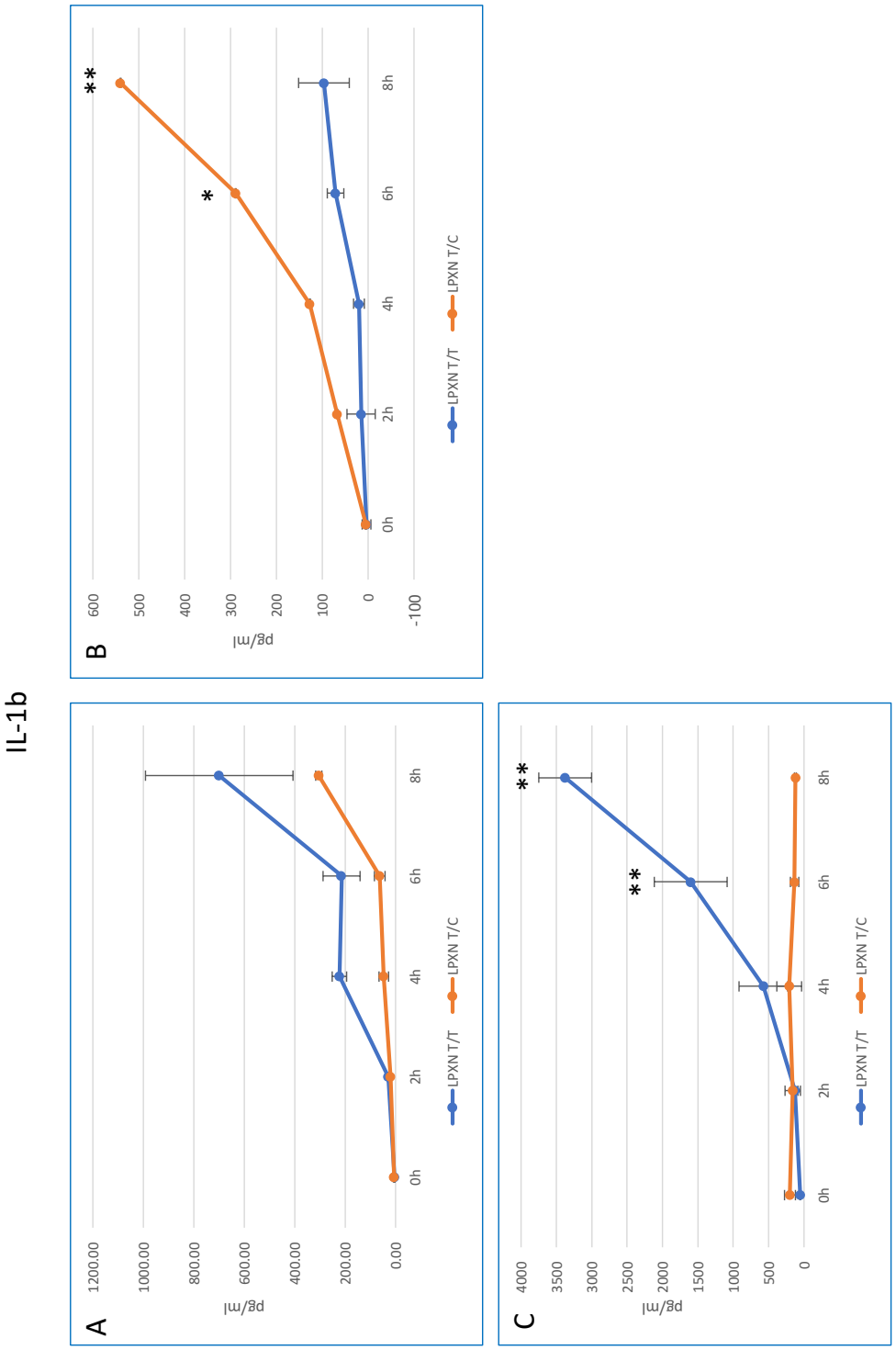


Figure 3-20 IL-1B quantified using Legend Plex Assay. ***A****; genotyped samples from normal colonic biopsies in a pIVOC which were homozygous for LPXN risk allele (Blue)or heterozygotes at the LPXN SNP site(red). SEM bars shown. Two tailed t-test was not significant.* ***B****; IL-1B quantified from normal colonic biopsies in a pIVOC which were homozygous for LPXN risk allele (Blue)or heterozygotes at the LPXN SNP site(red). SEM bars shown.* ***C****: IL-1b quantified from quiescent UC colonic biopsies in a pIVOC which were homozygous for LPXN risk allele (Blue)or heterozygotes at the LPXN SNP site(red). SEM bars shown. Two tailed T- test: *= p<0.05 ** = p<0.001*

3.6.4.2.2 IL18

In normal colonic biopsies IL-18 production did not differ significantly over 8 hours from each of the sample baseline levels. The *LPXN* risk allele homozygte samples had significantly more IL-18 compared to *LPXN* T/C heterozygote samples (p=0.0003 at 2 and 6 hours, p=0.02 at 4 hours) (Figure 3.20). There was no significant difference between any of the UCQ samples either from baseline or between the *LPXN* T/T homozygotes or T/C heterozygotes. The *LPXN* T/C inflamed UC samples produced significantly less IL-18 at 4,6 and 8 hours than the *LPXN* T/T samples. Interestingly, excluding the baseline inflamed samples, the IL-18 levels in the inflamed samples were directly comparable to those in the normal samples for both the *LPXN* T/C and *LPXN* T/T. This suggests changes in production of IL-18 seen with the *LPXN* SNP may not be directly involved with the acute inflammatory process in UC.

3.6.4.2.3 IL-6

The *LPXN* T/T had no significant effect on IL6 in the normal colons (Figure 3-22)compared to the *LPXN* T/C samples. Variability of the measured IL6 in UCQ samples meant that there was no significant difference between the *LPXN* T/C and *LPXN* T/T samples, however in the inflamed colonic samples there were significantly greater levels of IL6 production compared to the UCQ samples. Between the *LPXN* T/T and *LPXN* T/C inflamed samples there was a 6-8 fold difference (p=<0.0001) in IL-6 production.

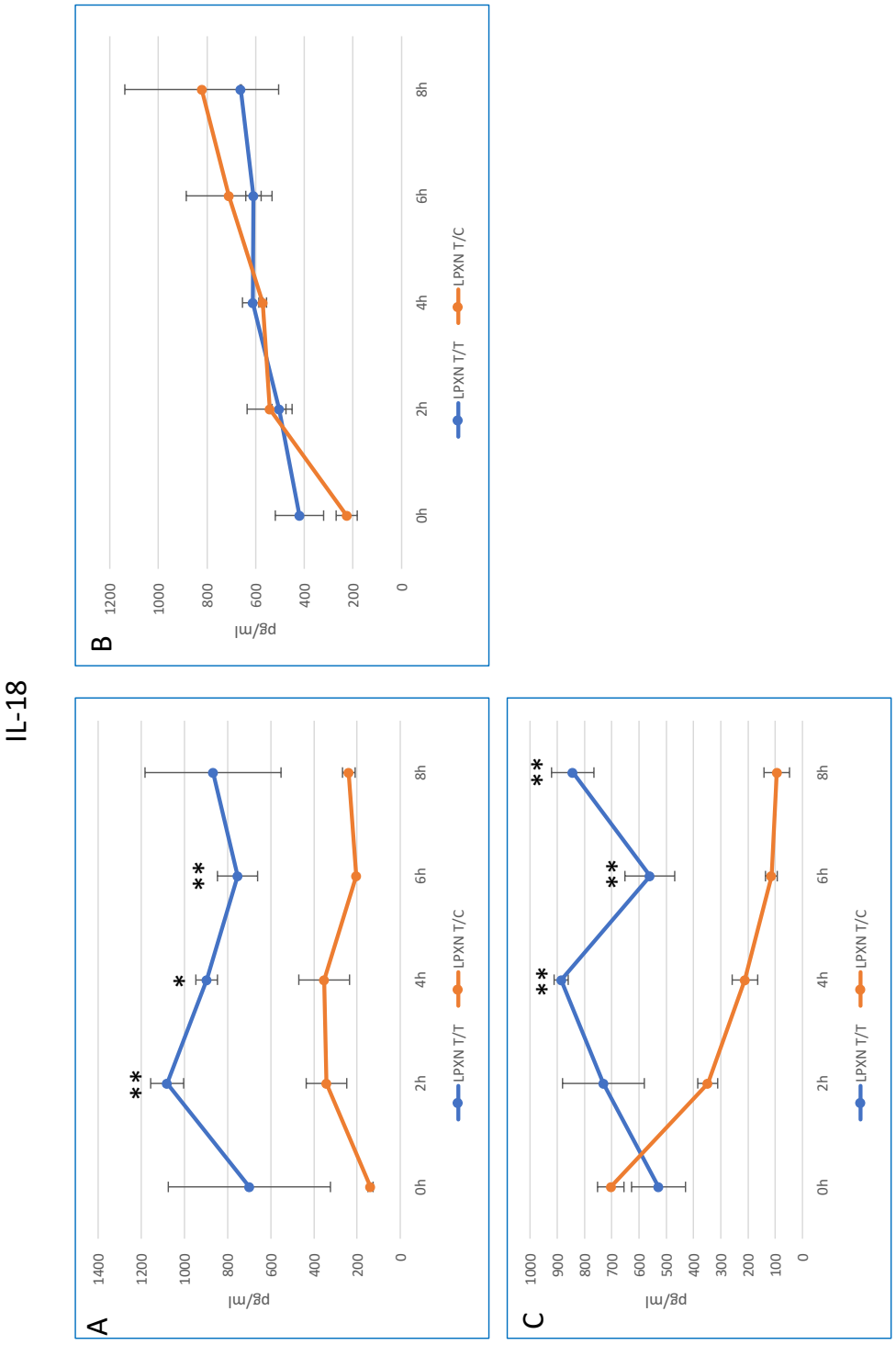


*Figure 3-21 IL-18 quantified using Legend Plex Assay **A**: normal colonic biopsies in a pIVOC which were homozygous for LPXN risk allele (Blue)or heterozygotes at the LPXN SNP site(red). SEM bars shown. **B:** IL-18 quantified from UCQ colonic biopsies in a pIVOC which were homozygous for LPXN risk allele (Blue)or heterozygotes at the LPXN SNP site(red). SEM bars shown. **C:** IL-18 quantified from inflamed UC colonic biopsies in a pIVOC which were homozygous for LPXN risk allele (Blue)or heterozygotes at the LPXN SNP site(red). SEM bars shown. Two tailed t-test; *= p<0.05 compared to same hour point in LPXN risk allele homozygote samples, ** = p<0.001 compared to same hour point in LPXN risk allele homozygote sample.*

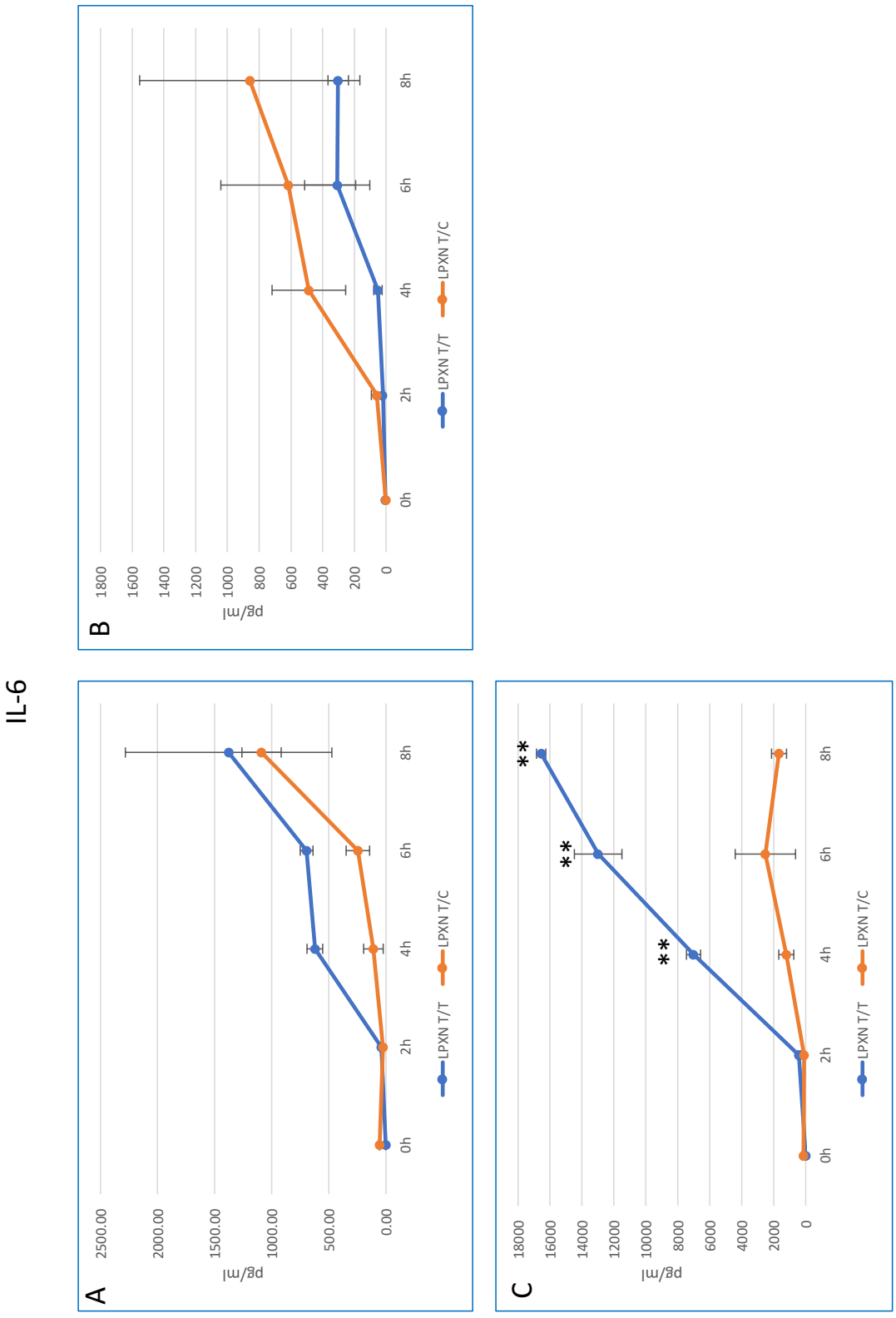


*Figure 3-22 IL-6 quantified using Legend Plex Assay **A:** IL6 quantified from normal colonic biopsies in a pIVOC which were homozygous for LPXN risk allele (Blue)or heterozygotes at the LPXN SNP site(red). SEM bars shown. **B:** IL-6 quantified from UCQ colonic biopsies in a pIVOC which were homozygous for LPXN risk allele (blue)or heterozygotes at the LPXN SNP site(red). SEM bars shown.**C:** IL-6 quantified from Inflamed UC colonic biopsies in a pIVOC which were homozygous for LPXN risk allele (blue)or heterozygotes at the LPXN SNP site(red). SEM bars shown. Two tailed t-test; ** = p<0.001 campared to same hour point in LPXN risk allele homozygote sample.*

3.6.4.2.4 IL-8

The trend in IL8 production in normal biopsies was directly comparable to IL6 production in normal biopsies, with significantly more IL8 being produced in *LPXN* T/T samples compared to *LPXN* T/C samples (Figure 3-23) (*p=0.04, **p=0.0009). Like IL6, the UCQ samples were the opposite to the normal samples, with *LPXN* T/C showing a greater rise in IL8 production over time compared to *LPXN* T/T (p=0.0024 at 4 hrs, p=0.0003 at 6 and 8 hours). In the inflamed UC samples, there was significantly less IL8 produced in the *LPXN* T/T samples than the T/C samples, for the *LPXN* T/C samples there was double the amount of IL8 produced from the inflamed samples compared to the normal samples. There was no significant difference between the normal *LPXN* T/T samples and the inflamed UC *LPXN* T/T samples. This indicates that the *LPXN* T/T may reduce the pro-inflammatory production of IL8 in an inflamed colon.

3.6.4.2.5 MCP1

MCP-1 production was consistent higher in the *LPXN* T/T samples for all 3 types of colonic biopsy (normal, UCQ and UC inflamed) compared to the *LPXN* T/C samples (Figure 3-24). In the inflamed samples this reached significance, in the UCQ samples this did not. The results were analysed via a Pearson correlation to identify if the changes seen in MCP1 production were due to or separate from changes in IL6 production (Figure 3-25). MCP1 and IL6 production correlated for all samples except for the normal *LPXN* T/T colonic sample, indicating like in the cell line cytokine analysis that IL6 has a role in MCP1 cytokine changes seen with these results. However, although MCP-1 and IL6 levels did correlate, when taking into account the scale of production, all of the *LPXN* T/T samples produced more MCP-1 than the heterogeneous counterparts, but this is not the case for IL6 production. This indicates that other factors are involved with MCP-1 production, such as the *LPXN* risk allele and patient specific features such as treatments and the presence of other genetic and environmental factors which could not be controlled for in this study.

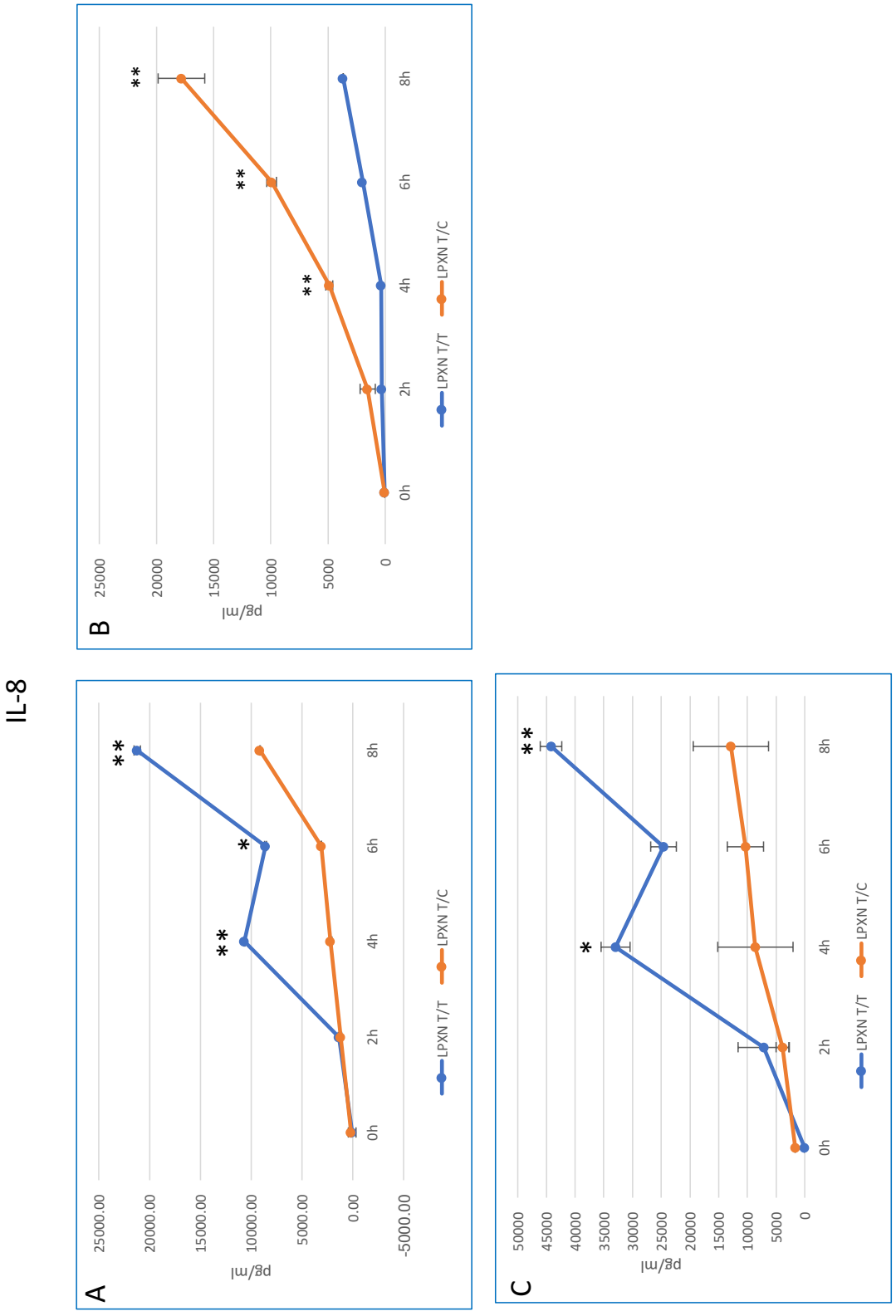


*Figure 3-23 IL-8 quantified using Legend Plex Assay **A:** IL-8 quantified from normal colonic biopsies in a pIVOC which were homozygous for LPXN risk allele (Blue)or heterozygotes at the LPXN SNP site(red). SEM bars shown. **B:** IL-8 quantified from UCQ colonic biopsies in a pIVOC which were homozygous for LPXN risk allele (blue)or heterozygotes at the LPXN SNP site(red). SEM bars shown.**C:** IL-8 quantified from Inflamed UC colonic biopsies in a pIVOC which were homozygous for LPXN risk allele (blue)or heterozygotes at the LPXN SNP site(red). SEM bars shown. Two tailed t-test; *= p<0.05 compared to same hour point in LPXN risk allele homozygote samples, ** = p<0.001 campared to same hour point in LPXN risk allele homozygote sample*

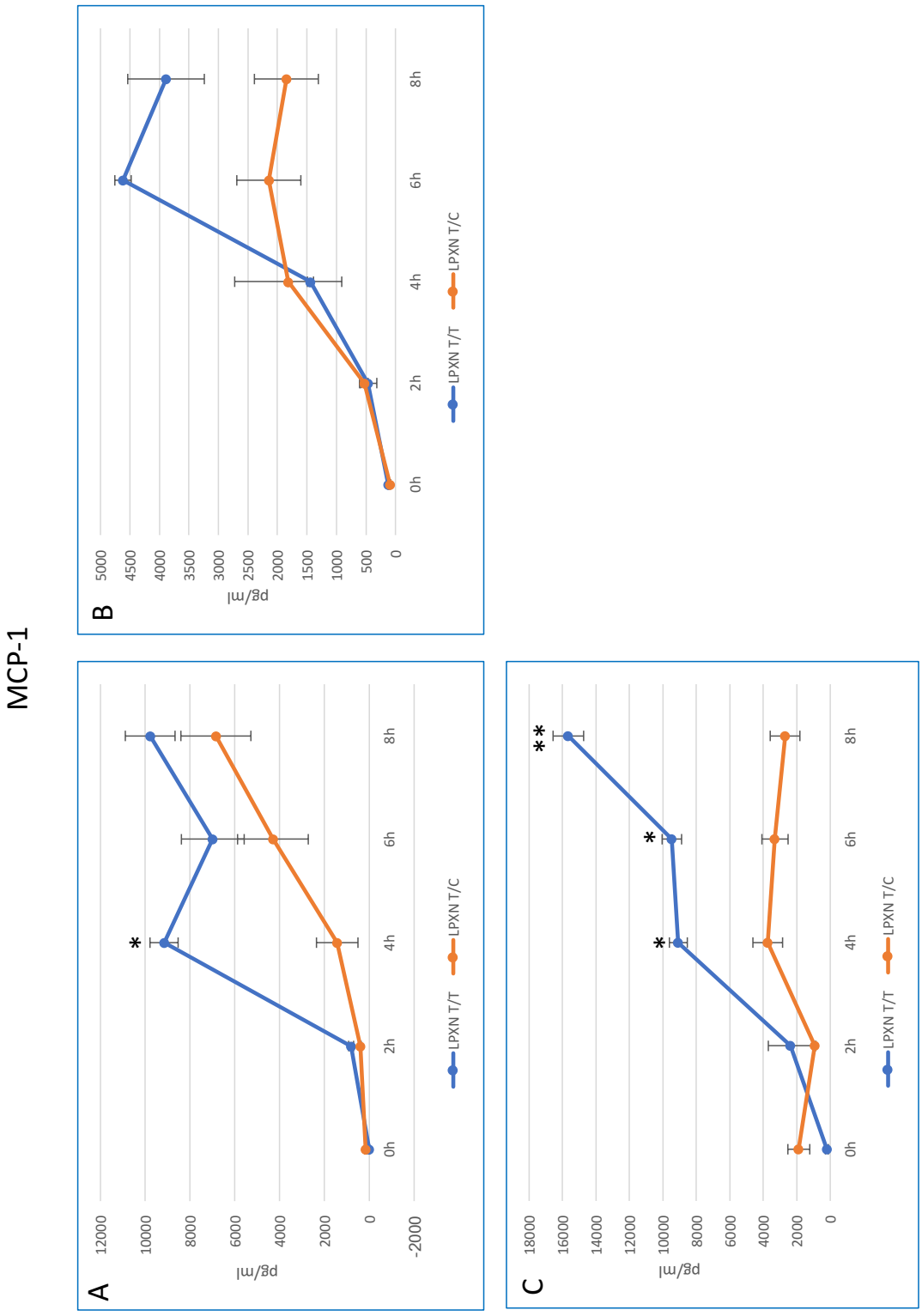


*Figure 3-24 MCP-1 quantified using Legend Plex Assay **A:** MCP-1 quantified from normal colonic biopsies in a pIVOC which were homozygous for LPXN risk allele (Blue)or heterozygotes at the LPXN SNP site(red). SEM bars shown. **B:** MCP-1 quantified from UCQ colonic biopsies in a pIVOC which were homozygous for LPXN risk allele (blue)or heterozygotes at the LPXN SNP site(red). SEM bars shown.**C:** MCP-1 quantified from Inflamed UC colonic biopsies in a pIVOC which were homozygous for LPXN risk allele (blue)or heterozygotes at the LPXN SNP site(red). SEM bars shown. Two tailed t-test; *= p<0.05 compared to same hour point in LPXN risk allele homozygote samples, ** = p<0.001 campared to same hour point in LPXN risk allele homozygote sample*

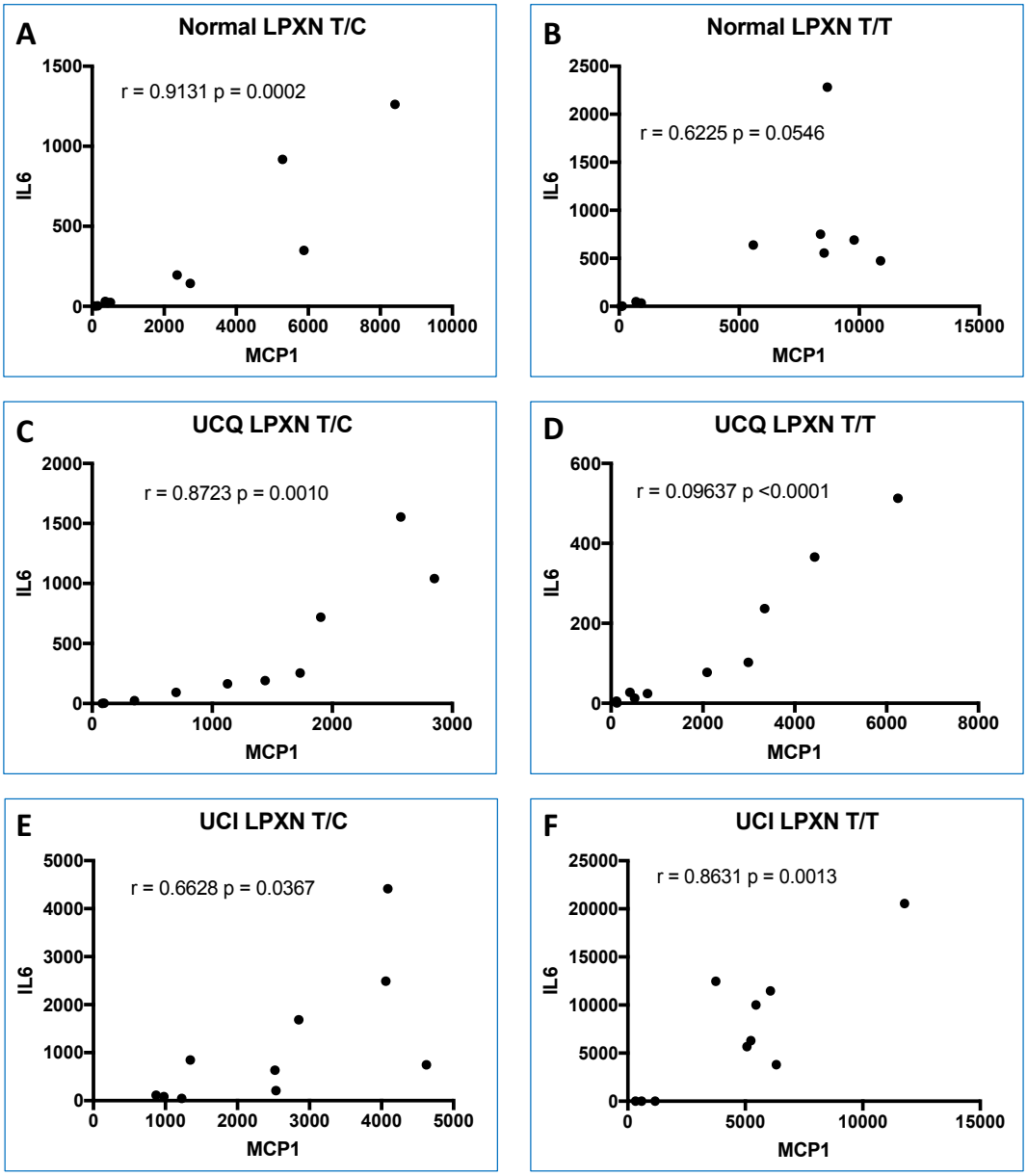


Figure 3-25 Pearson correlation of MCP1 and IL6 in pg/ml in pIVOC samples over 8 hours of incubation. ***A****: normal colonic samples heterogeneous at the LPXN SNP site.* ***B****: normal colonic samples homozygous for the risk allele at the LPXN SNP site.* ***C****: UCQ colonic samples heterogeneous at the LPXN SNP.* ***D****: UCQ colonic samples homozygous for the risk allele at the LPXN SNP site.* ***E****:UCI colonic samples heterogeneous at the LPXN SNP.* ***F****: UCl colonic samples homozygous for the risk allele at the LPXN SNP site.*

3.6.4.3 Cytokine response to bacterial ligand stress in genotyped colonic biopsies

To start to determine if a bacterial ligand stress was the driver of inflammation in patients with the LPXN risk allele, genotyped UCQ samples were analysed using "self" controls (half the biopsy stressed, the other half with no ligand stress). These were single samples, therefore statistical analysis was not undertaken; this experiment was undertaken as a pilot experiment to ascertain if bacterial ligand stimulation was feasible in a polarised biopsy with a theoretically intact mucosal layer and its own microbiota.

The markers of NLRP3 inflammasome activation, IL1-b and IL-18 both showed an initial heightened response to supra-physiological levels (10ug/ml) of bacterial ligand stimulation. The 1ug/ml LPS, PGN and MDP cocktail stimulation elicited a trend of higher IL-1b and IL18 compared to the non- stimulated controls (Figure 3-26).

IL6, IL8 and MCP-1 production all had a trend of increased production in the supra-physiologically stimulated samples (Figure 3-27) compared to their controls. More samples are required to characterise this further.

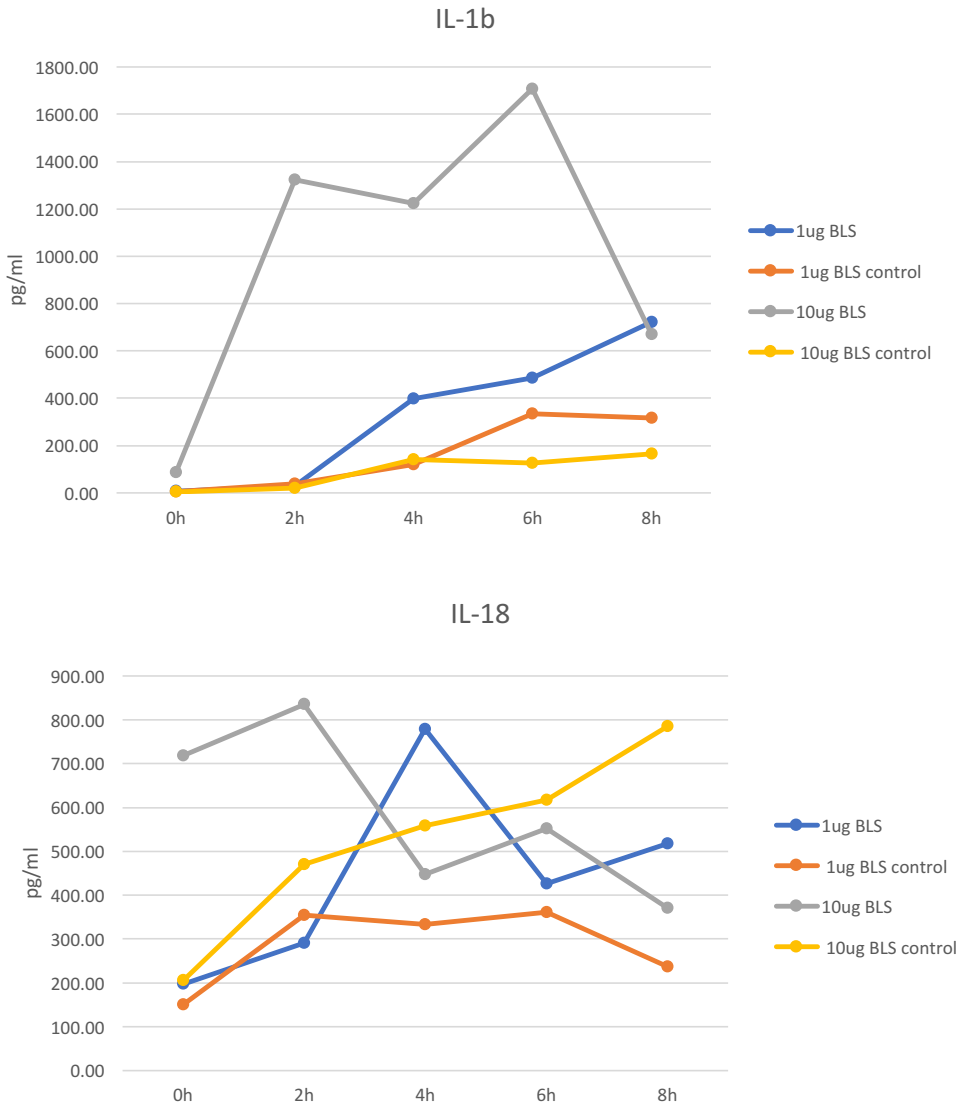


Figure 3-26 IL-1B and IL-18 production in UCQ samples containing the LPXN risk allele T/T. The figure shows both stimulated and non-stimulated controls from the same biopsy.

Figure 3-27 IL-6, IL-8 and MCP-1 production in UCQ samples containing the LPXN risk allele T/T. The figure shows both stimulated(BLS) and non-stimulated(Non-BLS) controls from the same biopsy.

3.7 Discussion and conclusion

3.7.1 Cell lines

The creation of multiple different *LPXN* expression cell lines was difficult. Low transfection rates of plasmids into intestinal epithelial cell lines was overcome by using an easier to transfect cell line (HeLa) and trialling multiple transfection reagents until the most effective for the plasmids used was found. 3 CRISPR plasmids were trialled with differing selection methods (antibiotic resistance, CD4+ expression for bead selection and GFP expression for flow-cytometry), the best performing in this circumstance were the antibiotic positive selection plasmids. It was decided to use the same selection technique for each of the plasmids to reduce the confounding effect of selection variables for both optimisation of the technique and results analysis.

There were persistent difficulties with the *LPXN* knock out cell lines, as the transfection rates were slightly lower than the sham controls for reasons that are not clear, and the growth of the knock out cell lines was significantly slower which can be explained by the lack of LPXN, with the stress of the transfection and individual cells not surviving without other cells in contact with them. Optimisation led to the use of 50% conditioned cell media until the cell lines were growing in 6 wells which meant that after 6 weeks, we could analyse the *LPXN* knock out cells by qPCR, but were still unable to undertake wound healing as the cells would die before confluency. *LPXN* expression has not been characterised in Hela cells before, one paper undertook protein quantification in cell types using western blot which suggested that Hela did not contain LPXN. Using commercial antibodies, our qPCR results indicate that *LPXN* is expressed in Hela cells and has appreciable protein production as shown by Western blot.

LPXN is a poorly understood LIM binding adapter protein whose role within the focal adhesion complex and role as a biomarker of cancer prognosis (breast and prostate) is just beginning to be evaluated. There have been 27 papers characterising the role of LPXN since 1998, compared to the 2747 papers since 1990 for Paxillin, a more ubiquitous LIM binding adapter protein.

LPXN was identified as a potential mediator in the susceptibility to or pathogenesis of UC from the network analysis of the functional genomics of the UC risk susceptibility SNPs. It was unclear from the literature which cell type would have the predominant affect from the *LPXN* SNP. We have shown that not only is LPXN present in epithelial cell lines, but knocking out LPXN in immortalised cell lines had a significantly detrimental effect on cell

growth. Overexpression of *LPXN* in Hela cell lines significantly reduced wound healing and with integrin activation significantly reduced the amount of inflammatory cytokines IL6, IL8 and MCP1 secreted into the media. This indicates that *LPXN* overexpression may impact on the ability of epithelial cells to signal to other cells of the innate immune system. Where we would have expected an increased IL6 and IL8 response to bacterial stimulation, this was not seen in the overexpression cells lines with integrin activation, highlighting a potential inability of these epithelial cells to respond appropriately to bacterial stimuli.

There are several intestinal epithelial cell models such as Caco2, HT29 and T84 – these are all immortalised colorectal cancer cell lines and whilst the Hela cells were easier to transfect they do not have ideal features of intestinal cells lines such as polarisation. Immortalised cell lines have been used in the literature as an in vitro model of colitis such as using DSS on caco2 cells (338)Specific cells such as dendritic cells and cell lines such as THP1 have been used to investigate pathogenesis pathways of colitis such as enteric cell education of dendritic cells (339) and identifying how carrageenan works on THP1 cells to cause the phenotype seen in carrageenan induced colitis mouse models(340). Hela cells have not been used in this context, for obvious reasons, but this case have provided a good indicator for how an epithelial cell would respond to changes in *LPXN* expression. There are limitations to utilising over expression and gene silencing to ascertain the phenotype of a SNP, most pertinently, given that the phenotype is unlikely to be pronounced, creating severe perturbations in the expression of LPXN, an adaptor protein with significant roles in signalling cascades is likely to have a more pronounced effect than a single SNP within LPXN will do. Future work on cell lines would include completion of the CRISPR-Cas9 homologous recombination technique to create immortal 'SNP' cell lines, other parallel techniques could include the creation of primary cell lines from genotyped individuals for further characterisation of the *LPXN* SNP within specific cell types such as enterocytes.

3.7.2 Polarised in vitro organ culture

The use of the polarised in vitro organ culture system overcomes the inherent problems of immortalised epithelial cell lines. The pIVOC is a snap shot of all the interacting cell types in the mucosa at the time of biopsy. It has been designed and used to examine adherence of and host responses to *E.coli* species (361, 367). We have utilised the technique to attempt to characterise the impact of genetic changes to the cellular response to mechanical stress and start to characterise the impact of bacterial ligand

stress. A limitation of this approach is that we have not shown that LPXN expression is altered in the genotyped biopsies that display SNP homozygosity. This could be addressed by qPCR and should be a priority for future SNP assessment using pIVOC on genotyped samples.

However, our results suggest that the *LPXN* risk allele in the cells present within the mucosa, may lead to altered cytokine response in normal colonic mucosa, quiescent UC mucosa and inflamed UC mucosa. An interesting question that needs to be answered raised by the data from patients with the normal mucosa but the *LPXN* T/T alleles who clearly had a different cytokine response to those with the T/C alleles, yet they do not have UC – what are the additional factors that translate a genetic mutation to a disease? The presence of a single SNP alone does not do this. Identification of other pertinent SNPS within the biopsies, such as the ATG16L1 T300A needs to be undertaken given the known effect of the SNP on IL1beta expression. Consistent with this would be further work to identify the synergistic effect of multiple SNPs and environmental factors such as the response to bacterial products needs to be done. The quiescent UC patients with the *LPXN* T/T allele had a consistently dampened pro inflammatory and NLRP3 cytokine response (IL-6, IL-8, IL-1b, IL-18), and even when flooded with bacterial ligand stimulation did not raise a cytokine response that approached the inflamed samples cytokine response. An explanation for this may be that all the patients with UCQ had well controlled disease, with minimal immune cell infiltrate within the mucosa and therefore the cells that would usually be recruited in an inflammatory response e.g. granulocytes and lymphocytes were unable to be recruited in the ex vivo sample. However, I would have expected the UCQ samples to respond in a similar manner to the normal samples if the SNP was the only variable.

To conclude; we hypothesised that *LPXN* over expression in cell lines would reduce wound healing and dampen the cytokine response to bacterial ligand stimulation. *LPXN* over expression reduced wound healing and reduced the secreted cytokine and chemokine response from the cells. Consistent with this, the UC susceptibility SNP rs10896794, when homozygous for the risk allele, produced dampened cytokine responses in UCQ samples. Given that the aim of this experimental work was to validate an *in silico* prediction, these findings add weight to the suggestion that the non-coding SNP rs10896794 may impact on the pathogenesis and ongoing inflammation found in UC, but does not show this conclusively. The data indicates that there are multiple factors working synergistically to

produce the dysregulated inflammatory response seen in UC, therefore, further work to outline and identify the role of synergistic SNP effects on individual patients has been undertaken and is outlined in the next chapter.

The role of the focal adhesion complex in UC pathogenesis remains intriguing with increasing interest in the role integrins play in IBD pathogenesis (164) and treatment, further work to examine this complex signalling cascade and the cross signalling with the NLRP3 inflamasome may provide new avenues for therapeutic options in patients with SNPs affecting the FAC.

With regard to pIVOC with UC biopsies, with undergraduate students and in collaboration with the Hall group (QIB), further work has been undertaken looking at the effect of adding protective microbiota species on cytokine responses in UC and cytokine responses to other bacterial products in UC and normal colons.

4. Moving towards personalised medicine by creating patient SNP 'footprints'

Overarching aim:

To identify patient-specific pathogenic pathways to disease from their genotype

---

4.1 Acknowledgements

Given the large and complex nature of the UC Interactome and creating 58 individual patient interactomes, automation of the SNP workflow was necessary. This was undertaken by employing a software engineer supported by a Norwich Research Park Translation Fund, and working with bioinformaticians including David Fazekas, and Dr Paddy Sudhakar, from the Korcsmaros group. The workflow was called the integrative SNP network platform (iSNP). Analysis and clustering of the networks was undertaken with Dr Dezso Modos, a collaborator from Professor Bender's Group, Cambridge University.

I am grateful to the UK IBD Genetics Consortium for allowing me access to their expertise, and data repository and I am especially grateful to Jeff Barrett, Miles Parkes and Dan Rice.

4.2 Introduction

There are multiple clinical subphenotypes of UC e.g. proctitis vs pancolitis, with or without extra-intestinal manifestations; it can even be argued that those who 'fail' treatments also have a different phenotype to those who have 'burnt out' disease and require no further pharmacological or surgical input, with graduations in between. Whilst GWAS has highlighted SNPs associated with UC, not every patient has every SNP as evidenced by the significantly variable minor allele frequency.

One of the major goals of GWAS research is to identifying the pathogenic mechanism of the genetic predisposition with early research focussed on <10% of SNPs which occur in exonic regions and alter the amino acid structure of the translated protein. It is clear, however, that complex genetic diseases such as UC these SNPs, although theoretically interesting, do not confer the pathogenic effect that was expected (368). Moreover, the

individual SNPs identified do not explain the missing heritability of UC (369) suggesting that there is a further unidentified factor in the 'genetic predisposition'. >90% of UC associated SNPs are non exonic SNPs, which include intronic, intergenic or regulatory region SNPs. Functional interpretation of non-coding SNPs can be challenging (43). Non coding SNPs can occur in functional DNA elements including long non coding RNAs (lncRNAs) (370, 371), microRNAs (miRNAs) (371, 372), miRNA binding sites on mRNA (373), and transcription factor binding sites (374). These functional DNA elements have a unique role in controlling cellular regulatory cascades in a dynamic, complex and temporally mediated manner (375). It has been hypothesised that SNPs within regulatory elements can function to fine tune the regulation. Validated pipelines such as CAUSEL (376) are evidence of the difficulty of identifying causal phenotypic effects from individual non-coding polymorphisms on gene regulation.

4.3 Hypothesis, Aims and Objectives

What hasn't been clear is if there is a summative or combined effect of regulatory gene changes caused by the presence of SNPs. We hypothesised that using a network biology approach to functionally annotate SNPs we could identify not only key pathogenic pathways to disease, but by using individual patient SNP genotypes, identify a mechanism to stratify patients based on their genotype and clinical parameters.

I further hypothesised that individual patient genotypes have an impact on or correlate with their phenotype. This is not without precedent, for example; individual SNPs associated with an increased risk of pancreatitis with Azathioprine use has already been identified as well as SNPs associated with IBD prognosis (343). However in-depth analysis of individual patient SNP profiles and biological mechanisms and pathways affected by the UC associated SNPs is lacking.

To test the hypothesis that there is a correlation between SNP function, associated SNP burden and disease, individual patient SNP profiles have to be constructed from publically available SNP datasets.

The aims for this project were twofold:

1. Utilise the iSNP workflow to identify key pathogenic pathways to disease from individual patient data - their 'footprint';
2. Correlate the patient footprint with their clinical data.

In order to do this, three objectives were identified:

1. Create individual footprints based on genomic data;
2. Undertake unsupervised clustering and gene ontology analysis of the clustering;
3. Undertake supervised clustering to identify if particular clinical phenotypes correspond with the clusters.

4.4 Methods

We obtained 58 individual Immunochip (IC) data sets as part of the UK IBD Genetics consortium, corresponding to a cohort of Norwich based patients with confirmed endoscopic and histological diagnosis of UC, all of whom are under the care of IBD physicians at the Norfolk and Norwich University Hospital. Two were excluded as their diagnosis had changed from UC to CD since the original data collection. Each patient notes were reviewed by an IBD physician to document the clinical parameters of demographics, disease site, disease treatments, extending back 20 years or to time of diagnosis depending or whichever was sooner, documented side effects of IBD treatment, extraintestinal manifestations and comorbidities.

We extracted each individual patient immunochip data set from the PLINK based format. Patient specific UC associated SNPs were identified against a combined list of IC UC associated (finemapped) SNPs, and UC associated finemapped (FM)SNPs. We utilised all of the finemapped SNPs with the highest PIC values, enriched to the colonic mucosa or with no enrichment. We identified GWAS UC associated SNPs if the IC SNP to GWAS SNP R^2 value was <0.8. If GWAS SNPs were included (p=<5E-7), we linked their highest scoring (R^2 = 1) linkage disequilibrium partners (via HapMap and 1000genomes) with them to undergo analysis *en bloc*. If the IC and FM SNPs overlapped and were concordant, both SNPs were put forward for analysis. If there was discordance, then both SNPs went forward, but were linked and analysed together.

The SNPs from the patient matrix then underwent the automated iSNP workflow. The methodology of the SNP workflow remained the same as in Chapter 2, except that it used the most up to date versions of each database with the highest stringencies and was automated. Splice alteration, lncRNAs and mature miRNAS were not used within the methodology due to the potential high false positive rates.

To correlate the anonymised patient data with the demographic and clinical data regarding IBD treatment, smoking, surgery that had already been collected as part of the original data collection, I worked with Dr Mark Tremelling at the Norfolk and Norwich

University Hospital to correlate the anonymised labels with the patient data. This data was then updated by myself as a clinician and re-anonymised and re-encrypted. The patient identifiable data remained at the Norfolk and Norwich University Hospital on a password protected spreadsheet, on a password protected computer that was only accessible via an ID swipe in system by clinical staff.

The Patient SNP matrix was input with the functional annotation and first neighbours into initially Excel and then Access to create an interaction matrix. The Hamming distance was calculated between each patient. Unsupervised clustering analysis was undertaken, creating Tree Clusters based on the Hamming distance and average summarization.

Supervised clustering with clinical data based on disease site, severity, treatment and comorbidities on the unsupervised clusters was then undertaken. Statistical significance was assessed using the Fisher Exact Test, in Microsoft excel.

Gene Enrichment and Pathway analysis was undertaken using the Panther Classification System (377). Each individual patient interaction matrix uniprot IDs were uploaded to the Pantherdb web service. Statistical overrepresentation was undertaken using GO Biological Process and Bonferroni correction, reference list Homo Sapiens, $P<0.05$ was considered significant. Functional classification was undertaken and list converted to a Panther Pathway List.

The network was analysed to identify the top 75% commonest first neighbour to the SNPs to identify converging pathways. Each converging pattern underwent GIANT analysis for cell type specificity and gene ontology.

An overview of the iSNP workflow can be seen in Figure 4-1.

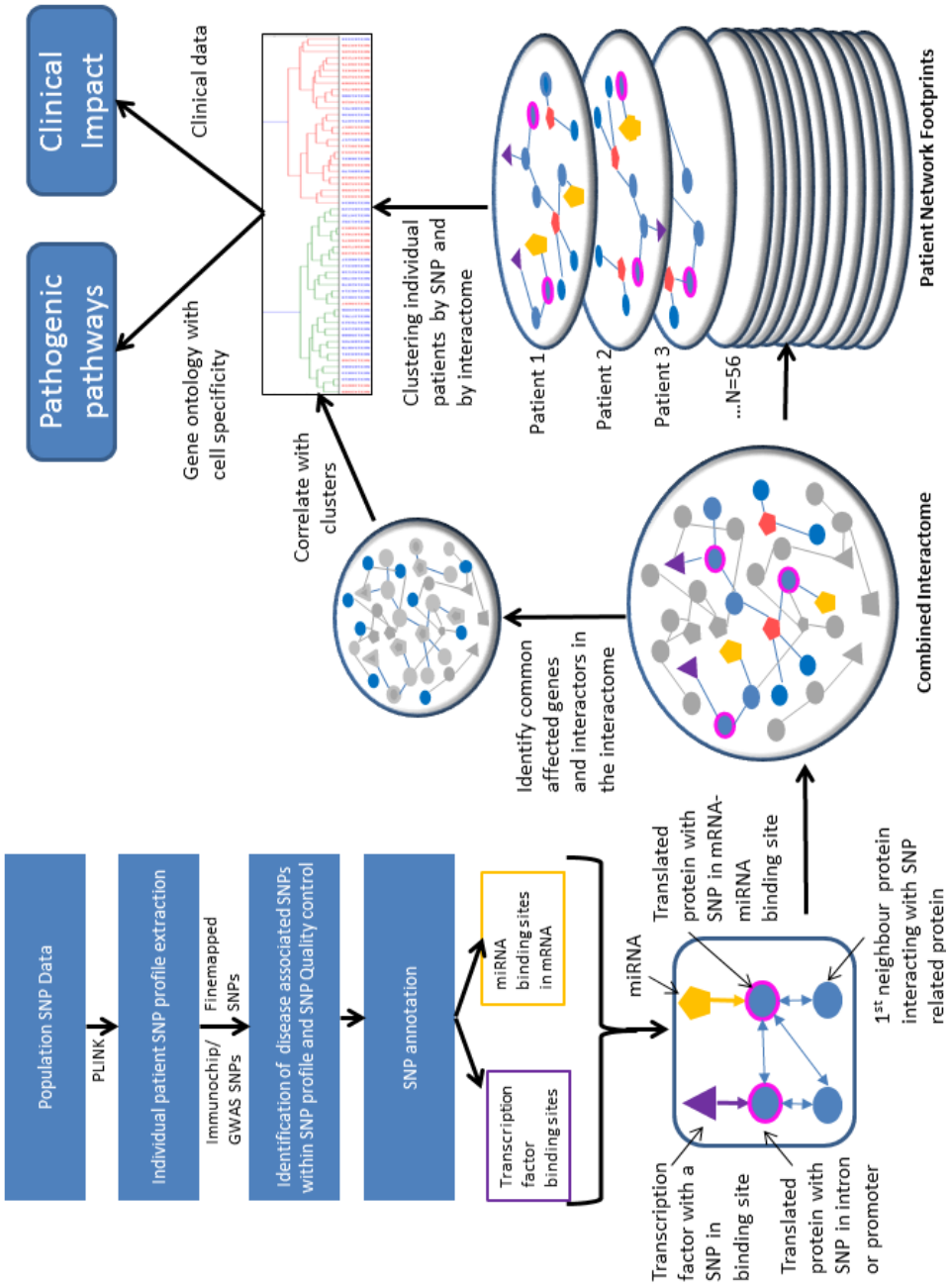


Figure 4-1 Overview of the Norwich cohort iSNP workflow from retrieval of the SNP data using PLINK (top left), through extraction of disease associated SNPs, identification of transcription factor or miRNA binding sites and first neighbour proteins to creation of the combined UC-ome with subsequent clustering and pathway analysis downstream.

4.5 Results

The summary demographics are in Table 4-1. The numbers requiring a thiopurine (e.g. azathioprine) was as expected, however we had very low numbers of patients requiring biologic therapy within the cohort. The disease extent in terms of Montreal classification was as expected (E1 – proctitis, E2 – left sided disease, E3 – pancolitis). Of the 385 SNPs 71 were represented in the IC Norwich cohort. The range of SNP burden was 19-41 (mean 28.5, median 28, mode 28).

Age Range	Sex (%)	Site of Disease (Montreal classification)	Management	Side effects
24-89y	Male n=34 (59%) Female n=24 (41%)	**E1** n=16/58 (28%) **E2** n= 20/58(34%) **E3** n= 22/58(38%)	5ASA n= 58 (100%) 5ASA only n = 30(52%) Azathioprine n=27 (47%) 6MP n= 5 (8%) Methotrexate n =5(8%) Tacrolimus n=1 (2%) IV Methylprednisolone n = 7 (12%) Infliximab n = 3(5%) Ciclosporin n =2 (3%) Surgery n = 2(3%)	Myelosuppression n=5 (8%) Abnormal Liver function tests n =4 (7%) Alopecia n= 2 (3%)

Table 4-1Summary Demographics for the UKIBDGC Norwich Cohort n=56

Given the low impact of the ELM results had on the UC interactome in chapter 2, the patient UC interactomes focused on transcription factor binding and miRNA binding sites. From the complete group we found two transcription factor binding sites, and 34 miRNA binding sites, 5 were in long non coding RNAs, 1 in a miRNA. The constructed network with the OmniPath interactors had 247 protein and 1297 edges (Figure 4-2). The network

consisted of one giant cluster and 9 single interactors which were not associated with the giant cluster and were excluded from future analysis. The giant network contained two large hubs with the highest degrees of centrality; NFKB1 and PRKCB. These hubs, not unsurprisingly formed the two large modules within the giant cluster. There were also 18 further SNP affected proteins that formed smaller modules within the giant cluster (Figure 4-3).

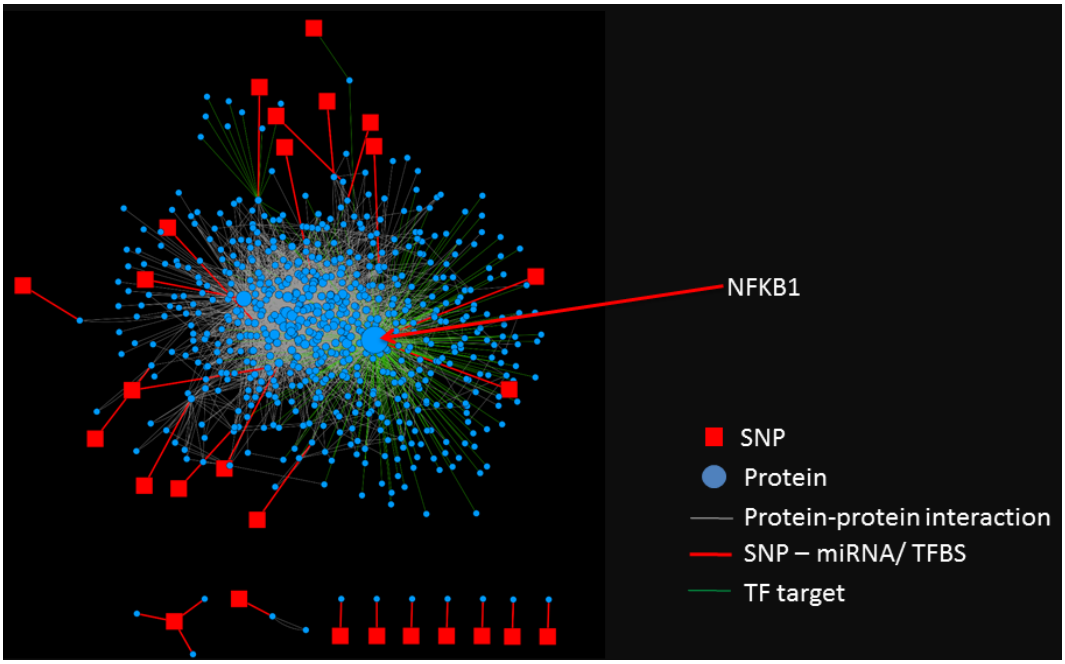


Figure 4-2 Diagramatic representation of the Norwich Patient Cohort UC interactome created from integrating each patients annotated SNP burden with Omnipath. The annotation of the SNPs focused on SNPs affecting transcription factor binding sites and miRNA binding sites. The diagram shows the SNPs as red boxes linked to the protein that they affect – such as NFKB1 highlighted by the red arrow. It also highlights the transcription factor targets and protein-protein interactions of the SNP affected proteins (green and grey lines respectively). There is clearly one large network (Giant cluster) and 9 SNPs which have no interaction with the larger network. These were removed from further analysis.

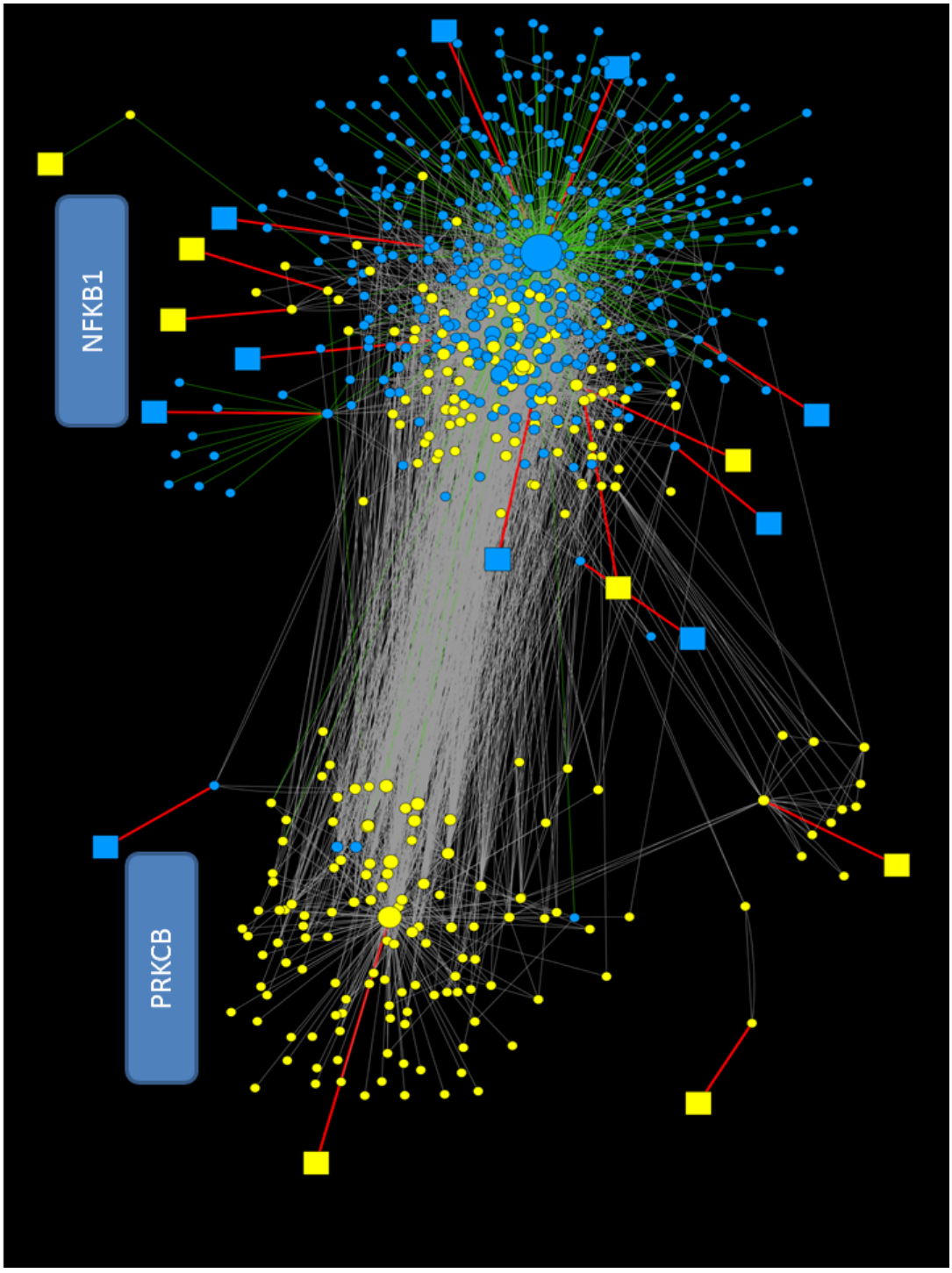


Figure 4-3 Modularisation of the Norwich Cohort UC interactome Giant cluster using a patient example. Using modularisation techniques, the giant cluster was separated into different modules, each important to the network. There were two large modules, NFKB1 and PRKCB and multiple smaller modules within the giant cluster. The smaller modules comprised connecting nodes to the two large modules. In this patient example, the yellow colouring denotes the parts of the network the patient has, therefore they are PRKCB+ NFKB1- but contain many of the connecting nodes to NFKB1. As before, the squares are the SNPs, the circles the proteins..

4.5.1 UC Patients cluster into one of four pathological footprints

Unsupervised clustering based on the Hamming distance between each patient dependant on their SNP burden identified that there were four distinct clusters or 'footprints' of patients (Figure 4-4). Cluster 1 encompassed both large modules of the giant cluster; NFKB1 and PRKCB with their binding targets (n=22). One patient was later excluded as they had a diagnosis of CD on the most recent histology. Cluster 2 was based around the NFKB1 node and binding targets (n=13). Cluster 3 was based around the NFKB1 and PRKCB binding targets and smaller modules, but not the major nodes themselves (n=14). Cluster 4 was the PRKCB node and binding targets (n=10) (Figure 4-5).

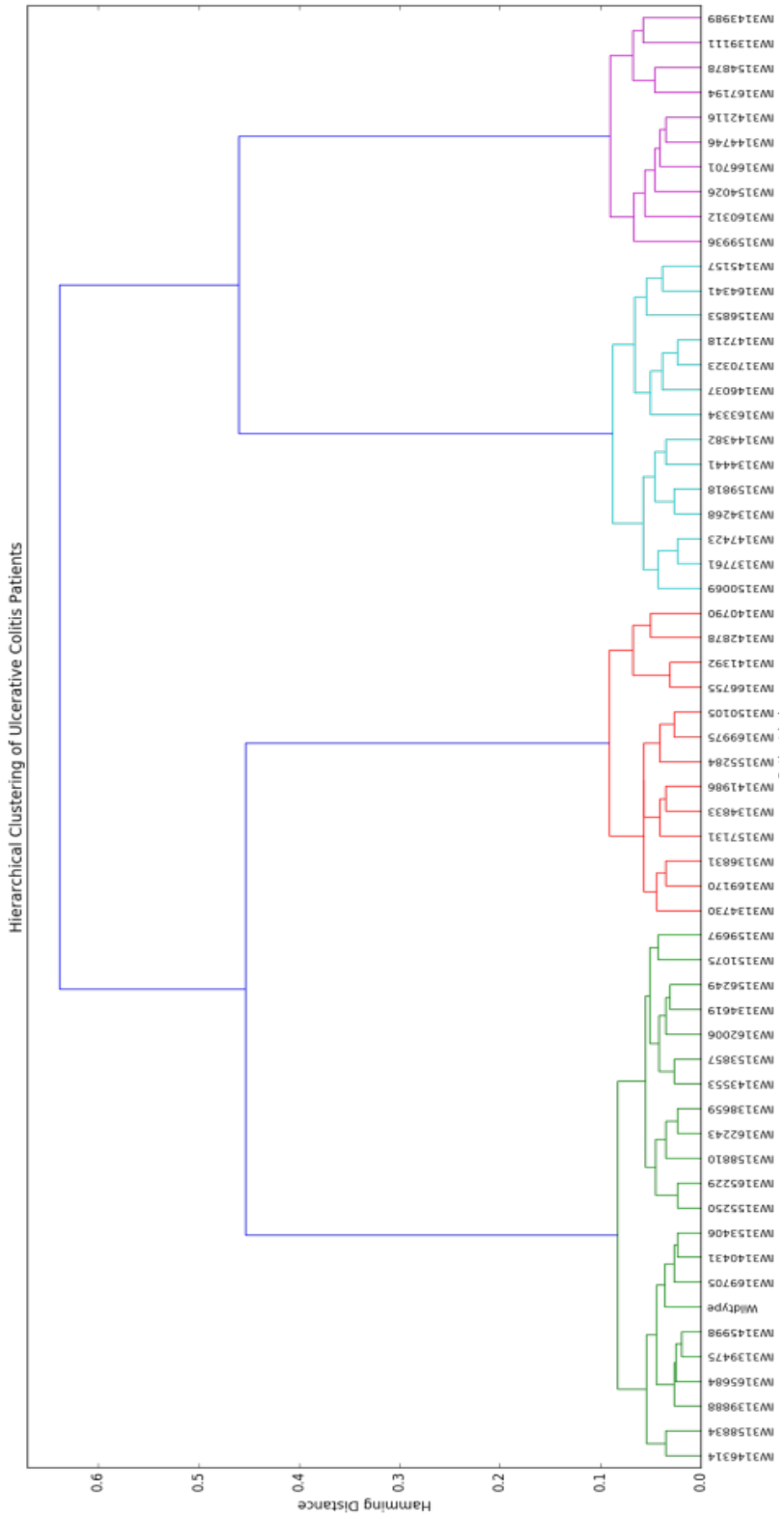


Figure 4-4 Unsupervised clustering based upon the Hamming distance between patients. Hamming distance calculates how similar each string of information – in this case the SNPs and first neighbours in one patient is to the next string of information (SNPs and first neighbours) to the next patient. With the patients, there were four clear cohorts (green, red, turquoise and purple).

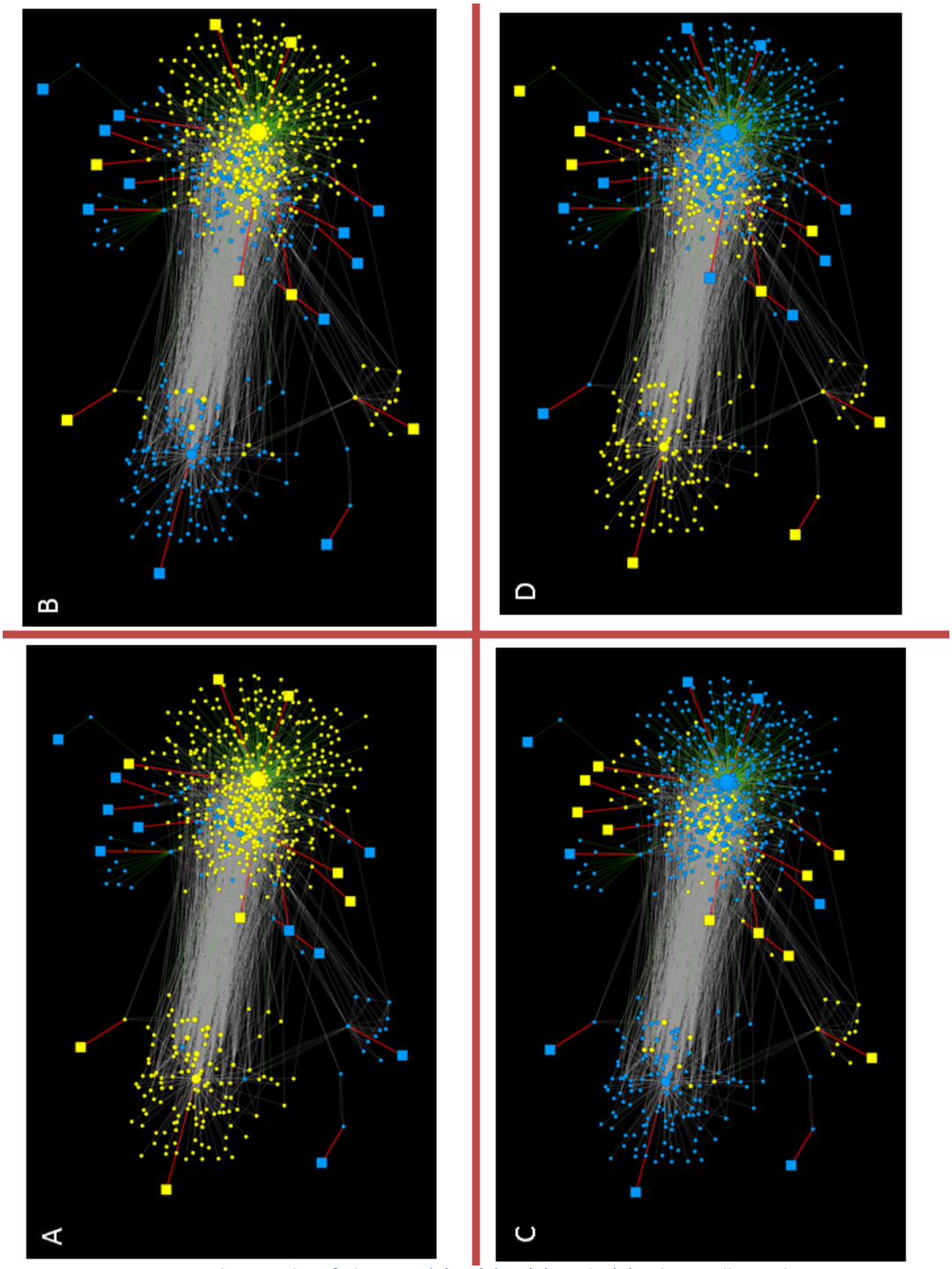


Figure 4-5 Patient network examples of Clusters A(1), B (2), C (3), and D(4), where yellow colouration identifies the nodes each particular patient has from within the interactome. Cluster 1 (A) is the PRKCB+NFKB1+ cohort, 2 (B) is PRCKB-NFKB1+, 3(C) is PRKCB-NFKB1- and cluster 4(D) is PRKCB+ NFKB1-.

Outside of the two large modules of NFKB1 and PRKCB, the clusters differed significantly with regard to the presence of SNPs affecting HDAC7, ZGPAT, C5orf66, MAML2 and DNMT3B with each cluster. In the Chi squared testing of % individual patients with individual proteins in their networks, Cluster 1 and 2, not unsurprisingly looked statistically similar for all proteins, except for ZGPAT, MAML2 and DMNT3B. Cluster 3 did not statistically differ from Cluster 1 in terms of non NFKB1 or PRKCB proteins, except for ZGPAT which appeared under-represented in cluster 1 (essentially Cluster 3 looked like Cluster 1 without the NFKB1 and PRKCB nodes). Cluster 4 (PRKCB node) and Cluster 2 (NFKB1) differed the most significantly in terms of non-major node SNP involvement withDNMT3B and HDAC7 being under-represented in Cluster 2 and C5orf66 being over represented (same as Cluster 1).

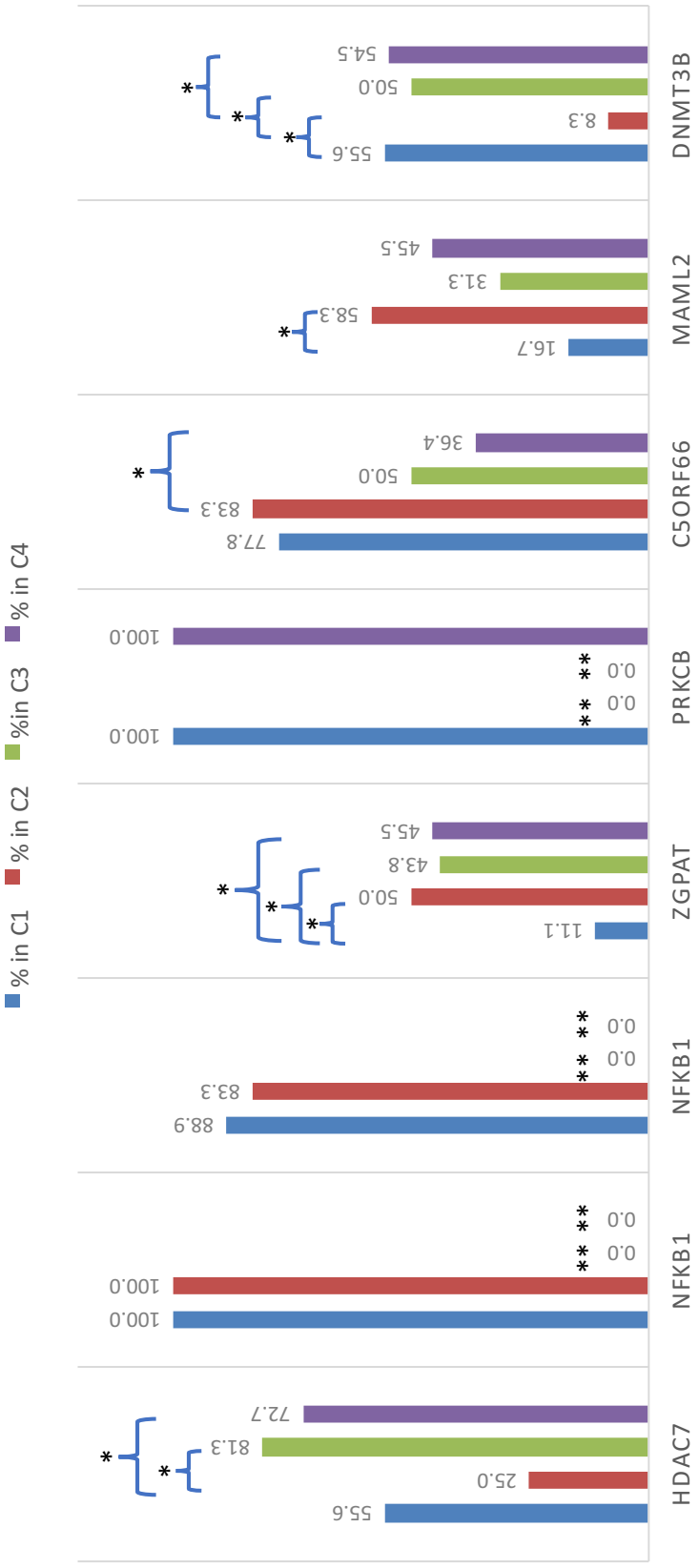


*Figure 4-6 Percentage of patients within each cluster with SNPs in specific proteins crucial to the network. There are 2 NFKB1 SNPs. Nonewithstanding the NFKB1-PRKCB status of the clusters, other SNPs such as HDAC7, ZGPAT, C5ORF66, MAML2 and DNM3TB also significantly differ between cohort.s Only significant differences are shown. * P= <0.05, **P= < 0.001 via Chi Squared Test.*

The clusters are partly a function of the SNP volume of the patients; patients within clusters 1 or 2 (NFKB1 containing) had significantly more SNPs and therefore interaction partners than patients within clusters 3 or 4, due to the large number of targets of NFKB1 and PRKCB. Cluster 3 contained the least number of SNPs and interacting proteins indicating there may be a minimum SNP burden to get the disease (

Figure 4-7). The majority of cluster 3 patients had a SNP in HDAC7 (86%), and/or ZNF831 (62%) or CCNY (56%).

Given the significant difference in SNP burden, further analysis of the unsupervised cluster was required in terms of gene ontology and pathway analysis, as well as a more in-depth analysis of the commonest convergent binding targets for each of the clusters.

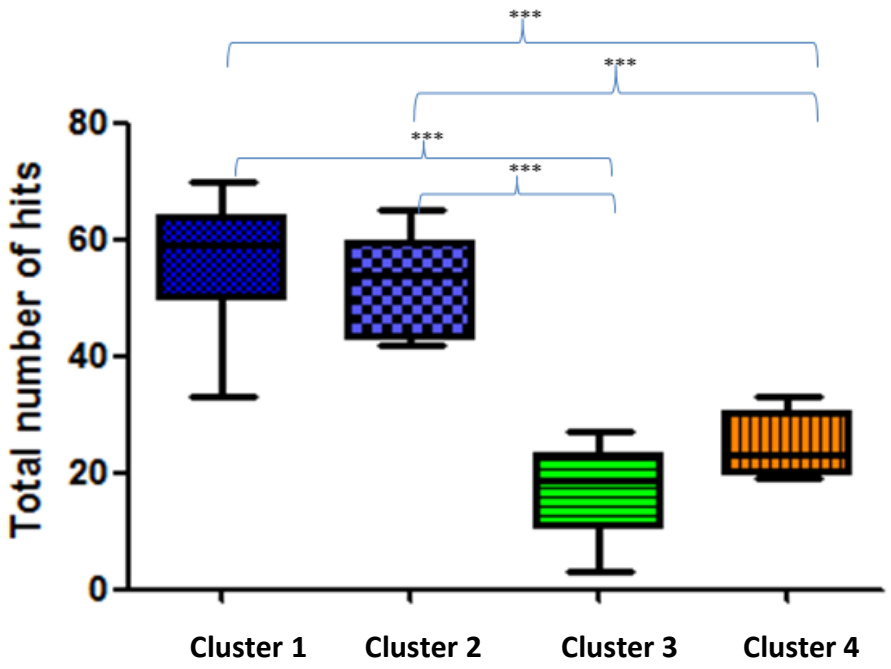


*Figure 4-7 Box-Whisker plot showing the number of SNPs (hits) per cluster, identifying that clusters 1 and 2 (NFKB1+ clusters) have a significantly higher SNP burden than clusters 3 and 4. *** = p=<0.001*

4.5.2 UC Footprints have overlapping pathway enrichment

Gene ontology and Pathway enrichment analysis using Panther (377) identified 61 commonly occurring significant pathways across all four clusters), with the apoptosis pathway featuring in all but 2 patients (Table 4.2). There is considerable commonality between the most frequently affected pathway enrichments between the clusters. Cluster 4 pathways predictably could be subsumed into cluster 1, however the frequency with which each pathway occurs within the clusters are subtly different as seen with analysis of the top ten pathways for each cluster (Figure 4.8). This highlights that although cluster 3 is significantly different to the other clusters, they share pathogenic pathways.

Cluster 1	%	Cluster 2	%	Cluster 3	%	Cluster 4	%
Apoptosis signalling pathway	100	Apoptosis signalling pathway	100	Endothelin signalling	100	Apoptosis signalling pathway	90.909
Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway	100	Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway	100	Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway	100	Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway	90.909
T cell activation	94.444	PDGF signalling	90.909	Apoptosis signalling pathway	93.333	Plasminogen activating cascade	81.818
CCKR signalling map	88.889	Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway	90.909	CCKR signalling map	93.333	B cell activation	81.818
VEGF signalling	88.889	wnt signalling pathway	81.818	angiogenesis	93.333	T cell activation	81.818
Endothelin signalling	88.889	CCKR signalling map	81.818	Plasminogen activating cascade	93.333	Toll receptor signalling pathway	81.818
Plasminogen activating cascade	88.889	VEGF signalling	81.818	Inflammation mediated by chemokine and cytokine signaling pathway	93.333	CCKR signalling map	72.727
Inflammation mediated by chemokine and cytokine signaling pathway	88.889	Endothelin signalling	81.818	wnt signalling pathway	86.667	de novo purine biosynthesis	72.727
Toll receptor signalling pathway	88.889	angiogenesis	81.818	fas signalling pathway	86.667	Endothelin signalling	72.727
de novo purine biosynthesis	83.333	alzheimer disease-presenilin pathway	81.818	de novo purine biosynthesis	86.667	Ras Pathway	72.727
angiogenesis	83.333	Plasminogen activating cascade	81.818	VEGF signalling	86.667	EGF receptor signalling pathway	63.636
blood coagulation	83.333	Inflammation mediated by chemokine and cytokine signaling pathway	81.818	p53 pathway	86.667	VEGF signalling	63.636
Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway	83.333	blood coagulation	81.818	alzheimer disease-presenilin pathway	86.667	angiogenesis	63.636
EGF receptor signalling pathway	77.778	Toll receptor signalling pathway	81.818	T cell activation	86.667	alzheimer disease-presenilin pathway	63.636
PDGF signalling	77.778	Gonadotropin releasing hormone receptor pathway	72.727	blood coagulation	86.667	Inflammation mediated by chemokine and cytokine signaling pathway	63.636
alzheimer disease-presenilin pathway	77.778	de novo purine biosynthesis	72.727	Toll receptor signalling pathway	86.667	blood coagulation	63.636
B cell activation	77.778	interleukin signalling pathway	72.727	Ras Pathway	86.667	Gonadotropin releasing hormone receptor pathway	54.545
FGF singalling pathway	77.778	p53 pathway	72.727	TGF-beta singalling pathway	86.667	huntingdon disease	54.545
Gonadotropin releasing hormone receptor pathway	72.222	integrin signalling pathway	72.727	p38 MAPK pathway	86.667	PDGF signalling	54.545
interleukin signalling pathway	72.222	B cell activation	72.727	huntingdon disease	80	p53 pathway	54.545
integrin signalling pathway	72.222	T cell activation	72.727	interleukin signalling pathway	80	integrin signalling pathway	54.545
Ras Pathway	72.222	Alzheimer disease-amyloid secretase pathway	72.727	integrin signalling pathway	80	Alzheimer disease-amyloid secretase pathway	54.545
Alzheimer disease-amyloid secretase pathway	72.222	FGF singalling pathway	72.727	Alzheimer disease-amyloid secretase pathway	80	FGF singalling pathway	54.545
wnt signalling pathway	66.667	EGF receptor signalling pathway	63.636	DNA replication	80	N-acetylglucosamine metabolism	54.545
huntingdon disease	66.667	Ras Pathway	63.636	Parkinson disease	80	DNA replication	54.545
p53 pathway	66.667	fas signalling pathway	54.545	EGF receptor signalling pathway	73.333	p38 MAPK pathway	54.545
JAK/STAT signalling	61.111	TGF-beta singalling pathway	54.545	PDGF signalling	73.333	wnt signalling pathway	45.455
TGF-beta singalling pathway	61.111	Thyrotropin-releasing hormone receptor signaling pathway	54.545	FGF singalling pathway	73.333	fas signalling pathway	45.455
Thyrotropin-releasing hormone receptor signaling pathway	61.111	Histamine H1 receptor mediated signaling pathway	54.545	N-acetylglucosamine metabolism	73.333	Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway	45.455
oxytocin recetor mediatd signaling pathway	61.111	5HT2 type receptor mediated signalling pathway	54.545	oxidative stress response	73.333	JAK/STAT signalling	45.455
p38 MAPK pathway	61.111	huntingdon disease	45.455	O-antigen biosynthesis	73.333	sulphate assimilation	45.455
muscarinic acetylcholine receptor 1 and 3 signaling pathway	61.111	cadherin signalling pathway	45.455	Gonadotropin releasing hormone receptor pathway	66.667	TGF-beta singalling pathway	45.455
metabotropic glutamate receptor group I pathway	61.111	JAK/STAT signalling	45.455	B cell activation	66.667	Parkinson disease	45.455

Table 4-2 Panther outputs for gene ontology for each cluster and the percentage of patients containing that pathway within each cluster.

Cluster 1	%	Cluster 2	%	Cluster 3	%	Cluster 4	%
fas signalling pathway	55.556	Parkinson disease	45.455	Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway	66.667	O-antigen biosynthesis	45.455
cell cycle	55.556	Insulin/IGF pathway-protein kinase B signaling cascade	45.455	insulin/IGF MAPKK/MAPK cascade	66.667	cadherin signalling pathway	36.364
Flavin biosynthesis	55.556	oxytocin recetor mediatd signaling pathway	45.455	JAK/STAT signalling	60	interleukin signalling pathway	36.364
Histamine H1 receptor mediated signaling pathway	55.556	p38 MAPK pathway	45.455	Opioid proopiomelanocortin pathway	60	PI3kinase pathway	36.364
5HT2 type receptor mediated signalling pathway	55.556	muscarinic acetylcholine receptor 1 and 3 signaling pathway	45.455	5HT1 type receptor mediated signaling pathway	60	Interferon gamma signalling pathway	36.364
N-acetylglucosamine metabolism	50	metabotropic glutamate receptor group I pathway	45.455	Beta2 adrenergic receptor signaling pathway	60	Insulin/IGF pathway-protein kinase B signaling cascade	36.364
oxidative stress response	50	sulphate assimilation	36.364	Opioid proenkephalin pathway	60	Thyrotropin-releasing hormone receptor signaling pathway	36.364
insulin/IGF MAPKK/MAPK cascade	50	cell cycle	36.364	Enkephalin release	60	Histamine H1 receptor mediated signaling pathway	36.364
PI3kinase pathway	44.444	Opioid proopiomelanocortin pathway	36.364	Dopamine receptor mediated signaling pathway	60	muscarinic acetylcholine receptor 1 and 3 signaling pathway	36.364
DNA replication	44.444	5HT1 type receptor mediated signaling pathway	36.364	GABA-B_receptor_II_signaling	60	5HT2 type receptor mediated signalling pathway	36.364
Interferon gamma signalling pathway	44.444	Beta2 adrenergic receptor signaling pathway	36.364	Beta1 adrenergic receptor signaling pathway	60	cell cycle	27.273
Parkinson disease	44.444	N-acetylglucosamine metabolism	36.364	Flavin biosynthesis	60	5HT1 type receptor mediated signaling pathway	27.273
O-antigen biosynthesis	44.444	Opioid proenkephalin pathway	36.364	Beta3 adrenergic receptr signalling pathway	60	Beta2 adrenergic receptor signaling pathway	27.273
sulphate assimilation	38.889	Enkephalin release	36.364	Opioid prodynorphin pathway	60	Opioid proenkephalin pathway	27.273
Opioid proopiomelanocortin pathway	38.889	Dopamine receptor mediated signaling pathway	36.364	5HT4 type receptor signaling pathway	60	Enkephalin release	27.273
5HT1 type receptor mediated signaling pathway	38.889	GABA-B_receptor_II_signaling	36.364	cell cycle	53.333	Flavin biosynthesis	27.273
Beta2 adrenergic receptor signaling pathway	38.889	DNA replication	36.364	Interferon gamma signalling pathway	53.333	oxidative stress response	27.273
Opioid proenkephalin pathway	38.889	Beta1 adrenergic receptor signaling pathway	36.364	Insulin/IGF pathway-protein kinase B signaling cascade	53.333	oxytocin recetor mediatd signaling pathway	27.273
Histamine H2 receptor mediated signaling pathway	38.889	Interferon gamma signalling pathway	36.364	PI3kinase pathway	46.667	insulin/IGF MAPKK/MAPK cascade	27.273
Enkephalin release	38.889	Beta3 adrenergic receptr signalling pathway	36.364	Thyrotropin-releasing hormone receptor signaling pathway	46.667	metabotropic glutamate receptor group I pathway	27.273
Dopamine receptor mediated signaling pathway	38.889	Opioid prodynorphin pathway	36.364	oxytocin recetor mediatd signaling pathway	46.667	Opioid proopiomelanocortin pathway	18.182
GABA-B_receptor_II_signaling	38.889	5HT4 type receptor signaling pathway	36.364	metabotropic glutamate receptor group I pathway	46.667	Dopamine receptor mediated signaling pathway	18.182
Beta1 adrenergic receptor signaling pathway	38.889	O-antigen biosynthesis	36.364	sulphate assimilation	40	GABA-B_receptor_II_signaling	18.182
Beta3 adrenergic receptr signalling pathway	38.889	PI3kinase pathway	27.273	Histamine H2 receptor mediated signaling pathway	40	Beta1 adrenergic receptor signaling pathway	18.182
Opioid prodynorphin pathway	33.333	Histamine H2 receptor mediated signaling pathway	27.273	muscarinic acetylcholine receptor 1 and 3 signaling pathway	40	Beta3 adrenergic receptr signalling pathway	18.182
5HT4 type receptor signaling pathway	33.333	Flavin biosynthesis	27.273	5HT2 type receptor mediated signalling pathway	40	Opioid prodynorphin pathway	18.182
Insulin/IGF pathway-protein kinase B signaling cascade	27.778	oxidative stress response	27.273	cadherin signalling pathway	33.333	5HT4 type receptor signaling pathway	18.182
cadherin signalling pathway	22.222	insulin/IGF MAPKK/MAPK cascade	27.273	Histamine H1 receptor mediated signaling pathway	33.333	Histamine H2 receptor mediated signaling pathway	9.0909

Table 4-3 Continued; Panther outputs for gene ontology for each cluster and the percentage of patients containing that pathway within each cluster.

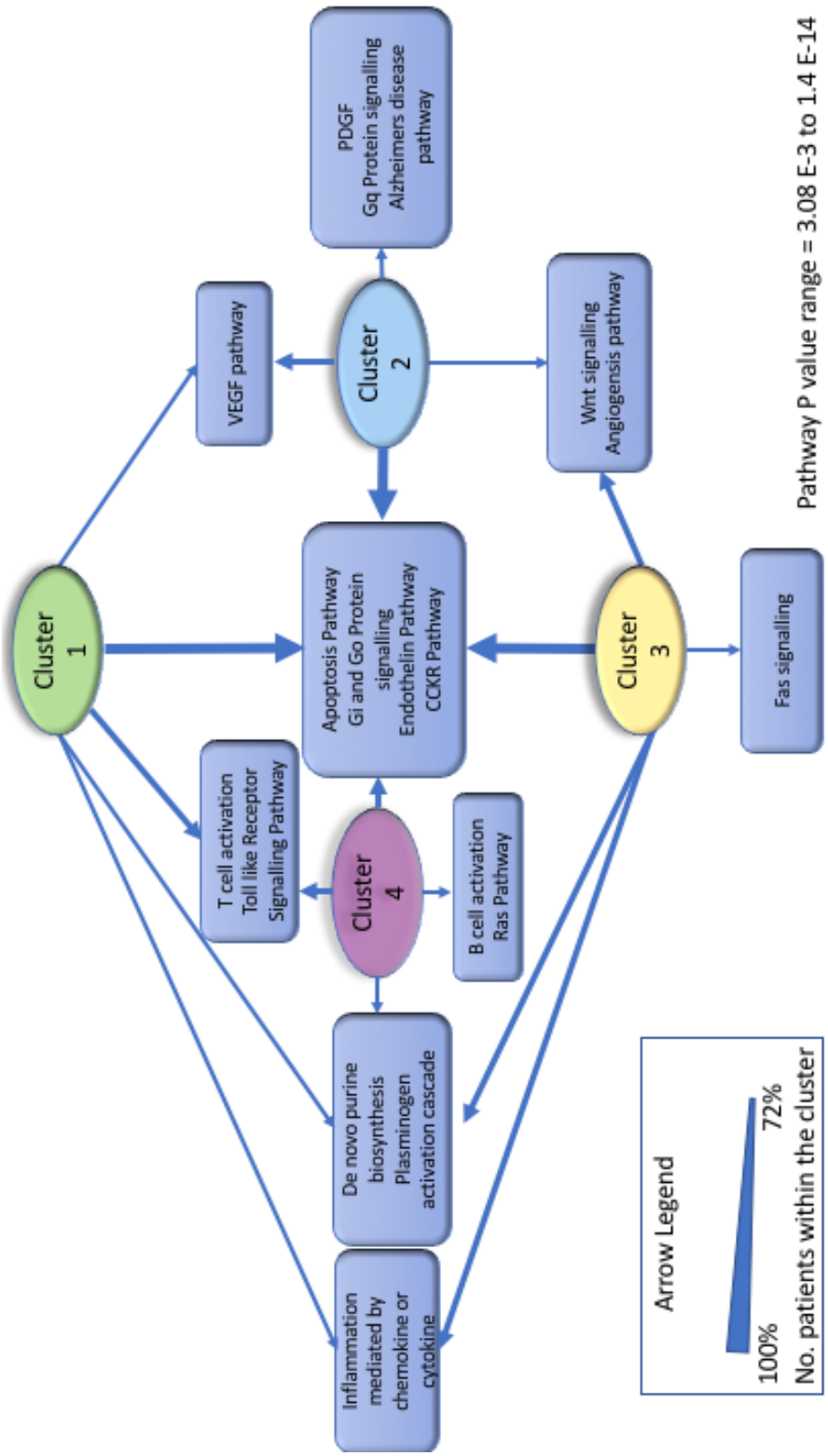


Figure 4-8 Diagram of the top 10 enriched gene ontology pathways across the clusters with an indicator of prevalence of each pathway within the cluster given by arrow thickness highlighting the commonality between the clusters, but also the subtle differences.

The top ten pathways across all 4 clusters highlighted a total of 18 pathogenic pathways. These pathways have components identified as being involved in the aetiology of colitis – either mouse models or in human studies, but not necessarily from a genetic perspective as shown in the table below (Table 4-3).

Pathway	References to UC
Apoptosis	Lv B 2014, Li L 2015, Seidelin JB 2015, Zhao 2012, Wu 2016
Gi, Go signalling	Hornquist 1997, Ohman 2000
Endothelin	Nakamura 2004, McCartney 2002, Letizia 1998, Padol 2000, Murch 1992, Rachmilewitz 1992
CCKR Pathway	Zietek 2016, Moran 2008
VEGF Pathway	Tolstanova 2009, Frysz-Naglak 2011, Cromer 2013, Ramadass 2016, Deng 2013
PDGF Pathway	Tohoku 2001, Lawrance 2001, Deng 2013
Gq signalling	Watanabe 2016
Wnt signalling	Cosín-Roger 2013, Kini 2015, Soubh 2015. NB colitis associated cancer
Alzheimers disease pathway	Zou 2012
Angiogenesis pathway	Bakirtzi 2016, Zak 2008 (Nb non VEGF, non CAC)
Fas signalling	Seidelin 2013, Seidelin 2015
B cell activation	Yeung 2000, Wang 2017, Uo 2013
Ras Pathway	Lyda 2000, Takahashi 2016, Burmer 1990 NB colitis associated cancer
T cell activation	Lord 2015, Mann 2014, Dahlen 2013, Shih 2008
Toll Like Receptor Pathway	Günaltay 2014, Carvalho 2012, Torok 2004, Dong 2012, Grabig 2006, Singh 2005
De novo purine biosynthesis	Chiaro 2017, Kurtz 2014
Plasminogen activation cascade	Munakata 2015, Kurose 1992, Kume 2007
Inflammation mediated by chemokine or cytokine	26139 citations...

Table 4-4 Pathways identified by Panther from UC patients' clusters, with current literature references to their involvement in colitis.

This gives confidence to the genetic underpinning we are trying to identify in UC, although the lack of novel pathogenic pathways for UC is striking.

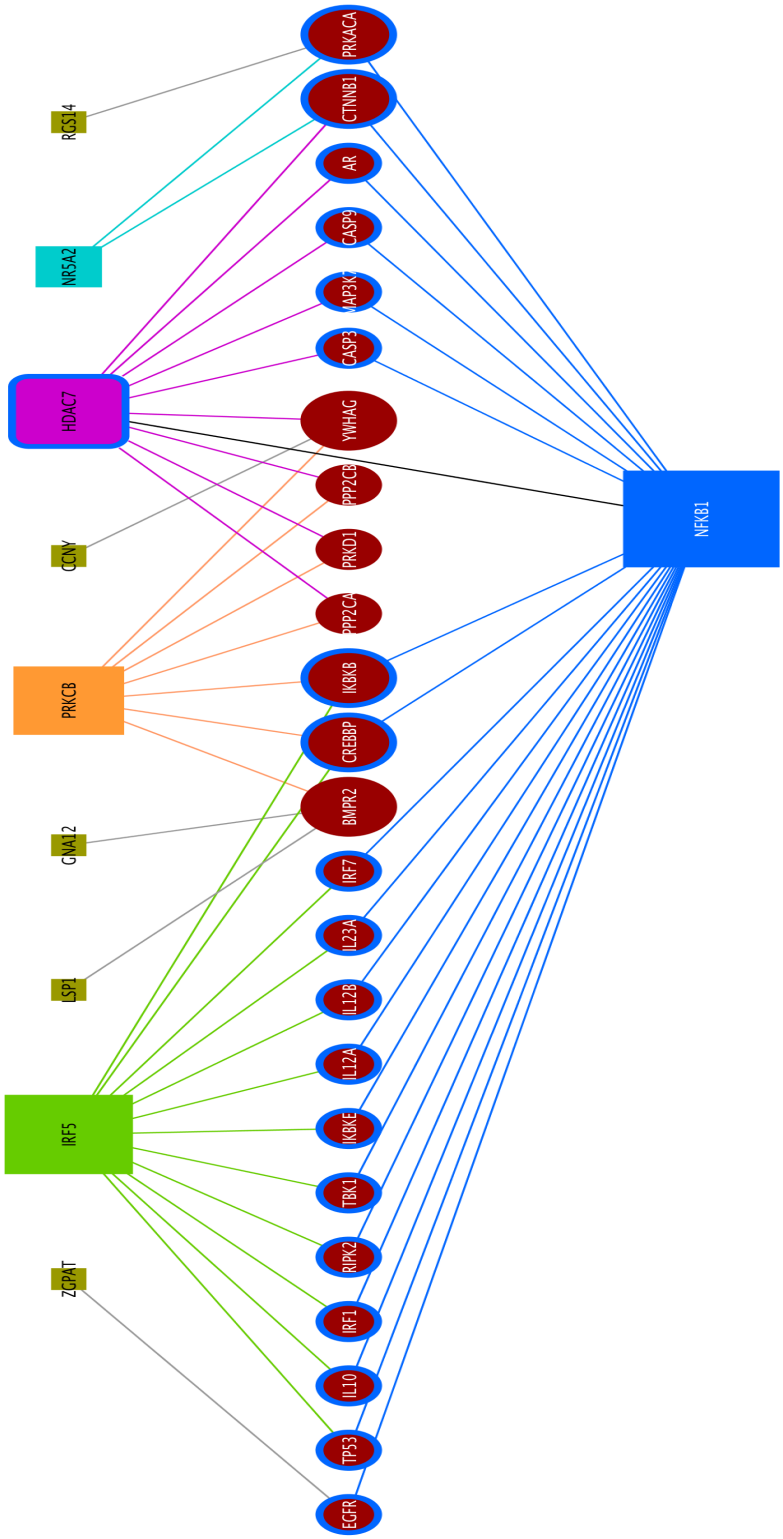
4.5.3 UC Patient SNPs converge to twenty-four first neighbour proteins

Gene ontology identified significant overlap between the clusters in terms of pathogenic pathways. To explore this further and identify SNP epistasis and convergent protein-protein interactions within the first neighbours, we examined the top 66% commonest affected first neighbours. The SNPs affected proteins (e.g. proteins whose regulation has been putatively affected by a SNP) that led to the convergent first neighbours within our cohort were NFKB1, PRKCB, IRF5, HDAC7, NR5A2, LSP1, CCNY, RGS14, GNA12 and ZGPAT. Analysing these against their minor allele frequency (Table 4-5), there was a wide variety of expected frequencies from 31% to 88%, with most below 50% highlighting the importance of the convergence of effects.

Globally, these 10 SNP affected proteins had between 1 and 20 protein-protein interactions with 24 first neighbour proteins (Figure 4-8). These first neighbour proteins were chosen from the highest frequency of cross-hits with other SNP affected proteins, therefore each first neighbour protein had ≥2 SNP affected proteins. Figure 4-9 highlights the convergent first neighbours in the context of the unsupervised clusters. Cluster 3 did not have any first neighbours with >2 hits, indicating a different mechanism of disease. Cluster 1 and 2 look very similar given the significant impact NFKB1 has on the network.

	IC SNP	Fine mapped SNP	IC MAF	Finemapped MAF
LSP1	rs11041476	rs907611	NA	0.318
NFKB1	rs1598859	rs3774959	NA	0.332
NFKB1	rs3774937	rs3774959	0.332	0.332
PRKCB1	rs7404095	NA	0.576	NA
IRF5	rs4728142	NA	0.437	NA
HDAC7	rs11168249	NA	0.461	NA
CCNY	NA	rs12261843	NA	*0.374*
RGS14	NA	rs4976646	NA	*0.404*
GNA12	rs1182188	rs798502	NA	0.7
ZGPAT	rs6062504	NA	0.702	NA
NR5A2	rs2816958	NA	0.887	NA

Table 4-5 Commonest SNP affected proteins in the Norwich UC cohort. Highlighted in yellow are the corresponding SNPs that appeared on Immunochip (IC) and in the patient cohort. Blue highlights indicate a finemapped SNP on Immunochip that was not in the Norwich cohort. Italics denote a minor allele frequency for the risk allele obtained from 1000 genomes, as the SNP was a finemapped UC SNP from the Broad Institute that was on Immunochip but identified as another disease susceptibility SNP.



.

Figure 4-9 Diagram highlighting the 10 commonest SNPs in the entire Norwich UC cohort with their convergent binding partners. The size of the square denotes the commonality, blue outlines indicate a first neighbour to NFKB1

4.5.3.1 GIANT Enrichment analysis of the convergent nodes identifies pathogenic pathways

To identify the role of the convergent proteins within the network, to further examine and test the hypothesis that there is a role of synergism between the SNPs that creates a SNP burden greater than the sum of the individual SNPs themselves, we undertook an enrichment analysis of the SNP affected proteins and convergent proteins in GIANT. In order to focus on the convergent proteins within the clusters, each patient footprint from within the cluster underwent subgroup analysis to identify if there were differences within the clusters which would bias or sway cluster-wide GIANT analysis. Cluster 3 was excluded from this analysis as it did not have convergent nodes. An example subgroup within clusters can be seen in Cluster 1, which has networks which contain NFKB1 and PRCKB hub nodes, but within the cluster are patients who have IRF5 SNPs and those who do not. This makes a significant difference to pathway enrichment as can be seen in comparing Table 4-6 with Table 4-7 .

The enrichment was undertaken across 4 tissue/cell types: B lymphocytes, Colon, Dendritic cells and T lymphocytes. The tables (4-6 to 4-11) indicate the significance of each pathway from the patient gene set within each cell or tissue type. Of note, regardless of the cell types, intracellular infections (viral infections, Toxoplasmosis, Legionellosis, TB, Leishmaniasis, Pertussis) featured in all the pathway enrichments (p=1.07E-11 to p=0.01). The pathways with the lower (but still significant) pathway enrichments came from smaller networks e.g. IW3134619: HDAC7, NFKB1, PRKCB and their first neighbours. Also enriched were pathways involved with receptor signalling associated with extracellular infective ligand signalling, adaptive, and innate immune system processes, the NFKB1 pathway, STAT4 regulation, cancer processes. Interestingly, when NFKB1 was not present (Cluster 4), cellular processes such as Hippo signalling, TGF beta signalling, cell development, apoptosis and cellular junctions rose to the fore. There was no specific cell type of those tested that consistently showed the most significant enrichment across the clusters.

Table 4-6 GIANT Enrichment analysis of Cluster 1 example with LSP1, HDAC7, NFKB1, IRF5 and PRKCB SNP affected proteins

B lymphocyte		Colon		Dendritic Cell		T Lymphocyte	
Pathways/Processes/Diseases	P-Value (FDR Corrected)	Pathways/Processes/Diseases	P-Value (FDR Corrected)	Pathways/Processes/Diseases	P-Value (FDR Corrected)	Pathways/Processes/Diseases	P-Value (FDR Corrected)
KEGG-Pathway-hsa04620: Toll-like receptor signaling pathway - Homo sapiens (human)	1.07E-11	KEGG-Pathway-hsa04622: RIG-I-like receptor signaling pathway - Homo sapiens (human)	1.19E-10	KEGG-Pathway-hsa05168: Herpes simplex infection - Homo sapiens (human)	8.07E-09	KEGG-Pathway-hsa05168: Herpes simplex infection - Homo sapiens (human)	7.19E-13
KEGG-Pathway-hsa05162: Measles - Homo sapiens (human)	3.15E-11	KEGG-Pathway-hsa05160: Hepatitis C - Homo sapiens (human)	3.15E-10	KEGG-Pathway-hsa05160: Hepatitis C - Homo sapiens (human)	9.07E-09	KEGG-Pathway-hsa04620: Toll-like receptor signaling pathway - Homo sapiens (human)	1.30E-11
KEGG-Pathway-hsa05160: Hepatitis C - Homo sapiens (human)	3.15E-11	KEGG-Pathway-hsa05168: Herpes simplex infection - Homo sapiens (human)	3.19E-10	KEGG-Pathway-hsa05161: Hepatitis B - Homo sapiens (human)	1.59E-08	KEGG-Pathway-hsa05134: Legionellosis - Homo sapiens (human)	1.42E-11
KEGG-Pathway-hsa05145: Toxoplasmosis - Homo sapiens (human)	3.15E-11	KEGG-Pathway-hsa04620: Toll-like receptor signaling pathway - Homo sapiens (human)	1.40E-09	GO-BP-0001817:regulation of cytokine production	1.62E-08	KEGG-Pathway-hsa04064: NF-kappa B signaling pathway - Homo sapiens (human)	2.34E-11
KEGG-Pathway-hsa05161: Hepatitis B - Homo sapiens (human)	3.15E-11	KEGG-Pathway-hsa05161: Hepatitis B - Homo sapiens (human)	1.53E-09	KEGG-Pathway-hsa04622: RIG-I-like receptor signaling pathway - Homo sapiens (human)	1.74E-08	KEGG-Pathway-hsa05164: Influenza A - Homo sapiens (human)	2.96E-11
KEGG-Pathway-hsa05164: Influenza A - Homo sapiens (human)	3.47E-11	KEGG-Pathway-hsa05145: Toxoplasmosis - Homo sapiens (human)	4.81E-09	GO-BP-0001816:cytokine production	2.08E-08	KEGG-Pathway-hsa05162: Measles - Homo sapiens (human)	3.15E-11
KEGG-Pathway-hsa05142: Chagas disease (American trypanosomiasis) - Homo sapiens (human)	3.54E-11	KEGG-Pathway-hsa05152: Tuberculosis - Homo sapiens (human)	8.37E-09	KEGG-Pathway-hsa05133: Pertussis - Homo sapiens (human)	2.20E-08	KEGG-Pathway-hsa05160: Hepatitis C - Homo sapiens (human)	3.15E-11
KEGG-Pathway-hsa05169: Epstein-Barr virus infection - Homo sapiens (human)	3.61E-11	KEGG-Pathway-hsa05164: Influenza A - Homo sapiens (human)	8.41E-09	KEGG-Pathway-hsa05145: Toxoplasmosis - Homo sapiens (human)	3.36E-08	KEGG-Pathway-hsa05145: Toxoplasmosis - Homo sapiens (human)	3.15E-11
KEGG-Pathway-hsa05168: Herpes simplex infection - Homo sapiens (human)	3.72E-11	KEGG-Pathway-hsa05142: Chagas disease (American trypanosomiasis) - Homo sapiens (human)	2.23E-08	KEGG-Pathway-hsa05164: Influenza A - Homo sapiens (human)	4.04E-08	KEGG-Pathway-hsa04622: RIG-I-like receptor signaling pathway - Homo sapiens (human)	3.24E-11
KEGG-Pathway-hsa04622: RIG-I-like receptor signaling pathway - Homo sapiens (human)	3.80E-11	KEGG-Pathway-hsa05215: Prostate cancer - Homo sapiens (human)	1.71E-07	KEGG-Pathway-hsa05152: Tuberculosis - Homo sapiens (human)	4.05E-08	GO-BP-0007249:I-kappaB kinase/NF-kappaB signaling	3.29E-11
KEGG-Pathway-hsa05134: Legionellosis - Homo sapiens (human)	1.00E-10	KEGG-Pathway-hsa05162: Measles - Homo sapiens (human)	1.78E-07	KEGG-Pathway-hsa05215: Prostate cancer - Homo sapiens (human)	5.66E-08	KEGG-Pathway-hsa05169: Epstein-Barr virus infection - Homo sapiens (human)	3.54E-11
KEGG-Pathway-hsa04064: NF-kappa B signaling pathway - Homo sapiens (human)	2.34E-10	GO-BP-0051091:positive regulation of sequence-specific DNA binding transcription factor activity	7.26E-07	KEGG-Pathway-hsa05162: Measles - Homo sapiens (human)	5.21E-08	KEGG-Pathway-hsa05161: Hepatitis B - Homo sapiens (human)	3.72E-11
KEGG-Pathway-hsa05152: Tuberculosis - Homo sapiens (human)	4.91E-10	GO-BP-0050776:regulation of immune response	9.51E-07	KEGG-Pathway-hsa05142: Chagas disease (American trypanosomiasis) - Homo sapiens (human)	1.68E-07	GO-BP-0043122:regulation of I-kappaB kinase/NF-kappaB signaling	4.81E-11
KEGG-Pathway-hsa05140: Leishmaniasis - Homo sapiens (human)	1.23E-09	KEGG-Pathway-hsa05133: Pertussis - Homo sapiens (human)	1.10E-06	KEGG-Pathway-hsa04620: Toll-like receptor signaling pathway - Homo sapiens (human)	1.82E-07	KEGG-Pathway-hsa05133: Pertussis - Homo sapiens (human)	5.18E-11
GO-BP-0043122:regulation of I-kappaB kinase/NF-kappaB signaling	1.27E-09	GO-BP-0006955:immune response	1.12E-06	GO-BP-0050776:regulation of immune response	2.66E-07	KEGG-Pathway-hsa05152: Tuberculosis - Homo sapiens (human)	5.81E-11
KEGG-Pathway-hsa04668: TNF signaling pathway - Homo sapiens (human)	1.27E-09	GO-BP-0097190:apoptotic signaling pathway	1.39E-06	GO-BP-0006955:immune response	2.98E-07	KEGG-Pathway-hsa04621: NOD-like receptor signaling pathway - Homo sapiens (human)	1.18E-10
KEGG-Pathway-hsa05133: Pertussis - Homo sapiens (human)	1.31E-09	GO-BP-0006952:defense response	1.66E-06	GO-BP-0001819:positive regulation of cytokine production	3.09E-07	KEGG-Pathway-hsa05142: Chagas disease (American trypanosomiasis) - Homo sapiens (human)	8.64E-10
GO-BP-0007249:I-kappaB kinase/NF-kappaB signaling	1.38E-09	GO-BP-0042504:tyrosine phosphorylation of Stat4 protein	2.78E-06	KEGG-Pathway-hsa05169: Epstein-Barr virus infection - Homo sapiens (human)	1.37E-06	KEGG-Pathway-hsa04668: TNF signaling pathway - Homo sapiens (human)	1.36E-09
KEGG-Pathway-hsa04210: Apoptosis - Homo sapiens (human)	3.43E-09	GO-BP-0042520:positive regulation of tyrosine phosphorylation of Stat4 protein	2.91E-06	GO-BP-0002819:regulation of adaptive immune response	2.15E-06	KEGG-Pathway-hsa05140: Leishmaniasis - Homo sapiens (human)	1.44E-09
KEGG-Pathway-hsa04621: NOD-like receptor signaling pathway - Homo sapiens (human)	3.81E-09	GO-BP-0042519:regulation of tyrosine phosphorylation of Stat4 protein	3.04E-06	GO-BP-0051132:NK T cell activation	2.20E-06	KEGG-Pathway-hsa04210: Apoptosis - Homo sapiens (human)	4.02E-09
KEGG-Pathway-hsa05321: Inflammatory bowel disease (IBD) - Homo sapiens (human)	1.77E-08	GO-BP-0002819:regulation of adaptive immune response	3.04E-06	GO-BP-0051133:regulation of NK T cell activation	2.09E-06	KEGG-Pathway-hsa04623: Cytosolic DNA-sensing pathway - Homo sapiens (human)	1.14E-08
KEGG-Pathway-hsa05203: Viral carcinogenesis - Homo sapiens (human)	2.43E-08	KEGG-Pathway-hsa05169: Epstein-Barr virus infection - Homo sapiens (human)	3.13E-06	GO-BP-0051135:positive regulation of NK T cell activation	2.00E-06	KEGG-Pathway-hsa05222: Small cell lung cancer - Homo sapiens (human)	1.22E-07
GO-BP-0001817:regulation of cytokine production	2.54E-08	GO-BP-0051135:positive regulation of NK T cell activation	3.20E-06	GO-BP-0042519:regulation of tyrosine phosphorylation of Stat4 protein	1.91E-06	KEGG-Pathway-hsa05215: Prostate cancer - Homo sapiens (human)	1.54E-07

IW3134268

Receptor signaling associated with external infective ligand signal
Infective process
Immune system process
Internal signaling: NFKB
Internal signaling: STAT4
Cancer Process
Apoptosis or cell death process

Table 4-7 GIANT Enrichment analysis of Cluster 1: Patient example with HDAC7, NFKB1 and PRKCB as SNP affected proteins

B Lymphocyte		Colon		Dendritic Cell		T lymphocyte	
Pathways/Processes/Diseases	P-Value (FDR Corrected)	Pathways/Processes/Diseases	P-Value (FDR Corrected)	Pathways/Processes/Diseases	P-Value (FDR Corrected)	Pathways/Processes/Diseases	P-Value (FDR Corrected)
KEGG-Pathway-hsa04114: Oocyte meiosis - Homo sapiens (human)	6.57E-06	KEGG-Pathway-hsa04114: Oocyte meiosis - Homo sapiens (human)	2.32E-09	KEGG-Pathway-hsa04530: Tight junction - Homo sapiens (human)	1.34E-02	KEGG-Pathway-hsa03015: mRNA surveillance pathway - Homo sapiens	2.84E-05
KEGG-Pathway-hsa04390: Hippo signaling pathway - Homo sapiens	3.40E-05	KEGG-Pathway-hsa03015: mRNA surveillance pathway - Homo sapiens	6.45E-07	KEGG-Pathway-hsa04115: p53 signaling pathway - Homo sapiens (human)	8.47E-03	KEGG-Pathway-hsa04114: Oocyte meiosis - Homo sapiens (human)	1.38E-03
KEGG-Pathway-hsa03015: mRNA surveillance pathway - Homo sapiens	7.46E-04	KEGG-Pathway-hsa05160: Hepatitis C - Homo sapiens (human)	6.18E-06	GO-BP-0060548:negative regulation of cell death	7.20E-03	KEGG-Pathway-hsa05160: Hepatitis C - Homo sapiens (human)	2.33E-03
KEGG-Pathway-hsa05160: Hepatitis C - Homo sapiens (human)	3.59E-03	KEGG-Pathway-hsa04390: Hippo signaling pathway - Homo sapiens	1.28E-05	KEGG-Pathway-hsa04390: Hippo signaling pathway - Homo sapiens	6.53E-03	KEGG-Pathway-hsa04390: Hippo signaling pathway - Homo sapiens (human)	3.57E-03
KEGG-Pathway-hsa05142: Chagas disease (American trypanosomiasis) - Homo sapiens (human)	1.91E-02	KEGG-Pathway-hsa04730: Long-term depression - Homo sapiens (human)	4.61E-05	KEGG-Pathway-hsa04210: Apoptosis - Homo sapiens (human)	8.57E-03	KEGG-Pathway-hsa05142: Chagas disease (American trypanosomiasis) - Homo sapiens (human)	1.08E-02
KEGG-Pathway-hsa05203: Viral carcinogenesis - Homo sapiens (human)	1.92E-02	KEGG-Pathway-hsa05145: Toxoplasmosis - Homo sapiens (human)	5.24E-05	KEGG-Pathway-hsa05215: Prostate cancer - Homo sapiens (human)	8.17E-03	KEGG-Pathway-hsa04728: Dopaminergic synapse - Homo sapiens (human)	2.20E-02
KEGG-Pathway-hsa04728: Dopaminergic synapse - Homo sapiens (human)	3.32E-02	KEGG-Pathway-hsa04728: Dopaminergic synapse - Homo sapiens (human)	6.85E-05	KEGG-Pathway-hsa04114: Oocyte meiosis - Homo sapiens (human)	1.60E-02	KEGG-Pathway-hsa04530: Tight junction - Homo sapiens (human)	2.06E-02
KEGG-Pathway-hsa05134: Legionellosis - Homo sapiens (human)	3.16E-02	KEGG-Pathway-hsa04261: Adrenergic signaling in cardiomyocytes - Homo	1.28E-04	KEGG-Pathway-hsa05145: Toxoplasmosis - Homo sapiens (human)	2.08E-02	DO-0050686:organ system cancer	2.56E-02
KEGG-Pathway-hsa04530: Tight junction - Homo sapiens (human)	2.81E-02	KEGG-Pathway-hsa05142: Chagas disease (American trypanosomiasis) -	4.00E-04	GO-BP-0043066:negative regulation of apoptotic process	1.98E-02	KEGG-Pathway-hsa04730: Long-term depression - Homo sapiens (human)	2.34E-02
KEGG-Pathway-hsa04730: Long-term depression - Homo sapiens (human)	3.27E-02	KEGG-Pathway-hsa04530: Tight junction - Homo sapiens (human)	1.24E-03	GO-BP-0043069:negative regulation of programmed cell death	1.96E-02	KEGG-Pathway-hsa04261: Adrenergic signaling in cardiomyocytes - Homo	2.16E-02
KEGG-Pathway-hsa05161: Hepatitis B - Homo sapiens (human)	3.18E-02	KEGG-Pathway-hsa04350: TGF-beta signaling pathway - Homo sapiens	2.72E-03	KEGG-Pathway-hsa05014: Amyotrophic lateral sclerosis (ALS) - Homo sapiens	1.80E-02	DO-162:cancer	2.12E-02
KEGG-Pathway-hsa04261: Adrenergic signaling in cardiomyocytes - Homo	3.15E-02	KEGG-Pathway-hsa05215: Prostate cancer - Homo sapiens (human)	3.79E-03	KEGG-Pathway-hsa05213: Endometrial cancer - Homo sapiens (human)	1.75E-02	GO-BP-0071900:regulation of protein serine/threonine kinase activity	2.20E-02
GO-BP-0043536:positive regulation of blood vessel endothelial cell migration	3.86E-02	KEGG-Pathway-hsa05169: Epstein-Barr virus infection - Homo sapiens (human)	6.94E-03	KEGG-Pathway-hsa05160: Hepatitis C - Homo sapiens (human)	1.78E-02	DO-14566:disease of cellular proliferation	2.18E-02
DO-0050615:respiratory system cancer	3.58E-02	KEGG-Pathway-hsa05203: Viral carcinogenesis - Homo sapiens (human)	6.91E-03	KEGG-Pathway-hsa05134: Legionellosis - Homo sapiens (human)	1.77E-02	GO-BP-0043536:positive regulation of blood vessel endothelial cell migration	2.63E-02
DO-1324:lung cancer	3.35E-02	GO-BP-0080135:regulation of cellular response to stress	6.90E-03	GO-BP-0060284:regulation of cell development	1.68E-02	DO-0050615:respiratory system cancer	2.46E-02
GO-BP-0032663:regulation of interleukin-2 production	4.45E-02	GO-BP-0030334:regulation of cell migration	1.21E-02	DO-0050686:organ system cancer	2.12E-02	DO-1324:lung cancer	2.31E-02
KEGG-Pathway-hsa04350: TGF-beta signaling pathway - Homo sapiens (human)	4.45E-02	GO-BP-0097190:apoptotic signaling pathway	1.16E-02	KEGG-Pathway-hsa05210: Colorectal cancer - Homo sapiens (human)	2.08E-02	KEGG-Pathway-hsa04350: TGF-beta signaling pathway - Homo sapiens (human)	2.89E-02
GO-BP-0032623:interleukin-2 production	4.39E-02	GO-BP-2000145:regulation of cell motility	1.30E-02	DO-162:cancer	2.14E-02	GO-BP-0032663:regulation of interleukin-2 production	2.91E-02
		GO-BP-0060284:regulation of cell development	1.23E-02	DO-14566:disease of cellular proliferation	2.46E-02	GO-BP-0032623:interleukin-2 production	3.06E-02
		GO-BP-0060548:negative regulation of cell death	1.19E-02	GO-BP-0043536:positive regulation of blood vessel endothelial cell migration	2.56E-02	KEGG-Pathway-hsa04210: Apoptosis - Homo sapiens (human)	3.03E-02
		GO-BP-0030335:positive regulation of cell migration	1.42E-02	KEGG-Pathway-hsa04520: Adherens junction - Homo sapiens (human)	2.71E-02	KEGG-Pathway-hsa05215: Prostate cancer - Homo sapiens (human)	3.19E-02
		GO-BP-0045859:regulation of protein kinase activity	1.37E-02	GO-BP-0032663:regulation of interleukin-2 production	3.31E-02	GO-BP-0071902:positive regulation of protein serine/threonine kinase activity	4.39E-02
		KEGG-Pathway-hsa05161: Hepatitis B - Homo sapiens (human)	1.33E-02	GO-BP-0032355:response to estradiol	3.16E-02		

IW3134619

mRNA surveillance
Infective process
Immune system process
Internal signaling: kinase activity
Cellular junctions
Cellular proliferation and migration
Apoptosis or cell death process
Cancer Process
Disease or unable to characterise

Table 4-8 GIANT Enrichment analysis of Cluster 2: Patient example from LSP1, NFKB1, RGS14 SNP affected proteins

B Lymphocyte		Colon		Dendritic Cell		T lymphocyte	
Pathways/Processes/Diseases	P-Value (FDR Corrected)	Pathways/Processes/Diseases	P-Value (FDR Corrected)	Pathways/Processes/Diseases	P-Value (FDR Corrected)	Pathways/Processes/Diseases	P-Value (FDR Corrected)
KEGG-Pathway-hsa04622: RIG-I-like receptor signaling pathway - Homo sapiens (human)	7.12E-11	KEGG-Pathway-hsa04622: RIG-I-like receptor signaling pathway - Homo sapiens (human)	3.71E-08	GO-BP-0001817:regulation of cytokine production	5.93E-07	KEGG-Pathway-hsa05168: Herpes simplex infection - Homo sapiens (human)	2.60E-09
KEGG-Pathway-hsa05162: Measles - Homo sapiens (human)	7.31E-11	KEGG-Pathway-hsa05168: Herpes simplex infection - Homo sapiens (human)	9.92E-08	GO-BP-0001816:cytokine production	5.10E-07	KEGG-Pathway-hsa05164: Influenza A - Homo sapiens (human)	2.70E-09
KEGG-Pathway-hsa05168: Herpes simplex infection - Homo sapiens (human)	7.32E-11	KEGG-Pathway-hsa05162: Measles - Homo sapiens (human)	7.88E-08	GO-BP-0009607:response to biotic stimulus	4.86E-07	KEGG-Pathway-hsa05162: Measles - Homo sapiens (human)	2.71E-09
KEGG-Pathway-hsa05164: Influenza A - Homo sapiens (human)	7.34E-11	GO-BP-0001816:cytokine production	2.22E-07	GO-BP-0009615:response to virus	2.28E-06	GO-BP-0001817:regulation of cytokine production	1.60E-07
KEGG-Pathway-hsa04620: Toll-like receptor signaling pathway - Homo sapiens (human)	1.81E-10	KEGG-Pathway-hsa04620: Toll-like receptor signaling pathway - Homo sapiens (human)	2.18E-07	GO-BP-0001818:negative regulation of cytokine production	1.98E-06	GO-BP-0001816:cytokine production	2.33E-07
KEGG-Pathway-hsa05160: Hepatitis C - Homo sapiens (human)	1.36E-09	KEGG-Pathway-hsa05133: Pertussis - Homo sapiens (human)	5.14E-07	GO-BP-0051132:NK T cell activation	1.97E-06	GO-BP-0009607:response to biotic stimulus	2.88E-07
KEGG-Pathway-hsa04064: NF-kappa B signaling pathway - Homo sapiens (human)	1.38E-09	GO-BP-0001817:regulation of cytokine production	1.55E-06	GO-BP-0051133:regulation of NK T cell activation	1.69E-06	KEGG-Pathway-hsa05133: Pertussis - Homo sapiens (human)	5.12E-07
KEGG-Pathway-hsa05169: Epstein-Barr virus infection - Homo sapiens (human)	3.02E-09	GO-BP-0046635:positive regulation of alpha-beta T cell activation	1.46E-06	GO-BP-0051135:positive regulation of NK T cell activation	1.48E-06	KEGG-Pathway-hsa05152: Tuberculosis - Homo sapiens (human)	5.44E-07
GO-BP-0001817:regulation of cytokine production	3.19E-09	GO-BP-0009607:response to biotic stimulus	2.91E-06	GO-BP-0042519:regulation of tyrosine phosphorylation of Stat4 protein	1.31E-06	GO-BP-0002252:immune effector process	1.07E-06
GO-BP-0001816:cytokine production	5.54E-09	GO-BP-0046634:regulation of alpha-beta T cell activation	3.04E-06	GO-BP-0042520:positive regulation of tyrosine phosphorylation of Stat4 protein	1.18E-06	GO-BP-0051707:response to other organism	2.05E-06
KEGG-Pathway-hsa05133: Pertussis - Homo sapiens (human)	7.21E-09	GO-BP-0051132:NK T cell activation	2.93E-06	GO-BP-0042504:tyrosine phosphorylation of Stat4 protein	1.07E-06	GO-BP-0032660:regulation of interleukin-17 production	2.87E-06
KEGG-Pathway-hsa05140: Leishmaniasis - Homo sapiens (human)	7.43E-09	GO-BP-0051133:regulation of NK T cell activation	2.69E-06	GO-BP-0051241:negative regulation of multicellular organismal process	1.13E-06	GO-BP-0032620:interleukin-17 production	2.63E-06
GO-BP-0009607:response to biotic stimulus	7.79E-09	GO-BP-0051135:positive regulation of NK T cell activation	2.48E-06	KEGG-Pathway-hsa05133: Pertussis - Homo sapiens (human)	1.39E-06	GO-BP-0051132:NK T cell activation	2.49E-06
KEGG-Pathway-hsa05145: Toxoplasmosis - Homo sapiens (human)	1.00E-08	GO-BP-0042519:regulation of tyrosine phosphorylation of Stat4 protein	2.30E-06	GO-BP-0051707:response to other organism	1.70E-06	GO-BP-0051133:regulation of NK T cell activation	2.31E-06
GO-BP-0006952:defense response	1.30E-08	GO-BP-0042520:positive regulation of tyrosine phosphorylation of Stat4 protein	2.15E-06	KEGG-Pathway-hsa05168: Herpes simplex infection - Homo sapiens (human)	1.36E-05	GO-BP-0051135:positive regulation of NK T cell activation	2.16E-06
GO-BP-0002252:immune effector process	2.33E-08	GO-BP-0042504:tyrosine phosphorylation of Stat4 protein	2.02E-06	GO-BP-0032816:positive regulation of natural killer cell activation	1.47E-05	GO-BP-0042519:regulation of tyrosine phosphorylation of Stat4 protein	2.02E-06
KEGG-Pathway-hsa05134: Legionellosis - Homo sapiens (human)	2.47E-08	GO-BP-0046631:alpha-beta T cell activation	3.85E-06	GO-BP-0001819:positive regulation of cytokine production	1.46E-05	GO-BP-0042520:positive regulation of tyrosine phosphorylation of Stat4 protein	1.90E-06
KEGG-Pathway-hsa05161: Hepatitis B - Homo sapiens (human)	3.92E-08	GO-BP-0001818:negative regulation of cytokine production	3.71E-06	GO-BP-0032814:regulation of natural killer cell activation	2.28E-05	GO-BP-0042504:tyrosine phosphorylation of Stat4 protein	1.80E-06
GO-BP-0043122:regulation of I-kappaB kinase/NF-kappaB signaling	5.02E-08	KEGG-Pathway-hsa05164: Influenza A - Homo sapiens (human)	3.64E-06	GO-BP-0007050:cell cycle arrest	2.56E-05	KEGG-Pathway-hsa04620: Toll-like receptor signaling pathway - Homo sapiens (human)	2.16E-06
GO-BP-0051707:response to other organism	5.58E-08	KEGG-Pathway-hsa05152: Tuberculosis - Homo sapiens (human)	3.77E-06	GO-BP-0002819:regulation of adaptive immune response	2.49E-05	GO-BP-0006952:defense response	3.16E-06
GO-BP-0007249:I-kappaB kinase/NF-kappaB signaling	5.74E-08	KEGG-Pathway-hsa05160: Hepatitis C - Homo sapiens (human)	8.12E-06	KEGG-Pathway-hsa05160: Hepatitis C - Homo sapiens (human)	2.67E-05	KEGG-Pathway-hsa05143: African trypanosomiasis - Homo sapiens (human)	3.49E-06
KEGG-Pathway-hsa04668: TNF signaling pathway - Homo sapiens (human)	8.84E-08	GO-BP-0051707:response to other organism	1.34E-05	KEGG-Pathway-hsa04622: RIG-I-like receptor signaling pathway - Homo sapiens (human)	2.56E-05	GO-BP-0001818:negative regulation of cytokine production	3.39E-06
KEGG-Pathway-hsa05321: Inflammatory bowel disease (IBD) - Homo sapiens (human)	9.41E-08	GO-BP-0008285:negative regulation of cell proliferation	1.33E-05	KEGG-Pathway-hsa05162: Measles - Homo sapiens (human)	2.55E-05	GO-BP-0001819:positive regulation of cytokine production	3.34E-06
GO-BP-0006955:immune response	9.72E-08	KEGG-Pathway-hsa05161: Hepatitis B - Homo sapiens (human)	1.35E-05	GO-BP-0006955:immune response	3.63E-05	KEGG-Pathway-hsa04622: RIG-I-like receptor signaling pathway - Homo sapiens (human)	3.70E-06

IW3137761

Receptor signaling associated with external infective ligand signal
Infective process
Immune system process
Internal signaling: NFKB
Internal signaling: STAT4
Cancer Process
Apoptosis or cell death process

Table 4-9 GIANT Enrichment analysis of Cluster 2: Patient example from HDAC7, IRF5, GNA12 and NFKB1 SNP affected proteins

B lymphocyte		Colon		Dendritic cell		T Lymphocyte	
Pathways/Processes/Diseases	P-Value (FDR Corrected)	Pathways/Processes/Diseases	P-Value (FDR Corrected)	Pathways/Processes/Diseases	P-Value (FDR Corrected)	Pathways/Processes/Diseases	P-Value (FDR Corrected)
KEGG-Pathway-hsa05162: Measles - Homo sapiens (human)	3.15E-11	KEGG-Pathway-hsa04622: RIG-I-like receptor signaling pathway - Homo sapiens (human)	3.71E-08	GO-BP-0001817:regulation of cytokine production	5.93E-07	KEGG-Pathway-hsa05168: Herpes simplex infection - Homo sapiens (human)	2.60E-09
KEGG-Pathway-hsa05161: Hepatitis B - Homo sapiens (human)	3.15E-11	KEGG-Pathway-hsa05168: Herpes simplex infection - Homo sapiens (human)	9.92E-08	GO-BP-0001816:cytokine production	5.10E-07	KEGG-Pathway-hsa05164: Influenza A - Homo sapiens (human)	2.70E-09
KEGG-Pathway-hsa05145: Toxoplasmosis - Homo sapiens (human)	3.15E-11	KEGG-Pathway-hsa05162: Measles - Homo sapiens (human)	7.88E-08	GO-BP-0009607:response to biotic stimulus	4.86E-07	KEGG-Pathway-hsa05162: Measles - Homo sapiens (human)	2.71E-09
KEGG-Pathway-hsa05164: Influenza A - Homo sapiens (human)	3.15E-11	GO-BP-0001816:cytokine production	2.22E-07	GO-BP-0009615:response to virus	2.28E-06	GO-BP-0001817:regulation of cytokine production	1.60E-07
KEGG-Pathway-hsa05160: Hepatitis C - Homo sapiens (human)	3.15E-11	KEGG-Pathway-hsa04620: Toll-like receptor signaling pathway - Homo sapiens (human)	2.18E-07	GO-BP-0001818:negative regulation of cytokine production	1.98E-06	GO-BP-0001816:cytokine production	2.33E-07
KEGG-Pathway-hsa04620: Toll-like receptor signaling pathway - Homo sapiens (human)	3.32E-11	KEGG-Pathway-hsa05133: Pertussis - Homo sapiens (human)	5.14E-07	GO-BP-0051132:NK T cell activation	1.97E-06	GO-BP-0009607:response to biotic stimulus	2.88E-07
KEGG-Pathway-hsa05168: Herpes simplex infection - Homo sapiens (human)	3.47E-11	GO-BP-0001817:regulation of cytokine production	1.55E-06	GO-BP-0051133:regulation of NK T cell activation	1.69E-06	KEGG-Pathway-hsa05133: Pertussis - Homo sapiens (human)	5.12E-07
KEGG-Pathway-hsa04622: RIG-I-like receptor signaling pathway - Homo sapiens (human)	3.85E-11	GO-BP-0046635:positive regulation of alpha-beta T cell activation	1.46E-06	GO-BP-0051135:positive regulation of NK T cell activation	1.48E-06	KEGG-Pathway-hsa05152: Tuberculosis - Homo sapiens (human)	5.44E-07
KEGG-Pathway-hsa05134: Legionellosis - Homo sapiens (human)	4.29E-11	GO-BP-0009607:response to biotic stimulus	2.91E-06	GO-BP-0042519:regulation of tyrosine phosphorylation of Stat4 protein	1.31E-06	GO-BP-0002252:immune effector process	1.07E-06
KEGG-Pathway-hsa05169: Epstein-Barr virus infection - Homo sapiens (human)	3.29E-10	GO-BP-0046634:regulation of alpha-beta T cell activation	3.04E-06	GO-BP-0042520:positive regulation of tyrosine phosphorylation of Stat4 protein	1.18E-06	GO-BP-0051707:response to other organism	2.05E-06
KEGG-Pathway-hsa05133: Pertussis - Homo sapiens (human)	3.90E-10	GO-BP-0051132:NK T cell activation	2.93E-06	GO-BP-0042504:tyrosine phosphorylation of Stat4 protein	1.07E-06	GO-BP-0032660:regulation of interleukin-17 production	2.87E-06
KEGG-Pathway-hsa05140: Leishmaniasis - Homo sapiens (human)	4.09E-10	GO-BP-0051133:regulation of NK T cell activation	2.69E-06	GO-BP-0051241:negative regulation of multicellular organismal process	1.13E-06	GO-BP-0032620:interleukin-17 production	2.63E-06
GO-BP-0001817:regulation of cytokine production	4.13E-10	GO-BP-0051135:positive regulation of NK T cell activation	2.48E-06	KEGG-Pathway-hsa05133: Pertussis - Homo sapiens (human)	1.39E-06	GO-BP-0051132:NK T cell activation	2.49E-06
GO-BP-0001816:cytokine production	5.96E-10	GO-BP-0042519:regulation of tyrosine phosphorylation of Stat4 protein	2.30E-06	GO-BP-0051707:response to other organism	1.70E-06	GO-BP-0051133:regulation of NK T cell activation	2.31E-06
KEGG-Pathway-hsa05152: Tuberculosis - Homo sapiens (human)	1.68E-09	GO-BP-0042520:positive regulation of tyrosine phosphorylation of Stat4 protein	2.15E-06	KEGG-Pathway-hsa05168: Herpes simplex infection - Homo sapiens (human)	1.36E-05	GO-BP-0051135:positive regulation of NK T cell activation	2.16E-06
KEGG-Pathway-hsa04064: NF-kappa B signaling pathway - Homo sapiens (human)	1.72E-09	GO-BP-0042504:tyrosine phosphorylation of Stat4 protein	2.02E-06	GO-BP-0032816:positive regulation of natural killer cell activation	1.47E-05	GO-BP-0042519:regulation of tyrosine phosphorylation of Stat4 protein	2.02E-06
GO-BP-0002252:immune effector process	2.07E-09	GO-BP-0046631:alpha-beta T cell activation	3.85E-06	GO-BP-0001819:positive regulation of cytokine production	1.46E-05	GO-BP-0042520:positive regulation of tyrosine phosphorylation of Stat4 protein	1.90E-06
GO-BP-0043122:regulation of I-kappaB kinase/NF-kappaB signaling	5.70E-09	GO-BP-0001818:negative regulation of cytokine production	3.71E-06	GO-BP-0032814:regulation of natural killer cell activation	2.28E-05	GO-BP-0042504:tyrosine phosphorylation of Stat4 protein	1.80E-06
KEGG-Pathway-hsa05321: Inflammatory bowel disease (IBD) - Homo sapiens (human)	5.77E-09	KEGG-Pathway-hsa05164: Influenza A - Homo sapiens (human)	3.64E-06	GO-BP-0007050:cell cycle arrest	2.56E-05	KEGG-Pathway-hsa04620: Toll-like receptor signaling pathway - Homo sapiens (human)	2.16E-06
GO-BP-0007249:I-kappaB kinase/NF-kappaB signaling	6.61E-09	KEGG-Pathway-hsa05152: Tuberculosis - Homo sapiens (human)	3.77E-06	GO-BP-0002819:regulation of adaptive immune response	2.49E-05	GO-BP-0006952:defense response	3.16E-06
KEGG-Pathway-hsa04668: TNF signaling pathway - Homo sapiens (human)	7.38E-09	KEGG-Pathway-hsa05160: Hepatitis C - Homo sapiens (human)	8.12E-06	KEGG-Pathway-hsa05160: Hepatitis C - Homo sapiens (human)	2.67E-05	KEGG-Pathway-hsa05143: African trypanosomiasis - Homo sapiens (human)	3.49E-06
GO-BP-0050776:regulation of immune response	1.12E-08	GO-BP-0051707:response to other organism	1.34E-05	KEGG-Pathway-hsa04622: RIG-I-like receptor signaling pathway - Homo sapiens (human)	2.56E-05	GO-BP-0001818:negative regulation of cytokine production	3.39E-06
GO-BP-0009607:response to biotic stimulus	1.32E-08	GO-BP-0008285:negative regulation of cell proliferation	1.33E-05	KEGG-Pathway-hsa05162: Measles - Homo sapiens (human)	2.55E-05	GO-BP-0001819:positive regulation of cytokine production	3.34E-06
GO-BP-0006955:immune response	1.60E-08	KEGG-Pathway-hsa05161: Hepatitis B - Homo sapiens (human)	1.35E-05	GO-BP-0006955:immune response	3.63E-05	KEGG-Pathway-hsa04622: RIG-I-like receptor signaling pathway - Homo sapiens (human)	3.70E-06

IW3138659

Receptor signaling associated with external infective ligand signal
Infective process
Immune system process
Internal signaling: NFKB
Internal signaling: STAT4
Cancer Process
Apoptosis or cell death process

Table 4-10 GIANT Enrichment analysis of Cluster 4: Patient example from LSP1, HDAC7, CCNY and PRKCB SNP affected proteins

B lymphocyte		Colon		Dendritic Cell		T Lymphocyte	
Pathways/Processes/Diseases	P-Value (FDR Corrected	Pathways/Processes/Diseases	P-Value (FDR Corrected)	Pathways/Processes/Diseases	P-Value (FDR Corrected)	Pathways/Processes/Diseases	P-Value (FDR Corrected)
KEGG-Pathway-hsa04390: Hippo signaling pathway - Homo sapiens	1.00E-04	KEGG-Pathway-hsa03015: mRNA surveillance pathway - Homo sapiens	1.01E-09	KEGG-Pathway-hsa04530: Tight junction - Homo sapiens (human)	3.16E-05	KEGG-Pathway-hsa04390: Hippo signaling pathway - Homo sapiens	1.14E-06
KEGG-Pathway-hsa03015: mRNA surveillance pathway - Homo	9.93E-05	KEGG-Pathway-hsa04728: Dopaminergic synapse - Homo sapiens	9.66E-09	KEGG-Pathway-hsa03015: mRNA surveillance pathway - Homo sapiens	7.65E-05	KEGG-Pathway-hsa03015: mRNA surveillance pathway - Homo sapiens	9.55E-07
KEGG-Pathway-hsa04114: Oocyte meiosis - Homo sapiens (human)	1.70E-04	KEGG-Pathway-hsa04114: Oocyte meiosis - Homo sapiens (human)	1.02E-07	KEGG-Pathway-hsa04114: Oocyte meiosis - Homo sapiens (human)	1.31E-04	KEGG-Pathway-hsa04114: Oocyte meiosis - Homo sapiens (human)	2.01E-06
KEGG-Pathway-hsa05142: Chagas disease (American trypanosomiasis) -	2.99E-03	KEGG-Pathway-hsa04261: Adrenergic signaling in cardiomyocytes - Homo	6.46E-07	KEGG-Pathway-hsa04728: Dopaminergic synapse - Homo	2.34E-04	KEGG-Pathway-hsa04530: Tight junction - Homo sapiens (human)	4.92E-06
KEGG-Pathway-hsa04728: Dopaminergic synapse - Homo sapiens (human)	5.88E-03	KEGG-Pathway-hsa04390: Hippo signaling pathway - Homo sapiens (human)	2.36E-05	KEGG-Pathway-hsa04261: Adrenergic signaling in cardiomyocytes - Homo sapiens (human)	3.52E-04	KEGG-Pathway-hsa05142: Chagas disease (American trypanosomiasis) - Homo sapiens (human)	4.01E-05
KEGG-Pathway-hsa05160: Hepatitis C - Homo sapiens (human)	5.20E-03	KEGG-Pathway-hsa05142: Chagas disease (American trypanosomiasis) - Homo sapiens (human)	8.12E-05	KEGG-Pathway-hsa04390: Hippo signaling pathway - Homo sapiens (human)	3.44E-04	KEGG-Pathway-hsa04728: Dopaminergic synapse - Homo sapiens (human)	1.05E-04
KEGG-Pathway-hsa04530: Tight junction - Homo sapiens (human)	4.58E-03	KEGG-Pathway-hsa05160: Hepatitis C - Homo sapiens (human)	2.35E-04	KEGG-Pathway-hsa05142: Chagas disease (American trypanosomiasis) - Homo sapiens (human)	1.41E-03	KEGG-Pathway-hsa05160: Hepatitis C - Homo sapiens (human)	9.70E-05
GO-BP-0060284:regulation of cell development	4.25E-03	KEGG-Pathway-hsa04530: Tight junction - Homo sapiens (human)	2.13E-04	KEGG-Pathway-hsa05160: Hepatitis C - Homo sapiens (human)	3.23E-03	KEGG-Pathway-hsa04730: Long-term depression - Homo sapiens (human)	9.96E-05
GO-BP-0048468:cell development	4.43E-03	KEGG-Pathway-hsa04730: Long-term depression - Homo sapiens (human)	2.00E-04	KEGG-Pathway-hsa04730: Long-term depression - Homo sapiens (human)	5.34E-03	KEGG-Pathway-hsa04261: Adrenergic signaling in	1.32E-04
KEGG-Pathway-hsa04261: Adrenergic signaling in	4.83E-03	KEGG-Pathway-hsa04350: TGF-beta signaling pathway - Homo sapiens	5.71E-04	KEGG-Pathway-hsa04350: TGF-beta signaling pathway - Homo sapiens	1.13E-02	KEGG-Pathway-hsa04350: TGF-beta signaling pathway - Homo sapiens	2.53E-04
KEGG-Pathway-hsa04730: Long-term depression - Homo sapiens	4.98E-03	GO-BP-0060284:regulation of cell development	4.18E-03	GO-BP-0048468:cell development	3.50E-02	GO-BP-0060284:regulation of cell development	1.86E-03
neuron differentiation	6.33E-03	GO-BP-0048468:cell development	4.26E-03	development	4.38E-02	GO-BP-0048468:cell development	1.78E-03
GO-BP-0050767:regulation of neurogenesis	9.50E-03	GO-BP-0045664:regulation of neuron differentiation	8.39E-03			GO-BP-0045664:regulation of neuron differentiation	3.97E-03
KEGG-Pathway-hsa04350: TGF-beta signaling pathway - Homo sapiens	9.15E-03	GO-BP-0050767:regulation of neurogenesis	1.27E-02			GO-BP-0050767:regulation of neurogenesis	6.00E-03
GO-BP-0051960:regulation of nervous system development	9.85E-03	GO-BP-0051960:regulation of nervous system development	1.42E-02			GO-BP-0051960:regulation of nervous system development	6.71E-03
differentiation	2.73E-02	differentiation	3.93E-02			differentiation	1.87E-02
GO-BP-0048699:generation of	3.50E-02					GO-BP-0048699:generation of	2.40E-02
GO-BP-0022008:neurogenesis	4.53E-02					GO-BP-0022008:neurogenesis	3.12E-02

IW3141986

mRNA surveillance
Infective process
Cellular proliferation and migration
Apoptosis or cell death process
Neurological pathways
Cellular junctions

Table 4-11 GIANT Enrichment analysis of Cluster 4: Patient example from HDAC7, GNA12, IRF5, PRKCB and ZGPAT SNP affected proteins

B lymphocyte		Colon		Dendritic cell		T lymphocte	
Pathways/Processes/Diseases	P-Value (FDR Corrected)	Pathways/Processes/Diseases	P-Value (FDR Corrected)	Pathways/Processes/Diseases	P-Value (FDR Corrected)	Pathways/Processes/Diseases	P-Value (FDR Corrected)
KEGG-Pathway-hsa03015: mRNA surveillance pathway - Homo sapiens (human)	1.98E-06	KEGG-Pathway-hsa03015: mRNA surveillance pathway - Homo sapiens (human)	9.67E-11	KEGG-Pathway-hsa03015: mRNA surveillance pathway - Homo sapiens (human)	8.52E-05	KEGG-Pathway-hsa03015: mRNA surveillance pathway - Homo sapiens (human)	4.64E-07
KEGG-Pathway-hsa04390: Hippo signaling pathway - Homo sapiens (human)	2.34E-05	KEGG-Pathway-hsa04114: Oocyte meiosis - Homo sapiens (human)	9.78E-11	KEGG-Pathway-hsa04728: Dopaminergic synapse - Homo sapiens (human)	2.62E-04	KEGG-Pathway-hsa04114: Oocyte meiosis - Homo sapiens (human)	7.35E-07
KEGG-Pathway-hsa05142: Chagas disease (American trypanosomiasis) - Homo sapiens	6.95E-05	KEGG-Pathway-hsa04728: Dopaminergic synapse - Homo sapiens (human)	2.72E-10	KEGG-Pathway-hsa04261: Adrenergic signaling in cardiomyocytes - Homo sapiens	3.30E-04	KEGG-Pathway-hsa04390: Hippo signaling pathway - Homo sapiens (human)	3.72E-06
KEGG-Pathway-hsa04114: Oocyte meiosis - Homo sapiens (human)	6.89E-05	KEGG-Pathway-hsa04261: Adrenergic signaling in cardiomyocytes - Homo sapiens	5.55E-10	KEGG-Pathway-hsa05142: Chagas disease (American trypanosomiasis) - Homo	1.61E-03	KEGG-Pathway-hsa04350: TGF-beta signaling pathway - Homo sapiens (human)	4.34E-06
KEGG-Pathway-hsa04728: Dopaminergic synapse - Homo sapiens (human)	1.31E-04	KEGG-Pathway-hsa04390: Hippo signaling pathway - Homo sapiens (human)	2.48E-06	KEGG-Pathway-hsa04114: Oocyte meiosis - Homo sapiens (human)	1.60E-03	KEGG-Pathway-hsa05142: Chagas disease (American trypanosomiasis) - Homo sapiens	1.30E-05
KEGG-Pathway-hsa05160: Hepatitis C - Homo sapiens (human)	1.18E-04	KEGG-Pathway-hsa05142: Chagas disease (American trypanosomiasis) - Homo sapiens	1.21E-05	KEGG-Pathway-hsa05160: Hepatitis C - Homo sapiens (human)	2.82E-03	KEGG-Pathway-hsa04728: Dopaminergic synapse - Homo sapiens (human)	3.44E-05
KEGG-Pathway-hsa04530: Tight junction - Homo sapiens	1.05E-04	KEGG-Pathway-hsa05160: Hepatitis C - Homo sapiens	3.53E-05	KEGG-Pathway-hsa04530: Tight junction - Homo sapiens	2.49E-03	KEGG-Pathway-hsa05160: Hepatitis C - Homo sapiens	3.18E-05
KEGG-Pathway-hsa04730: Long-term depression - Homo sapiens (human)	1.03E-04	KEGG-Pathway-hsa04530: Tight junction - Homo sapiens (human)	3.21E-05	KEGG-Pathway-hsa04390: Hippo signaling pathway - Homo sapiens (human)	3.74E-03	KEGG-Pathway-hsa04530: Tight junction - Homo sapiens (human)	2.89E-05
KEGG-Pathway-hsa04261: Adrenergic signaling in cardiomyocytes - Homo sapiens (human)	1.37E-04	KEGG-Pathway-hsa04730: Long-term depression - Homo sapiens (human)	4.11E-05	KEGG-Pathway-hsa04730: Long-term depression - Homo sapiens (human)	3.98E-03	KEGG-Pathway-hsa04730: Long-term depression - Homo sapiens (human)	3.70E-05
KEGG-Pathway-hsa04350: TGF-beta signaling pathway - Homo sapiens (human)	2.63E-04	KEGG-Pathway-hsa04350: TGF-beta signaling pathway - Homo sapiens (human)	1.18E-04	KEGG-Pathway-hsa04350: TGF-beta signaling pathway - Homo sapiens (human)	8.42E-03	KEGG-Pathway-hsa04261: Adrenergic signaling in cardiomyocytes - Homo sapiens	3.91E-05
GO-BP-0048468:cell development	2.68E-02	GO-BP-0048468:cell development	1.26E-02	GO-BP-2001238:positive regulation of extrinsic apoptotic	2.13E-02	GO-BP-0048468:cell development	1.14E-02
GO-BP-0060284:regulation of cell development	3.54E-02	GO-BP-0060284:regulation of cell development	2.05E-02	GO-BP-0048468:cell development	2.15E-02	GO-BP-0060284:regulation of cell development	1.84E-02

mRNA surveillance
Infective process
Cellular proliferation and migration
Apoptosis or cell death process
Neurological pathways
Cellular junctions

4.5.3.2 GIANT cell specificity analysis using convergent nodes

Broadly B cells, T cells, dendritic cells, neutrophils, macrophages and enterocytes have all been implicated in the pathogenesis of UC. It is unclear whether UC is specifically T cell driven, or as is more likely, multiple cell types involved with the pathogenesis of UC with SNPs have slightly different phenotypic effects depending on their cell type (as seen above). To identify the key cell players, a weight needs to be applied to the strength of the protein-protein interaction within specific cell types. This was undertaken using GIANT, and conditional formatting applied to identify high strength interactions. There was no specific cell type for the clusters footprints as they all had low to mid-level interactions.

4.5.4 Supervised analysis of the unsupervised clustering

The clusters were not significantly different with regard to severity of disease by 5 years or 10 years of diagnosis (e.g. requirement of thiopurines or above), nor site of disease. Some patients were excluded from the 10-year analysis as they had not yet reached this time point, reducing the sensitivity of the analysis further due to low numbers. The numbers were also too small to analyse for genomic effects on extra-intestinal manifestations. There was significant gender disparity between the groups with NFKB1 hub disease being a disease predominately of males.

We could however show that the clusters were significantly different in terms of the age the patients were diagnosed in ach cluster (Figure 4-10, with cluster 2 (NFKB1 Predominant disease) being the youngest cohort by age of diagnosis, and cluster 3 containing some of the eldest patients to be diagnosed of the cohort.

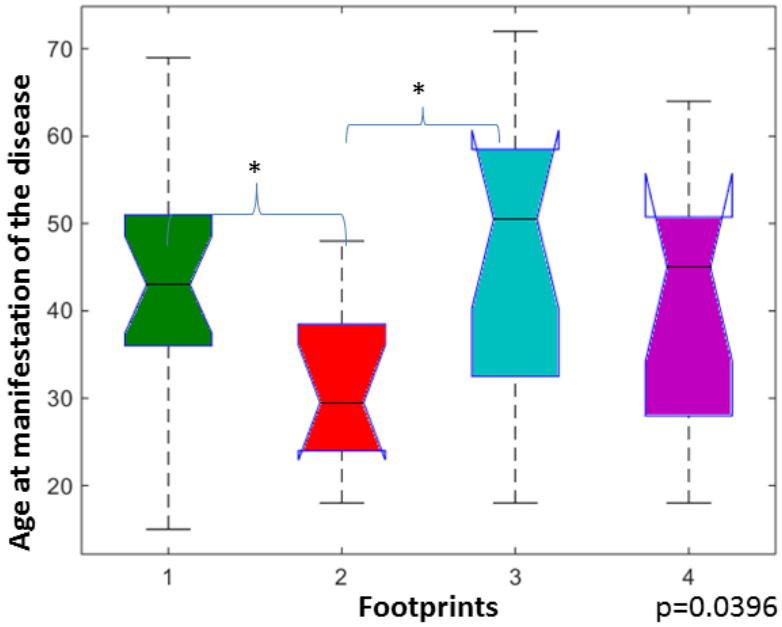


Figure 4-10 Age at manifestation of the disease across the clusters (footprints). Green is cluster 1, red is cluster 2, turquoise is cluster 3 and purple is cluster 4.

If this was due to a critical mass of pathways affected, cluster 1 patients should have manifested the disease earlier, having the largest SNP burden, highlighting the potential importance of an environmental trigger within the cohort and the complexity of delineating phenotype from SNP interactions. The need for a further trigger is also suggested by cluster 3 having the lowest SNP burden in terms of convergent proteins. It could be hypothesised that an environmental trigger or other pathological pathway needs to be activated to develop the disease. This is a plausible conclusion as it was observed that the SNP in MAML2 was identified as being associated with severe disease in non-

NFKB1 containing cohorts (cohorts 3 and 4, p=0.0001 via Fishers Exact test). The SNP in MAML2 is annotated to cause an increase in MAML2 expression. MAMAL2 is known to activate the Notch Receptor, thereby increasing NOTCH1 activation, which as part of the downstream effects, regulates the cell cycle and cellular proliferation. It also actives the inhibitor of the NFKB inhibitor (the IKK complex consisting of IKKB, IKKA and NEMO), which phosphorylates the inhibitor of NFKB (IkB) leading to a translocation of NFKB1 from the cytosol to the nucleus where it is transcriptionally active (Figure 4-11), as well as being a transcriptional activator of NFKB1.

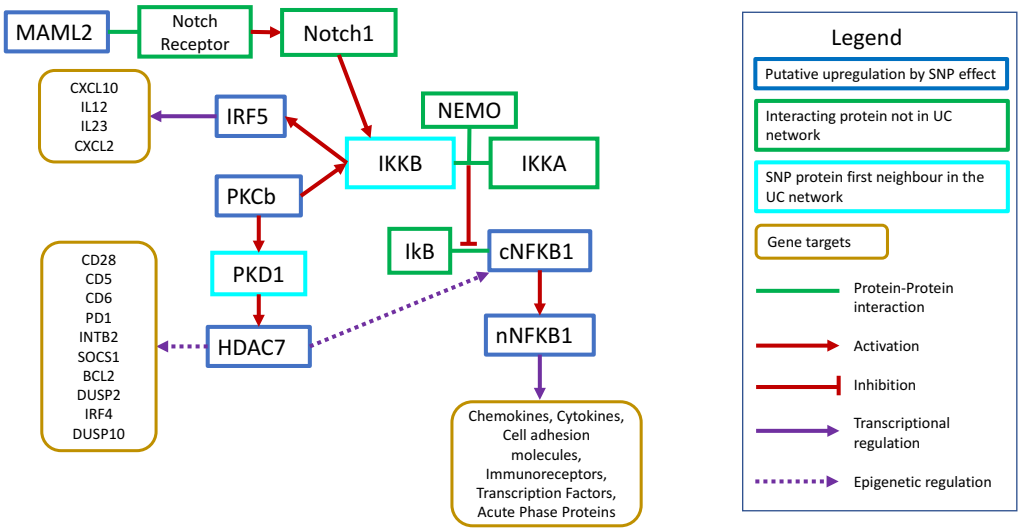


Figure 4-11 The NFKB1-MAML2 interaction. Diagram showing the interaction and convergence between UC associated SNP affected genes and the NFKB1 pathway. cNFKB1 = cytoplasmic NFKB1, nNFKB1 = nuclear NFKB1. All 5 SNPs are annotated to lead to an increase in NFKB1.

This indicates that even without direct NFBK1 involvement via a SNP, these MAML2 patients are activating the NFKB1 Pathway and this associated with more severe disease. The indicator of severity did not hold true for the MAML2 containing patients in cohort 1 and 2, indicating that this pathway is only part of the pathogenesis, but warrants further investigation in NFKB1 SNP negative cohorts for therapeutic targeting.

4.6 Discussion

Using patient specific data, we have shown that broadly, UC patients have different SNP burdens and different SNP convergence pathways, with the global effect on well documented pathways including pathogen handling and downstream immune response, cellular proliferation and apoptosis. We have documented how the putative SNP effects on gene regulation can have a convergent effect on downstream pathways specifically the NFKB1 pathway. We have observed that this convergence, albeit in a small cohort, does not show cell specificity, but appears to be present in all cell types that were analysed, pertinent to UC.

Our findings highlight the importance of other environmental factors that lead to the development of UC. The pathway enrichment indicates that this could be of viral, bacterial or parasite aetiology as all of these pathways were enriched. This is consistent with the current literature, showing that the tripartite of the epithelial barrier, the immune system and the human microbiota all have significant roles to play in the development of inflammatory bowel disease (378-380). Elucidating role of the virome in IBD is still in its infancy, however alteration in bacteriophage communities has been shown in IBD patients (139). Our research indicates that viral handling of DNA and RNA viruses (specifically herpes simplex, influenza virus, measles virus and hepatitis C) may play a role in the pathogenesis in IBD. Epidemiological evidence suggested a role for persistent measles virus infection in CD, but no causal role for measles virus in IBD has been found(381-383). No causal role for influenza virus or hepatitis C virus has been found for IBD. Herpes simplex virus can cause colitis, but is usually a superinfection in an immunocompromised host(384, 385). One potential role for these viruses in IBD pathogenesis is via the common denominator that these viruses all have the capacity to infect dendritic cells (DC) and modulate the inflammatory response via IL12 to alter dendritic cell ability to stimulate T cells (386-393). Hepatitis C, measles virus, influenza virus and herpes simplex all utilise an adhesion molecule called dendritic cell-specific intercellular adhesion molecule-3 Grabbing Non –integrin (DC-SIGN) (394-400). DC-SIGN is present on intestinal epithelial cells and epithelial cells with expression of DC-SIGN of can induce and control T cell differentiation and proliferation. DC-SIGN expression was strongly correlated with disease severity in children with IBD (401). The role of viruses (both human specific and bacteriophage) in the pathogenesis of UC, is an area that warrants further investigation.

4.6.1 Hidden players in the UC-regulome: Notch

Notch signalling (*via* the non-canonical pathway) is known to interact with multiple inflammatory responses including NFKB, hypoxia, epithelial-to-mesenchymal transition (involving TGF-beta), the Wnt signalling pathway and the mitogen activated protein kinase (402). TGF beta featured significantly in the pathway enrichment for cluster 4, but not clusters 1 or 2. Cluster 3 in the global gene ontology featured Wnt signalling as did cluster 2. Within the intestinal mucosa, Notch activation is necessary for epithelial regeneration after an inflammatory injury when there is depletion of goblet cells (403). Conversely it decreases goblet cell number in response in MMP9 and p-ERK signalling in response to upregulation of Claudin 1, a structural protein of the tight junction which gets upregulated in an inflammatory state (404). We know from Chapter 2 that tight junction regulation is a key feature of the function of the UC associated SNPs.

This is not the only pathway NOTCH impacts on from within the UC-ome; Notch has cell specific effects, for example, in macrophages NOTCH1 activation leads to the development of the inflammatory M1 phenotype (405, 406), and NOTCH1 deficiency has been shown to regulate the VEGF receptor and therefore downstream VEGF signalling (407), as well as inflammatory cytokine expression from macrophages including IL6 and IL12 (408). The VEGF pathway is a significant component of the pathways enrichment of both cluster 1 and 2. IL12 is impacted directly by SNPs in the UC-regulome, as well as the cytokine response and regulation in the pathway enrichment for clusters 1 and 2. In this way the Notch pathway is impacted in all 4 clusters.

Notch signalling is required for intestinal epithelial homeostasis(409) and goblet cell number . Goblet cell depletion is seen in UC (410) with overexpression of the Notch intracellular domain leading to depletion of secretory cells in intestinal crypt (411). Notch impacts on mucosal immune function by regulating the production of cytokines that bridge the innate and adaptive immunity in the gut (412). It is plausible that Notch can therefore contribute to the development or maintenance of the mucosal inflammation that characterises UC by modulating mucosal immune function and controlling physical and chemical attributes of the epithelial barrier.

4.6.2 Completion of aims and objectives and conclusions

This was the first test of iSNP and larger cohorts are needed to utilise it to its full potential. We used the iSNP workflow to create individual footprints for each patient based on their genomic data and identify key pathogenic pathways to disease from that individual patient data including highlighting a potential role of intracellular pathogens in the aetiology of UC. Using unsupervised, enrichment analysis of the data we have confirmed the role of the NFKB1 pathway, as well as identifying the Notch pathway as warranting further investigation for the cohort of patients who have the MAML2 SNP, and also as a pathogenic pathway to disease. This has been done and forms the basis of Brooks *et al* Systems genomics of UC: An integrative network analysis reveals the Notch pathway as a hidden player in disease pathogenesis under review.

We have also added weight to the literature on multiple other pathways previously proposed as pathogenic including VEGF and PDGF pathways. We have begun to see how this method can be used to correlate the patient footprint with their clinical data, however larger numbers are needed for significance and further experimental validation.

5. Conclusions and Future Work

The focus of this PhD was to increase the understanding of the pathogenic pathways that underpin UC using a combination of systems biology, in vitro and ex vivo techniques. To encompass the above, the first part of this thesis explored the bioinformatics and network biology techniques used to annotate the function of UC and IBD associated risk alleles and identified their role within a larger interactome. This allowed stratification of a SNP for experimental validation using both in vitro molecular biology techniques and ex vivo organ culture, which was highlighted in the central portion of the thesis. The final section of this thesis utilised the bioinformatics techniques honed in chapter 2 to identify patient specific UC networks, cluster patients together based on their genetics and to identify the convergence of SNP effects, elucidating the different pathogenic pathways to disease based on a patient's genetic landscape. This chapter summarises the findings from these studies, their impact in a wider field, and future directions of work.

Previously, GWAS and functional analysis of risk associated SNPs have looked for the identification of causal SNPs, with fine mapping, deep sequencing, and annotation of SNPs from within transcription factor binding sites or microRNA binding sites but very few have been ascribed to a definite causal variant with clear insight into the underlying disease pathogenesis. SNP epistasis has also been explored, identifying SNPs that affect interleukin production as working in concert. However, the hypothesis free global examination of multiple SNP functions with downstream protein-protein interactions has not been undertaken. As a result, the understanding of how SNPs may work in concert to drive pathogenesis is still poorly understood. In our study, the extensive, but exhaustive, functional annotation of SNPs utilising multiple validated techniques allowed us to create a UC interactome which could then be probed for pathogenic pathways, both novel and hypothesis driven.

Overall, the UC interactome identified UC to be a disease of regulation with a significant number of SNPs annotated to be in splicing sites, miRNA binding sites in genes and long non-coding RNAs or TFBS. This is not surprising as >90% SNPs are present in non-coding regions that would represent sites of regulation. The identification of TFBS motifs is in accordance with previous studies which have identified transcription factor motif disruption in RTEL1/TNFRSF6B confirmed with epigenetic peaks (43). Whilst other studies utilise epigenetic peaks to confirm, we utilised enhancer regions to confirm the validity of transcription factor binding sites. Enhancers are distinct genomic regions that contain

binding sites sequences for transcription factors which show the histone modifications that are identified in epigenetic peaks (413).

The simplicity of using nucleotide complementarity to identify mRNA targets gives rise to bioinformatics tools which are based on complementarity to the seed, evolutionary conservation and free energy binding. The best algorithms have false positive rates of 20-40%. Using the most stringent measures, we identified 107 miRNA binding sites from 56 SNPs. Creation and loss of miRNA binding sites by SNPs has been identified and validated previously (414, 415) at 3'UTR sites, but we also identified further sites in introns. These have been previously identified as imperfect centred, non-seed site, binding sites that mediate repression of target mRNAs and have been validated as genuine binding sites (416).

We also showed that a significant number of SNPs were found in splicing sites. Validating splicing site mutations using in vitro experiments is costly and time consuming and may not be practical, however in silico prediction methods are regarded as essential for analysing these variants (417). We utilized 3 validated mechanisms for splice enhancing sites (418), but were limited in splice enhancing sites. There was also no way to predict the phenotypic effect of the individual splicing sites. Splicing mutations they cause highly penetrant Mendelian diseases such as a SNP in HMBS causes exon skipping leading to acute intermittent porphyria, is an example at the severe end of a spectrum of functional variants that produce a gradation of phenotypic effects. We aimed to elucidate the potentially mild phenotype of the significant number of exon and intron splicing sites using network biology. The genes affected by splicing sites were first neighbours of regulators within the tight junction network, the autophagy network and the apoptosis network, and were first neighbours of regulatory proteins, or regulatory kinases/phosphatases within the focal adhesion network.

Unlike other studies using in silico methods, we did not find many deleterious effects of missense SNPs. This fits with UC being a disease of regulation, but it is at odds with the paradigm of missense mutations playing major roles in diseases such as cancer (419). it does fit with the concept that each risk associated locus only makes a small contribution to the disease phenotype, therefore subtle effects at a protein level can still be deleterious. ELM is a method, which utilises peptide motifs, but makes no assignation to the effect of such motif changes, however it is a tool to identify minor changes in protein

motifs, which would otherwise be missed by high impact assessments such as Polyphen, SNPs3D or SNPeff.

Consistent with the current GWAS literature, pathway analysis of the UC interactome indicated host-microbe interactions as a key global pathway (53, 171)but this is to be expected, as we are using the same proteins as was examined in these prior papers. What we have added with this section of the data is the potential key pathways and signalling cascades within host microbe interactions, that are regulated by multiple SNPs e.g. the tight junctions, autophagy and the focal adhesion complex. We have also identified, that unlike other functional annotation studies, when multiple modalities are examined there are hub SNPs that have multiple avenues of deleterious effect e.g. TFBS, MiRNA BS and splicing site alterations, which can have multiple impacts from gene expression, mRNA processing and mRNA repression depending on the dynamics of the cellular environment. This has to be taken in the context of wider IBD. There is a pathogenic overlap between CD and UC and by utilising IBD SNPs to create the UC interactome, we may have further illustrated the overlap. Future work, therefore analysing an equivalent CD cohort and comparing it against the UC cohort would both confirm the overlapping and highlight CD and UC specific pathways.

Using these networks, we were able to stratify a SNP for validation, a putative TFBS SNP in *LPXN*. This was, by necessity, hypothesis based and was influenced by the available resources to experimentally validate said SNP. There are validated pipelines for establishing the function of non-coding GWAS variants by fine mapping, epigenomic profiling, epigenome editing and then creating of isogenic cell lines for phenotypic characterisation (376), however we took a slightly different route. Utilising CRISPR-Cas9 we created leupaxin knock out and over expression epithelial cell lines. This technique has been successfully utilised elsewhere to identify and validate regulatory SNPs in prostate cancer (420) and for identification and validation of drug resistant mutations in mammalian cells (421). Whilst we were not able to create a SNP cell line, we were able to show that *LPXN* overexpression in cell lines impairs wound healing and modulates cytokine response via an integrin mediated and TLR mediated effect. This has not been shown in epithelial cells before. We then went on to identify that the presence of the *LPXN* risk allele homozygosity may alterscolonic biopsy cytokine response to mechanical stress in the polarised in vitro organ culture system, as well as in response to bacterial ligand stress. Albeit, supraphysiological levels of LPS, PGN and MDP were used most likely

required due to the presence of a mucus layer. Further work looking at the response to TLR3 ligands such as Poly(I:C) or other PRR ligands and inflammasome inducers would elucidate the role of LPXN in the colonic mucosa further. Given the data of SNPs working in concert, analysis of the several SNPs that affect the focal adhesion complex, in the pIVOC system would be advantageous to confirm this cumulative effect leading to a stronger phenotype. This however would require a significantly larger study, with both controls and UC patients. The focal adhesion complex provides a wealth of biomarkers for the oncology field (422), further work is needed to identify its role in the pathogenesis of inflammatory diseases such as UC, and thereby identifying potential biomarkers of disease progression, or treatment response.

To extend this work further, we showed that it was possible to identify and visualize patient specific genetic footprints, and consistent with the literature have shown that it is the first neighbours (423), the protein-protein interactions who are hidden players in the disease pathogenesis. We have identified MAML2 as a potential severity marker in patients who don't have UC associated NFKB1 SNPs, which via the Notch pathway can activate NFKB1. This work would benefit from validation in a larger patient cohort, which has been undertaken. The work has also shown that despite the thousands on dynamic interactions that occur within a cell at any one time, the seemingly disparate SNP affected proteins converge on key regulatory proteins, some of which, like NFKB1, IL10, IL12a and IL23 are well known and explored, others such as the IRF family, RIPK2 require more work to characterise their role in UC. By clustering the patients based on their genetics, we have identified, in a small cohort, the minimum SNP burden and genetic footprint that leads to UC as shown in Cluster 3.

We have also clearly shown that viral handling is a pervasive, significant theme from the gene enrichment across the entire patient cohort. Work recently published has identified Epstein Barr Virus as present within the colonic mucosa of UC patients, irrespective of their disease severity or treatment course(143). Further work identifying the mechanism of host genetics-virus interaction would be beneficial to understand this aspect of the disease instigation.

This work would benefit from larger cohort studies, as the numbers were not large enough to make any significant contribution to identifying prognostic indicators or markers of disease severity, this is currently being undertaken. Nor was it large enough to identify drug repurposing targets for the majority of UC patients. It would also benefit

from large control cohorts to confirm that the pathways identified are valid and not just a function of the method used.

This work is novel as it attempts to take a global overview of complex GWAS data and place the SNPS into a physiological, dynamic system in an attempt to explain genetic susceptibility, explain inter-person variability of disease processes and identify hidden players in the pathogenesis of UC that could be used for either biomarker identification or drug targeted/repurposing. From a pathogenesis perspective, the systems biology approach has provided a wealth of potential hypothesis driven avenues to pursue from the role of non enteropathogenic viruses such as EBV in the instigation of UC, the role of the focal adhesion complex in UC, to the role of the SNPS in colitis associated cancer. From a personalised medicine perspective, the ability to create individual patient's genetic footprints opens the way for individualised drug targeting and disease modulation. The approach has identified the importance of combining multiple modalities, to gain a clearer, yet more complex overview of the genetic landscape of complex diseases.

6. References

1. Pabst O. Trafficking of regulatory T cells in the intestinal immune system. Int Immunol. 2013;25(3):139-43.

2. Gonnella PA, Chen Y, Inobe J, Komagata Y, Quartulli M, Weiner HL. In situ immune response in gut-associated lymphoid tissue (GALT) following oral antigen in TCR-transgenic mice. J Immunol. 1998;160(10):4708-18.

3. Solberg IC, Lygren I, Jahnsen J, Aadland E, Hoie O, Cvancarova M, et al. Clinical course during the first 10 years of ulcerative colitis: results from a population-based inception cohort (IBSEN Study). Scand J Gastroenterol. 2009;44(4):431-40.

4. Truelove SC, Witts LJ. Cortisone in ulcerative colitis; preliminary report on a therapeutic trial. Br Med J. 1954;2(4884):375-8.

5. Ransford RA, Langman MJ. Sulphasalazine and mesalazine: serious adverse reactions re-evaluated on the basis of suspected adverse reaction reports to the Committee on Safety of Medicines. Gut. 2002;51(4):536-9.

6. Elseviers MM, D'Haens G, Lerebours E, Plane C, Stolear JC, Riegler G, et al. Renal impairment in patients with inflammatory bowel disease: association with aminosalicylate therapy? Clin Nephrol. 2004;61(2):83-9.

7. Patel VN, Kaelber DC. Using aggregated, de-identified electronic health record data for multivariate pharmacosurveillance: a case study of azathioprine. J Biomed Inform. 2014;52:36-42.

8. Long MD, Herfarth HH, Pipkin CA, Porter CQ, Sandler RS, Kappelman MD. Increased risk for non-melanoma skin cancer in patients with inflammatory bowel disease. Clin Gastroenterol Hepatol. 2010;8(3):268-74.

9. Toruner M, Loftus EV, Jr., Harmsen WS, Zinsmeister AR, Orenstein R, Sandborn WJ, et al. Risk factors for opportunistic infections in patients with inflammatory bowel disease. Gastroenterology. 2008;134(4):929-36.

10. Lichtenstein GR, Hanauer SB, Sandborn WJ. Risk of Biologic Therapy-Associated Progressive Multifocal Leukoencephalopathy: Use of the JC Virus Antibody Assay in the Treatment of Moderate-to-Severe Crohn's Disease. Gastroenterol Hepatol (N Y). 2012;8(11 Suppl 8):1-20.

11. McLean MH, Neurath MF, Durum SK. Targeting interleukins for the treatment of inflammatory bowel disease-what lies beyond anti-TNF therapy? Inflamm Bowel Dis. 2014;20(2):389-97.

12. Colombel JF, Sandborn WJ, Reinisch W, Mantzaris GJ, Kornbluth A, Rachmilewitz D, et al. Infliximab, azathioprine, or combination therapy for Crohn's disease. N Engl J Med. 2010;362(15):1383-95.

13. Kotlyar DS, Blonski W, Diamond RH, Wasik M, Lichtenstein GR. Hepatosplenic T-cell lymphoma in inflammatory bowel disease: a possible thiopurine-induced chromosomal abnormality. Am J Gastroenterol. 2010;105(10):2299-301.

14. Monsen U, Brostrom O, Nordenvall B, Sorstad J, Hellers G. Prevalence of inflammatory bowel disease among relatives of patients with ulcerative colitis. Scand J Gastroenterol. 1987;22(2):214-8.

15. Kobayashi K, Atoh M, Konoeda Y, Yagita A, Inoko H, Sekiguchi S. HLA-DR, DQ and T cell antigen receptor constant beta genes in Japanese patients with ulcerative colitis. Clin Exp Immunol. 1990;80(3):400-3.

16. Purrmann J, Bertrams J, Knapp M, Cleveland S, Gemsa R, Hengels KJ, et al. Investigation of genetic markers in patients with Crohn's disease and ulcerative colitis. Z Gastroenterol. 1989;27(7):366-9.

17. Sugimura K, Asakura H, Mizuki N, Inoue M, Hibi T, Yagita A, et al. Analysis of genes within the HLA region affecting susceptibility to ulcerative colitis. Hum Immunol. 1993;36(2):112-8.

18. Yang H, Rotter JI, Toyoda H, Landers C, Tyran D, McElree CK, et al. Ulcerative colitis: a genetically heterogeneous disorder defined by genetic (HLA class II) and subclinical (antineutrophil cytoplasmic antibodies) markers. J Clin Invest. 1993;92(2):1080-4.

19. de la Concha EG, Arroyo R, Crusius JB, Campillo JA, Martin C, Varela de Seijas E, et al. Combined effect of HLA-DRB1*1501 and interleukin-1 receptor antagonist gene allele 2 in susceptibility to relapsing/remitting multiple sclerosis. J Neuroimmunol. 1997;80(1-2):172-8.

20. Bouma G, Oudkerk Pool M, Crusius JB, Schreuder GM, Hellemans HP, Meijer BU, et al. Evidence for genetic heterogeneity in inflammatory bowel disease (IBD); HLA genes in the predisposition to suffer from ulcerative colitis (UC) and Crohn's disease (CD). Clin Exp Immunol. 1997;109(1):175-9.

21. Lee KW, Steiner N, Hurley CK. Clarification of HLA-B serologically ambiguous types by automated DNA sequencing. Tissue Antigens. 1998;51(5):536-40.

22. Satsangi J, Landers CJ, Welsh KI, Koss K, Targan S, Jewell DP. The presence of anti-neutrophil antibodies reflects clinical and genetic heterogeneity within inflammatory bowel disease. Inflamm Bowel Dis. 1998;4(1):18-26.

23. Mansfield JC, Holden H, Tarlow JK, Di Giovine FS, McDowell TL, Wilson AG, et al. Novel genetic association between ulcerative colitis and the anti-inflammatory cytokine interleukin-1 receptor antagonist. Gastroenterology. 1994;106(3):637-42.

24. Bioque G, Crusius JB, Koutroubakis I, Bouma G, Kostense PJ, Meuwissen SG, et al. Allelic polymorphism in IL-1 beta and IL-1 receptor antagonist (IL-1Ra) genes in inflammatory bowel disease. Clin Exp Immunol. 1995;102(2):379-83.

25. Heresbach D, Alizadeh M, Bretagne JF, Dabadie A, Colombel JF, Pagenault M, et al. TAP gene transporter polymorphism in inflammatory bowel diseases. Scand J Gastroenterol. 1997;32(10):1022-7.

26. Parkes M, Satsangi J, Jewell D. Contribution of the IL-2 and IL-10 genes to inflammatory bowel disease (IBD) susceptibility. Clin Exp Immunol. 1998;113(1):28-32.

27. Yang H. Analysis of ICAM-1 gene polymorphism in immunologic subsets of inflammatory bowel disease. Exp Clin Immunogenet. 1997;14(3):214-25.

28. Kyo K, Parkes M, Takei Y, Nishimori H, Vyas P, Satsangi J, et al. Association of ulcerative colitis with rare VNTR alleles of the human intestinal mucin gene, MUC3. Hum Mol Genet. 1999;8(2):307-11.

29. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet. 1996;58(6):1347-63.

30. Satsangi J, Parkes M, Jewell DP, Bell JI. Genetics of inflammatory bowel disease. Clin Sci. 1998;94(5):473-8.

31. Mirza MM, Lee J, Teare D, Hugot JP, Laurent-Puig P, Colombel JF, et al. Evidence of linkage of the inflammatory bowel disease susceptibility locus on chromosome 16 (IBD1) to ulcerative colitis. J Med Genet. 1998;35(3):218-21.

32. Duerr RH, Barmada MM, Zhang L, Davis S, Preston RA, Chensny LJ, et al. Linkage and association between inflammatory bowel disease and a locus on chromosome 12. Am J Hum Genet. 1998;63(1):95-100.

33. Cho JH, Nicolae DL, Gold LH, Fields CT, LaBuda MC, Rohal PM, et al. Identification of novel susceptibility loci for inflammatory bowel disease on chromosomes 1p, 3q, and 4q: evidence for epistasis between 1p and IBD1. Proc Natl Acad Sci U S A. 1998;95(13):7502-7.

34. Stokkers PC, Huibregtse K, Jr., Leegwater AC, Reitsma PH, Tytgat GN, van Deventer SJ. Analysis of a positional candidate gene for inflammatory bowel disease: NRAMP2. Inflamm Bowel Dis. 2000;6(2):92-8.

35. Duerr RH, Barmada MM, Zhang L, Pfutzer R, Weeks DE. High-density genome scan in Crohn disease shows confirmed linkage to chromosome 14q11-12. Am J Hum Genet. 2000;66(6):1857-62.

36. van Rheenen PF, Van de Vijver E, Fidler V. Faecal calprotectin for screening of patients with suspected inflammatory bowel disease: diagnostic meta-analysis. BMJ. 2010;341:c3369.

37. Health NIf. The Human Genome Project. 2003.

38. The International HapMap Project. 2005.

39. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012;491(7422):119-24.

40. Lees CW, Satsangi J. Genetics of inflammatory bowel disease: implications for disease pathogenesis and natural history. Expert Rev Gastroenterol Hepatol. 2009;3(5):513-34.

41. Lees CW, Barrett JC, Parkes M, Satsangi J. New IBD genetics: common pathways with other diseases. Gut. 2011;60(12):1739-53.

42. Luo Y, de Lange KM, Jostins L, Moutsianas L, Randall J, Kennedy NA, et al. Exploring the genetic architecture of inflammatory bowel disease by whole-genome sequencing identifies association at ADCY7. Nat Genet. 2017;49(2):186-92.

43. Huang H, Fang M, Jostins L, Umicevic Mirkov M, Boucher G, Anderson CA, et al. Fine-mapping inflammatory bowel disease loci to single-variant resolution. Nature. 2017;547(7662):173-8.

44. Silverberg MS, Cho JH, Rioux JD, McGovern DPB, Wu J, Annese V, et al. Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. Nature Genetics. 2009;41(2):216-20.

45. Yang SK, Jung Y, Kim H, Hong M, Ye BD, Song K. Association of FCGR2A, JAK2 or HNF4A variants with ulcerative colitis in Koreans. Digestive and Liver Disease. 2011;43(11):856-61.

46. Sandborn WJ, Ghosh S, Panes J, Vranic I, Su C, Rousell S, et al. Tofacitinib, an oral Janus kinase inhibitor, in active ulcerative colitis. N Engl J Med. 2012;367(7):616-24.

47. Cohen LB, Nanau RM, Delzor F, Neuman MG. Biologic therapies in inflammatory bowel disease. Transl Res. 2014;163(6):533-56.

48. Cardinale CJ, Wei Z, Li J, Zhu J, Gu M, Baldassano RN, et al. Transcriptome profiling of human ulcerative colitis mucosa reveals altered expression of pathways enriched in genetic susceptibility Loci. PLoS One. 2014;9(5):e96153.

49. Hasler R, Feng Z, Backdahl L, Spehlmann ME, Franke A, Teschendorff A, et al. A functional methylome map of ulcerative colitis. Genome Research. 2012;22(11):2130-7.

50. Lepage P, Hasler R, Spehlmann ME, Rehman A, Zvirbliene A, Begun A, et al. Twin Study Indicates Loss of Interaction Between Microbiota and Mucosa of Patients With Ulcerative Colitis. Gastroenterology. 2011;141(1):227-36.

51. Gensollen T, Iyer SS, Kasper DL, Blumberg RS. How colonization by microbiota in early life shapes the immune system. Science. 2016;352(6285):539-44.

52. Wang MH, Fiocchi C, Zhu X, Ripke S, Kamboh MI, Rebert N, et al. Gene-gene and gene-environment interactions in ulcerative colitis. Hum Genet. 2014;133(5):547-58.

53. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease.

Nature. 2012;491(7422):119-24 %8 Nov %! Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease %@ 1476-4687.

54. Peters LA, Perrigoue J, Mortha A, Iuga A, Song WM, Neiman EM, et al. A functional genomics predictive network model identifies regulators of inflammatory bowel disease. Nat Genet. 2017;49(10):1437-49.

55. Sonnenburg JL, Backhed F. Diet-microbiota interactions as moderators of human metabolism. Nature. 2016;535(7610):56-64.

56. Boccia S, Torre I, Santarpia L, Iervolino C, Del Piano C, Puggina A, et al. Intestinal microbiota in adult patients with Short Bowel Syndrome: Preliminary results from a pilot study. Clin Nutr. 2016.

57. Mittal R, Debs LH, Patel AP, Nguyen D, Patel K, O'Connor G, et al. Neurotransmitters: The Critical Modulators Regulating Gut-Brain Axis. Journal of cellular physiology. 2016.

58. Schroeder BO, Backhed F. Signals from the gut microbiota to distant organs in physiology and disease. Nat Med. 2016;22(10):1079-89.

59. Lankelma JM, Belzer C, Hoogendijk AJ, de Vos AF, de Vos WM, van der Poll T, et al. Antibiotic-Induced Gut Microbiota Disruption Decreases TNF-alpha Release by Mononuclear Cells in Healthy Adults. Clin Transl Gastroenterol. 2016;7(8):e186.

60. Fujimura KE, Sitarik AR, Havstad S, Lin DL, Levan S, Fadrosh D, et al. Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation. Nat Med. 2016;22(10):1187-91.

61. Bernstein CN. Antibiotic use and the risk of Crohn's disease. Gastroenterol Hepatol (N Y). 2013;9(6):393-5.

62. Kronman MP, Zaoutis TE, Haynes K, Feng R, Coffin SE. Antibiotic exposure and IBD development among children: a population-based cohort study. Pediatrics. 2012;130(4):e794-803.

63. Lucke K, Miehlke S, Jacobs E, Schuppler M. Prevalence of Bacteroides and Prevotella spp. in ulcerative colitis. J Med Microbiol. 2006;55(Pt 5):617-24.

64. Sasaki I, Funayama Y, Fukushima K, Watanabe K. [Recent progress in surgical treatment of IBD]. Nihon Rinsho. 2012;70 Suppl 1:31-6.

65. Johansson ME, Larsson JM, Hansson GC. The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. Proc Natl Acad Sci U S A. 2011;108 Suppl 1:4659-65.

66. Van der Sluis M, De Koning BA, De Bruijn AC, Velcich A, Meijerink JP, Van Goudoever JB, et al. Muc2-deficient mice spontaneously develop colitis, indicating that MUC2 is critical for colonic protection. Gastroenterology. 2006;131(1):117-29.

67. Velcich A, Yang W, Heyer J, Fragale A, Nicholas C, Viani S, et al. Colorectal cancer in mice genetically deficient in the mucin Muc2. Science. 2002;295(5560):1726-9.

68. Strugala V, Dettmar PW, Pearson JP. Thickness and continuity of the adherent colonic mucus barrier in active and quiescent ulcerative colitis and Crohn's disease. Int J Clin Pract. 2008;62(5):762-9.

69. Png CW, Linden SK, Gilshenan KS, Zoetendal EG, McSweeney CS, Sly LI, et al. Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. Am J Gastroenterol. 2010;105(11):2420-8.

70. Breslin NP, Nash C, Hilsden RJ, Hershfield NB, Price LM, Meddings JB, et al. Intestinal permeability is increased in a proportion of spouses of patients with Crohn's disease. Am J Gastroenterol. 2001;96(10):2934-8.

71. Peeters M, Geypens B, Claus D, Nevens H, Ghoos Y, Verbeke G, et al. Clustering of increased small intestinal permeability in families with Crohn's disease. Gastroenterology. 1997;113(3):802-7.

72. Söderholm JD, Olaison G, Lindberg E, Hannestad U, Vindels A, Tysk C, et al. Different intestinal permeability patterns in relatives and spouses of patients with Crohn's disease: an inherited defect in mucosal defence? Gut. 1999;44(1):96-100.

73. Huang C, Wang JJ, Jing G, Li J, Jin C, Yu Q, et al. Erp29 Attenuates Cigarette Smoke Extract-Induced Endoplasmic Reticulum Stress and Mitigates Tight Junction Damage in Retinal Pigment Epithelial Cells. Invest Ophthalmol Vis Sci. 2015;56(11):6196-207.

74. Gassler N, Rohr C, Schneider A, Kartenbeck J, Bach A, Obermuller N, et al. Inflammatory bowel disease is associated with changes of enterocytic junctions. Am J Physiol Gastrointest Liver Physiol. 2001;281(1):G216-28.

75. Wyatt J, Vogelsang H, Hubl W, Waldhoer T, Lochs H. Intestinal permeability and the prediction of relapse in Crohn's disease. Lancet. 1993;341(8858):1437-9.

76. Youakim A, Ahdieh M. Interferon-gamma decreases barrier function in T84 cells by reducing ZO-1 levels and disrupting apical actin. Am J Physiol. 1999;276(5 Pt 1):G1279-88.

77. Al-Sadi R, Guo S, Ye D, Ma TY. TNF-alpha modulation of intestinal epithelial tight junction barrier is regulated by ERK1/2 activation of Elk-1. Am J Pathol. 2013;183(6):1871-84.

78. Edelblum KL, Turner JR. The tight junction in inflammatory disease: communication breakdown. Curr Opin Pharmacol. 2009;9(6):715-20.

79. Jiao X, He P, Li Y, Fan Z, Si M, Xie Q, et al. The Role of Circulating Tight Junction Proteins in Evaluating Blood Brain Barrier Disruption following Intracranial Hemorrhage. Dis Markers. 2015;2015:860120.

80. Ma TY, Iwamoto GK, Hoa NT, Akotia V, Pedram A, Boivin MA, et al. TNF-alpha-induced increase in intestinal epithelial tight junction permeability requires NF-kappa B activation. Am J Physiol Gastrointest Liver Physiol. 2004;286(3):G367-76.

81. Cao B, Zhou X, Ma J, Zhou W, Yang W, Fan D, et al. Role of MiRNAs in Inflammatory Bowel Disease. Dig Dis Sci. 2017;62(6):1426-38.

82. Cario E. Toll-like receptors in inflammatory bowel diseases: a decade later. Inflamm Bowel Dis. 2010;16(9):1583-97.

83. Frantz AL, Rogier EW, Weber CR, Shen L, Cohen DA, Fenton LA, et al. Targeted deletion of MyD88 in intestinal epithelial cells results in compromised antibacterial immunity associated with downregulation of polymeric immunoglobulin receptor, mucin-2, and antibacterial peptides. Mucosal Immunol. 2012;5(5):501-12.

84. Kubler I, Koslowski MJ, Gersemann M, Fellermann K, Beisner J, Becker S, et al. Influence of standard treatment on ileal and colonic antimicrobial defensin expression in active Crohn's disease. Aliment Pharmacol Ther. 2009;30(6):621-33.

85. Zilbauer M, Jenke A, Wenzel G, Goedde D, Postberg J, Phillips AD, et al. Intestinal alpha-defensin expression in pediatric inflammatory bowel disease. Inflamm Bowel Dis. 2011;17(10):2076-86.

86. Kalus AA, Fredericks LP, Hacker BM, Dommisch H, Presland RB, Kimball JR, et al. Association of a genetic polymorphism (-44 C/G SNP) in the human DEFB1 gene with expression and inducibility of multiple beta-defensins in gingival keratinocytes. BMC Oral Health. 2009;9:21.

87. Kocsis AK, Lakatos PL, Somogyvari F, Fuszek P, Papp J, Fischer S, et al. Association of beta-defensin 1 single nucleotide polymorphisms with Crohn's disease. Scand J Gastroenterol. 2008;43(3):299-307.

88. Wehkamp J, Harder J, Weichenthal M, Mueller O, Herrlinger KR, Fellermann K, et al. Inducible and constitutive beta-defensins are differentially expressed in Crohn's disease and ulcerative colitis. Inflamm Bowel Dis. 2003;9(4):215-23.

89. Jones GR, Kennedy NA, Lees CW, Arnott ID, Satsangi J. Letter: faecal calprotectin and lactoferrin - accurate biomarkers in post-operative Crohn's disease - authors' reply. Letter: biologic therapies are effective for prevention of post-operative Crohn's disease recurrence - authors' reply. Aliment Pharmacol Ther. 2014;40(3):323.

90. Dai J, Liu WZ, Zhao YP, Hu YB, Ge ZZ. Relationship between fecal lactoferrin and inflammatory bowel disease. Scand J Gastroenterol. 2007;42(12):1440-4.

91. Lamb CA, Mansfield JC. Measurement of faecal calprotectin and lactoferrin in inflammatory bowel disease. Frontline Gastroenterol. 2011;2(1):13-8.

92. Zhou XL, Xu W, Tang XX, Luo LS, Tu JF, Zhang CJ, et al. Fecal lactoferrin in discriminating inflammatory bowel disease from irritable bowel syndrome: a diagnostic meta-analysis. BMC Gastroenterol. 2014;14:121.

93. Bouma G, Strober W. The immunological and genetic basis of inflammatory bowel disease. Nat Rev Immunol. 2003;3(7):521-33.

94. Vamadevan AS, Fukata M, Arnold ET, Thomas LS, Hsu D, Abreu MT. Regulation of Toll-like receptor 4-associated MD-2 in intestinal epithelial cells: a comprehensive analysis. Innate Immun. 2010;16(2):93-103.

95. Eaves-Pyles T, Bu HF, Tan XD, Cong Y, Patel J, Davey RA, et al. Luminal-applied flagellin is internalized by polarized intestinal epithelial cells and elicits immune responses via the TLR5 dependent mechanism. PLoS One. 2011;6(9):e24869.

96. Lee J, Mo JH, Katakura K, Alkalay I, Rucker AN, Liu YT, et al. Maintenance of colonic homeostasis by distinctive apical TLR9 signalling in intestinal epithelial cells. Nat Cell Biol. 2006;8(12):1327-36.

97. Gasteiger G, Fan X, Dikiy S, Lee SY, Rudensky AY. Tissue residency of innate lymphoid cells in lymphoid and nonlymphoid organs. Science. 2015;350(6263):981-5.

98. Bahrami B, Child MW, Macfarlane S, Macfarlane GT. Adherence and cytokine induction in Caco-2 cells by bacterial populations from a three-stage continuous-culture model of the large intestine. Appl Environ Microbiol. 2011;77(9):2934-42.

99. Wells JM, Rossi O, Meijerink M, van Baarlen P. Epithelial crosstalk at the microbiota-mucosal interface. Proc Natl Acad Sci U S A. 2011;108 Suppl 1:4607-14.

100. Hepworth MR, Monticelli LA, Fung TC, Ziegler CG, Grunberg S, Sinha R, et al. Innate lymphoid cells regulate CD4+ T-cell responses to intestinal commensal bacteria. Nature. 2013;498(7452):113-7.

101. Waggoner SN, Kumar V. Evolving role of 2B4/CD244 in T and NK cell responses during virus infection. Front Immunol. 2012;3:377.

102. Lang PA, Lang KS, Xu HC, Grusdat M, Parish IA, Recher M, et al. Natural killer cell activation enhances immune pathology and promotes chronic infection by limiting CD8+ T-cell immunity. Proc Natl Acad Sci U S A. 2012;109(4):1210-5.

103. Jarry A, Bossard C, Bou-Hanna C, Masson D, Espaze E, Denis MG, et al. Mucosal IL-10 and TGF-beta play crucial roles in preventing LPS-driven, IFN-gamma-mediated epithelial damage in human colon explants. J Clin Invest. 2008;118(3):1132-42.

104. Tomasello E, Pollet E, Vu Manh TP, Uze G, Dalod M. Harnessing Mechanistic Knowledge on Beneficial Versus Deleterious IFN-I Effects to Design Innovative Immunotherapies Targeting Cytokine Activity to Specific Cell Types. Front Immunol. 2014;5:526.

105. Coombes JL, Powrie F. Dendritic cells in intestinal immune regulation. Nat Rev Immunol. 2008;8(6):435-46.

106. Bell SJ, Rigby R, English N, Mann SD, Knight SC, Kamm MA, et al. Migration and maturation of human colonic dendritic cells. J Immunol. 2001;166(8):4958-67.

107. Yamazaki S, Dudziak D, Heidkamp GF, Fiorese C, Bonito AJ, Inaba K, et al. CD8+ CD205+ splenic dendritic cells are specialized to induce Foxp3+ regulatory T cells. J Immunol. 2008;181(10):6923-33.

108. Stagg AJ, Kamm MA, Knight SC. Intestinal dendritic cells increase T cell expression of alpha4beta7 integrin. Eur J Immunol. 2002;32(5):1445-54.

109. Mora JR, Iwata M, Eksteen B, Song SY, Junt T, Senman B, et al. Generation of gut-homing IgA-secreting B cells by intestinal dendritic cells. Science. 2006;314(5802):1157-60.

110. Follows GA, Munk ME, Gatrill AJ, Conradt P, Kaufmann SH. Gamma interferon and interleukin 2, but not interleukin 4, are detectable in gamma/delta T-cell cultures after activation with bacteria. Infect Immun. 1992;60(3):1229-31.

111. Barnes PF, Abrams JS, Lu S, Sieling PA, Rea TH, Modlin RL. Patterns of cytokine production by mycobacterium-reactive human T-cell clones. Infect Immun. 1993;61(1):197-203.

112. Kober OI, Ahl D, Pin C, Holm L, Carding SR, Juge N. gammadelta T-cell-deficient mice show alterations in mucin expression, glycosylation, and goblet cells but maintain an intact mucus layer. Am J Physiol Gastrointest Liver Physiol. 2014;306(7):G582-93.

113. Probert CS, Chott A, Turner JR, Saubermann LJ, Stevens AC, Bodinaku K, et al. Persistent clonal expansions of peripheral blood CD4+ lymphocytes in chronic inflammatory bowel disease. J Immunol. 1996;157(7):3183-91.

114. Hueber W, Sands BE, Lewitzky S, Vandemeulebroecke M, Reinisch W, Higgins PD, et al. Secukinumab, a human anti-IL-17A monoclonal antibody, for moderate to severe Crohn's disease: unexpected results of a randomised, double-blind placebo-controlled trial. Gut. 2012;61(12):1693-700.

115. Symons A, Budelsky AL, Towne JE. Are Th17 cells in the gut pathogenic or protective? Mucosal Immunol. 2012;5(1):4-6.

116. Zenewicz LA, Antov A, Flavell RA. CD4 T-cell differentiation and inflammatory bowel disease. Trends Mol Med. 2009;15(5):199-207.

117. Park JY, Kim HY, Lee JY, Kim KH, Jang MK, Lee JH, et al. Macrolide-affected Toll-like receptor 4 expression from Helicobacter pylori-infected monocytes does not modify interleukin-8 production. FEMS Immunol Med Microbiol. 2005;44(2):171-6.

118. Witowski J, Pawlaczyk K, Breborowicz A, Scheuren A, Kuzlan-Pawlaczyk M, Wisniewska J, et al. IL-17 stimulates intraperitoneal neutrophil infiltration through the release of GRO alpha chemokine from mesothelial cells. J Immunol. 2000;165(10):5814-21.

119. Kim MS, Kim WS, Piao ZH, Yun S, Lee SH, Lee S, et al. IL-22 producing NKp46+ innate lymphoid cells can differentiate from hematopoietic precursor cells. Immunol Lett. 2011;141(1):61-7.

120. Rendon JL, Li X, Akhtar S, Choudhry MA. Interleukin-22 modulates gut epithelial and immune barrier functions following acute alcohol exposure and burn injury. Shock. 2013;39(1):11-8.

121. Kolls JK, McCray PB, Jr., Chan YR. Cytokine-mediated regulation of antimicrobial proteins. Nat Rev Immunol. 2008;8(11):829-35.

122. Zheng Y, Valdez PA, Danilenko DM, Hu Y, Sa SM, Gong Q, et al. Interleukin-22 mediates early host defense against attaching and effacing bacterial pathogens. Nat Med. 2008;14(3):282-9.

123. Cong Y, Feng T, Fujihashi K, Schoeb TR, Elson CO. A dominant, coordinated T regulatory cell-IgA response to the intestinal microbiota. Proc Natl Acad Sci U S A. 2009;106(46):19256-61.

124. Batista FD, Harwood NE. The who, how and where of antigen presentation to B cells. Nat Rev Immunol. 2009;9(1):15-27.

125. Amu S, Saunders SP, Kronenberg M, Mangan NE, Atzberger A, Fallon PG. Regulatory B cells prevent and reverse allergic airway inflammation via FoxP3-positive T regulatory cells in a murine model. J Allergy Clin Immunol. 2010;125(5):1114-24 e8.

126. Carter NA, Vasconcellos R, Rosser EC, Tulone C, Munoz-Suano A, Kamanaka M, et al. Mice lacking endogenous IL-10-producing regulatory B cells develop exacerbated disease and present with an increased frequency of Th1/Th17 but a decrease in regulatory T cells. J Immunol. 2011;186(10):5569-79.

127. Lippert C, Listgarten J, Davidson RI, Baxter S, Poon H, Kadie CM, et al. An exhaustive epistatic SNP association analysis on expanded Wellcome Trust data. Sci Rep. 2013;3:1099.

128. Heath RJ, Goel G, Baxt LA, Rush JS, Mohanan V, Paulus GLC, et al. RNF166 Determines Recruitment of Adaptor Proteins during Antibacterial Autophagy. Cell Rep. 2016;17(9):2183-94.

129. Tsuboi K, Nishitani M, Takakura A, Imai Y, Komatsu M, Kawashima H. Autophagy Protects against Colitis by the Maintenance of Normal Gut Microflora and Secretion of Mucus. J Biol Chem. 2015;290(33):20511-26.

130. Patel KK, Miyoshi H, Beatty WL, Head RD, Malvin NP, Cadwell K, et al. Autophagy proteins control goblet cell function by potentiating reactive oxygen species production. EMBO J. 2013;32(24):3130-44.

131. Lassen KG, Kuballa P, Conway KL, Patel KK, Becker CE, Peloquin JM, et al. Atg16L1 T300A variant decreases selective autophagy resulting in altered cytokine signaling and decreased antibacterial defense. Proc Natl Acad Sci U S A. 2014;111(21):7741-6.

132. Cooney R, Baker J, Brain O, Danis B, Pichulik T, Allan P, et al. NOD2 stimulation induces autophagy in dendritic cells influencing bacterial handling and antigen presentation. Nat Med. 2010;16(1):90-7.

133. Schuster AT, Homer CR, Kemp JR, Nickerson KP, Deutschman E, Kim Y, et al. Chromosome-associated protein D3 promotes bacterial clearance in human intestinal epithelial cells by repressing expression of amino acid transporters. Gastroenterology. 2015;148(7):1405-16 e3.

134. Hao J, Pei Y, Ji G, Li W, Feng S, Qiu S. Autophagy is induced by 3beta-O-succinyl-lupeol (LD9-4) in A549 cells via up-regulation of Beclin 1 and down-regulation mTOR pathway. Eur J Pharmacol. 2011;670(1):29-38.

135. Zhang R, Chi X, Wang S, Qi B, Yu X, Chen JL. The regulation of autophagy by influenza A virus. Biomed Res Int. 2014;2014:498083.

136. Wang L, Ou JH. Hepatitis C virus and autophagy. Biol Chem. 2015;396(11):1215-22.

137. Shrivastava S, Devhare P, Sujijantarat N, Steele R, Kwon YC, Ray R, et al. Knockdown of Autophagy Inhibits Infectious Hepatitis C Virus Release by the Exosomal Pathway. J Virol. 2015;90(3):1387-96.

138. Yakoub AM, Shukla D. Basal Autophagy Is Required for Herpes simplex Virus-2 Infection. Sci Rep. 2015;5:12985.

139. Norman JM, Handley SA, Baldridge MT, Droit L, Liu CY, Keller BC, et al. Disease-specific alterations in the enteric virome in inflammatory bowel disease. Cell. 2015;160(3):447-60.

140. Masclee GM, Penders J, Pierik M, Wolffs P, Jonkers D. Enteropathogenic viruses: triggers for exacerbation in IBD? A prospective cohort study using real-time quantitative polymerase chain reaction. Inflamm Bowel Dis. 2013;19(1):124-31.

141. Cadwell K, Patel KK, Maloney NS, Liu TC, Ng AC, Storer CE, et al. Virus-plus-susceptibility gene interaction determines Crohn's disease gene Atg16L1 phenotypes in intestine. Cell. 2010;141(7):1135-45.

142. Kolho KL, Klemola P, Simonen-Tikka ML, Ollonen ML, Roivainen M. Enteric viral pathogens in children with inflammatory bowel disease. J Med Virol. 2012;84(2):345-7.

143. Lopes S, Andrade P, Conde S, Liberal R, Dias CC, Fernandes S, et al. Looking into Enteric Virome in Patients with IBD: Defining Guilty or Innocence? Inflamm Bowel Dis. 2017;23(8):1278-84.

144. Severa M, Giacomini E, Gafa V, Anastasiadou E, Rizzo F, Corazzari M, et al. EBV stimulates TLR- and autophagy-dependent pathways and impairs maturation in plasmacytoid dendritic cells: implications for viral immune escape. Eur J Immunol. 2013;43(1):147-58.

145. Granato M, Santarelli R, Farina A, Gonnella R, Lotti LV, Faggioni A, et al. Epstein-barr virus blocks the autophagic flux and appropriates the autophagic machinery to enhance viral replication. J Virol. 2014;88(21):12715-26.

146. Shannon-Lowe C, Rowe M. Epstein-Barr virus infection of polarized epithelial cells via the basolateral surface by memory B cell-mediated transfer infection. PLoS Pathog. 2011;7(5):e1001338.

147. Geiger B, Spatz JP, Bershadsky AD. Environmental sensing through focal adhesions. Nat Rev Mol Cell Bio. 2009;10(1):21-33.

148. MacGillivray MK, Cruz TF, McCulloch CA. The recruitment of the interleukin-1 (IL-1) receptor-associated kinase (IRAK) into focal adhesion complexes is required for IL-1beta -induced ERK activation. J Biol Chem. 2000;275(31):23509-15.

149. Wang Q, Wang Y, Downey GP, Plotnikov S, McCulloch CA. A ternary complex comprising FAK, PTPalpha and IP3 receptor 1 functionally engages focal adhesions and the endoplasmic reticulum to mediate IL-1-induced Ca2+ signalling in fibroblasts. Biochem J. 2016;473(4):397-410.

150. Dixit N, Kim MH, Rossaint J, Yamayoshi I, Zarbock A, Simon SI. Leukocyte function antigen-1, kindlin-3, and calcium flux orchestrate neutrophil recruitment during inflammation. J Immunol. 2012;189(12):5954-64.

151. Klimova Z, Braborec V, Maninova M, Caslavsky J, Weber MJ, Vomastek T. Symmetry breaking in spreading RAT2 fibroblasts requires the MAPK/ERK pathway scaffold RACK1 that integrates FAK, p190A-RhoGAP and ERK2 signaling. Biochim Biophys Acta. 2016;1863(9):2189-200.

152. Bouchard V, Harnois C, Demers MJ, Thibodeau S, Laquerre V, Gauthier R, et al. B1 integrin/Fak/Src signaling in intestinal epithelial crypt cell survival: integration of complex regulatory mechanisms. Apoptosis. 2008;13(4):531-42.

153. Yu Y, Wu J, Wang Y, Zhao T, Ma B, Liu Y, et al. Kindlin 2 forms a transcriptional complex with beta-catenin and TCF4 to enhance Wnt signalling. EMBO Rep. 2012;13(8):750-8.

154. Beausejour M, Noel D, Thibodeau S, Bouchard V, Harnois C, Beaulieu JF, et al. Integrin/Fak/Src-mediated regulation of cell survival and anoikis in human intestinal epithelial crypt cells: selective engagement and roles of PI3-K isoform complexes. Apoptosis. 2012;17(6):566-78.

155. Johnson HE, King SJ, Asokan SB, Rotty JD, Bear JE, Haugh JM. F-actin bundles direct the initiation and orientation of lamellipodia through adhesion-based signaling. J Cell Biol. 2015;208(4):443-55.

156. Fuste NP, Fernandez-Hernandez R, Cemeli T, Mirantes C, Pedraza N, Rafel M, et al. Cytoplasmic cyclin D1 regulates cell invasion and metastasis through the phosphorylation of paxillin. Nat Commun. 2016;7:11581.

157. Shi J, Wu WJ, Hu G, Yu X, Yu GS, Lu H, et al. Regulation of beta-catenin transcription activity by leupaxin in hepatocellular carcinoma. Tumour Biol. 2016;37(2):2313-20.

158. May M, Wang TB, Muller M, Genth H. Difference in F-Actin Depolymerization Induced by Toxin B from the Clostridium difficile Strain VPI 10463 and Toxin B from the Variant Clostridium difficile Serotype F Strain 1470. Toxins. 2013;5(1):106-19.

159. Khan MRI, Yazawa T, Anisuzzaman ASM, Semba S, Ma YJ, Uwada J, et al. Activation of focal adhesion kinase via M1 muscarinic acetylcholine receptor is required in restitution of intestinal barrier function after epithelial injury. Bba-Mol Basis Dis. 2014;1842(4):635-45.

160. Guo S, Nighot M, Al-Sadi R, Alhmoud T, Nighot P, Ma TY. Lipopolysaccharide Regulation of Intestinal Tight Junction Permeability Is Mediated by TLR4 Signal Transduction Pathway Activation of FAK and MyD88. J Immunol. 2015;195(10):4999-5010 %8 Nov %! Lipopolysaccharide Regulation of Intestinal Tight Junction Permeability Is Mediated by TLR4 Signal Transduction Pathway Activation of FAK and MyD88 %@ 1550-6606.

161. Eitel J, Meixenberger K, van Laak C, Orlovski C, Hocke A, Schmeck B, et al. Rac1 Regulates the NLRP3 Inflammasome Which Mediates IL-1beta Production in Chlamydophila pneumoniae Infected Human Mononuclear Cells. Plos One. 2012;7(1).

162. Spalinger MR, Kasper S, Gottier C, Lang S, Atrott K, Vavricka SR, et al. NLRP3 tyrosine phosphorylation is controlled by protein tyrosine phosphatase PTPN22. Journal of Clinical Investigation. 2016;126(5):1783-800.

163. Elinav E, Strowig T, Kau AL, Henao-Mejia J, Thaiss CA, Booth CJ, et al. NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. Cell. 2011;145(5):745-57.

164. de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nat Genet. 2017;49(2):256-61.

165. Venkatesan K, Rual JF, Vazquez A, Stelzl U, Lemmens I, Hirozane-Kishikawa T, et al. An empirical framework for binary interactome mapping. Nat Methods. 2009;6(1):83-90.

166. Nabhan AR, Sarkar IN. Structural network analysis of biological networks for assessment of potential disease model organisms. J Biomed Inform. 2014;47:178-91.

167. Ali S, Nafis S, Kalaiarasan P, Rai E, Sharma S, Bamezai RN. Understanding Genetic Heterogeneity in Type 2 Diabetes by Delineating Physiological Phenotypes: SIRT1 and its Gene Network in Impaired Insulin Secretion. Rev Diabet Stud. 2016;13(1):17-34.

168. Donn R, De Leonibus C, Meyer S, Stevens A. Network analysis and juvenile idiopathic arthritis (JIA): a new horizon for the understanding of disease pathogenesis and therapeutic target identification. Pediatr Rheumatol Online J. 2016;14(1):40.

169. Wang Z, Liu TA, Lin ZW, Hegarty J, Koltun WA, Wu RL. A General Model for Multilocus Epistatic Interactions in Case-Control Studies. Plos One. 2010;5(8 %8 Aug 18 %! A General Model for Multilocus Epistatic Interactions in Case-Control Studies).

170. Diegelmann J, Czamara D, Le Bras E, Zimmermann E, Olszak T, Bedynek A, et al. Intestinal DMBT1 expression is modulated by Crohn's disease-associated IL23R variants and by a DMBT1 variant which influences binding of the transcription factors CREB1 and ATF-2. PLoS One. 2013;8(11):e77773.

171. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature. 2014.

172. Anderson CA, Boucher G, Lees CW, Franke A, D'Amato M, Taylor KD, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. Nature Genetics. 2011;43(3):246-U94.

173. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003;31(13):3812-4.

174. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in

the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6(2):80-92.

175. Dayem Ullah AZ, Lemoine NR, Chelala C. SNPnexus: a web server for functional annotation of novel and publicly known genetic variants (2012 update). Nucleic Acids Res. 2012;40(Web Server issue):W65-70.

176. Yamazaki K, Onouchi Y, Takazoe M, Kubo M, Nakamura Y, Hata A. Association analysis of genetic variants in IL23R, ATG16L1 and 5p13.1 loci with Crohn's disease in Japanese patients. J Hum Genet. 2007;52(7):575-83 %8 Jun %! Association analysis of genetic variants in IL23R, ATG16L1 and 5p13.1 loci with Crohn's disease in Japanese patients.

177. Palomino-Morales RJ, Oliver J, Gomez-Garcia M, Lopez-Nevot MA, Rodrigo L, Nieto A, et al. Association of ATG16L1 and IRGM genes polymorphisms with inflammatory bowel disease: a meta-analysis approach. Genes Immun. 2009;10(4):356-64 %8 Jun %! Association of ATG16L1 and IRGM genes polymorphisms with inflammatory bowel disease: a meta-analysis approach.

178. Dusatkova P, Hradsky O, Lenicek M, Bronsky J, Nevoral J, Kotalova R, et al. Association of IL23R p.381Gln and ATG16L1 p.197Ala With Crohn Disease in the Czech Population. J Pediatr Gastr Nutr. 2009;49(4):405-10 %8 Oct %! Association of IL23R p.381Gln and ATG16L1 p.197Ala With Crohn Disease in the Czech Population.

179. Morgan AR, Lam WJ, Han DY, Fraser AG, Ferguson LR. Association Analysis of ULK1 with Crohn's Disease in a New Zealand Population. Gastroenterology Research and Practice %! Association Analysis of ULK1 with Crohn's Disease in a New Zealand Population. 2012.

180. Hirano A, Yamazaki K, Umeno J, Ashikawa K, Aoki M, Matsumoto T, et al. Association Study of 71 European Crohn's Disease Susceptibility Loci in a Japanese Population. Inflammatory Bowel Diseases. 2013;19(3):526-33 %8 Mar %! Association Study of 71 European Crohn's Disease Susceptibility Loci in a Japanese Population.

181. Lakatos PL, Szamosi T, Szilvasi A, Molnar E, Lakatos L, Kovacs A, et al. ATG16L1 and IL23 receptor (IL23R) genes are associated with disease susceptibility in Hungarian

CD patients. Digest Liver Dis. 2008;40(11):867-73 %8 Nov %! ATG16L1 and IL23 receptor (IL23R) genes are associated with disease susceptibility in Hungarian CD patients.

182. Fowler EV, Doecke J, Simms LA, Zhao ZZ, Webb PM, Hayward NK, et al. ATG16L1 T300A Shows Strong Associations With Disease Subgroups in a Large Australian IBD Population: Further Support for Significant Disease Heterogeneity. American Journal of Gastroenterology. 2008;103(10):2519-26 %8 Oct %! ATG16L1 T300A Shows Strong Associations With Disease Subgroups in a Large Australian IBD Population: Further Support for Significant Disease Heterogeneity.

183. Amre DK, Mack DR, Morgan K, Krupoves A, Costea I, Lambrette P, et al. Autophagy Gene ATG16L1 But Not IRGM Is Associated with Crohn's Disease in Canadian Children. Inflammatory Bowel Diseases. 2009;15(4):501-7 %8 Apr %! Autophagy Gene ATG16L1 But Not IRGM Is Associated with Crohn's Disease in Canadian Children.

184. Huebner C, Petermann I, Lam WJ, Shelling AN, Ferguson LR. Characterization of Single-Nucleotide Polymorphisms Relevant to Inflammatory Bowel Disease in Commonly Used Gastrointestinal Cell Lines. Inflammatory Bowel Diseases. 2010;16(2):282-95 %8 Feb %! Characterization of Single-Nucleotide Polymorphisms Relevant to Inflammatory Bowel Disease in Commonly Used Gastrointestinal Cell Lines.

185. Hancock L, Beckly J, Geremia A, Cooney R, Cummings F, Pathan S, et al. Clinical and Molecular Characteristics of Isolated Colonic Crohn's Disease. Inflammatory Bowel Diseases. 2008;14(12):1667-77 %8 Dec %! Clinical and Molecular Characteristics of Isolated Colonic Crohn's Disease.

186. Plevy S, Silverberg MS, Lockton S, Stockfisch T, Croner L, Stachelski J, et al. Combined serological, genetic, and inflammatory markers differentiate non-IBD, Crohn's disease, and ulcerative colitis patients. Inflammatory Bowel Diseases. 2013;19(6):1139-48.

187. Weersma RK, Stokkers PCF, Cleynen I, Wolfkamp SCS, Henckaerts L, Schreiber S, et al. Confirmation of Multiple Crohn's Disease Susceptibility Loci in a Large Dutch-Belgian Cohort. American Journal of Gastroenterology. 2009;104(3):630-8.

188. Cummings JRF, Cooney R, Pathan S, Anderson CA, Barrett JC, Beckly J, et al. Confirmation of the role of ATG16L1 as a Crohn's disease susceptibility gene. Inflammatory Bowel Diseases. 2007;13(8):941-6.

189. Okazaki T, Wang MH, Rawsthorne P, Sargent M, Datta LW, Shugart YY, et al. Contributions of IBD5, IL23R, ATG16L1, and NOD2 to Crohn's Disease Risk in a Population-Based Case-Control Study: Evidence of Gene-Gene Interactions. Inflammatory Bowel Diseases. 2008;14(11):1528-41 %8 Nov %! Contributions of IBD5, IL23R, ATG16L1, and NOD2 to Crohn's Disease Risk in a Population-Based Case-Control Study: Evidence of Gene-Gene Interactions.

190. Peter I, Mitchell AA, Ozelius L, Erazo M, Hu JZ, Doheny D, et al. Evaluation of 22 genetic variants with Crohn's Disease risk in the Ashkenazi Jewish population: a case-control study. Bmc Med Genet. 2011;12 %8 May 6 %! Evaluation of 22 genetic variants with Crohn's Disease risk in the Ashkenazi Jewish population: a case-control study.

191. Parkes M. Evidence from Genetics for a Role of Autophagy and Innate Immunity in IBD Pathogenesis. Digestive Diseases. 2012;30(4):330-3 %! Evidence from Genetics for a Role of Autophagy and Innate Immunity in IBD Pathogenesis.

192. Hu JZ, Peter I. Evidence of expression variation and allelic imbalance in Crohn's disease susceptibility genes NOD2 and ATG16L1 in human dendritic cells. Gene. 2013;527(2):496-502.

193. Doecke JD, Simms LA, Zhao ZZ, Huang N, Hanigan K, Krishnaprasad K, et al. Genetic Susceptibility in IBD: Overlap Between Ulcerative Colitis and Crohn's Disease. Inflammatory Bowel Diseases. 2013;19(2):240-5 %8 Feb %! Genetic Susceptibility in IBD: Overlap Between Ulcerative Colitis and Crohn's Disease.

194. Henckaerts L, Cleynen I, Brinar M, John JM, Van Steen K, Rutgeerts P, et al. Genetic Variation in the Autophagy Gene ULK1 and Risk of Crohn's Disease. Inflammatory Bowel Diseases. 2011;17(6):1392-7 %8 Jun %! Genetic Variation in the Autophagy Gene ULK1 and Risk of Crohn's Disease.

195. Van Limbergen J, Wilson DC, Satsangi J. The Genetics of Crohn's Disease. Annu Rev Genom Hum G. 2009;10:89-116 %! The Genetics of Crohn's Disease.

196. Hampe J, Franke A, Rosenstiel P, Till A, Teuber M, Huse K, et al. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. Nature Genetics. 2007;39(2):207-11 %8 Feb %! A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1.

197. Yamazaki K, Umeno J, Takahashi A, Hirano A, Johnson TA, Kumasaka N, et al. A Genome-Wide Association Study Identifies 2 Susceptibility Loci for Crohn's Disease in a Japanese Population. Gastroenterology. 2013;144(4):781-8 %8 Apr %! A Genome-Wide Association Study Identifies 2 Susceptibility Loci for Crohn's Disease in a Japanese Population.

198. Jung C, Colombel JF, Lemann M, Beaugerie L, Allez M, Cosnes J, et al. Genotype/Phenotype Analyses for 53 Crohn's Disease Associated Genetic Polymorphisms. Plos One. 2012;7(12 %8 Dec 27 %! Genotype/Phenotype Analyses for 53 Crohn's Disease Associated Genetic Polymorphisms).

199. Ballal SA, Gallini CA, Segata N, Huttenhower C, Garrett WS. Host and gut microbiota symbiotic factors: lessons from inflammatory bowel disease and successful symbionts. Cell Microbiol. 2011;13(4):508-17 %8 Apr %! Host and gut microbiota symbiotic factors: lessons from inflammatory bowel disease and successful symbionts.

200. Roberts RL, Gearry RB, Hollis-Moffatt JE, Miller AL, Reid J, Stat M, et al. IL23R R381Q and ATG16L1 T300A are strongly associated with Crohn's disease in a study of New Zealand Caucasians with inflammatory bowel disease. American Journal of Gastroenterology. 2007;102(12):2754-61 %8 Dec %! IL23R R381Q and ATG16L1 T300A are strongly associated with Crohn's disease in a study of New Zealand Caucasians with inflammatory bowel disease.

201. Lauriola M, Ugolini G, Rivetti S, Nani S, Rosati G, Zanotti S, et al. IL23R, NOD2/CARD15, ATG16L1 and PHOX2B polymorphisms in a group of patients with Crohn's disease and correlation with sub-phenotypes. International Journal of Molecular Medicine. 2011;27(3):469-77 %8 Mar %! IL23R, NOD2/CARD15, ATG16L1 and PHOX2B polymorphisms in a group of patients with Crohn's disease and correlation with sub-phenotypes.

202. Srivastava B, Mells GF, Cordell HJ, Muriithi A, Brown M, Ellinghaus E, et al. Fine mapping and replication of genetic risk loci in primary sclerosing cholangitis. Scandinavian Journal of Gastroenterology. 2012;47(7):820-6 %8 Jul %! Fine mapping and replication of genetic risk loci in primary sclerosing cholangitis.

203. Goyette P, Lefebvre C, Ng A, Brant SR, Cho JH, Duerr RH, et al. Gene-centric association mapping of chromosome 3p implicates MST1 in IBD pathogenesis. Mucosal Immunol. 2008;1(2):131-8 %8 Mar %! Gene-centric association mapping of chromosome 3p implicates MST1 in IBD pathogenesis.

204. Fisher SA, Tremelling M, Anderson CA, Gwilliam R, Bumpstead S, Prescott NJ, et al. Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease. Nature Genetics. 2008;40(6):710-2 %7 2008/04/29 %8 Jun %9 Multicenter Study Research Support, Non-U.S. Gov't %! Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease %@ 1546-718 (Electronic) 061-4036 (Linking).

205. Gorlatova N, Chao K, Pal LR, Araj RH, Galkin A, Turko I, et al. Protein characterization of a candidate mechanism SNP for Crohn's disease: the macrophage stimulating protein R689C substitution. PLoS One. 2011;6(11):e27269.

206. Beckly JB, Hancock L, Geremia A, Cummings JRF, Morris A, Cooney R, et al. Two-stage candidate gene study of chromosome 3p demonstrates an association between nonsynonymous variants in the MST1R gene and Crohn's disease. Inflammatory Bowel Diseases. 2008;14(4):500-7 %8 Apr %! Two-stage candidate gene study of chromosome 3p demonstrates an association between nonsynonymous variants in the MST1R gene and Crohn's disease.

207. Hauser F, Deyle C, Berard D, Neukirch C, Glowacki C, Bickmann JK, et al. Macrophage-stimulating protein polymorphism rs3197999 is associated with a gain of function: implications for inflammatory bowel disease. Genes Immun. 2012;13(4):321-7.

208. Sanchez E, Rueda B, Callejas JL, Sabio JM, Ortego-Centen N, Jimenez-Alonso J, et al. Analysis of interieukin-23 receptor (IL23R) gene polymorphisms in systemic lupus erythematosus. Tissue Antigens. 2007;70(3):233-7 %8 Sep %! Analysis of interieukin-23 receptor (IL23R) gene polymorphisms in systemic lupus erythematosus.

209. Amre DK, Mack D, Israel D, Morgan K, Lambrette P, Law L, et al. Association Between Genetic Variants in the IL-23R Gene and Early-Onset Crohn's Disease: Results From a Case-Control and Family-Based Study Among Canadian Children. American Journal of Gastroenterology. 2008;103(3):615-20 %8 Mar %! Association Between Genetic Variants in the IL-23R Gene and Early-Onset Crohn's Disease: Results From a Case-Control and Family-Based Study Among Canadian Children.

210. Yu PL, Shen FC, Zhang XF, Cao RS, Zhao XD, Liu PF, et al. Association of Single Nucleotide Polymorphisms of IL23R and IL17 with Ulcerative Colitis Risk in a Chinese Han Population. Plos One. 2012;7(9 %8 Sep 11 %! Association of Single Nucleotide Polymorphisms of IL23R and IL17 with Ulcerative Colitis Risk in a Chinese Han Population).

211. Lacher M, Schroepf S, Helmbrecht J, von Schweinitz D, Ballauff A, Koch I, et al. Association of the interleukin-23 receptor gene variant rs11209026 with Crohn's disease in German children. Acta Paediatr. 2010;99(5):727-33 %8 May %! Association of the interleukin-23 receptor gene variant rs11209026 with Crohn's disease in German children.

212. Baldassano RN, Bradfield JP, Monos DS, Kim CE, Glessner JT, Casalunovo T, et al. Association of variants of the interleukin-23 receptor gene with susceptibility to pediatric Crohn's disease. Clin Gastroenterol H. 2007;5(8):972-6 %8 Aug %! Association of variants of the interleukin-23 receptor gene with susceptibility to pediatric Crohn's disease.

213. Weersma RK, Zhernakova A, Nolte IM, Lefebvre C, Rioux JD, Mulder F, et al. ATG16L1 and IL23R Are Associated With Inflammatory Bowel Diseases but Not With Celiac Disease in The Netherlands. American Journal of Gastroenterology. 2008;103(3):621-7 %8 Mar %! ATG16L1 and IL23R Are Associated With Inflammatory Bowel Diseases but Not With Celiac Disease in The Netherlands.

214. Moon CM, Shin DJ, Kim SW, Son NH, Park A, Park B, et al. Associations Between Genetic Variants in the IRGM Gene and Inflammatory Bowel Diseases in the Korean Population. Inflammatory Bowel Diseases. 2013;19(1):106-14 %8 Jan %! Associations Between Genetic Variants in the IRGM Gene and Inflammatory Bowel Diseases in the Korean Population.

215. Roberts RL, Hollis-Moffatt JE, Gearry RB, Kennedy MA, Barclay ML, Merriman TR. Confirmation of association of IRGM and NCF4 with ileal Crohn's disease in a population-based cohort. Genes Immun. 2008;9(6):561-5 %8 Sep %! Confirmation of association of IRGM and NCF4 with ileal Crohn's disease in a population-based cohort.

216. Weersma RK, Stokkers PCF, Cleynen I, Wolfkamp SCS, Henckaerts L, Schreiber S, et al. Confirmation of Multiple Crohn's Disease Susceptibility Loci in a Large Dutch-Belgian Cohort. American Journal of Gastroenterology. 2009;104(3):630-8 %8 Mar %! Confirmation of Multiple Crohn's Disease Susceptibility Loci in a Large Dutch-Belgian Cohort.

217. Lapaquette P, Glasser AL, Huett A, Xavier RJ, Darfeuille-Michaud A. Crohn's disease-associated adherent-invasive E. coli are selectively favoured by impaired autophagy to replicate intracellularly. Cell Microbiol. 2010;12(1):99-113 %8 Jan %! Crohn's disease-associated adherent-invasive E. coli are selectively favoured by impaired autophagy to replicate intracellularly.

218. Bentley RW, Cleynen I, Gearry RB, Barclay ML, Rutgeerts P, Merriman TR, et al. Evidence that glioma-associated oncogene homolog 1 is not a universal risk gene for inflammatory bowel disease in Caucasians. Genes Immun. 2010;11(6):509-14 %8 Sep %! Evidence that glioma-associated oncogene homolog 1 is not a universal risk gene for inflammatory bowel disease in Caucasians.

219. Andersen V, Ernst A, Sventoraityte J, Kupcinskas L, Jacobsen BA, Krarup HB, et al. Assessment of heterogeneity between European Populations: a Baltic and Danish replication case-control study of SNPs from a recent European ulcerative colitis genome wide association study. Bmc Med Genet. 2011;12 %8 Oct 13 %! Assessment of heterogeneity between European Populations: a Baltic and Danish replication case-control study of SNPs from a recent European ulcerative colitis genome wide association study.

220. Gazouli M, Zacharatos P, Mantzaris GJ, Barbatis C, Ikonomopoulos L, Archimandritis AJ, et al. Association of NOD/CARD15 variants with Crohn's disease in a Greek population. Eur J Gastroen Hepat. 2004;16(11):1177-82.

221. van der Linde K, Boor PPC, Houwing-Duistermaat JJ, Kuipers EJ, Wilson JHP, de Rooij FWM. CARD15 and Crohn's disease: Healthy homozygous carriers of the 3020insC

frameshift mutation. American Journal of Gastroenterology. 2003;98(3):613-7 %8 Mar %! CARD15 and Crohn's disease: Healthy homozygous carriers of the 3020insC frameshift mutation.

222. Mahurkar S, Banerjee R, Rani VS, Thakur N, Rao GV, Reddy DN, et al. Common variants in NOD2 and IL23R are not associated with inflammatory bowel disease in Indians. J Gastroen Hepatol. 2011;26(4):694-+ %8 Apr %! Common variants in NOD2 and IL23R are not associated with inflammatory bowel disease in Indians.

223. Nakagome S, Mano S, Kozlowski L, Bujnicki JM, Shibata H, Fukumaki Y, et al. Crohn's Disease Risk Alleles on the NOD2 Locus Have Been Maintained by Natural Selection on Standing Variation. Mol Biol Evol. 2012;29(6):1569-85.

224. Hradsky O, Dusatkova P, Lenicek M, Bronsky J, Nevoral J, Vitek L, et al. The CTLA4 variants may interact with the IL23R-and NOD2-conferred risk in development of Crohn's disease. Bmc Med Genet. 2010;11 %8 Jun 10 %! The CTLA4 variants may interact with the IL23R-and NOD2-conferred risk in development of Crohn's disease.

225. Hu JZ, Peter I. Evidence of expression variation and allelic imbalance in Crohn's disease susceptibility genes NOD2 and ATG16L1 in human dendritic cells. Gene. 2013;527(2):496-502 %8 Sep 25 %! Evidence of expression variation and allelic imbalance in Crohn's disease susceptibility genes NOD2 and ATG16L1 in human dendritic cells.

226. Lin ZW, Hegarty JP, John G, Berg A, Wang Z, Sehgal R, et al. NOD2 Mutations Affect Muramyl Dipeptide Stimulation of Human B Lymphocytes and Interact with Other IBD-Associated Genes. Digest Dis Sci. 2013;58(9):2599-607.

227. Sventoraityte J, Zvirbliene A, Franke A, Kwiatkowski R, Kiudelis G, Kupcinskas L, et al. NOD2, IL23R and ATG16L1 polymorphisms in Lithuanian patients with inflammatory bowel disease. World Journal of Gastroenterology. 2010;16(3):359-64.

228. Hama I, Ratbi I, Reggoug S, Elkerch F, Kharrasse G, Errabih I, et al. Non-association of Crohn's disease with NOD2 gene variants in Moroccan patients. Gene. 2012;499(1):121-3.

229. Underhill DM, Shimada T. A pair of 9s: it's in the CARDs. Nat Immunol. 2007;8(2):122-4.

230. Rivas MA, Beaudoin M, Gardet A, Stevens C, Sharma Y, Zhang CK, et al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nature Genetics. 2011;43(11):1066-U50 %8 Nov %! Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease.

231. Beaudoin M, Goyette P, Boucher G, Lo KS, Rivas MA, Stevens C, et al. Deep Resequencing of GWAS Loci Identifies Rare Variants in CARD9, IL23R and RNF186 That Are Associated with Ulcerative Colitis. Plos Genet. 2013;9(9 %8 Sep %! Deep Resequencing of GWAS Loci Identifies Rare Variants in CARD9, IL23R and RNF186 That Are Associated with Ulcerative Colitis).

232. Yasukawa S, Miyazaki Y, Yoshii C, Nakaya M, Ozaki N, Toda S, et al. An ITAM-Syk-CARD9 signalling axis triggers contact hypersensitivity by stimulating IL-1 production in dendritic cells. Nat Commun. 2014;5:3755.

233. Cooke J, Zhang H, Greger L, Silva AL, Massey D, Dawson C, et al. Mucosal genome-wide methylation changes in inflammatory bowel disease. Inflammatory Bowel Diseases. 2012;18(11):2128-37 %8 Nov %! Mucosal genome-wide methylation changes in inflammatory bowel disease.

234. Lee YH, Song GG. Pathway analysis of a genome-wide association study of ileal Crohn's disease. DNA Cell Biol. 2012;31(10):1549-54.

235. Roth S, Rottach A, Lotz-Havla AS, Laux V, Muschaweckh A, Gersting SW, et al. Rad50-CARD9 interactions link cytosolic DNA sensing to IL-1beta production. Nat Immunol. 2014;15(6):538-45.

236. Juyal G, Prasad P, Senapati S, Midha V, Sood A, Amre D, et al. An Investigation of Genome-Wide Studies Reported Susceptibility Loci for Ulcerative Colitis Shows Limited Replication in North Indians. Plos One. 2011;6(1 %8 Jan 31 %! An Investigation of Genome-Wide Studies Reported Susceptibility Loci for Ulcerative Colitis Shows Limited Replication in North Indians).

237. Wang P, Wu Y, Li Y, Zheng J, Tang J. A novel RING finger E3 ligase RNF186 regulate ER stress-mediated apoptosis through interaction with BNip1. Cell Signal.

2013;25(11):2320-33 %8 Nov %! A novel RING finger E3 ligase RNF186 regulate ER stress-mediated apoptosis through interaction with BNip1 %@ 1873-3913.

238. Anderson CA, Boucher G, Lees CW, Franke A, D'Amato M, Taylor KD, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. Nature Genetics. 2011;43(3):246-52 %7 2011/02/08 %8 Mar %9 Meta-Analysis Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't %! Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47 %@ 1546-718 (Electronic) 061-4036 (Linking).

239. Lees CW, Barrett JC, Parkes M, Satsangi J. New IBD genetics: common pathways with other diseases. Gut. 2011;60(12):1739-53 %8 Dec %! New IBD genetics: common pathways with other diseases %@ 468-3288.

240. Wouters MM. New insight in the pathogenesis of functional gastrointestinal disorders: association between genetics and colonic transit. Neurogastroent Motil. 2011;23(10):893-7 %8 Oct %! New insight in the pathogenesis of functional gastrointestinal disorders: association between genetics and colonic transit.

241. Coetzee SG, Pierce S, Brundin P, Brundin L, Hazelett DJ, Coetzee GA. Enrichment of risk SNPs in regulatory regions implicate diverse tissues in Parkinson's disease etiology. Sci Rep. 2016;6:30509.

242. Fullard JF, Giambartolomei C, Hauberg ME, Xu K, Voloudakis G, Shao Z, et al. Open chromatin profiling of human postmortem brain infers functional roles for non-coding schizophrenia loci. Hum Mol Genet. 2017.

243. Sawalha AH, Dozmorov MG. Epigenomic functional characterization of genetic susceptibility variants in systemic vasculitis. J Autoimmun. 2016;67:76-81.

244. Chen H, Yu H, Wang J, Zhang Z, Gao Z, Chen Z, et al. Systematic enrichment analysis of potentially functional regions for 103 prostate cancer risk-associated loci. Prostate. 2015;75(12):1264-76.

245. Gurtan AM, Lu V, Bhutkar A, Sharp PA. In vivo structure-function analysis of human Dicer reveals directional processing of precursor miRNAs. RNA. 2012;18(6):1116-22.

246. Gurtan AM, Sharp PA. The role of miRNAs in regulating gene expression networks. J Mol Biol. 2013;425(19):3582-600.

247. Gong J, Tong Y, Zhang HM, Wang K, Hu T, Shan G, et al. Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. Hum Mutat. 2012;33(1):254-63.

248. Chen K, Rajewsky N. Natural selection on human microRNA binding sites inferred from SNP data. Nat Genet. 2006;38(12):1452-6.

249. Gong Y, Wu CN, Xu J, Feng G, Xing QH, Fu W, et al. Polymorphisms in microRNA target sites influence susceptibility to schizophrenia by altering the binding of miRNAs to their targets. Eur Neuropsychopharmacol. 2013;23(10):1182-9.

250. Wu C, Gong Y, Sun A, Zhang Y, Zhang C, Zhang W, et al. The human MTHFR rs4846049 polymorphism increases coronary heart disease risk through modifying miRNA binding. Nutr Metab Cardiovasc Dis. 2013;23(7):693-8.

251. Brest P, Lapaquette P, Souidi M, Lebrigand K, Cesaro A, Vouret-Craviari V, et al. A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. Nat Genet. 2011;43(3):242-5.

252. Biton M, Levin A, Slyper M, Alkalay I, Horwitz E, Mor H, et al. Epithelial microRNAs regulate gut mucosal immunity via epithelium-T cell crosstalk. Nat Immunol. 2011;12(3):239-46.

253. Paraskevi A, Theodoropoulos G, Papaconstantinou I, Mantzaris G, Nikiteas N, Gazouli M. Circulating MicroRNA in inflammatory bowel disease. J Crohns Colitis. 2012;6(9):900-4.

254. Buroker NE. Regulatory SNPs and transcriptional factor binding sites in ADRBK1, AKT3, ATF3, DIO2, TBXA2R and VEGFA. Transcription. 2014;5(4):e964559.

255. Glas J, Wagner J, Seiderer J, Olszak T, Wetzke M, Beigel F, et al. PTPN2 gene variants are associated with susceptibility to both Crohn's disease and ulcerative colitis supporting a common genetic disease background. PLoS One. 2012;7(3):e33682.

256. John G, Hegarty JP, Yu W, Berg A, Pastor DM, Kelly AA, et al. NKX2-3 variant rs11190140 is associated with IBD and alters binding of NFAT. Mol Genet Metab. 2011;104(1-2):174-9.

257. Mesbah-Uddin M, Elango R, Banaganapalli B, Shaik NA, Al-Abbasi FA. In-silico analysis of inflammatory bowel disease (IBD) GWAS loci to novel connections. PLoS One. 2015;10(3):e0119420.

258. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 2012;22(9):1775-89.

259. Geisler S, Coller J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. Nat Rev Mol Cell Biol. 2013;14(11):699-712.

260. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annu Rev Biochem. 2012;81:145-66.

261. Mirza AH, Kaur S, Brorsson CA, Pociot F. Effects of GWAS-associated genetic variants on lncRNAs within IBD and T1D candidate loci. PLoS One. 2014;9(8):e105723.

262. Johnson CM, Traherne JA, Jamieson SE, Tremelling M, Bingham S, Parkes M, et al. Analysis of the BTNL2 truncating splice site mutation in tuberculosis, leprosy and Crohn's disease. Tissue Antigens. 2007;69(3):236-41.

263. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2012;40(Database issue):D13-25.

264. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. 2009;37(9):e67.

265. Eng L, Coutinho G, Nahas S, Yeo G, Tanouye R, Babaei M, et al. Nonclassical splicing mutations in the coding and noncoding regions of the ATM Gene: maximum entropy estimates of splice junction strengths. Hum Mutat. 2004;23(1):67-76.

266. Wang M, Marin A. Characterization and prediction of alternative splice sites. Gene. 2006;366(2):219-27.

267. Turei D, Korcsmaros T, Saez-Rodriguez J. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. Nat Methods. 2016;13(12):966-7.

268. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics. 2008;24(24):2938-9.

269. Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, et al. Ensembl 2016. Nucleic Acids Res. 2016;44(D1):D710-6.

270. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013;Chapter 7:Unit7 20.

271. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248-9.

272. Kalvari I, Tsompanis S, Mulakkal NC, Osgood R, Johansen T, Nezis IP, et al. iLIR: A web resource for prediction of Atg8-family interacting proteins. Autophagy. 2014;10(5):913-25.

273. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015;43(Database issue):D447-52.

274. The UniProt C. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2017;45(D1):D158-D69.

275. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-504.

276. Tang R, Prosser DO, Love DR. Evaluation of Bioinformatic Programmes for the Analysis of Variants within Splice Site Consensus Regions. Adv Bioinformatics. 2016;2016:5614058.

277. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 2014;42(Database issue):D68-73.

278. Wong N, Wang X. miRDB: an online resource for microRNA target prediction and functional annotations. Nucleic Acids Res. 2015;43(Database issue):D146-52.

279. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics. 2005;21(16):3448-9.

280. Greene CS, Krishnan A, Wong AK, Ricciotti E, Zelaya RA, Himmelstein DS, et al. Understanding multicellular function and disease with human tissue-specific networks. Nat Genet. 2015;47(6):569-76.

281. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. Human MicroRNA targets. PLoS Biol. 2004;2(11):e363.

282. Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. Nucleic Acids Res. 2008;36(Database issue):D149-53.

283. Mathelier A, Fornes O, Arenillas DJ, Chen CY, Denay G, Lee J, et al. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 2016;44(D1):D110-5.

284. Mathelier A, Zhao X, Zhang AW, Parcy F, Worsley-Hunt R, Arenillas DJ, et al. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Res. 2014;42(Database issue):D142-7.

285. Turatsinze JV, Thomas-Chollier M, Defrance M, van Helden J. Using RSAT to scan genome sequences for transcription factor binding sites and cis-regulatory modules. Nat Protoc. 2008;3(10):1578-88.

286. Zaidel-Bar R, Itzkovitz S, Ma'ayan A, Iyengar R, Geiger B. Functional atlas of the integrin adhesome. Nature Cell Biology. 2007;9(8):858-68.

287. Barrett GC, Elmore DT. Amino Acids and Peptides: Cambridge University Press; First Edition edition; 2009. 244 p.

288. Matthews JM, Bhati M, Lehtomaki E, Mansfield RE, Cubeddu L, Mackay JP. It takes two to tango: the structure and function of LIM, RING, PHD and MYND domains. Curr Pharm Des. 2009;15(31):3681-96.

289. Yang W, Paschen W. SUMO proteomics to decipher the SUMO-modified proteome regulated by various diseases. Proteomics. 2015;15(5-6):1181-91.

290. Barber LJ, Youds JL, Ward JD, McIlwraith MJ, O'Neil NJ, Petalcorin MI, et al. RTEL1 maintains genomic stability by suppressing homologous recombination. Cell. 2008;135(2):261-71.

291. Uringa EJ, Youds JL, Lisaingo K, Lansdorp PM, Boulton SJ. RTEL1: an essential helicase for telomere maintenance and the regulation of homologous recombination. Nucleic Acids Res. 2011;39(5):1647-55.

292. Youds JL, Mets DG, McIlwraith MJ, Martin JS, Ward JD, NJ ON, et al. RTEL-1 enforces meiotic crossover interference and homeostasis. Science. 2010;327(5970):1254-8.

293. Vannier JB, Pavicic-Kaltenbrunner V, Petalcorin MI, Ding H, Boulton SJ. RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. Cell. 2012;149(4):795-806.

294. Baretic D, Williams RL. PIKKs--the solenoid nest where partners and kinases meet. Curr Opin Struct Biol. 2014;29:134-42.

295. Seidah NG, Sadr MS, Chretien M, Mbikay M. The multifaceted proprotein convertases: their unique, redundant, complementary, and opposite functions. J Biol Chem. 2013;288(30):21473-81.

296. Sarrias MR, Farnos M, Mota R, Sanchez-Barbero F, Ibanez A, Gimferrer I, et al. CD6 binds to pathogen-associated molecular patterns and protects from LPS-induced septic shock. Proc Natl Acad Sci U S A. 2007;104(28):11724-9.

297. Deville C, Girard-Blanc C, Assrir N, Nhiri N, Jacquet E, Bontems F, et al. Mutations in actin used for structural studies partially disrupt beta-thymosin/WH2 domains interaction. FEBS Lett. 2016;590(20):3690-9.

298. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, et al. Towards a knowledge-based Human Protein Atlas. Nat Biotechnol. 2010;28(12):1248-50.

299. Gremel G, Wanders A, Cedernaes J, Fagerberg L, Hallstrom B, Edlund K, et al. The human gastrointestinal tract-specific transcriptome and proteome as defined by RNA sequencing and antibody-based profiling. J Gastroenterol. 2015;50(1):46-57.

300. Ke S, Zhang XH, Chasin LA. Positive selection acting on splicing motifs reflects compensatory evolution. Genome Res. 2008;18(4):533-43.

301. Fairbrother WG, Holste D, Burge CB, Sharp PA. Single nucleotide polymorphism-based validation of exonic splicing enhancers. PLoS Biol. 2004;2(9):E268.

302. Carlini DB, Genut JE. Synonymous SNPs provide evidence for selective constraint on human exonic splicing enhancers. J Mol Evol. 2006;62(1):89-98.

303. Kawase T, Akatsuka Y, Torikai H, Morishima S, Oka A, Tsujimura A, et al. Alternative splicing due to an intronic SNP in HMSD generates a novel minor histocompatibility antigen. Blood. 2007;110(3):1055-63.

304. Seo S, Takayama K, Uno K, Ohi K, Hashimoto R, Nishizawa D, et al. Functional Analysis of Deep Intronic SNP rs13438494 in Intron 24 of PCLO Gene. PLoS One. 2013;8(10):e76960.

305. Pettigrew CA, Wayte N, Wronski A, Lovelock PK, Spurdle AB, Brown MA. Colocalisation of predicted exonic splicing enhancers in BRCA2 with reported sequence variants. Breast Cancer Res Treat. 2008;110(2):227-34.

306. Lim KH, Ferraris L, Filloux ME, Raphael BJ, Fairbrother WG. Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. Proc Natl Acad Sci U S A. 2011;108(27):11093-8.

307. Duursma AM, Kedde M, Schrier M, le Sage C, Agami R. miR-148 targets human DNMT3b protein coding region. RNA. 2008;14(5):872-7.

308. Elcheva I, Goswami S, Noubissi FK, Spiegelman VS. CRD-BP protects the coding region of betaTrCP1 mRNA from miR-183-mediated degradation. Mol Cell. 2009;35(2):240-6.

309. Moretti F, Thermann R, Hentze MW. Mechanism of translational regulation by miR-2 from sites in the 5' untranslated region or the open reading frame. RNA. 2010;16(12):2493-502.

310. Tsai NP, Lin YL, Wei LN. MicroRNA mir-346 targets the 5'-untranslated region of receptor-interacting protein 140 (RIP140) mRNA and up-regulates its protein expression. Biochem J. 2009;424(3):411-8.

311. Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, et al. Combinatorial microRNA target predictions. Nat Genet. 2005;37(5):495-500.

312. Brennecke J, Stark A, Russell RB, Cohen SM. Principles of microRNA-target recognition. PLoS Biol. 2005;3(3):e85.

313. Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. Widespread changes in protein synthesis induced by microRNAs. Nature. 2008;455(7209):58-63.

314. Lai X, Schmitz U, Gupta SK, Bhattacharya A, Kunz M, Wolkenhauer O, et al. Computational analysis of target hub gene repression regulated by multiple and cooperative miRNAs. Nucleic Acids Res. 2012;40(18):8818-34.

315. Shalgi R, Lieber D, Oren M, Pilpel Y. Global and local architecture of the mammalian microRNA-transcription factor regulatory network. PLoS Comput Biol. 2007;3(7):e131.

316. Peterson SM, Thompson JA, Ufkin ML, Sathyanarayana P, Liaw L, Congdon CB. Common features of microRNA target prediction tools. Front Genet. 2014;5:23.

317. Yamamura S, Imai-Sumida M, Tanaka Y, Dahiya R. Interaction and cross-talk between non-coding RNAs. Cell Mol Life Sci. 2017.

318. Choi SO, Cho YS, Kim HL, Park JW. ROS mediate the hypoxic repression of the hepcidin gene by inhibiting C/EBPalpha and STAT-3. Biochem Biophys Res Commun. 2007;356(1):312-7.

319. Kheradpour P, Kellis M. Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. Nucleic Acids Res. 2014;42(5):2976-87.

320. Siersbaek R, Rabiee A, Nielsen R, Sidoli S, Traynor S, Loft A, et al. Transcription factor cooperativity in early adipogenic hotspots and super-enhancers. Cell Rep. 2014;7(5):1443-55.

321. Bonta J, Zeitz J, Frei P, Biedermann L, Sulz MC, Vavricka SR, et al. Cytomegalovirus disease in inflammatory bowel disease: epidemiology and disease characteristics in a large single-centre experience. Eur J Gastroenterol Hepatol. 2016;28(11):1329-34.

322. Mirsepasi-Lauridsen HC, Du Z, Struve C, Charbon G, Karczewski J, Krogfelt KA, et al. Secretion of Alpha-Hemolysin by Escherichia coli Disrupts Tight Junctions in Ulcerative Colitis Patients. Clin Transl Gastroenterol. 2016;7:e149.

323. Schultz BM, Paduro CA, Salazar GA, Salazar-Echegarai FJ, Sebastian VP, Riedel CA, et al. A Potential Role of Salmonella Infection in the Onset of Inflammatory Bowel Diseases. Front Immunol. 2017;8:191.

324. Zamani S, Zali MR, Aghdaei HA, Sechi LA, Niegowska M, Caggiu E, et al. Mycobacterium avium subsp. paratuberculosis and associated risk factors for inflammatory bowel disease in Iranian patients. Gut Pathog. 2017;9:1.

325. Oh SH, You CR, Kim EO, Bae SH, Choi JY, Yoon SK, et al. A Case of Ulcerative Colitis Following Acute Hepatitis Induced by Epstein-Barr Virus Infection. Korean J Gastroenterol. 2016;68(2):104-8.

326. Rokkas T, Gisbert JP, Niv Y, O'Morain C. The association between Helicobacter pylori infection and inflammatory bowel disease based on meta-analysis. United European Gastroenterol J. 2015;3(6):539-50.

327. Yu Q, Zhang S, Li L, Xiong L, Chao K, Zhong B, et al. Enterohepatic Helicobacter Species as a Potential Causative Factor in Inflammatory Bowel Disease: A Meta-Analysis. Medicine (Baltimore). 2015;94(45):e1773.

328. Nishimura S, Yoshino T, Fujikawa Y, Watanabe M, Yazumi S. A rare case of ulcerative colitis exacerbated by VZV infection. Clin J Gastroenterol. 2015;8(6):390-2.

329. Iyer VH, Augustine J, Pulimood AB, Ajjampur SS, Ramakrishna BS. Correlation between coinfection with parasites, cytomegalovirus, and Clostridium difficile and disease severity in patients with ulcerative colitis. Indian J Gastroenterol. 2013;32(2):115-8.

330. Sabath E, Negoro H, Beaudry S, Paniagua M, Angelow S, Shah J, et al. Galpha12 regulates protein interactions within the MDCK cell tight junction and inhibits tight-junction assembly. J Cell Sci. 2008;121(Pt 6):814-24 %8 Mar %! Galpha12 regulates protein interactions within the MDCK cell tight junction and inhibits tight-junction assembly %@ 0021-9533.

331. Herroeder S, Reichardt P, Sassmann A, Zimmermann B, Jaeneke D, Hoeckner J, et al. Guanine nucleotide-binding proteins of the G12 family shape immune functions by controlling CD4+ T cell adhesiveness and motility. Immunity. 2009;30(5):708-20.

332. Chen PW, Kroog GS. Leupaxin is similar to paxillin in focal adhesion targeting and tyrosine phosphorylation but has distinct roles in cell adhesion and spreading. Cell Adh Migr. 2010;4(4):527-40.

333. Dierks S, von Hardenberg S, Schmidt T, Bremmer F, Burfeind P, Kaulfuss S. Leupaxin stimulates adhesion and migration of prostate cancer cells through modulation of the phosphorylation status of the actin-binding protein caldesmon. Oncotarget. 2015;6(15):13591-606.

334. Chew V, Lam KP. Leupaxin negatively regulates B cell receptor signaling. J Biol Chem. 2007;282(37):27181-91.

335. Heller F, Fromm A, Gitter AH, Mankertz J, Schulzke JD. Epithelial apoptosis is a prominent feature of the epithelial barrier disturbance in intestinal inflammation: effect of pro-inflammatory interleukin-13 on epithelial cell function. Mucosal Immunol. 2008;1 Suppl 1:S58-61.

336. Nighot P, Ma T. Role of autophagy in the regulation of epithelial cell junctions. Tissue Barriers. 2016;4(3):e1171284.

337. Emtage AL, Mistry SN, Fischer PM, Kellam B, Laughton CA. GPCRs through the keyhole: the role of protein flexibility in ligand binding to beta-adrenoceptors. J Biomol Struct Dyn. 2016:1-16.

338. Araki Y, Sugihara H, Hattori T. In vitro effects of dextran sulfate sodium on a Caco-2 cell line and plausible mechanisms for dextran sulfate sodium-induced colitis. Oncol Rep. 2006;16(6):1357-62.

339. Iliev ID, Spadoni I, Mileti E, Matteoli G, Sonzogni A, Sampietro GM, et al. Human intestinal epithelial cells promote the differentiation of tolerogenic dendritic cells. Gut. 2009;58(11):1481-9.

340. Benard C, Cultrone A, Michel C, Rosales C, Segain JP, Lahaye M, et al. Degraded carrageenan causing colitis in rats induces TNF secretion and ICAM-1 upregulation in monocytes through NF-kappaB activation. PLoS One. 2010;5(1):e8666.

341. Yu C, Liu Y, Ma T, Liu K, Xu S, Zhang Y, et al. Small molecules enhance CRISPR genome editing in pluripotent stem cells. Cell Stem Cell. 2015;16(2):142-7.

342. Zhang JP, Li XL, Li GH, Chen W, Arakaki C, Botimer GD, et al. Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. Genome Biol. 2017;18(1):35.

343. Lee JC, Espeli M, Anderson CA, Linterman MA, Pocock JM, Williams NJ, et al. Human SNP links differential outcomes in inflammatory and infectious disease to a FOXO3-regulated pathway. Cell. 2013;155(1):57-69.

344. Bowcutt R, Malter LB, Chen LA, Wolff MJ, Robertson I, Rifkin DB, et al. Isolation and cytokine analysis of lamina propria lymphocytes from mucosal biopsies of the human colon. J Immunol Methods. 2015;421:27-35.

345. Vadstrup K, Galsgaard ED, Gerwien J, Vester-Andersen MK, Pedersen JS, Rasmussen J, et al. Validation and Optimization of an Ex Vivo Assay of Intestinal Mucosal Biopsies in Crohn's Disease: Reflects Inflammation and Drug Effects. PLoS One. 2016;11(5):e0155335.

346. Lipsky BP, Beals CR, Staunton DE. Leupaxin is a novel LIM domain protein that forms a complex with PYK2. J Biol Chem. 1998;273(19):11709-13.

347. Sahu SN, Nunez S, Bai G, Gupta A. Interaction of Pyk2 and PTP-PEST with leupaxin in prostate cancer cells. Am J Physiol Cell Physiol. 2007;292(6):C2288-96.

348. Sahu SN, Khadeer MA, Robertson BW, Nunez SM, Bai G, Gupta A. Association of leupaxin with Src in osteoclasts. Am J Physiol Cell Physiol. 2007;292(1):C581-90.

349. Okigaki M, Davis C, Falasca M, Harroch S, Felsenfeld DP, Sheetz MP, et al. Pyk2 regulates multiple signaling events crucial for macrophage morphology and migration. Proc Natl Acad Sci U S A. 2003;100(19):10740-5.

350. Gupta A, Lee BS, Khadeer MA, Tang Z, Chellaiah M, Abu-Amer Y, et al. Leupaxin is a critical adaptor protein in the adhesion zone of the osteoclast. J Bone Miner Res. 2003;18(4):669-85.

351. Kaulfuss S, Grzmil M, Hemmerlein B, Thelen P, Schweyer S, Neesen J, et al. Leupaxin, a novel coactivator of the androgen receptor, is expressed in prostate cancer and plays a role in adhesion and invasion of prostate carcinoma cells. Mol Endocrinol. 2008;22(7):1606-21.

352. Tanaka T, Moriwaki K, Murata S, Miyasaka M. LIM domain-containing adaptor, leupaxin, localizes in focal adhesion and suppresses the integrin-induced tyrosine phosphorylation of paxillin. Cancer Sci. 2010;101(2):363-8.

353. Herard AL, Pierrot D, Hinnrasky J, Kaplan H, Sheppard D, Puchelle E, et al. Fibronectin and its alpha 5 beta 1-integrin receptor are involved in the wound-repair process of airway epithelium. Am J Physiol. 1996;271(5 Pt 1):L726-33.

354. Case LB, Waterman CM. Integration of actin dynamics and cell adhesion by a three-dimensional, mechanosensitive molecular clutch. Nat Cell Biol. 2015;17(8):955-63.

355. Thinwa J, Segovia JA, Bose S, Dube PH. Integrin-mediated first signal for inflammasome activation in intestinal epithelial cells. J Immunol. 2014;193(3):1373-82.

356. Chung IC, OuYang CN, Yuan SN, Li HP, Chen JT, Shieh HR, et al. Pyk2 activates the NLRP3 inflammasome by directly phosphorylating ASC and contributes to inflammasome-dependent peritonitis. Sci Rep. 2016;6:36214.

357. Jun HK, Lee SH, Lee HR, Choi BK. Integrin alpha5beta1 activates the NLRP3 inflammasome by direct interaction with a bacterial surface protein. Immunity. 2012;36(5):755-68.

358. Itani S, Watanabe T, Nadatani Y, Sugimura N, Shimada S, Takeda S, et al. NLRP3 inflammasome has a protective effect against oxazolone-induced colitis: a possible role in ulcerative colitis. Sci Rep. 2016;6:39075.

359. Lazaridis LD, Pistiki A, Giamarellos-Bourboulis EJ, Georgitsi M, Damoraki G, Polymeros D, et al. Activation of NLRP3 Inflammasome in Inflammatory Bowel Disease: Differences Between Crohn's Disease and Ulcerative Colitis. Dig Dis Sci. 2017.

360. Assi K, Patterson S, Dedhar S, Owen D, Levings M, Salh B. Role of epithelial integrin-linked kinase in promoting intestinal inflammation: effects on CCL2, fibronectin and the T cell repertoire. Bmc Immunol. 2011;12:42.

361. Schuller S, Lucas M, Kaper JB, Giron JA, Phillips AD. The ex vivo response of human intestinal mucosa to enteropathogenic Escherichia coli infection. Cell Microbiol. 2009;11(3):521-30.

362. Michael KE, Dumbauld DW, Burns KL, Hanks SK, Garcia AJ. Focal adhesion kinase modulates cell adhesion strengthening via integrin activation. Mol Biol Cell. 2009;20(9):2508-19.

363. Watanabe Y, Tamura M, Osajima A, Anai H, Kabashima N, Serino R, et al. Integrins induce expression of monocyte chemoattractant protein-1 via focal adhesion kinase in mesangial cells. Kidney Int. 2003;64(2):431-40.

364. Wong VW, Garg RK, Sorkin M, Rustad KC, Akaishi S, Levi K, et al. Loss of keratinocyte focal adhesion kinase stimulates dermal proteolysis through upregulation of MMP9 in wound healing. Ann Surg. 2014;260(6):1138-46.

365. Angrisano T, Pero R, Peluso S, Keller S, Sacchetti S, Bruni CB, et al. LPS-induced IL-8 activation in human intestinal epithelial cells is accompanied by specific histone H3 acetylation and methylation changes. BMC Microbiol. 2010;10:172.

366. Jijon HB, Panenka WJ, Madsen KL, Parsons HG. MAP kinases contribute to IL-8 secretion by intestinal epithelial cells via a posttranscriptional mechanism. Am J Physiol Cell Physiol. 2002;283(1):C31-41.

367. Lewis SB, Cook V, Tighe R, Schüller S. Enterohemorrhagic Escherichia coli colonization of human colonic epithelium in vitro and ex vivo. Infect Immun. 2015;83(3):942-9.

368. Prager M, Buettner J, Buening C. Genes involved in the regulation of intestinal permeability and their role in ulcerative colitis. J Dig Dis. 2015;16(12):713-22.

369. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747-53.

370. Gong J, Liu W, Zhang J, Miao X, Guo AY. lncRNASNP: a database of SNPs in lncRNAs and their potential functions in human and mouse. Nucleic Acids Res. 2015;43(Database issue):D181-6.

371. Hrdlickova B, de Almeida RC, Borek Z, Withoff S. Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease. Biochim Biophys Acta. 2014;1842(10):1910-22.

372. Mu XJ, Lu ZJ, Kong Y, Lam HY, Gerstein MB. Analysis of genomic variation in non-coding elements using population-scale sequencing data from the 1000 Genomes Project. Nucleic Acids Res. 2011;39(16):7058-76.

373. Gong J, Tong Y, Zhang HM, Wang K, Hu T, Shan G, et al. Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. Hum Mutat. 2012;33(1):254-63 %8 Jan %! Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis %@ 1098-04.

374. Riva A. Large-scale computational identification of regulatory SNPs with rSNP-MAPPER. Bmc Genomics. 2012;13 Suppl 4:S7.

375. Smale ST, Plevy SE, Weinmann AS, Zhou L, Ramirez-Carrozzi VR, Pope SD, et al. Toward an understanding of the gene-specific and global logic of inducible gene transcription. Cold Spring Harb Symp Quant Biol. 2013;78:61-8.

376. Spisak S, Lawrenson K, Fu Y, Csabai I, Cottman RT, Seo JH, et al. CAUSEL: an epigenome- and genome-editing pipeline for establishing function of noncoding GWAS variants. Nat Med. 2015;21(11):1357-63.

377. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res. 2013;41(Database issue):D377-86.

378. Chu H. Host gene-microbiome interactions: molecular mechanisms in inflammatory bowel disease. Genome Med. 2017;9(1):69.

379. Halfvarson J, Brislawn CJ, Lamendella R, Vazquez-Baeza Y, Walters WA, Bramer LM, et al. Dynamics of the human gut microbiome in inflammatory bowel disease. Nat Microbiol. 2017;2:17004.

380. Knights D, Silverberg MS, Weersma RK, Gevers D, Dijkstra G, Huang H, et al. Complex host genetics influence the microbiome in inflammatory bowel disease. Genome Med. 2014;6(12):107.

381. Ghosh S, Armitage E, Wilson D, Minor PD, Afzal MA. Detection of persistent measles virus infection in Crohn's disease: current status of experimental work. Gut. 2001;48(6):748-52.

382. Lavy A, Broide E, Reif S, Keter D, Niv Y, Odes S, et al. Measles is more prevalent in Crohn's disease patients. A multicentre Israeli study. Dig Liver Dis. 2001;33(6):472-6.

383. Polymeros D, Tsiamoulos ZP, Koutsoumpas AL, Smyk DS, Mytilinaiou MG, Triantafyllou K, et al. Bioinformatic and immunological analysis reveals lack of support for measles virus related mimicry in Crohn's disease. Bmc Med. 2014;12:139.

384. Chevaux JB, Peyrin-Biroulet L. Herpes simplex virus colitis complicating the course of a patient with Crohn's disease and cirrhosis: an underestimated association? Gastroenterol Hepatol (N Y). 2010;6(2):122-4.

385. Phadke VK, Friedman-Moraco RJ, Quigley BC, Farris AB, Norvell JP. Concomitant herpes simplex virus colitis and hepatitis in a man with ulcerative colitis: Case report and review of the literature. Medicine (Baltimore). 2016;95(42):e5082.

386. Goldwich A, Prechtel AT, Muhl-Zurbes P, Pangratz NM, Stossel H, Romani N, et al. Herpes simplex virus type I (HSV-1) replicates in mature dendritic cells but can only be transferred in a cell-cell contact-dependent manner. J Leukoc Biol. 2011;89(6):973-9.

387. Hahm B, Cho JH, Oldstone MB. Measles virus-dendritic cell interaction via SLAM inhibits innate immunity: selective signaling through TLR4 but not other TLRs mediates suppression of IL-12 synthesis. Virology. 2007;358(2):251-7.

388. Pang IK, Ichinohe T, Iwasaki A. IL-1R signaling in dendritic cells replaces pattern-recognition receptors in promoting CD8(+) T cell responses to influenza A virus. Nat Immunol. 2013;14(3):246-53.

389. Ryan EJ, Stevenson NJ, Hegarty JE, O'Farrelly C. Chronic hepatitis C infection blocks the ability of dendritic cells to secrete IFN-alpha and stimulate T-cell proliferation. J Viral Hepat. 2011;18(12):840-51.

390. Servet-Delprat C, Vidalain PO, Azocar O, Le Deist F, Fischer A, Rabourdin-Combe C. Consequences of Fas-mediated human dendritic cell apoptosis induced by measles virus. J Virol. 2000;74(9):4387-93.

391. Sheridan PA, Beck MA. The dendritic and T cell responses to herpes simplex virus-1 are modulated by dietary vitamin E. Free Radic Biol Med. 2009;46(12):1581-8.

392. Stone AE, Mitchell A, Brownell J, Miklin DJ, Golden-Mason L, Polyak SJ, et al. Hepatitis C virus core protein inhibits interferon production by a human plasmacytoid

dendritic cell line and dysregulates interferon regulatory factor-7 and signal transducer and activator of transcription (STAT) 1 protein expression. PLoS One. 2014;9(5):e95627.

393. Zaslavsky E, Hayot F, Sealfon SC. Computational approaches to understanding dendritic cell responses to influenza virus infection. Immunol Res. 2012;54(1-3):160-8.

394. Avota E, Koethe S, Schneider-Schaulies S. Membrane dynamics and interactions in measles virus dendritic cell infections. Cell Microbiol. 2013;15(2):161-9.

395. Chan MC, Lee N, Chan PK, To KF, Wong RY, Law CO, et al. Intestinal binding of seasonal influenza A viruses to DC-SIGN(+) CD68(+) cells. Influenza Other Respir Viruses. 2013;7(3):228-30.

396. de Jong MA, de Witte L, Bolmstedt A, van Kooyk Y, Geijtenbeek TB. Dendritic cells mediate herpes simplex virus infection and transmission through the C-type lectin DC-SIGN. J Gen Virol. 2008;89(Pt 10):2398-409.

397. de Witte L, Abt M, Schneider-Schaulies S, van Kooyk Y, Geijtenbeek TB. Measles virus targets DC-SIGN to enhance dendritic cell infection. J Virol. 2006;80(7):3477-86.

398. Lozach PY, Lortat-Jacob H, de Lacroix de Lavalette A, Staropoli I, Foung S, Amara A, et al. DC-SIGN and L-SIGN are high affinity binding receptors for hepatitis C virus glycoprotein E2. J Biol Chem. 2003;278(22):20358-66.

399. Ludwig IS, Lekkerkerker AN, Depla E, Bosman F, Musters RJ, Depraetere S, et al. Hepatitis C virus targets DC-SIGN and L-SIGN to escape lysosomal degradation. J Virol. 2004;78(15):8322-32.

400. Mesman AW, Zijlstra-Willems EM, Kaptein TM, de Swart RL, Davis ME, Ludlow M, et al. Measles virus suppresses RIG-I-like receptor activation in dendritic cells via DC-SIGN-mediated inhibition of PP1 phosphatases. Cell Host Microbe. 2014;16(1):31-42.

401. Zeng JQ, Xu CD, Zhou T, Wu J, Lin K, Liu W, et al. Enterocyte dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin expression in inflammatory bowel disease. World J Gastroenterol. 2015;21(1):187-95.

402. Fazio C, Ricciardiello L. Inflammation and Notch signaling: a crosstalk with opposite effects on tumorigenesis. Cell Death Dis. 2016;7(12):e2515.

403. Bray SJ. Notch signalling: a simple pathway becomes complex. Nat Rev Mol Cell Biol. 2006;7(9):678-89.

404. Shinoda M, Shin-Ya M, Naito Y, Kishida T, Ito R, Suzuki N, et al. Early-stage blocking of Notch signaling inhibits the depletion of goblet cells in dextran sodium sulfate-induced colitis in mice. J Gastroenterol. 2010;45(6):608-17.

405. Outtz HH, Tattersall IW, Kofler NM, Steinbach N, Kitajewski J. Notch1 controls macrophage recruitment and Notch signaling is activated at sites of endothelial cell anastomosis during retinal angiogenesis in mice. Blood. 2011;118(12):3436-9.

406. Outtz HH, Wu JK, Wang X, Kitajewski J. Notch1 deficiency results in decreased inflammation during wound healing and regulates vascular endothelial growth factor receptor-1 and inflammatory cytokine expression in macrophages. J Immunol. 2010;185(7):4363-73.

407. Xu JY, Meng QH, Chong Y, Jiao Y, Zhao L, Rosen EM, et al. Sanguinarine is a novel VEGF inhibitor involved in the suppression of angiogenesis and cell migration. Mol Clin Oncol. 2013;1(2):331-6.

408. Kueanjinda P, Roytrakul S, Palaga T. A Novel Role of Numb as A Regulator of Pro-inflammatory Cytokine Production in Macrophages in Response to Toll-like Receptor 4. Sci Rep. 2015;5:12784.

409. Crosnier C, Stamataki D, Lewis J. Organizing cell renewal in the intestine: stem cells, signals and combinatorial control. Nat Rev Genet. 2006;7(5):349-59.

410. Kim YS, Ho SB. Intestinal goblet cells and mucins in health and disease: recent insights and progress. Curr Gastroenterol Rep. 2010;12(5):319-30.

411. Fre S, Huyghe M, Mourikis P, Robine S, Louvard D, Artavanis-Tsakonas S. Notch signals control the fate of immature progenitor cells in the intestine. Nature. 2005;435(7044):964-8.

412. Mathern DR, Laitman LE, Hovhannisyan Z, Dunkin D, Farsio S, Malik TJ, et al. Mouse and human Notch-1 regulate mucosal immune responses. Mucosal Immunol. 2014;7(4):995-1005.

413. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet. 2014;15(4):272-86.

414. Sethupathy P, Collins FS. MicroRNA target site polymorphisms and human disease. Trends Genet. 2008;24(10):489-97.

415. Moszynska A, Gebert M, Collawn JF, Bartoszewski R. SNPs in microRNA target sites and their potential role in human disease. Open Biol. 2017;7(4).

416. Martin EC, Rhodes LV, Elliott S, Krebs AE, Nephew KP, Flemington EK, et al. microRNA regulation of mammalian target of rapamycin expression and activity controls estrogen receptor function and RAD001 sensitivity. Mol Cancer. 2014;13:229.

417. Caminsky N, Mucaki EJ, Rogan PK. Interpretation of mRNA splicing mutations in genetic disease: review of the literature and guidelines for information-theoretical analysis. F1000Res. 2014;3:282.

418. Soukarieh O, Gaildrat P, Hamieh M, Drouet A, Baert-Desurmont S, Frebourg T, et al. Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. Plos Genet. 2016;12(1):e1005756.

419. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. Nat Genet. 2013;45(10):1127-33.

420. Jin HJ, Jung S, DebRoy AR, Davuluri RV. Identification and validation of regulatory SNPs that modulate transcription factor chromatin binding and gene expression in prostate cancer. Oncotarget. 2016;7(34):54616-26.

421. Smurnyy Y, Cai M, Wu H, McWhinnie E, Tallarico JA, Yang Y, et al. DNA sequencing and CRISPR-Cas9 gene editing for target validation in mammalian cells. Nat Chem Biol. 2014;10(8):623-5.

422. Brooks J, Watson A, Korcsmaros T. Omics Approaches to Identify Potential Biomarkers of Inflammatory Diseases in the Focal Adhesion Complex. Genomics Proteomics Bioinformatics. 2017;15(2):101-9.

423. Modos D, Bulusu KC, Fazekas D, Kubisch J, Brooks J, Marczell I, et al. Neighbours of cancer-related proteins have key influence on pathogenesis and could increase the drug target space for anticancer therapies. NPJ Syst Biol Appl. 2017;3:2.

7. Glossary

Ensembl (https://www.ensembl.org) Genome Database of the genome sequence

for vertebrates of the genome sequence.

dbSNP (ncbi.nlm.nih.gov) Free public archive/database of Single Nucleotide Polymorphisms hosted by the National Center for Biotechnology Information (NCBI) in collaboration with the National Human Genome Research Institute (NHGRI)

JASPAR largest open-access database of curated and non-redundant transcription factor binding profiles. The JASPAR core database contains curated, nonredundant set of profiles, derived from published collections of experimentally defined transcription factor binding sites for eukaryotes.

miRBASE archive of microRNA sequences and annotations. It is a primary repository for published microRNA sequence and annotation database.

TARBASE v7.0 DIANA-TarBase v7.0 (http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=tarbase/index) Database of published experimentally validated miRNA:gene interactions.

MiRANDA Target prediction software to identify potential microRNA target sites in genomic sequences.

ELM (elm.eu.org) computational biology resource for annotation and detection of eukaryotic linear motifs

Omnipath (omnipathdb.org) a comprehensive collection of literature curated human signalling pathways.

ASSP (wangcomputing.com/assp/) sequence analysis tool for the prediction and classification of splice sites. It predicts putative alternative exon isoform, cryptic and constitutive splice sites of internal (coding) exons. Non canonical splice sites are not detected.

MES (gene.mit.edu/burgelab/maxent/Xmaxentsca_scoreseq.html) MaxEntScan is

based on the approach for modelling the sequences of short sequence motifs.

HSF (www.umd.be/HSF/) tool for calculating consensus values of potential splice sites

Cytoscape (www.cytoscape.org) open source bioinformatics software platform for visualizing molecular interaction networks and integrating with gene expression profiles and other state data.

Uniprot (www.uniprot.org) The Universal Protein Resource; comprehensive resource for protein sequence and annotation data.

STRING (string-db.org) database of known and predicted protein-protein interactions. Ween organisms and interactions aggregated from other primary databases. The interactions include direct and indirect associations. The data comes from computational prediction, knowledge transfer between organisms and fro interactions aggregated from other databases.

CHESEL Curated hexamer exonic splice enhancing ligands – the motifs curated and collated by De Haerty, Earlham Institute.

FASTA a suite of programs for searching nucleotide or protein databases with a query sequence. It performs a heuristic search of a protein or nucleotide database for a query of the same type.

MirSVR a predicted target site scoring method. Allows for comprehensive modelling of miRNA target predicting functional non-conserved and non-canonical sites.

ORegAnno (www.oreganno.org) The Open Regulatory Annotation database. Curated database of regulatory regions, transcription factor binding sites, regulatory variants and haplotypes.

GeneCards (www.genecards.org) searchable, integrated, database of human genes that provides concise genomic related information on all known and predicted human genes.

UCSC Genome Browser (https://genome.ucsc.edu/) on-line genome browser hosted by the university of California, Santa Cruz. Allows access to genome sequence data integrated with a large collection of aligned annotations.

Adhesome (adhesome.org) Literature based protein-protein interaction network that was developed from the biomedical literature of known interactions and cellular components constituting the focal adhesion complex in mammalian cells.

8. Appendix

8.1 SNPs used for the UC Interactome and Patient Footprints – the Norwich Cohort

This appendix contains all the UC and IBD SNPs used to create the UC Interactome as described in Chapter 2. This appendix also contains all the UC and IBD SNPs used to create the UC Interactome as described in Chapter 4. These SNPs were identified from the original SNPs used in Chapter 2 that were present in the Norwich Patient cohort and are highlighted in yellow throughout the tables.

The tables detail the risk allele used, the source of the SNP e.g. immunochip, broad institute fine mapping. If the Broad institute data was used their PICS value was identified, or p-value if immunochip data was used. Tissue enhancers were identified if available. The SNP site annotation, gene name and Ensembl ID are included if relevant.

The acronyms used in this appendix are as follows:

Reverse strand	rev
Forward strand	fwd
Not applicable	NA
Immunochip data UC associated	ICUC
Sanger GWAS data UC associated	UCS
Immunochip data IBD associated	ICIBD
Sanger GWAS data IBD associated	IBDS
Broad Institute UC finemapped index SNP	BII
Broad Institute UC finemapped SNP	BIFM
T helper 1 Cells	Th1
T helper 2 Cells	Th2

Please note that the Broad Institute index SNPs that are not identified on Immunochip or the Sanger GWAS data were identified from those enhancing to the colonic mucosa or non- enhancing. The Broad institute finemapped SNPs were taken from the entire UC associated cohort if there was not a SNP within the subset of Broad Institute Index SNPs defined above.

SNP/ID	Strand	Risk Allele	WT allele	Source	PICS	P value	SNP if FM	Enhancer	Annotation SNP/ID	Gene	Ensembl ID
rs1182188	rev	T	C	ICUC	NA	1.23E-09	rs798502		intronic	GNA12	ENSG00000146535
rs1654644	fwd	G	T	ICUC	NA	6.05E-11	rs11672983		intronic	KIR3DL2	ENSG00000240403
rs3774937	fwd	A	G	ICUC	NA	2.23E-09	rs3774959		intronic	NFKB1	ENSG00000109320
rs3774959	fwd	A	G	UCS BII BIFM	0.0664	NA	rs3774959	Th1	intronic	NFKB1	ENSG00000109320
rs2816958	rev	G	A	UCS ICUC BII BIFM	0.4543	1.98E-17	rs2816958	none	intronic	NR5A2	ENSG00000116833
rs254560	rev	A	G	UCS ICUC BII BIFM	0.3469	2.55E-09	rs254560	None	intronic	C5orf66	ENSG00000224186
rs17229285	fwd	C	T	UCS ICUC BII BIFM	0.2125	1.73E-18	rs17229285	none	intronic	lincRNA	ENSG00000225421
rs11168249	fwd	C	T	IBDS ICUC	NA	7.78E-09	rs11168249		intronic	HDAC7	ENSG00000061273
rs4743820	fwd	T	C	IBDS ICUC	NA	3.60E-09	rs4743820		intronic	lincRNA	ENSG00000229694
rs7134599	fwd	A	G	IBDS ICUC BIFM BII	0.0503	8.51E-32	rs7134599	None	intronic	IFNG-AS1	ENSG00000255733
rs6920220	fwd	A	G	IBDS ICUC BIFM BII	0.2286	1.40E-21	rs6920220	none	intronic	lincRNA	ENSG00000230533
rs941823	rev	C	T	IBDS ICUC BII BIFM	0.2447	2.95E-11	rs941823	colonic mucosa	intronic	lincRNA	ENSG00000215483
rs79755370	fwd	A	C	BIFM	0.1769	NA	rs11209026	none	intronic	IL23R	ENSG00000162594
rs11581607	fwd	A	G	BIFM	0.1769	NA	rs11209026	none	intronic	IL23R	ENSG00000162594
rs113935720	fwd	C	T	BIFM	0.1769	NA	rs11209026	none	intronic	IL23R	ENSG00000162594

12 UC Associated SNPs utilised in Chapter 2 UC Interactome – Intronic SNPS. Highlighted SNPs used in Chapter 4.

SNP/ID	Strand	Risk Allele	WT allele	Source	PICS	P value	SNP if FM	Enhancer	Annotation SNP/ID	Gene	Ensembl ID
rs12261843	fwd	G	T	BII BIFM	0.1173	NA	rs12261843	none	intronic	CCNY	ENSG00000108100
rs6481950	fwd	T	C	BIFM	0.1173	NA	rs12261843	none	intronic	CCNY	ENSG00000108100
rs12254167	fwd	G	C	BIFM	0.1173	NA	rs12261843	none	intronic	CCNY	ENSG00000108100
rs1297265	fwd	A	G	BII BIFM	0.1531	NA	rs1297265	HEpG2	intronic	lincRNA	ENSG00000229425
rs4657041	fwd	C	T	BIFM	0.444	NA	rs1801274		intronic	FCGR2A	ENSG00000143226
rs254562	rev	C	T	BIFM	0.1447	NA	rs254560-A	none	intronic	C5orf66	ENSG00000224186
rs267939	rev	C	G	BII BIFM	0.1752	NA	rs267939	none	intronic	DAP	ENSG00000112977
rs267984	rev	A	T	BIFM	0.0303	NA	rs267939	none	intronic	DAP	ENSG00000112977
rs28671712	fwd	A	G	BIFM	0.1403	NA	rs28374715	none	intronic	CHP1	ENSG00000187446
rs3024495	rev	A	G	BIFM	0.4133	NA	rs3024505	Th2	intronic	IL10	ENSG00000136634
rs11567701	fwd	T	G	BIFM	0.034	NA	rs3194051-G	colonic mucosa	intronic	IL17R	ENSG00000168685

13 UC Associated SNPs utilised in Chapter 2 UC Interactome – Intronic SNPs. Highlighted SNPs used in Chapter 4.

SNP/ID	Strand	Risk Allele	WT allele	Source	PICS	P value	SNP if FM	Enhancer	Annotation SNP/ID	Gene	Ensembl ID
rs11567699	fwd	G	C	BIFM	0.034	NA	rs3194051	colonic mucosa	intronic	IL17R	ENSG00000168685
rs1598859	fwd	C	T	BIFM	0.0664	NA	rs3774959	Th1	intronic	NFKB1	ENSG00000109320
rs9770544	fwd	C	G	BIFM	0.4201	NA	rs4722672	colonic mucosa	intronic	lincRNA	ENSG00000253508
rs543104	rev	A	G	BIFM	0.0647	NA	rs483905	colonic mucosa	intronic	MAML2	ENSG00000184384
rs2425019	fwd	A	G	BIFM	0.0996	NA	rs6088765	none	intronic	MMP24	ENSG00000125966
rs6584283	fwd	T	C	BII BIFM	0.2242	NA	rs6584283	colonic mucosa	intronic	lincRNA	ENSG00000257582
rs6911490	fwd	T	C	BII BIFM	0.9378	NA	rs6911490	none	intronic	ATG5	ENSG00000057663
rs11757201	fwd	C	G	BIFM	0.2286	NA	rs6920220	none	intronic	lincRNA	ENSG00000230533
rs17264332	fwd	G	A	BIFM	0.2286	NA	rs6920220	none	intronic	lincRNA	ENSG00000230533
rs6927172	fwd	G	C	BIFM	0.2286	NA	rs6920220	none	intronic	lincRNA	ENSG00000230533
rs34902013	fwd	G	A	BIFM	0.0503	NA	rs7134599	None	intronic	IFNG-AS1	ENSG00000255733
rs723403	rev	C	T	BIFM	0.0503	NA	rs7134599	None	intronic	IFNG-AS1	ENSG00000255733

14 UC Associated SNPs utilised in Chapter 2 UC Interactome – Intronic SNPs. Highlighted SNPs used in Chapter 4.

SNP/ID	Strand	Risk Allele	WT allele	Source	PICS	P value	SNP if FM	Enhancer	Annotation SNP/ID	Gene	Ensembl ID
rs2193041	fwd	G	A	BIFM	0.0503	NA	rs7134599	None	intronic	IFNG-AS1	ENSG00000255733
rs12829089	fwd	G	T	BIFM	0.0503	NA	rs7134599	None	intronic	IFNG-AS1	ENSG00000255733
rs200073939	fwd	G	C	BIFM	0.0503	NA	rs7134599	None	intronic	IFNG-AS1	ENSG00000255733
rs12318183	fwd	A	C	BIFM	0.0503	NA	rs7134599	None	intronic	IFNG-AS1	ENSG00000255733
rs1558743	fwd	G	T	BIFM	0.0503	NA	rs7134599	None	intronic	IFNG-AS1	ENSG00000255733
rs1558746	fwd	A	G	BIFM	0.0503	NA	rs7134599	None	intronic	IFNG-AS1	ENSG00000255733
rs12132298	fwd	C	T	BIFM	0.1139	NA	rs7554511	none	intronic	C1orf107	ENSG00000163362
rs59655222	fwd	C	T	BIFM	0.1139	NA	rs7554511	none	intronic	C1orf108	ENSG00000163362
rs41299637	fwd	G	T	BIFM	0.1139	NA	rs7554511	none	intronic	C1orf109	ENSG00000163362
rs12131796	fwd	A	G	BIFM	0.1139	NA	rs7554511	none	intronic	C1orf110	ENSG00000163362
rs7608697	fwd	C	A	BIFM	0.1946	NA	rs7608910	none	intronic	PUS10	ENSG00000162927
rs7596362	fwd	C	T	BIFM	0.1946	NA	rs7608910	none	intronic	PUS10	ENSG00000162927

15 UC Associated SNPs utilised in Chapter 2 UC Interactome – Intronic SNPs

SNP/ID	Strand	Risk Allele	WT allele	Source	PICS	P value	SNP if FM	Enhancer	Annotation SNP/ID	Gene	Ensembl ID
rs4560096	fwd	G	T	BIFM	0.1946	NA	rs7608910	none	intronic	PUS10	ENSG00000162927
rs798502	rev	A	C	BII BIFM	0.1681	NA	rs798502	none	intronic	GNA12	ENSG00000146535
rs11041476	fwd	A	G	BIFM	0.1404	NA	rs907611	colonic mucosa	intronic	LSP1	ENSG00000130592
rs943072	rev	G	T	BII BIFM	0.141	NA	rs943072	colonic mucosa	intronic	lincRNA	ENSG00000283573
rs6940798	fwd	A	G	BIFM	0.0742		rs943072	colonic mucosa	intronic	lincRNA	ENSG00000283573
rs2396087	fwd	T	C	BIFM	0.141	NA	rs943072	colonic mucosa	intronic	lincRNA	ENSG00000283573
rs2396088	fwd	A	G	BIFM	0.141	NA	rs943072-G	colonic mucosa	intronic	lincRNA	ENSG00000283573
rs4273687	fwd	G	A	BIFM	0.141	NA	rs943072-G	colonic mucosa	intronic	lincRNA	ENSG00000283573
rs9822268	fwd	A	G	BII BIFM	0.1386	NA	rs9822268	none	intronic	APEH	ENSG00000164062
rs11130213	fwd	T	C	BIFM	0.1386	NA	rs9822268	none	intronic	APEH	ENSG00000164062
rs2581817	rev	G	C	BIFM	0.1017	NA	rs9847710	none	intronic	SFMBT1	ENSG00000163935
rs2564956	rev	G	A	BIFM	0.1017	NA	rs9847710	colonic mucosa	intronic	SFMBT1	ENSG00000163935

16 UC Associated SNPs utilised in Chapter 2 UC Interactome – Intronic SNPs. Highlighted are SNPs used in Chapter 4.

SNP/ID	Strand	Risk Allele	WT allele	Source	PICS	P value	SNP if FM	Enhancer	Annotation SNP/ID	Gene	Ensembl ID
rs6871626	fwd	A	C	IBDS ICIBD BIFM BII	0.7608	1.43E-42	rs6871626	none	intronic	AC008697.1	ENSG00000249738
rs7554511	fwd	C	A	IBDS ICIBD BIFM BII	0.1139	1.24E-32	rs7554511	none	intronic	C1orf106	ENSG00000163362
rs7608910	fwd	G	A	IBDS ICIBD BIFM BII	0.1946	8.65E-32	rs7608910	none	intronic	PUS10	ENSG00000162927
rs2823286	fwd	G	A	IBDS ICIBD	NA	9.28E-30	rs2823286		intronic	lincRNA	ENSG00000229425
rs1893217	rev	G	A	IBDS ICIBD	NA	3.05E-26	rs1893217		intronic	PTPN2	ENSG00000175354
rs6062504	fwd	G	A	IBDS ICIBD	NA	1.09E-23	rs6062504		intronic	ZGPAT	ENSG00000197114
rs12942547	fwd	A	G	IBDS ICIBD	NA	5.51E-22	rs12942547		intronic	STAT3	ENSG00000168610
rs11879191	fwd	G	A	IBDS ICIBD	NA	2.04E-18	rs11879191		intronic	CDC37	ENSG00000105401
rs2266959	fwd	T	G	IBDS ICIBD	NA	1.39E-16	rs2266959		intronic	UBE2L3	ENSG00000185651
rs3766606	rev	G	T	BIFM ICIBD	0.1119	1.12E-15	rs35675666	colonic mucosa	intronic	PARK7	ENSG00000116288
rs8005161	fwd	T	C	IBDS ICIBD	NA	2.35E-14	rs8005161		intronic	GPR65	ENSG00000140030
rs5763767	fwd	A	G	ICIBD	NA	2.70E-14	rs2412970		intronic	HORMAD2	ENSG00000176635

2 IBD Associated SNPs utilised in Chapter 2 UC Interactome – Intronic SNPs. Highlighted are SNPs used in Chapter 4.

SNP/ID	Strand	Risk Allele	WT allele	Source	PICS	P value	SNP if FM	Enhancer	Annotation SNP/ID	Gene	Ensembl ID
rs3851228	fwd	T	A	IBDS ICIBD	NA	1.08E-13	rs3851228		intronic	TRAF3IP2-AS1	ENSG00000231889
rs7657746	fwd	A	G	IBDS ICIBD	NA	2.76E-13	rs7657746		intronic	KIAA1109	ENSG00000138688
rs259964	fwd	A	G	IBDS ICIBD	NA	1.01E-12	rs259964		intronic	ZNF831	ENSG00000124203
rs17119	rev	A	G	IBDS ICIBD	NA	3.08E-11	rs17119		intronic	lincRNA	ENSG00000234261
rs1517352	fwd	C	A	IBDS ICIBD	NA	3.28E-11	rs1517352		intronic	STAT4	ENSG00000138378
rs7495132	fwd	C	T	IBDS ICIBD	NA	9.48E-11	rs7495132		intronic	CRTC3	ENSG00000140577
rs11229555	fwd	G	T	ICIBD	NA	6.80E-10	rs10896794		intronic	GLYAT	ENSG00000149124
rs7404095	fwd	C	T	IBDS ICIBD	NA	9.68E-10	rs7404095		intronic	PRKCB	ENSG00000166501
rs13277237	fwd	G	A	ICIBD	NA	1.65E-09	rs1991866		intronic	lincRNA	ENSG00000229140
rs10896794	fwd	T	C	IBDS	NA	6.10E-08	rs10896794		Intronic	LPXN	ENSG00000110031

2 IBD Associated SNPs utilised in Chapter 2 UC Interactome – Intronic SNPs. Highlighted are SNPs used in Chapter 4.

SNP/ID	Strand	Risk Allele	WT allele	Source	PICS	P value	SNP if FM	Enhancer	Annotation SNP/ID	Gene	Ensembl ID
rs1801274	rev	A	G	UCS ICUC BIFM	0.444	1.44E-22	rs1801274		missense	FCGR2A	ENSG00000143226
rs3749171	fwd	T	C	IBDS ICUC	NA	3.07E-21	rs3749171		missense	GPR35	ENSG00000178623
rs11209026	fwd	G	A	ICIBD IBDS BIFM BII	0.1769	8.12E-161	rs11209026	none	missense	IL23R	ENSG00000162594
rs3197999	rev	A	G	ICIBD IBDS	NA	1.01E-47	rs3197999		missense	MST1	ENSG00000173531
rs3742130	rev	G	A	ICIBD	NA	2.37E-14	rs9557195		synonymous	GPR18	ENSG00000125245
rs11230563	fwd	C	T	IBDS ICIBD	NA	9.03E-13	rs11230563		missense	CD6	ENSG00000013725
rs12103	fwd	A	G	IBDS ICIBD	NA	7.66E-13	rs12103		synonymous	CPSF3L	ENSG00000127054
rs5771069	fwd	G	A	BII BIFM	0.0956	NA	rs5771069	colonic mucosa	missense	IL17REL	ENSG00000188263
rs10781499	fwd	A	G	BII BIFM	0.2	NA	rs10781499		synonymous	CARD9	ENSG00000187796
rs2257440	fwd	T	C	BIFM	0.0769	NA	rs2297441	colonic mucosa	synonymous	RTEL1	ENSG00000258366

3 UC and IBD Associated SNPs utilised in Chapter 2 UC Interactome – Exonic SNPs. Highlighted are SNPs used in Chapter 4.

SNP/ID	Strand	Risk Allele	WT allele	Source	PICS	P value	SNP if FM	Enhancer	Annotation SNP/ID	Gene	Ensembl ID
rs3208008	fwd	A	C	BIFM	0.0478	NA	rs2297441	colonic mucosa	missense	RTEL1	ENSG00000258366
rs1131095	fwd	C	T	BIFM	0.1386	NA	rs9822268	none	synonymous	APEH	ENSG00000164062
rs13085791	fwd	A	C	BIFM	0.1386	NA	rs9822268	colonic mucosa	synonymous	MST1	ENSG00000173531

3 UC and IBD Associated SNPs utilised in Chapter 2 UC Interactome – Exonic SNPs

SNP/ID	Strand	Risk Allele	WT allele	Source	PICS	P value	SNP if FM	Enhancer	Annotation SNP/ID	Gene	Ensembl ID
rs17780256	fwd	A	C	ICUC BIFM	0.1314	1.89E-09	rs7210086-A	none	3'UTR	SLC39A11	ENSG00000133195
rs2382817	fwd	A	C	ICIBD IBDS	NA	3.70E-12	rs2382817		5'UTR	TMBIM1	ENSG00000135926
rs727088	fwd	G	A	ICIBD IBDS	NA	4.65E-09	rs727088		3'UTR	CD226	ENSG00000150637
rs1126510	rev	G	A	BIFM BII	0.3898	NA	rs1126510	none	3'UTR	PTGIR	ENSG00000160013
rs2297441	fwd	A	G	BIFM BII	0.0769	NA	rs2297441	colonic mucosa	3'UTR	RTEL1	ENSG00000258366
rs35675666	fwd	G	T	BIFM BII	0.1119	NA	rs35675666	colonic mucosa	5'UTR	PARK7	ENSG00000116288
rs10114470	fwd	T	C	BIFM	0.0511	NA	rs4246905	colonic mucosa	3'UTR	TNFSF15	ENSG00000181634
rs11567685	fwd	C	T	BIFM	0.034	NA	rs3194051	colonic mucosa	5'UTR	IL17R	ENSG00000168685

4 UC and IBD Associated SNPs utilised in Chapter 2 UC Interactome – 3'UTR and 5'UTR SNPs. Highlighted are SNPs used in Chapter 4.

SNP/ID	Strand	Risk Allele	WT allele	Source	PICS	P value	SNP if FM	Enhancer	Annotation SNP/ID	Gene	Ensembl ID
rs10797432	fwd	C	T	UCS BIFM BII	0.1121	5.18E-09	rs10797432	none	DGV	TNFRSF14	ENSG00000157873
rs11150589	fwd	T	C	UCS ICUC BIFM BII	0.2236	6.04E-10	rs11150589	Th2	UGV	ITGAL	ENSG00000005844
rs561722	rev	C	T	UCS ICUC BIFM BII	0.1751	5.15E-17	rs561722	none	UGV	NXPE2P1	ENSG00000255982
rs6667605	fwd	C	T	ICUC	NA	2.62E-12	rs10797433		DGV	RP3-395M20.7	ENSG00000225931
rs7210086	fwd	A	C	BII BIFM	0.2636	NA	rs7210086	none	DGV	SLC39A11	ENSG00000133195
rs6451493	fwd	T	G	BII BIFM	0.0382	NA	rs6451493	colonic mucosa	UGV	lincRNA	ENSG00000283286
rs16940202	fwd	C	T	BII BIFM	0.9638	NA	rs16940202	colonic mucosa	UGV	lincRNA	ENSG00000269667
rs2155219	fwd	T	G	BII BIFM	0.3783	NA	rs2155219	colonic mucosa	UGV	pseudogene	ENSG00000254755
rs7562334	fwd	A	G	BIFM	0.0422	NA	rs11676348		DGV	CXCR2	ENSG00000180871
rs4672873	fwd	G	A	BIFM	0.0523	NA	rs11676348		DGV	CXCR2	ENSG00000180871

5 UC and IBD Associated SNPs utilised in Chapter 2 UC Interactome – Upstream Gene variants (UGV) and Downstream gene variants (DGV). Highlighted are SNPs used in Chapter 4.

SNP/ID	Strand	Risk Allele	WT allele	Source	PICS	P value	SNP if FM	Enhancer	Annotation SNP/ID	Gene	Ensembl ID
rs10228276	fwd	G	A	BIFM	0.0528	NA	rs4722672	none	DGV	HOTTIP	ENSG00000243766
rs10910092	fwd	A	G	BIFM	0.1121	NA	rs10797432	none	DGV	TNFRSF15	ENSG00000157873
rs12598978	fwd	T	G	BIFM	0.2236	NA	rs11150589	th2	UGV	ITGAL	ENSG00000005844
rs12716977	fwd	T	C	BIFM	0.2236	NA	rs11150589	th2	UGV	ITGAL	ENSG00000005844
rs6451494	fwd	C	T	BIFM	0.0382	NA	rs6451493	colonic mucosa	UGV	lincRNA	ENSG00000283286
rs6883964	fwd	G	A	BIFM	0.0382	NA	rs6451493	colonic mucosa	UGV	lincRNA	ENSG00000283286
rs12655810	fwd	T	C	BIFM	0.0382	NA	rs6451493	colonic mucosa	UGV	lincRNA	ENSG00000283286
rs6888952	fwd	G	A	BIFM	0.0382	NA	rs6451493	colonic mucosa	UGV	lincRNA	ENSG00000283286
rs11742570	fwd	C	T	BIFM	0.0382	NA	rs6451493	colonic mucosa	UGV	lincRNA	ENSG00000283286
rs6890268	fwd	T	A	BIFM	0.0382	NA	rs6451493	colonic mucosa	UGV	lincRNA	ENSG00000283286
rs6871591	fwd	A	T	BIFM	0.0382	NA	rs6451493	colonic mucosa	UGV	lincRNA	ENSG00000283286
rs1445004	rev	T	C	BIFM	0.0382	NA	rs6451493	colonic mucosa	UGV	lincRNA	ENSG00000283286
rs1992660	rev	T	C	BIFM	0.0382	NA	rs6451493	colonic mucosa	UGV	lincRNA	ENSG00000283286

rs7117324	fwd	A	G	BIFM	0.1751	NA	rs561722	none	UGV	NXPE2P1	ENSG00000255982

5 UC and IBD Associated SNPs utilised in Chapter 2 UC Interactome – Upstream Gene variants (UGV) and Downstream gene variants (DGV). Highlighted are SNPs used in Chapter 4.

SNP/ID	Strand	Risk Allele	WT allele	Source	PICS	P value	SNP if FM	Enhancer	Annotation SNP/ID	Gene	Ensembl ID
rs34414754	fwd	C	A	BIFM	0.1056	NA	rs7524102	none	UGV		ENSG00000279625
rs3024505	rev	A	G	ICIBD IBDS BIFM BII	0.4133	6.66E-42	rs3024505	Th2	DGV	IL10	ENSG00000136634
rs907611	fwd	A	G	ICIBD IBDS BIFM BII	0.2159	2.70E-10	rs907611	colonic mucosa	UGV	LSP1	ENSG00000130592
rs4656958	fwd	G	A	ICIBD IBDS	NA	3.80E-09	rs4656958		UGV	ITLN1	ENSG00000179914
rs12946510	fwd	T	C	ICIBD IBDS	NA	4.10E-38	rs12946510		DGV	IKZF3	ENSG00000161405
rs10758669	fwd	C	A	ICIBD IBDS	0.9823	1.29E-26	rs10758669	Th1	UGV	JAK2	ENSG00000096968
rs6087990	fwd	C	T	ICIBD	NA	1.20E-09	rs4911259		UGV	DNMT3B	ENSG00000088305

5 UC and IBD Associated SNPs utilised in Chapter 2 UC Interactome – Upstream Gene variants (UGV) and Downstream gene variants (DGV)

SNP/ID	Strand	Risk Allele	WT allele	Source	PICS	P value	SNP if FM	Enhancer	Annotation SNP/ID
rs6017342	fwd	C	A	UCS ICUC BIFM BII	1	1.43E-43	rs6017342	colonic mucosa	non coding transcript variant
rs2836878	fwd	G	A	IBDS ICUC BIFM BII	0.179	4.62E-48	rs2836878	CD14+ cells	RRV
rs12568930	fwd	T	C	IBDS ICUC BIFM	0.0569	1.26E-17	rs12568930	colonic mucosa	RRV
rs2838519	fwd	G	A	BIFM BII	0.1914	NA	rs2838519	colonic mucosa	RRV
rs2310173	fwd	T	G	BIFM BII	0.5088	NA	rs2310173	colonic mucosa	RRV
rs7282490	fwd	G	A	BIFM	0.0678	NA	rs2838519	CD20+ cells	RRV
rs4817986	fwd	T	G	BIFM	0.179	NA	rs2836878	CD14+ cells	RRV
rs4817987	fwd	T	C	BIFM	0.179	NA	rs2836878	CD14+ cells	RRV
rs913678	fwd	T	C	IBDS ICIBD	NA	4.59E-08	rs913678		RRV
rs7282490	fwd	G	A	IBDS ICIBD	NA	2.35E-26	rs7282490		RRV

6 UC and IBD Associated SNPs utilised in Chapter 2 UC Interactome – regulatory region variants (RRV). Highlighted are SNPs used in Chapter 4.

SNP/ID	Strand	Risk Allele	WT allele	Source	PICS	P value	SNP if FM	Enhancer	Annotation SNP/ID
rs6426833	fwd	A	G	UCS ICUC BIFM BII	0.4767	2.39E-68	rs6426833	colonic mucosa	intergenic
rs477515	rev	G	A	UCS ICUC	NA	6.96E-62	rs6927022		intergenic
rs4380874	fwd	T	C	ICUC BII BIFM	0.7608	2.07E-26	rs4380874		intergenic
rs2413583	fwd	C	T	ICIBD IBDS	NA	4.40E-33	rs2413583		intergenic
rs17085007	fwd	C	T	IBDS ICUC BIFM BII	0.703	1.38E-11	rs17085007	colonic mucosa	intergenic
rs17694108	fwd	A	G	IBDS ICIBD	NA	5.85E-15	rs17694108		intergenic
rs9297145	fwd	C	A	IBDs ICIBD	NA	8.21E-12	rs9297145		intergenic
rs559928	fwd	C	T	IBDS ICIBD	NA	4.19E-11	rs559928		intergenic
rs4243971	fwd	G	T	IBDS ICIBD	NA	6.05E-10	rs6142618		intergenic
rs4957048	rev	C	T	BII BIFM	0.0634	NA	rs4957048	none	intergenic
rs7524102	fwd	A	G	BII BIFM	0.1056	NA	rs7524102	none	intergenic

rs7809799	fwd	G	A	BII BIFM	0.2779	NA	rs7809799	colonic mucosa	intergenic
rs11676348	fwd	T	C	BIFM BII	0.1055	NA	rs11676348	none	intergenic

7 UC and IBD Associated SNPs utilised in Chapter 2 UC Interactome – intergenic SNPs. Highlighted are SNPs used in Chapter 4.

SNP/ID	Strand	Risk Allele	WT allele	Source	PICS	P value	SNP if FM	Enhancer	Annotation SNP/ID
rs72703046	fwd	C	T	BIFM	0.0634	NA	rs4957048	none	intergenic
rs11955068	fwd	T	C	BIFM	0.0634	NA	rs4957048	none	intergenic
rs72703050	fwd	C	A	BIFM	0.0634	NA	rs4957048	none	intergenic
rs56410216	fwd	A	C	BIFM	0.0634	NA	rs4957048	none	intergenic
rs72703058	fwd	A	G	BIFM	0.0634	NA	rs4957048	none	intergenic
rs10737481	fwd	T	G	BIFM	0.4767	NA	rs6426833	none	intergenic

7 UC and IBD Associated SNPs utilised in Chapter 2 UC Interactome – intergenic SNPs

8.2 Cytoscape SNP inputs

SNP_ID	ID of effecting gene	Name of effecting gene	Type of interaction1	ID of SNP affected gene	Name of effected gene	loss_up/ gain_down
Unique ID of each SNP	miRNA or TF which has a binding site in the gene, empty if the gene is a mature miRNA which is lost		Type of interaction of the SNP	Effected gene by changed TFBS or miRNA BS or the mID/name of lost mature miRNA itself		outcome of SNP
	UNIROT ID, MIMAT ID IDs are separated by only a comma		miRNA BS, TFBS, mature miRNA loss_up	UNIROT ID MIMAT ID IDs are separated by only a comma		up reg or down reg of protein
rs1131095	hsa-miR-369-3p		mirna BS	P13798	APEH	loss_up
rs6911490			ESE	Q9H1Y0	ATG5	loss_up
rs41299637	hsa-miR-6511b-5p		mirna BS	Q3KP66	C1orf 106	gain_down
rs12132298	hsa-miR-7159-5p		mirna BS	Q3KP66	C1orf106	loss_up
rs59655222	hsa-miR-199a-5p		mirna BS	Q3KP66	C1orf106	gain_down
rs59655222	hsa-miR-199b-5p		mirna BS	Q3KP66	C1orf106	gain_down
rs59655222	hsa-miR-4733-5p		mirna BS	Q3KP66	C1orf106	loss_up
rs59655222			ESE	Q3KP66	C1orf106	gain_down
rs254562	hsa-miR-1229-5p		mirna BS	Q9H5L9	C5orf66	gain_down
rs254562			ESE	Q9H5L9	C5orf66	gain_down
rs10781499	hsa-miR-4502		mirna BS	Q9H257	CARD9	gain_down
rs10781499			splice site	Q9H257	CARD9	gain_down
rs10781499			ESE	Q9H257	CARD9	loss_up
rs12261843	hsa-miR-6870-3p		mirna BS	Q8ND76	CCNY	gain_down
rs6481950	hsa-miR-433-5p		mirna BS	Q8ND76	CCNY	loss_up
rs6481950	hsa-miR-4430		mirna BS	Q8ND76	CCNY	gain_down
rs12261843	hsa-miR-6870-3p		mirna BS	Q8ND76	CCNY	gain_down
rs727088	hsa-miR-2392		mirna BS	Q15762	CD226	loss_up

rs11230563	hsa-miR-4783-3p		mirna BS	P30203	CD6	gain_down
rs11230563	hsa-miR-4281		mirna BS	P30203	CD6	gain_down
rs11230563	hsa-miR-6849-5p		mirna BS	P30203	CD6	loss_up
rs11230563			ESS	P30203	CD6	gain_up
rs28374715	hsa-miR-5589-5p		mirna BS	Q99653	CHP1	gain_down
rs28671712	P10914	IRF1	TFBS	Q99653	CHP1	gain_up
rs28671712			ESE	Q99653	CHP1	gain_down
rs12103	hsa-miR-7113-3p		mirna BS	Q5TA45	CPSF3L	gain_down
rs12103	hsa-miR-7113-3p		mirna BS	Q5TA45	CPSF3L	gain_down
rs7495132	P55317	FOXA1	TFBS	Q6UUV7	CRTC3	gain_up
rs7495132	O60806	TBX19	TFBS	Q6UUV7	CRTC3	gain_up
rs7495132			ESE	Q6UUV7	CRTC3	gain_down
rs267984			ESE	P51397	DAP	gain_down
rs1801274	hsa-miR-204-5p		mirna BS	P12318	FCGR2A	gain_down
rs1801274	hsa-miR-6867-3p		mirna BS	P12318	FCGR2A	gain_down
rs4657041	A6NLW8	DUXA	TFBS	P12318	FCGR2A	loss_down
rs4657041	P10276	RARA	TFBS	P12318	FCGR2A	loss_down
rs1801274	Q9H0R8	ATG8	ELM	P12318	FCGR2A	gain
rs1801274	P49840	GSK3	ELM	P12318	FCGR2A	gain
rs1801274	P38398	BRCA1	ELM	P12318	FCGR2A	gain
rs11229555			ESE	Q6IB77	GLYAT	loss_up
rs1182188	hsa-miR-4433a-3p		mirna BS	Q03113	GNA12	loss_up
rs1182188	hsa-miR-6880-5p		mirna BS	Q03113	GNA12	loss_up
rs1182188	hsa-miR-4510		mirna BS	Q03113	GNA12	loss_up
rs1182188	hsa-miR-6760-5p		mirna BS	Q03113	GNA12	loss_up
rs1182188	hsa-miR-7847-3p		mirna BS	Q03113	GNA12	loss_up
rs1182188			ESE	Q03113	GNA12	loss_up
rs3749171			ELM	Q9HC97	GPR35	
rs8005161			ESE	Q8IYL9	GPR65	loss_up

rs11168249	Q9UJU2	LEF1	TFBS	Q8WUI4	HDAC7	loss_down
rs11168249	Q9NQB0	TCF7L2	TFBS	Q8WUI4	HDAC7	loss_down
rs12946510	Q06413	MEF2C	TFBS	Q9UKT9	IKZF3	gain_up
rs3024495	hsa-miR-4647		mirna BS	P22301	IL10	gain_down
rs3024495	O14978	ZNF263	TFBS	P22301	IL10	gain_up
rs11567699	Q99592	ZBTB18	TFBS	Q96F46	IL17R	loss_down
rs5771069	hsa-miR-6747-3p		mirna BS	Q6ZVW7	IL17REL	gain_down
rs5771069			ELM	Q6ZVW7	IL17REL	loss
rs5771069			splice site	Q6ZVW7	IL17REL	gain_down
rs11209026	Q29120	PCSK7	ELM	Q5VWK5	IL23R	loss
rs11209026			ESS	Q5VWK5	IL23R	gain_up
rs11150589	hsa-miR-548aa		mirna BS	P20701	ITGAL	gain_down
rs11150589	hsa-miR-548t-3p		mirna BS	P20701	ITGAL	gain_down
rs11150589	hsa-miR-548ay-3p		mirna BS	P20701	ITGAL	loss_up
rs11150589	hsa-miR-548at-3p		mirna BS	P20701	ITGAL	loss_up
rs12716977	hsa-miR-1268b		mirna BS	P20701	ITGAL	loss_up
rs12716977	hsa-miR-1268a		mirna BS	P20701	ITGAL	loss_up
rs11150589	P10914	IRF1	TFBS	P20701	ITGAL	gain_up
rs7657746	hsa-miR-3183		mirna BS	Q2LD37	KIAA1109	loss_up
rs7657746	hsa-miR-2114-5p		mirna BS	Q2LD37	KIAA1109	loss_up
rs7657746	hsa-miR-642a-5p		mirna BS	Q2LD37	KIAA1109	loss_up
rs7657746	hsa-miR-3184-3p		mirna BS	Q2LD37	KIAA1109	gain_down
rs1654644	hsa-miR-625-5p		mirna BS	P43630	KIR3DL2	loss_up
rs1654644	hsa-miR-4716-3p		mirna BS	P43630	KIR3DL2	loss_up
rs1654644	O14978	ZNF263	TFBS	P43630	KIR3DL2	loss_down
rs10896794	P10914	IRF1	TFBS	O60711	LPXN	gain_up

rs10896794	P52630	STAT1::STAT2	TFBS	O60711	LPXN	gain_up
rs10896794	P42224	STAT1::STAT2	TFBS	O60711	LPXN	gain_up
rs10896794			ESE	O60711	LPXN	gain_down
rs11041476	hsa-miR-3941		mirna BS	P33241	LSP1	loss_up
rs11041476			ESE	P33241	LSP1	loss_up
rs483905	hsa-miR-6839-5p		mirna BS	Q8IZL2	MAML2	loss_up
rs543104			ESE	Q8IZL2	MAML2	gain_down
rs2425019			ESE	Q9Y5R2	MMP24	gain_down
rs13085791	hsa-miR-6746-5p		mirna BS	P26927	MST1	loss_up
rs13085791	hsa-miR-8085		mirna BS	P26927	MST1	loss_up
rs9822268	hsa-miR-6769a-3p		mirna BS	P26927	MST1	gain_down
rs3774937			ESE	P19838	NFKB1	gain_down
rs2816958	hsa-miR-619-5p		mirna BS	O00482	NR5A2	gain_down
rs2816958	hsa-miR-6513-5p		mirna BS	O00482	NR5A2	loss_up
rs10891692	hsa-miR-3653-5p		mirna BS	Q8N323	NXPE1	gain_down
rs10891692	hsa-miR-1200		mirna BS	Q8N323	NXPE1	gain_down
rs10891692	hsa-miR-3192-5p		mirna BS	Q8N323	NXPE1	loss_up
rs10891692	hsa-miR-6761-5p		mirna BS	Q8N323	NXPE1	gain_down
rs10891692	P68133	ACTA1	ELM	Q8N323	NXPE1	gain
rs10891692			splice site	Q8N323	NXPE1	gain_down
rs10891692			ESS	Q8N323	NXPE1	gain_up
rs3766606			ESE	Q99497	PARK7	gain_down
rs1893217			ESE	P17706	PTPN2	loss_up
rs7596362			ESE	Q3MIT2	PUS10	loss_up
rs7608697			ESE	Q3MIT2	PUS10	loss_up
rs7608910			ESE	Q3MIT2	PUS10	loss_up

rs3208008	hsa-miR-6759-5p		mirna BS	Q9NZ71	RTEL1	loss_up
rs2257440	hsa-miR-661		mirna BS	Q9NZ71	RTEL1	loss_up
rs2297441	Q92826	HOXB13	TFBS	Q9NZ71	RTEL1	gain_up
rs2297441	P35453	HOXD13	TFBS	Q9NZ71	RTEL1	gain_up
rs3208008			ELM	Q9NZ71	RTEL1	loss
rs3208008			ESS	Q9NZ71	RTEL1	gain_up
rs2581817	hsa-miR-4538		mirna BS	Q9UHJ3	SFMBT1	gain_down
rs2581817			ESE	Q9UHJ3	SFMBT1	loss_up
rs12942547			ESE	P40763	STAT3	loss_up
rs2382817	hsa-miR-1291		mirna BS	Q969X1	TMBIM1	gain_down
rs10797432	hsa-miR-8073		mirna BS	Q92956	TNFRSF14	loss_up
rs10910092	hsa-miR-7157-3p		mirna BS	Q92956	TNFRSF14	loss_up
rs10797432	Q01101	INSM1	TFBS	Q92956	TNFRSF14	loss_down
rs2266959			ESE	P68036	UBE2L3	gain_down
rs6062504			ESE	Q8N5A5	ZGPAT	gain_down