The role of the hedgehog signalling pathway in inflammatory bowel disease.

Charles William Lees

PhD thesis presented to the University of Edinburgh - 2009

For Liz, Jessie, Holly and Cameron

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF PUBLICATIONS ARISING FROM THIS THESIS

Prizes

- ASNEMGE European Rising Star in Gastroenterology 2009
- Best abstract presented at European Crohn's and Colitis Organisation meeting 2007 (BSG fellow)
- Anne Ferguson prize (Scottish Society of Gastroenterology) 2006
- Sir James Black research prize (Scottish Society of Experimental Medicine) 2006
- Molecular Medicine Symposium prize (University of Edinburgh) 2006

Publications (peer-reviewed)

- **Lees CW**, Zacharias W, *et al.* Analysis of germline *GLI1* variation implicates hedgehog signalling in the regulation of intestinal inflammatory pathways. ***PLoS Medicine*** 2008; 5(12):e239
- Noble CL, Abbas A, Cornelius J, **Lees CW** *et al.* Characterisation of Dysregulated Intestinal Gene Expression Profiles in Ulcerative Colitis by Microarray Analysis. ***GUT*** 2008;57(10):1398-405.
- **Lees CW**, Ironside J, Wallace W, Satsangi J. Resolution of Non-small Cell Lung Cancer following Withdrawal of Anti-TNF Therapy. ***New England Journal of Medicine*** 2008 Jul 17;359(3):320-1.
- **Lees CW**, Satsangi J. Early combined immunosuppression in Crohn's disease. ***Lancet*** 2008;371(9629):1995.
- Anderson CA, Massey DC, Barrett JC, Prescott NJ, Tremelling M, Fisher SA, Gwilliam R, Jacob J, Nimmo ER, Drummond H, **Lees CW,** *et al.* Investigation of Crohn's Disease Risk loci in Ulcerative Colitis further Defines their Molecular Relationship. ***Gastroenterology*** 2008; Oct 25 [*Epub ahead of print*]
- Fisher SA, Tremelling M, Anderson CA, Gwilliam R, Bumpstead S, Prescott NJ, Nimmo ER, Massey D, Berzuini C, Johnson C, Barrett JC, Cummings FR, Drummond H, **Lees CW**, *et al.* New genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease. ***Nature Genetics*** 2008;40(6):710-2.
- Wellcome Trust Case Control Consortium (***Lees CW,*** *in authors*). Association scan of 14,500 non-synonymous SNP's in four diseases identifies autoimmunity variants. ***Nature Genetics*** 2007;39(11):1329-37

- **The Wellcome Trust Case Control Consortium** (***Lees CW,*** *in authors*). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. ***Nature*** 2007; 447(7145):661-678
- Parkes M, Barrett JC, Prescott NJ, Tremelling M, Anderson CA, Fisher SA, Roberts RG, Nimmo ER, Cummings FR, Soars D, Drummond H, **Lees CW,** *et al.* Sequence variants in the autophagy gene *IRGM* and multiple other replicating loci contribute to Crohn's disease susceptibility. ***Nature Genetics*** 2007; 39(7):830-2
- Ho GT, Mowat A, Potts L, Cahill A, Mowat C. **Lees CW,** *et al.* Efficacy and complications of Adalimumab treatment for medically refractive Crohn's disease: analysis of nationwide experience in Scotland (2004-2008). ***Alimentary Pharmacology & Therapeutics*** 2008; Dec 17 [*Epub ahead of print*]
- **Lees CW,** *et al.* The safety profile of anti-TNF therapy in inflammatory bowel disease in clinical practice: analysis of 620 patient-years follow-up. ***Alimentary Pharmacology & Therapeutics*** 2009; 29(3):286-297
- **Lees CW**, *et al.* Tolerability and safety of mercaptopurine in azathioprine intolerant patients with inflammatory bowel disease. ***Alimentary Pharmacology & Therapeutics*** 2008; 27(3):220-7
- **Lees CW**, *et al.* A retrospective analysis of the efficacy and safety of infliximab as rescue therapy in acute severe ulcerative colitis. ***Alimentary Pharmacology & Therapeutics*** 2007; 26(3):411-9
- Noble CL, McCullough J, Ho W, **Lees CW** *et al.* Low body mass not vitamin D receptor polymorphism predicts osteoporosis in patients with inflammatory bowel disease. ***Alimentary Pharmacology & Therapeutics*** 2008; 27(7):588-96
- Ho GT, Smith L, Aitken S, Lee HM, Ting T, Fennell J, **Lees CW** *et al.* The use of adalimumab in the management of refractory Crohn's disease. ***Alimentary Pharmacology & Therapeutics*** 2008; 27(4):308-15
- Van Limbergen J, Russell RK, Nimmo ER, Torkvist L, **Lees CW** *et al.* Contribution of the *NOD1/CARD4* insertion/deletion polymorphism +32656 to inflammatory bowel disease in Northern Europe. ***Inflammatory Bowel Diseases*** 2007; 13(7):882-9
- **Lees CW** and Satsangi J. Hedgehog, Paneth cells and colon cancer: a cautionary note for the use of systemic agonists/antagonists. ***Gastroenterology*** 2006; 131(5):1657-8
- **Lees CW**, *et al.* The role of infliximab in ulcerative colitis: further questions. ***Inflammatory Bowel Diseases*** 2006; 12(4):335-7

- **Lees CW**, *et al.* The hedgehog signalling pathway in the gastrointestinal tract: Implications for development, homeostasis and disease. ***Gastroenterology*** 2005; 129:1696-1710
- Ho GT, **Lees CW**, *et al.* Pharmacogenetics and Inflammatory Bowel Disease: Progress and Prospects. ***Inflammatory Bowel Diseases*** 2004; 10(2):148-158

Invited reviews (peer reviewed)

- **Lees CW,** Satsangi J. Autophagy and Crohn's disease. ***IBD Monitor*** 2008; 9(2):49-55
- **Lees CW**, Ho GT, Satsangi J. The Genetics of Inflammatory Bowel Disease. ***Encyclopaedia of Life Sciences*** 2007. DOI: 10.1002/9780470015902.a0020220
- Ho GT, **Lees CW**, Satsangi J. Ulcerative Colitis. ***Medicine*** 2007; 35(5):277-282

Book chapters

- Gaya DR*, **Lees CW***, Satsangi J. Recent progress in inflammatory bowel disease. In: Rhodes J (ed), ***Horizons in Medicine*** volume 18. London: Royal College of Physicians, 2006. **joint first authors*

LIST OF PUBLISHED ABSTRACTS

Oral presentations

- **Lees CW**, *et al.* Expression analysis of all genes implicated in susceptibility to Crohn's disease from genome-wide association studies. ***Gastroenterology*** 2008;134:A41. ***Digestive Diseases Week (DDW) 2008***
- Zacharias WJ, **Lees CW**, *et al.* Epithelial hedgehog signals modulate the inflammatory response of the intestinal lamina propria. ***Gastroenterology*** 2008;134:A255. ***DDW 2008***
- **Lees CW**, *et al.* Cervical dysplasia and IBD: no effect of disease status or immunosuppressants on analysis of 2,199 smear records. ***Gastroenterology*** 2008;134:A143-144. ***DDW 2008***
- **Lees CW**, *et al.* The safety profile of infliximab for IBD in Edinburgh: final analysis of over 200 patients. ***GUT*** 2008;57:A35-36. ***British Society of Gastroenterology (BSG) 2008***
- **Lees CW**, *et al.* A retrospective analysis of small bowel MRI in evaluation of disease activity in ileal Crohn's disease. ***GUT*** 2008;57:A34-35. ***BSG 2008***
- **Lees CW**, *et al.* Haplotypic variation of the *GLI1* locus determines susceptibility and phenotype in ulcerative colitis. ***GUT*** 2007;56:A21-A22. ***BSG 2007, European Crohn's Colitis Organisation (ECCO) 2007, U.K. Association of Physicians Annual Meeting 2007***, ***DDW 2006***
- **Lees CW,** *et al.* Infliximab as rescue therapy in acute severe UC: a survey of the Scottish Society of Gastroenterology (SSG). ***GUT*** 2007;56:A1. ***BSG 2007***
- **Lees CW**, *et al.* Pregnancy outcome and fertility in inflammatory bowel disease is determined by disease phenotype. ***GUT*** 2007;56:A21. ***BSG 2007***
- **Lees CW**, *et al.* Tolerability of mercaptopurine in azathioprine intolerant patients with inflammatory bowel disease. ***GUT*** 2007;56:A24. ***BSG 2007***
- van Limbergen J, **Lees CW**, *et al.* Association of a complex insertion/deletion polymorphism of NOD1/CARD4 with susceptibility to inflammatory bowel disease in the Scottish population. ***Gastroenterology*** 2006;130:A64. ***DDW 2006***
- **Lees CW**, *et al.* Analysis of CCL20 variants in IBD provides further evidence for genetic heterogeneity in disease susceptibility. ***GUT*** 2006;55:A1. ***BSG 2006***

Poster presentations

- **Lees CW**, *et al.* Infliximab is effective as rescue therapy for acute severe ulcerative colitis - The initial Edinburgh experience. ***GUT*** 2006;55:A79. ***DDW 2006; BSG 2006***

- **Lees CW**, *et al.* The hedgehog (Hh) signalling pathway is differentially expressed in the healthy adult colon, and upregulated in ulcerative colitis and Crohn's disease. ***Gastroenterology*** 2005;128:A504. ***DDW 2005***

DECLARATION OF ORIGINALITY

I declare that all of the work in this thesis unless otherwise indicated is entirely my own, performed in the gastrointestinal laboratory (University of Edinburgh, Western General Hospital) and the Centre for Inflammation Research (University of Edinburgh) between August 2003 and August 2007. This work has not been submitted for any other professional degree or qualification.

The experiemtal work and subsequent analysis was all carried out by Charlie Lees with the exception of:

- Generation of raw microarray dataset and RT-PCR on biopsies performed by Colin Noble and Jenine Cornelius (at Genentech Inc., San Francisco, CA, USA).
- TaqMan genotyping performed at the Wellcome Trust Clinical Research Facility by Angie Fawkes and Anglea White.
- Sequencing reactions performed at MRC Human Genetics Unit, University of Edinburgh.
- Assistance with statistical programming from Albert Tenesa and Harry Campbell (MRC HGU), specifically the logistic regression analysis in 'R', the meta-analysis of *GLI1* SNP rs2228226, and the estimation of the false positive reporter probablility for *GLI1*.
- William Zacharias assisted with the phylogeny analysis of *GLI1*.
- Patient phenotyping was performed jointly by various members of the clincial and research team in the gastrointestinal unit.

ACKNOWLEDGEMENTS

I am indebted to the supervision and guidance of my supervisors, Professor Jack Satsangi and Professor Sarah Howie. Jack has been fiercely supportive and an inspiration both during my time in his laboratory and in helping me prepare thereafter for a career in academic gastroenterology and as an inflammatory bowel disease specialist.

I would like to acknowledge the help and support of all those in the GI Unit and at the Centre for Inflammation Research in Edinburgh: Elaine Nimmo for her patience in teaching key molecular biology methods, Marian Aldhous for teaching me tissue culture techniques, Norman Anderson for general instruction in the lab, Hazel Drummond for maintaining the database, Linda Smith and Janice Fennell for recruiting patients, Colette McColl for invaluable administrative support, Anne Grant for help with immunohistochemistry and Robert Benson and Paul Fitch for patiently teaching me basic and advanced immunology as well as endless discussions about hedgehog signalling. I would particularly like to thank the other research fellows with whom I have shared this voyage of discovery, and for making the journey so much fun: Colin Noble, Richard Russell, Gwo-Tzer Ho and Johan van Limbergen. I must also acknowledge the support and tolerance of the clincial gastroenterologists at the Western General Hospital: Kelvin Palmer, Ian Penman, Alan Shand and Ian Arnott.

Much of the work presented here would not have been possible without key national and international collaborators: Miles Parkes and Mark Tremelling in Cambridge, Leif Torkvist in Stockholm, Nobuhide Oshitani in Osaka and William Zacharias and Deborah Gumucio at Ann Arbor. Will has been a truly fantastic collaborator and, despite the endless revisions and rejections, a rock as we prepared the GLI1 manuscript for publication. Deb has been a true inspiration and I am thoroughly looking forward to working in her lab and with her team in Ann Arbor over the next 2 to 3 years.

My thanks to all the patients and families with inflammatory bowel disease who have consented to the various studies described in this thesis.

Finally, but most importantly, my heartfelt thanks to my family: Liz, Jessie, Holly and Cameron who have demonstrated patience and tolerability in its extreme whilst I have stressed over data, deadlines, abstract, paper and grant submissions, and thesis writing.

ABBREVIATIONS

APC	Adenomatous polyposis coli
5-ASA	5-amino salicylic acid
ATG16L1	Autophagy related 16-like 1
AZA	Azathioprine
CCL20	Chemokine (C-C motif) ligand 20
CD	Crohn's disease
CI	95% confidence interval
CNV	Copy number variation
CTNNB1	Catenin (cadherin-associated protein), beta 1
DC	Dendritic cell
DHH	Desert hedgehog
DISP1	Dispatched
DKK1	Dickopf homolog 1
DKK2	Dickopf homolog 2
DKK3	Dickopf homolog 3
DKK4	Dickopf homolog 4
DSS	Dextran sodium sulphate
ENS	Enteric nervous system
ER	Endoplasmic reticulum
FZD4	Frizzled homolog 4
FZD5	Frizzled homolog 5
FZD6	Frizzled homolog 6
FZD7	Frizzled homolog 7
GF	Germ-free
GI	Gastrointestinal
GLI1	Glioma-associated oncogene homolog 1
GLI2	Glioma-associated oncogene homolog 2
GLI3	Glioma-associated oncogene homolog 3
GO	Gene ontology
GSK3α	Glycogen synthase kinase-3 alpha
GSK3β	Glycogen synthase kinase-3 beta
GWAS	Genome-wide association study
HC	Healthy controls
HHIP	Hedgehog-interacting protein
H_2O_2	Hydrogen peroxide
I	Inflamed
IBD	Inflammatory bowel disease

IBDU	Colonic inflammatory bowel disease – type unspecified
IECs	Intestinal epithelial cells
IFNγ	Interferon gamma
IL	Interleukin
IHH	Indian hedgehog
IQR	Inter-quartile range
IRGM	Immunity-related GTPase family, M
KCTD11	Potassium channel tetramerisation domain containing 11
LD	Linkage disequilibrium
LPS	Lipopolysaccharide
LRP	Low-density lipoprotein receptor-related
LRR	Leucine rich region
mLN	Mesenteric lymph node
MDP	Muramyl dipeptide
MDR1	Multi-drug resistance 1
MIM	Mendelian inheritance in man
MP	Mercaptopurine
mTOR	Mammalian target of rapamycin
MTX	Methotrexate
NFκB	Nuclear factor kappa B
N-I	Non-inflamed
NKX2.3	NK2 transcription factor related, locus 3
NLRP3	NLR family, pyrin domain containing 3
NNT	Number needed to treat
NOD2	Nucleotide-binding oligomerizatoin domain containing 2
nsSNP	non-synonymous single nucleotide polymorphism
OMIM	Online Mendelian inheritance in man
OR	Odds ratio
PAR	Population attributable risk
PBS	Phosphate buffered saline
PGLYRP	Peptidoglycan recognition protein
PPAR	Peroxisome proliferator-activated receptor gamma
PTCH	Patched
PTCH2	Patched-2
PTPN2	Protein tyrosine phosphatase non-receptor type 2
Q-PCR	Real-time reverse trancriptase PCR
RCLB	Red cell lysis buffer
rSHH	Recombinant sonic hedgehog

RT	Room temperature
sFRP1	Secreted frizzled-related protein 1
sFRP5	Secreted frizzled-related protein 5
siRNA	Small interefering RNA
SHH	Sonic hedgehog
SMO	Smoothened
SNP	Single nucleotide polymorphism
SPF	Specific pathogen free
SUFU	Suppressor of Fused
TBS	Tris-buffered saline
TCF1	Transcription factor 1
TCF2	Transcription factor 2
TCF3	Transcription factor 3
TCF4	Transcription factor 4
TCPTP	T cell protein tyrosine phosphatase
TDT	Transmission disequilibrium testing
TG	Transgenic
TLR	Toll-like receptor
TNFα	Tumour necrosis factor alpha
tSNP	Tagging single nucleotide polymorphism
UC	Ulcerative colitis
WT	Wild-type
WTCCC	Wellcome trust case control consortium
WTCRF	Wellcome trust clinical research facility (Genetics core)
XBP1	X-box-binding protein 1

Human and murine genes are indicated throughout the text in italicised capitals and italicised lower case respectively. Human and murine gene products are in capitals and lower case respectively.

ABSTRACT

Introduction. The inflammatory bowel diseases (IBD), Crohn's disease (CD) and ulcerative colitis (UC), are common in Western Europe (200-400 cases /100,000) and associated with substantial morbidity, although mortality is now low. There is presently a great unmet need for novel therapeutics in IBD as present agents are limited by lack of efficacy, toxicity and poor patient acceptance. Recent findings from genome-wide association studies (GWAS) have characterised the genetic architecture of CD and UC. Defects in innate and adaptive immunity have been clearly established, and substantial novel insights into disease pathogenesis have been gained. Over 30 genes / loci are now associated with CD; a number of these, along with a few specific loci, are also associated with UC. The hedgehog (HH) signalling pathway is critical to gastrointestinal development and plays key roles in intestinal and immune homeostasis. Furthermore, in addition to well described roles in tumorigenesis, it is evident that recapitulation of embryonic HH signals play critical roles in response to acute and chronic inflammatory challenge in diverse tissues.

Aims. The main aims of the work presented in this thesis were to characterise the expression of key HH signalling components in the healthy and inflamed human intestine, establish whether germline variation in HH genes is associated with IBD and describe the *in vitro* responses of intestinal epithelial cells to pathogen associated molecular patterns. The WNT pathway, antagonised by HH in the intestine, and two HH target genes (*NKX2.3* and *CCL20*) were also analysed for evidence of association with IBD.

Methods. Expression of HH and WNT signalling components was described by immunohistochemistry and microarray analysis in healthy controls (HC), CD, UC, and non-IBD inflamed terminal ileal and colonic samples. Gene-wide haplotype-tagging studies were performed for *GLI1* in Scottish, English and Swedish CD and UC, and Scottish early-onset colo-rectal cancer, *IHH* in Scottish IBD, *NKX2.3* in Scottish and UK IBD, and *CCL20* in Scottish, Swedish and Japanese IBD. Evidence for association of all HH (n=13) and WNT (n=27) signalling genes in CD was established by analysis of UK GWAS data and meta-analysis from UK, French/Belgium and N American studies. The effect of lipopolysaccharide (LPS) and muramyl dipeptide (MDP) on HH signalling was assessed in colonic epithelial cells (SW480). The effect of HH pathway agonists and antagonists on NFκB activity and cytokine expression was analysed in SW480 cells and peripheral blood mononuclear cells (HC and IBD patients) *in vitro*.

Results. The expression of HH pathway ligand is present in the intestinal epithelium and the pathway response network in the lamina propria demonstrating the paracrine nature of HH

signalling in the intestine. Immunohistochemical studies and microarray analysis demonstrates that HH pathway activity is decreased in all forms of colonic inflammation studied in man. Variation in *Glioma-associated oncogene homolog 1* (*GLI1*), a key HH transcription factor located at 12q13 (IBD2), was associated with IBD ($p<0.0001$), UC ($p<0.0001$) and to a lesser extent CD ($p=0.03$) in Scotland, a finding replicated in English IBD and UC. This association was attributed to a non-synonymous SNP (rs2228226C→G) with pools odds ratio of 1.194 in meta-analysis of over 5000 individuals from Scotland, England and Sweden ($p=0.0002$). There was association of this SNP with early-onset colo-rectal cancer, but of borderline significance ($p=0.05$). The variant protein (Q1100E) is 50% less active than wild-type protein *in vitro*. *IHH* was not associated with CD or UC. Preliminary evidence was produced for association at *SUFU* (10q24; $p=0.005$), a GLI1-binding protein, and at the *WNT3 / WNT9B* locus (17q21; $p=0.0005$). MDP stimulation of colonic epithelial cells decreased HH pathway activity. Exogenous HH increased expression of CCL20. *CCL20* promoter polymorphisms were associated with UC in Japanese patients ($p=0.018$) but not in Scotland or Sweden. *NKX2.3* was associated with IBD in Scotland (UC>CD), but there was insufficient power for fine-mapping of causative variants.

Conclusions. Multiple lines of evidence presented here demonstrate that the HH signalling pathway is involved in IBD pathogenesis. In key complementary *in vivo* studies (conceived by CWL; conducted in collaboration with the Gumucio lab in Ann Arbor) we have demonstrated that *Gli1*$^{+/-}$ mice develop early, severe colitis with high mortality in response to acute inflammatory challenge. Furthermore, lamina propria antigen presenting cells are identified as the key HH target cells. With HH agonists and antagonists in extensive pre-clinical and early clinical testing, these studies have real potential to translate into novel therapeutics for patients with IBD.

1 Introduction

1.1 *Inflammatory Bowel Disease - Background*

Crohn's disease (CD; MIM 266600) and ulcerative colitis (UC; MIM 191390) are idiopathic, chronic relapsing inflammatory bowel diseases (IBD) of high prevalence in the UK, N Europe and N America (200-400 IBD cases per 100,000).[1] After sharp increases in the past 50 years, incidence rates are now stabilising in these countries. In contrast, rates continue to rise in S Europe, Asia and other developing countries where IBD was essentially unheard up until a few decades ago.[2] Changes in prevalence after migration have been documented in Jewish populations, the Chinese (moving from the mainland to Hong Kong) and in migrants from the Indian sub-continent to England.

Whilst mortality rates have fallen considerably in the last 50 years, significant morbidity associated with disabling symptoms (abdominal pain, diarrhoea and lethargy), medical therapies and surgical intervention is common. This is of particular concern in patients with early-onset disease, where problems with growth, puberty, schooling, employment prospects, social and sexual well-being are all too frequently documented. IBD is associated with extra-intestinal manifestations affecting the joints, skin, eyes and liver, and with an increased risk of colo-rectal cancer, related to the duration and extent of active colonic inflammation.

1.2 *Treatment of IBD*

Since the first trials of corticosteroids,[3] medical management has evolved with the only paradigm shift in the past two decades following the licensing of anti-tumour necrosis factor alpha (TNFα) antibody therapy for CD in the past 10 years. The 'traditional' therapeutic options available for the management of IBD consist of 5-aminosalicylates (5-ASA), corticosteroids, immunosuppressants and surgery (**Figure 1-1**). Corticosteroid dependence and resistance,[4] and complications with existing therapies remain important limiting factors in clinical practice. There is very little evidence to demonstrate that any existing therapies alter the natural history of disease, with the exception of colectomy which is curative in UC (even now, some commentators are challenging this supposition, describing pouchitis as 'recurrent IBD' of the pouch).[5] However, there is great excitement and hope for novel therapies following a revolution in our understanding in the pathogenesis of IBD, largely catalysed by recent insights resulting from genetic studies.

1.2.1 Thiopurines

Immunosuppression with azathioprine (AZA) or mercaptopurine (MP) is now regarded as the standard of care for patients with steroid refractory or steroid dependant IBD. There is proven efficacy for the induction of remission (OR 2.36, NNT=5) and there is also a potent steroid sparing effect (OR 3.86, NNT=3).[6] For patients entering a medically induced remission, AZA or MP are an effective maintenance strategy (OR 3.17, NNT=3.3 favouring AZA 2mg/kg over placebo),[7] offering an acceptable balance of efficacy, tolerability and cost. However, ability to tolerate thiopurines is not universal with between 5.7% and 22.0% of patients will need to stop AZA or MP due to adverse events.[8] Myelosuppresssion, a potentially life threatening adverse event, is fortunately relatively infrequent with treatment more often limited by nausea, allergy and a flu like illness.[8] The toxicities of thiopurine therapy have long been of concern to physicians, with patients routinely counselled about short term concerns of leucopenia and infectious complications as well as the small, but definite increased risk of lymphoma in later life.[9-11] The choice of whether to use AZA or MP has largely been based on historical preference rather than clinical evidence, with AZA use commonplace in Europe, whilst MP has been favoured in N America. We have shown that 59% of AZA intolerant patients were subsequently able to tolerate MP (Lees *et al*, *AP&T* 2008).[12]

Inter-individual variations in drug metabolism are important determinants of tolerability to AZA and MP. The thioguanine derivative AZA is a prodrug, which is rapidly, and almost completely (88%) broken down in the liver by non-enzymatic cleavage of the methylnitroimidazole ring to MP. This reaction is aided by glutathione and other endogenous sulphydryl-containing proteins, levels of which decrease in smokers and with advancing age.[13] Assuming 100% bioavailability, a conversion factor of 2.07 should be used to convert a dose of MP to that of AZA.[14] There is then extensive metabolic transformation by competing catabolic (xanthine oxidase and thiopurine methyltransferase (TPMT)) and anabolic (hypoxanthine phosphoribosyl transferase) enzymatic pathways.[15] Once formed, 6-thiosine-5′-monophosphate may be transformed either into 6-thioguanine nucleotides (6-TGN) by inosine monophosphate dehydrogenase and guanine monophosphate synthetase enzymatic pathways or methylated to 6-methylmercaptopurine (6-MMPR).

The thioguanine nucleotide (TGN) metabolites act as purine antagonists and induce cytotoxicity and immunosuppression by inhibition of RNA, DNA and protein synthesis.

These cytotoxic properties are, at least in part, due to the direct incorporation of TGN into DNA.[16] There is also direct evidence that AZA/MP induces T cell apoptosis by blocking CD28 signalling via Rac-1. It is thought that AZA metabolite 6-thioguanine triphosphate binds directly to Rac-1 and on CD28 costimulation brings about apoptosis and thus suppresses the activation of Rac1 target genes, such as mitogen-activated protein kinase kinase, NFκB, and $bclx_L$.[17, 18]

The assessment of the activity of TPMT, one of the key metabolising enzymes, has allowed clinicians to identify the 1 in 300 patients who are deficient and at are at very high risk of potentially fatal neutropenia on AZA.[13] Mutations in the TMPT gene have been shown to correlate with enzyme activity. The commonest variant alleles are TPMT*3A, *3C and *2. In a healthy population these occurred at a frequency of 4.4%, 0.4% and 0.2% respectively. All individuals that were deficient in TPMT were homozygous or compound heterozygous at one or more of these sites.[19] Emerging evidence suggests that dose reduction in TPMT carriers with intermediate enzyme activity can limit treatment related adverse events[20, 21] but other studies have failed to demonstrate this association.[22, 23]

1.2.2 Anti-TNF therapy

There are presently two biological agents licenced for the treatment of IBD in the UK; both are monoclonal antibodies against TNFα. Infliximab is a chimeric anti-TNF antibody, consisting of 75% human IgG and 25% murine component that actively binds membrane bound and soluble TNFa. Infliximab is given by intravenous infusion only. Adalimumab is a humanized anti-TNF antibody, given by sub-cutaneous injection only. The efficacy of anti-TNF antibodies in CD is well established with infliximab, adalimumab and certolizumab pegol (N America only) presently licensed for luminal and fistulising disease.[24-31] Infliximab is also licensed for moderate UC based on data from ACT1/2 trials,[5] and has become an alternative to ciclosporine as rescue therapy in acute severe UC failing first-line medical therapy.[32, 33] Infliximab is the only anti-TNF antibody licensed for paediatric use in North America and Europe.[34] In CD, most current guidelines recommend that these agents are reserved for those patients refractory to standard medical therapy, thus maximising the relative benefit in any particular patient.

1.2.3 Safety of anti-TNF therapy

The most common reported side-effects with anti-TNF therapy are acute infusion reactions, delayed hypersensitivity reactions and infections.[35, 36] In particular, anti-TNF therapy has

been shown to be associated with tuberculosis[37] and opportunistic infections.[38, 39] While concomitant immunosuppression has been shown to decrease formation of antibodies to infliximab and hence the likelihood of infusion reactions,[40, 41] it has also been shown to increase the likelihood of opportunistic infections.[38] Anti-TNF therapy has also been associated with drug-induced lupus and demyelinating complications.

We have recently reported on the full safety profile of the use of anti-TNF agents in IBD in Edinburgh (Lees *et al*, *AP&T* 2009).[42] This cohort comprised 202 patients (157 CD, 42 UC, and 3 coeliac disease), with median follow-up of 2.4 years (IQR 1.0-4.9) and a total of 620 patient-years. 19.1% of CD patients were subsequently treated with adalimumab. There were 7 deaths (3.3%) during follow-up – only one death was <1 year post-infliximab (at day 72, from lung cancer). 6 malignancies (3 haematological, 3 bronchogenic) and 6 cases of suspected demyelination (3 with confirmed neurological disease) were reported. In the 90 days following infliximab, 95 adverse events (36 serious) occurred in 58/202 (28.7%) patients. 42/202 (20.8%) had an infectious event (22 serious). 27/202 (13.4%) of patients had an infusion reaction - 19 acute (4 serious); 8 delayed (3 serious).

New signals for real concern have continued to emerge over recent years. Most recently, a series of 8 cases of hepatosplenic T cell lymphoma (HSTCL) were reported in young CD patients concomitantly treated with infliximab and a thiopurine;[43] this had risen to 13 in predominantly paediatric IBD with such concomitant treatment by November 2007 (letter from Schering-Plough). The aggressive and near universal fatality of these lymphomas has coloured physicians approach to maintenance therapy with both agents in this patient group.

There are also reports of lung cancers in smokers treated with anti-TNF therapy. In the Mayo Clinic series of 500 CD patients, two lung cancers, deemed 'possibly related' to infliximab, were reported in elderly smokers.[35] A 24-week trial in COPD reported on 157 infliximab treated patients; in total 11 malignancies including 6 lung cancers were diagnosed in this group.[44] In our Unit, we have seen three lung cancers in anti-TNF treated patients, all smokers/ex-smokers over the age of 65.[42]

One of these was in a 69 year old female with CD colitis, diagnosed in 2000, who had required treatment with methotrexate (2003-present), together with anti-TNF therapies - initially episodic infliximab (2000-2004), and thereafter maintenance adalimumab (Lees *et*

al, *NEJM* 2008).[45] In 1999, the patient, an ex-smoker (35 pack years) was noted to have finger clubbing. CT scan demonstrated interstitial lung disease; she had normal pulmonary function tests, but remained under respiratory review. On repeat CT in June 2006, a 2.5x2.5cm pleural-based spiculated mass was noted in the right lower lobe (**Figure 1-2a**) with two small satellite appendages. Mediastinal and hilar lymphadenopathy were present. CT-guided biopsy of the main lesion confirmed non-small cell lung cancer (**Figure 1-2b**), TNFR1 and TNFR2 positive on immunostaining (**Figure 1-2b**). The tumour stage was T4N2M0. Adalimumab was withdrawn but methotrexate was continued. She received no surgery, radiotherapy, chemotherapy or biological therapy. In April 2007, on repeat CT scan there was virtually no evidence of primary lung tumour, nodules or lymphadenopathy (**Figure 1-2c**). To date, she remains in complete clinical and radiological remission.

The remarkable aspect of this case is the sustained remission on withdrawal of anti-TNF therapy. The over-expression of TNFR1/R2 implicates TNF blockade in carcinogenesis. This hypothesis is supported by data demonstrating a critical role for adaptive immunity in maintaining tumour dormancy following exposure to a carcinogen in cigarette smoke.[46] We recommend special vigilance in the use of anti-TNF therapy in patients with smoking histories, or chronic lung disease, especially those over 65.[45]

1.2.4 Step-up versus top-down therapy in CD

Some investigators have recently suggested inverting the therapeutic pyramid in CD, in much the same way as has been successfully implemented in rheumatoid arthritis. In the first trial to report on an early combined immunosuppression strategy in CD, patients were randomised to 'top-down' (AZA and infliximab) or 'step-up' (convential treatment escalation from corticosteroids to AZA to anti-TNF therapy only if required).[47] In this trial, there was an increased likelihood of remission at 54 weeks in patients treated with this strategy, compared with a conventional approach. However, on grounds of limited efficacy as well as unproven safety, caution should be exercised in any attempts to translate these data directly to clinical practice.

Firstly, the data provided no evidence for sustained benefit. At all time points beyond 54 weeks, remission rates were the same, regardless of treatment strategies. Indeed, the 'top-down' cohort is disadvantaged in the treatment of a subsequent relapse, as these patients will have an increased risk of a hypersensitivity reaction to infliximab on repeat exposure.

Secondly, a 'top-down' strategy for all newly-diagnosed patients will expose a significant number to risks of anti-TNF therapy unnecessarily. The long-term safety of anti-TNF strategy is unproven, and has not been adequately addressed by existing registries. We advise strongly against changing treatment algorithms in CD until we are able to identify, by clinical, serological, proteomic or genetic testing, those patients who will derive most benefit, and thereby protect those at most risk (Lees and Satsangi, *Lancet* 2008).[48]

1.2.5 Medical treatment of acute severe UC

The use of corticosteroids and appropriate early surgical intervention for acute severe ulcerative colitis (UC) has reduced the mortality from 31-60% to 1-2%.[49] Reproducible data have demonstrated that 30-40% of patients will not respond to corticosteroid therapy and will need urgent colectomy.[50-56] This may be a life saving procedure, but is feared by patients: quality of life is diminished,[57] psychological morbidity may be considerable,[58] and a 48% reduction in female fecundity has been reported following pouch surgery.[59] There is therefore a need for medical 'rescue' therapy for patients failing steroids. Historically, cyclosporine has been the only option, but recently infliximab has emerged as an alternative.

1.2.6 Rescue therapy in acute severe UC: infliximab or ciclosporine?

In 1994 Lichtiger and colleagues analysed the use of ciclosporine as rescue therapy for acute severe UC. In a small RCT of 20 patients they observed significantly lower colectomy rates in patients given active drug (18%) rather than placebo (100%).[60] Further RCTs have failed to fully replicate these results and longer-term outcomes have been disappointing.[61] A recent Cochrane meta-analysis of the use of ciclosporine in acute severe UC concluded that "…there is limited evidence that cyclosporine is more effective than standard treatment alone for severe UC. The relatively quick response makes the short-term use of cyclosporine potentially attractive, but the long-term benefit is unclear…".[62] This uncertainty, together with widely-held concerns regarding toxicity, has led to only sporadic use in UK or indeed worldwide.

The first RCT of infliximab in UC assessed 42 patients with moderately active steroid resistant disease.[63] This was a negative but probably underpowered study. Observational case series implied modest response rates in heterogeneous groups of UC patients.[64-70] In December 2005, the Centocor sponsored ACT1 and ACT2 trials demonstrated the efficacy of infliximab in 728 patients with moderately active UC (mean Mayo score 8.4).[5] Patients were included if there was evidence of active UC together with previous lack of efficacy or

intolerance of at least one treatment. In ACT2 this may have been 5-ASA alone. Combined ACT1 and ACT2 data in patients treated with infliximab 5mg/kg demonstrated an 8-week remission rate of 33.2%, a response in 66.9% (the primary efficacy endpoint) and mucosal healing (including mild inflammation) in 61.1%. There were no differences between the 5mg/kg and 10mg/kg doses. At 1 year (ACT1 only) sustained remission was seen in 19.8% and sustained response in 38.8% of patients; significantly better results being seen in patients on active treatment rather than placebo. It is of note that steroid-free remission was seen in only 22%. Colectomy data are presently not available for this trial.

Jarnerot and colleagues assessed the use of infliximab in acute severe UC in a publicly funded RCT in Scandinavia.[32] 45 patients who were admitted to hospital with acute severe UC (defined by the Seo index) were randomised to receive a single infusion of infliximab 5mg/kg (24 patients) or placebo (21 patients) either with fulminant colitis on day 4 or severe colitis on days 6-8. The primary outcome analysis was colectomy or death at 90 days. Patients who received infliximab required colectomy less frequently than those who received placebo (7/24 [29.2%] v 14/21 [66.7%], p=0.017, OR 4.9, C.I. 1-4-17). We observed very similar response rates at the Western General Hospital (Edinburgh) in a preliminary report in 2006; 6 out of 9 patients with acute severe UC (Truelove and Witts criteria) responded and avoided colectomy with relatively few adverse events.[71]

We have subsequently reported a retrospective cohort of 39 patients (median age 31.7 years) from throughout Scotland who received infliximab as rescue therapy for acute severe UC (Lees et al, *AP&T* 2007).[33] 26/39 (66%) responded, avoided colectomy during the acute admission, and were followed up for a median of 203 days [IQR 135.5–328.5]). Hypoalbuminaemia was a consistent predictor of non-response on univariate and multivariate analysis. At day 3 of iv steroids, 9/18 (50.0%) with serum albumin <34g/l had urgent colectomy versus 1/13 (7.7%) ≥34g/l (p = 0.02, OR 12.0, C.I. 1.28-112.7). Two serious adverse events occurred in this study – one death due to Pseudomonas pneumonia, and one post-operative fungal septicaemia.

1.3 *Pathogenesis of IBD*

The intestine has the primary function of absorbing the nutrients required for all post-natal developmental and homeostatic processes. Having been colonised by ~10^{14} bacteria during early life, intestinal homeostasis is critical to simultaneously protect the host against

pathogens whilst avoiding excessive inflammatory responses to commensal flora. The single-layered epithelium, consisting of enterocytes, goblet cells, Paneth cells, and microfold (M) cells, is all that stands between the intestinal lumen and the mucosal immune system. Key epithelial-microbial interactions occur across the epithelium. This active barrier, expressing a host of receptors designed to interact with different bacterial motifs (e.g. toll-like receptors (TLRs)), is in constant contact with the luminal contents and communicates with the richly diverse commensal flora under healthy conditions to establish immune homeostasis. Paneth cells secrete specialised antimicrobial peptides known as defensins. M cells and lamina propria dendritic cells (DCs) can directly sample intraluminal microbial components.

The precise aetiology of CD and UC remains unknown. However, recent advances in basic molecular science, animal models and clinical studies, have highlighted the relative importance of genetic and environmental contributions to the pathogenesis of IBD (**Figure 1-3** and **Table 1-1**).[72, 73] The currently accepted hypothesis is that specific defects in barrier function, dysregulated immune responses and/or ineffective bacterial clearance in genetically susceptible hosts, may lead to the establishment of T-cell mediated chronic mucosal inflammation. Factors known to disturb intestinal homeostasis that likely play a role in IBD pathogenesis include alterations in the barrier function of epithelial cells, innate immune cells (including macrophages and DCs), lymphocyte function and stromally-derived factors (e.g. TGFβ).[74] Professional antigen presenting cells (APCs), notably DCs, activated by either pathogenic bacteria or disturbed epithelium (via direct cell contact or indirectly through chemokines and other signalling molecules) translocate to the mesenteric lymph node (mLN) where they instruct naïve T cells to adopt a pro-inflammatory phenotype. Recent data demonstrates the importance of IL17 producing T cells (Th17) in intestinal inflammation.

The most important 'environmental' factor contributing to IBD pathogenesis is undoubtedly the intestinal flora. This is perhaps best illustrated in animal models of colitis. In the vast majority, gut inflammation does not develop when animals are raised in germ-free (GF) conditions (see below). It is also supported by the efficacy of antibiotics and probiotics in various clinical settings (e.g. pouchitis). Epidemiological data add further compelling evidence: IBD is more prevalent in urban populations, where increased industrialisation and sanitation, combine in a variation on the 'hygeine hypothesis'. In essence the theory goes that lack of exposure to environemental antigens during childhood leads to sup-optimal priming of gut immune responses. Whilst direct evidence for this is presently lacking, lower

rates of CD are reported for lower birth rank, absence of tap / hot water, large / poor families, and crowded living conditions.[2]

Other environemental stimuli implicated in IBD pathogenesis include cigarette smoking, appendicectomy, breast feeding, oral contraceptive pill use, sunlight / vitamin D exposure and various dietary factors. The evidence for some of these factors is rather limited and the literature conflicting. However, consistent themes are established for cigarette consumption. Smoking is detrimental in CD; not only is CD more common in smokers but also in children born to mothers who smoke during pregnancy and at birth.[75] After diagnosis, patients that continue to smoke do worse with, for example, increased rates of resectional surgery. Furthermore, smoking cessation is a very effective intervention in CD; although incompletely documented, the clinical benefit is probably equivalent to immunosuppression with a thiopurine or biologic. In contrast, smoking has a protective effect in UC as best illustrated by the significant proportion who first present within months to years of cigarette cessation. Anecdotaly, whilst these patients have a more aggressive disease course, many of them can subsequently control their disease by a 'prescribed' number of daily cigarettes. However, the failure of nicotine replacement therapy, both via transdermal and rectal routes,[76, 77] opens to question which alternative chemical component of cigarette smoke is responsible for these powerful effects.

The genetic contribution to IBD pathogenesis was conclusively demonstrated by twin studies, and further refined to multiple replicated loci at certain chromosomal locations by genome-wide linkage studies in sib-pairs and multiply-affected families.[78] Fine mapping of the IBD1 locus on chromosome 16 led to the discovery that *NOD2* variants are associated with CD.[79, 80] Our most recent understanding of the genetic architecture of these diseases strongly indicates that many disease genes confer susceptibility to CD (nucleotide-binding oligomerization domain containing 2 [*NOD2*], autophagy-related 16-like 1 [*ATG16L1*], immunity-related GTPase family, M [*IRGM*], leucine-rich repeat kinase 2 [*LRRK2*], CC chemokine receptor 6 [*CCR6*], inducible T cell co-stimulator ligand [*ICOSLG*], intelectin 1 [*ITLN1*]); or specifically to UC (extracellular matrix protein 1 [*ECM1*] and *IL10*); whilst many invoke risk to both (interleukin-23 receptor [*IL23R*], NK2 transcription factor related locus 3 (*Drosophila*) [*NKX2.3*], janus kinase 2 [*JAK2*], signal transducer and activator of transcription 3 [*STAT3*], *IL12B*, cyclin fold protein-1 [*CCNY*], IL18 receptor accessory protein [*IL18RAP*], *LYRM4*, and *CDKAL1*).[81-90]

1.4 *Importance of gene-environment interactions*

Defects in the intestinal epithelium, the mucosal immune system, and bacterial receptors may be genetic or environmental, with the complex interactions between these critical to the establishment and phenotype of disease (**Figure 1-3**). Gene-environmental interactions therefore appear critical to the initiation of the exaggerated and inappropriate inflammatory response that characterises these relapses and remitting inflammatory disorders. The mucosal interface where commensal and pathogenic bacterial antigens interact, directly and indirectly, with the epithelium is the front-line to a healthy well-regulated mucosal immune system. Defects in barrier genes and those that interact directly with bacterial antigen, may predispose to the development of IBD (**Table 1-1**). Dysregulation of innate immune signals are increasingly recognised to be important in pathogenesis. Epithelial-mesenchymal interactions via a raft of cytokines (e.g. TNFα, IFNγ and IL23) and chemokines (e.g. CCL20), initiate further downstream immune responses that in the healthy state bring about resolution of the inflammatory episode. However, in CD and UC a loss of immune tolerance leads to a breakdown in this homeostasis paving the way for chronic inflammation and the sequelae thereof.

Perhaps the most compelling recent evidence in support of gene-environment interactions comes from Sartor's work in animal models of colitis.[91] Both mice with targeted deleted of the *IL10* gene (*Il10*$^{-/-}$) and *Hla-B27* transgenic (TG) rats develop colitis in specific pathogen free (SPF) conditions, but not when kept completely germ-free (GF). However, it is now clear that selective colonisation with different bacterial species critically affects aspects of disease phenotype, notably severity and anatomical location. In *Il10*$^{-/-}$ mice monoassociated with *Enterococcus faecalis* or *Escherichia coli* a progressive chronic colitis develops, although the regional distribution and kinetics of this colitis varies with the bacteria. It is noteworthy that interferon gamma (IFNγ) secretion from meseneteric lymph node CD4+ T cells precedes the onset of microscopic inflammation. Mice monoassociated with *Pseudomonas fluorescens* or *Bacteroides vulgatus* do not develop colitis. In stark contrast, *Hla-B27* TG rats monoassociated with *B vulgatus* develop an aggressive colitis, but no inflammation with *E coli*. Clearly the interaction of both genetic factors (*Il10*$^{-/-}$ or *Hla-B27* TG) and environmental factors (*E coli* or *B vulgatus*) is fundamentally importance for the establishment of colitis in these animal models.

1.5 *Genetic basis of IBD*

There are considerable epidemiological data, which implicate genetic susceptibility in the pathogenesis of CD and UC. Most notably these include the familial prevalence of IBD, concordance rates in twin pairs, and ethnic differences in disease susceptibility. In fact, the studies of twin pairs have provided the strongest impetus towards further investigation of genetic susceptibility in IBD. Three studies have been carried out in Europe, including Tysk's important review of the Swedish Twin Registry in 1988.[92-94] The data from these studies, in Sweden, Denmark and the UK combine to provide powerful evidence for the role of both genetic and environmental factors in disease susceptibility. The concordance rates for CD in monozygotic and dizygotic twin pairs from these studies are estimated as, respectively 37% and 7%; in UC, the equivalent results are 10% and 3%.

1.5.1 Whole-genome studies: from linkage to association

The establishment of a linkage map of the human genome using informative microsatellite markers in the 1990s paved the way for hypothesis-free scanning for loci of association in monogenic and complex genetic disorders. Using this model 9 IBD susceptibility loci (designated IBD1-9, **Figure 1-4**) have been discovered and replicated to varying extent. Some of these loci appear to be relatively specific for CD (e.g. IBD1 on 16q - OMIM 266600)[95, 96] and UC (e.g. IBD2 on 12q - OMIM 601458)[97-99] whereas others are associated with IBD as a whole (e.g. IBD3 on 6p – OMIM 604519).[100-102] Very recently, microarray chip technology has become sophisticated enough to allow genome-wide association studies (GWAS) detailing variation up to ~500,000 SNPs across the human genome. This approach has in the past 12 months yielded up to 10 new CD susceptibility genes including *IL23R*,[83] *ATG16L1*,[87] *IRGM*,[81, 89] *NKX2.3*, *PTPN2* and a gene desert on chromosome 5.[88] Furthermore, meta-analysis of the GWAS studies from N America, U.K. and Belgium has identified a total of 30 CD genes / loci many of which did not meet genome-wide levels of significance in the individual studies ($p < 10^{-7}$).[103]

1.5.1.1 Nucleotide-binding oligomerization domain containing 2 *(NOD2)*

The first genome-wide scan, reported in 1996, provided the strongest evidence for linkage on the pericentromeric region of chromosome 16 (IBD1),[96] a region that has subsequently been widely replicated. Fine-mapping of the IBD1 region, and two positional candidate gene studies, identified *NOD2* (formerly *CARD15*) as a CD susceptibility gene in 2001.[79, 80, 104] Structural changes are induced in the muramyl dipeptide (MDP)-sensing NOD2 protein by two SNPs (Gly908Arg/rs2066845 and Arg702Trp/rs2066844) and one frameshift mutation

(Leu1007fsincC/rs2066847), located in the leucine-rich repeat (LRR) region of the gene. Carriage rates range widely (0-50%) with the highest rates seen in central Europe.[79, 105] The contribution of *NOD2* to disease susceptibility is relatively lower in Northern European (Scottish[106, 107] and Scandinavian[108, 109]) populations, where the population attributable risk (PAR) ranges from 7.9% (early-onset Scottish CD) to 11.4% (Swedish CD). In Japanese,[110] Chinese,[111] and S Korean[112] populations these *NOD2* mutations (Gly908Arg, Arg702Trp, and Leu1007fsincC) are absent (**Figure 1-5**).

Recent studies have further shown a number of associations between genotype and phenotype,[113, 114] thus adding to the complexities of IBD and reinforcing the point that the term 'IBD' may in fact represent several forms of diseases. Cuthbert and colleagues demonstrated the association between *NOD2* mutations and IBD are primarily based on the phenotype of ileal CD.[114] This association was also shown in the Oxford dataset by Ahmad and colleagues who reported detail genotype-phenotype analysis of 244 patients with CD.[113] The phenotype of fibro-stenosing CD was also associated with *NOD2* mutations, but this association was not shown to be independent of the association with ileal involvement. The *NOD2* gene does not appear to be associated with colonic involvement, and may indeed be protective against colonic involvement in CD and UC.

There has been much debate arising out of apparently conflicting experimental data as to whether the *NOD2* mutations represent a 'loss-of-function' or 'gain-of-function' phenotype. Two different animal models (*Nod2*$^{-/-}$ and *Nod2*^{2939iC}) published simultaneously in *Science* in 2005 added extra insight but failed to resolve this unsettling paradox.[115, 116] Neither mutant developed chronic intestinal inflammation spontaneously; indeed, neither model demonstrated any abnormality in intestinal microstructure. However, when stressed with dextran sodium sulphate (DSS) the *Nod2*^{2939iC} mouse behaved very differently to wild-type (WT) littermates, developing much more severe colitis with increased mortality (35% vs. 0%). However, the Karin group was unable to breed this Nod2 knock-in mouse (*Nod2*^{2939iC}) onto a pure Black-6 background.[116] These *in vivo* data from a mixed background therefore need to be interpreted with some caution. *Nod2*$^{-/-}$ mice responded less well than WT mice to challenge with oral *Listeria monocytogenes*,[115] an observation which may be related to decreased Paneth cell antimicrobial activity. NOD2 is constitutively expressed by macrophages, DCs and Paneth cells,[117] the latter being specialised epithelial cells located in the small intestine that secrete antimicrobial peptides, including alpha defensins (HD5 and

HD6) into the base of the crypt. Patients with ileal CD have decreased HD5 and HD6 expression, particularly those carrying *NOD2* mutations,[118] although this latter observation has not replicated in a subsuequent study from Australia.[119] Peripheral blood monocytes from CD patients with *NOD2* mutations have defective cytokine production and NFκB activity in response to MDP,[120] lending further support to the loss-of-function hypothesis.

1.5.1.2 *IBD5 locus (5q31)*

The first evidence for a susceptibility locus on the chromosome 5q31 cytokine gene cluster was generated by a Canadian genome-wide scan using ~300 microsatellite markers in 158 affected sibling pair families.[121] The linkage signal covered 18 cM including the *IL4*, *IL13*, *IL5* and interferon regulatory factor-1 (*IRF1*) genes. Further fine-mapping of this region with 56 additional markers demonstrated significant over-transmission of two adjacent markers using transmission disequilibrium testing (TDT).[122] However, due to very tight linkage disequilibrium (LD) across this region it was not possible to narrow the signal associated with CD to less than ~500 kb. In the pre-HapMap era (http://www.hapmap.org), re-sequencing of this region (in 8 individuals) was required to define haplotype blocks; this approach revealed a region of very limited recombination.[122, 123] Association studies performed with haplotype-tagging SNPs (tSNPs) in at least 6 large, independent CD populations has found that a single common haplotype spanning ~250 kb was associated with CD.[124-127]

The IBD5 risk haplotype has proved to be the most widely and robustly replicated inherited risk for IBD after *NOD2*.[128-133] A report in 2004 suggested that two variants within the *OCTN1* (*SLC22A4*-1672C/T) and *OCTN2* (*SLC22A5-207G/C*) genes, proposed to alter gene function and expression, were associated with CD independent of the IBD5 risk haplotype.[134] Whilst this may yet prove to be a valid finding, a series of replication studies have discredited the suggestion from the index study that these variants are independent of the IBD5 risk haplotype.[128, 130-133, 135] The tight LD across this region remains the limiting factor in further delineating the specific CD risk in IBD5. It has been calculated that cohorts with several thousand CD patients and controls will be required to demonstrate independence of the causative IBD5 variant(s) from the background haplotype. There are 5 genes within this region: *IRF1*, *OCTN1*, *OCTN2*, *PDLIM4* and *P4HA2* (for a detailed review of the expression and function of each of these genes see Reinhard and Rioux, 2006).[130] Detailed expression and functional studies of these genes in CD compared with healthy controls may informative

in further determining the causative IBD5 gene. Finally, it is noteworthy that 5q31 is also associated with psoriasis, although the variants typed to date appear to be distinct to those associated with CD.[136-138]

1.5.1.3 *IBD2 (12q13)*

A susceptibility locus on chromosome 12q13 was discovered in a two-stage U.K. genome-wide linkage scan in 186 affected sibling pairs from 160 nuclear families.[99] The initial screening of 260 autosomal microsatellite markers in 89 affected siblings suggested evidence for linkage with IBD across four adjacent markers on chromosome 12 spanning 16 cM (**Figure 1-6**). This was confirmed in a second cohort of 97 additional affected siblings. The best scoring marker on chromosome 12 was D12S83 ($p = 2.66 \times 10^{-7}$, lod score 5.47). Subsequent re-analysis of the IBD2 region (with 12 microsatellite markers) in a larger cohort (252 CD-only, 138 UC-only and 191 mixed families) demonstrated that the locus makes a major contribution to UC susceptibility (lod score 3.91) but has only a mild influence on CD (lod score 1.66).[98] A large linkage study, stratified for phenotype, has demonstrated strong evidence for linkage with extensive UC ($p < 0.001$, lod score 3.27).[97] Candidate gene studies in IBD2 have largely been negative (e.g. *STAT6*[139, 140] and *AVIL*[141]). However, a very recent UC GWAS has provided evidence for association at the edge of the IBD2 linkage region at 12q15, in the region of *IL26* and *IFNγ* (see **1.5.4.6** below).[142]

1.5.1.4 *HBD2 (DEF4B)* and Copy Number Variation

It is becoming increasingly clear that copy number variation (CNV) of DNA sequences has important functional consequences and likely plays a large role in both microdeletion and microduplication disorders as well as susceptibility to complex diseases. CNVs can influence gene expression directly and indirectly through position effects, alter gene dosage, and predispose to deleterious genetic changes. Recently, CNV across the genome has been described in the four HapMap populations by both SNP genotyping array and clone-based comparative genomic hybridization.[143] These data provide a map of human genomic CNV that will provide a great resource to future genetic studies. Indeed, it is suggested by these authors that assessment of CNV should become a standard component of studies into the genetic basis of disease susceptibility.

In the first study of its kind in IBD, Fellermann and colleagues performed an array-based comparative genomic hybridization analysis using ~8000 genomic fragments to give genome-wide coverage with an average resolution of ~0.5Mb.[144] Comparing 10 patients

with colonic CD and 10 HCs, they demonstrated that the β-defensin cluster at 8p23.1 showed altered copy number between CD and HC. This region, spanning ~750kb, did not extend into the adjacent α-defensin cluster. 8/10 CD patients showed small copy number loss compared with HC, whereas none showed copy number gain. On quantitative PCR of the *HBD2* (*DEF4B*) gene in 169 HC there was a median of 4 copies (range 2-10). A U.S. cohort of 85 CD patients who had either ileal or colonic surgical resections were analysed. Patients with ileal CD had *HBD2* copy number identical to HC. However, in the colonic CD group the median copy number was 3, with 72% having <4 copies of the gene (ileal vs. colonic copy number p = 0.008). These findings were confirmed in a separate cohort of CD from Europe. Whilst the difference in copy number appears small, HBD2 mRNA analysis from patients with active colonic CD demonstrated decreased mucosal HBD2 expression in patients with copy number <4 compared with those with ≥4 copies of the gene (IL8 levels were not significantly different in these groups).

1.5.1.5 NLR family, pyrin domain containing 3 *(NLRP3)*

The *NLRP3* gene (1q44) encodes cryopyrin, a protein that controls the inflammasome and thus regulates activation of caspase-1 and IL1β. *NLRP3*, like *NOD2*, is a CATERPILLER gene, in that it has both NOD and LRR domains. Gain-of-function mutations in the NOD domain of *NLRP3* cause Muckle-Wells, familial cold autoinflammatory syndrome and neonatal-onset multisystem inflammatory disease, three hereditary periodic fever syndromes.[145, 146] Villani and colleagues recently took a candidate gene approach to study the contribution of germline variation in *NLRP3* to CD.[147]

3 tSNPs were associated with CD in the first phase, where 47 tSNPs were genotyped in 296 trios from Leuven [Belgium]. In phase two, these 3 tSNPs were genotyped in 3 further family-based cohorts (Liege [Belgium], Quebec and Toronto). All 3 tSNPs, spanning a 5.3kb region replicated, with the strongest association at rs10733113 ($p_{combined} = 3.49 \times 10^{-9}$, OR 1.78, CI 1.47 - 2.16) reaching genome-wide association levels of significance (see below). None of these SNPs was in LD with any genotyped SNPs within *NLRP3*. Deep resequencing of a 9kb region (from the *NLRP3* 3′ UTR to the associated 5.3kb region) in 16 CD and 8 HC (selected on the basis of genotype) yielded 79 SNPs, 60 of which had a MAF ≥ 0.05 and 14 of which were novel to dbSNP 129. Following a further detailed tagging study of this region in the Leuven cohort and subsequent replication, 3 additional SNPs were consistently associated with CD ($p = 10^{-6} – 10^{-7}$).

Villani and colleagues went on to analyse the effect of the 6 CD associated mutations on *NLRP3* expression in peripheral blood cells and monocytes.[147] rs4353135 genotype was significantly associated with altered expression, with homozygosity at the risk allele conferring low expression levels. In a separate assay in cultured monocytes, homozygosity of the risk allele at a different SNP (rs6672995) was associated with decreased levels of IL1β following stimulation with LPS ($p = 0.0059$). Finally, the authors demonstrate significant elevation of NLRP3 expression in human CD biopsies (fold change = 4.08, $p < 0.0028$), and acute TNBS –induced colitis (fold change = 9.38, $p < 0.0009$).

1.5.1.6 X-box binding protein 1 *(XBP1)*

In a recent report in Cell, Kaser and colleagues have provided *in vivo*, cellular and genetic evidence for a role of the endoplasmic reticulum (ER) stress reponse in IECs.[148] The cellular response that allows cells to survive ER stress involves a highly coordinated unfolded protein response (UPR), and the transcription factor X-box-binding protein 1 (XBP1). XBP1, activated by excision of 26 bp from its mRNA, induces transcription in a subset of UPR target genes.[149] ER stress and XBP1 mRNA expression is increased in the inflamed and non-inflamed terminal ileum of patients with CD and inflamed and non-inflamed colons of patients with CD and UC. Mice with deletion of *Xbp1* specifically in IECs ($Xbp1^{flox/flox}VCre$) demonstrate increased ER stress in small intestinal epithelial cells and subsequent mild, patchy, spontaneous mucosal inflammation. These animals were almost completely devoid of Paneth cells, in addition to a minor defect in small intestinal (but not colonic) goblet cells. Oral challenge with *Listeria monocytogenes*, was impaired with significantly increased numbers of bacteria retrieved in faeces and the liver, but not spleen compared with WT animals. Although the basal phenotype in the colon of these mice demonstrated increased ER stress, there was no evidence of spontaneous inflammation. However, stressing these animals with 4.5% DSS for 5 days led to a marked increase in the severity of colitis compared with WT.

The *XBP1* gene is an excellent biological candidate for IBD susceptibility based on expression, function (as above) and position. It is located on 22q12 in a previously identified linkage region.[150-152] The Schreiber group performed a tSNP study of *XBP1* variation in German IBD with subsequent replication in two independent panels.[148] In the index cohort, 3

of 20 tSNPs were significant, all of which replicated. In the combined cohort of 4389 IBD patients and 5322 HC, and after correction for multiple testing, 6 *XBP1* SNPs were significantly associated with IBD after 10,000 permutations. rs35873774 was the strongest SNP with OR 0.74 (C.I. 0.66-0.84, $p = 1.6 \times 10^{-5}$). The complex haplotype structure of *XBP1*, and deep resequencing in 564 IBD patients and 282 HC, strongly indicated that multiple, rare, private SNPs contribute to the association with IBD. Indeed, several of these rare nsSNPs demonstrated decreased transactivating function in a series of *in vitro* assays.

1.5.2 Genome-wide association studies (GWAS)

Up until about 2 years ago, researchers were becoming increasing frustrated by the limitations of genome-wide linkage studies, fine-mapping of these large regions and candidate gene approaches to describing the inherited risk of complex genetic diseases. Despite a few notable exceptions, including *NOD2* in CD, progress had become slow and the field muddied by the inability of new disease 'susceptibility genes' to consistently replicate in different populations (e.g. *DLG5* in CD). It was clear that a radical new approach was required. However, before genome-wide association studies (GWAS) could be implemented, a number of scientific and technological hurdles had to be passed.[153] The first step followed quickly on from the premature announcement by the popular media of the sequencing of the human genome in 2000, with the discovery of over 10×10^6 SNPs. Second was the generation of the hapmap (www.hapmap.org), a high-resolution map of genome variation that utilised measures of association between SNPs (LD) to define haplotypes. This provided a means by which ~70% of variation across the genome (in populations of European ancestry) could be described by genotyping 300-500k SNPs. Third was the development of high-throughput genotyping platforms with capability of sequencing up to 1×10^6 SNPs simulataneously on a single chip, all at a reasonable cost.[153] The two commonly used platforms to date are the Illumina 317k (used in the N American CD GWAS)[83] and Affymetrix 500k (WTCCC study)[81] chips. Other studies have taken a more limited (but more cost effective) initial approach by typing just nsSNPs (~14k). Whilst much less extensive in their coverage of the genome, these studies have the advantage that any 'hits' have almost immediate functional potential (although we now recognise that a large number of causative SNPs function without altering the amino acid sequence). This approach has been successfully employed by the Schreiber group in CD and the UK IBD genetics consortium in UC.[84, 87]

When considering GWAS design, it is widely agreed that a large sample size of cases and controls is critical to enable sufficient power to detect association of variants with modest increase in disease risk at a genome-wide significance threshold (now widely established at $p < 1 \times 10^{-7}$). The GWAS approach is particularly effective at detecting relatively common variants, as the ability to detect association at a particular SNP decreases with the frequency of its minor allele (↓ *frequency* = ↓ *power*). Several studies have adopted different techniques to improve study power by attempting to enrich for specific disease-predisposing alleles. Two common approaches have been to minimise population heterogeneity (successfully adopted in the N American CD GWAS [ileal diseae only])[83] and to focus on extreme / familial cases (e.g. early-onset IBD; N American UC GWAS [excluding proctitis]).[142, 154] The true value of such approaches has yet to be determined and is cast partially into doubt by the high success of CD in the WTCCC study where this was not adopted.[81, 155]

Selection of appropriate HCs remains controversial. The WTCCC study demonstrated the success of a common pool of controls for comparison with 7 different complex diseases. The WTCCC HCs consisted of both a population-based birth cohort (1958) and opportunistic blood donors. Comparison of these two sources was reassuring in the minimal impact that ascertainment, selection and survival biases had on genotype distributions. Another issue with control populations that will decrease study power is that in most studies they are not extensively screened for latent disease. For most relatively rare complex diseases this is not such a big issue as it can be circumvented by sufficiently increased sample size. For common traits, selection of extreme phenotypes (e.g. very obese or very tall) can also minimise this error.

Once the cohort has been collected and genotyped on the chosen platform, a series of important quality control steps are implemented prior to a case-control association study. The most critical next step is to distinguish between artefact and a true positive result through validation of positive hits by replication in suitably powered independent cohorts of cases and controls.[156] The hard challenge following this is fine-mapping (a particular challenge in areas of tight LD) to determine causative mutations followed by functional studies to understand the true biology behind the increased disease risk (or protection). McCarthy and colleagues have recently published a very thorough review of the technical and statistical aspects of all stages of GWAS design, implementation and follow-up.[155]

In the past 2 years, more than 300 replicated associations have been reported in over 70 common diseases and traits.[157] Along with CD, particular success has been obtained in type II diabetes mellitus (~20 loci)[158] and height (over 40 loci),[159, 160] a number of common cancers (including colo-rectal,[161-163] and prostate[164-169]) and other continuous traits (e.g. obesity).[170-173] Furthermore, it is of great interest that many disease associated genes have demonstrated overlap between dinstinct disease types.[157] The best characterised overlap to date is with coeliac disease and type I diabetes mellitus.[174] It was previously known that both diseases are associated with class II HLA genes. In their comprehensive study, Smyth and colleagues discovered shared association of 6 additional loci with both diseases (*RSG1* at 1q31, *IL18RAP* at 2q12, and *TAGAP* at 6q25, a 32-bp insertion-deletion variant at *3p21*, *PTPN2* at 18p11, *CTLA4* at 2q33).[174] Intriguingly, not all the effects were in the same direction. Variants in *IL18RAP* and *TAGAP* had protective effects in type I diabetes whilst conferring susceptibility to coeliac disease. As discussed further below, both *IL18RAP* and *PTPN2* are also confirmed CD suscpetiblity loci.[103] Other examples of disease overlap include a region at 8q24 associated with both prostate and colo-rectal cancer, and the *TCF2* and *JAZF1* loci conferring risk (or protection) to both type II diabetes and prostate cancer.[157]

Despite this undisputed success, the known variants to date only explain a fraction of the observed familial aggregation.[175] For example, Barrett estimates that the 30 known CD loci account for only ~20% of the genetic contribution to CD.[103] There are several explanations for this. Firstly, the present GWAS provide only relatively limited surveys of potential sequence variation. Secondly, many 'hits' from the index studies are surrogate markers for the true disease causative mutations that will in most cases confer greater risk. Thirdly, the risk at some loci will be attributed to multiple independent mutations. Some GWAS are limited to specific phenotypes (e.g. ileal only CD, or adult-onset only IBD). In addition, very few studies into CNVs have been performed on a genome-wide scale; time will tell whether these will add siginificantly to the genetic risk (see **1.5.1.4**). Finally, it is unclear yet how the newly discovered disease mutations will interact with important environmental factors. For several or all of these reasons, we are probably currently underestimating the known genetic risk already accounted for. However, the shortfall that remains is (and almost certainly should) presently limit the early application of genetics to determinging individual disease risk.

The immediate future involves more GWAS in different disease types, sub-phenotypes and common traits (the Wellcome Trust is spending over £30 million typing DNA from 120,000 individuals over 27 phenotypes)[157], meta-analysis of existing individual GWAS (as has been particularly successful in CD[103], colo-rectal cancer[176] and type II diabetes[158]), fine-mapping of loci (a major on-going phase of the WTCCC follow-up studies, with CD prioritised for deep resequencing of associated loci), and detailed molecular and functional studies to truly understand the biology of various disease associated mutations. In the meantime, the major evolving paradigm for complex disease genetics, is that the biological insights gained (and subsequent abundance of novel therapeutic targets) will turn out to have far greater importance than the relatively small contribution to disease risk conferred by each individual mutation or locus, as will now be illustrated for first CD and subsequently UC.

1.5.2.1 Interleukin 23 receptor *(IL23R)*

Duerr and colleagues performed the first full GWAS in IBD, analysing over 300,000 SNPs on the Illumina HumanHap300 Genotyping BeadChip.[83] To minimize pathogenic heterogeneity their study population was limited to ileal CD, analysing 567 cases and 571 controls, all of non-Jewish, European ancestry. It was noteworthy that after applying Bonferroni correction, only 3 SNPs remained significant at the 0.05 level (and nearly two orders of magnitude more significant than the next best scoring marker). Two of these SNPs were in *NOD2* (rs2066843 and rs2076756) effectively providing proof-of-principle for this technique in discovering CD susceptibility genes. The third was a non-synonymous SNP in the *IL23R* gene (located on chromosome 1p31), rs11209026 Arg381Gln ($p = 5.05 \times 10^{-9}$, corrected $p = 1.56 \times 10^{-3}$). Several other SNPs within *IL23R* were subsequently shown to be associated with ileal CD, including some that were independent of the rs11209026 variant. It is proposed that the functional significance of these multiple variants within *IL23R* may in part be due to differential splicing, as at least six alternatively spliced mRNAs of *IL23R* are described.[83] Duerr and colleagues demonstrated replication of the *IL23R* association in an independent case-control association study of non-Jewish and Jewish ileal CD patients, and in a family-based association study in 833 nuclear families (data on both parents plus affected offspring – CD and UC – were available). Further replication studies for *IL23R* have been published in adult and paediatric populations from the U.K.,[177-179] France / Belgium,[88] Quebec founder population,[180] Italy,[181, 182] Canada,[183] Holland,[184, 185] Spain,[186] Finland,[187] Hungary,[188] and Brazil.[189] There was no consistent genotype-phenotype

substratification apparent in these studies.[178, 190, 191] *IL23R* has also recently been confirmed as a UC susceptibility gene.[84, 85, 142]

In addition, the CD meta-analysis (see **1.5.3**) has confirmed 3 addition members of the IL23 signalling pathway as CD susceptibility genes.[103] *IL12B* (5q33; OR 1.11, $p = 3.86 \times 10^{-13}$), *JAK2* (9p24; OR 1.12, $p = 3.46 \times 10^{-9}$) and *STAT3* (17q21; OR 1.18, $p = 6.82 \times 10^{-12}$) were all shown to be associated with CD. Furthermore, they are also UC susceptibility genes,[84, 85] confirming that defects in IL23 signalling confer risk to IBD as a whole, without any specific phenotypic associations.

Furthermore, *IL23R* variants have been associated with ankylosing spondylosis,[192, 193] and both *IL23R* and *IL12B* with psoriasis,[194-197] but not with rheumatoid arthritis,[198, 199] or systemic sclerosis.[200]

The timing of the initial *IL23R* gene discovery study in CD was interesting, as it coincided with detailed murine studies of bacterial induced intestinal inflammation demonstrating a critical role for IL23 in innate immune pathology.[201, 202] The proinflammatory cytokine IL12 has long been thought to be a critical element in the development of pathogenic Th1 CD4+ effector cells. The discovery that IL23 is a heterodimer of p19 and the IL12p40 (IL12B) subunit has led to a critical reappraisal of the relative roles that IL12 (a heterodimer of p40 and p35 subunits) and IL23 play in inflammatory disorders.[203] IL23 is secreted by activated DCs, monocytes and macrophages. Transgenic mice that constitutively overexpress IL23p19 develop a fatal multiorgan inflammation.[204] IL23 supports the development of the Th17 subset of CD4+ inflammatory T cells (that produce IL17, IL6 and T and TNFα)[205-207] and has potent effects on innate immune cells (monocytes and macrophages) by inducing production of IL1, IL6 and TNFα.[207, 208] IL23 and IL17 synthesis and secretion is increased in the inflamed murine intestine (*Helicobacter hepaticus* infected $Rag^{-/-}$ mice that develop a chronic typhlocolitis characterised by the accumulation of innate immune cells), whereas IL12 expression is unchanged.[201] IL23 blockade with an anti-p19 monoclonal antibody significantly decreased *H hepaticus* induced intestinal inflammation as well as decreasing proinflammatory cytokine (TNFα, IFNγ, IL6, IL1β and IL17) in the intestine. Equivalent findings have been demonstrated with anti-IL10R antibody treatment of *H hepaticus* infected T-cell sufficient hosts and CD4+ T cell transfer into various $Rag^{-/-}$ recipients.[202] Combined these data demonstrate the crucial role for IL23, but not IL12 in intestinal inflammation, by

sustained activation of adaptive and / or innate immune mechanisms. These data make critical re-evaluation of early clinical studies using anti-p40 monoclonal antibody therapy in CD of paramount importance.[209] Anti-p19 monoclonal antibodies, that are likely to provide a more specific anti-inflammatory intervention in IBD, are presesntly in development.

1.5.2.2 Autophagy genes – *ATG16L1* and *IRGM*

Perhaps the most exciting and novel insight to arise from the GWAS reports in CD is the discovery that defective autophagy may play a critical role in disease pathogenesis. Two autophagy genes are now known to be associated with CD: *ATG16L1* and *IRGM*.

Autophagy, the process by which cells digest parts of their own cytoplasm for removal, functions as a homeostatic "house-keeping" mechanism, a pathogenic process in a variety of cancers and neurodegenerative disorders, and as an innate immune mechanism against intracellular bacteria.[210] It is a degradative, membrane-based system by which long-lived intracellular products are removed and turned-over via the lysosome. As such, it complements the ubiquitin–proteasome pathway that degrades short-lived proteins. Autophagy is highly conserved; in a unicellular world, it probably evolved as an essential survival mechanism to provide an alternative energy source during periods of starvation, thus ensuring continued minimal cellular function. Most of the upstream regulation of autophagy occurs directly through the mammalian target of rapamycin (mTOR). The autophagy-related (ATG) proteins control the execution of autophagy, as outlined in **Figure 1-7**.

In the same way that damaged organelles are homeostatically removed by autophagy, intracellular bacteria are engulfed by autophagosomes and transferred to lysosomes for degradation; this process is termed xenophagy. This has been demonstrated *in vitro* for several pathogens, including group A *Streptococcus*, *Mycobacterium tuberculosis*, *Shigella flexneri*, and *Salmonella enterica*, as well as viruses such as herpes simplex virus type-1, although direct *in vivo* evidence is presently lacking.

ATG16L1 was originally identified as a CD gene in a German nsSNP GWAS, and has subsequently been confirmed in UK, North American, and French/Belgium GWAS as well as a handful of independent replication cohorts.[81, 87-90] Hampe and colleagues analyzed approximately 20,000 coding SNPs (SNPlex Genotyping system) in a German cohort of 735 CD patients and 368 HCs.[87] Of the 16,360 successful assays, 7159 had a MAF of >1%. On

allele-based testing for association, 72 of these 7159 were significant at $p<0.01$. These SNPs were then tested in a second panel consisting of 380 CD trios, 498 CD singletons, and 1032 HC. Three of 72 SNPs satisfied a pre-defined statistical cut-off of $p<0.05$: rs2066845 in *NOD2*, rs1050152 (L503F) in *SLC22A4*, and rs2241880 (Thr300Ala) in *ATG16L1* at 2q37 near a locus of susceptibility ($p_{case-control} = 1.6 \times 10^{-5}$; $p_{TDT} = 2.7 \times 10^{-5}$). After re-sequencing of all exons, splice sites, and the promoter region of *ATG16L1* in 47 CD samples failed to detect any further coding or splice site variants, 28 tSNPs ($r^2 > 0.8$; MAF >1%) were selected from the CEU HapMap data. The results of logistic regression and haplotype analysis suggested that the CD risk from *ATG16L1* was confined to the G allele of Thr300Ala. Analysis of the combined cohorts gave an OR of 1.45 (C.I. 1.21–1.74) and a PAR of 0.26 for carriage of this variant, with some evidence of a gene dosage effect (homozygote OR 1.77, CI 1.43–2.18; PAR 0.17). It is noteworthy that the variant confers an amino acid change at the evolutionary conserved position 300 of the *N* terminus, located in exon 9, which is translated into the same reading frame of all six known splice variants of *ATG16L1*.

Subsequent to this index publication, the full results of the N American GWAS were published.[90] In this study, the authors reported association of Thr300Ala with CD at $p = 6.4 \times 10^{-8}$, with additional replication in two further cohorts (530 CD trios; 353 patients with ileal CD and 207 HC) at $p = 4.1 \times 10^{-8}$. Furthermore, expression of ATG16L1 (along with ATG5 and ATG7) was demonstrated in a variety of epithelial and immune cell lines. ATG16L1 was additionally expressed in a number of primary human immune cells, notably $CD4^+$ and $CD8^+$ T cells and $CD19^+$ B cells. Finally, the requirement of ATG16L1 for mammalian autophagy was demonstrated by small interfering RNA (siRNA) knock-down of *ATG16L1* in an immortalized cell line (HeLa cells) subsequently challenged with *Salmonella typhimurium*. In this system there was decreased targeting of intracellular bacteria to autophagic $LC3^+$ vacuoles within 1 hour (2% compared with 17.5% in WT cells).

Hampe and colleagues demonstrated expression of *ATG16L1* mRNA and protein in colon, small intestine, intestinal epithelial cells, and leukocytes. However, they did not detect any difference in protein expression in intestinal tissue between CD and HC, although the numbers for analysis were small and the methods descriptive in nature.[87] Furthermore, protein and cDNA expression levels were independent of Thr300Ala genotype. We have demonstrated downregulation of *ATG16L1* mRNA in colonic CD biopsies compared with

healthy controls in our large microarray dataset (Noble, Lees *et al*, manuscript in preparation).

Three different murine models of ATG16L1 deficiency have recently been described in two publications in *Nature*.[211, 212] Cadwell and colleagues generated two mouse lines with gene-trap mutations (which introduce a false splice acceptor into introns to inhibit expression of intact mRNA) in *Atg16l1* (*Atg6l1*^{HM1} and *Atg16l1*^{HM2}).[211] *Atg16l1*^{HM} mice had defective autophagy. *In vitro* experiments with murine embryonic fibroblasts demonstrated diminished rapamycin-induced and autophagy-dependent degradation of the adaptor protein sequestonome (SQSTM1 or p62), reduced production of LC3-II, and decreased autophagosome induction. *In vivo*, *Atg16l1*^{HM} mice expressed Atg16l1 at 23–37% of the expected level in the intestine. Ileal and colonic morphology was normal. In addition to providing evidence of defective intestinal autophagy (increased LC3-II and p62 levels), the investigators demonstrated that these mice have abnormalities in Paneth cells (decreased lysozyme in mucus, aberrant disorganized granules, and reduced granule number) indicating that Atg16l1 is required for maintenance of the Paneth cell granule exocytosis pathway. Given this key observation, it is noteworthy that in contrast to *Nod2*$^{-/-}$ mice, *Atg16l1*^{HM} animals had no alterations in handling oral *Listeria monocytogenes* challenge (bacterial titers in the spleen, liver, and mesenteric lymph nodes were the same as those in wild-type mice). This suggests that the phenotypes of *ATG16L1* and *NOD2* mutations appear to be distinct. The Paneth cell defects seen in *Atg16l1*^{HM} mice on transmission electron microscopy (including degenerating mitochondria, loss of granules, and frequent absence of apical microvilli) were not present in epithelial progenitors or enterocytes, indicating that Paneth cells have a unique sensitivity to autophagy mutation.

Cadwell and colleagues proceeded to examine RNA profiles in laser microdissected Paneth cells from mutant and wild-type mice.[211] Approximately 1.5% of the probe set was altered by greater or less than 1.3-fold in *Atg16l1*^{HM} Paneth cells including those related to peroxisome proliferator-activated receptor (PPAR) signaling, adipocytokine signaling, and lipid metabolism, as well as the acute phase response (serum amyloid A, haptoglobin, and complement factors D and I). The same analysis was performed in T cells; remarkably, only one gene had similarly altered expression profiles in T cells and Paneth cells. Therefore, the transcriptional signature of *ATG16L1* deficiency is specific to Paneth cells.

Saitoh and colleagues generated mice that lack the entire coiled-coil domain of *Atg16l1*.[212] These *Atg16l1*$^{-/-}$ mice die on the first postnatal day. Interestingly, the authors speculate that this is because intact autophagy may be essential for surviving through neonatal starvation. To circumvent this lethal mutation, a chimeric mouse was generated by transplanting *Atg16l1*$^{-/-}$ fetal liver cells into lethally irradiated CD45.1^{+} mice. Under SPF conditions these chimeric mice did not develop spontaneous colitis. However, when stressed with DSS, the mortality rate was 100% at day 7 compared with 0% in WT mice. The mutant animals had severe inflammation in the distal colon with large areas of ulceration and increased lymphocytic infiltration.

The investigators then stimulated peritoneal and bone-marrow derived macrophages from chimeric and WT animals with LPS, MDP, and different bacterial strains.[212] A robust increase in IL1β levels was demonstrated upon stimulation with LPS and non-invasive Gram-negative bacteria (such as *Escherichia coli*) but not with invasive strains (such as *Salmonella typhimurium*). Intriguingly, a normal inflammatory cytokine response was noted following MDP stimulation, providing further evidence that ATG16L1 is not involved in NOD2 signaling. In addition, stimulation with TLR3, TLR4, TLR7, and TLR9 ligands (but not TLR2 or TLR5 ligands) also induced IL1β production. In *Atg16l1*$^{-/-}$ macrophages, autophagosomes were hardly detected and p62 was markedly increased under nutrient-rich conditions. Together, these data indicate that reductions in basal autophagy induce IL-1β overproduction. The relevance of this was demonstrated *in vivo* as injection of neutralizing anti-IL1β antibodies significantly improved mortality rate and weight loss in the DSS-stressed chimeric animals.

Translating these mechanistic studies directly back to man, Cadwell and colleagues tested the effect of the CD-associated *ATG16L1* mutation on the histopathology of ileal resection specimens from patients who were either homozygous mutant or WT at Thr300Ala (none of these patients had *NOD2* or *IL23R* mutations).[211] Analysis of healthy tissue at resection margins in patients with the *ATG16L1* mutation demonstrated that Paneth cells had disorganized granules with reduced granule formation, diffuse cytoplasmic staining of lysozyme, and positive staining for leptin. This was strikingly concordant with the phenotype of *Atg16l1*^{HM} mice.

1.5.2.3 WTCCC study and immunity-related GTPase family, M (*IRGM)*

In a landmark study published in the *Nature* journal in June 2007, the WTCCC performed a GWAS on 500,000 markers in 2000 patients with seven different diseases (CD, rheumatoid arthritis, type I and II diabetes mellitus, coronary artery disease, bipolar disorder and hypertension) and 3000 shared HC.[81] This study was particularly successful in CD, with nine genes or loci reaching genome-wide significance at $p < 10^{-7}$. These included established CD genes (including *NOD2*, *IL23R*, and *ATG16L1*) and four novel associations (*IRGM*, *NKX2.3*, *PTPN2*, and *3p21*). All were subsequently replicated in an independent UK population (1182 patients with CD and 2024 independent HC).[89]

Two SNPs (rs13361189 and rs4958847) flanking *IRGM* on chromosome 5q33.1 were strongly associated with CD on the index WTCCC scan and subsequent replication cohort ($p_{combined} = 2.1 \times 10^{-10}$ and 3.8×10^{-9}). Neither of these SNPs is known to be functional. Re-sequencing of the major *IRGM* exon identified two new SNPs, one that confers an amino acid change, and another that does not. Only the silent of these two SNPs was associated with CD, and was noted to be in near perfect LD with rs11361189. This strongly suggests that the disease-associated variants in *IRGM* do not cause simple amino acid substitutions, but more likely function via regulation of gene expression, gene splicing, or the rate of protein translation.

Most recently, a 20kb deletion immediately upstream of *IRGM*, and in perfect LD with rs13361189, has been described by the laboratory of Ramnik Xavier.[213] This insertion/deletion polymorphism was typed in a N American case–control population consisting of 685 individuals. The frequency of the deletion in the reference population was 10%, compared with 15% in CD patients (OR 1.6; $p < 0.01$) and 14% in UC (OR 1.4; $p < 0.05$). It is intriguing that the deletion was associated with UC in this study, given the absence of association noted for *IRGM* SNPs in previous studies.[84, 85] Replication of this observation in independent UC cohorts is required to validate or refute this point. McCarroll and colleagues went on to demonstrate *in vitro* that the CD risk (deletion) and protective (reference) haplotypes differentially activate IRGM expression in distinctive cellular contexts.[213]

The immunity-related guanosine triphosphatase (IRG or p47 GTPase) genes are critical to innate immunity against intracellular pathogens[214]. Whilst there are 23 complete Irg genes in

mice, only three have been identified in man[214, 215]. Of these, *IRGM* (OMIM 608212) is the human homologue of murine Irgm1 (Lrg-47), a key mediator of IFNγ-induced autophagy[216]. Human IRGM is constitutively expressed and contains no interferon-inducible elements in the promoter, leading to initial reports that humans lacked a p47 resistance system[215]. However, despite being unresponsive to IFNγ, IRGM plays a critical role in IFNγ-induced or conventionally induced autophagy in human macrophages.[216] siRNA knockdown of IRGM in human macrophages infected with mycobacterium leads to defective autophagy, with increased bacterial survival compared with control cells.[216] Lrg-47 knockout mice (*Lrg-47*$^{-/-}$) fail to control mycobacterium tuberculosis infection (succumbing after several weeks following aerosol or intravenous inoculation), despite the formation of well-organized granulomas replete with infiltrating lympocytes[217]. *Lrg-47*$^{-/-}$ macrophages have defective bacterial killing with impaired maturation of mycobacterium tuberculosis containing phagosomes; Lrg-47 is recruited to these vesicles in WT cells.[216] Similarly, *Lrg-47*$^{-/-}$ mice infected with mycobacterium avium are unable to control bacterial replication; although surviving the acute illness they succumb 11-16 weeks later.[218] The phenotype of these animals suggests a regulatory role for Lrg-47 on lymphocyte survival, as they have a profound systemic anaemia and lymphopenia, with marked lymphocyte deficiency in mycobacterium granulomas.[218]

1.5.2.4 Autophagy, innate immunity and intestinal luminal bacteria

It is highly likely that autophagy and innate immunity, both ancient and highly conserved processes, co-evolved as primitive defenses against pathogens. Direct links have now been established between TLR signaling, autophagy, and phagocytosis[219] and, although the relative contributions of each in different tissues remains to be established, both are likely to be important for clearance of intracellular pathogens in the gastrointestinal tract. It has been demonstrated that microtubule-associated protein 1 light chain 3 (LC3) and beclin 1 are recruited to phagosomes under the control of ATG5 and ATG7, but without the formation of double-membrane structure around the TLR-induced phagosomes (in contrast to conventional autophagosomes). This process results in more rapid and extensive acidification of the phagosomes and enhanced pathogenic killing.[219]

It is intriguing to speculate that there might be a link between NOD2 signalling and autophagy in GI tract homeostasis and in CD pathogenesis. This is being actively investigated in several laboratories internationally. It has already been demonstrated that

NOD-like receptor (NLR) signalling interacts with autophagy. The activation of caspase 1 and subsequent cell death in macrophages following *Shigella* infection is mediated via the NLR-family member IL1-converting enzyme-protease activating factor (Ipaf) and the adaptor protein apoptosis-associated speck-like protein containing a C-terminal caspase recruitment domain (ASC). *Shigella*-infected macrophages undergo increased autophagy in the absence of caspase 1 and Ipaf, but not ASC.[220] This demonstrates that autophagy in this context is regulated by the Ipaf inflammasome.

In recent years, much attention has focused on the role of enteroadherent strains of *E coli* and *M avium* subsp. *paratuberculosis* (MAP). Whilst different international groups have isolated very similar strains of *E coli* from patients with CD, the role of MAP remains highly contentious. This is largely due to the wide-ranging recovery rates reported for MAP culture and DNA in cases and controls (0–100%), and the lack of a functional role for bacteria recovered from ulcerated mucosa.[221] However, given the defective mycobacterial killing demonstrated with *IRGM* knock-down, the MAP "story" may see a resurgence. It will be of great interest to establish whether this effect is seen in those with the CD-associated *IRGM* insertion-deletion polymorphism and, furthermore, whether MAP is isolated more frequently from the intestines of these patients.

Determining whether defects in autophagy are related to specific bacterial strains will be an exciting avenue of future exploration. Indeed, careful dissection of the mechanisms by which the host (with a variety of disease-associated mutations in *ATG16L1*, *IRGM*, and *LRRK2*) handles different strains of intracellular pathogens could provide fundamental insights into disease pathogenesis. Furthermore, autophagy looks like a readily 'druggable' target for CD therapeutics. It is noteworthy that rapamycin, the most commonly used laboratory agent to induce autophagy (via mTOR), has been in clinical use in a number of different arenas for several years, where it is now known as sirolimus (Rapamune; Wyeth). Sirolimus is frequently used as an immunosuppressant post-solid organ transplantation, either as monotherapy or in conjunction with calcineurin inhibitors and/or mycophenolate mofetil. Sirolimus-eluting coronary stents (Cypher; Cordis Corporation) used in coronary artery disease patients are reported to have lower re-stenosis rates when compared with bare metal stents. Sirolimus has also been shown to shrink tumors (astrocytomas and renal tumors) in patients with tuberous sclerosis complex. Indeed, there is much hope that the anti-

proliferative effects of agents such as this will be of clinical benefit in a number of different malignancies.

Very recently, a single case report has described the clinical use of sirolimus in a patient with severe CD, with the drug inducing a marked and sustained improvement in symptoms and endoscopic appearances.[222] In animal models, rapamycin has been demonstrated to be as effective as cyclosporine in reducing experimental chronic colitis induced by DSS.[223] These observations lead to the prospect of early clinical trials of these agents to induce autophagy in patients with CD. However, due to the wide-ranging effects of autophagy in both homeostasis and pathogenesis in a variety of organs and tissues, not to mention the GI tract itself, it is likely that more specific targeting of defined autophagy pathways will be required, and delivery mechanisms developed to limit unwanted or dangerous systemic side effects.

1.5.2.5 NK2 transcription factor related, locus 3 *(NKX2.3)*

The WTCCC study identified a cluster of associated SNPs around rs10883365 ($p = 1.4 \times 10^{-8}$) at 10q24.2, the location of *NKX2.3*.[81] Evidence for association at *NKX2.3* was also obtained in the UK CD replication study of rs10883365 ($p = 0.0037$, $p_{combined} = 3.7 \times 10^{-10}$; OR 1.18, C.I. 1.05-1.32)[89] and in a German tSNP study where the best marker was rs11190140 ($p = 8.87 \times 10^{-6}$; OR 1.22, C.I.1.05-1.42).[85] Recently, rs10883365 has also been shown to be associated with Japanese CD ($p = 0.0065$; OR 1.29, C.I. 1.07-1.54).[224] Of note, these two markers from the UK (rs10883365) and German (rs11190140) studies are in perfect LD.[85] *NKX2.3* has also been confirmed as a susceptibility gene for UC, in both the UK ($p = 2.4 \times 10^{-6}$; OR 1.20, C.I. 1.05-1.38, at rs10883365) and Germany ($p = 1.33 \times 10^{-8}$; OR 1.39, C.I. 1.16-1.37, at rs11190140).[84, 85]

NKX2.3 is a member of the NKX family of homeodomain-containing transcription factors implicated in cell type specification and maintenance of differentiation in a number of different tissue types. In mouse, Nkx2.3 is expressed in gut mesoderm and branchial arches from E9.5, persisting in the hindgut until adulthood.[225] During later stages of murine development, expression is confined to the inner ring of gut mesoderm (smooth muscle), immediately underlying the mucosal endodermal epithelium.[226] In the adult, highest levels of expression are noted in the ileum with lowest levels of expression in the duodenum and rectum. In addition to smooth muscle expression, NKX2.3 is expressed in numerous

mesenchymal cells in the intestinal lamina propria and in endothelial cells of small blood vessels.

A significant proportion of *Nkx2.3*$^{-/-}$ mice die from acute intestinal malaborption within two weeks of birth.[226, 227] In the early postnatal period, these animals have steatorrhoea and bloody diarrhoea, with lipid-staining vesicles predominant in the intestinal mucosa. Whilst TNFα expression was increased in one study of mutant intestine, no acute inflammation was detected at any post-natal stage.[226] Those mice that survive this period have reduced intestinal length and intestinal distension secondary to hyperplasia (not dilatation), but otherwise appear healthy as adults and are fertile.[226] Meanwhile, heterozygotes are viable and appear to develop normally.[226, 227] The homozygous mutant animals have severe defects in both the spleen and intestine. Of note, 1 in 3 of these mice are asplenic; the spleens in the remaining animals are approximately 10-fold smaller than wild-type animals and demonstrate severe structural abnormalities.[226, 227] They lack a marginal zone, have abnormal segregation of B and T lymphocytes, and a paucity of macrophages in red pulp. They are noted to have a smaller caecum than WT mice, from E15.5 onwards.[226] In the small intestine, villus formation is delayed during development; however, normal differentiation of epithelial cell types is present in surviving adult mutants.[227] Furthermore, these animals have significant abnormalities of gut-associated lymphoid tissue with fewer and smaller Peyer's patches and lymphoid aggregates than WT. Moreover, MadCAM-1 expression is significantly down-regulated (about 10-fold reduction) in areas that normally express NKX2.3, providing a potential mechanism by which they have deranged lymphocyte homing.[226] The importance of lymphocyte homing via MadCAM-1 in IBD pathogenesis has been clearly demonstrated as testified by the clinical efficacy of the anti-alpha4 monoclonal antibody, natalizumab, in CD.

1.5.2.6 Protein tyrosine phosphatase non-receptor type 2 *(PTPN2)*

The strongest hit at 18p11 in the WTCCC study was at rs2542151 ($p = 4.6 \times 10^{-8}$), 5.5kb upstream of *PTPN2*.[81] Whilst only modest evidence for replication at rs2542151 was provided in both the index study ($p = 0.048$)[89] and in the German cohort ($p = 0.01$)[85], this SNP was robustly replicated in the meta-analysis ($p_{replication} = 2.41 \times 10^{-7}$, $p_{combined} = 5.10 \times 10^{-17}$).[103] There was no evidence for association at *PTPN2* in the smaller Japanese CD cohort.[224] As mentioned above (**1.5.2**), *PTPN2* has also subsequently been associated with coeliac disease and type I diabetes mellitus.

PTPN2 (OMIM 176887) encodes the T cell protein tyrosine phosphatase (TCPTP) that is a key negative regulator of inflammatory responses. TCPTP is located intracellularly, containing no transmembrane domains.[228] It is expressed in all tissues and at all stages of development. There are two splice variants in humans (48kDa: ER targeted; 45kDa: nuclear and cytoplasm targeted) but only one is mice (45kDa). Expression is ubiquitous, but highest in cells of haematopoietic origin; the knock-out mouse dies of severe anaemia by post-natal weeks 3-5.[229] Interestingly, this mouse, although apparently normal through embryonic development, develops a severe systemic inflammatory disease / wasting disease characterised by runting, hunched posture, diarrhoea, splenomegaly (caused by macrophage infiltrates) and weight loss.[229] The severe immunosuppression is characterised by defective T-cell-dependent B-cell responses. TCPTP is an important regulator of CSF1 (colony stimulating factor-1) and mononuclear-phagocyte development (*Tcptp*$^{-/-}$ show increased granulocyte-macrophage precursors).[230] Furthermore, the *Tcptp*$^{-/-}$ mice show widespread lymphocytic infiltrates in nonlymphoid tissues correlating with increased IFNγ, TNFα, IL12, nitric oxide, and increased LPS sensitivity.[231] Increased cytokine production is detected as early as day 3 post-partum, preceding the onset of systemic disease (symptoms develop at day 10-14 post-partum), suggesting this is a primary abnormality in these animals. IL12/IL23-p40 expression is massively up-regulated in the liver and to a lesser extent in the spleen at day 3. Intra-peritoneal injection of 4μg LPS to 20 day old *Tcptp*$^{-/-}$ mice led to the development of septic shock, not seen in WT littermates, with massive induction of serum IFNγ.

TCPTP has been shown to dephosphorylate distinct substrates, including the insulin receptor,[232] EGFR,[233] JAK1,[234] JAK3,[234] STAT1,[235] STAT3,[236] and most recently STAT6.[237, 238] As a result, it can regulate signalling pathways that are induced by various growth factors (e.g. EGF) and cytokines (e.g. TNFα). In the case of TNFα, MAPK signalling is suppressed by TCPTP (it interacts with TRAF2 to dephosphorylate and hence inactive Src tyrosine kinases).[239] TCPTP-deficient cells show enhanced IL-6 production in response to TNFα, an effect blocked with inhibition of ERK1/2 activation.[239] TCPTP dephosphorylates nuclear STAT6 (PTPN1 does the same to cytoplasmic STAT6), inducing the expression of IL4 target genes.[237] The 45kDa TCPTP recognises mutant EGFR (a common rearrangement with a truncated protein with an in-frame deletion of 267 amino acids that is detected in glioblastoma, breast, lung and prostate cancer) as a cellular substrate and dephosphorylates it

leading to decreased MAPK ERK2 and PI-3K, suppressing the tumorigenicity of mutant EGFR-expressing glioblastoma cells.[240]

1.5.3 CD GWAS meta-analysis

The individual GWAS reports mentioned were powered to identify associations with OR of ≥1.3 (WTCCC) or ≥1.5 (N American and French / Belgium). Combining these datasets gives a total population of 3230 CD patients and 4829 HC and as a result 74% power to detecte OR of ≥1.2. Barrett and colleagues recently performed a meta-analysis on this combined cohort.[103] This would previously have been technically impossible as the different scans were done on two contrasting platforms. However, recently developed algorithms (IMPUTE and MACH) were employed to impute genotypes on the relevant 'missing' SNPs for each platform to generate data for all studies on 635,547 SNPs. This process relies on the observed haplotype patterns in the HapMap cohort, used as a reference population. Imputation was performed on all samples irrespective of case status to avoid introducing artefact. As the 3 different GWAS had carefully matched population controls, Barrett performed the meta-analysis on the results on each study, rather than combining the raw data.

526 SNPs in 74 distinct loci achieved a nominal significance threshold for replication ($p < 5 \times 10^{-5}$). 11/74 loci were previously reported. Therefore 63 loci were genotyped in a replication panel consisting 2325 CD patients, 1809 HCs, and 1339 trios for case-control and TDT analysis. 21/63 loci were successfully replicated (19 meeting replication threshold of $p < 0.0008$; 2 additional loci otherwise satisfying genome-wide significance levels – $p < 5 \times 10^{-8}$ - in combined cohort), thus bringing to 32 the number of confirmed loci for CD.

In addition to those results already mentioned above in the context of IL23 signalling, the meta-analysis data provide substantial new insights into CD pathogenesis. *ICOSLG* is expressed in intestinal epithelial cells where it acts as a costimulatory signal for T-cell proliferation and cytokine secretion.[241] In addition, ICOSLG is upregulated in mature plasmacytoid DCs; as a result they can induce IL10 producing T_{REGS} from naieve CD4 T cells.[242]

Intelectin-1 (*ITLN1*) is a lectin that recognises sugar motifs specific to bacterial cells walls. ITLN1 is expressed in the gut, specifically Paneth cells, and induced by parasitic infection in a STAT6 dependent manner.[243, 244] It has also recently been shown to be decreased in lung

epithelium from both 'healthy' and diseased lungs from cigarette smokers (there was no change in the expression of all other lectin genes studied).[245] It will therefore be of interest to see whether ITLN1 expression is decreased in the terminal ileum or colon from smokers, potentially decreasing pathogen clearance.

LRRK2 is implicated in the pathogenesis of Parkinson's disease, is expressed in myeloid cells, and is strongly implicated in autophagy.[246] *ZNF365* is a zinc finger protein that is a cause of susceptibility to uric acid neprolithiasis in a founder population in Sardinia.[247] Little is known about *CDKAL1*, but of note it has also been confirmed as a susceptibility gene for both type II diabetes mellitus and psoriasis.[248, 249] *CCR6* is the receptor for the chemokines CCL20, which is discussed in much more detail in Chapter 10.

1.5.4 Ulcerative colitis genetics

The genetic architecture of IBD genetics has become better defined in recent months, thanks to a series of large studies in UC, previously lagging significantly behind the progress in CD. As already discussed, whilst several loci are common to both CD and UC, several (notably *NOD2* and the autophagy genes) are distinct to CD. In addition to historic associations in the HLA region and with *MDR1*, 3 UC GWAS studies from the UK, Germany and N America have added a number of novel, specific UC loci.[84, 86, 142]

1.5.4.1 *Human Leukocyte Antigen (HLA) region*

In UC, a stronger association exists between genes of the HLA region (6p21), involved in regulating the immune response, than for CD. Despite confounding effects of ethnic origin and disease heterogeneity, this association is strongest in patients with extensive UC; a positive association with *DR2* (in particular, the *DRB1*1502* subtype) and the rare alleles *DRB1**0103 and *DRB1**12, and a negative association with *DR4* and *DRw6* have been reported.[250] The *HLA DRB1*0103* allele situated at the HLA Class II region has been shown to be associated with extensive disease, the need for surgery and extraintestinal manifestation in the Caucasian population. Although a rare allele in the general population, this association has been consistently replicated in at least three independent studies and a meta analysis.[251-253] While the allele is only carried by 1-3% of the general population, carriage is 7.4-11% in UC patients with extensive or pan-colitis. Furthermore, in a study of patients who have undergone surgery and ileo-anal anastamosis for severe UC, *HLA DRB1*0103* allele was present in 15.8% of patients with extensive colitis at time of surgery, and in 22.8% of a subset of patients who also had extraintestinal manifestations. An association between the

*HLA DRB1*1502* (DR2) with UC has also been observed in several studies, notably in USA and Japan.[254-256] Moreover, in the Japanese population, *HLA DRB1*1502* (a sub-group of *DR1*15*) is common and has been associated with intractable disease. *HLA-DR4* allele has been shown in several studies to be protective against UC, and this was confirmed in a meta-analysis.[257]

In the UK nsSNP study MHC tSNPs were also genotyped in the discovery cohort.[84] The association in this region peaked at rs6927022 ($p = 4.7 \times 10^{-8}$). As this SNP failed genotyping in the replication cohort, a surrogate marker (rs660895) in tight LD with rs6927022 was used with evidence of replication ($p = 0.0035$, $p_{combined} = 2.8 \times 10^{-8}$). This SNP lies in a 400kb haplotype block containing *BTNL2*, *HLA-DRA*, *HLA-DRB5*, *HLA-DRB1*, and *HLA-DQA1*. The design of this study, with high marker density in this region, demonstrated that the association is confined to this block and does not extend beyond the recombination hotspots at either end. In an initial attempt to identify the causative mutations within this large region, the *BTLN2* variants were stratified by *HLA-DRB1*1502* status providing clear evidence for residual association at *BTLN2* ($p = 0.0036$). The German UC GWAS (see **1.5.4.5**) demonstrated genome-wide levels of significance at 3 SNPs in this region: rs9268877 ($p = 5.23 \times 10^{-7}$, $p_{combined} = 6.48 \times 10^{-18}$) and rs9268858 ($p = 5.41 \times 10^{-7}$, $p_{combined} = 2.58 \times 10^{-12}$) in *HLA-DRA*, and rs9268480 in *BTLN2* ($p = 2.21 \times 10^{-6}$, $p_{combined} = 3.15 \times 10^{-9}$).[86] The HLA SNPs in the WTCCC study were stratified by disease location (colonic only and ileal only). This provided additional evidence that the HLA region confers risk of colonic IBD – i.e. both UC and colonic CD.[84]

1.5.4.2 Multi-drug resistance gene 1 *(MDR1/ABCB1)*

MDR1 (*ABCB1*; encoding P-glycoprotein 170) is located at chromosome 7q22, in a susceptibility locus described by a UK genome-wide scan in 1996.[99] The knockout mouse (*Mdr1*$^{-/-}$) develops a spontaneous colitis in specific pathogen-free conditions that is reversed by antibiotics and absent in germ-free housing.[258] Expression of MDR1 decreases with inflammation in various animal models of colitis, and in colonic biopsies from patients with UC.[259] Inherited variation in *MDR1* has been associated with UC susceptibility in most, but not all, populations studied. A Scottish gene-wide haplotype-tagging study has demonstrated a significant association in *MDR1* variation with UC ($p = 4.22 \times 10^{-7}$), but not CD.[260] The variants most extensively studied in association studies are C3435T and A2677G/T, although

results are variable and it is not clear if either of these SNPs is causative. Phenotypic sub-studies have reported a strong association with extensive disease in UC.

1.5.4.3 Myosin-IXb *(MYO9B)*

Recently, inherited variations of the *myosin-IXb* (*MYO9B*) gene have also been shown to be associated with UC in four independent populations.[261, 262] It is of interest to note that the initial association was primarily observed with coeliac disease in the Dutch population.[263] The mechanism underlying this is unclear. Overexpression of rat myosin-IXb leads to actin filament related morphological changes in epithelial cells, and human myosin-IXb is expressed in intestinal epithelial cells. It is postulated that variations of the *MYO9B* gene may influence intestinal permeability leading to inflammatory intestinal disease susceptibility.

1.5.4.4 Extracellular matrix protein 1 *(ECM1)*

In the UK nsSNP GWAS, the strongest novel association was with a locus on 1q21.2 containing *ECM1*, *MRPS21*, *PRPF3* and *TARS2*.[84] The strongest markers in this region, rs3737240 and rs13294) withstood replication in two independent cohort ($p_{combined} = 2.3 \times 10^{-6}$ and 7.9×10^{-6} respectively). Whilst additional tagging at ECM1 provided evidence of association in two snSNPs within the gene, conditional regression analysis showed that these did not fully explain the association. Therefore, further fine-mapping and deep re-sequencing is required to determine whether the causative mutations lie within *ECM1* or elsewhere in the haplotype block. Despite this, *ECM1* remains an excellent biological candidate for a UC susceptibility gene. It is expression throughout the intestine, interacts with the basement membrane, inhibits MMP9 and strongly activates NFκB. It is noteworthy that rs3737240 and rs13294 were also associated with ankylosing spondylitis ($p = 0.0041$ and 0.0044 respectively)[192] given the overlap between these two diseases clinically.

1.5.4.5 *Interleukin 10 (IL10)*

Despite only finding GWAS levels of significance at the HLA region in the gene discovery cohort in the German GWAS, the extensive replication stages in this study (3 additional large cohorts from Germany, UK and Belgium/Netherlands, totalling 3022 UC cases and 3867 HCs) also provided robust evidence for association at IL10.[86] Following on from the limited UK nsSNP study, this was the first full GWAS in UC using the Affymetrix 500k chip and leaving 440k SNPs for analysis after quality control. With a gene discovery cohort of 1,167 UC cases and 777 HC, Franke estimates 80% power to detect a variant with OR ≥ 1.5 ($\alpha = 0.05$) assuming MAF of ≥ 0.20 in HC.[86] The 20 most significant SNPs in this cohort were

replicated across all 3 additional panels. The lead SNP in *IL10* was rs3024505 located 1kb downstream of the 3 UTR ($p_{combined}$ = 1.35 x 10^{-12}, OR 1.46, C.I. 1.31-1.62). Following an *IL10* tSNP (n=22) supplementary study in all 4 panels, 2 additional SNPs (both intronic) were associated with UC: rs3024495 ($p_{combined}$ = 2.69 x 10^{-11}) and rs3024493 ($p_{combined}$ = 6.16 x 10^{-12}). The 3 associated SNPs were in perfect LD; regression analysis suggested they might act independently of each other. Of note, rs3024505 was also associated with German CD, albeit weakly (p = 0.013, OR 1.17, C.I. 1.01-1.34). The location of this SNP – 79bp from a highly conserved stretch of DNA at the 3′ end of *IL10* – is of some interest, at it contains a putative AP-1 binding site.

As many as 10 historic studies (from 1999 to 2006) in multiple different populations have analysed the *IL10* promoter (at -592 [rs1800872], -819 [1800871], and -1082 [rs1800896]) for evidence of association with UC and CD with very variable results. None of these 3 promoter SNPs were associated with UC in the German GWAS.[86] This does provide a salutary reminder of the strength of GWAS design, compared with these older candidate gene studies.

As previously mentioned (see **1.4**), *Il10*$^{-/-}$ mice develop spontaneous colitis in SPF conditions due to defective counter-regulatory anti-cytokine responses. Indeed, these animals provide one of the oldest (first reported in 1993), most widely used and best characterised genetic animal models of colitis. Physiologically important defects in IL10 signalling in LPMCs from UC patients were described in 1995,[264] and by 1998 very early clinical trials of subcutaneous recombinant IL10 produced some evidence of benefit in UC patients.[265] For some reason, the phase II trials of this protein (rHuIL-10) were in CD, and when they failed to demonstrate efficacy this avenue was quietly dropped from further investigation.[266, 267] However, given these new genetic data, further exploration of IL10 delivery should be explored as a therapeutic avenue in UC, perhaps utilising Larry Steidler's ingenious delivery mechanism via recombinant IL10 producing *Lactococcus lactis*.[268]

1.5.4.6 N American UC GWAS: 1p36 and 12q15

The N American group recently reported their GWAS in UC.[142] In this study they only included patients with inflammation extending proximal to the rectum (Montreal E2 or E3). In the gene discovery cohort, 1052 patients with UC and 2571 geographically matched HC were included. Two independent replication panels were used: 786 UC and 721 HC from N

America and 619 UC and 394 HC from S Italy. Two novel regions met genome-wide levels of significance. At 1p36 there were two independent signals ($r^2 < 0.01$) from rs6426833 ($p = 6.8 \times 10^{-10}$, pcombined = 5.1×10^{-13}; OR 0.73) and rs3806308 ($p = 4.7 \times 10^{-8}$; pcombined = 6.7×10^{-9}; OR 0.78), separated by recombination hotspots. rs6426833 maps to *OTUD3* and *PLA2G2E* (phospholipase A2, group IIE), whilst rs3806308 maps to *RNF186* (ring finger protein 186).

At 12q15, rs1558744 ($p = 5.5 \times 10^{-10}$, pcombined = 2.5×10^{-12}; OR 1.35) and rs2870946 ($p = 1.0 \times 10^{-5}$, pcombined = 4.8×10^{-7}; OR 1.54) were independently associated with UC ($r^2 = 0.031$).[142] This region is particularly noteworthy as it maps to the edge of the IBD2 region (see **1.5.1.3**), and contains *IFNγ*, *IL26*, and *IL22*. These genes are 44 – 137kb from rs1558744, whilst rs2870946 is located in *IL26*. *IL26* and *IL22* are both secreted by Th17 cells.[269] The role of *IFNγ* in the regulation of multiple areas of mucosal immune homeostasis and in the response to pathogens is well established.

1.6 *The hedgehog (HH) signalling pathway in gastrointestinal development, homoestasis and disease*

It is becoming increasingly clear that signalling pathways important in pre-natal development continue to have vital roles in adult life by directing differentiation, determining cell fate, maintaining stem cell niches, and co-ordinating appropriate cellular responses to injury. These pathways influence many disease processes; in some situations the normal critical balance of signalling components is upset by inherited variants, whereas in others ligand-driven upregulation of signalling is a sufficient driving force. In the GI tract, the hedgehog (HH) signalling pathway has been shown to have critical roles in development, homeostasis and disease (**Figure 1-8**) (Lees *et al*, *Gastroenterology* 2005).[270]

1.6.1 HH signalling (Figure 1-8)

The HH signalling pathway was originally described in the development of *Drosophila melanogaster* as a segment polarity gene required for embryonic patterning,[271] a discovery that led to Nusslein-Volhard and Wieschaus winning the 1995 Nobel Prize for Medicine. The genes involved in Drosophila are *Hedgehog (Hh), Patched (Ptc), Smoothened (Smo), Hedgehog-Interacting Protein (Hhip), Costal-2 (Cos-2), Fused (Fu), Suppressor of Fused (Su(Fu)), and Cubitus interruptus (Ci)*. The pathway components demonstrate high inter-species conservation.[272] There are three vertebrate homologues of Hh, *Sonic hedgehog (SHH), Indian hedgehog (IHH), and Desert hedgehog (DHH)*. These demonstrate different,

but frequently overlapping, expression patterns. They have remarkably similar biological properties, albeit with differing potency [*SHH>IHH>DHH*] noted in some, but not all, experimental assays.[273] Increasing evidence implicates accessory molecules in mediating HH activity,[274-276] and there may be a role for these in modulating this potency. Most research has centred on *SHH*. The two homologues for patched, *patched-1 (PTCH) and patched-2 (PTCH-2),* both bind vertebrate HH with similar affinity. *PTCH* is found in target cells and upregulated by HH signalling.[277] *PTCH-2*, although little studied, is co-expressed with HH and does not depend on it for transcription.[278] The three homologues for *Ci, GLI1, GLI2, and GLI3*, are responsible for many of the refined complexities and intricacies observed in mammalian HH signalling.

HH is synthesised as a 45 kDa precursor protein that undergoes autoproteolysis.[279-281] The active N-terminal signalling domain (N-HH) is released once the catalytic C-terminal portion has been removed and a cholesterol molecule has been added (N-HHp).[282] With the addition of cholesterol, a modification to N-HHp that is unique amongst signalling molecules, N-HHp is rendered hydrophobic and thus able to bind to the cell membrane where it mediates local signals. In an independent step, the N-terminus can be further modified by palmitoylation, increasing hydrophobicity.[283] Furthermore, N-HHp, once released from the cell via the transmembrane protein Dispatched (DISP1),[284] functions to provide a long-range, paracrine signal to target cells.[285] At the present time, the mechanisms involved in short-range and long-range signalling remain contentious and under investigation. The transduction of the HH signal has many interesting and unique facets (**Figure 1-9**), that share some similarities with WNT signalling.[286-288]

As with most cell signalling pathways, regulatory mechanisms exist for HH signalling. HH activity induces further expression of its transmembrane receptor PTCH, thus simultaneously initiating signalling whilst restricting the range of movement by sequestering HH protein. Another cell-surface protein that binds to and sequesters HH, HHIP,[289, 290] is also upregulated by pathway activation, thus serving to downregulate HH activity. The phenotype of HHIP knockout mice is consistent with increased HH signalling.[291] GLI1 is a direct transcriptional target of HH and a robust indicator of pathway activity. Megalin, a LDL receptor-related protein, is also active on the cell surface and controls the endocytic uptake of the HH protein by acting as a direct binding partner *in vitro,* although little is known of its function *in vivo*.[292, 293] The regulation of HH signalling at different levels indicates that tight control is

crucial to its proper function, analogous to regulation of other key signalling pathways, such as NFκB, whose activation induces inhibitory molecules (IκBα and others).[294]

1.7 *HH signalling in gastrointestinal development*

The gut develops from a primitive endodermal tube that gives rise to the pharynx, oesophagus, stomach, small intestine and colon. Endodermal buds grow into mesenchyme closely associated with the tube and form the liver and pancreas. The key molecular pathways involved are the HH, bone morphogenetic protein (BMP), Notch, and WNT/β-catenin signalling pathways, the Hox and Sox transcription factors, and the Eph receptor/ephrin ligand signalling system.[295] Patterned gene expression within the endoderm and surrounding mesoderm regulates the morphogenesis, differentiation and boundaries of these organ primordial.[296] The cross-talk generated between endoderm and mesoderm is critical to this process and later establishes radial patterning within the developing digestive tract. HH signals are a vital component of this cross-talk and thus essential to the normal patterning of the GI track along anterior-posterior (A-P), dorsal-ventral (D-V) and radial axes.[297] In the animal models studied high levels of the HH receptor, PTCH, in the mesenchyme allows for processing of the signals generated endodermally.[298-302] Paracrine signalling in the late gestational (E18.5), murine small intestine has been elegantly demonstrated after careful separation of these components.[303] SHH and IHH are present in epithelial cells, with PTCH and GLI1-3 mRNAs detected almost exclusively in the mesenchymal compartment. Outside of the intestine, HH signals have been reported to act cell-autonomously; in developing and adult pancreatic islets IHH and PTCH co-localize.[304, 305] Of great relevance, manipulation / interruption of HH components in animal models bears a close resemblance to many human congenital malformations arising from abnormalities in fore- and hindgut development.

Developmental patterning by Hh in Drosophila is largely dependent on concentration gradients.[306] In the mammalian ventral neural tube, the morphogenic gradients of SHH setup in the notochord and floorplate determine the differentiation of progenitors into several different cell types.[307] In the developing gut, the overlap in expression of two HH homologues, SHH and IHH, may serve to generate greater variation in concentration gradients. Additionally, the different phenotypes of the knockout mice ($Shh^{-/-}$ and $Ihh^{-/-}$) suggests that the two homologues have different functions during intestinal development.

HH proteins are highly expressed in the murine embryonic gut and decrease after birth.[302] Shh and Ihh are expressed in mouse gut endoderm in overlapping patterns. From day 8.5 of gestation (E8.5) the proteins are present in two ventrolateral strips in gut endoderm.[308] This is noted first in the caudal hindgut, followed by foregut pocket and then hindgut. Shh is downregulated in two critical areas – in the prospective pancreatic endoderm to allow normal pancreatic development, and along the small intestine, possibly to allow normal intestinal epithelial differentiation.[309] By E18.5, there is expression of Shh and Ihh in the glandular stomach, small intestine and colon. Shh is mostly restricted to colonic crypts at this time.

1.7.1 Shh, Ihh, and Gli knockout mice

Shh$^{-/-}$ & *Ihh*$^{-/-}$ mice die peri- or immediately post-natally.[299] Both mutant embryos (examined at E18.5) show significant gastrointestinal defects, both common (smaller, overtly malrotated GI tracts) and distinct (**Table 1-2**). There are no data reporting any such defects in *Dhh*$^{-/-}$ mice (males are viable, but sterile due to defective spermatogenesis).[310]

The *Shh*$^{-/-}$ embryo displays a hyperplastic stomach epithelium with some intestinal transformation of the glandular epithelium. Overgrown villi in the duodenum are seen to cause an occlusion similar to duodenal stenosis in humans. On dissection, the colon terminates in a blind dilatation leading to imperforate anus.[299] Furthermore, *Gli2*$^{-/-}$ and *Gli3*$^{-/-}$ mice exhibit, respectively, imperforate anus with recto-urethral fistula and anal stenosis.[311] Mice with a truncated *Gli3* (*Gli3*$^{\Delta 699/\Delta 699}$), resembling the aberrant transcription factor in Pallister-Hall syndrome, display imperforate anus, reduced size of small intestinal villi, dilated intestine and thin walled colons.[312] *Gli1*$^{-/-}$*;Gli2*$^{-/-}$ double knockouts have a persistent cloaca whilst the mixed homo-heterozygous mutants show interim changes consistent with a gene dose-dependent effect.[311] Hh signalling is involved in the development of a normal hindgut and although probably important to the pathogenesis of human anorectal malformations, this has yet to be studied in detail. Of great interest, the *Gli1*$^{-/-}$ mouse appears to develop normally, although the published literature does not extend to a detailed analysis of gastro-intestinal ultrastructure.[313]

In both *Hh* knockouts there is decreased thickness of the circular smooth muscle layer along the small intestine (34% reduction for *Ihh*$^{-/-}$ and 21% for *Shh*$^{-/-}$ when compared with WT).[299] Both also show abnormalities of the enteric nervous system (ENS). *Shh*$^{-/-}$ mice have neurons that differentiate abnormally under the epithelium and migrate into the villi. The healthy

epithelium is known to inhibit proliferation of enteric neurons, an effect that is lost when Shh signalling is blocked with the general Hh pathway-inhibitor cyclopamine (**Figure 1-9**).[314] Grafting Shh-expressing cells into the mesenchyme limits neural proliferation in the vicinity. Furthermore, Shh promotes proliferation of neural crest cells whilst inhibiting differentiation and modulating their responsiveness to glial cell line-derived neurotrophic factor.[315] Taken together, these data strongly implicate Shh signalling in proper radial patterning of the ENS.

Ihh$^{-/-}$ mice (and to a lesser extent *Gli3*$^{\Delta 699/\Delta 699}$ mice)[312] macroscopically exhibit marked dilatation of the small intestine and parts of the colon, corresponding to a microscopic absence of neurones.[299] Some neurones are present in a normal pattern in patches of non-dilated colon. This colonic phenotype is observed with a penetrance of about 50%. The inference is that neural crest cells can migrate into the gut and differentiate, but in the absence of Ihh locally they fail to survive or proliferate. Interestingly, this phenotype resembles Hirschsprung disease in humans. The one study to assess a genetic link failed to associate *IHH* polymorphisms with Hirschsprung's, but was probably under-powered.[316]

Ihh is normally expressed in the intervilli region of the small intestine, the presumed stem cell compartment.[299, 317] In *Ihh*$^{-/-}$ animals there is decreased size of the villi, and a 54% reduction in cell proliferation in the stem cell compartment. Abnormalities are also noted in the colon. The normal epithelial monoloayer with organised crypts is replaced by multilayered epithelial cells lacking crypts.[318] These data implicate Ihh in the regulation of stem cell proliferation and possibly epithelial cell migration in the intestine.

The *Shh* and *Ihh* double knockout embryo arrests development at early somite stages (E8.0). Utilising the pan-Hh inhibitor Hhip under control of the mouse villin promoter, Madison and Gumucio have generated a viable murine model to study the effects of abrogating all Hh signals in the postnatal small intestine.[303] The epithelium was noted to be flattened, with significant interference of villus formation and epithelial remodelling. Changes in the mesenchyme (expansion of smooth muscle progenitors, and mislocalization of intestinal subepithelial myofibroblasts (ISEMFs)) were found, in turn, to impact further on the epithelium with increased proliferation and enhanced β-catenin/Tcf4 activity. In this manner, the Hh pathway patterns the intestinal crypt-villus axis via a paracrine signal.

A further role for Hh signalling in the development and/or function of the murine postnatal intestine has been suggested by Wang and co-workers.[302] In view of the lethality of *Shh*$^{-/-}$ and *Ihh*$^{-/-}$ knockout mice above, these authors studied WT animals, and blocked Hh signalling at late stages of pregnancy and immediately post-natally with tail-vein injection of the anti-Hh monoclonal antibody 5E1, which blocks Shh, Ihh and Dhh. All pups died by 3 weeks, having displayed a runting and wasting phenotype post-natally. It is of some interest that this corresponded with the development of significant diarrhoea. The small intestines of these animals showed significant histopathological abnormalities, notably disorganised villi projecting into the lumen, and hyperplastic crypts with a 73% increase in proliferation compared with control animals. Furthermore, a prominent vesicular vacuolation was noted in enterocytes, predominantly affecting ileum and caecum. This defect became apparent only on post-natal day two (having been absent one day previously even in animals treated with 5E1 from E12.5), suggesting it might be dependent on colonisation of the GI tract with commensal bacteria. It was also noted to be an intestinal specific phenomenon with no such vacuolation identified in other cell types, including hepatocytes. Further characterisation showed that these vacuoles represent an intracellular accumulation of neutral lipid (Oil Red O positive, PAS negative). The possible involvement of Hh signalling post-natally in lipid absorption and secretion is also suggested by the numerous microscopic fat droplets in the stool of treated mice; and by decreases in serum apolipoprotein A-IV, total cholesterol and high-density lipoprotein cholesterol, in these animals.

This is further supported by a study in adults, demonstrating that monoclonal antibody-mediated inactivation of Hh signalling in mice fed a high-fat diet protects against weight gain[319]. Similarly treated leptin-deficient mice (*ob-ob*) that ordinarily gain weight on a low-fat diet, showed decreased weight gain compared to controls. Whilst total lipid absorption was normal, the rate of triglyceride absorption was significantly slowed and there was increased faecal free fatty acid excretion in the mAb treated animals. Indeed, the hepatic steatosis noted in the high-fat control group was abolished with abrogation of Hh signalling. However, these data need to be interpreted with caution, as detailed studies of cyclopamine treated animals demonstrate increased WNT/β-catenin activity and colonic cancers (discussed in detail in **1.9.1.2**).[318] This could account for at least a proportion of the weight loss demonstrated here.

1.7.2 Shh, foegut development and oesophageal fistula / tracheoesophageal fistula (EA/TEF)

The development of EA/TEF in *Shh*$^{-/-}$ mice suggests a role for Hh signalling in the pathogenesis of this condition and in normal foregut development.[298, 299] Furthermore, the human *GLI* genes are implicated in congenital foregut malformations by the significant abnormalities noted in the *Gli2*$^{-/-}$ and *Gli3*$^{-/-}$ mice (**Table 1-2**).[320] In normal murine development, foregut expression of Shh is particularly high when the trachea and oesophagus split.[321, 322] At E10.5 the normal undivided foregut has a ventrally placed prospective trachea that is positive for Shh.[323] By E11.5 this pattern is reversed with the separated trachea now negative for Shh.[298, 323] However, analysis of the adriamycin-induced rat model of EA/TEF shows that when the trachea is still undivided (at E11.5 in the rat, a slightly different developmental time point to the mouse), there is a diffuse Shh staining pattern that lacks any dorsal-ventral gradient.[323] The rat fistula tract shows much lower Gli2 levels than the adjacent oesophagus.[324] Furthermore, culture of fistula tract with exogenous Shh induces branching, a known Shh effect on the developing lung.[324] Examination of one resected human fistula tract in a newborn showed SHH in the proximal oesophageal pouch, but not in the distal fistula tract.[325]

Combined, these data show that the fistula tract is of respiratory, not oesophageal, origin. Shh is vital to the normal development of the oesophagus and trachea, and heavily involved in the pathogenesis of EA/TEF and the VACTERL association that the adriamycin model mimics. In humans, both foregut and hindgut anomalies are seen in some patients with holoprosencephaly (where mutant *Shh* causes cyclopia). As EA/TEF in humans do not appear to be caused by genetic defects, an environmental association in the early embryo that manipulates HH signalling is more probable than a *HH* mutation. What the relevant trigger is remains unclear, although it has been shown to be unrelated to maternal cocaine use as had been hypothesised.[326]

1.8 *HH signalling in the healthy adult gastrointestinal tract*

1.8.1 Patterns of expression in the adult GI tract

The main published survey of the HH signalling pathway in the gastrointestinal system in adult humans and rodents was carried out by van den Brink *et al*.[327] These authors demonstrated SHH mRNA to be abundantly expressed in the gastric fundus, in small quantities in the crypts of the small intestine, and in a few colonic crypts, but not in the

oesophagus or gastric antrum.[327] When they used commercially available goat polyclonal anti-SHH antibody they did not find any positive staining for SHH protein anywhere apart from in the gastric fundus.[327] Subsequently, work from three separate teams has now shown that SHH protein is present in the human colon, as will be discussed in more detail in Chapter 4.[328-330]

1.8.2 Functional aspects of HH signalling in the adult GI tract

1.8.2.1 SHH and fundic gland differentiation in the stomach

Within the stomach there are two compartments relative to the position of the epithelial stem cell. In the pit region cells are migrating toward the gastric lumen, and in the gland region they migrate in the opposite direction. SHH expression is restricted to the glandular portion of both human and murine stomach (utilising gastric mucins as genetic markers for each compartment). In humans, SHH expression has been shown to be greatest at the pit-gland transition and restricted to parietal cells (as confirmed by double staining with H^+K^+ATPase).[328, 331] PTCH is expressed on the epithelial cells of the gastric gland region and some of the interstitial cells. Pit cells are PTCH negative but parietal cells and epithelial cells at the base of the glands express PTCH.[331]

SHH expression correlates with fundic gland type, as seen where there is loss or gain of fundic gland differentiation. In chronic gastritis, where there is intestinal metaplasia (as identified by the presence of MUC2) there is a loss of SHH expression.[331] However, Dimmler *et al* suggest that this may be related to the process of atrophy, as SHH is highly expressed in non-atrophic gastritis.[328] In the oesophagus, SHH is present in parietal cells in Barrett's resection specimens, and in areas of Meckel's diverticulum.[331]

These data show that SHH may play a role as an essential polarising signal for fundic gland differentiation. It has been postulated that an intestinal epithelium may be the default state during development. SHH expression is part of a regulatory network that induces the character of gastric epithelium compared with intestinal type. This is then maintained by SHH signalling throughout adult life.

1.8.2.2 IHH and colonic enterocyte differentiation.

Cyclopamine treatment markedly alters the slender nucleus and cytoplasm of normal terminally differentiated enterocytes.[318] Without Hh pathway activity the enterocytes at the

luminal end of the crypt have an enlarged nucleus and large cytoplasm. In addition, they show significantly altered expression of three molecular markers of enterocyte differentiation, with redistribution of villin to the cytoplasm, decreased carbonic anhydrase and increased intestinal trefoil factor.[318] In the HT-29 colon carcinoma differentiation model, treatment of cells with butyrate induces expression of villin and Cip-1, and induces Ihh expression. Blocking Hh signalling with cyclopamine substantially decreases the Cip-1 and villin induction with butyrate. In addition, giving recombinant N-terminal Shh (91% homologous to Ihh) leads to induction of Cip-1 and villin. These data suggest a role for Ihh in regulating colonic epithelial differentiation *in vivo* and *in vitro*.

Furthermore Hh signalling restricts expression of Bmp4 (a Dpp homolog co-expressed with Hh genes during development) and engrailed-1 (a WNT target gene) to the precursor cell compartment at the base of the colonic crypt. β-catenin-TCF (a marker of WNT pathway activity) signalling can be completely abrogated by Ihh in colon cancer cells.[318]

Where WNT pathway activity is constitutively overexpressed (APC gene mutations in familial adenomatous polyposis) IHH expression is lost. This phenomenon is observed in FAP resection specimens that contain both normal and dysplastic epithelial cells (corresponding to APC mutations and WNT over-activity). Normal IHH expression is noted in the normal epithelial cells, but this is lost in dysplastic cells. Inhibiting β-catenin-TCF signalling in colon cancer cells causes rapid induction of IHH expression. These data together indicate that IHH expression is negatively regulated by β-catenin-TCF signalling. Furthermore, decreased IHH expression in colonic polyps is seen in response to the WNT-activating APC mutation. Because the loss of differentiation observed in the cyclopamine treated colonic epithelial cells mirrors that seen in colorectal carcinogenesis, loss of IHH may have a role to play in the development of dysplasia.[318]

As discussed, all methods currently available to manipulate HH pathway activity are non-specific for SHH or IHH. The reported data are heavily reliant on expression studies, and as previously shown there remains some conflict with other published literature (as is further discussed below in the specific context of colon cancer). It is therefore once again difficult to ascribe the effects seen here to one HH homologue or another. Greater specificity could now be achieved through the appropriate application of RNA interference technology, utilising specific sequences of double-stranded RNA to knock down the expression of

complementary genes, and determine the relative contributions of SHH and IHH in homeostasis of the adult digestive tract.

1.8.2.3 HH and Paneth cell differentiation

Varnat and colleagues have described a role for PPARβ and HH signalling in Paneth cell differentiation.[332] First, they demonstrated the widespread expression of PPARβ in the murine small intestine. Examination of the *Pparβ*$^{-/-}$ mouse revealed it to have not only a decreased number of Paneth cells and altered intestinal luminal flora when compared with WT, but also a 3-fold induction of IHH expression. Conversely, treatment of WT mice with L-165041, a PPARβ-specific agonist, led to increased IHH expression. PPARβ, therefore appears to negatively regulate IHH in the murine small intestine.

Of potential importance to future trials of clinical HH antagonists, intraperitoneal delivery of cyclopamine to WT mice led to a 30% increase in the number of mature Paneth cells in the small intestine.[333] In a series of elegant co-localisation experiments the authors demonstrated expression of IHH in mature Paneth cells, with the pathway response elements, PTCH and HHIP, located in Paneth cell precursors.[332] The data support a working model whereby IHH, synthesised and secreted by mature Paneth cells, is processed by HH response elements in precursor cells, thus negatively regulating their differentiation. However, it should be noted that this model is out of keeping with the present over-riding body of evidence that HH signalling is purely paracrine from epithelium to mesenchyme in the intestine. The Varnat studies used polyclonal HH antibodies which may be a source of error as some researchers have found them to be potentially non-specific.[334]

1.9 *HH signalling and gastrointestinal disease*

We have seen that HH signalling is critical to normal gastrointestinal development, and that it may have an important role to play in homeostatic processes of the adult stomach, colon and pancreas. These processes rely on time-dependent and tissue-specific signalling at distinct concentrations. This is made possible by differential expression of HH signalling components, the tight positive transcriptional and negative feedback loops (created by ligand-induced activation of negative regulators, PTCH and HHIP), regulatory interactions with WNT/β-catenin signalling, and by the pathway's ability to function as a bistable genetic switch, flipping cell fates at precise, threshold SHH concentrations.[335] However, as a result of this reliance on such tight control, increases or decreases in pathway activity can result in severe defects. It is important to consider how a developmental pathway that normally

directs differentiation and proliferation may, with the loss of its normal strict controls, contribute to cancer formation and maintenance, and chronic intestinal inflammation.

1.9.1 HH signalling and tumorigenesis

The HH signalling pathway interacts directly with cell-cycle components to increase cell proliferation. The G1-S transition is promoted by cyclins D and E, both transcriptional targets of Ci in Drosophila,[336] a finding confirmed in mammalian cells.[337] The G2-M transition is in part controlled by PTCH, which regulates the activation of cyclin B (part of the mitosis-promoting-factor complex).[338] SHH also blocks p21(CIP1/WAF1)-induced growth arrest.[339] Furthermore, there is some evidence that PTCH, like DCC, UNC5 and RET, functions as a dependence receptor.[340, 341] Dependence receptors are characterised by a cellular state of dependence on their ligand, such that absence of ligand induces apoptosis.[342] The presence of ligand (HH) prevents the induction of apoptosis. In the absence of HH, caspase-3-mediated cleavage of the intra-cellular PTCH domain exposes a receptor region that transduces the apoptotic signal via GLI3. It is noteworthy that many of the HH pathway activating *PTCH* mutations implicated in CNS, skin and muscle tumours map to the carboxy-terminus in the vicinity of the caspase-3 site.[343] Such critical functions as cell cycle control help to explain the importance of tight, multi-layered control on the pathway. It is perhaps, however, not surprising to learn that HH signalling is implicated in the induction and maintenance of cancer. This is, in part, by promoting cell cycle proliferation and opposing the normal stimuli for cell cycle arrest.

To date, aberrant HH signalling has been described in tumours of the skin, brain, lung, and digestive tracts.[344-349] A subset of HH-responsive cancers is, in part, caused by mutations in HH pathway components. This phenomenon was first identified in patients with Gorlin syndrome who, along with generalised body overgrowth, cysts and skeletal developmental abnormalities, have a predisposition to benign and malignant neoplasia, notably multiple basal cell carcinomas (BCCs).[350, 351] Patients inherit a mutant *PTCH* that permits constitutive pathway activation. A mutation in *PTCH* is also present in a proportion of medulloblastomas and rhabdomyosarcomas, where there is evidence that this confers on tumour cells the ability to resist apoptosis.[352] This loss of function phenotype defines *PTCH* as a tumour suppressor gene, and is consistent with its role as a dependence receptor. Somatic mutations in the *PTCH* gene are reported in oesophageal squamous cell carcinomas and transitional cell carcinomas of the bladder.[353, 354] Other oncogenic mutations include *SMO* in BCC (where

the pathway becomes independent of HH-PTCH binding),[355] and *SUFU* in medulloblastoma,[356] whereas ectopic expression of *GLI* causes glioma.[357]

It is now known that the HH pathway is also involved in the formation and maintenance of some sporadic tumours without implicating pathway-activating mutations. This is seen to be the case for small cell lung cancer,[349] prostate cancer,[358] pancreatic cancer,[348] cholangiocarcinoma and other tumours of the digestive tract.[346]

1.9.1.1 Cancers of the upper GI tract

The HH signalling pathway is active in many digestive tract cell lines. SHH and IHH mRNA is present in virtually all cell lines from oesophageal, stomach, biliary tract, pancreatic and colonic carcinomas. However, PTCH and GLI1, used as markers of HH pathway activity, are co-expressed only in oesophageal (4/6), stomach (6/6), biliary tract (5/9) and pancreatic (5/6) tumours. In these cell lines the GLI-luciferase reporter assay shows a high level of activity and therefore confirms autonomous pathway activity. Of the PTCH mRNA positive cell lines, treatment with cyclopamine resulted in a 75-95% decrease in growth. PTCH negative cell lines show no difference in growth with cyclopamine treatment.

The dramatic *in vivo* responses of pancreatic tumours to HH blockade have been paralleled by experiments described by Berman and colleagues, who have studied murine xenografts created from a human metastatic cholangiocarcinoma cell line.[346] 180mm^3 tumours treated with cyclopamine regressed completely in 12 days and remained histologically undetectable in mice that by 3 months post-treatment had shown no ill-effects of the cyclopamine treatment. The control animals euthanized at 22 days had tumours averaging 800mm^3. In contrast to the constitutional activation of HH signalling caused by mutant *PTCH* in Gorlin syndrome, these tumours are dependent on HH-ligand for growth. This has been demonstrated by the concentration-dependent decrease in tumour growth and luciferase activity with the ligand blockade (SHH and IHH) achieved using the 5E1 monoclonal antibody.[346] Tumour growth was partly rescued with the addition of endogenous SHH.

1.9.1.2 Colon cancer

Although the survey of digestive tract tumours by Berman *et al* failed to establish active HH signalling within a panel of 11 cell lines, SHH and IHH mRNA were detected in all cell lines, and GLI was present in 4 out of 11.[346] Furthermore, Qualtrough *et al* have shown expression of mRNA and protein of SHH, IHH, PTCH, SMO and GLI1 in all five cell lines

they tested.[359] Of these, the adenoma-derived cell lines (AA/C1 & RG/C2) have higher levels of IHH expression than the adenocarcinoma-derived lines (CaCo2, HT29, & SW480), but slightly lower levels of PTCH, SMO and GLI. These observations may reflect the evolving role of HH signalling with tumour phenotypic progression. All these cell lines, however, show a dose-dependent response to cyclopamine treatment, with decreased cell yield, increased apoptosis, and a decrease in autocrine HH signalling (decreased GLI-luciferase reporter activity). This effect is partially rescued with endogenous SHH-N protein. Similarly, Oniscu *et al* demonstrate SHH to have a mitogenic effect on primary murine colonocytes that is reversible with cyclopamine.[330] Furthermore, these authors were able to show increased SHH, PTCH and SMO expression in resected human adenocarcinomas, with intermediate levels in dysplastic tissue and benign adenomas.

In a subsequent study published in 2007, Chatel and colleagues concluded that HH signalling in inactive in colonc cancer cell lines following a detailed survey in 7 different lines.[360] None of the lines they examined had a full complement of HH signalling components. No HH ligand mRNA was expressed in Colo320 and HCT116 cells, no SMO in Colo205, HT29 and WiDr cells, no GLI1 in their SW480 cells and no GLI2 or GLI3 in Colo205 and Caco-2 cells. In their experience, cyclopamine treatment did not alter PTCH or GLI1 mRNA expression.

These data certainly appear conflicting and hard to reconcile with the van den Brink studies on IHH in colon cancer. However, Yauch and colleagues from Genentech have very recently published in *Nature* an extensive series of *in vitro* and *in vivo* studies that clarify matters substantially.[361] In essence, they demonstrated that HH signalling in colonic and pancreatic cancer is paracrine, just as Madison and Gumucio had demonstrated in gut development.[303] Their studies are probably conclusive, although they have been met by initial scepticism in some quarters.[362] The paracrine model clearly demonstrates that ligand driven autonomous HH signalling does not occur in colonic or pancreatic cancer. It remains to be seen whether or not these observations also hold for all epithelial-derived tumours and other solid organ malignancies. Whilst they definitely cast into serious doubt the accuracy of a series of high-profile papers in *Nature* published between 2003 and 2004,[346, 348, 349, 358] Yauch's paper has other important sequelae. Firstly, it demonstrates the fundamental mechanisms by which HH signalling drives colon cancer development and progression. This has provided the basis for a phase II clinical trial of a small molecule HH antagonist produced by Curis Inc.

(Cambridge, MA, USA) in advanced colo-rectal cancer (with concurrent chemotherapy and Bevacizumab) for which Genentech are currently actively recruiting patients (www.clinicaltrials.gov). Secondly, it demonstres the futility of studying HH signalling in cancer cells in isolation, certainly for intestinal lines and quite probably for a number of other tissues.

Yauch first studied the effect of a potent small molecule SMO-acting HH antagonist (HhAntag) on a large number of cancer cell lines from the colon, pancreas and lung.[361] Not only was a high concentration of HhAntag required for growth inhibition in these cells (~400 times higher than in a mesenchymal cell line expressing a *GLI* luciferase reporter construct), but no correlation was observed with markers of HH pathway activation (including GLI1 expression), strongly suggesting that the observed effect was indirect. This effect was also observed with cyclopamine (also SMO-acting), but not with alternative methods of antagonism (forskolin [PKA activation inhibits HH] or 5E1 antibody) (**Figure 1-9**). Importantly, they used mRNA analysis of GLI1 in this study in comparison to the previous reports where a *GLI* luciferase reporter was transfected to monitor pathway activity.

Whilst Yauch had shown no evidence of HH pathway activity in epithelial cells *in vitro*, RT-PCR on a panel of colo-rectal, pancreatic and ovarian (but not lung) cancers demonstrated significant upregulation of HH ligand.[361] The paracrine nature of HH signalling in these tumours was first illustrated in a co-culture model. In this system, epithelial cancer cells expressing HH ligand were shown to induce HH pathway activity in fibroblasts (10T1/2 cells) stably transfected with a *GLI* reporter construct. HH pathway activity was subsequently assessed in vivo using Ptch-LacZ mice to give an accurate readout of HH pathway activity in xenografts. The stromal compartments of pancreatic xenografts demonstrated high levels of β-galactosidase staining but only where the cancer cells were known to express high levels of HH ligand. Separate experiments in xenografts from human cancers (surgical biopsies of patients) confirmed using species-specific primer sets that the activation of HH pathway activity in tumour stroma was not an artefact of cell culture. Therefore, HH ligand in tumour cells activates the pathway in surrounding stroma in a paracrine manner. Finally, it was demonstrated that tumour growth was dependent on paracrine activation of HH pathway. HhAntag induced growth inhibition of pancreatic and colonic cancer xenografts was observed and directly related to HH pathway inhibition in the

stromal microenvironment. Canonical HH target genes were unaltered in tumour epithelium but down-regulated in stroma, strongly supporting the paracrine model.

1.9.2 HH and the immune system

Shh has been shown to be an important regulator of T-cell differentiation at the transitions from double negative (DN) 1 to DN2 cells and from DN3 to the double positive (DP) stage, as well as in T-cell receptor (TCR) repertoire selection (DP to SP stage).[363] Ihh and Shh have distinct but overlapping functions in the thymus. Whereas Shh is secreted by the epithelium, Ihh is produced by thymocytes and functions to control thymocyte number by providing concentration-dependent feedback on the production of DP cells.[364] The Howie lab in Edinburgh has shown SHH-mediated signalling to be a physiological component of peripheral T lymphocyte responses,[365-367] (reviewed in Benson *et al*, 2004[368]). SHH acts to modulate CD4+ T cell effector function in man,[366] with endogenously produced SHH playing a role in sustaining normal CD4+ T cell proliferation.[365] This response was enhanced by adding exogenous SHH with increased synthesis and secretion of of IL2, IFNγ and IL10 in CD3/CD28 activated CD4+ T cells.[366] The neutralising 5E1 antibody subsequently reduced IL2 and IFNγ expression, but not IL10. IFNγ has also been shown to induce STAT1-dependent expression of SHH in medulloblastomas.[369]

Varas and colleagues recently demonstrated physiological effects of autocrine HH signalling on thymic DCs in man.[370] They showed expression of PTCH, SMO and GLI1-3 mRNA in thymic DCs. On flow cytometric analysis, 50-60% of DCs expressed SHH protetin. Exogenous SHH led to increased expression of BCL2 and $BCLX_L$ and increased cell survival. CD40 activation of DCs, which increases HLA-DR, CD86, CD80 and CD83, down-regulated SHH and PTCH expression. Cyclopamine, on the other hand, increased apoptosis as well as abrogating the up-regulation of the co-stimulatory molecules as above.

Intriguingly, the pro-inflammatory mediator NFκB appears to upregulate SHH expression. In pancreatic cancer, blockade of NFκB was shown to suppress SHH mRNA.[371] Furthermore, IL1β, TNFα and LPS treatment of pancreatic cancer cells increased levels of SHH as well as NFκB, activated HH pathway activity and increased cell proliferation. NFκB inhibition abrogated these effects. NFκB is also required for the induction of SHH critical in hair placode down growth.[372] Furthermore, putative NFκB binding motifs have been identified in the SHH promoter region (at position +139), p65 binds to the SHH promoter *in vivo*, and

SHH expression is up-regulated on induction of NFκB *in vivo* in a genetic mouse model of inducible NFκB expression.[373]

1.9.3 HH signalling and inflammation

There is increasing evidence to implicate a recapitulation of embryonic HH signalling in normal and pathogenic inflammation. This has been shown in response to several insults, including acute epithelial injury in the lung[349] and prostate,[358] skeletal muscle ischaemia,[374] bone fracture,[375] multiple sclerosis / experimental autoimmune encephalomyelitis (MS / EAE),[376] chronic pancreatitis / pancreatic regeneration,[377, 378] liver / biliary injury,[379, 380] as well as in chronic GI tract inflammation.[329] In chronic pancreatitis there is increased IHH, PTCH and SMO with expression identified in the cells forming tubular complexes, and in the islets (with loss of the punctuate staining pattern seen previously).[378] In the lung epithelium there is extensive activation of the HH pathway in response to acute airway injury,[349] and in chronic lung fibrosis.[367] Much of the work done in the lung suggests that SHH (like KGF and FGF10)[381, 382] is involved in repair to epithelial injury, with the primary aim of restoring continuity to areas of denuded epithelium, thus minimising input of pathogenic agents across this barrier. In psoriasis, carefully controlled microarray experiments have shown down-regulation of HH pathway activity,[383] contradicting earlier reports where GLI1 protein expression was increased,[384] and a small uncontrolled 'proof of concept' study where cyclopamine 'cleared' psoriatic skin lesions.[385]

1.9.3.1 Liver and biliary injury

The Diehl laboratory at Duke University (N Carolina USA) has in the past 2 years published a series of papers outlining the role of the HH signalling pathway in liver and biliary injury.[386-392] With respect biliary injury, they first demonstrated that both cholangiocytes and fibroblastic cells have active HH pathway signalling in experimental models and patients with primary biliary cirrhosis (PBC).[389, 391] Subsequently, they showed in a rat model of reversible chronic bile duct obstruction that injured cholangiocytes express HH proteins, which are also co-expressed with mesenchymal markers in ductular cells in patients with PBC.[390] In a co-culture model (cholangiocyte and hepatic stellate cell lines) it was shown that hepatic stellate cells influence cholangiocyte migratory behaviour via soluble HH protein paracrine signalling.[390] In *Ptch*$^{+/-}$ mice (with increased Hh signalling activity), enhanced fibrogenesis with increased expression of various mesenchymal markers and Gli2, was evidenced in response to bile duct ligation.[390] This study suggests that HH has both

physiological and pathogenic actions in biliary injury, depending on its relative effects in cholangiocytes (as a survival factor) and as a profibrogenic agent.[387]

In a separate study from the same group, the role of the HH signalling pathway in alcohol mediated liver injury was assessed in mice and man.[388] The number of Hh-responsive cells were increased in mice fed high-fat [HF] diets and ethanol [Etoh] (in progenitor and stromal cell populations) and in patients with alcoholic liver disease [ALD] (in immature ductular cells). Fatty liver disease by itself (mice fed HF chow) led to increased expression of Shh, Ptch, Gli2 and sFrp1. The addition of Etoh to this model potentiated the increase in Shh, Ptch and Gli2 but not sFrp1. As predicted by the previous studies, Hh target gene expression (Ptch and Gli2) was restricted to rare bile ductular or hepatocytic cells in control mice. Hh-responsive cells were increased in HF mice (localised in or around the portal tract), with much greater expansion noted in HF + Etoh animals. The Hh-responsive ductular type progenitors expanded in HF + Etoh mice expressed high levels of TGFβ1. Subsequent *in vitro* experiments demonstrated that TGFβ1 promoted the survival of Hh-responsive immature liver cells, whilst increasing the production of Hh ligand (Ihh mRNA and protein, but not Hh target genes) in surviving mature hepatocytes undergoing epithelial-mesenchymal transition. On parallel analysis of liver biopsies specimens from patients with ALD, expansion of liver cell populations expressing PTCH and GLI2 were identified. GLI2 expression was co-localised with α-smooth muscle actin and vimentin in ductular structures and the epithelial progenitor marker CK7. Interestingly, there was close correlation between GLI2/CK7 expressing liver progenitors and Maddrey discriminant function (cut-off 32) (r^2 = 0.71).

1.9.3.2 Multiple sclerosis / EAE

The Gli1 expression data in the EAE model of MS is interesting.[376] Whilst Shh and Gli1 were increased during early inflammatory stages in EAE, Gli1 was subsequently decreased in spinal cord oligodendrocyte precursor cells in the chronic phases. GLI1 expression in MS lesions mirrored this with high levels in acutely active lesions but low levels in chronic active and quiescent lesions.[376] These effects were predominantly downstream of IFNγ. These data demonstrate paradoxical regulation of HH signalling during different phases of demyelination in man and mouse.

1.9.3.3 Chronic GI tract inflammation

Preliminary findings from Nielson and co-workers suggest that SHH may be upregulated in areas of chronic gastrointestinal inflammation.[329] This was shown by immunohistochemistry and *in situ* hybridization in Barrett's oesophagus, gastritis, CD and UC. While not quantitative, their findings show that during chronic inflammation mRNA expression was strong throughout the epithelium from base to lumen, with loss of the crypt-villous pattern normally identified in normal intestinal tissues. They also report increased SHH protein expression, although this is not readily evident in the published photographs. PTCH mRNA and protein was expressed in metaplastic and regenerating epithelial cells and in the crypts, but not in the inflamed colonic surface epithelium. Inflammatory cells in inflamed mucosa stained for SHH and PTCH protein and mRNA. The extension of SHH mRNA from crypt base to include luminal epithelium, may represent either expansion of the stem cell compartment or delayed differentiation.

1.9.3.4 HH, chronic inflammation and cancer

Could HH signalling play a role in cancers arising in areas of chronic inflammation? Beachy and colleagues hypothesize that these cancers represent the continuous operation of an upregulated state of tissue repair associated with chronic activation of pathways such as HH and WNT.[393] This is compatible with the probability that increasing cellular resistance to injury and death creates an epithelium populated by genetically damaged cells. Additionally, the persistence of stem cells or the abnormal distribution of SHH may contribute to the malignant transformation. Clinically relevant scenarios where present data suggest that HH could play such a role include hepatomas, cholangiocarcinomas and pancreatic neoplasms arising on a background of cirrhosis, primary sclerosing cholangitis and chronic pancreatitis respectively. However, as will be shown in this thesis, the situation is rather different in colonic inflammation, and an alternative model will be argued for colonic cancer resulting from chronic active UC.

1.10 *WNT signalling pathway*

The fly *Wingless* (*Wg*) gene, homologous to mammalian WNTs, was originally described by Nusslein-Volhard and Wieschaus in the same seminal segment polarity study that led to the discovery of Hh.[271] The first mammalian gene homologous to Wg, *Integrase-1* (*Int1*), was subsequently identified by Nusse and Varmus in 1982.[394] The term WNT was originally coined by combining *Wingless* and *Integrase*.[395] The WNT signalling pathway consists of a family of small secreted WNT proteins present in all animal genomes (~20 in mammals).

They are involved in virtually all aspects of embryonic development (notably establishment of the dorsal axis), homeostatic self-renewal, and disease pathogenesis (frequently via germline and / or somatic mutation of critical pathway components).[396] Like HH, they are morphogens.[397] 3 different modalities of WNT signalling are described: canonical WNT/β-catenin, non-canonical planar cell polarity (PCP) and WNT/Ca^{2+} signalling. All discussion from here on refers to the former which is detailed in the schematic presented in **Figure 1-10**.

WNT signals activate transcriptional programs via a raft of different WNT/TCF target genes depending on the developmental identity of the responding cell (i.e. cell specific). As such, disparate events – cell proliferation, fate determination or terminal differentiation – can be activated by WNT within the same organ (e.g. intestine or hair follicle). Like HH, WNT target genes include both positive and negative regulators of the pathway (including FZD, LRP, HSPG, AXIN2 and TCF/LEF).

In the GI tract, WNT is the dominant force in controlling cell fate along the crypt-villus axis. WNT proteins are expressed by crypt epithelial cells, rather than in the surrounding mesenchyme.[398] They are pro-mitogenic to crypt progenitors and promote terminal differentiation of Paneth cells.[399] *Tcf4*$^{-/-}$ mice have absent crypt progenitor compartments, whilst maintaining normally differentiated epithelium.[400] Transgenic overexpression of the WNT inhibitor Dkk1 results in complete absence of crypts.[401]

The WNT signalling pathway has recently been shown to have multiple and varied roles in immunology, from self-renewal of haematopoietic stem cells, to T and B cell development, DC maturation, peripheral T-cell activation, and the induction/maintenance of a regulatory immune response.[402] In the thymus, WNTs regulate the survival of both DN and DP T cells, as well as promoting T cell development at various points in between. In the periphery, naïve T cells continue to express LEF1, TCF1 and FZD (in a dynamic fashion reflecting the activation status of the cell), and there is active WNT signalling in mature splenic lymphocytes.[403] Furthermore, T_{REGS} with constitutively active WNT signalling (overexpression of stabilised β-catenin) demonstrate increased survival with the ability to induce an anergic phenotype in CD4+CD25- effector T cells.[404] Finally, WNTs from endothelial cells signal to activated T cells to synthesis and secrete MMPs and aid extravasation into sites of tissue injury / inflammation. Blockade of WNT signalling (by FZD5) decreases T cell migration.

Myeloid DCs and macrophages, but not lymphocytes, stimulated with TLR ligands (LPS and Pam_3CSK_4) or IL4 induced WNT5A expression in a NFκB dependent manner.[405, 406] In response to infection, WNT5A (and FZD5) led to the generation of IFNγ-producing T cells by inhibiting IL12 production.[405]

Defects in:	GENETIC	ENVIRONMENTAL
Intestinal epithelium / barrier function	***MDR1*** ***PTGER4*** ***(DLG5)*** ***MYOIXb***	**NSAIDs** **Smoking**
Bacteria / receptors, innate immunity and autophagy	***NOD2/CARD15*** ***ATG16L1*** ***IRGM*** ***LRRK2*** ***HBD2*** ***NLRP3*** ***ITLN1***	**Antibiotics** **Commensal flora** **Pathogenic bacteria** **Smoking** **Population defect in innate immunity**
Adaptive immunity	***IL23R*** ***IL12B*** ***JAK2*** ***STAT3*** ***HLA*** ***IL10***	**Commensal flora** **Pathogenic bacteria**

Table 1-1 Genetic and Environmental factors contributing to pathogenesis of IBD.

Genotype	PHENOTYPE	
	Lethality	**GI defects**
***Shh*$^{-/-}$** [298, 299, 304]	Peri-natal death	**General:** Body size 30% of WT; small, malrotated GI tract. **Foregut:** EA/TEF with severe lung hypoplasia. **Stomach:** Hyperplastic gastric epithelium, increased number of glucagon-positive cells. **Small intestine:** Decreased smooth muscle thickness; overgrowth of duodenal villi. **Pancreas:** Annular; increased size and endocrine cell number. **Lower GI:** Imperforate anus. **ENS:** Abnormal differentiation of neurons under epithelium, with migration into villi.
***Ihh*$^{-/-}$** [299, 304]		**General:** Body size 67% of WT; small, malrotated GI tract. **Pancreas:** Annular pancreas. **Lower GI:** Dilated segments resemble Hirschprung's. **ENS:** Microscopic absence of neurons, corresponding to dilated segments of SI and colon.
***Shh*$^{-/-}$*/Ihh*$^{-/-}$** [299]	Die: E8.0	
Ipf1/Pdx1-Shh † [407]	Viable	**Pancreas:** Disrupted morphogenesis; pancreatic mesoderm develops into intestinal mesenchyme.
***Ptch1*$^{-/-}$** [304, 408]	Die: E9-10.5	**Pancreas:** Pdx1, glucagon absent from pancreas at E9.5.
***Ptch1*$^{+/-}$** [304]	Viable	**Pancreas:** Males have impaired glucose-tolerance.
***Hhip*$^{-/-}$** [290, 291]	Post-natal death	**Pancreas:** Impaired morphogenesis, islet formation and endocrine cell proliferation (enhanced effect in *HIP*$^{-/-}$*;Ptc*$^{+/-}$ that die before E13); small deformed spleen. **Stomach:** Altered ratio of epithelial-mesenchymal thickness in posterior stomach.
Villin-Hhip [303] ‡	Viable	**Small intestine:** Flattened hyperproliferative epithelium; mislocalized ISEMFs; ectopic precrypt structures
***Gli1*$^{-/-}$**	Viable	No obvious defects phenotypically*
***Gli2*$^{-/-}$** [320]	Peri-natal death	**Foregut:** Hypoplastic esophagus, with no development of smooth muscle and small lumen. **Hindgut:** Imperforate anus with recto-urethral fistula.
***Gli3*$^{-/-}$** [311]		**Hindgut:** Anal stenosis.
***Gli2*$^{+/-}$*/Gli3*$^{+/-}$** [311, 320]	Viable	**Foregut & Hindgut:** No obvious defects.
***Gli2*$^{-/-}$*/Gli3*$^{+/-}$** ***Gli3*$^{-/-}$*/Gli2*$^{+/-}$** [311, 320]	Peri-natal.	**Foregut:** EA/TEF **Hindgut:** Persistent cloaca, less severe than double knockout
***Gli-2*$^{-/-}$*/Gli-3*$^{-/-}$** [311, 320]	Die: E10.5-E13.5	**Foregut:** Endoderm does not develop into oesophagus, trachea and lungs. Small hepatic and pancreatic buds. **Hindgut:** Persistent cloaca, same severity as *Shh*$^{-/-}$

Table 1-2 Summary of Hh signalling component knockout and transgenic phenotypes**.**
Note the similarity between the phenotypes of *Shh*$^{-/-}$, *Gli2*$^{-/-}$ and *Gli3*$^{-/-}$ mice, with the severest defects demonstrated in the *Gli2*/*Gli3* double ko. † In the *Ipf1/Pdx1-Shh* mouse Shh is selectively expressed in developing pancreatic epithelium. ‡ The villin-*Hhip* mouse puts the pan-Hh inhibitor HIP under the control of the intestinal specific villin promoter. In this manner, only intestinal Hh signals are blocked from mid-gestation onwards, allowing viability post-natally. *The Gli1 null mutant is the only mouse with deletion of a major hedgehog family gene that appears to develop normally. However, the published data do not extend to a detailed analysis of the intestines of these mice. Our collaborators at Ann Arbor have pilot data indicating an increased basal level of inflammation in the intestines of these mice as they age (Zacharias, Gummucio unpublished data), but as discussed in chapter 4 it may be only in response to injury / inflammation that the apparent redundancy in Gli1 signalling is lost. EA/TEF: Oesophageal atresia / tracheo-oesophageal fistula. ISEMFs: intestinal subepithelial myofibroblasts. WT: wild-type.

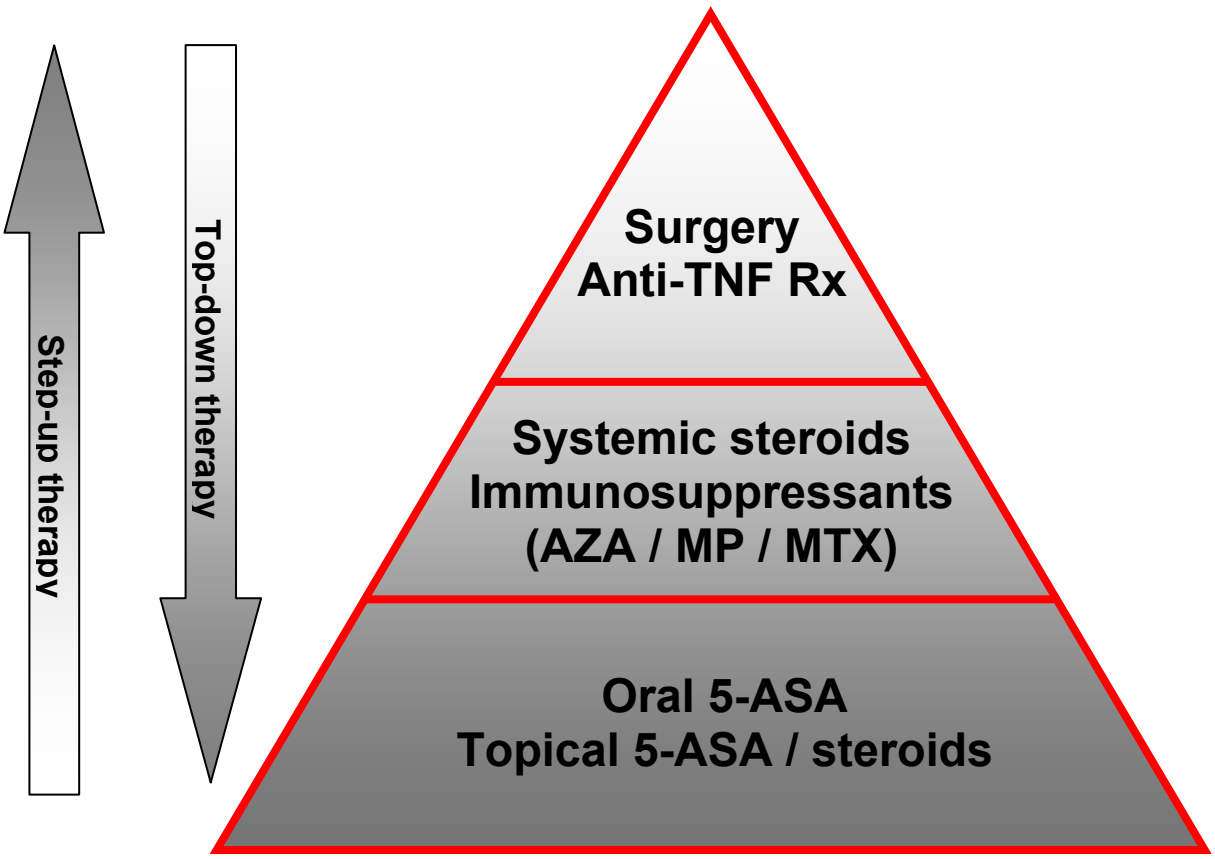

Figure 1-1 Therapeutic 'pyramid' in IBD.

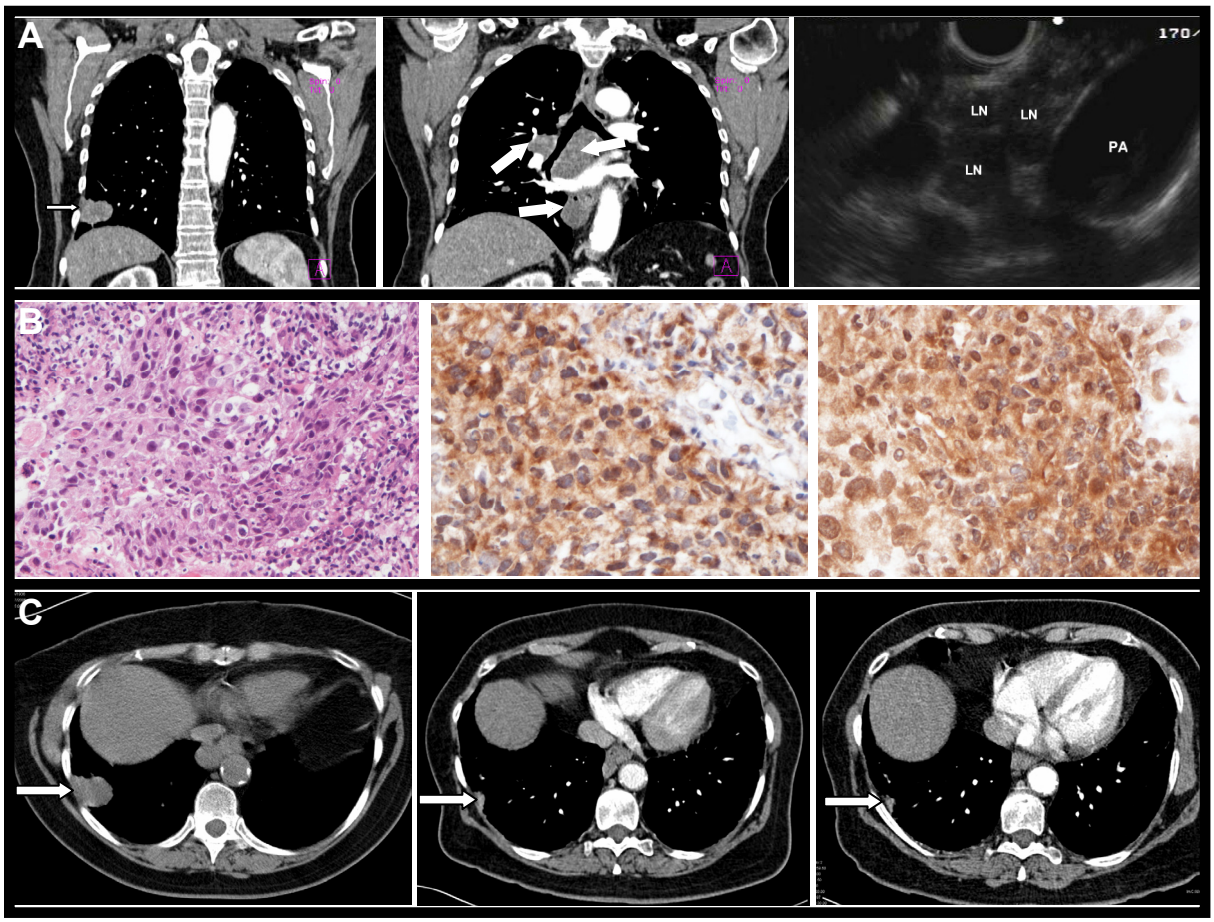

Figure 1-2 Resolution of non-small cell lung cancer on withdrawal of anti-TNF therapy.
A) CT coronal reformat from June 2006 (whilst the patient was on anti-TNF therapy) of the primary lung tumour in the right lower lobe (solid white arrow, left panel). The middle panel shows the mediastinal lymph node enlargement (hollow white arrows) on the same CR scan. Endoscopic ultrasound from July 2006 shows mediastinal lymph node enlargement (right panel). **B)** CT-guided core biopsy of the primary tumour demonstrates classical features of non-small cell lung cancer on haematoxylin and eosin staining (left panel, original magnification x200), and TNFR1 (middle panel) and TNFR2 (right panel) positivity on immunostaining. Endoscopic ultrasound core biopsy of the mediastinal lymph nodes showed necrotic debris consistent with necrotic tumour within malignant lymph nodes (not shown). **C)** Representative CT images are shown before (June 2006, left panel) and after withdrawal of adalimumab therapy (April 2007, middle panel and July 2007, right panel) demonstrating no measurable disease at the site of the primary lung tumour (solid white arrow).

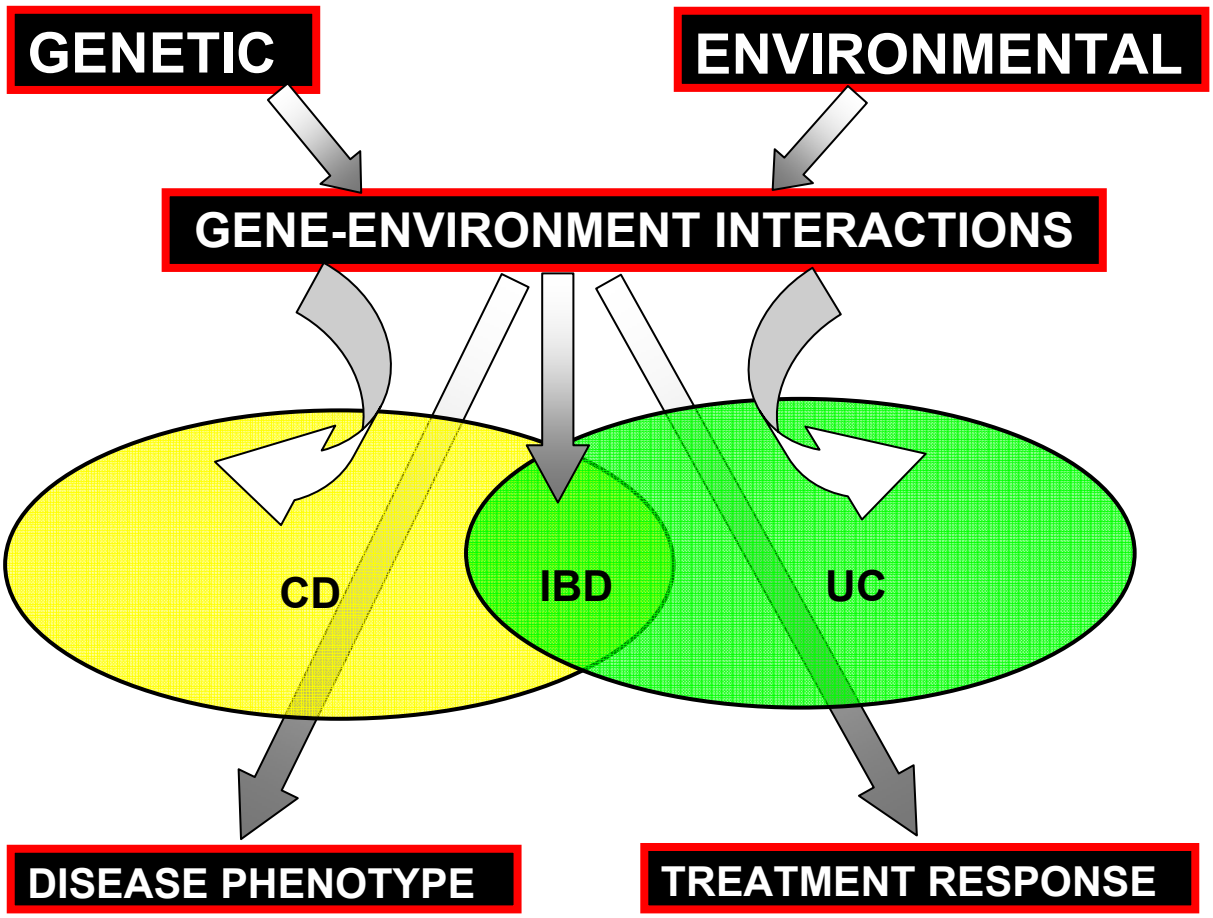

Figure 1-3 Model of IBD pathogenesis.
UC and CD can be thought of as overlapping syndromes sharing some phenotypes, and additionally sharing some genetic and environmental susceptibility and modifying stimuli. However, some phenotypic features and environmental/genetic factors are specific to one or other disorder.

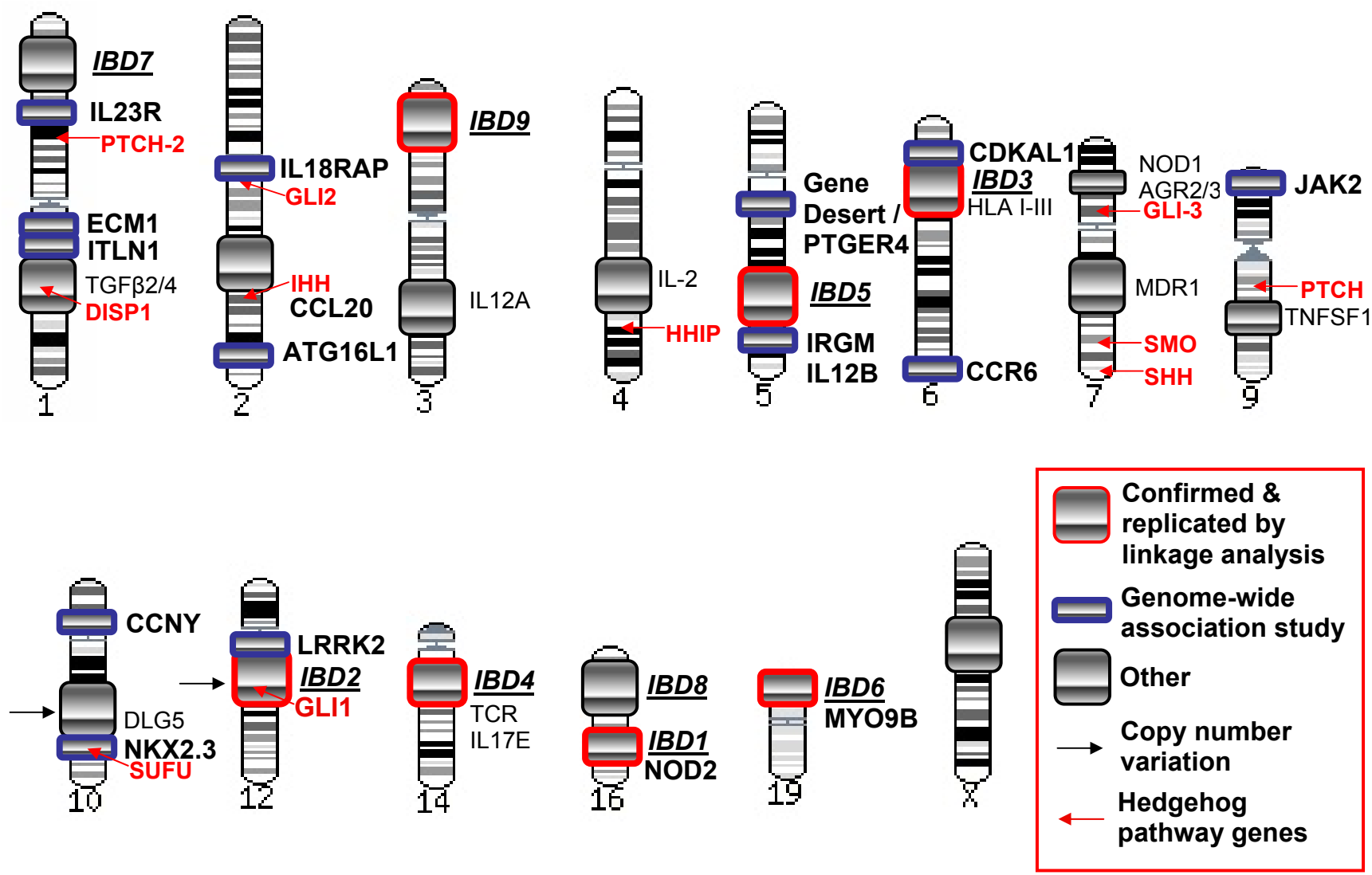

Figure 1-4 IBD Genetics Map.
Susceptibility loci (IBD1-9), confirmed and replicated by linkage analysis, candidate genes (including *GLI1* and *CCL20*) and IBD susceptibility genes / loci identified by GWAS and meta-analysis. Some genes are not included due to space constraints, including *ICOSLG* (21q22.3) and *STAT3* (17q21).

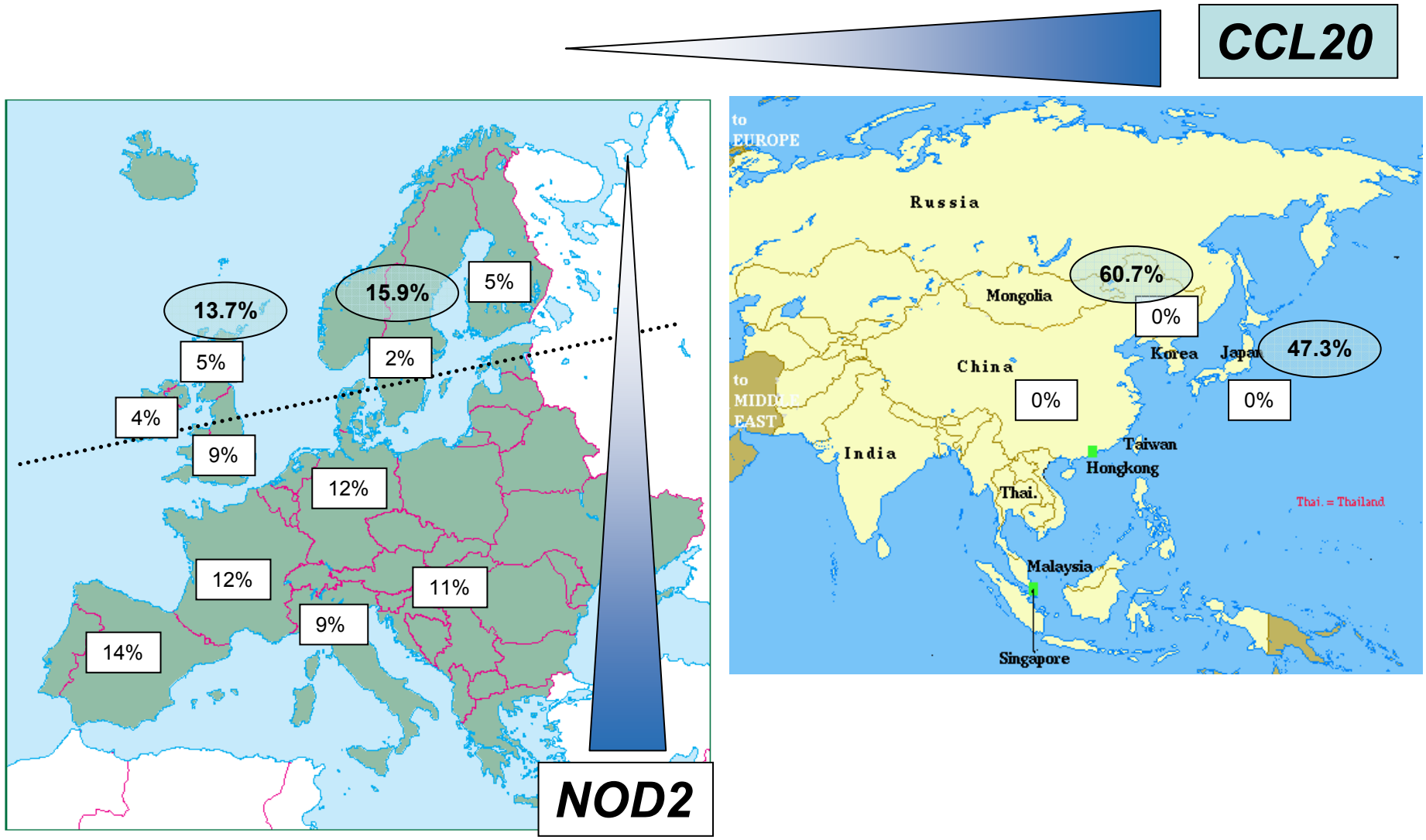

Figure 1-5 Population heterogeneity of IBD susceptibility genes: *NOD2* and *CCL20*.
Allelic frequencies of *NOD2* frameshift mutation (Leu1007fsincC/rs2066847) [white boxes] and *CCL20* promoter polymorphism [blue circles] in Western Europe and the far East. Of particular note is the North – South gradient of increasing frequency of *NOD2* frameshift mutation in Europe, and the absence of all CD-associated *NOD2* mutations in the far East.

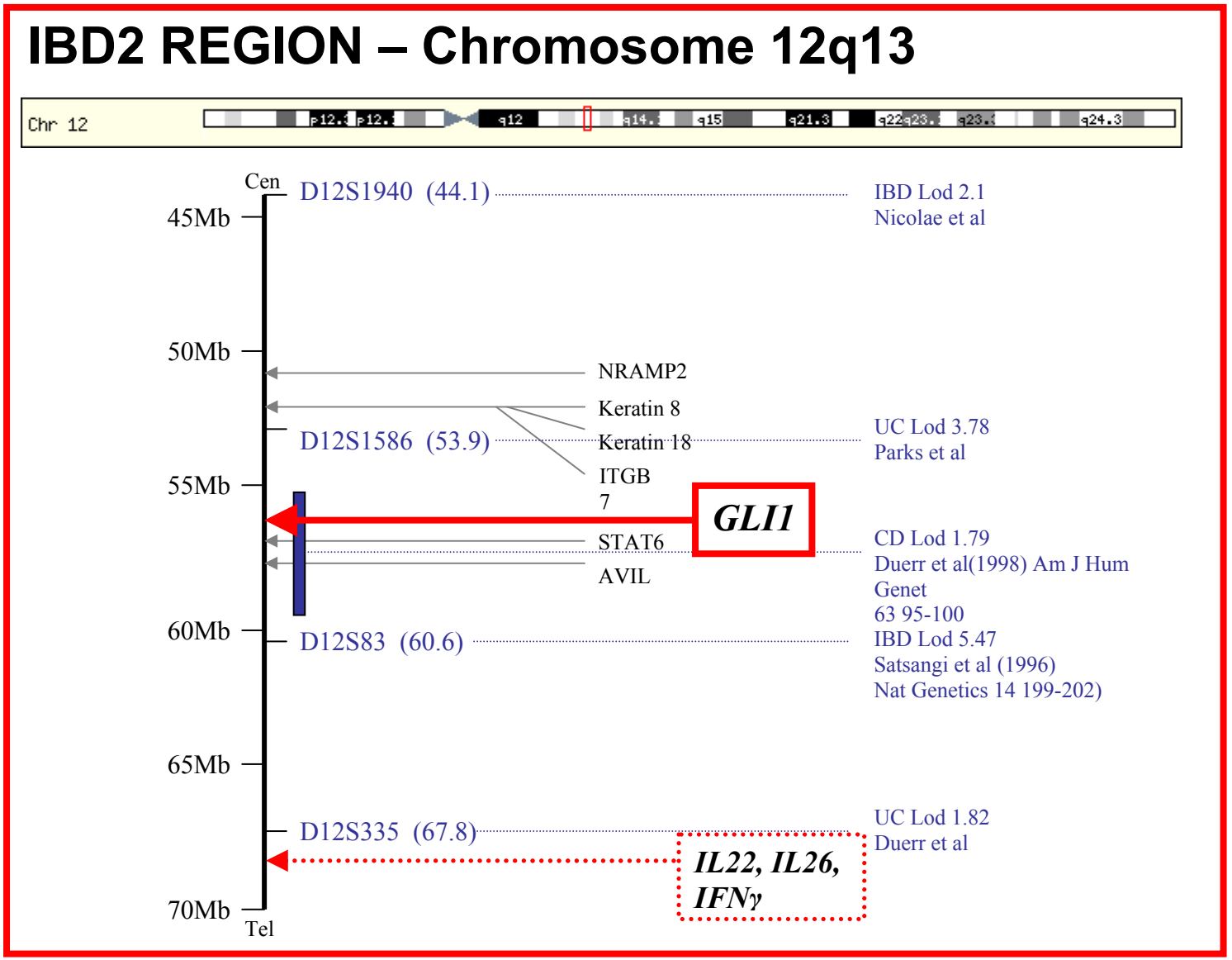

Figure 1-6 The IBD2 locus on chromosome 12q13.
The position of associated micro-satellite markers and various candidate genes, including *GLI1,* are demonstrated on this schematic of IBD2.

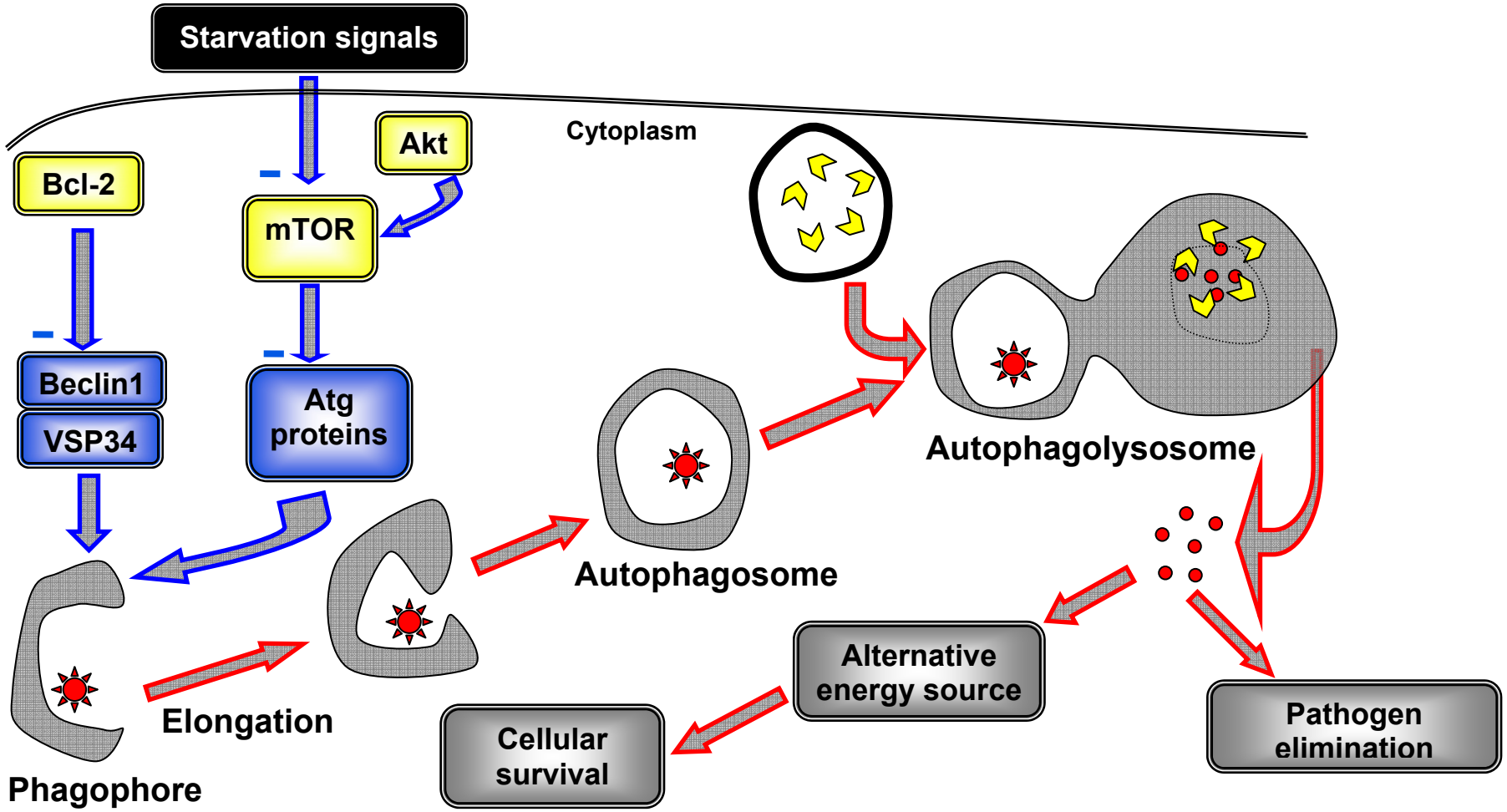

Figure 1-7 Mechanisms of autophagy.
The majority of upstream regulation of autophagy is through the mTOR, and subsequently the autophagy-related proteins (including ATG16L1), the class III PI3K VPS34, and beclin 1. In the steady-state and with a plentiful supply of amino acids, VPS34 activates mTOR with subsequent inhibition of autophagy. Under starvation conditions, mTOR inhibition leads to activation of the autophagy cascade to provide an alternative energy source by recycling intracellular organelles. The autophagy cascade may also be activated by damaged cellular products (e.g. mitochondria) and intracellular pathogens, thus avoiding apoptosis. Intracellular organelles or pathogens are enclosed into an autophagosome. This forms from a double-layered membrane called a phagophore (formed out of various large molecules in the cytoplasm) that elongates to surround the pathogen before closing in on itself. This elongation step is controlled by two protein (and lipid) conjugation systems: the ATG12 and the LC3 pathways. The formed autophagosome then fuses with a lysosome (forming an autophagolysosome). Lysosomal enzymes first cut through the inner membrane of the autophagosome and then break down the enclosed cellular contents or pathogen, finally releasing amino acids back into the cytoplasm for reuse.

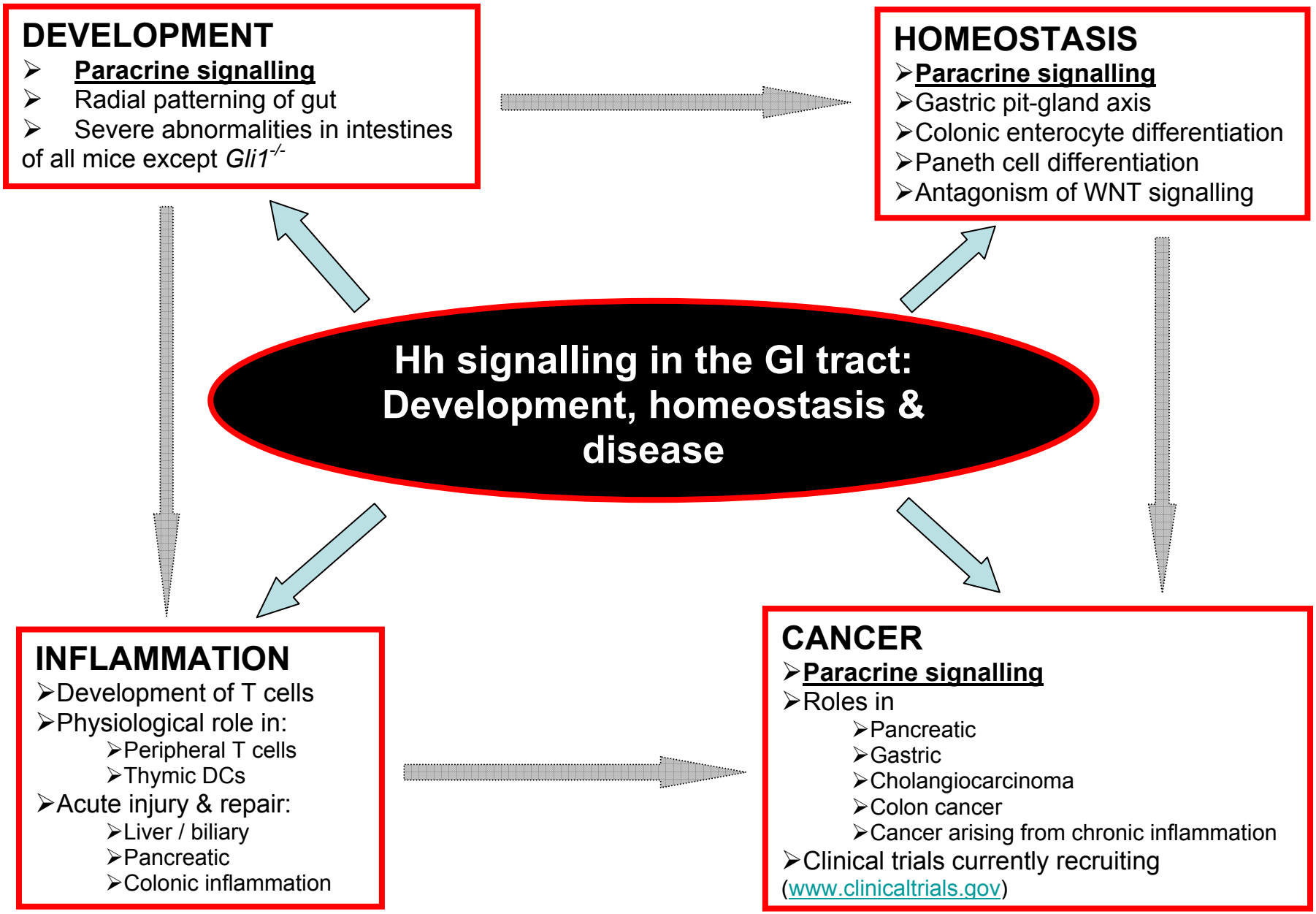

Figure 1-8 The role of the HH signalling pathway in GI tract development, homeostasis and disease

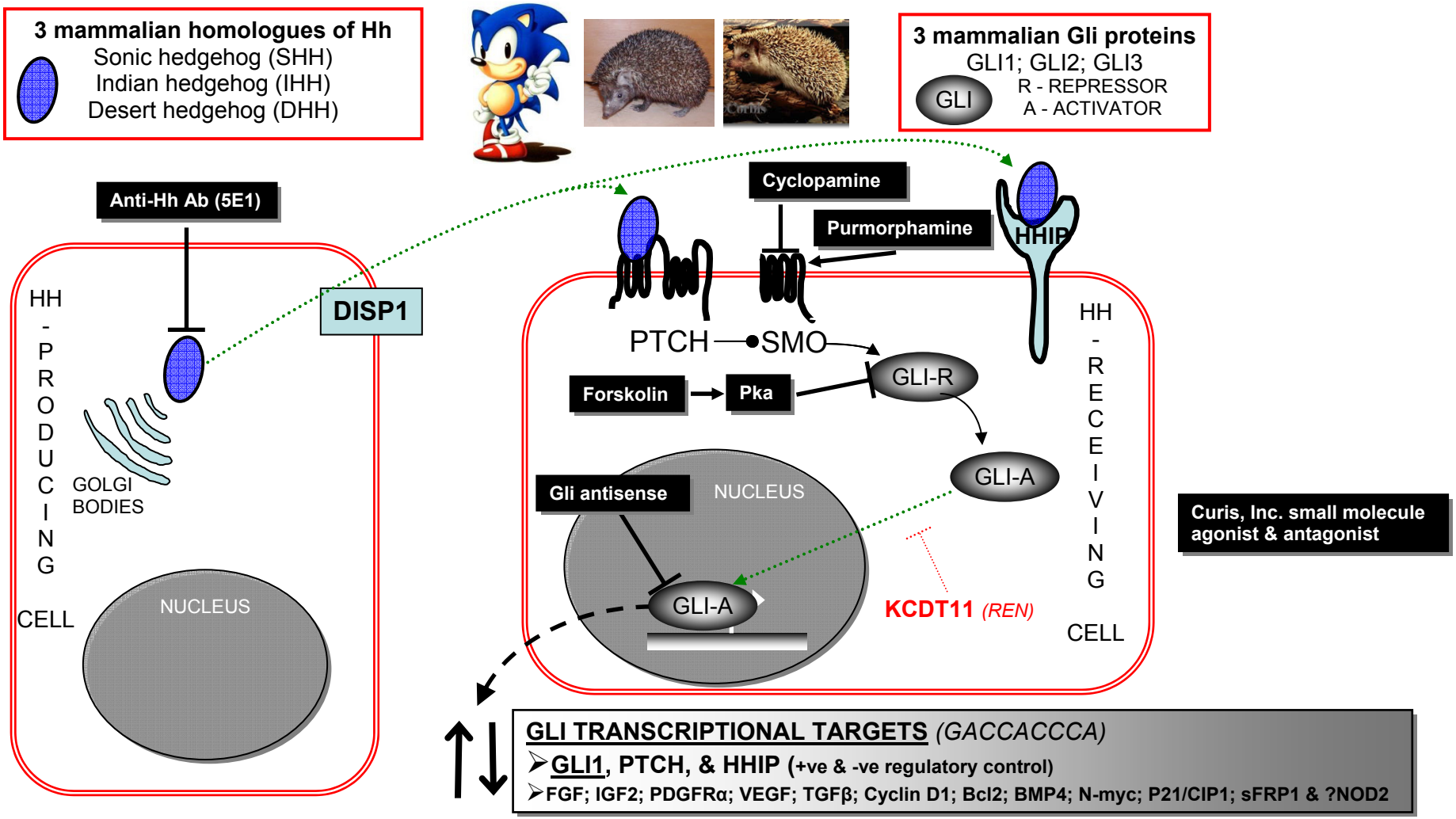

Figure 1-9 Schematic of hedgehog (HH) signalling pathway.
Fully processed HH protein is manufactured in the Golgi apparatus and released via the transmembrane protein dispatched (DISP1) (dotted green arrow). On binding to the HH receptor patched (PTCH), the inhibition of PTCH on another transmembrane protein smoothened (SMO) is released. Subsequent signalling occurs via the 3 zinc-fingered nuclear transcription factors (GLI1, GLI2 and GLI3). GLI2 and GLI3 exist in both repressor and activator forms. When the pathway is activated by HH binding to PTCH, GLI1 shuttles from the cytoplasm to the nucleus (indicated by dotted green arrow), whereupon transcription is activated in a raft of downstream targets. These include positive and negative regulatory pathway components (GLI1, PTCH and hedgehog-interacting protein (HHIP)). HH pathway agonists and antagonists are indicated (white text in black boxes) including the small molecule agonists and antogonists under development by Curis, Inc.

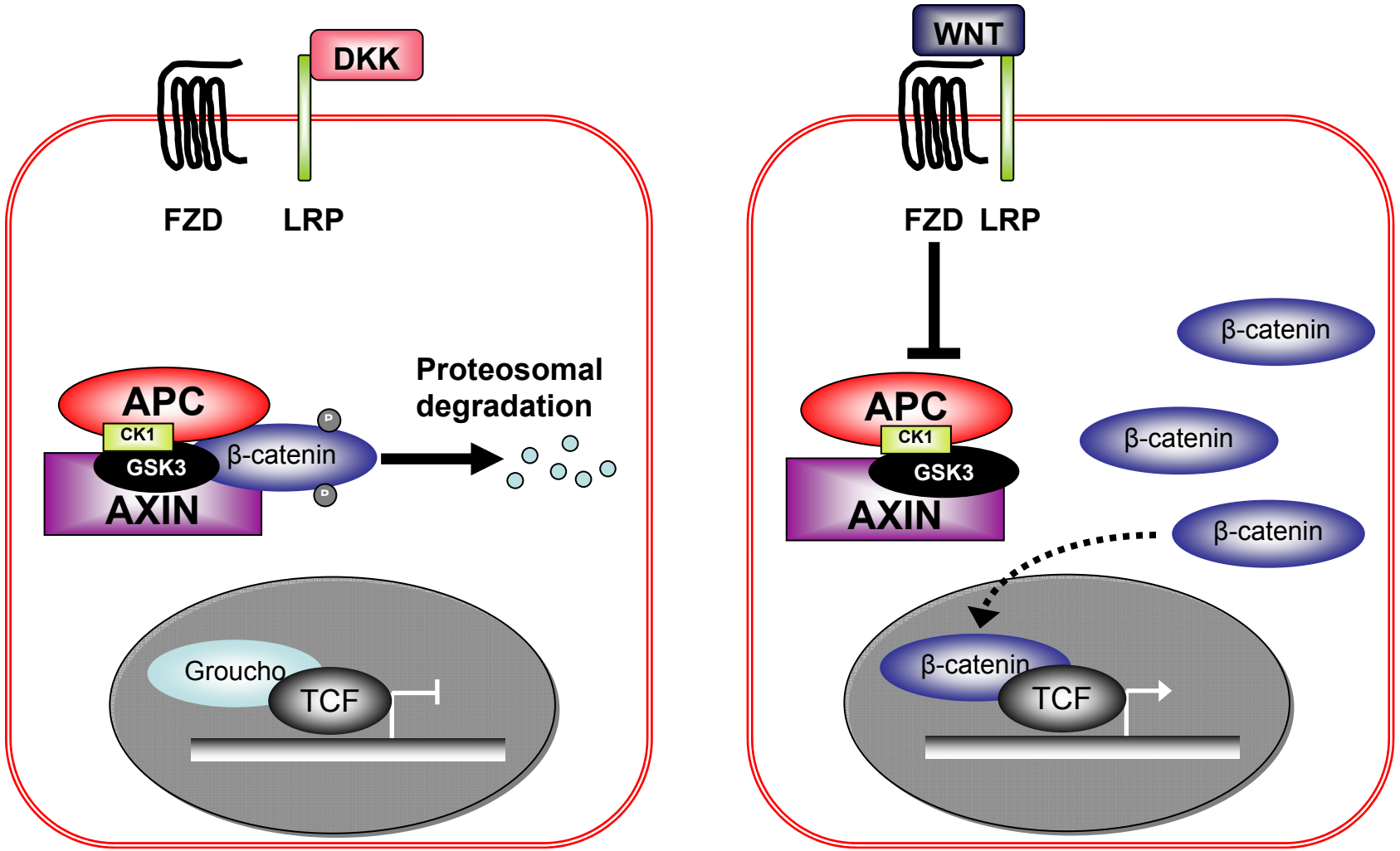

Figure 1-10 Canonical WNT signalling pathway.
In the absence of WNT ligand (**left panel**), β-catenin is targeted for proteosomal degradation via phosphorylation; this process is mediated via a dedicated destruction complex consisting adenomatous polyposis coli (APC), casein kinase I (CKI), glycogen synthase kinase 3 (GSK3) and axin. When WNT binds to the transmembrane receptors frizzled and low-density lipoprotein receptor-related protein (LRP) (**right panel**), the pathway is activated. β-catenin is stabilised, levels rise, it translocates to the nucleus, displaces groucho from the T cell factor (TCF) transcription factors, and initiates transcription of a variety of WNT target genes.

2 Rationale, hypothesis and aims

2.1 *Rationale*

As detailed throughout Chapter 1, multiple strands of evidence support a role for the HH signalling pathway in the pathogenesis of IBD (**Table 2-1**). This pathway is critical to normal gastrointestinal development,[270] as well as enterocyte and Paneth cell differentiation.[318, 332] Chronic injury, inflammation and repair are critical aspects of IBD, and thus it is pertinent that the HH pathway is centrally involved in these processes in several other tissues, including muscle,[374] bone,[409] liver,[388-391] and lung.[349, 367] Indeed, HH signalling may play a central role in the inflammatory response since SHH is critical for T lymphocyte development,[410] adult human CD4+ T cell activation,[365, 366] and myeloid cell maturation in the spleen.[370] Furthermore, SHH has recently been suggested to be a direct transcriptional target of NFκB.[373] Dysregulation of components of the HH pathway has also been noted in inflammatory diseases of the gut, including Barrett's esophagus, chronic gastritis and IBD.[329] Of great interest, there are two GLI consensus motifs upstream of the *NOD2* gene.[411] Furthermore, an alternative splice variant of the NOD2-interacting Centaurin beta-1 (CENTB1) protein[412] is KCTD11 (REN), a negative regulator of the HH pathway that prohibits GLI1 nuclear translocation.[413] In addition, there exists a large Ashkenazi pedigree that has a high degree of overlap between basal cell naevus syndrome (MIM 109400; cause by truncating mutations of the HH receptor, *PTCH*) and IBD,[414, 415] and an isolated case report of a 14 year old Japanese patient with the same syndrome who also developed early-onset UC.[416]

WNT and HH are both secreted morphogens with roles in developmental processes including cell-fate specification and progenitor cell differentiation. Physiological and pathological links between the HH and WNT signalling pathways have been established in the GI tract and other tissues. Both pathways play critical roles in the healthy development of the intestine. In villin-Hhip mice, where Hh is inhibited in the intestine, cytoplasmic and nuclear β-catenin is noted in the intervillus epithelial cells, compared with low levels of membrane associated expression in WT villus tips.[303] WNT/Tcf4 target gene expression is also increased. In colonic epithelial cell differentiation and in colo-rectal cancer pathogenesis, IHH - GLI1 signalling has been shown to antagonise the WNT pathway.[318, 417] A similar phenomenon has been observed in skin, hair follicles and BCC.[418] There is a conserved GLI1 binding site in

sFRP1,[419] and sFRP1 expression in murine fibroblasts in dependent on GLI1 and GLI2.[420] Furthermore, it has been shown in gastric cancer cell lines (SIIA) that the HH antagonist cyclopamine leads to WNT1-mediated β-catenin accumulation in the cytosol.

The WNT signalling intestinal expression signature has been described in full allowing for a detailed, targeted assessment of this pathway in IBD.[398] Recent reports have described reduced WNT/TCF4, a known regulator of Paneth cell differentiation, in ileal CD with a direct and functional correlation to α-defensin levels.[399, 421]

With the advent of GWAS and their successful application now to both CD and UC, a more reliable and detailed understanding of the genetic architecture of IBD is unfolding (see **1.5**). However, the 32 CD susceptibility genes / loci identified in the meta-analysis only accounts for about 20% of the genetic contribution to disease pathogenesis (see **1.5.3**). Therefore, there remains a role for alternative approaches to identifying candidate susceptibility genes in IBD.

2.2 *Hypothesis*

On this background, the initial working hypothesis was that the HH signalling pathway plays a critical role in the pathogenesis of IBD. Furthermore, the genomic location of the *GLI1* gene within IBD2 led to the supplementary hypothesis that inherited variation in *GLI1* is associated with the development of IBD.

2.3 *Specific Aims*

1. To describe the expression of HH signalling components by immunohistochemistry, RT-PCR and microarray analysis in the healthy adult human colon (**Chapter 4**).
2. To determine whether the expression of HH ligands (SHH and IHH) and pathway response elements (GLI1, PTCH, HHIP) are dysregulated in intestinal tissue from IBD patients compared with healthy adult controls (**Chapter 4**).
3. To describe the expression patterns of all WNT pathway genes with documented intestinal expression in healthy intestine and IBD (**Chapter 4**).
4. To analyse the contribution of germ-line variation in *GLI1*, a key transcriptional regulator of HH signalling, to disease susceptibility in a large Scottish cohort of

UC and CD patients, with replication in 2 independent N European populations and 1 from the Far East (**Chapter 5**).

5. To assess the contribution of germ-line variation in *GLI1*, dysregulated in colon cancer, to the inherited susceptibility of early-onset colo-rectal cancer (**Chapter 6**).
6. To determine whether germ-line variation in *IHH*, down-regulated in UC (irrespective of inflammation), is associated with IBD pathogenesis (**Chapter 7**).
7. To analyse WTCCC GWAS and meta-analysis data for association of all HH and WNT pathway genes with CD (**Chapter 8**).
8. To analyse the effect of PAMPs (LPS and MDP) on HH pathway activity in colonic epithelial cells *in vitro* (**Chapter 9**).
9. To assess the effect of HH agonists / antagonists on NFκB signalling and cytokine synthesis and secretion in colonic epithelial cells and PBMCs from patients with acute severe UC *in vitro* (**Chapter 9**).
10. To assess the contribution of promoter polymorphisms in *CCL20*, a key pro-inflammatory cytokine that is downstream of HH (demonstrated in **9**), to disease susceptibility in N European and Far Eastern populations with UC (**Chapter 10**).
11. To further assess the contribution of genetic variation in *NKX2.3* to adult and early-onset IBD (**Chapter 11**). This homeobox gene, critical to normal GI tract development, is a target of HH and was identified as a CD susceptibility gene in the WTCCC GWAS.

Hh pathway is critical to normal gut development [299, 303]
Present in normal adult GI tract with gradients mirroring increasing bacterial burden and clinical distribution of UC (Chapter 4)
Role in enterocyte and Paneth cell differentiation[318, 332]
Recapitulation of embryonic signalling in response to injury: – skeletal muscle ischaemia[374] – liver[388-391] – acute pulmonary epithelial injury[349] – prostate epithelial regeneration[358]
Essential mediator of progenitor T cell differentiation[410] and myeloid cell maturation in the spleen.[370] Physiological role in human peripheral CD4+ T lymphocytes, increasing the proliferative index and altering cytokine profiles.[365, 366, 368, 422] FOXP3 is a predicted HH-GLI1 target gene.[411]
IFNγ activates SHH expression in the brain[369] SHH is a direct transcriptional target of NFκB,[373] and is downstream of NFκB in epidermal keratinocytes and the pancreas[371, 372]
Two GLI1 consensus sequences identified upstream of *NOD2* gene. NOD2 expression levels in healthy colon correlate closely with HH pathway activity (GLI1 expression) (Chapter 4).
Large Ashkenazi pedigree that has a high degree of overlap between basal cell naevus syndrome (MIM 109400; caused by truncating mutations of the HH receptor, *PTCH*) and IBD;[414, 415] case reports of mutated HH signalling components and UC [416]
GLI1 is located in the IBD2 locus (OMIM 601458; spanning 25 Mb on chromosome 12q13),[98, 99] most strongly implicated in UC.[97, 98]
HH response elements (GLI1, PTCH and HHIP) are decreased in UC (Chapter 4)
***GLI1* variation is associated with UC, specifically a transactivationally defective non-synonymous SNP in a highly conserved domain of exon 12 near a known activation domain (Chapter 5)**
***Gli1*$^{+/LacZ}$ mice develop early, severe colitis with high mortality when stressed with DSS. Lamina propria myeloid cells are direct targets of HH in reponse to this acute inflammatory challenge (Lees, Zacharias *et al*, PLoS Med 2008).[423]**

Table 2-1 Summary of evidence implicating HH signalling in IBD.
The evidence compiled from published literature and data generated within this thesis (**bold text**) is summarised.

3 Materials and Methods

3.1 *Definitions*

The diagnosis of IBD, and specifically UC or CD, was made on the basis of Lennard-Jones criteria.[424] This seminal classification originated from St Mark's Hospital, London in 1989 and remains in widespread use both clinically and for research purposes. Essentially, the system first details alternative causes for bowel inflammation (including infection, ischaemia and trauma) that must be excluded prior to establishing a positive diagnosis of IBD. The diagnosis of UC or CD is then made on the basis of macroscopic (from clinical and endoscopic features, imaging and inspection of surgical resection specimens) and microscopic (analsysed by an expert histopathologist) changes. In his original classification, Lennard-Jones did not allow for indeterminate IBD to be diagnosed. Later systems designed to classify IBD phenotype primarily for research purposes (e.g. Vienna and Montreal)[425, 426] allow for this. The Montreal classification system, detailed below (**3.5.2**), is used for phenotyping throughout the clinical studies presented here.[426]

3.2 *Ethical approval*

Written, informed consent was obtained from all patients and controls. The study protocol was approved by the Lothian Research and Ethics Committee (LREC-2000/4/192 and 2004/S1103/22); Cambridge (LREC-01/418; MREC-03/5/012) and Regional Ethics Committee, Karolinska Institutet.

3.3 *Patients for expression studies*

3.3.1 Patients for immunohistochemistry

Suitable cases were identified retrospectively (2000-2005) from the archives of the Department of Pathology, WGH, Edinburgh. Samples were taken at endoscopy and surgery for CD (n=22) and UC (n=17), with colonic location clearly established by the endoscopist. Where accurate location could not be established, the sample was excluded. Healthy controls (HC) (n=16) were taken from patients with normal histology following colonoscopy for investigation of diarrhoea (full colonoscopic series, n=6), and following surgery for non-inflammatory, non-ischaemic obstruction (n=4). Patients were initially consented retrospectively for permission to use historical pathology specimens for research. Subsequently a new system was

developed, such that all patients recruited onto the IBD genetics database were all asked if they would consent for use of historical and future pathology specimens. After selection of cases, samples numbers were provided to the pathology department who retrieved wax blocks and cut samples 'in-house' so as to retain wax blocks in the central pathology stores (located at NRIE, Edinburgh) for future use. All histological slides were reviewed by a consultant gastrointestinal pathologist (J Loane), and used for immunohistochemistry as early as possible after cutting (maximum 3 months) to avoid excess degradation of protein epitopes once exposed to air.

3.3.2 Patients for microarray

The cohort of patients used in the microarray studies consisted of 67 patients with UC, 53 with CD and 31 HC all undergoing colonoscopy ± ileoscopy at WGH, Edinburgh (**Table 3-1**). Mucosal biopsies from specified anatomical locations (terminal ileum, caecum, ascending colon, transverse colon, descending colon, sigmoid colon and rectum; location determined by operator experience, distance from insertion and colonoscope configuration on a Scope Guide^{TM}) were immediately flash frozen in liquid nitrogen before storage at -80°C. Inflammatory status was graded (inflamed or non-inflamed) on paired biopsies, formalin-fixed and stained with haematoxylin and eosin for histological analysis of lamina propria inflammatory cell infiltrate, by an experienced gastrointestinal pathologist.

For UC patients, 8 were recruited at diagnosis (treatment naïve), 18 had active disease and 41 quiescent disease; 10 were on systemic corticosteroid therapy, 11 on immunosuppressants (AZA, MP or MTX) and 40 on 5-ASA therapy (**Table 3-1**). 4/53 patients with CD were on corticosteroids, 13/53 on immunosuppressants and 21/53 on a 5-ASA. 8/31 HC had non-IBD inflammation: 2 scattered lymphoid aggregates with history of gastroenteritis, 2 microscopic colitis, 1 pseudomembranous colitis, 1 diverticulitis, 1 amoebiasis and 1 eosinophilic infiltrate.

3.4 *Expression studies on human intestinal tissue*

3.4.1 General immunohistochemistry protocol

3μm sections from formalin-fixed, paraffin-embedded samples were cut onto Superfrost Plus slides (BDH laboratory supplies), dewaxed in xylene and rehydrated in graded alcohols. Antigen retrieval was performed in unmasking solution (H-3300,

Vector Labs, Burlingame, CA) at 1000W for 15 minutes, and sections were cooled in running tap water for 20 minutes. To enhance antibody-binding specificity and decrease non-specific background staining several blocking steps were employed. Endogenous peroxidase activity was blocked with 2% hydrogen peroxide (H_20_2) for 15 minutes, endogenous avidin and biotin was blocked for 10 minutes (Avidin/Biotin Blocking Kit, SP-2001, Vector Labs), and non-specific antibody-binding was blocked with a commercially available serum-free protein block for 10 minutes (X0909, DakoCytomation, Denmark). All antibodies were diluted in DAKO diluent. Following H_20_2 treatment, slides were loaded into Sequenza™ (Thermo Shandon) immuno-staining chambers. In between all steps, except following application of protein block, slides were washed twice with PBS. Sections were incubated with goat polyclonal primary antibodies overnight at 4°C, and mouse monoclonal antibodies for 30-60 minutes at room temperature (RT). Prior to addition of secondary antibody, slides incubated at 4°C overnight were allowed to come to RT on the laboratory bench. Biotinylated secondary antibodies were applied at RT for 30 minutes, followed by avidin-biotin complex for a further 30 minutes (Vector R.T.U. Vectastain Elite ABC Reagent, PK-7100; Vectastain ABC Kit Alkaline Phosphatase Standard, AK-5000). Substrate was visualised with diaminobenzidine (Liquid DAB+, DAKO), developed for 5 minutes to reveal a brown precipitate under light microscopy. Finally, samples were washed in PBS, transferred to a staining rack, counterstained with Harris' haematoxylin (15 seconds), 'blued' in tap water, dehydrated through serial alcohols and xylene (x2) before mounting coverslips with DPX glue (Sigma). Negative controls were included for all samples with omission of primary antibody.

3.4.2 Dual immunofluorescence and con-focal microscopy

For double immunofluorescent staining, antigen retrieval was performed as above, and 2% H_20_2 and protein block were applied. Both primary antibodies (diluted in PBS) were applied on their own with both secondary antibodies, in first and second positions sequentially, and with PBS as a negative control. All washing steps were in Tris-buffered saline (TBS). Secondary antibodies were applied for 2 hours at RT. TO-PRO-3 (Molecular Probes, Invitrogen) was applied for 10 minutes as a nuclear counter-stain, slides were washed in de-ionised water and mounted in Mowiol medium (gift from Anne Grant, MRC CIR, University of Edinburgh). Digital images

were captured on a Leica confocal laser scanning microscope with assistance from Linda Sharp (IMPACT facility, University of Edinburgh).

3.4.3 Primary and secondary antibodies for immunohistochemistry

All antibodies were optimised prior to use by running serial dilutions on slides from tissues with known expression. SHH and PTCH were optimised on rat lung, IHH on human kidney, GLI1-3 and CD56 on human glioblastoma, and immune stains on murine spleen and human tonsil. Primary antibodies used were anti-SHH (N-19; 1:30 *see note below), anti-IHH (C-15; 1:25), anti-PTCH (C-20; 1:30), anti-HHIP (M-17; 1:400) (all goat polyclonal antibodies from Santa Cruz Biotechnology, California, USA), anti-GLI1 (ab7523-50; 1:400), anti-GLI2 (ab7181; 1:1000), anti-GLI3 (ab6050; 1:1000) (all rabbit polyclonal antibodies from Abcam, Cambridge, UK); anti-CD3 (clone F7.2.38; 1:20; DAKO), anti-CD4 (clone 4B12; Vector; VP-C319), anti-CD8, anti-CD20 (M755; 1:50; DAKO), anti-CD56 (MCA1814T; Serotec), anti-CD68 (clone PG-M1, 1:100; DAKO), anti-p65 (SC370; Santa-Cruz), and anti-chromogranin A (clone DAK-A3, DAKO; 1:200). Secondary antibodies used were rabbit anti-goat (1:400), rabbit anti-mouse (1:200), and donkey anti-rabbit (1:300) (all DAKO) for DAB, and donkey anti-goat (588) Alexa Fluor and rabbit anti-mouse (488) Alexa Fluor (both Molecular Probes) for immunofluoresence.

3.4.3.1 SHH immunohistochemistry

Using the N-19 antibody at a dilution of 1:250 SHH was not detectable in the caecum, ascending colon, transverse or descending colon, whilst maintaining a strong signal in the sigmoid colon and rectum (data not shown). The precise anatomical locations studied may explain the differences in this data compared with the original published study, which did not specify colonic location of samples.[327, 427] Utilising various commercial blocking agents described above, N-19 was used at a 1:30 titre, detecting low levels of expression, without non-specific background staining. At this concentration, SHH protein was detectable in the proximal colon, albeit with a much lower signal compared with the rectum within the same individual colonoscopic series (**Figure 4-3**). SHH expression was blocked following pre-incubation (overnight at 4°C) with 5x blocking peptide (sc-1194P); no signal was detected with omission of the primary antibody (**Figure 3-1**). However the close homology between SHH and

IHH amino acid sequences means that the probability that N-19 is also detecting IHH protein cannot be excluded.

3.4.4 RT-PCR

Q-PCR on human intestinal tissue (microarray cohort) was performed by Jenine Cornelius (Genentech, Inc.) in close collaboration, after unsuccessful attempts to complete a Materials Transfer Agreement to transfer the cDNA back to Edinburgh for further analysis. One RNA amplification cycle was carried out using the MessageAmp™ II aRNA Amplification Kit protocol (Ambion technologies, Austin, TX). The amplified cRNA was purified using the RNeasy Mini Kit protocol (Qiagen) and 1 µl of amplified cRNA was quantified using the NanoDrop ND-1000 Spectrophotometer. RT-PCR was then performed on 50ng of RNA using Stratagene model MX4000 (La Jolla, Ca, USA). The primers and probes were designed in conjunction with JC using Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) and are as follows: SHH forward cgcagctgctctaccaaata; SHH reverse ggcttcagctggacttgac; SHH probe tcctggacagcgaggccctg; IHH forward cttcagcgatgtgctcattt; IHH reverse ctgagtctcgatgacctgga; IHH probe tactggaccgcgagccccac.

3.4.5 Microarray methodology

The full details of the microarray methods are described in Noble, Cornelius, Abbas, Lees *et al*, GUT 2008.[428] Briefly, 750ng of the test sample labelled with Cy-5 and 750ng of Agilent universal human reference labelled with Cy-3 were fragmented for 30 minutes at 60°C and then loaded onto Agilent whole human genome oligo microarray chips G4112A (Agilent technologies, Palo Alto, CA). The samples were hybridized for 18 hours at 60°C with constant rotation. Slides were washed and dried using the Agilent stabilization and drying solution protocol (Agilent technologies), and scanned using the Agilent G2505B model (Agilent Technologies). Expression signals were calculated using the Agilent feature extraction software (version 7.5, Agilent Technologies). The distribution of log intensities for each sample were plotted and outlying samples (greater than 2 standard deviations from the mean) were excluded from analysis.

3.4.6 Microarray data analysis

Raw microarray data was provided by Alex Abbas (Genentech, Inc.) as 3 Excel spreedsheets (UC, CD and HC). Each Agilent probe was listed with Agilent universal reference value given for every sample. The Agilent universal reference describes the value of expression for each sample relative to an arbitrary reference value set by Agilent. As a crude measure, values of <1.0 indicate low levels of expression; however, this is poorly characterised for many of the probes in the genechip dataset. Importantly, it provides a standard reference point for each probe by which expression changes can be measured. Agilent probe IDs for genes of interest were selected from Ensembl (www.ensembl.org). The raw microarray data for each probe of interest was then copied into a separate Excel spreadsheet and transformed into a workable format, for transfer into GraphPad Prism 4.04 (GraphPad Software Inc., San Diego, USA).

Expression profiles were first plotted for all probes in non-inflamed (N-I) HC samples to establish the presence or absence of differential expression along the length of the healthy human colon. Where no significant gradient (Kruskall-Wallis test) was evident the whole dataset was used for analysis of UC vs. HC and CD vs. HC (Mann-Whitney U test). Where expression gradients were present, this analysis was performed in samples from the sigmoid colon only (standardised for all datasets; most frequent sampling point). For CD, analysis was also performed in terminal ileal tissue. All IBD analyses were compared against N-I HC samples from the same location. Finally, comparison was made between I and N-I HC samples from the same sampling point. All datasets were plotted against the Agilent universal reference as individual datapoints with medians represented by horizontal lines.

3.5 *Patients for genetics studies*

3.5.1 DNA extraction and storage

DNA was extracted from whole blood (collected in 10ml EDTA tubes) 'in-house' by a modified salting-out technique.[429] 40mls of red cell lysis buffer (RCLB) was added to 10mls whole blood in a 50ml conical tube and re-suspended for 5 minutes. After centrifuging for 10 minutes at 3000rpm the supernatant was removed. The resulting pellet was re-suspended in 40mls RCLB and centrifuged for a futher 5 minutes. The pellet was resuspended in 3mls nuclear lysis buffer and sodium dodecyl sulphate, plus 1ml of 6M NaCl and 3mls chloroform before thorough mixing and centrifuging for 20

minutes. The middle layer of the resulting solution was removed by pipette and added to 20mls of 100% ethanol. The precipitated DNA was dried at RT for 5 minutes, re-suspended in 0.5mls TE at 4°C. The DNA was quantified, the stock stored at -80°C and working solutions stored at 100ng/ml at 4°C.

Recently, the DNA extraction procedure has been out-sourced to the MRC Human Genetics Unit. Storage of stock has also been transferred to the Wellcome Trust Clinical Research Facility (WTCRF) Genetics Core. At WTCRF, DNA is quantified by nanodrop technology and plated out onto 384 well stock plates, stored at 4°C for robotic transfer to working 384 well plates for genotyping. These plates were previously ordered by disease (i.e. HC1-5, CD1-7, UC1-8), but after careful consideration were fully randomised in 2007 to avoid significant biais from single plate errors and to ensure WTCRF personal are fully blinded to sample identification.

3.5.2 Disease phenotyping

Phenotypic data were collected by patient questionnaire, interview, and case-note review. Age at diagnosis, disease behaviour and disease location were classified initially by Vienna[425] and since 2005 by Montreal criteria (both classifications are stored on the database for all patients).[426] Data consisted of demographics, date of onset and diagnosis, disease location, disease behaviour and severity, progression of disease type, extraintestinal manifestations, surgery, smoking history and family history.

In CD, disease location is defined as terminal ileal (L1), colonic (L2), ileo-colonic (L3) or upper GI (L4). L4 is also applied as a modifier to L1, L2 and L3 when upper GI disease is also present. Disease behaviour is defined as purely inflammatory (B1), stricturing (B2), or penetrating (B3), with perianal disease added as a modifier (p). In UC, disease extent is defined by endoscopic appearance as the maximum extent during follow-up and is categorised as ulcerative proctitis (E1; disease limited to the rectum), left-sided (E2; involvement limited to the colorectum distal to the splenic flexure), or extensive colitis (E3; involvement extends proximal to the splenic flexure). Disease severity is defined as patients suffering at least one acute severe attack of UC, satisfying Truelove and Witts criteria, and requiring intensive in-patient

medical therapy. Colectomy is surgery for active UC not adequately responding to medical therapy, and does not include surgery for dysplasia or neoplasia.

3.5.3 Databases

The Scottish adult and early-onset cohorts are both stored as password-protected files in separate, purpose-designed Microsoft Access databases kept on a single non-networked PC. Data is backed up once a week onto CD-ROM and stored in a locked, fire-proof case. The databases are presently being transferred onto the secure University of Edinburgh network and stored as encrypted files for additional security of personal data. They are managed by a dedicated database manager (H Drummond).

3.5.4 Scottish adult-onset IBD population and healthy controls

For the Scottish adult IBD database patients were recruited from the Western General Hospital (Edinburgh) by clinical research fellows and research nurses within the GI unit. HCs consisted of volunteers from hospital staff and blood donors from the Lothian region. They were not extensively screened for latent disease. Detailed demographics and phenotype are presented in Chapters 5, 7, 10 and 11.

3.5.5 Scottish early-onset IBD population

The Scottish early-onset cohort recruited patients with IBD diagnosed at <16 years of age from the specialist paediatric hospitals in Edinburgh, Glasgow and Aberdeen (recruitment primarily by R Russell and J van Limbergen). Detailed demographics and phenotype are presented in Chapters 10 and 11.

3.5.6 English IBD population and healthy controls

The study of *GLI1* in the English IBD population described in Chapter 5 was performed in collaboration with Miles Parkes and Mark Tremelling (IBD Research Group, Addenbrooke's Hospital, University of Cambridge). Genotyping was done in Cambridge, with raw data provided for analysis. This cohort is from East Anglia with HCs recruited from the EPIC study. Full details presented in Chapter 5.

3.5.7 Swedish IBD population and healthy controls

The Swedish IBD and HC population was provided by Leif Torkvist and Robert Lofberg (Karolinska Institutet, Stockholm, Sweden). Extracted DNA was shipped to Edinburgh for storage. All genotyping was performed as for Scottish samples at the WTCRF. Full details are provided in Chapters 5 and 10.

3.5.8 Japanese IBD population and healthy controls

The Japanese population of IBD patients and HC was provided by Nobuhide Oshitani (Department of Gastroenterology, Osaka City University Graduate School of Medicine, Osaka, Japan), Nagamu Inoue and Toshifumi Hibi (Department of Internal Medicine, Keio University, Tokyo, Japan). DNA was shipped to Edinburgh for genotyping at the WTCRF. Details are provided in Chapter 10 (limited phenotypic data available on this cohort).

3.5.9 Scottish colo-rectal cancer population and healthy controls

The Scottish colo-rectal cancer population and HC were provided by Malcolm Dunlop (MRC Human Genetics Unit, University of Edinburgh). DNA from this cohort is stored in the WTCRF. Full details are provided in Chapter 6. The HCs from this population were also used for aspects of the IBD *GLI1* study in Chapter 5.

3.6 *Methods for genetic studies*

3.6.1 Polymerase chain reaction (PCR)

All primer pairs (Sigma, **Table 3-2**) for PCR, shipped in lyophilised form, were first reconstituted in TE to make 100μM solutions. Optimisation PCRs were run using a gradient (60°CΔ10) to find optimal annealing temperatures. The standard annealing time was 30s, increased to 60s for products of ~1kb. The polymerase amplification was performed in a total volume of 20μL containing 10× PCR buffer (160mM $(NH_4)_2SO_4$, 670mM Tris-HCl [pH 8.8], 0.1% Tween-20 (Sigma)) 50mM $MgCl_2$ (Sigma), 1μM dNTP (Bioline) and 5U/μl Biotaq DNA polymerase (gift from Elaine Nimmo.) The PCR program using a Techne PCR machine was as follows: 35 cycles of 30s at 92ºC, 1-3min at 55-65ºC, 45s at 72ºC, followed by a 5 min 72ºC extension and then a 4ºC hold. These samples were then electrophoresed (150V, 400mA, 30 min) on a 1% agarose gel (agarose multi-purpose, Bioline; 2.25g to 150ml half-strength TBE, microwaved for 2-3m and cooled to 50°C) containing 1μg/ml of ethidium bromide (Sigma) for visualisation of the bands. Orange G dye (Sigma) and ethidium bromide in a ratio of 25:1 were added to the samples before being pipetted onto the gel, adjacent a 100 base pair marker lane (Bioline). Visualisation of bands was achieved in a high performance UV transilluminator with a Kodak Edas 290 Digital Zoom camera and a Kodak 1D-Scientific Imaging Systems computer package.

3.6.2 DNA sequencing and SNP identification

Direct sequencing was performed by at the MRC Human Genetics Unit (University of Edinburgh) on a 7900HT sequence detection system. Sequence was analysed with Sequencher v4.5 (Gene Codes Coporation, Ann Arbor, MI, USA. SNP frequencies were scored on a purpose-designed spreadsheet.

3.6.3 Taqman methodology

The Taqman method is a PCR-based assay. Purpose-designed Taqman probes contain fluorescent dyes that are excited by a built-in laser. The reporter dye fluoresces when a quencher is removed from a specific DNA fragment during PCR. This allows the detection of the increasing amount of PCR product to be quantified in real-time. This is done cycle-by-cycle, giving a cycling threshold (Ct) required to detect the product, of relevance in expression studies. For genotyping studies, allele specific primers are designed such that one allele of the SNP is assigned to VIC and the other to FAM. The primers and corresponding probes for Taqman genotyping are ordered from ABIs Assay-on-Demand and Assay-by-Design services. Taqman genotyping was performed at the WTCRF on ABI PRISM 7900HT machine with the following thermal cycling conditions: 50°C for 2minutes, 95°C for 10 minutes, then 40-cycle program of alternating 95°C for 15 seconds and 60°C for 1 minute.

For each Taqman run, the allelic discrimination assay is visualised by plotting FAM and on opposing axes, given clusters for each genotype (AA, AB, BB) (**Figure 3-2**). At WTCRF alleles are called by two personnel blinded to the study's aims. Where a genotype cannot be called the sample is dropped from analysis. Taqman genotyping accuracy was assessed by duplicate analysis for all SNPs (5-10% of samples) and direct sequencing of 10% of all samples (for *GLI1* and *CCL20*).

3.7 *Statistics for genetic studies*

Individual SNP analysis was performed on GraphPad InStat 3 statistical package (GraphPad Software Inc., San Diego, USA), using chi^2 or Fisher's exact test, where appropriate, with two-tailed p-values given and odds ratios (OR) presented with 95% confidence intervals (C.I.).

3.7.1 Gene-wide haplotype-tagging methodology

The gene-wide haplotype-tagging studies presented in this thesis each use slightly different methodologies reflecting the timescale over which these experiments were performed (2004-2008). All use the same premise. tSNPs were selected that describe the variation across the gene or region of choice. This requires knowledge of the LD across this region. In the pre-Hapmap era this meant resequencing in a select number of controls or cases. Hence, for *GLI1*, coding regions were sequenced in 32 cases (16 UC, 16 CD), SNPs identified, genotypes scored, and then analysed in Haploview (vers 3.2) to select multi-marker tSNPs (Chapter 5). De Bakker has demonstrated that multi-marker tSNPs have increased tagging efficiency compared with single-markers without loss of power.[430] Subsequent to the publication of Phase I Hapmap data, re-sequencing was not a pre-requisite.[431] For *IHH* and *NKX2.3* (Chapters 7 and 11), multi-marker tSNPs were selected on the basis of phase II Hapmap data using de Bakker's Tagger algorithm embedded in Haploview ($r^2 > 0.8$).[430]

3.7.2 Log-likelihood statistic

Haplotype frequencies of the tSNPs were inferred using the expectation-maximization algorithm and used to test whether haplotype frequencies were different in cases and controls as implemented in the EH and PM programmes (log-likelihood test). The test statistic 2*(In(Lcase) + In(Lcontrol) – In(Lcase/Lcontrol)), which has a chi^2 distribution with n-1 degrees of freedom (where n = number of possible haplotypes) was calculated and empirical p values obtained by permuting the data either 1000 or 10,000 times.

3.7.3 Population attributable risk (PAR)

The PAR is defined as the excess rate of disease in individuals with a mutation compared with those without. This is calculated from the attributable risk as a function of prevalence in the exposed population (100/100,000), divided by the incidence of IBD.[432]

3.7.4 Haplotype analysis

Haplotype analysis was performed in Haploview (versions as indicated throughout text).

3.7.5 Transmission disequilibrium testing (TDT)

TDT was performed using TRANSMIT software for *CCL20* (Chapter 10, with assistance from Niall Anderson, Department of Statistics, University of Edinburgh) and in Haploview for *NKX2.3* (Chapter 11).

3.7.6 Meta-analysis

The meta-analysis for rs222826 in *GLI1* was performed using the Mantel-Haenszel method (fixed-effects model) embedded in the R softward package (assistance from Albert Tenesa, MRC Human Genetics Unit, University of Edinburgh).

3.7.7 False positive reporter probability (FPRP)

We estimated the posterior probability that the observed association (meta-analysis of *GLI1* SNP rs2228226) with disease risk represents a true positive by estimating 1 – FPRP using the method described by Wacholder[433] and the prior probabilities for a candidate gene study proposed by Newton-Cheh and Hirschhorn (Chapter 5; assistance from H Campbell, MRC Human Genetics Unit).[434] This approach is based on the fact that the probability of a false positive is dependent not only on the observed p value but also the statistical power of the study and the prior probability that the association is true. We made calculations using the spreadsheet written by El Ghormli and Wacholder and made freely available for non-commercial use.[433]

3.8 *Phylogeny analysis and protein alignment*

GLI1 protein sequences from several mammalian species (human, rhesus, mouse, rat, dog, and cow) were obtained from UCSC (genome.ucsc.edu) or Entrez Protein (http://www.ncbi.nlm.nih.gov). Sequences were loaded into MacVector (www.macvector.com) for alignment and analysis. Regional conservations scores were calculated individually by comparing the most different mammalian GLI1 proteins (human and mouse) (CL and WZ in collaboration).

3.9 *Cell culture*

3.9.1 SW480 cells

Cells from the adherent transformed colonic epithelial adenocarcinoma cell line SW480 (gift from Marian Aldhous) were cultured in L-15 Leibovitz medium (Gibco). Each 500ml of medium contained 2mM L-glutamine (5ml of 200mM solution, Sigma), supplemented with 5ml penicillin-streptomycin solution (10,000 U/ml

penicillin, 10mg/ml streptomycin, Sigma), 10% foetal calf serum (FCS) (56ml, Sigma-Aldrich; Gibco) and 2.5mg/ml plasmosin (Invivogen) for prophylaxis against mycoplasma infection. Cells were kept in 75cm^2 tissue culture flasks at 37ºC (5% CO_2). Once at confluence, cells were washed with PBS and split 1 in 8 into 25cm^2 flasks with 3ml trypsin/EDTA (10×, Sigma). These flasks were left at 37ºC to achieve 70-80% confluence prior to stimulation.

3.9.2 Peripheral blood mononuclear cells (PBMCs)

PBMCs were extracted over a Ficoll-histopaque gradient from 30ml of heparinised blood collected form healthy volunteers and patients in Lithium-heparin tubes. 30ml blood was carefully layed over 15ml Histopaque (H8889, Sigma) in a 50ml conical tube. This was centrifuged at 2100rpm (brake off) for 30 minutes. The interface layer was removed with a Pasteur pipette and added to a new 50ml conical tube with at least an equal volume of PBS. This was centrifuged at 1200rpm for 5 minutes. The resulting cell pellet was re-suspended in 50ml PBS and centrifuged for a further 5 minutes. This last step was repeated. After decanting of supernatant and cell counting, (see below) PBMCs were resuspending in RPMI (1640, Irvine Scientific) and 10% FCS (Gibco). Extraction of PBMCs was optimised in blood samples from 3 healthy volunteers before use in patients with active UC (patient details in Chapter 9).

3.9.3 Cell counting

To count cells, 5ml of RPMI was added to the cell pellet, which was re-suspended and mixed thoroughly. 50µl was added to 50µl of 0.4% tryptan blue (T8154, Sigma). A fresh coverslip was placed on top a haemcounter and a drop of the resulting mixture added to the groove on either end. The number of cells was counted on the in-built 5 x 5 grid. This was multiplied by 10^5 to give the total cell count (i.e. 10^4 for volume plus x2 for dilution with tryptan blue and x5 for dilution in medium).

3.9.4 RNA extraction from cells

The culture supernatant was retained, centrifuged (1000 rpm, 5 min) and stored in aliquots at -20ºC for later protein analysis by ELISA. SW480 cells were dislodged from the base of tissue culture flasks using cell scrapers (Corning Inc.) and centrifuged at 14,000 rpm for 15 seconds. RLT lysis buffer (RNeasy kit, Qiagen) and 2-mercapto-ethanol (≥98%; Sigma) in a ratio of 100:1 was added to the cellular fraction, before storage at -80ºC.

The SW480 cellular suspensions in RLT lysis buffer were pipetted into QIAshredder spin columns (RNeasy kit, Qiagen) and centrifuged at 14,000 rpm for 2 minutes, to homogenise the samples. The homogenised lysates plus 70% ethanol were transferred to RNeasy mini columns (Qiagen) and centrifuged for a further 15 seconds at ≥ 10,000 rpm, allowing the RNA to adsorb to the silica-gel membrane. After washing with buffer RW1 (Qiagen, 15s ≥ 10,000 rpm) the RNA was treated with DNase I (RNase-Free DNase set). The solid DNase I (Qiagen, 1500 Kunitz units) was reconstituted with RNase-free water (Qiagen) and pipetted directly onto the RNeasy silica-gel membrane in a ratio of 1:7 with the DNA digest buffer RDD (Qiagen.) After 15 minutes the DNase was removed by a second wash with buffer RW1. Contaminants were removed with two wash spins using buffer RPE (Qiagen, 2min ≥ 10,000 rpm to dry the membrane.) The RNA was eluted with RNase free water (Qiagen, 1min ≥ 10,000 rpm) and the second elution step was performed with the first eluate to achieve a higher total RNA concentration. The RNA concentrations were determined by measuring the absorbance at 260nm in a spectrophotometer. The ratio of readings at 260nm and 280 nm and 260nm and 230nm provide an estimate of the purity of RNA with respect to DNA contamination and protein contamination respectively.

3.9.5 RNA purity

To check that no residual DNA remained with the RNA, a GAPDH PCR was conducted on the RNA samples, DNA from untreated SW480 cells (positive controls) and water (negative control). GAPDH primers were: forward primer (fp): 5'-TCATCTCTGCCCCCTCTGCT-3' and reverse primer (rp): 5'-CGACGCCTGCTTCACCACCT-3'.

3.9.6 cDNA synthesis

Each sample was diluted to an RNA concentration of 100ng/µl and 400ng was reverse transcribed using the Taqman Multiscribe Reverse Transcriptase kit (Applied Biosystems) consisting of RT buffer (10×), 25mM $MgCl_2$, 10mM dNTP mix, 50µM random hexamers, 20U/µl RNase inhibitor, 50U/µl multiscribe reverse transcriptase and nuclease free water. Samples were incubated for 10min at 25ºC, 40 min at 48ºC and 5min at 90ºC, diluted 1:5 in nuclease free water for Q-PCR and frozen at –80ºC.

3.9.7 Q-PCR on SW480 cells

The real-time mastermix consisted of Taqman Universal PCR mastermix (Applied Biosystems), a primer/probe mix specific for the SHH or PTCH gene sequence (fp, rp, probe and water) with the probe labelled with the fluorochrome FAM, a primer/probe mix specific to 18s rRNA control reagent, with the probe labelled with the fluorochrome VIC (Applied Biosystems) and nuclease free water (Promega.) The following primer/probe sequences were used SHH fp: taaggacaagttgaacgctttgc, SHH rp: tcggtcacccgcagtttcac; SHH probe: catctcggtgatgaaccagtggcca; PTCH fp: tgcaaaccggcagccgcgataag; PTCH rp: ttaatgatgccatctgcatcca; PTCH probe: atcgacatcagccagttgactaaacagcgtc. 2.5µl of each cDNA sample was run in duplicate 25µl volumes on a capped 96-well optical reaction plate (Applied Biosystems.) The plate was run in the Applied Biosystems Prism 7700 sequence detector using SDS software. The PCR conditions were as follows: 50ºC for 2 min, 95ºC for 10 min, then 40 cycles of 15s at 95ºC and 1 min at 60ºC. The software analysed the data and output a Ct for the gene of interest and 18s. The raw Ct values were processed as follows: Ct(FAM) – Ct(VIC) to give ΔCt; sample (ΔCt) was divided by control (ΔCt) to give ΔΔCt; this was transposed into a fold change by the function $2^{-\Delta\Delta Ct}$.

3.9.8 RT-PCR on SW480 cells.

A standard RT-PCR reaction was run on the Q-PCR primer pairs used above. The bands were resolved on a Nusieve 3:1 agarose gel (Cambrex BioScience) which is used for analysis of nucleic acids which are ≤ 1kb. Product sizes were as follows: SHH 211bp; PTCH 462bp; GLI1 244bp.

3.9.9 Immunohistochemistry on cells

SW480 cells were grown in Lab-Tek II chamber slide systems (Nunc) at 37˚C. After stimulation, the culture medium was washed off with PBS and the cells were air dried and fixed in anhydrous acetone (90%, Fissure Chemicals) and methanol (10%, Fissure Chemicals) for 10 minutes at RT. Blocking of endogenous peroxidase, non-specific binding and biotin as for tissue sections was not required. Slides were incubated with the primary antibody (anti-GLI1 [1:400]; anti-p65 [1:100]) diluted with antibody diluent (Dako) and left covered for an hour at RT.

After washing with PBS the biotinylated secondary goat anti-rabbit antibody (Dako) was added to the sections for 30 minutes. The slides were washed again and Vector

RTU ABC was applied for another 30 minutes, followed by DAB+ for 5 minutes. The wells were snapped off and the slides were counterstained. After a final deionised water wash, coverslips were mounted over each well indentation on the slide using aqueous mounting medium (Biømeda.) Images were captured using an Axioskop microscope (× 400) and a Q-imaging 5.0 RTV camera and computer package. Cells counts were performed at ×63 magnification, on 5 fields /well. The total number of cells and the number with nuclear staining were recorded per field.

3.9.10 Enzyme-linked immunosorbent assay (ELISA) on cell culture supernatant

ELISAs were performed for CCL20 (DT360), IL1β (DY201), IL8 (DY208), IL10 (DY217B), TNFα (DY210), and IFNγ (DY285) using DuoSet ELISA Development System (R&D Systems) according the to the manufacturers' instructions for each kit. All samples were run in duplicate. As an example, the CCL20 protocol was as follows. 100µl of capture antibody was applied to each well (COSTAR 9018 high-binding, polystyrene-coated 96-well plate) and incubated overnight at RT. 300µl /well of blocking solution (1% BSA in PBS) was added for 1 hour. 100µl /well of standards and samples were added for 2 hours at RT. Standards underwent a 7-point 2-fold dilution (top 1000pg/ml). ≥2 wells were left blank as controls. 100µl /well of detection antibody was added for 2 hours at RT. In between each of these stages, all wells were thoroughly washed with PBS x3. 100µl /well streptavidin-HRP solution (1:200) was added to each well, followed by 100µl /well 3′,3′,4,51-Tetramethylbenzidine (TMB, T0440, Sigma), both for 20 minutes in the dark. The reaction was stopped with 50µl of stop solution /well and read at 540 or 570nm.

3.9.10.1 Buffers and solutions for ELISA

- Wash buffer (all)
    - 0.05% Tween 20 in PBS, pH 7.2-7.4
- Reagent diluent (TGFβ2 & CCL20)
    - 1% BSA in PBS, pH7.2-7.4, 0.2µm filtered
- Block buffer (IFNγ)
    - 1% BSA in PBS with 0.05% NaN3
- TBS
    - 20mM Trizma base, 150mM NaCl)
    - To 800ml dH20 add 8g NaCl and 0.2g KCl and 3g Tris Base

        - Adjust pH with conc HCl
        - dH20- to 1000mL
- Reagent diluent (IFNγ)
    - 0.1% BSA, 0.05% Tween 20 in TBS, pH7.2-7.4, 0.2μm filtered
- 1N HCl
    - To 91.67mL dH20 slowly add 8.33mL 12 N HCl. Mix well.
- 1.2 N NaOH / 0.5M HEPES (100Ml)
    - To 75mL dH20 slowly add 12mL 10N NaOH. Mix well. Add 11.9g HEPES. Mix will. Bring final volume to 100mL with dH20.

3.9.11 NFκB luciferase reporter assay

For the NFκB luciferase reporter assay, SW480 cells were grown to ~50% confluence in 6-well plates with 5 ml medium. The ConA NFκB and δκB constructs were gifts from Lesley Stark (MRC Human Genetics Unit, University of Edinburgh). For the transfections, 10μl of lipofectamine 2000 (1mg/ml Invitrogen 11668-019) plus 100μl Optimem (serum-free medium; Gibco 51985-026) per well were mixed and added to an equal volume of both NFκB and δκB (3μg construct plus 1.5μg pCMVβ in 100μl Optimem). This solution was mixed and made up to 1ml/well with Optimem. For each control and treatment, samples were duplicated to allow for both NFκB and δκB constructs. The transfections were incubated at 37°C for 5 hours, before the medium was changed for 0.5% serum.

After 24 hours the cells were stimulated for the desired time (all 24 hours for the data reported here in Chapter 9). At timepoint 0 the cells were harvested and incubated with reporter lysis buffer through a series of centrifugations and washing steps. Samples were then immediately analysed on a luminometer or frozen at -80°C. The samples were normalised for the transfection efficacy by dividing the luminescence reading by the concentration of pCMVβ calculated by ELISA. This was then 'zero'd' to the experimental control to generate fold changes in NFκB activity.

Total number	**150** **30 HC; 67 UC; 53 CD**
Sex (% male)	HC 33.3%; UC 49.3%; CD 50.9%
Median age at diagnosis/ recruitment (years)	HC 43.0 UC 37.0 CD 28.6
CD location	
Terminal ileum (L1)	6/53 (11.3%)
Colon (L2)	28/53 (52.8%)
Ileocolon (L3)	19/53 (35.8%)
UC location	
Proctitis (E1)	15/67 (22.4%)
Left-sided colitis (E2)	27/67 (40.3%)
Extensive colitis (E3)	25/67 (37.3%)
CD 5 year behaviour	
Inflammatory (B1)	32/53
Stricturing (B2)	8/53
Penetrating (B3)	12/53
Surgery	CD 20/53 (37.7%) UC 2/67 (3.0%)
Medical therapy for microarray cohort	
5-ASA	CD 21/53 (39.6%); UC 40/67 (59.7%)
Corticosteroids	4/53 (7.5%); UC 10/67 (14.9%)
Immunosuppressants	13/53 (24.5%); 11/67 (16.4%)

Table 3-1 Demographics and phenotype of patients in microarray cohort.

Gene	Forward primer	Reverse primer	Annealing conditions
GLI1	GTCTCAGCTTGTGTGTAATTATG	TTCTGGAAGGATAGAAACCC	59°C; 1min
GLI1	CAACTGACCTCAGGTGATCCTC	AGAGTCATGGGGACCACAAG	62°C; 3min
GLI1	GGAGATGTGAGGCGTCAGAG	GTTGATGAAAGCTACGAGGGAG	60°C; 1min
GLI1	TCCTTCTGCTTACTTCCACCCTC	GGAAGAGACCCTGGACTTGG	60°C; 1min
GLI1	CAGTCACTGGGACACAGGC	CAAGGGTGACTTCCTCCTCTC	60°C; 1min
GLI1	TCTCTGATGTGTGTCCTGTTGG	GTGGGTGCTGGGCTAAGG	60°C; 1min
GLI1	GTCTCAAGCCCTCAAACTACC	CTAGGGCAAGAGAGGCAATC	64°C; 1min
GLI1	GGGCACTTAGGGCAGGAA	CTTTATGCCAACACAGTCACAC	63°C; 1min
GLI1	CATGGGAGAAGTAGGAGAC	GCACTTGTCCATAATGTTC	59°C; 1min
GLI1	CCCAGCAGGCCTCATATC	CCAGCCCATACCTCCCATCC	64°C; 1min
GLI1	GTAAAACGACGGCCAGAGCATCA ATAGGGCAAAGCAG	CAGGAAACAGCTATGACCTTGGG ATTTGTAGTGGTTG	50°C; 1min
GLI1	GTAAAACGACGGCCAGCCAGGA GTTCGAGGCTATG	CAGGAAACAGCTATGACGGGACA GAGTTGCGGTTTG	61°C; 1min
CCL20	TATGTGGTTTCCTTTCTGTCTG	GTAACACAAAATTAACCCATTGG	60°C; 1min
CCL20	ATTTCCTATTCTTCCCTTCAATTC	CTCAAACCTCAGCTTCACC	60°C; 1min
NKX2.3	GCCAAGCAAGAGCTGTCG	TTACAGCCACTGCAGGGACTG	64°C; 1min
NKX2.3	CCACATGAGCCAGAACCTATAG	TTTGTGGAGCTGTCGGTGAC	64°C; 1min
NKX2.3	GGCGAGGAGGATGACTGGG	TCGCACCTTTTCCTCCCCTTC	64°C; 1min

Table 3-2 Primer pairs for DNA sequencing

Figure 3-1 Controls for HH (N-19) immunohistochemsitry
The specificity of the N-19 HH antibody is demonstrated with application of the primary antibody, primary antibody + SHH blocking peptide pre-absorbed overnight, and with primary antibody omitted. This demonstrates that the antibody reacts with the peptide but does not preclude it from cross-reacting with other proteins. Indeed, we would agree with other commentators in urging caution in the use of this and other presently available HH antibodies for immunohistochemistry.[334]

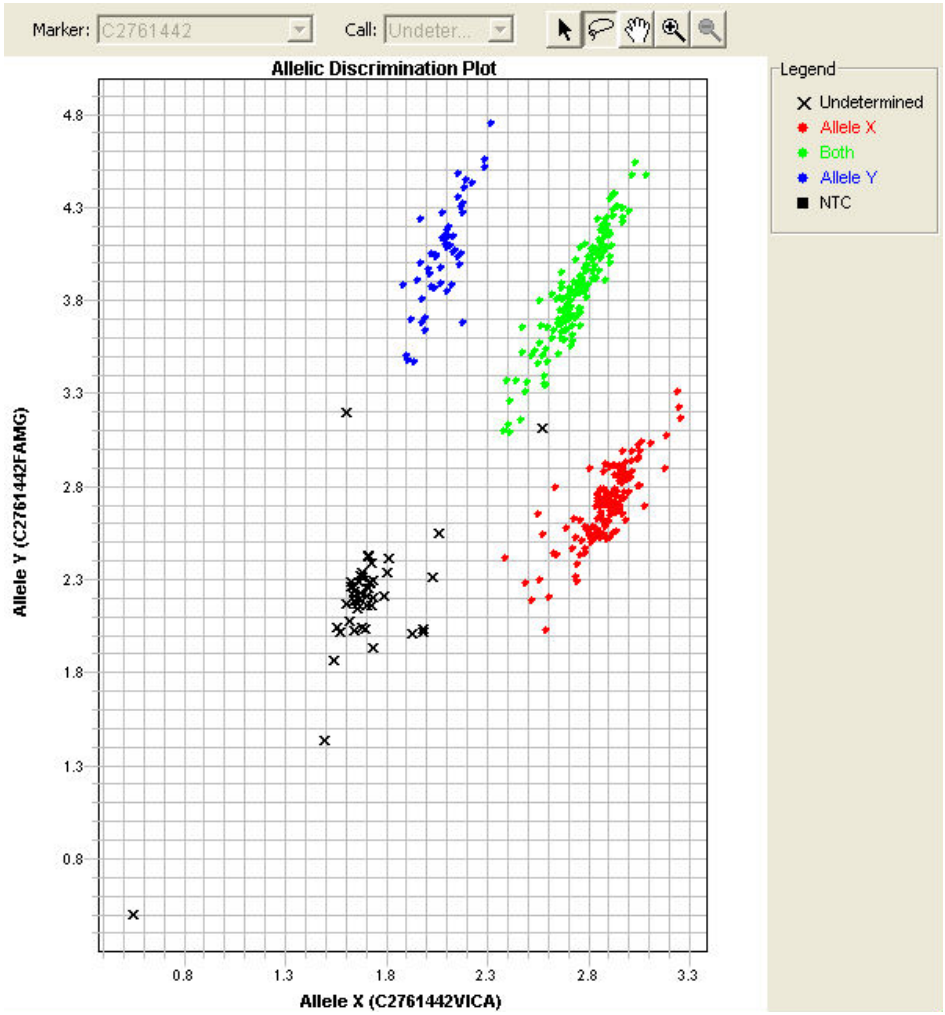

Figure 3-2 Allelic discrimination in Taqman genotyping

4 Expression analysis of key HH and WNT signalling components in the healthy GI tract and IBD.

4.1 *Abstract*

Introduction. The HH and WNT pathways have critical and overlapping roles in the development of a healthy mammalian gastrointestinal tract. The expression of various HH signalling components has been previously reported in the healthy adult small intestine and colon. However, these existing reports are somewhat limited and conflicting in their nature. In contrast, the profile of WNT pathway genes has been documented in great detail in the murine intestine. Multiple lines of evidence suggest that HH may play an important role in the pathogenesis of IBD.

Aims. The aim of the present study was therefore to document the expression profiles of HH and WNT signalling components in the health small bowel and colon, in UC and CD, and in non-IBD colonic inflammation.

Methods. Immunohistochemistry, microarray and RT-PCR were used to document the protein and mRNA expression of key HH signalling components in endoscopic biopsies and surgical resection specimens from healthy controls and patients with UC and CD. GLI1 expression, a direct target of HH signalling and robust indicator of pathway activity, was compared by linear regression analysis with expression of known and potential HH target genes. In addition, the microarray dataset was mined for all 27 WNT signalling components with known intestinal expression to provide a detailed profile of this pathway in IBD.

Results. In the healthy adult colon, HH protein is primarily expressed in terminally differentiated enterocytes on the luminal surface. There is a gradient of increasing HH protein expression along the length of the colon, with greatest expression in the rectum. This is mirrored by increased mRNA of the HH response network (GLI1, PTCH and HHIP) in the distal colon. GLI1, PTCH and HHIP mRNA was lower in inflamed UC compared with HC tissues. IHH mRNA was decreased in UC regardless of inflammatory status. GLI1 and PTCH were also decreased in colonic CD and non-IBD inflammation. GLI1 expression closely correlated with BCL2, BMP4, P21/CIP1, NOD2, a raft of inflammatory cytokines, and NKX2.3. There were less significant alterations in the WNT signalling pathway noted in IBD, although several WNT genes did exhibit differential expression. TCF4, FZD4, DKK3 and LRP5 expression profiles were closely correlated with GLI1 in non-inflamed control samples. The GSK3 genes, with key roles in the phosphorylation of WNT and HH pathway

components, were dysregulated in inflamed UC and CD but not in non-IBD inflammation.

Conclusions. This study provides to date the most detailed expression analysis of the HH signalling pathway in the healthy colon and in IBD. Of note, the gradients of increasing HH pathway activity in the healthy colon mirror the clinical distribution of UC and the increasing bacterial burden in the colon. The pathway is down-regulated in IBD and non-IBD inflammation, suggesting a key role in the regulation of colonic inflammatory pathways. The down-regulation of IHH in UC, regardless of inflammatory status, suggests a possible role for germline variation in *IHH* in disease pathogenesis. The WNT pathway analysis indicates multiple over-lapping roles for both pathways in IBD.

4.2 *Introduction*

Various groups have demonstrated expression of key HH signalling components in the small intestine and colon; however, the published literature is somewhat conflicting. Oniscu and co-workers in Edinburgh found protein in the cytoplasm at the top of normal colonic crypts.[330] This localisation was confirmed by the presence of SHH mRNA in laser microdissected crypts. Similar protein findings were reported by Dimmler *et al*, who showed SHH expression in the distal colon only.[328] Most recently, Nielson *et al* have reported the presence of SHH mRNA by *in situ* hybridization and protein by immunohistochemistry in small intestinal and colonic crypts, predominantly in the bases but with some focal surface epithelial expression.[329] The discrepancy in protein data likely reflects varying protein detection sensitivities in differing immunohistochemistry protocols, since different labs used the same antibody, and may also reflect differences in colonic sampling. Colonic gradients of HH signalling are potentially of functional relevance. It is of note that the differential expression of many different genes in the right and left colon has recently been described.[435]

Van den Brink's group has subsequently shown expression of IHH mRNA and protein in the human colon localising to surface absorptive enterocytes.[318] These same cells were negative for SHH mRNA by *in situ* hybridization.[327] There are no published data examining DHH expression in the adult mammalian digestive tract.

PTCH mRNA and protein have previously been reported to be expressed in similar locations to SHH in the adult human gut. mRNA has been identified in the base of the villi and the lamina propria of the small intestine, with protein also detected in tips of villi.[329] In the colon PTCH mRNA and protein has been detected in the luminal epithelium and subadjacent lamina propria,[318] and also basally in isolated crypt cells.[330] GLI1 protein expression has been described at the surface epithelium and the top of colonic crypts.[417] Analysis of foetal and adult colonic mRNAs by RT-PCR confirms expression, if not location, of all essential Hh signalling components (SHH, PTCH, SMO, GLI1, GLI2 and GLI3).[436]

Recent data reported by Nielson *et al* indicates that SHH and PTCH are increased in a variety of intestinal inflammatory conditions, including IBD, but their findings are rather preliminary and warrant further investigation.[329]

Several lines of evidence implicate the WNT signalling pathway in IBD pathogenesis. Furthermore, WNT and HH pathways have been shown to interact in colo-rectal cancer pathogenesis.[318, 417, 437] In contrast to HH, the murine intestinal expression profiles of all WNT signalling components has previously been described in great detail by carefully conducted in situ hybridisation studies.[398]

GSK3 is a key enzyme in WNT signalling. It main role is phosphorylation of β-catenin leading to its degradation. Additionally, it is known to phosphorylate the tail of LRP, AXIN, APC and TCF. Furthermore, it has important roles in HH signalling, as well as insulin signalling and NFAT phosphorylation.

In the present study, a detailed analysis of protein and mRNA expression of key HH and WNT signalling components in the healthy human colon and small intestine was performed, along with comparison between inflamed and non-inflamed UC, CD, non-IBD inflammation and HC.

4.3 *Methods*

To investigate the functional integrity of the HH and WNT signalling pathways, expression of pathway components in the intestine by microarray, RT-PCR and immunohistochemistry was examined. Microarray data from 129 UC, 99 CD and 76 HC intestinal biopsies were available for all HH signalling components except IHH, which was analysed by RT-PCR (these experiments were performed at Genentech, Inc; all data analysis performed by CWL). Immunohistochemsitry was performed on formalin-fixed and paraffin-embedded human small intestine and colon from HC and patients with CD and UC collected at colonoscopy and surgery.

4.3.1 Immunohistochemistry, microarray and Q-PCR

The full details of the methodology for immunohistochemistry, microarray and Q-PCR presented here are detailed in Chapter 3, including all reagents, antibodies, primers and probes, and statistical analysis.

4.3.2 Analysis of HH target genes

GLI1 expression is a robust indicator of HH pathway activity. This readout was used to assess a panel of known and potential HH target genes for correlation with HH pathway activity in the non-inflamed terminal ileum and colon from the microarray dataset. HH target genes were selected from a number of different sources. Firstly, a detailed survey of the published literature was performed, including a study by Hallikas *et al*, where a high-throughput bioinformatics model (enhancer element locator) was used to examine gene promoters for DNA with homology to the GLI consensus sequence.[411] Secondly, Professor Gumucio (Ann Arbor, MI) provided a list of inflammatory genes (searched under gene ontology [GO] terms relating to inflammation) that were dysregulated in a microarray performed on total small intestine from villin-Hhip mice vs. WT at birth (18 days) (full list of inflammatory genes and fold-changes listed in **Table 13-1**).[303] The TG promoter (villin) is switched on at embryonic day 14-15 in this model, resulting in over-expression of the pan-Hh inhibitor Hhip, and subsequent down-regulation of Hh pathway activity (further discussion of this model and the results in Chapter 12). The target gene of choice was plotted individually (x-axis) against GLI1 (y-axis) in GraphPad Prism, and correlation assessed with linear regression (95% confidence intervals plotted) to obtain r^2 and p values.

4.4 *Results*

In healthy colons, HH protein (N-19 antibody) was primarily present in terminally differentiated enterocytes on the luminal surface, with extension into the upper two-thirds of colonic crypts (**Figure 4-1a-c** and **4-2a**). This antibody is meant to detect SHH and was successfully blocked with the specific blocking peptide (**3.4.3.1** and **Figure 3-1**). However, it is widely thought that it also detects IHH at higher concentrations due to the tight homology of SHH and IHH amino acid sequences, and it is therefore referred hereafter as simply HH (N-19). IHH, PTCH and HHIP were expressed in a proportion of basal crypt cells, identified as entero-endocrine cells by morphology and co-localisation of PTCH with chromogranin A (**Figure 4-1d-g;h and Figure 4-2b-c**). PTCH protein was also detected in the mesenchyme - in immune cells below the epithelium and in germinal centres of some lymphoid follicles (**Figure 4-1g and Figure 4-2d-g**). These expression patterns map to those of macrophages and lymphocytes (**Figure 4-2i**). Smoothened was expressed only in colonic

epithelium (**Figure 4-1h**). The transmembrane pathway inhibitor HHIP was expressed in enterocytes, predominantly at the luminal surface, in entero-endocrine cells, and in the lamina propria (**Figure 4-1i**). GLI1 immunoreactivity was detected in predominantly cytoplasmic compartments of epithelial and lamina propria cells (**Figure 4-1j**), although some nuclear staining was also evident. GLI2 expression was exclusively nuclear in both epithelium and mesenchyme (**Figure 4-1k**), whilst GLI3 was only present in endothelial cells in blood vessel walls (**Figure 4-1l**).

4.4.1 Developmental gradients of HH expression persist in the healthy adult colon

In HC tissue, it was evident that HH gradients present during development[270] persist in the healthy adult colon. Basal GLI1 ($p<0.0001$, Kruskal-Wallis), GLI3 ($p<0.0001$), PTCH ($p=0.002$), and HHIP ($p=0.03$) mRNA levels were greater in the distal than proximal colon (**Table 4-1** and **Figure 4-3b**). GLI1, PTCH and HHIP are key HH response genes, providing positive and negative regulation; as direct transcriptional targets of HH pathway activation, mRNA levels predict pathway activity. HH protein (N-19) was similarly greater along the length of the colon (**Figure 4-3a**); however, SHH mRNA levels were similar whatever the location. Similarly, there was no differential expression of DHH, GLI2, SUFU, SMO or DISP1 expression along the colon, whereas KCTD11 ($p=0.0001$) and PTCH2 ($p=0.001$) levels were lower in distal compared with proximal colon (**Figure 4-3b**).

4.4.2 The HH response network (GLI1, PTCH, HHIP) is downregulated in UC

GLI1 ($p=0.0003$), PTCH ($p=0.002$), and HHIP ($p=0.0003$) mRNA transcripts were all lower in inflamed UC compared with HC tissue from equivalent colonic location; KCTD11 ($p=0.0002$) and PTCH2 ($p=0.0009$) were greater (**Figure 4-4**). These expression profiles were not altered by medical therapy for UC (5-ASA, thiopurines or corticosteroids) (**Figure 4-5**). We observed lower IHH mRNA in UC samples, irrespective of inflammatory status ($p=0.02$) (**Figure 4-4**). There was no difference in IHH between non-inflamed UC and HC. GLI1 and PTCH were also lower in CD (**Figure 4-6**), although with a less distinct response to inflammatory status, and in non-IBD inflammation (**Figure 4-7**).

In contrast with IHH expression, SHH mRNA was increased in UC (UC vs. HC p=0.0003) (**Figure 4-4**) and non-IBD colonic inflammation (**Figure 4-7**), but not in CD (**Figure 4-6**). These microarray data were confirmed by RT-PCR in a second UC cohort. Notably, the higher levels of SHH were present only with inflammation (p=0.02) (**Figure 4-4**). There was no change in levels of DHH, PTCH2, GLI2, GLI3, SUFU or DISP1 in either UC or CD compared with HC, or in non-IBD inflammation (**Table 4-1**).

HH protein (N-19) was increased in a proportion of UC and CD tissues, with expression identified in enterocytes throughout the crypt length, compared with HC tissue from the same colonic sampling point (**Fig 4-8a,** ascending colon). It was notable that in most areas of ulceration no significant HH expression was detected. PTCH expression was predominant in the inflammatory infiltrate, in lymphocytes and macrophages (**Figure 4-8c-e**). GLI1 expression was largely decreased in the epithelium; however, marked staining of expanded plasma cell populations was noted in 7/17 UC colectomy specimens (**Figure 4-8b**).

These data demonstrate down-regulation of IHH and the HH response genes (GLI1, PTCH, HHIP) with greater SHH expression in association with colonic inflammation.

4.4.3 HH pathway activity (GLI1 expression) vs. target genes

A panel of known and potential HH target genes (**Table 4-2**) was assessed using GLI1 expression as a readout of HH pathway activity. Close correlation of GLI1 expression was found with NOD2, BCL2, BMP4 and P21/CIP1 expression in the healthy adult intestine (**Figure 4-9a**). Given the presence of GLI1 enhancer elements upstream of *NOD2*, the correlation between GLI1 and NOD2 expression was analysed in more detail in control and disease biopsies (**Figure 4-10**). Of note, the gene expression profiles of GLI1 correlated closely with the expression of these target genes in non-inflamed samples from HC and CD. This correlation was lost in UC. Other genes of interest that significantly correlated with GLI1 included a panel of cytokines (IL10, IL5, IL1R1, IL17A, IL21, IL4 and TGFβ3), FOXP3, the peptidoglycan recognition proteins (PGLYRP1-4), and NKX genes (NKX2.8 and NKX2.3) (**Table 4-2**).

4.4.4 Expression of WNT signalling components in health and UC, and interactions with HH

Interactions between the HH and WNT signalling pathways, both critical to intestinal development, have been described in intestinal homeostasis,[303] and colon cancer pathogenesis.[318, 417] To explore the relative contribution of WNT signalling to IBD pathogenesis the microarray dataset was first interrogated for expression profiles of WNT components with previously documented intestinal expression profiles (**Table 4-3**).[398] Of note, differential expression along the length of the healthy adult colon was noted for WNT6 ($p=0.037$), FZD4 ($p<0.0001$), FZD7 ($p=0.0007$), LRP5 ($p<0.0001$), LRP6 ($p=0.0026$), TCF3 ($p=0.0018$), TCF4 ($p=0.028$), DKK3 ($p=0.0005$), APC ($p=0.0017$), CTNNB1 ($p=0.014$), and AXIN1 ($p<0.0001$) (**Table 4-3, Figure 4-11a** and **13-1 - 13-3**).

Comparison of UC and HC samples revealed relatively greater expression of WNT2B ($p=0.044$), WNT9B ($p=0.022$), FZD4 ($p=0.030$), DKK4 ($p=0.0026$), TCF1 ($p=0.0010$), TCF4 ($p=0.021$), and ATOH ($p=0.0017$) in inflamed UC compared with non-inflamed HC (**Table 4-3** and **Figure 4-11b**). Of these genes, DKK4 and TCF1 showed comparable changes in non-IBD colonic inflammation (**Table 4-3** and **Figures 13-7 – 13-9**). The expression of AXIN1 was significantly lower in both inflamed ($p<0.0001$) and non-inflamed ($p=0.0002$) UC samples compared with HC.

FZD5 expression was lower in both colonic ($p<0.0001$) and terminal ileal ($p=0.018$) CD compared with HC (**Table 4-3** and **Figures 13-10 – 13-15**). Other notable changes in colonic CD were increased expression of TCF4 ($p=0.0038$) and decreased expression of WNT6 ($p=0.0003$, non-inflamed samples only), LRP6 ($p=0.0010$) and TCF2 ($p=0.0025$). In terminal ileal CD, there was increased expression of WNT5a ($p=0.0016$) and decreased expression of FZD7 ($p=0.0008$). Of these altertations in WNT signalling components in CD, only that of FZD7 was mirrored by a decrease in non-IBD colonic inflammation ($p=0.022$).

GSK3, involved in both WNT and HH signalling, has two mammalian forms (GSK3α and GSK3β) that are redundant.[438] GSK3 expression did not alter along the length of the healthy human colon, nor with non-IBD inflammation (**Figure 4-12**). GSK3α

expression was lower in inflamed UC ($p<0.0001$) and inflamed terminal ileal CD samples ($p=0.036$) compared with HC. GSK3β expression was unchanged in UC, but increased in inflamed CD in both colonic ($p=0.0007$) and terminal ileal samples ($p=0.0076$) (**Figure 4-12**).

HH pathway activity has been reported to be modulated by the WNT components sFRP1[420] and beta-catenin.[318, 417] There was only minimal correlation between the expression profile of sFRP1 with GLI1, but close positive correlations with TCF4, FZD4, DKK3 and a negative correlation with LRP5 (**Figure 4-9b**). In contrast to the decrease in IHH and GLI1 expression in UC, beta-catenin staining is increased with greater nuclear accumulation (**Figure 4-13**). No clear correlation between GLI1 and beta-catenin staining was observed in a panel of 17 UC colectomy specimens.

4.5 *DISCUSSION*

The expression data from the healthy colon indicate that the HH pathway is more active in the distal compared with proximal colon, with similarities to the gradients of HH signalling previously reported only in development. The HH protein gradient is mirrored by pathway response genes, including GLI1, but not by SHH and DHH mRNA, suggesting that IHH is driving the greater HH activity in the distal colon. This is consistent with reports that IHH directs enterocyte differentiation and is the major epithelial HH signal in the adult colon.[318] Here lamina propria immune cells have been identified as targets of HH signalling. The presence of PTCH[329] and GLI1 in these cells indicates they have the ability to receive and process a HH signal. This is entirely consistent with *in vitro* data that SHH exhibits co-stimulatory activity on peripheral CD4 T cells that express PTCH on their surface, potentiating CD3-mediated proliferation and increasing cytokine production.[365, 366, 422] Furthermore, germinal centre B cells express key elements of the HH response network, and depend on SHH, produced by germinal centre DCs, as a survival signal to prevent apoptosis.[439]

In fact, an emerging body of data strongly supports the notion that the intestinal HH signal is purely paracrine, first suggested in murine development by Maddison and Gumucio in Ann Arbor,[303] and subsequently in homeostasis (Kolterud and Gumucio, personal communication) and malignancy.[361] We have now clearly demonstrated in

the adult mouse, in a series of closely collaborative experiments with Zacharias and Gumucio (initiated by CWL and designed with WZ and DG), that epithelial Hh signals to underlying mesenchymal cells, predominantly macrophages and DCs (Lees *et al*, *PLoS Medicine* 2008).[423] In *Gli1*$^{+/LacZ}$ animals, the colonic response to acute inflammatory challenge (3% DSS) is exclusively paracrine (further details and discussion in Chapter 12).[423] The strong advantage of this experimental system is that it allows identification of Hh responsive cells by LacZ staining. This avoids use of the polyclonal antibodies (against HH signalling components) described in this chapter, doubts over whose specificity have recently been raised by a number of researchers in the field.[334]

There are striking parallels between this distribution of HH signalling in the colon and the clinical pattern of UC. In contrast to CD, the rectum is invariably affected in UC, with variable proximal extension to involve the left side of the colon (40.5% in our cohort) and more extensively (42.2% with disease proximal to the splenic flexure). IHH – GLI1 signalling has been shown to antagonise the WNT pathway in the colon.[318, 417] The suppression of WNT signalling by high levels of IHH-PTCH-GLI1 in the distal colon may limit the capacity for effective response to injury and lead to chronic inflammation in susceptible individuals. However, it is difficult to extrapolate from the present study any clear functional link between HH and WNT. Certainly the present data in HC vs. UC indicate that WNT signalling appears to play a lesser role in the pathogenesis of UC than HH, in contrast to the critical role established for WNT in colo-rectal cancer.

It has recently been demonstrated that IHH regulates Paneth cell differentiation in the murine intestine.[332] It would appear that IHH, synthesised and secreted by mature Paneth cells, is processed by pathway response elements in Paneth cell precursors, thus negatively regulating their differentiation (see **1.8.2.3**).[332, 333] This observation, coupled with the role of HH in inflammation and repair in diverse tissues,[349, 358, 367, 374, 409] links with NFκB / IFNγ signalling,[369, 371, 372] the identification of GLI1 response elements upstream of the *NOD2* gene, and the Ashkenazi pedigree with co-existent IBD and basal cell naevus syndrome (BCNS),[414-416] strongly suggests a link between this developmental pathway and intestinal inflammation.[270] The decrease in IHH mRNA expression in UC, coupled with evidence of reduced pathway activity

(decreased HHIP, PTCH and GLI1) may have two important consequences – Paneth cell metaplasia and increased WNT pathway activity. This link is further enhanced by our report demonstrating striking upregulation of alpha-defensins 5 and 6 (HD5 & HD6) in UC (compared with HC), with strong correlation to Paneth cell metaplasia (Noble, Abbas, Cornelius, Lees *et al*, GUT 2008).[428]

In stark contrast to the decrease in IHH signalling, the present study and the published literature describes increased SHH protein and mRNA in IBD tissue.[329] The increased levels of SHH mRNA in UC were restricted to the inflamed group of biopsies, suggesting this is a response to NFκB activation. This induction of SHH mRNA with inflammation is all relative to normal tissue; the absolute contribution of SHH in this setting may therefore still be low. This is supported by the global decrease in PTCH, HHIP and GLI1, further suggesting that IHH is the dominant signal in the colon. However, we did note the presence of PTCH and GLI1 in the inflammatory infiltrate, variably in T lymphocytes, macrophages and plasma cells, further supporting the notion of paracrine HH signalling in the colon.

Several WNT signalling components demonstrated differential expression patterns in IBD tissue compared with HC. Whilst the present study was limited to microarray data, combined with the thorough pre-existing knowledge of WNT expression in the murine intestine this is a reasonably powerful approach.[398] Of note, WNT5a, increased in terminal ileal CD, is expressed in villus mesenchyme. WNT6, decreased in non-inflamed colonic CD, is expressed in colonic epithelium. FZD5, decreased in colonic and terminal ileal CD is expressed in crypt epithelium, as is FZD7, decreased in terminal ileal CD only. TCF1, increased in UC, is expressed in gut lymphocytes. Finally, we noted increased expression of TCF4 in colonic IBD where it is expressed in the epithelium. We did not, however, note differential expression in the terminal ileum as has been previously described. Of note, in this German study the lower TCF4 expression mirrored that of α-defensins in CD.[421] In a follow-up study, presently only in abstract form, Wehkamp's group found reduced TCF1 in ileal CD. In keeping with a strong correlation with TCF4 levels ($r^2 = 0.52$), TCF4 binding sites were identified in the TCF1 promoter region.[440]

Somewhat consistent with our results here for TCF1 and TCF4, our data are also discordant with Wehkamp's for the α-defensins. By microarray and Q-PCR analysis, α-defensin expression was unaltered in terminal ileal CD in our study (Noble, Abbas, Lees et al, manuscript in preparation). However, further to the discussion above in the context of reduced HH signals in colonic inflammation, the increase in TCF1 and TCF4 in UC is consistent with a role for both WNT and HH in Paneth cell metaplasia.

You and colleagues recently utilised a WNT-specific microarray to perform a limited survery of WNT gene expression in 6 patients with UC (paired biopsies from inflamed and non-inflamed areas) and 6 HCs.[441] In this study, and largely in contrast to the present data, they noted increased expression of WNT1, WNT2B, WNT3A, WNT5B and WNT9A in non-inflamed UC samples versus HC. Consistent with the present data, they noted higher levels of FZD4 and lower levels of FZD5 and, with inflammation, higher levels of DKK4. Their study design was limited by small numbers of patients and no control for position of biopsies within the colon.

GENE	HC gradient	HC (N-I) vs. HC (I)	UC (SC) vs. HC (SC)	CD (SC) vs. HC (SC)	CD (TI) vs. HC (TI)
SHH	↑*	↑ P=0.0004	↑^ P=0.006	-	-
IHH	N/A	N/A	↓† P=0.02	N/A	N/A
DHH	↓ P=0.002	↑ P=0.01	↑ P=0.001	-	-
DISP1	-	-	-	-	-
PTCH	↑ P=0.002	↓ P=0.005	↓ P=0.002	↓ P=0.007	-
PTCH2	↓ P=0.001	-	↑ P=0.0009	-	-
SMO	-	-	-	-	-
HHIP	↑ P=0.03	-	↓ P=0.0003	-	-
SUFU	↑ P=0.05	-	-	-	-
GLI1	↑ P<0.0001	↓ P=0.0005	↓ P=0.0003	↓ P=0.004	-
GLI2	-	-	-	-	-
GLI3	↑ P<0.0001	↓ P=0.044	-	↓ P=0.024	-
KCTD11	↓ P=0.0001	↑ P=0.0003	↑ P=0.0002	↑ P=0.0049	-

Table 4-1 Summary of HH pathway expression profiles in HC, UC and CD.
HC gradient is calculated by Krusall-Wallis test on expression values in ascending, descending and sigmoid colon. Non-inflamed (N-I) HC were analysed versus inflamed (I) HC samples, all from the sigmoid colon (SC) to take into account gradients of expression. UC analysis is presented for inflamed UC vs. HC (N-I) for SC samples. CD analysis is separated in SC and terminal ileal (TI) samples in CD and HC. All p-values (except HC gradient) are 2-tailed and calculated by Mann-Whitney U test.
* *HH protein expression increases along length of healthy human colon (see Figure 4-2); no change in SHH or DHH mRNA; no microarray data on IHH.*
^ *subsequently confirmed by Q-PCR (p=0.02)*
† *Performed by Q-PCR (IHH not represented on microarray chip)*

GENE	r^2 value	+ve or –ve	P-value
BCL2	0.540	+	<0.0001
BMP4	0.430	+	<0.0001
S100A4	0.420	+	<0.0001
NOD2	0.410	+	<0.0001
P21/CIP1	0.410	-	<0.0001
KCDT11	0.360	-	<0.0001
NKX2.5	0.330		<0.0001
CENTB1	0.260	-	<0.0001
IL10	0.250	-	<0.0001
NKX2.1	0.250		<0.0001
IL5	0.240	-	<0.0001
IL1R1	0.220	+	<0.0001
IL17A	0.190	-	<0.0001
PGLYRP2	0.190	-	0.0001
PGLYRP3	0.190	-	0.0001
NKX2.8	0.190		0.0009
MAP2K4	0.180	+	0.0002
IL21	0.180	-	0.0001
TRAF4	0.170	-	0.0004
THBD	0.170	+	0.0003
NKX2.3(B)	0.170		0.0017
ICAM1	0.160	+	0.0003
WNT8B	0.150	-	0.0006
TGFβ3	0.140	-	0.001
S100A6	0.140	-	0.0011
MADCAM1	0.140		0.0056
TGFβ1	0.130	+	0.0018
CD24a	0.130	-	0.0021
NKX2.3(A)	0.130		0.007
N-MYC	0.120	-	0.0029
DEFB4 (HBD2)	0.110	-	0.0054
SOX13	0.110	-	0.005
SP1	0.110	-	0.0048
IFNγ	0.110	-	0.0062
TNFα	0.110	+	0.0044
IL4	0.110	-	0.022
PGLYRP1	0.110	-	0.0037
RIPK2	0.100	+	0.0055
MMP10	0.100	+	0.0055
FOXP3	0.100	-	0.0059
VEGF	0.093	-	0.0085
PGLYRP4	0.090	-	0.0083

PPARβ	0.088		0.0007
DEFB1	0.084	-	0.013
RIPK1	0.079	-	0.016
MMP1	0.069	+	0.025
MMP12	0.069	+	0.025
IL12B	0.069	-	0.024
sFRP2	0.062	-	0.034
ATOH1	0.059	-	0.039
sFRP1	0.055	-	0.046
AREG	-	n/a	ns
APC	-	n/a	ns
AXIN2	-	n/a	ns
CCL20	-	n/a	ns
CCND1	-	n/a	ns
DEFA5	-	n/a	ns
DEFA6	-	n/a	ns
EGR3	-	n/a	ns
FGF10	-	n/a	ns
FGF8	-	n/a	ns
IFNγR1	-	n/a	ns
IL10RB	-	n/a	ns
IL17E	-	n/a	ns
IL1B	-	n/a	ns
IL1RA	-	n/a	ns
IL23 (P19)	-	n/a	ns
IL27	-	n/a	ns
IL6	-	n/a	ns
IL8	-	n/a	ns
JAG1	-	n/a	ns
JUN (AP-1)	-	n/a	ns
MMP13	-	n/a	ns
NKX2.2	-	n/a	ns
PAP	-	n/a	ns
PPARγ	-	n/a	ns
RELA	-	n/a	ns
S100A1	-	n/a	ns
S100A8	-	n/a	ns
S100A9	-	n/a	ns
SNAIL	-	n/a	ns
TGFβ2	-	n/a	ns
TRAF6	-	n/a	ns
WISP2	-	n/a	ns

Table 4-2 Correlation between GLI1 and a panel of known and potential HH target genes.
r^2 and p-values generated by linear regression analysis between GLI1 and target gene of choice. n/a = not applicable; ns = non-significant ($\alpha = 0.05$). '+' and '-' indicate positive and negative correlations with GLI1 respectively.

GENE	HC gradient	HC (N-I) vs. HC (I)	UC vs. HC	CD (colon) vs. HC (colon)	CD (TI) vs. HC (TI)
WNT2b	-	-	↑ p=0.044	-	-
WNT3	-	-	-	-	-
WNT4	-	-	-	-	↑ p=0.012
WNT5a	-	-	-	-	↑ p=0.0016
WNT5b	-	-	-	-	-
WNT6	↑ p=0.037	-	-	↓* p=0.0003	-
WNT9b	-	-	↑ p=0.022	-	-
FZD4	↑ p<0.0001	-	↑ p=0.030	↑ p=0.025	-
FZD5	-	-	-	↓ p<0.0001	↓ p=0.018
FZD6	-	↓ p=0.0071	↓ p=0.020	-	-
FZD7	↑ p=0.0007	↓ p=0.022	-	-	↓ p=0.0008
LRP5	↓ p<0.0001	↑ p=0.030	-	-	-
LRP6	↑ p=0.0026	-	-	↓ p=0.0010	-
sFRP1	-	-	-	-	-
sFRP5	-	-	-	-	-
DKK1	-	-	-	-	-
DKK2	-	-	-	-	-
DKK3	↑ p=0.0005	-	-	-	-
DKK4	-	↑ p=0.015	↑ p=0.0026	-	-
TCF1	-	↑ p=0.0012	↑ p=0.0010	-	-
TCF2	-	-	-	↓ p=0.0025	-
TCF3	↑ p=0.0018	-	-	-	-
TCF4	↑ p=0.028	-	↑ p=0.021	↑ p=0.0038	-
APC	↓ p=0.0017	↑ p=0.032	-	-	-
CTNNB1	↑ p=0.014	-	-	↑ p=0.029	↑ p=0.05
AXIN1	↓ p<0.0001	-	↓† p<0.0001	-	-
ATOH	-	-	↑ p=0.0017	-	-
GSK3α	-	-	↓ p<0.0001	-	↓ p=0.036
GSK3β	-	-	-	↑^ p=0.0007	↑ p=0.0076

Table 4-3 Summary of WNT pathway expression profiles in HC, UC and CD.
HC gradient is calculated by Krusall-Wallis test on expression values in ascending, descending and sigmoid colon. Non-inflamed (N-I) HC were analysed versus inflamed (I) HC samples, all colonic. UC analysis is presented for inflamed UC vs. HC (N-I) for all colonic samples unless where specifically indicated. CD analysis is separated in colonic and terminal ileal (TI) samples in CD and HC. All p-values (except HC gradient) are 2-tailed and calculated by Mann-Whitney U test.
† *Also significantly lower expression in UC(N-I) vs. HC (p=0.0002).*
* *Expression change between CD(N-I) and HC; no difference for CD(I) vs. HC.*
^ *Also significantly higher expression in CD(N-I) vs. HC (p=0.0001).*

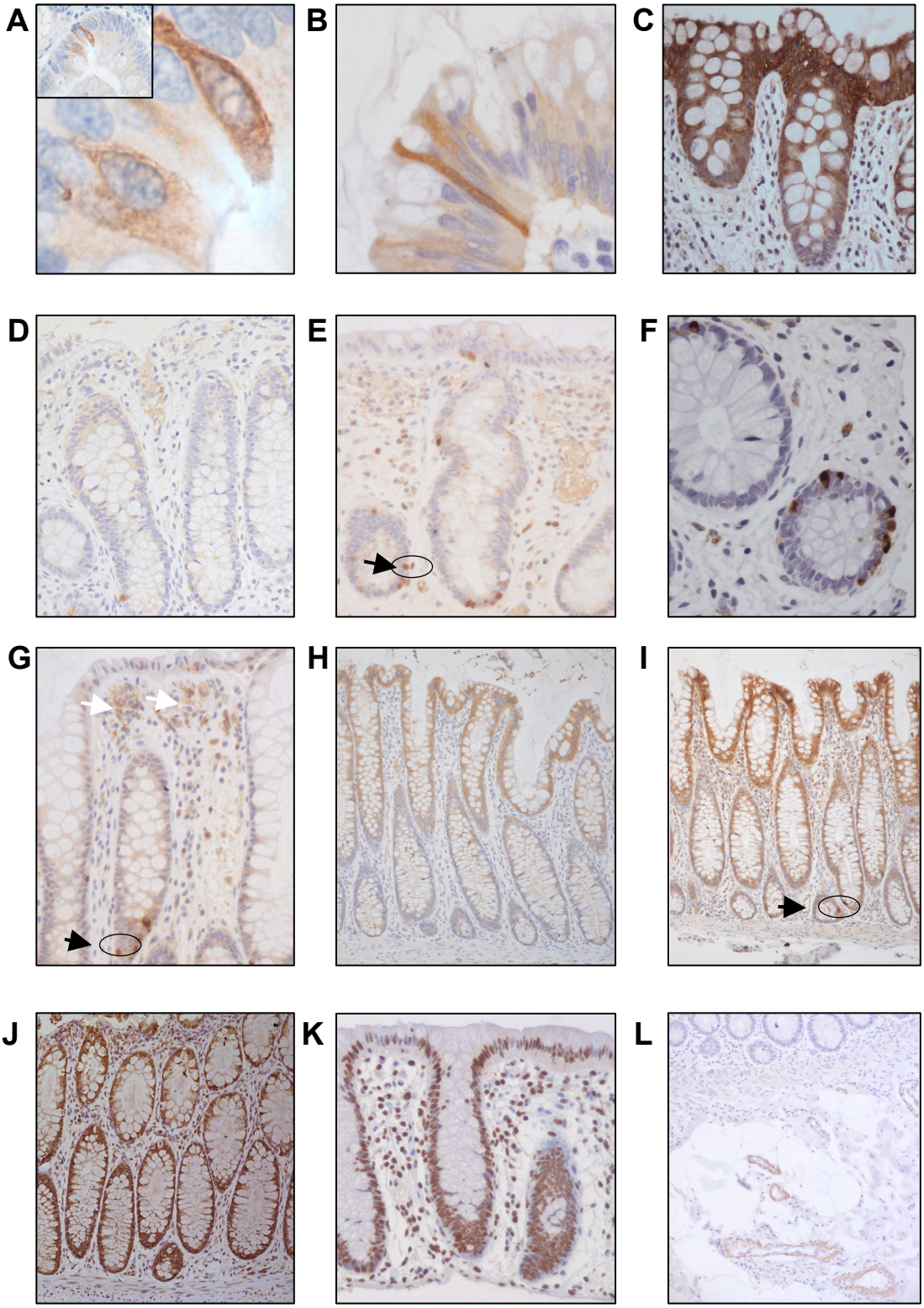

Figure 4-1 Protein expression of HH signalling components in healthy human adult colon.
Expression of SHH **(A-B)**, IHH **(D-F)**, PTCH **(G)**, SMO **(H)**, HHIP **(I)**, GLI1 **(J)**, GLI2 **(K)** and GLI3 **(L)** in the healthy human adult colon. SHH is expressed in selected enterocytes in the proximal colon (A-B) and throughout the epithelium in the distal colon (**C** and **Figure 4-2**). Weak epithelial expression of IHH is noted (**D**), although strong staining was noted in selected crypt cells in the distal colon (black arrows, **E-F**). The expression of PTCH (**G**) is predominantly in the mesenchyme (white arrow) but

selected crypt cells were strongly positive (black arrow). SMO (**H**) is expressed in the epithelium only. HHIP (**I**) is noted in epithelium and mesenchyme with strong staining in selected crypt cells (black arrow). GLI1 is expressed in epithelium and mesenchyme with predominant cytoplasmic staining (**J**), whilst GLI2 expression is exclusively nuclear (**K**) and GLI3 expression is limited to endothelial layers of blood vessels (**L**).

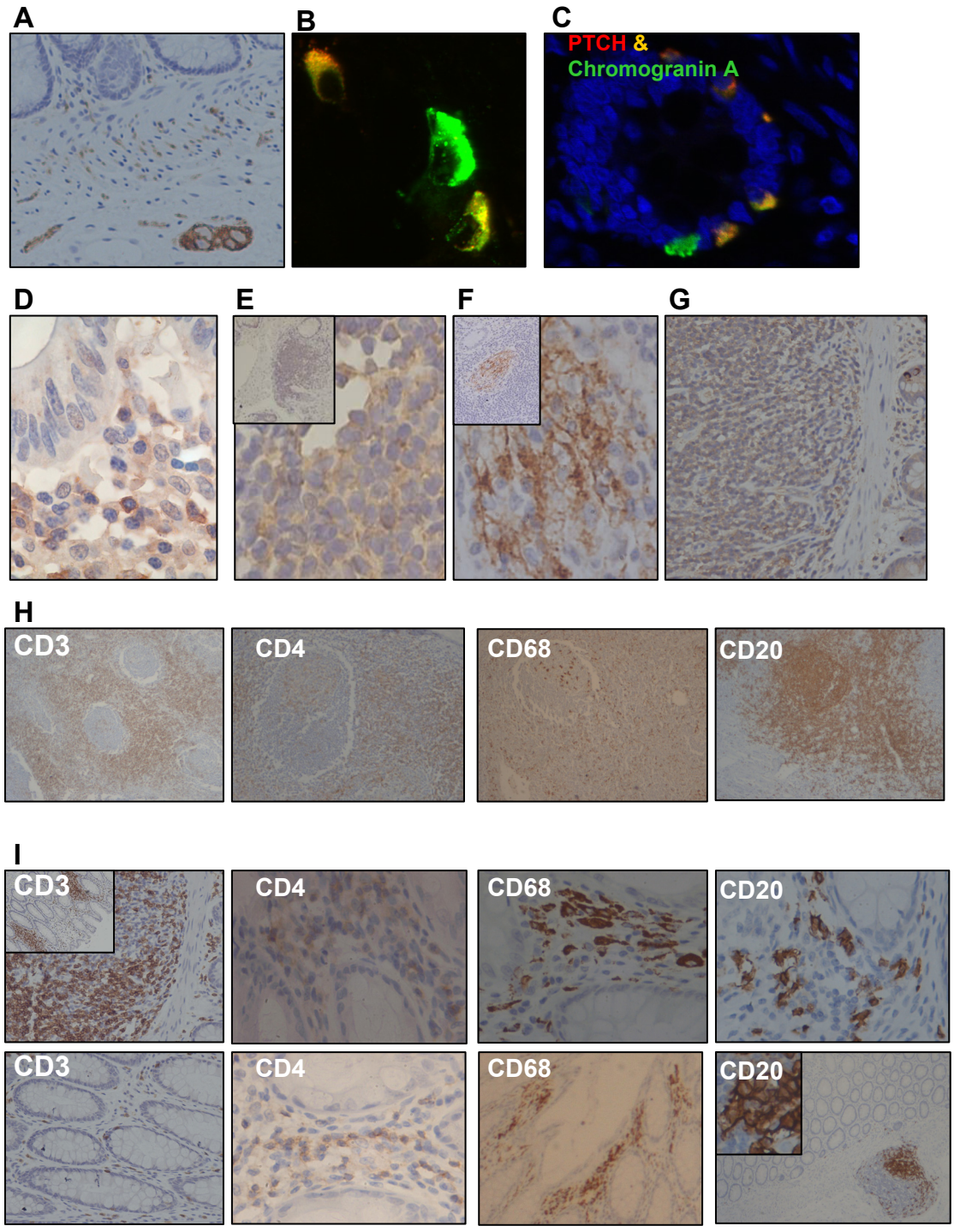

Figure 4-2 Detailed expression of PTCH protein in healthy human intestine.
PTCH, IHH and HHIP were seen to be expressed in a proportion of colonic crypt cells (**Figure 4-1**). Morphologically these had the appearance of neuro-endocrine cells. CD56 (**A**) was expressed in myenteric plexus only. Chromogranin A was expressed in neuro-endocrine cells in crypt bases. A proportion of these cells were PTCH positive, with co-localisation of PTCH and chromogranin A identified by con-focal microscopy (**B-C**). PTCH was expressed in the mesenchyme near the epithelium (**Figure 4-1** and **D**), and in germinal centres of lymphoid follicles (**E-G**). To identify immune subtypes in the colonic mesenchyme CD3, CD4, CD68 and CD20 expression was analysed. These antibodies were first optimised in human tonsil (**H**) and then in healthy human colon (**I**). The pattern of PTCH staining resembles macrophages identified by CD68 and is also consistent with CD3 staining.

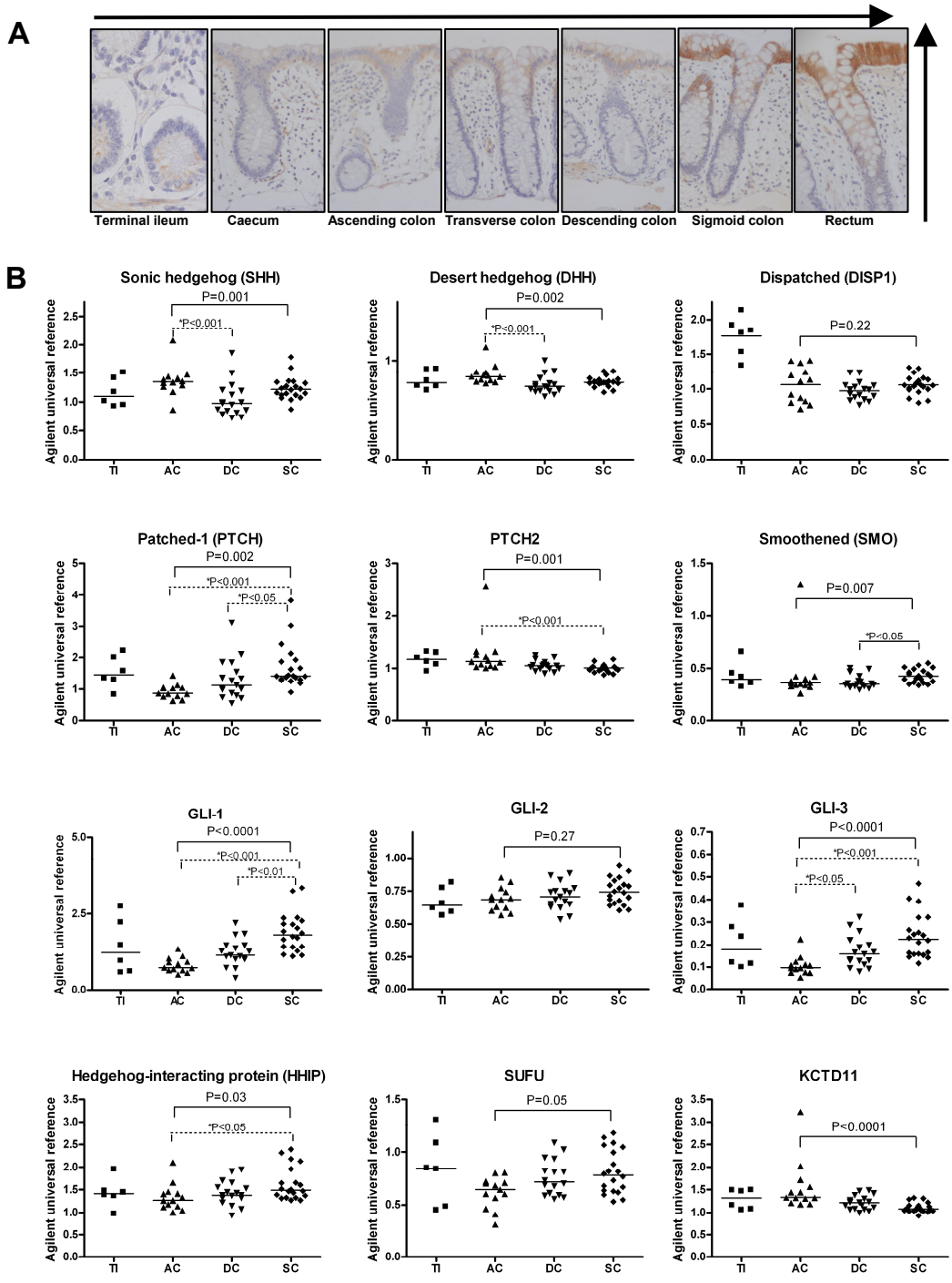

Figure 4-3 Gradients of HH expression along length of healthy human colon.
HH protein (N-19 antibody) expression (**A**) is greater in the distal colon compared with proximal colon (caecum) and terminal ileum (very weak expression in crypt bases only). mRNA expression (microarray data) for all other HH signalling components (**B**) is plotted in terminal ileum (TI), ascending colon (AC), descending colon (DC) and sigmoid colon (SC). For each dataset, p-values are given for Kruskal-Wallis test across AC, DC and SC (solid lines). Where $p<0.05$ a post-hoc Dunn's multiple comparison analysis was applied to compare all sets of columns (* indicated $p<0.05$ with dotted line to indicate columns compared). Of note SHH mRNA expression does not mirror that for protein, suggesting that IHH may be the predominant ligand in the colon (IHH not present on microarray chip). However, the gradient of increasing protein expression is paralleled by the HH response genes (GLI1, PTCH and HHIP, upregulated on pathway activation and thus predict pathway activity. KCTD11, a HH pathway inhibitor, is lower in distal compared with proximal colon.

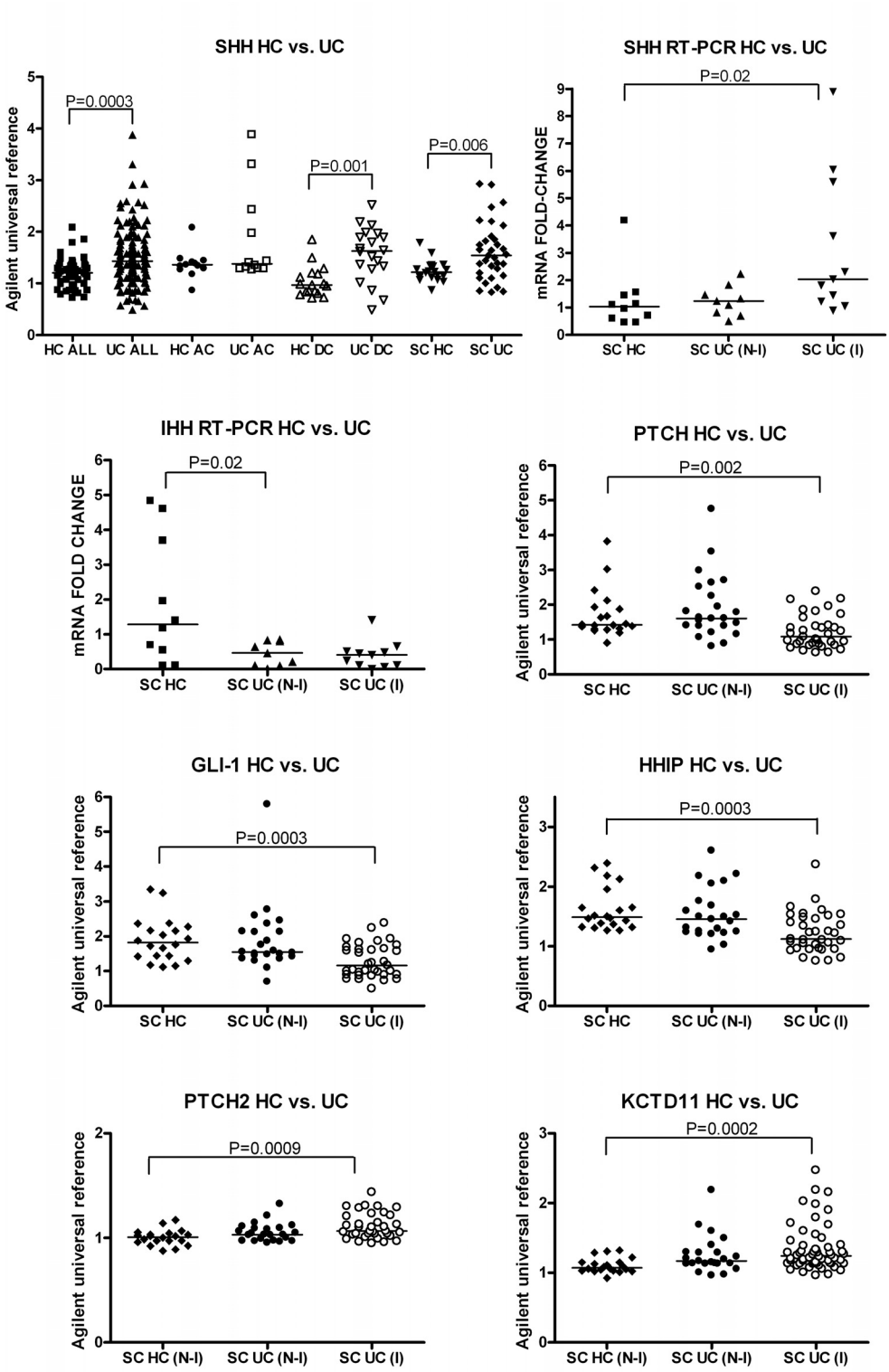

Figure 4-4 mRNA expression of HH pathway components in UC versus HC.
Quantitative analysis of mRNA levels of SHH, IHH, PTCH, GLI1, HHIP, PTCH2 and KCTD11 in UC compared with non-inflamed HC**.** To account for the gradients in healthy colon (**Figure 4-2**), SHH is plotted at each anatomical location studied (ascending (AC); descending (DC) and sigmoid colon (SC). These microarray data are further validated in SC by RT-PCR. Only SC RT-PCR data is presented for IHH as this gene was not present on the Agilent microarray chip. Microarray data for the other probes are presented for the SC only. Disease specimens, where indicated, are sub-categorised into non-inflamed (N-I) and inflamed (I) tissues. Individual data points are plotted with horizontal lines representing the medians for each dataset. P-values, plotted where significant ($p<0.05$) are derived from Mann-Whitney U-tests.

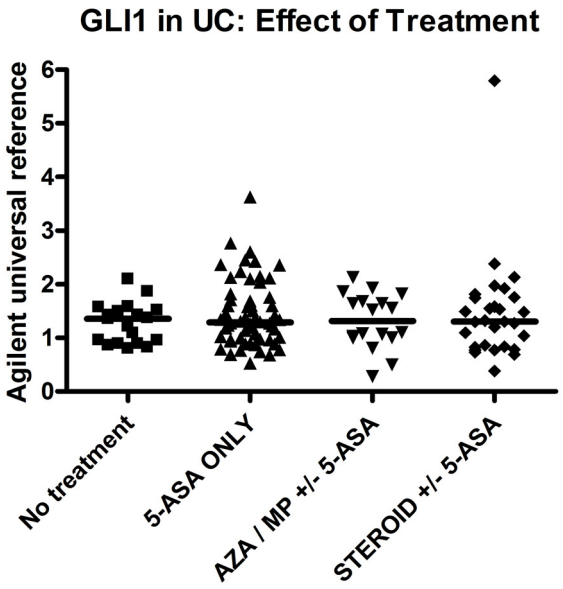

Figure 4-5 Analysis of GLI1 expression in UC by medical therapy
Data are presented for UC patients on no treatment (Rx), on a 5-amino salicylic acid (5-ASA) only, on azathioprine (AZA) or mercaptopurine (MP) with or without a 5-ASA and on steroids (oral and topical) with or without a 5-ASA. Individual data points are represented with horizontal lines representing medians. There was no difference between any of the groups on analysis by Mann-Whitney U testing.

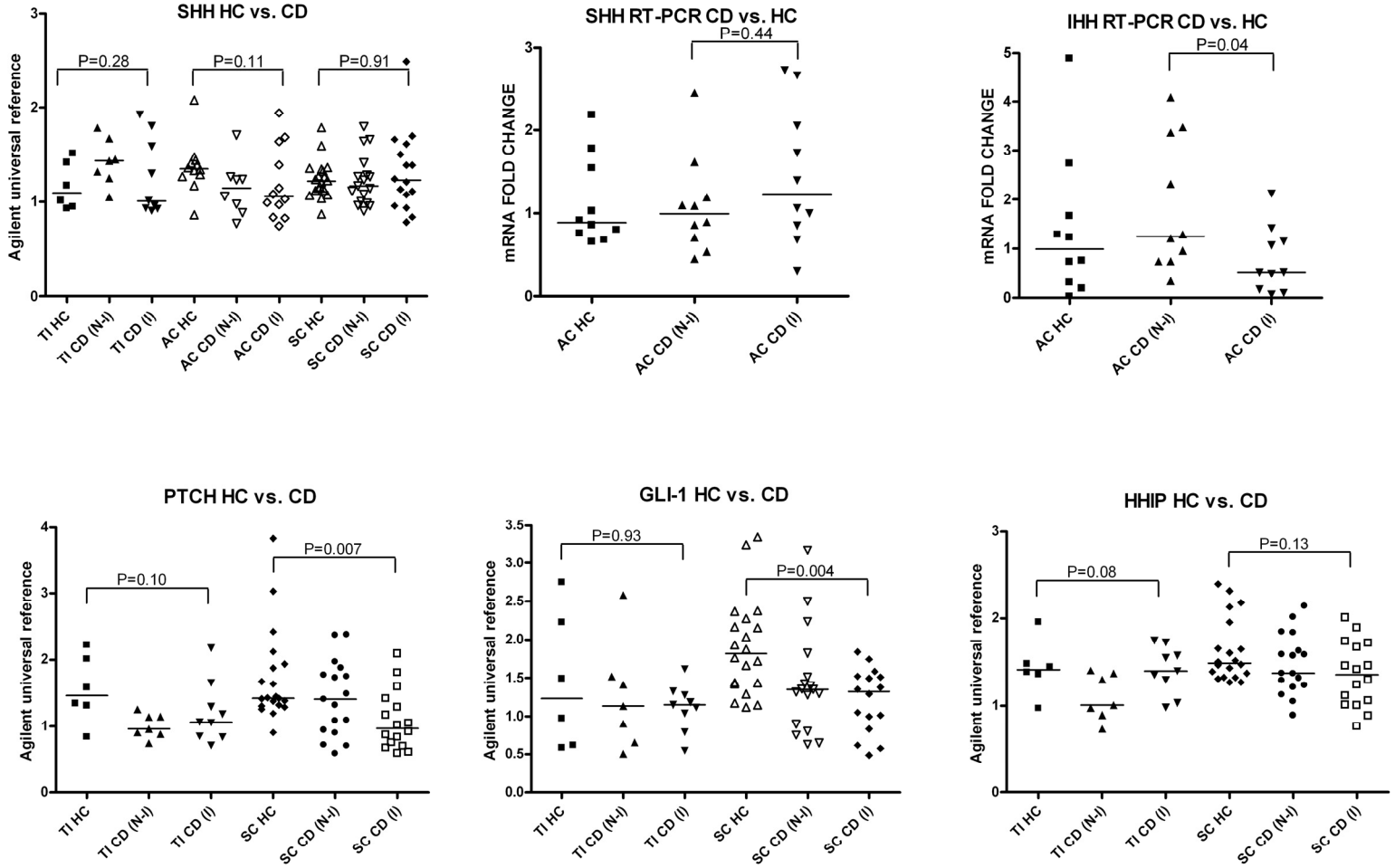

Figure 4-6 mRNA expression of HH signalling components in CD versus HC

Quantitative analysis of mRNA levels by microarray and RT-PCR of SHH, GLI1, PTCH, and HHIP in CD compared with non-inflamed HC. Disease specimens are taken from the terminal ileum (TI), ascending colon (AC) and sigmoid colon (SC), and are sub-categorised into non-inflamed (N-I) and inflamed (I) tissues. P-values represent Kruskal-Wallis analysis of HC, CD (N-I) and CD (I) for each anatomical location.

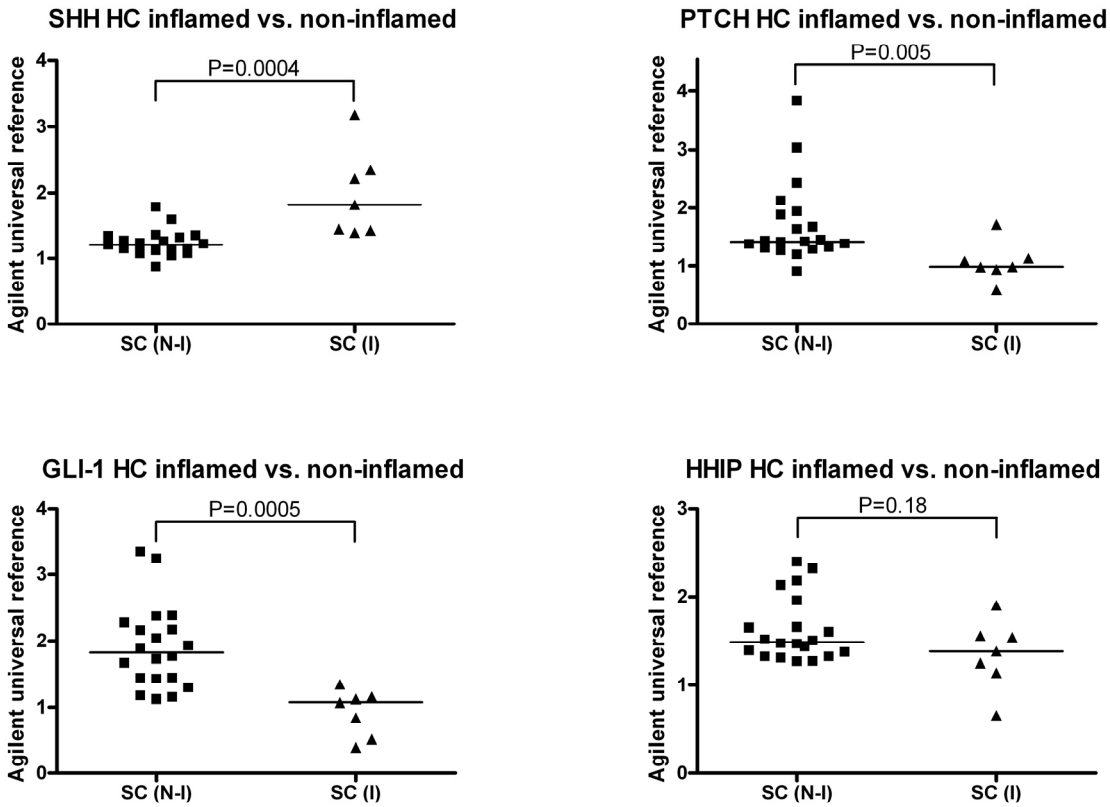

Figure 4-7 mRNA expression of HH signalling components in non-IBD colonic inflammation. Analysis of SHH, PTCH, GLI1 and HHIP mRNA levels by microarray in colonoscopic biopsies from the sigmoid colon (SC) of non-IBD patients with inflammation (I) compared with non-inflamed HC (NI).

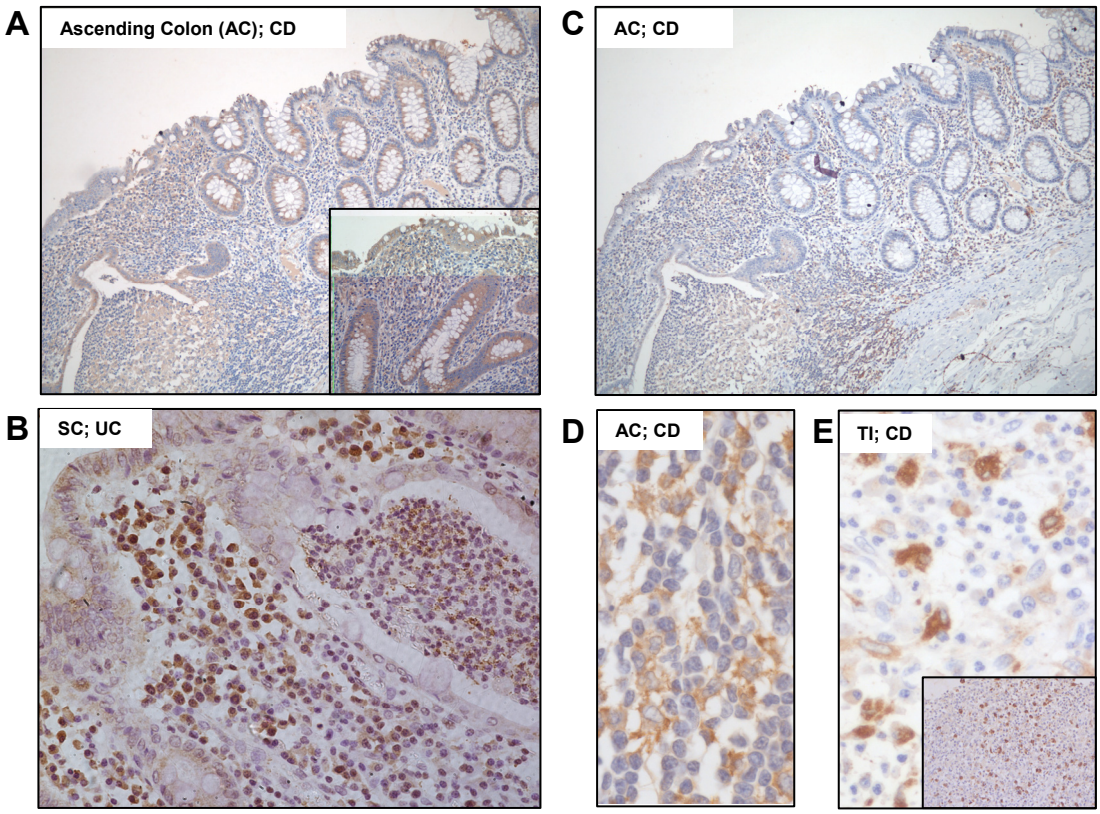

Figure 4-8 Protein expression of HH, PTCH and GLI1 in IBD.
Representative images demonstrate HH protein in the epithelium throughout the crypt length (**A**) in contrast to healthy colon where expression is restricted to the upper two-thirds of the crypt. There was, however, considerable variability in SHH staining; no expression was identified in some sections. GLI1 protein expression was notable in some UC specimens in plasma cells in the lamina propria (**B**). PTCH protein (**C-E**) is predominant in mesenchyme consistent with lymphocytes (**D**) and macrophages (**E**) as per staining of CD3 and CD68 in figure 3.

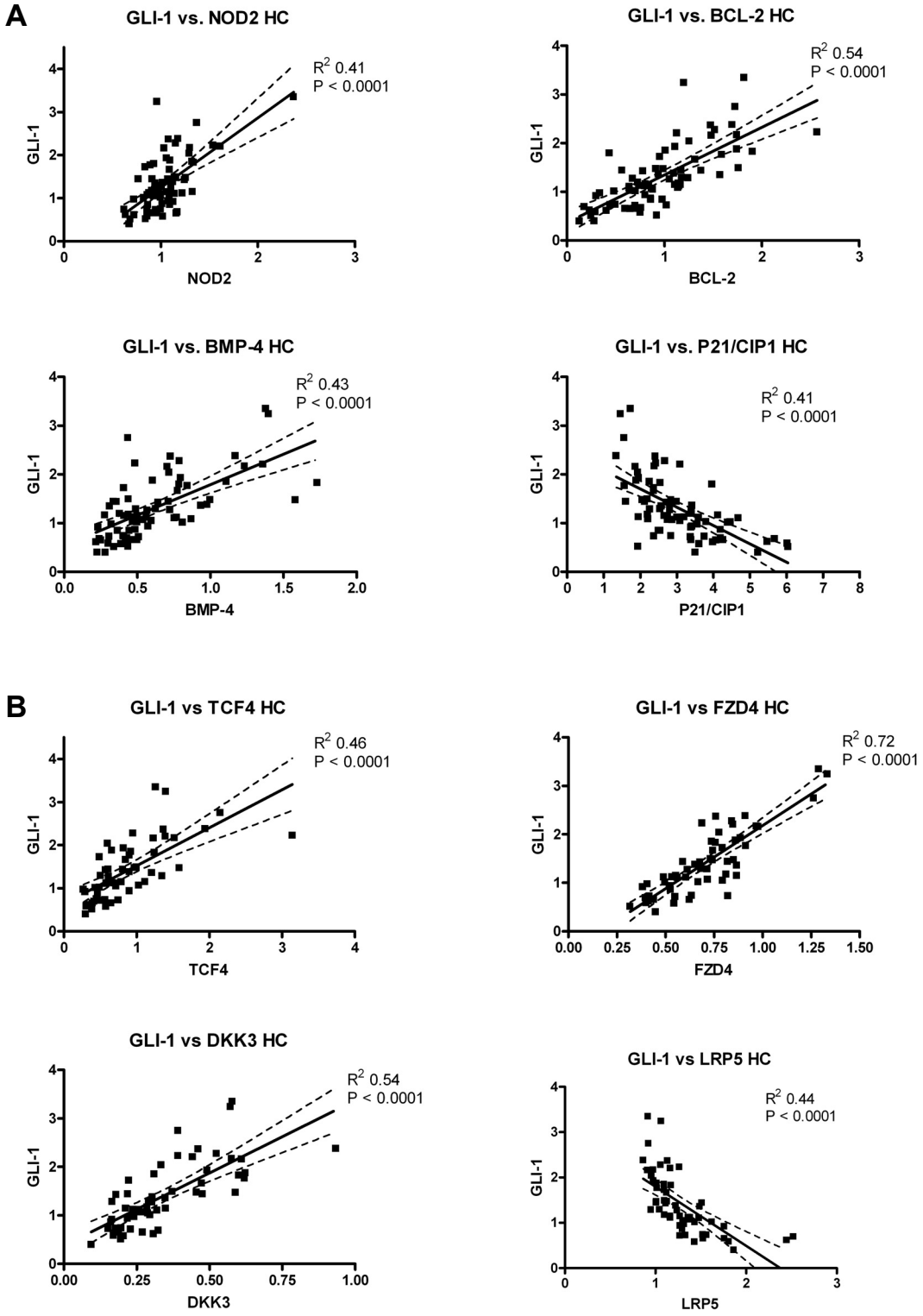

Figure 4-9 Correlation between HH pathway activity and known target genes.
Correlation between HH pathway activity (represented by GLI1 expression) in putative and known HH target genes (A) and WNT signalling components (B) in non-inflamed healthy colon. There is a strong positive correlation with NOD2, BCL-2, and BMP-4 with a negative correlation with P21/CIP1 (**A**). Analysis of all known WNT signalling components was performed in relation to GLI1 expression. Significant results are plotted, including positive correlations with TCF4, FZD4 and DKK3 and negative correlation with LRP5. Linear regression analysis is plotted with 95% confidence intervals for each analysis.

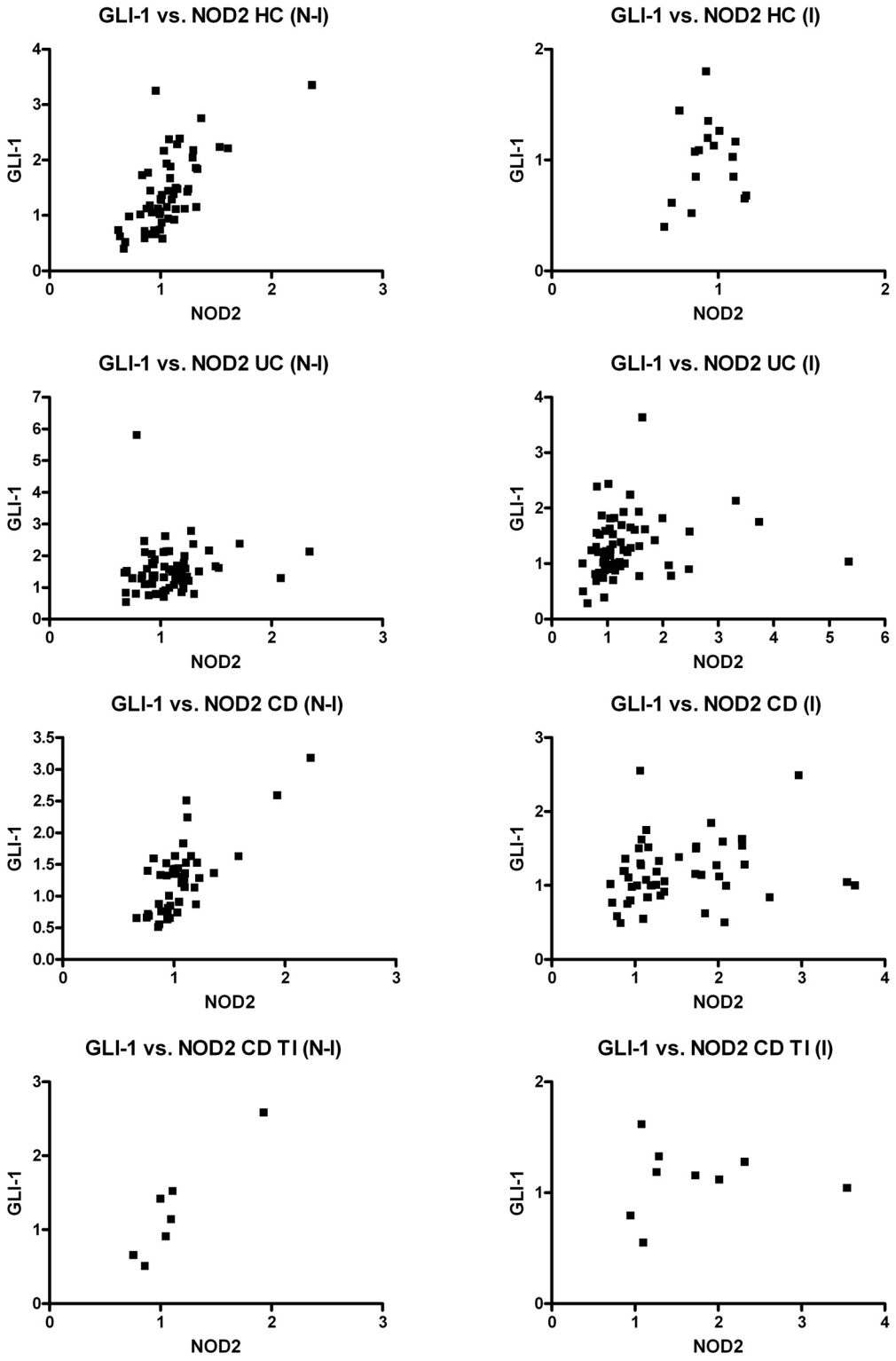

Figure 4-10 Correlation between GLI1 and NOD2 expression profiles.
Correlation between GLI1 and NOD2 mRNA expression profiles in non-inflamed (N-I) and inflamed (I) tissue from all HCs, all patients with UC, all patients with CD and those CD biopsies from the terminal ileum (TI) only.

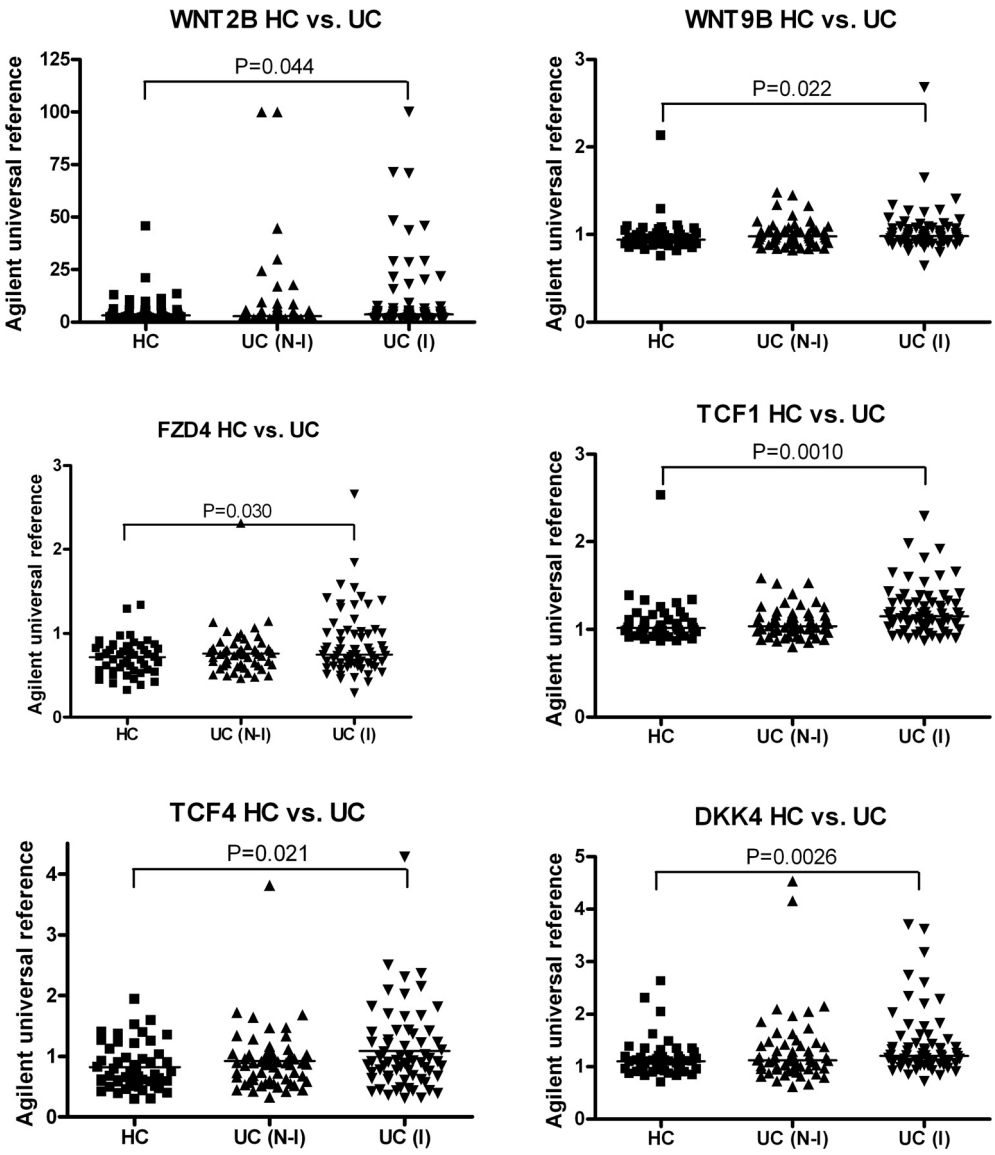

Figure 4-11 Expression of selected WNT signalling components in UC versus HC.
The expression of those WNT signalling components for which significant differences were noted between UC and HC are plotted. The full data series for the WNT pathway genes is presented in summary form in **Table 4-1** and in complete graphical form in the appendix (**Figures 13-1 – 13-15**) Two-tailed p-values comparing inflamed UC samples with non-inflamed HC samples (all from the colon) are calculated by Mann-Whitney U test.

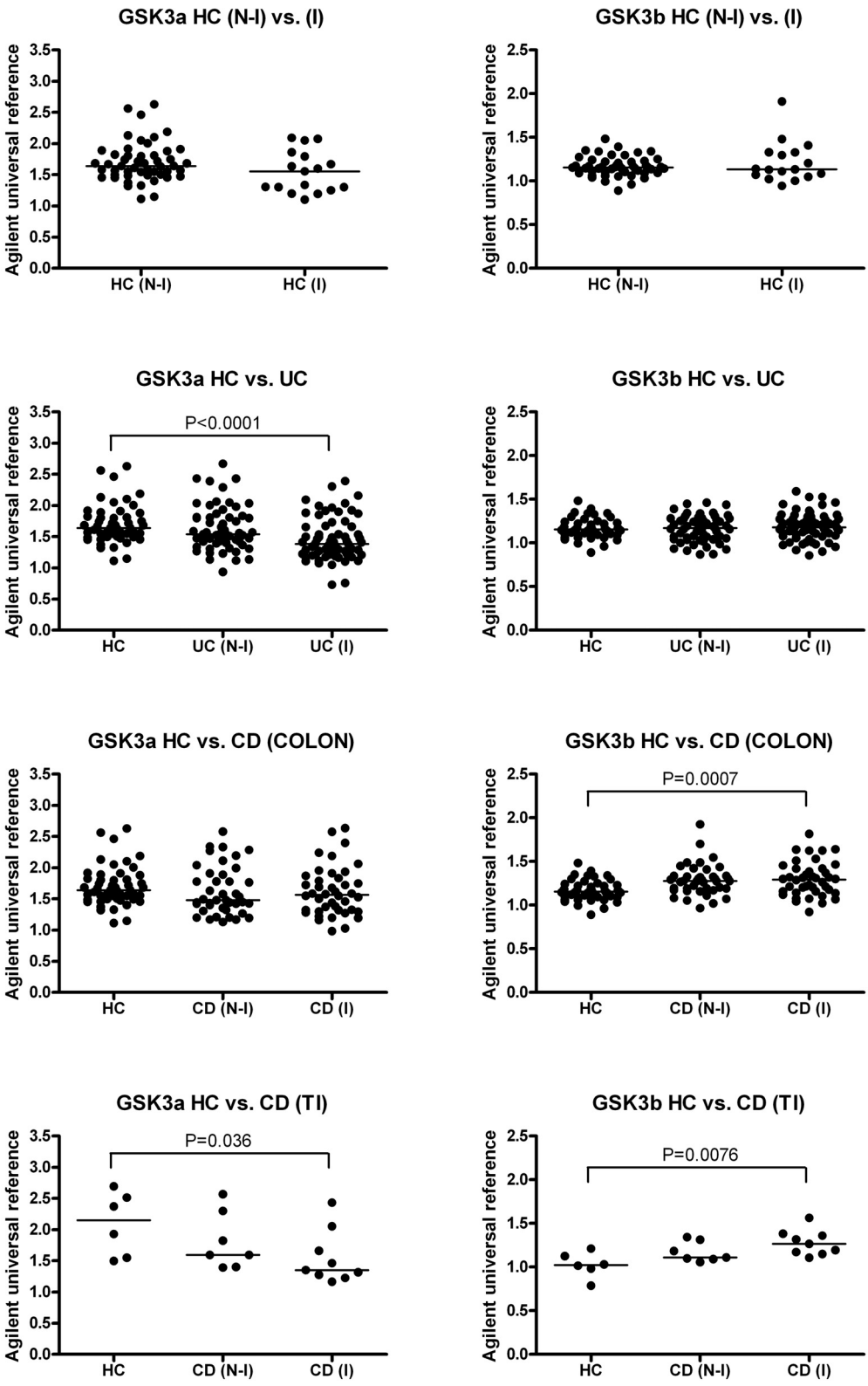

Figure 4-12 Expression of GSK3α and GSK3β in HC, UC and CD.
There was no gradient of expression of either GSK3 gene along the length of the healthy colon. Individual data points are plotted with horizontal lines representing the medians for each dataset. P-values, plotted where significant ($p<0.05$) are derived from Mann-Whitney U-tests.

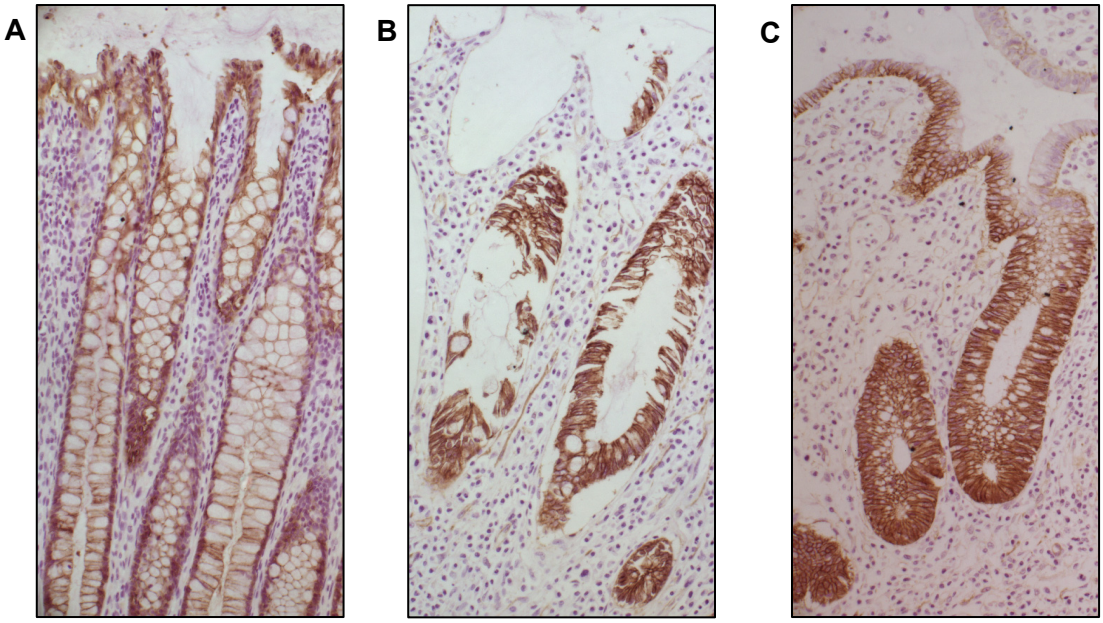

Figure 4-13 Protein expression of β-catenin in UC and healthy colon.
The expression of beta-catenin in the healthy colon (**A**) and UC (**B-C**).

5 Gene-wide haplotype-tagging study of *GLI1* in IBD.

5.1 *Abstract*

Background. Genome-wide linkage analysis has provided compelling evidence for a genetic susceptibility locus determining IBD susceptibility/phenotype within 12q13 (IBD2). The HH signalling pathway plays vital roles in gastrointestinal tract development and homeostasis, Paneth cell differentiation, inflammation and malignancy. *GLI1*, a key HH transcription factor, lies within the IBD2 linkage region, leading us to hypothesise that *GLI1* might play a role in disease causation. In detailed expression studies in Chapter 4 it was demonstrated that that IHH – PTCH – GLI1 signalling is down-regulated in UC, supporting a causal role for HH dysregulation in disease pathogenesis.

Aims. The aim of the present study was to establish whether germline variation in *GLI1* was associated with IBD.

Methods. A gene-wide haplotype-tagging approach was employed, with 4 multi-marker tSNPs selected in HaploView on the basis of sequencing across coding sequencing in 32 patients and controls. These were genotyped by TaqMan in 4 separate populations of IBD patients and controls, from Scotland, Cambridge (England), Sweden and Japan. Log-likelihood, haplotype, individual SNP analysis and meta-analysis of associated SNPs (Mantel-Haenszel) was performed.

Results. Germ-line variation in *GLI1* was associated with IBD in the Scottish population (n=2256, IBD vs. HC: $p<0.0001$; UC vs. HC: $p<0.0001$ on log-likelihood analysis). This effect was confirmed in an independent English population (n=2326, IBD vs. HC: $p=0.009$; UC vs. HC: $p<0.0001$). This association was confined to *GLI1*, and attributable to a non-synonymous SNP in exon 12 of *GLI1* (rs2228226C→G), which encodes an amino acid change at Q1100E. On meta-analysis of Scottish, English, and Swedish populations (n=5352), rs2228226 was associated with IBD with pooled odds ratio of 1.194 (C.I. 1.09-1.31, $p=0.0002$) and UC (OR 1.196, C.I. 1.077-1.327, $p=0.0008$).

Conclusions. These findings implicate common genetic variation in *GLI1*, a key component of the HH signalling pathway, in the pathogenesis of IBD as an IBD2 gene.

5.2 *Introduction*

The IBD2 locus (OMIM 601458), spanning 25 Mb on chromosome 12q13, was identified on a UK genome-wide linkage scan in 1996 (peak LOD score 5.47 at D12S83)[99] and has subsequently been most strongly implicated in UC (see **1.5.1.3**).[97, 98] Located within IBD2 is the gene encoding GLI1, a zinc finger protein that is the major transcriptional regulator of the HH signalling pathway (see **1.6.1** and **Figure 1-6**).

In Chapter 4, it was shown that HH expression is dysregulated in IBD, with lower levels of GLI1 expression in UC compared with HC. Based on this expression data, the function of the HH signalling pathway (see **Figure 1-8** and **Table 2-1**) and the position of *GLI1* in IBD2, it was hypothesised that inherited variation in *GLI1* is associated with the development of IBD. A gene-wide haplotype-tagging SNP approach was adopted to examine the contribution of the *GLI1* gene to IBD pathogenesis. These genetics studies have involved three independent Northern European populations (Scotland, England and Sweden) totalling 5352 individuals.

5.3 *Methods*

5.3.1 Subjects for genotyping

Scotland: The population consisted of 809 IBD cases (474 UC; 335 CD) and 1364 controls (**Table 5-1**). The population was 98.5% white, non-Jewish Caucasian. The median age at diagnosis was 31.0 (IQR 23.3-46.0). Controls were recruited from across Scotland; median age was 50.0 (IQR 43.0-55.0), ethnicity >99% white, non-Jewish Caucasian, and 48.7% male sex.

Low levels of genetic differentiation in terms of allele frequencies across mainland Scotland have previously been reported.[442] Since it has been shown that low levels of population stratification are likely only to result in weak associations,[443] it is unlikely that the low levels present in the Scottish population could account for the strong associations between *GLI1* haplotype and disease risk reported herein.

Cambridge, England: 1767 unrelated Caucasian IBD patients of north European origin attending IBD clinics in East Anglia, UK were recruited comprising 928 UC, 756 CD, and 83 with indeterminate colitis (IC) (**Table 5-1**). The population was

>99% non-Jewish. Median age at diagnosis was 36.7 and 26.1 years for UC and CD respectively. The 637 ethnically and geographically matched healthy controls were previously recruited in East Anglia for the European Prospective Investigation of Nutrition and Cancer (EPIC). Median age of controls was 60 years.

Sweden: 288 UC and 205 CD patients were recruited from various hospitals in Stockholm County (**Table 5-1**). 281 Swedish HC were healthy volunteers from the Karolinska University Hospital staff (n=170) and orthopaedic day-case surgery patients (n=111) with no previous medical conditions. The details of this cohort are given in methods chapter.

5.3.2 Sequencing

Sequencing of *GLI1* was performed in 32 patients with UC and CD, initially targeting known SNPs from Ensembl to enable selection of tagging SNPs (this stage performed prior to on-line publication of HapMap), and subsequently all 12 exons (including intron-exon boundaries) and 1kb of the promoter region was sequenced to search for novel mutations for genotyping (see **Table 3-2** for details of primer pairs used).

5.3.3 Genotyping

Scotland and Sweden: Genotypes were derived using the Taqman system for allelic discrimination; the assays were available from Applied Biosystems as Taqman SNP Genotyping Assays (7900HT sequence detection system; Applied Biosystems, Foster City, CA), except for SNPs rs10783819, rs3809114, rs507562, rs542278, rs730560, rs1669296, rs775322 which were genotyped on the Illumina platform. The accuracy of each Taqman assay was checked by repeat analysis in 5% of cases, with 100% concordance. Genotype distributions in control populations were consistent with Hardy-Weinberg Equilibrium ($p>0.01$) for all SNPs. Genotypes, in cases for the four tSNPs that could not be derived by TaqMan were obtained by direct sequencing.

Cambridge: DNA was extracted using Nucleon kits. Genotyping of CD cases and controls was performed using the Taqman biallelic discrimination system using an ABI 7900HT analyser; Applied Biosystems, Foster City, CA. Genotyping of UC cases was performed using a 1536 SNP Golden Gate bead array; Illumina, San Diega, CA. Concordance between platforms was assessed by genotyping 92 UC cases for SNPs rs2228224 and rs2228226 with concordance rates of 100% and 97.9% respectively and no evidence of systematic bias between platforms. Genotype

distributions in case and control populations were consistent with Hardy-Weinberg Equilibrium ($p>0.01$) for all SNPs.

5.3.4 Statistical analysis

Haplotype frequencies of the tSNPs were inferred using the expectation-maximization algorithm and imputed 10,000 times using the log-likelihood analysis described in Chapter 3 (**3.7.2**). Haplotypes were examined using the Haploview programme vers. 3.2 (www.hapmap.org). The meta-analysis of SNP rs2228226 was performed using the Mantel-Haenszel method using a fixed effects model (R software package; assistance from A Tenesa) (**3.7.6**).

5.4 *Results*

5.4.1 Gene-wide variation in *GLI1* is associated with IBD and attributable to a non-synonymous SNP (rs2228226C→G) in the Scottish population

Sequencing of the *GLI1* gene in 32 IBD cases, informed by Ensembl, confirmed the presence of variants from dbSNP in the Scottish population. Four multi-marker tagging SNPs (tSNPs; $r^2 \geq 0.8$) were identified in Haploview (rs3817474, rs2228225, rs2228224, and rs2228226), that described haplotypic variation of *GLI1*, detecting haplotypes of a frequency >1% (**Figure 5-1**). These 4 tSNPs were genotyped in a Scottish IBD population consisting of 474 UC and 335 CD cases, and 1364 well-matched healthy controls (**Table 5-1**). A model-free analysis was then employed to test the association of *GLI1* and IBD susceptibility. Haplotype frequencies were estimated separately in cases and controls and on the whole sample. The likelihood of the dataset assuming the same frequencies in cases and controls was compared with the likelihood when the frequencies were estimated separately in cases and controls. Empirical significance levels were obtained by permutation testing.

In the Scottish population, a highly significant association was demonstrated in IBD ($p<0.0001$) and UC ($p<0.0001$), and an association with CD of borderline significance ($p=0.03$) (**Table 5-2A**). On analysis of individual estimated haplotype frequencies in Haploview, 3 common haplotypes were described (**Table 5-2A**). Haplotype B (2222) was a risk haplotype for IBD (OR 1.16, C.I. 1.01-1.32, $p=0.036$), whereas there was a

protective effect with haplotype C (2221; OR 0.71, C.I. 0.54-0.93, p=0.014) (**Table 5-2A**). To determine whether this effect was confined to the *GLI1* gene an additional 7 haplotype-tagging SNPs, chosen from Phase II HapMap data to tag neighbouring blocks, were genotyped in 166 CD and 170 UC patients. This confirmed the presence of a *GLI1* spanning haplotype block that did not extend into neighbouring genes (*INHBE* and *ARHGAP9*) (**Figure 5-2**).

The association on haplotype testing and log-likelihood analysis was attributable to a non-synonymous SNP in exon 12 of *GLI1* (rs2228226C→G; $tSNP_4$), which encodes an amino acid change (glutamic acid to glutamine) at position E1100Q. rs2228226 was associated with IBD (allelic frequency OR=1.23, C.I. 1.07-1.40, p=0.0026; homozygotes OR=1.56, C.I. 1.15-2.11, p=0.0047), CD (allelic frequency OR=1.30, C.I. 1.08-1.55, p=0.0053, homozygotes OR=1.79, C.I. 1.21-2.65, p=0.0048) and UC (allelic frequency OR=1.19, C.I. 1.01-1.39, p=0.04) (**Table 5-3**). Mutation screening of the *GLI1* coding regions by direct sequencing failed to identify any novel SNPs. Therefore, in addition to the 4 tSNPs above a further 7 *GLI1* variants, selected as validated SNPs from dbSNP, were genotyped in the Scottish cohort. There was no association between any of these SNPs and IBD susceptibility, suggesting that rs2228226 is the causative *GLI1* variant (**Table 5-4**). Despite tight linkage disequilibrium (LD) across *GLI1* (**Figure 5-2**), there was notably less LD between rs2228226 and other tSNPs ($r^2 = 0.7$) in both cases and controls (**Figure 5-3**), and this is likely to provide the explanation for the lack of association in neighbouring SNPs.

5.4.2 Replication of *GLI1* association in two independent North European IBD cohorts

In order to replicate these findings, a large IBD panel from Cambridge, England was examined (n=1000 UC, 737 CD and 589 HC). The association with *GLI1* was replicated by log-likelihood analysis in IBD (p=0.009), UC (p<0.0001) and CD (p=0.002) (**Table 5-2B**). Haplotype C was present in the Cambridge populations with very similar frequencies to that in Scotland, and was protective for IBD (OR 0.76, C.I. 0.58-0.99, p=0.049) and UC (OR 0.66, C.I. 0.49-0.90, p=0.0093). Furthermore, rs2228226 was associated with IBD (OR 1.17, C.I. 1.00-1.36, p=0.042) and UC (OR 1.24, C.I. 1.05-1.46, p=0.011) but not CD in this population (**Table 5-3**). As in Scotland, there was no association with $tSNPs_{1-3}$ (**Table 5-5**).

In the Swedish cohort, there was a similar magnitude of effect with haplotype C in association with IBD (OR 0.75, C.I. 0.50-1.10), UC (OR 0.78, C.I. 0.50-1.21) and CD (OR 0.68, C.I. 0.41-1.12) (**Table 5-2C**) and with rs2228226 with IBD (OR 1.14, C.I. 0.90-1.45), UC (OR 1.11, C.I. 0.90-1.45) and CD (OR 1.19, C.I. 0.90-1.58), albeit without significance in this smaller cohort (n=770; **Table 5-3**).

Finally, a smaller Japanese cohort (n=300) was genotyped for the 4 tSNPs. There was a significant difference in HC minor allelic frequencies compared with the N European cohorts. Whilst analysis of HC vs. IBD, UC and CD was not significant, a similar trend existed for rs2228226 as for the other cohorts (**Table 5-3**).

5.4.3 Meta-analysis of rs2228226C→G

A meta-analysis, using the Mantel-Haenszel method with a fixed effects model on the IBD cases and healthy controls in Scotland, England and Sweden confirmed the association with rs2228226 (OR 1.194, C.I. 1.089-1.309, p=0.0002), achieving criteria for significance in a gene-centric study (**Figure 5-4A**).[192, 444] There was no evidence of heterogeneity in the contribution of rs2228226 between the 3 cohorts (p=0.825). The meta-analysis changed little with the addition of the smaller Japanese cohort (OR 1.188, C.I. 1.087-1.298, p=0.0001). The meta-analysis was repeated separately for UC cases versus healthy controls with a similar effect size noted (OR 1.196, C.I. 1.077-1.327, p=0.0008) (**Figure 5-4B**).

Recognising the current problem with the publication of false positive findings in genetic association studies we estimated the probability that the association with disease risk found in the meta-analysis of *GLI1* SNP rs2228226 represents a true (rather than false positive) association by adopting the false positive report probability (FPRP) approach described by Wacholder.[433] This gives an estimated probability that these findings represent a true finding of at least 92% (FPRP < 0.08). This method is designed to avoid over-interpretation of statistically significant findings that are not likely to signify a true positive but in the present study gives clear support to the interpretation of these data.

5.4.4 Genotype-phenotype analysis

On detailed haplotype-phenotype analysis in the Scottish population, the risk haplotype B (2222) was associated with left-sided (E2) and extensive UC (E3) compared with ulcerative proctitis (E1) (33.8% vs. 24.7%, p=0.034, OR 1.56, C.I. 1.05-2.32). Genotype-phenotype analysis of rs2228226 was consistent with this (E2+E3 35.1% vs. E1 26.0%, p=0.037, OR 1.54, C.I. 1.04-2.28). There was no association with disease location in CD, or with disease behaviour, surgery, family history or age at diagnosis in UC or CD (data not shown). None of these phenotypic associations were replicated in the Cambridge cohort (data not shown). Genotype-phenotype analysis was not performed in Swedish or Japanese cohorts due to limited sample size.

5.5 *Discussion*

Following on from the extensive expression studies described in Chapter 4, the data presented here provide the first evidence that intact HH signalling is critical in the mammalian gut response to inflammatory challenge, and that down-regulated GLI1 signalling is implicated in IBD pathogenesis. The inherited variation in the *GLI1* gene that we have detected is associated with IBD and UC, in both Scotland and England, with findings in individual cohorts confirmed by meta-analysis of over 5000 individuals, in 3 N European populations with a pooled odds ratio of 1.19. Evidence for an effect in CD is seen in the present study, but the predominant effect is clearly related to UC, consistent with the reported IBD2 linkage studies.[98, 99] The magnitude of this association is entirely in line with the effect size noted in a number of recent studies of complex disease genetics, including CD,[103] colo-rectal cancer,[162] and coeliac disease.[445] The level of significance attained satisfies suggested criteria of $p < 10^{-4}$ - 10^{-6} for gene-centric studies.[192, 444] Population heterogeneity has been previously described at the IBD2 linkage region.[98, 99, 152, 446] The three N European populations studied here have previously demonstrated similar contribution of other IBD susceptibility genes / loci, including *NOD2* and IBD5.[78]

Whilst the minor allelic frequencies for rs2228226 are very similar in Scotland and Sweden (30.3% and 30.6%), they differ by 3.9% between Scotland and Cambridge (30.3% and 26.4%). This difference is in keeping with that noted for a number of

SNPs analysed for population stratification in the recent WTCCC study.[81] It has been demonstrated that the *GLI1* effect is not secondary to LD across the IBD2 region.

rs2228226C→G, a mis-sense mutation in exon 12 of *GLI1*, encodes a change from glutamine to glutamic acid (Q1100E). The mutation falls within a highly conserved motif at the C-terminus of all mammalian GLI1 proteins, near a recognized transactivation domain (**Figure 5-5**).[447] The Q1100 residue is itself 100% conserved in all mammals examined. *In situ* mutagenesis studies, performing in collaboration with the Gumucio lab in Ann Arbor, demonstrate that this mutant protein (1100E) is a subfunctional transcriptional activator compared with WT GLI1 protein, although it is appropriately synthesised and localised. In a GLI luciferase reporter assay, the mutant protein demonstrated a 50% reduction in activity compared with WT protein. The Q1100E mutation causes a significant charge change in a conserved region directly adjacent to the known transactivation region of GLI1; this change could directly modify transactivation activity, disrupt the structure of the transactivation domain, or affect protein stabilization,[448] decreasing activity.

Additionally, the SNP and mutation analysis of human *GLI1* indicate that rs2228226C→G is the only GLI1 coding change observed in the study population. Taken together, these data suggest that Q1100E could be the causative *GLI1* variant associated with IBD. Whilst the resequencing efforts identify rs2228226 as the only coding variant associated with IBD, further deep re-sequencing will be important in the future as the haplotype analysis and log-likelihood analyses raise the possibility that other germ-line variants may also contribute to IBD risk. These need be explored formally – specifically the role of intronic variants, long-range promoter effects and / or CNVs. In this context, several complex disease genes, including *NOD2*,[79, 80] have multiple independent mutations conferring disease risk, some disease genes have no causative mutations within coding sequences (e.g. *IRGM* in CD),[89, 213] and synonymous SNPs may be associated with functional effects.[449]

It is of note that germline variants in other HH pathway components are associated with human disease. Heterozygous *SHH* mutations are the most common genetic lesion associated with holoprosencephaly (HPE).[450, 451] Loss-of-function *GLI2* mutations are associated with HPE-like features along with pan-hypopituitarism.[452]

Mutant *PTCH* leads to constitutive pathway activation in basal cell naevus syndrome (Gorlin's syndrome), characterised by multiple benign and malignant tumours.[350] Mutations in *GLI3* cause Greig cephalopolysyndactyly syndrome and Pallister-Hall syndrome.[453, 454] A rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism is caused by mutations in *GLIS3* (GLI similar 3).[455]

Most recently, germline variation in *IHH*, *PTCH* and *HHIP* has been associated with height in two large meta-analyses of GWAS data.[159, 160] In phase I of their study, Weedon and colleagues analysed data from 13,665 individuals of European ancestry from 5 separate GWAS datasets all performed on the Affymetrix 500k chip, including the WTCCC study.[160] They then took 39 SNPs to replication in a further 16,482 individuals (in phase II). 20/39 SNPs reached GWAS levels of significance in the combined analysis, including SNPs in *PTCH* (rs10512248; $p_{combined} = 4.2 \times 10^{-11}$), *HHIP* (rs6854783; $p_{combined} = 2.1 \times 10^{-9}$) and *IHH* (rs6724465; $p_{combined} = 2.1 \times 10^{-8}$). Lettre and colleagues performed a 3 phase study, initially in 15,821 individuals from 6 separate GWAS.[159] Imputation of SNP data was required prior to meta-analysis as the individual GWAS were performed on different platforms (Affy 500k, Illunina 317k and Illumina 500k). 78 SNPs from phase I were replicated in a panel of 2189 European Americans taken from the near-extremes of the normal height distribution (short: 5-10th centile; tall: 90-95th centile). 29/78 SNPs were then replicated in 4 further validation cohorts totalling 17,801 individuals. Of note, *HHIP* was also associated with height variation in their study (rs1492820; $p_{combined} = 1.2 \times 10^{-11}$). Together, these data strongly implicate IHH signalling in the regulation of human height. It is noteworthy that *Ihh*$^{-/-}$ mice and TG mice with overexpression of *Hhip* in cartilage, have very similar phenotypes. Severe skeletal defects are noted, including short-limbed dwarfism.[289, 456]

Intriguingly, whilst most Hh pathway knock-outs are fatal during mouse development or peri-natally[270] (*Shh* and *Ihh* knockouts have shortened, malrotated digestive tracts with abnormal epithelial proliferation),[299] *Gli1*$^{-/-}$ mice appear to develop normally under standard conditions (**Table 1-2**).[313] The apparently normal phenotype has not to date involved a detailed description of the intestinal microstucture.[313] Furthermore, stressing this model (e.g. with DSS) or crossing it with strains susceptible to colitis (e.g. *Il10*$^{-/-}$) may provide important insights into pathogenic mechanisms. This

strategy has been successful employed with *Nod2*$^{-/-}$ and *Nod2*^{2939iC} mice (**1.5.1.1**).[115, 116]

After hypothesising that GLI1 is largely functionally redundant in development but plays and important role in response to injury (CWL), we have now tested this experimentally in close collaboration with Zacharias and Gummucio (Ann Arbor) (Lees, Zacharias *et al*, *PLoS Medicine* 2008).[423] Intriguingly, we have shown that *Gli1*$^{+/LacZ}$ mice develop early, severe colitis with high mortality when stressed with DSS compared with WT (further details in Chapter 12). These data provide robust support for a critical role for HH signalling in the pathogenesis of IBD, and for *GLI1* as an IBD2 gene.

	Scottish Panel	*Cambridge Panel*	*Swedish Panel*
Total number	**2191** (1374 HC; 474 UC; 335 CD; 8 IBDU)	**2337** (589 HC; 928 UC; 737 CD; 83 IBDU)	**774** (281 HC; 288 UC; 205 CD; 0 IBDU)
Sex - % male	HC 48.7% UC 52.0% CD 39.6%	HC 45.0% UC 52.6% CD 37.0%	HC 45.3% UC 57.3% CD 49.8%
Median age at diagnosis/ recruitment -years (IQR)	HC 50.0 (43.0-55.0) UC 34.1 (25.2-49.9) CD 27.8 (20.8-41.1)	HC 60.0 (53.0-69.0) UC 36.7 (26.84-50.35) CD 26.1 (20.3-37.2)	HC 47.6 (36.4-60.7) UC 28.6 (20.6-41.6) CD 24.5 (19.1-36.7)
CD location			
Terminal ileum (L1)	35.9%	31.3%	17.4%
Colon (L2)	36.5%	36.5%	47.9%
Ileocolon (L3)	25.1%	30.3%	32.6%
Upper GI disease (L4)	2.5%	1.9%	2.1%
UC location			
Proctitis (E1)	17.3%	14.8%	10.4%
Left-sided colitis (E2)	40.5%	34.1%	49.5%
Extensive colitis (E3)	42.2%	51.1%	40.1%
CD 5 year behaviour			
Inflammatory (B1)	64.8%	52.3%	70.9%
Stricturing (B2)	14.8%	35.2%	21.5%
Penetrating (B3)	20.3%	12.5%	7.5%
Perianal involvement (p)	17.4%	24.4%	6.4%

Table 5-1 Detailed demographic and phenotypic data on Scottish, English and Swedish IBD populations.

A. HAPLOTYPE					IBD n=884	UC n=549	CD n=355	HC n=1374	IBD vs. HC p-value OR (C.I.)	UC vs. HC p-value OR (C.I.)	CD vs. HC p-value OR (C.I.)
	$tSNP_1$	$tSNP_2$	$tSNP_3$	$tSNP_4$							
A	1	1	1	1	59.2%	59.6%	58.7%	60.9%	0.26 0.93 (0.82-1.05)	0.50 0.95 (0.81-1.10)	0.30 0.91 (0.76-1.08)
B	2	2	2	2	33.9%	32.5%	35.9%	30.7%	**0.036** 1.16 (1.01-1.32)	0.32 1.09 (0.93-1.28)	**0.012** 1.26 (1.06-1.51)
C	2	2	2	1	4.9%	4.8%	5.1%	6.9%	**0.014** 0.71 (0.54-0.93)	**0.031** 0.68 (0.49-0.96)	0.13 0.74 (0.51-1.07)
Log-likelihood p-value									**< 0.0001**	**< 0.0001**	**0.03**

B. HAPLOTYPE					IBD n=1737	UC n=1000	CD n=737	HC n=589	IBD vs. HC p-value OR (C.I.)	UC vs. HC p-value OR (C.I.)	CD vs. HC p-value OR (C.I.)
	$tSNP_1$	$tSNP_2$	$tSNP_3$	$tSNP_4$							
A	1	1	1	1	62.8%	64.5%	60.4%	63.3%	0.94 0.99 (0.87-1.13)	0.48 1.06 (0.91-1.23)	0.25 0.91 (0.78-1.06)
B	2	2	2	2	29.7%	30.8%	28.2%	27.7%	0.14 1.12 (0.97-1.29)	0.067 1.16 (0.99-1.36)	0.61 1.05 (0.89-1.24)
C	2	2	2	1	5.3%	4.7%	6.1%	6.9%	**0.049** 0.76 (0.58-0.99)	**0.0093** 0.66 (0.49-0.90)	0.44 0.88 (0.65-1.19)
Log-likelihood p-value									**0.009**	**< 0.0001**	**0.002**

C. HAPLOTYPE					IBD n=493	UC n=288	CD n=205	HC n=277	IBD vs. HC p-value OR (C.I.)	UC vs. HC p-value OR (C.I.)	CD vs. HC p-value OR (C.I.)
	$tSNP_1$	$tSNP_2$	$tSNP_3$	$tSNP_4$							
A	1	1	1	1	58.2%	58.6%	57.6%	57.0%	0.68 1.05 (0.85-1.30)	0.62 1.07 (0.84-1.36)	0.86 1.03 (0.79-1.35)
B	2	2	2	2	35.1%	34.4%	36.1%	34.2%	0.81 1.03 (0.83-1.29)	0.97 1.01 (0.78-1.29)	0.55 1.10 (0.83-1.44)
C	2	2	2	1	6.7%	7.0%	6.3%	8.8%	0.17 0.75 (0.50-1.10)	0.32 0.78 (0.50-1.21)	0.16 0.68 (0.41-1.12)
Log-likelihood p-value									0.13	0.41	0.19

Table 5-2 *GLI1* haplotype frequencies in IBD, UC, CD and HC in A. Scotland, B. Cambridge, England, and C. Sweden.
$tSNP_1$: rs3817474; $tSNP_2$: rs2228225; tSNP3: rs2228224; $tSNP_4$: rs2228226. Differences between IBD, CD, UC and HC frequencies are shown, calculated by χ^2 test, with two-sided p-value, odds ratio (OR) and 95% confidence intervals (C.I.). Estimated haplotype frequencies were calculated using Haploview, vers. 3.2. Log-likelihood p-values (calculated on the PM/EH platform with 10,000 permutations and 15 degrees of freedom) are given for each analysis.

	HC		IBD			UC			CD		
	N	**%**	**N**	**%**	**p value** **OR (C.I.)**	**N**	**%**	**p value** **OR (C.I.)**	**N**	**%**	**p value** **OR (C.I.)**
Scotland	1374	30.3	884	34.8	0.0026 1.23 (1.07-1.40)	549	33.9	0.042 1.19 (1.01-1.39)	335	36.1	0.0053 1.30 (1.08-1.55)
Cambridge	589	26.4	1737	29.6	0.042 1.17 (1.00-1.36)	1000	30.8	0.011 1.24 (1.05-1.46)	737	27.9	0.40 1.08 (0.90-1.28)
Sweden	277	30.6	493	35.0	0.27 1.14 (0.90-1.45)	288	34.4	0.43 1.11 (0.90-1.45)	205	35.9	0.24 1.19 (0.90-1.58)
Japan		46.8		49.4	0.53						

Table 5-3 Minor allelic frequencies for *GLI1* non-synonymous SNP rs2228226 (tSNP4) in Scottish, English, and Swedish HC, IBD, CD and UC. Two-tailed p-values are given for χ^2 analysis of IBD vs. HC, UC vs. HC and CD vs. HC.

SNP	dbSNP ID & Position	Location	Allele 1/2		UC N=474	CD N=335	HC N=1374*	UC vs. HC P-value OR (C.I.)	CD vs. HC P-value OR (C.I.)
1	rs2242578 56139420	Promoter	C/G	*U/K* CC CG GG C G	*48 (8.7%)* 221 (44.1%) 225 (44.9%) 55 (11.0%) 667 (66.6%) 335 (33.4%)	*44 (9.9%)* 164 (40.9%) 189 (47.1%) 48 (12.0%) 517 (64.5%) 285 (35.5%)	*30 (8.1%)* 151 (44.4%) 138 (40.6%) 51 (15.0%) 440 (64.7%) 240 (35.3%)	22 vs 11 P=0.18 OR 0.74 (0.48-1.14) 2 vs 1 P=0.40 OR 0.92 (0.75-1.13)	22 vs 11 P=0.57 OR 0.87 (0.55-1.36) 2 vs 1 P=0.96 OR 1.01 (0.82-1.25)
2	rs3825077 56142281	Intron 1/2	T/C	*U/K* TT TC CC T C	*51 (9.3%)* 190 (38.2%) 249 (50.0%) 59 (11.8%) 629 (63.2%) 367 (36.8%)	*34 (7.6%)* 153 (37.2%) 201 (48.9%) 57 (13.9%) 507 (61.7%) 315 (38.3%)	*26 (7.0%)* 141 (41.0%) 152 (44.2%) 51 (14.8%) 434 (63.1%) 254 (36.9%)	22 vs 11 P=0.51 OR 0.86 (0.56-1.32) 2 vs 1 P=0.94 OR 1.00 (0.82-1.22)	22 vs 11 P=0.91 OR 1.03 (0.66-1.60) 2 vs 1 P=0.59 OR 1.06 (0.86-1.31)
3	rs4760148 56143043	Intron 1/2	G/C	*U/K* GG GC CC G C	*65 (11.8%)* 209 (43.2%) 231 (47.7%) 44 (9.1%) 649 (67.0%) 319 (33.0%)	*40 (9.0%)* 175 (43.2%) 183 (45.2%) 47 (11.6%) 533 (65.8%) 277 (34.2%)	*20 (5.4%)* 171 (48.9%) 138 (39.4%) 41 (11.7%) 480 (68.6%) 220 (31.4%)	22 vs 11 P=0.63 OR 0.88 (0.55-1.41) 2 vs 1 P=0.55 OR 1.07 (0.87-1.32)	22 vs 11 P=0.72 OR 1.12 (0.70-1.79) 2 vs 1 P=0.27 OR 1.13 (0.91-1.41)
4	rs3782126 56143200	Intron 1/2	T/C	*U/K* TT TC CC T C	*49 (8.9%)* 222 (44.4%) 233 (46.6%) 45 (9.0%) 677 (67.7%) 323 (32.3%)	*56 (12.6%)* 165 (42.4%) 185 (47.6%) 39 (10.0%) 515 (66.2%) 263 (33.8%)	*19 (5.1%)* 173 (49.3%) 137 (39.0%) 41 (11.7%) 483 (68.8%) 219 (31.2%)	22 vs 11 P=0.55 OR 0.86 (0.54-1.37) 2 vs 1 P=0.67 OR 1.05 (0.86-1.29)	22 vs 11 P=1.00 OR 0.99 (0.61-1.62) 2 vs 1 P=0.29 OR 1.13 (0.91-1.40)
5 ***tSNP1***	**rs3817474** **56145102**	**Intron 4/5**	**A/G**	***U/K*** **AA** **AG** **GG** **A** **G**	***26 (5.4%)*** **163 (35.9%)** **232 (51.1%)** **59 (13.0%)** **558 (61.5%)** **350 (38.5%)**	***32 (9.6%)*** **105 (34.8%)** **150 (49.7%)** **47 (15.6%)** **360 (59.6%)** **244 (40.4%)**	***95 (6.9%)*** **506 (39.6%)** **577 (45.2%)** **194 (15.2%)** **1589 (62.2%)** **965 (37.8%)**	**22 vs 11** **P=0.81** **OR 0.94** **(0.67-1.33)** **2 vs 1** **P=0.71** **OR 1.03** **(0.88-1.21)**	**22 vs 11** **P=0.49** **OR 1.17** **(0.80-1.71)** **2 vs 1** **P=0.25** **OR 1.12** **(0.93-1.14)**
6 ***tSNP2***	**rs2228225** **56145698**	**Exon 6**	**A/G**	***U/K*** **AA** **AG** **GG** **A** **G**	***37 (7.7%)*** **161 (36.2%)** **220 (49.4%)** **64 (14.4%)** **542 (60.9%)** **348 (39.1%)**	***4 (1.2%)*** **118 (35.2%)** **162 (48.4%)** **55 (16.4%)** **398 (59.4%)** **272 (40.6%)**	***91 (6.6%)*** **504 (39.3%)** **577 (45.0%)** **202 (15.7%)** **1585 (61.8%)** **981 (38.2%)**	**22 vs 11** **P=0.96** **OR 1.02** **(0.74-1.42)** **2 vs 1** **P=0.67** **OR 1.04** **(0.89-1.21)**	**22 vs 11** **P=0.46** **OR 1.16** **(0.81-1.67)** **2 vs 1** **P=0.28** **OR 1.10** **(0.93-1.31)**
7	rs2292657 56146199	Intron 7/8	A/G	*U/K* AA AG GG A G	*39 (7.1%)* 186 (36.5%) 256 (50.2%) 68 (13.3%) 628 (61.6%) 392 (38.4%)	*38 (8.5%)* 149 (36.6%) 201 (49.4%) 57 (14.0%) 499 (61.3%) 315 (38.7%)	*26 (7.0%)* 146 (42.4%) 145 (42.2%) 53 (15.4%) 437 (63.5%) 251 (36.5%)	22 vs 11 P=1.00 OR 1.01 (0.66-1.53) 2 vs 1 P=0.44 OR 1.09 (0.89-1.33)	22 vs 11 P=0.82 OR 1.05 (0.68-1.63) 2 vs 1 P=0.39 OR 1.10 (0.89-1.36)
8	rs11830874 56147239	Intron 8/9	T/C	*U/K* TT TC CC T C	*32 (5.8%)* 506 (97.9%) 11 (2.1%) 0 (0%) 1023 (98.9%) 11 (1.1%)	*29 (6.5%)* 404 (97.1%) 12 (2.9%) 0 (0%) 820 (98.6%) 12 (1.4%)	*16 (4.3%)* 340 (96.0%) 14 (4.0%) 0 (0%) 694 (98.0%) 14 (2.0%)	22 vs 11 N/A 2 vs 1 P=0.075 OR 0.53 (0.24-1.18)	22 vs 11 N/A 2 vs 1 P=0.61 OR 1.27 (0.64-2.52)
9	rs10783828 56147751	Intron 9/10	G/A	*U/K* GG GA AA G A	*48 (8.7%)* 225 (44.9%) 229 (45.7%) 47 (9.4%) 679 (67.8%) 323 (32.2%)	*58 (13.0%)* 172 (44.4%) 169 (43.7%) 46 (11.9%) 513 (66.3%) 261 (33.7%)	*23 (6.2%)* 179 (51.6%) 128 (36.9%) 40 (11.5%) 486 (70.0%) 208 (30.0%)	22 vs 11 P=0.81 OR 0.93 (0.59-1.49) 2 vs 1 P=0.35 OR 1.11 (0.90-1.37)	22 vs 11 P=0.47 OR 1.20 (0.75-1.92) 2 vs 1 P=0.13 OR 1.19 (0.95-1.48)
10 ***tSNP3***	**rs2228224** **56151588**	**Exon 12**	**A/G**	***U/K*** **AA** **AG** **GG** **A** **G**	***14 (2.9%)*** **175 (37.2%)** **230 (48.9%)** **65 (13.8%)** **580 (61.7%)** **360 (38.3%)**	***14 (4.1%)*** **109 (33.6%)** **158 (48.8%)** **57 (17.6%)** **376 (58.0%)** **272 (42.0%)**	***101 (7.4%)*** **506 (39.7%)** **560 (44.0%)** **207 (16.3%)** **1572 (61.7%)** **974 (38.3%)**	**22 vs 11** **P=0.62** **OR 0.91** **(0.65-1.26)** **2 vs 1** **P=0.98** **OR 1.00** **(0.86-1.17)**	**22 vs 11** **P=0.21** **OR 1.28** **(0.89-1.83)** **2 vs 1** **P=0.091** **OR 1.17** **(0.98-1.39)**
11 ***tSNP4***	**rs2228226** **56152088**	**Exon 12**	**C/G**	***U/K*** **CC** **CG** **GG** **C** **G**	***9 (1.9%)*** **205 (43.2%)** **217 (45.7%)** **53 (11.2%)** **627 (66.0%)** **323 (34.0%)**	***10 (2.9%)*** **137 (41.5%)** **148 (44.8%)** **45 (13.6%)** **422 (63.9%)** **238 (36.1%)**	***110 (8.0%)*** **610 (48.3%)** **542 (42.9%)** **112 (8.9%)** **1762 (69.7%)** **766 (30.3%)**	**22 vs 11** **P=0.079** **OR 1.41** **(0.98-2.03)** **2 vs 1** **P=0.040** **OR 1.19** **(1.01-1.39)**	**22 vs 11** **P=0.0048** **OR 1.79** **(1.21-2.65)** **2 vs 1** **P=0.0053** **OR 1.30** **(1.08-1.55)**

Table 5-4 Details of the 11 *GLI1* SNPs genotyped in the Scottish population.
*n=370 for SNPs 1-4 and 7-9. U/K = unknown.

SNP	dbSNP ID & Position	Location	Allele 1/2		UC N=1000	CD N=737	HC N=589	UC vs. HC P-value OR (C.I.)	CD vs. HC P-value OR (C.I.)
tSNP$_1$	rs3817474 56145102	Intron 4/5	A/G	*U/K* AA AG GG A G	 426 (42.6%) 441 (44.1%) 133 (13.3%) 1293 (64.7%) 707 (35.4%)	 311 (42.2%) 348 (47.2%) 78 (10.6%) 970 (65.8%) 504 (34.2%)	 265 (45.0%) 255 (43.3%) 69 (11.7%) 785 (66.6%) 393 (33.4%)	22 vs 11 P=0.32 OR 1.20 (0.86-1.67) 2 vs 1 P=0.27 OR 1.09 (0.94-1.27)	22 vs 11 P=0.91 OR 0.96 (0.67-1.39) 2 vs 1 P=0.68 OR 1.04 (00.88-1.22)
tSNP$_2$	rs2228225 56145698	Exon 6	A/G	*U/K* AA AG GG A G	 426 (42.6%) 441 (44.1%) 133 (13.3%) 1293 (64.7%) 707 (35.4%)	 235 (33.8%) 367 (52.7%) 94 (13.5%) 837 (60.1%) 555 (39.9%)	 228 (38.1%) 293 (48.9%) 78 (13.0%) 749 (62.5%) 449 (37.5%)	22 vs 11 P=0.64 OR 0.91 (0.66-1.26) 2 vs 1 P=0.24 OR 0.91 (0.79-1.06)	22 vs 11 P=0.43 OR 1.17 (0.82-1.66) 2 vs 1 P=0.23 OR 1.11 (0.94-1.30)
tSNP$_3$	rs2228224 56151588	Exon 12	A/G	*U/K* AA AG GG A G	 423 (42.3%) 444 (44.4%) 133 (13.3%) 1290 (64.5%) 710 (35.5%)	 261 (41.1%) 301 (47.4%) 73 (11.5%) 823 (64.8%) 447 (35.2%)	 272 (43.9%) 267 (43.1%) 80 (12.9%) 811 (65.5%) 427 (34.5%)	22 vs 11 P=0.74 OR 1.07 (0.78-1.47) 2 vs 1 P=0.58 OR 1.05 (0.90-1.21)	22 vs 11 P=0.86 OR 0.95 (0.66-1.36) 2 vs 1 P=0.74 OR 1.03 (0.88-1.22)
tSNP$_4$	rs2228226 56152088	Exon 12	C/G	*U/K* CC CG GG C G	 494 (49.4%) 397 (39.7%) 109 (10.9%) 1385 (69.2%) 615 (30.8%)	 368 (51.0%) 306 (42.4%) 48 (6.7%) 1042 (72.2%) 402 (27.8%)	 316 (55.5%) 206 (36.2%) 47 (8.26%) 838 (73.6%) 300 (26.4%)	22 vs 11 P=0.045 OR 1.48 (1.03-2.15) 2 vs 1 P=0.010 OR 1.24 (1.05-1.46)	22 vs 11 P=0.62 OR 0.88 (00.57-1.35) 2 vs 1 P=0.43 OR 1.08 (0.90-1.28)

Table 5-5 Details of the 4 *GLI1* tSNPs genotyped in the Cambridge population.
χ^2 analysis of allelic frequency and homozygosity for each tSNP for UC and CD versus HC is presented with corresponding p-values, odds ratios (OR) and 95% confidence intervals (C.I.).

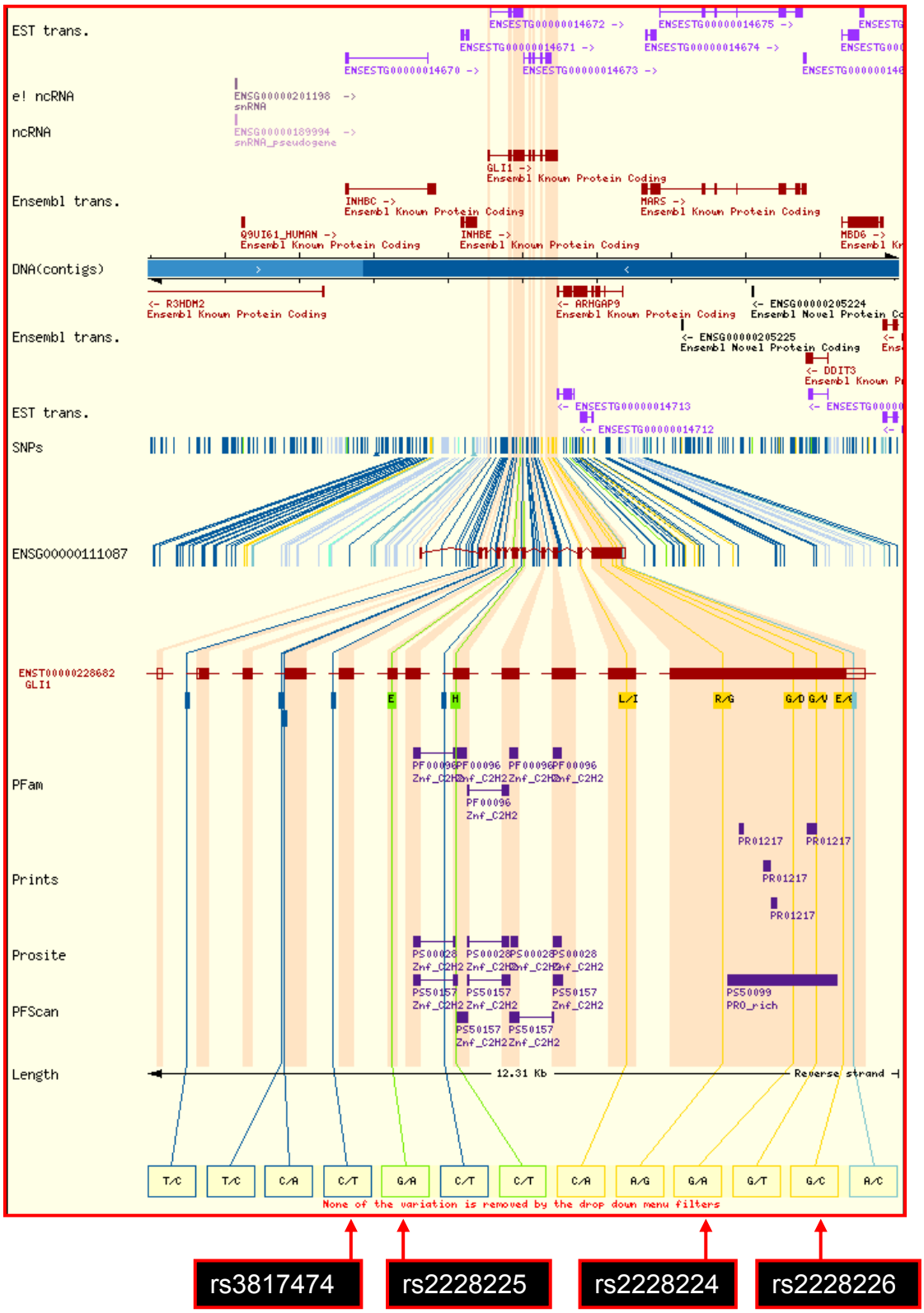

Figure 5-1 Structure of *GLI1* gene.
Data from Ensembl illustrate the structure of the *GLI1* gene, including position of tSNPs1-4 and of neighbouring genes.

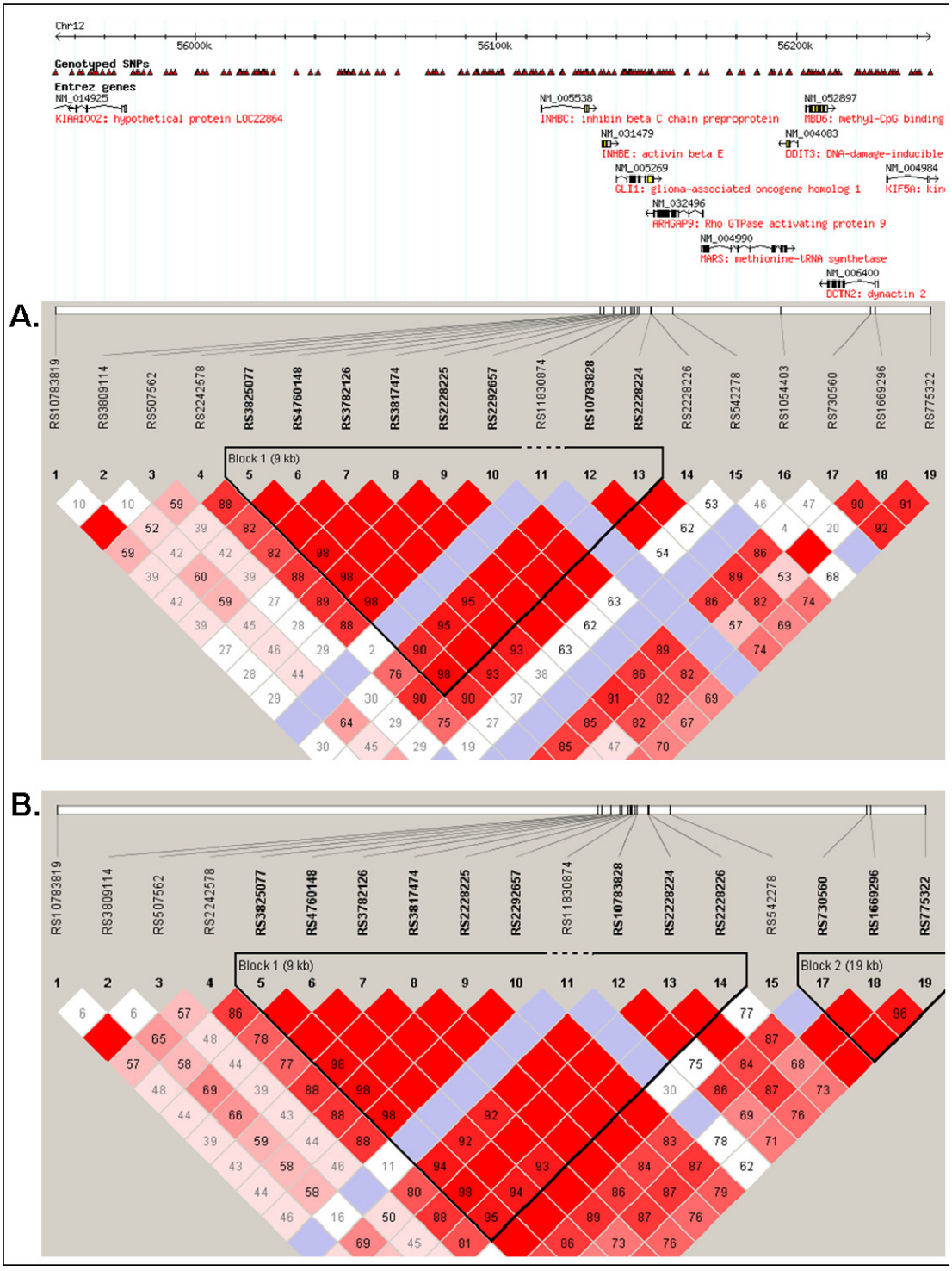

Figure 5-2 Haplotype structure at *GLI1* and surrounding region from Scottish data.
The haplotype structure in *GLI1* and surrounding haplotype blocks are presented in Haploview from genotyping in UC (**A**) and CD (**B**). There is a haplotype block spanning the *GLI1* gene that does not extend into neighbouring genes. LD is described here by D′ values.

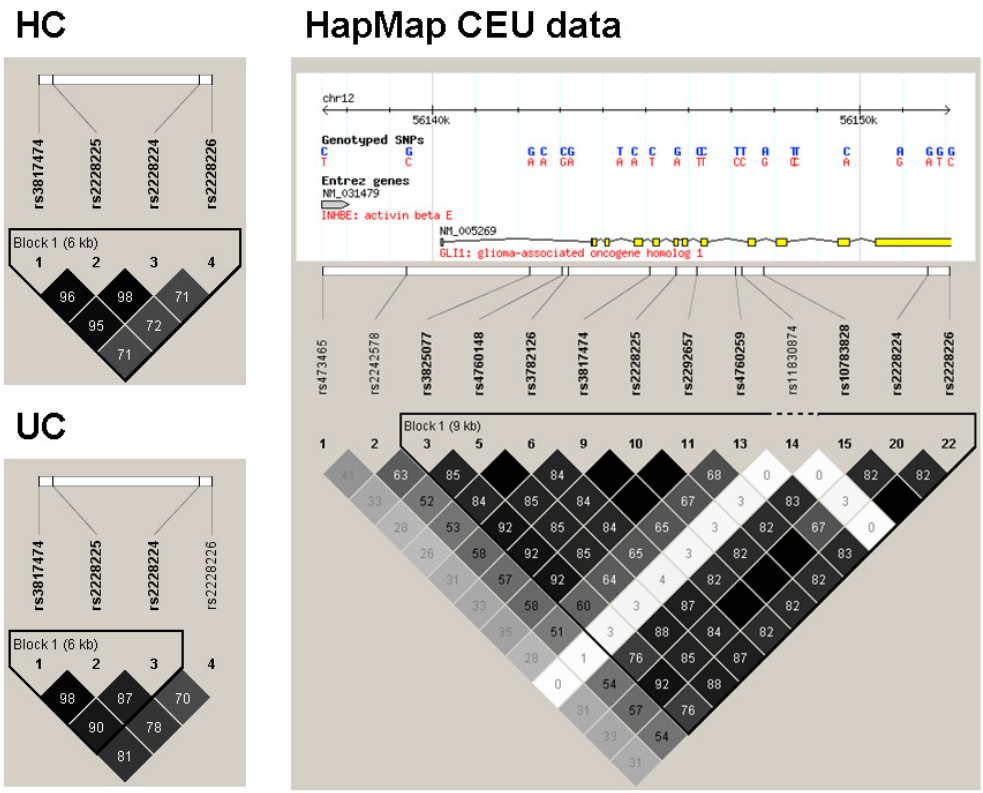

Figure 5-3 r^2 linkage disequilibrium between tSNPs1-4.

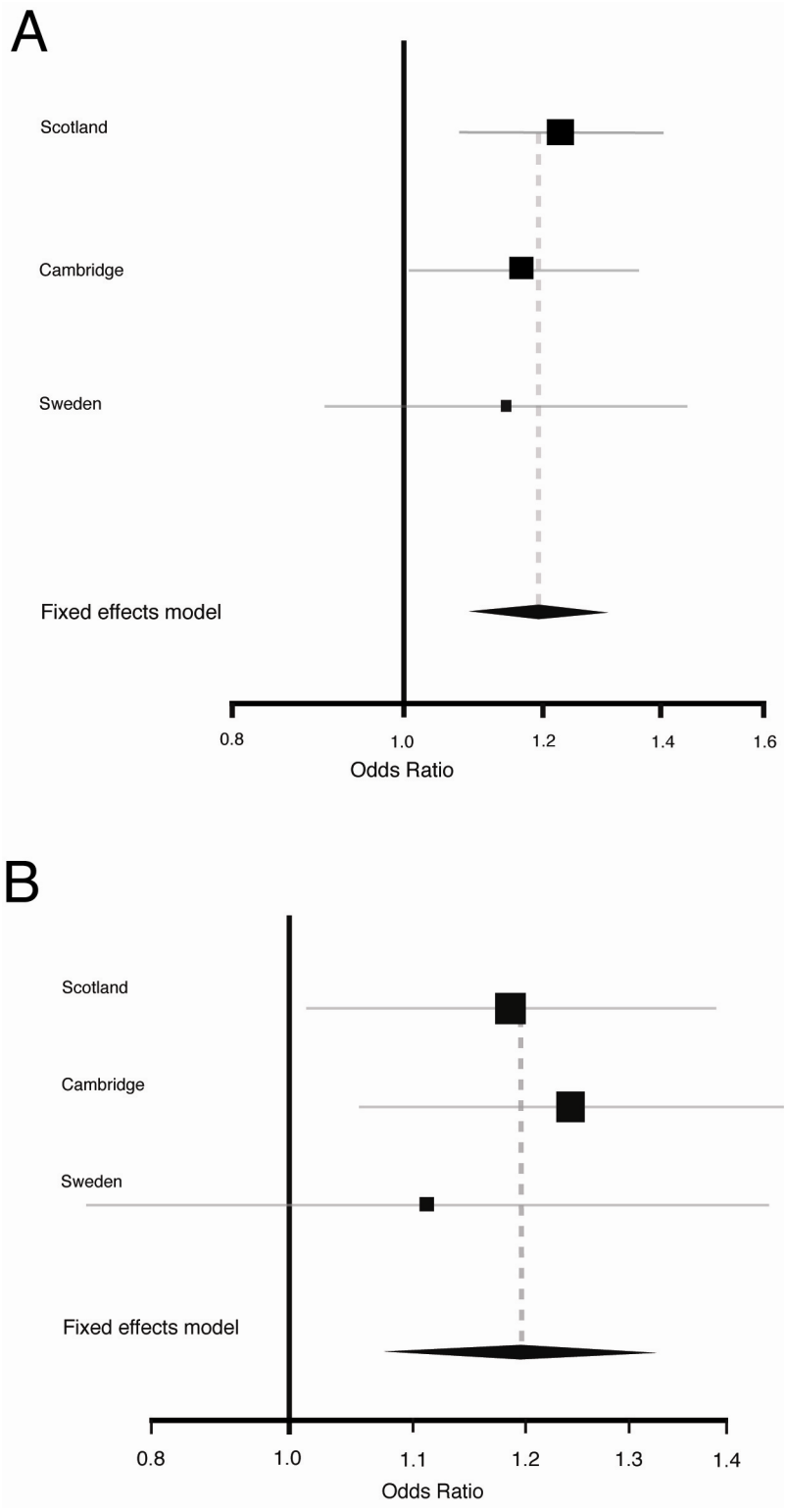

Figure 5-4 Mantel-Haenszel meta-analysis of rs2228226 in Scotland, Cambridge and Sweden. **A**. IBD vs. HC. **B**. UC vs. HC. There was no evidence of heterogeneity in the contribution of rs2228226 between the 3 cohorts (p=0.825). Confidence intervals for individual populations are represented by horizontal lines and population sizes by square boxes. The diamond represents the pooled odds ratio (fixed effect model) with 95% C.I. delineated by the diamond's width. Note the different range of the x axis for A and B.

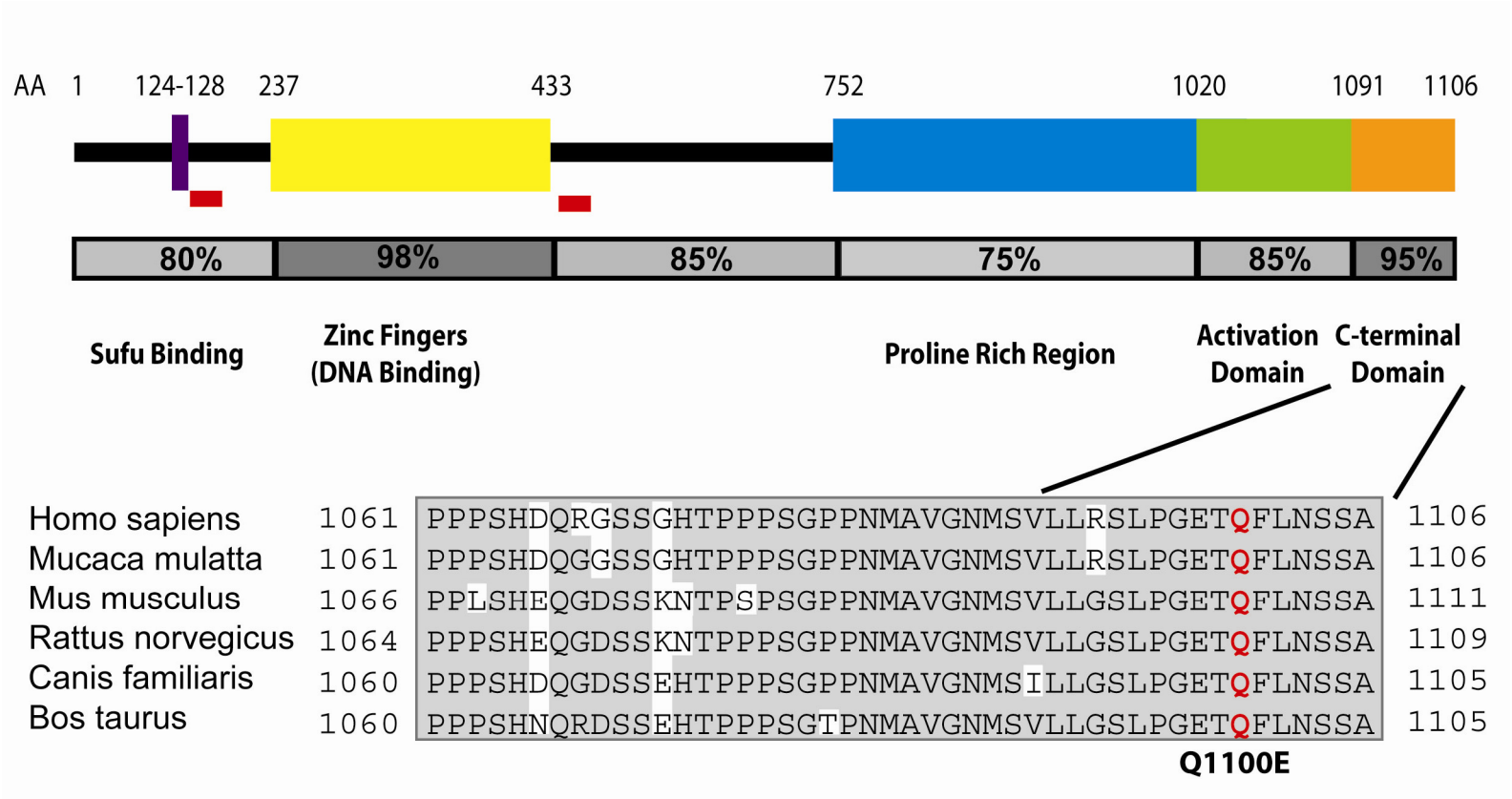

Figure 5-5 Protein structure of *GLI1*.
A) Conservation of known functional domains in the Gli1 protein. Previously described Sufu binding, DNA binding, and transactivation domains[447, 457, 458] are shown schematically. Amino acid conservation of each domain is represented numerically and by shading of the bar below the domain. The Sufu binding sequence is 100% conserved in examined mammalian GLI1 proteins. Red boxes indicate regions known to regulate GLI1 protein stability.[448] A highly conserved C-terminal domain which includes Q1100E (equating to rs2228226) is identified adjacent to a known transactivation domain. B) Alignment of the C-terminus of mammalian GLI1 proteins. This region (AA 1080-1106) is one of the most conserved portions of the GLI1 protein in mammalian lineages.

6 Germ-line variation in *GLI1* and colo-rectal cancer in Scotland.

6.1 *Abstract*

Background. Association of *GLI1* with IBD, predominantly UC, has been demonstrated. A putative role for HH signalling and GLI1 in colo-rectal cancer pathogenesis has been described. About one thirds of disease variance has been attributed to inherited risk factors. Only a small proportion of this is explained by genetic variants identified through candidate gene studies (e.g. DNA mismatch repair genes) and recent GWAS reports (e.g. *SMAD7*).

Aims. The aim of the present study was to determine if germ-line variation in *GLI1* was associated with colo-rectal cancer.

Methods. Four gene-wide haplotype-tagging SNPs (rs3817474, rs2228225, rs2228224 and rs2228226) were genotyped on the TaqMan platform in a cohort of Scottish early-onset colo-rectal cancer patients (n = 983, median age at diagnosis 50.0) and age and geographically matched healthy controls (n = 1004). Log-likelihood, haplotype and individual analysis was performed along with binomial logistic regression modelling and subphenotype analysis for location of colonic tumour.

Results. On log-likelihood analysis, there was overall association of *GLI1* with colo-rectal cancer ($p<0.001$). $tSNP_4$ (rs2228226) was associated with colo-rectal cancer on genotype analysis only ($p=0.05$). There was no association with $tSNPs_{1-3}$ or individual haplotypes. Binomial logistic regression analysis suggested that overall gene association was best explained by a combination of $tSNPs_3$ and $_4$ (rs2228224 and rs2228226). There was a trend to increased association with rectal tumours but this did reach significance.

Conclusions. These data suggest that germ-line variation in *GLI1* is associated with colo-rectal cancer. However, the effect is much less marked than that observed in UC and requires validation by replication in independent cohorts. Recent determination of the paracrine nature of HH signalling in colo-rectal cancer has catalysed clinical trials of a small molecule HH antagonist in patients with metastatic disease.

6.2 *Introduction*

Worldwide, colo-rectal cancer is the fourth leading cause of cancer death (about 655,000 deaths per annum) with a lifetime risk for an individual in Western Europe of about 5%. It is typically more common in men than women with peak onset in the sixth and seventh decades of life; onset of disease before 50 years of age is uncommon.[459] The location of tumours within the colon largely determines clinical features at presentation; the typical distribution is depicted in **Figure 6-1**. There is a significantly increased risk of colo-rectal cancer in patients with both UC and CD colitis.

Sequential molecular changes occur in colonic epithelial cells, including somatic mutations in the *K-ras* protooncogene, *APC* (18q), the deleted in colorectal cancer (*DCC*) gene and the *p53* tumour suppressor gene, and hypomethylation of DNA. These are known to be induced by a variety of stimuli including cigarette smoke, dietary and other environmental carcinogens, and chronic colonic inflammation. This accumulation of cellular changes drives the healthy colonic epithelium to adenocarcinoma via increasing dysplasia and subsequent adenoma. The colonic stem cell is the logical location for the accumulation of these mutations due to its protected niche in the colonic crypt (position +4) and subsequent longevity. Very recently, LGR5 has been identified as a putative intestinal stem cell marker.[460]

It has been calculated that around one third of colo-rectal cancer variance is attributable to inherited genetic factors.[461] However, in the era pre-GWAS only about 5% of disease risk could be explained by a series of rare, highly penetrant variants in a number of genes : DNA mismatch repair genes,[462] *APC*, *SMAD4*, *BMPR1A* and *MUTYH*.[463]

With the advent of GWAS a number of widely replicated novel loci have been identified. Most recently, Tenesa and colleagues from Scotland,[162] and Tomlinson and colleagues from across the United Kingdom[163] have reported the full details of GWAS each involving 3 - 4 stages of replication and tens of thousands of patients and controls. In phase I of the Scottish study, 555,510 SNPs were genotyped in 1,012 early-onset colo-rectal cancer patients (from the lowest 10th centile by age at

diagnosis) and a similar number of age and geographically matched controls.[162] After replication of 15,008 SNPs in phase II, the top-five ranked SNPs were genotyped in 14,500 cases from 7 geographically distinct populations (Scotland, England, Canada, Germany, Spain, Japan and Israel) along with 13,294 appropriate matched controls. In this final stage these regions (along with 2 previously identified) were additionally fine mapped. The best replicated loci to date are:

- rs7014346 at 8q24 (*POU5F1P1 / DQ515897*)
    - $p = 8.6 \times 10^{-26}$; OR 1.19
- rs4939827 at 18q21 (*SMAD7*)
    - $p = 7.77 \times 10^{-28}$; OR 1.20
- rs3802842 at 11q23 (high LD across multi-genic locus including *LOC120376 / FL54803 / C11orf53 / POU2AR1* plus polymorphic binding site target for miRNA)
    - $p = 5.82 \times 10^{-10}$; OR 1.11

In the other large GWAS, Tomlinson reported two additional well replicated loci[163]:

- rs10795668 at 10p14
    - $p = 2.5 \times 10^{-13}$; OR 0.89
- rs16892786 at 8q23.3 (*EIF3H*)
    - $p = 3.3 \times 10^{-18}$; OR 1.25

A meta-analysis of the Tenesa and Tomlinson GWAS has subsequently identified an additional four loci[161]:

- rs4444235 at 14q22.2 (*BMP4*)
    - $p = 8.1 \times 10^{-10}$; OR 1.11
- rs9929218 at 16q22.1 (*CDH1*)
    - $p = 1.2 \times 10^{-8}$; OR 0.91
- rs10411210 at 19q13.1 (*RHPN2*)
    - $p = 4.6 \times 10^{-9}$; OR 0.87
- rs961253 at 20p12.3
    - $p = 2.0 \times 10^{-10}$; OR 1.12

In Chapter 5, it was demonstrated that germ-line variation in *GLI1* was significantly associated with UC, and to a lesser extent CD. In recent years, there has been much debate as to the role HH signalling may or may not play in the pathogenesis of colo-

rectal cancer. This literature has been discussed in some detail in Chapter 1 (see **1.9.1.2**). Most pertinent to the present chapter is the demonstration by Akiyoshi and colleagues that GLI1 expression is dysregulated in colon cancer.[417] Given this potential link, the gene-wide haplotype tagging strategy described in the previous chapter was employed to ascertain whether *GLI1* variation was associated with colo-rectal cancer.

6.3 *METHODS*

6.3.1 Colo-rectal cancer patients

The patient population consisted of 983 patients with early-onset colo-rectal cancer from Scotland (median age at diagnosis 50.0 years, IQR 45.0-53.0) and 1,004 age and geographically matched controls. This cohort was provided by Professor MG Dunlop and, as such, derived from phase I of their GWAS described above.[162] The demographics and phenotype of the cohort are detailed in **Table 6-1**. Location of colonic tumours was defined as follows: rectal – rectum and sigmoid; left-sided – distal to splenic flexure excluding rectum and recto-sigmoid; right-sided – proximal to splenic flexure.

6.3.2 Genotyping

The four gene-wide haplotype-tagging SNPs employed in the previous chapter were genotyped on the same TaqMan platform.

6.3.3 Statistical analysis

In addition to the chi^2, haplotype and log-likelihood analysis as described in the previous chapter, binomial logistic regression was also performed on this cohort. Binomial logistic regression was used in R (by A Tenesa) to test for the overall association of the gene with the disease. All 4 SNPs were fitted at a time and dropped from the model one by one (dropping always the one less significant). The 2 nested models (i.e. with and without the SNP that was dropped from the model) were tested using R's ANOVA command. This tests whether having the extra variable improves the fit of the model. The final model included only $tSNP_3$ and $tSNP_4$. A genotype-phenotype analysis stratified by colo-rectal tumour location was performed for the four individual tSNPs.

6.4 *Results*

6.4.1 Germ-line variation in *GLI1* and colo-rectal pathogenesis.

On log-likelihood analysis with 1000 permutations, overall *GLI1* variation was associated with colo-rectal cancer ($p<0.001$). On analysis of individual SNPs (**Table 6-2**) there was association with $tSNP_4$ (rs2228226) genotype of borderline significance ($p=0.05$), but not on analysis of allelic frequency ($p=0.29$). There was no association with $tSNPs_{1-3}$. There was no difference in individual haplotype frequencies between cases and controls (**Table 6-3**). To account for potential confounding of the IBD cases within this cohort the analysis was repeated with these cases removed with no significant alteration of these results (data not shown).

Using binomial logistic regression to model the association with colo-rectal cancer, a best fit was observed when $tSNPs_{3\&4}$ were included in a genotypic model ($p=0.023$) (**Table 6-4**). The model was improved further with the addition of an iteration term ($p=0.0004$). Analysis of the genotype counts at these two tSNPs demonstrates why the interaction explains the overall association of *GLI1* on log-likelihood testing, whilst individual SNP analysis is borderline.

6.4.2 Genotype-phenotype analysis

A within-cases genotype-phenotype and haplotype-phenotype analysis was performed for colo-rectal cancer patients stratified by location of colonic tumour. A trend was noted towards increased risk with all tSNPs when comparing rectal tumour versus left-sided and right-sided tumour (**Table 6-5a**), but this did not reach statistical significance on individual SNP analysis or log-likelihood testing ($p=0.35$). In light of the gradients of HH expression noted in Chapter 4, a potential association between genotype and location of colonic tumour was explored. There was no evidence of a gradient along the length of the colon (rectum – left – right-sided) with increasing susceptibility to colonic cancer for individual tSNPs, although MAFs were consistently greater for rectal tumours (**Figure 6-2**).

6.5 *Discussion*

The data herein demonstrate overall association of *GLI1* with colo-rectal cancer in a single large population of early-onset patients from Scotland. However, the

association on permutation testing with log-likelihood analysis was not seen at either the haplotype or the individual SNP level. This discrepancy has a number of possible explanations. Firstly, as discussed in the previous chapter in the context of the contribution of germ-line *GLI1* variation to IBD pathogenesis, additional variation in the gene – promoter or intronic variants and/or CNVs – may explain the overall gene association. For IBD, it was shown firstly that the association was limited to *GLI1* and did not extend into neighbouring genes and secondly, that the association with $tSNP_4$ (rs2228226) explained most, but not all, of the genetic variation. In this colo-rectal cancer cohort, there is a trend towards association with this rs2228226, but this is only of very borderline significance and unlikely to explain the gene-wide association. However, the binomial logistic regression analysis performed provides evidence for an alternative explanation: that the contribution of $tSNP_4$ is dependent on the genotype of $tSNP_3$. These data require replication in an independent cohort of colo-rectal cancer patients and controls.

In the recent GWAS in colo-rectal cancer, genotype has been associated with subphenotypes, specifically the location of the tumour within the colon.[162] For example, rs3802842 at 11q23 and rs4939827 at 18q21 (*SMAD7*) were both associated with rectal tumours to a great extent than tumours in the rest of the colon (p=0.008 and p=0.009 respectively), although the magnitude of this effect is small (OR 1.07 for both SNPs).[162] A genotype-phenotype analysis was therefore performed within the colo-rectal cancer cases for *GLI1* in the present cohort. Whilst there was a trend towards greater association with rectal cancers, similar to 11q23 and 18q21, this fell short of statistical significance. However, this may be an issue of sample size and power: in the GWAS the subphenotype analysis was performed on a total of 14,500 cases, compared with 983 in the present study. It would be of great interest to see if this subphenotypic association was confirmed on a larger cohort given the gradients of HH pathway expression demonstrated in Chapter 4. There, it was shown that HH ligand plus pathway response elements (PTCH, HHIP and GLI1) were expressed at greater levels in the distal compared with the proximal colon. The inferred greater HH pathway activity in the distal colon is notable given the distribution of intestinal commensal flora in the healthy human colon and the clinical distribution of UC.

The colo-rectal cancer cohort in this study contained a total of 39/983 (4.0%) patients with a concomitant medical history of IBD (majority UC), compared with only 7/1004 (0.7%) of healthy controls ($p<0.0001$, OR 5.88, C.I. 2.62-13.22). This significant difference is expected as there is a well established increased risk of colo-rectal cancer in patients with known colitis (both UC and CD colitis, as discussed in more detail in Chapter 1). Although the overall number of patients with co-existent IBD is small, this may be sufficient to explain the overall association of *GLI1* with colo-rectal cancer. However, repeat log-likelihood, haplotype and individual SNP analysis with these 46 cases omitted did not significantly alter any of the data or conclusions.

In conclusion, these data suggest that further to the significant association seen between *GLI1* and IBD, this key transcriptional mediator of HH signalling may also be associated with inherited susceptibility to colo-rectal cancer. However, these data should be interpreted with some caution as the results are of borderline significance and require validation in independent cohorts. They do, at least, add to the weight of evidence suggesting that HH signalling does have a significant role to play in colo-rectal cancer pathogenesis (**1.9.1.2**).[361] Finally, the role of HH in the development of colo-rectal cancer in patients with established colitis merits detailed exploration.

	Colo-rectal cancer cases	**Healthy controls**
Total number	983	1004
Median (IQR) age (years) at diagnosis / recruitment	50.0 (45.0-53.0)	52.0 (47.0-55.0)
Male sex (%)	49.7%	52.6%
Co-existent IBD	UC 23 CD 9 IBDU 7	UC 3 CD 1 IBDU 3
Location of cancer		
Rectum	41.6%	
Left-sided	26.9%	
Right-sided	23.8%	
Not specified / unknown	7.7%	

Table 6-1 Demographics and phenotype of colo-rectal cancer cases and HCs.

SNP	dbSNP ID & Position	Location	Allele 1/2		CRC N=983	HC N=1004	CRC vs. HC P-value OR (C.I.)
tSNP1	rs3817474 56145102	Intron 4/5	A/G	*U/K* AA AG GG A G	79/983 (8.0%) 349 (38.6%) 420 (46.5%) 135 (14.9%) 1118 (61.8%) 690 (38.2%)	65/1004 (6.5%) 368 (39.2%) 427 (45.5%) 144 (15.3%) 1163 (61.9%) 715 (38.1%)	22 vs 11 P=0.94 OR 0.99 (0.75-1.30) 2 vs 1 P=0.98 OR 1.00 (0.88-1.15)
tSNP2	rs2228225 56145698	Exon 6	A/G	*U/K* AA AG GG A G	61/983 (6.2%) 344 (37.3%) 435 (47.2%) 143 (15.5%) 1123 (60.9%) 721 (39.1%)	79/1004 (7.9%) 354 (38.3%) 426 (46.1%) 145 (15.7%) 1134 (61.3%) 716 (38.7%)	22 vs 11 P=0.94 OR 1.02 (0.77-1.34) 2 vs 1 P=0.0.83 OR 1.02 (0.89-1.16)
tSNP3	rs2228224 56151588	Exon 12	A/G	*U/K* AA AG GG A G	80/983 (8.1%) 348 (38.5%) 417 (46.2%) 138 (15.3%) 1113 (61.6%) 693 (38.4%)	83/1004 (8.1%) 365 (39.6%) 408 (44.3%) 148 (16.1%) 1138 (61.8%) 704 (38.2%)	22 vs 11 P=0.89 OR 0.98 (0.74-1.29) 2 vs 1 P=0.95 OR 1.01 (0.88-1.15)
tSNP4	rs2228226 56152088	Exon 12	C/G	*U/K* CC CG GG C G	90/983 (9.6%) 419 (46.9%) 378 (42.3%) 96 (10.8%) 1216 (68.1%) 570 (31.9%)	101/1004 (10.1%) 426 (47.2%) 408 (45.2%) 69 (7.6%) 1260 (69.8%) 546 (30.2%)	22 vs 11 P=0.05 OR 1.42 (1.01-1.98) 2 vs 1 P=0.29 OR 1.08 (0.94-1.25)

Table 6-2 Details of the 4 *GLI1* tSNPs genotyped in the colo-rectal cancer (CRC) population and HCs. χ^2 analysis of allelic frequency and homozygosity for each tSNP for CRC versus HC is presented with corresponding p-values, odds ratios (OR) and 95% confidence intervals (C.I.).

HAPLOTYPE					CRC	HC	p-value	OR (C.I.)
	$tSNP_1$	$tSNP_2$	$tSNP_3$	$tSNP_4$				
A	1	1	1	1	60.4%	60.9%	0.74	0.98 (0.86-1.11)
B	2	2	2	2	31.4%	30.8%	0.65	1.03 (0.90-1.18)
C	2	2	2	1	6.8%	7.0%	0.80	0.97 (0.76-1.25)

Table 6-3 Haplotype analysis of colo-rectal cancer (CRC) cases versus HCs.
Differences between CRC and HC frequencies are shown, calculated by χ^2 test, with two-sided p-value, odds ratio (OR) and 95% confidence intervals (C.I.). Estimated haplotype frequencies were calculated using Haploview, vers. 3.2.

A.		**Genotypes**		
	$tSNP_4$	**11**	**12**	**22**
	$tSNP_3$			
OR	**11**	**1**	**8.7***	**∞**
	12	**1**	**0.8**	**2.5**
	22	**1**	**0.47***	**0.77**

B.		**Genotypes**		
	$tSNP_4$	**11**	**12**	**22**
	$tSNP_3$			
OR	**22**	**1.1**	**12.9***	**∞**
	12	**1.3**	**1.6**	**2.3**
	11	**1**	**1**	**1**

Table 6-4 Binomial logistic regression of *GLI1* in colo-rectal cancer compared with healthy controls.
A. Change of risk (odds ratio) for SNP4 given that a person is SNP3 genotype 11, 12 or 22.
B. Change of risk for SNP3 given that a person is SNP4 genotype 11, 12 or 22. For example, people with genotype 11 at SNP4 do not have a big increase in risk for being 22 homozygous for SNP3. However, there is a big increase in risk if the person is 12 or 22 for SNP4.
* *means is significant at the 5% level.*

Rectal vs. Left & Right-sided tumours	MAF		p-value	OR (C.I.)
	Rectum	Left & right		
$tSNP_1$	0.394	0.361	0.17	1.16 (0.95-1.41)
$tSNP_2$	0.403	0.374	0.23	1.13 (0.93-1.37)
$tSNP_3$	0.394	0.367	0.25	1.12 (0.92-1.37)
$tSNP_4$	0.329	0.301	0.23	1.14 (0.92-1.40)

Rectal vs. Left & Right-sided tumours	MAF		p-value	OR (C.I.)
	Rectum	Left & right		
$tSNP_1$	0.375	0.376	0.98	1.00 (0.79-1.25)
$tSNP_2$	0.388	0.387	0.97	1.00 (0.80-1.26)
$tSNP_3$	0.378	0.380	0.97	0.99 (0.79-1.25)
$tSNP_4$	0.302	0.318	0.53	0.93 (0.73-1.18)

Table 6-5 Genotype-phenotype analysis of colo-rectal cancer cases (within-cases analysis) by tumour location of colonic tumour.
Top panel: rectal tumours versus left and right-sided tumours
Bottom panel: right-sided tumours versus rectal and left-sided tumours

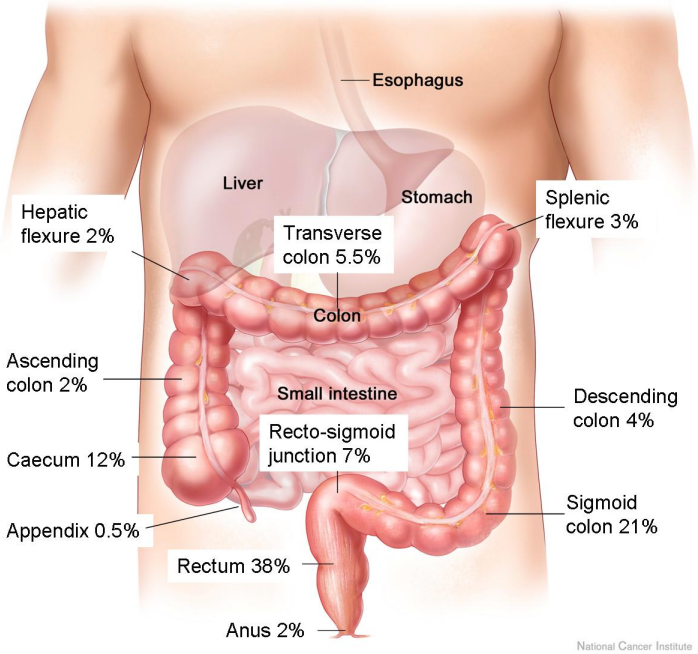

Figure 6-1 Distribution of colo-rectal tumours in Scottish early-onset population.

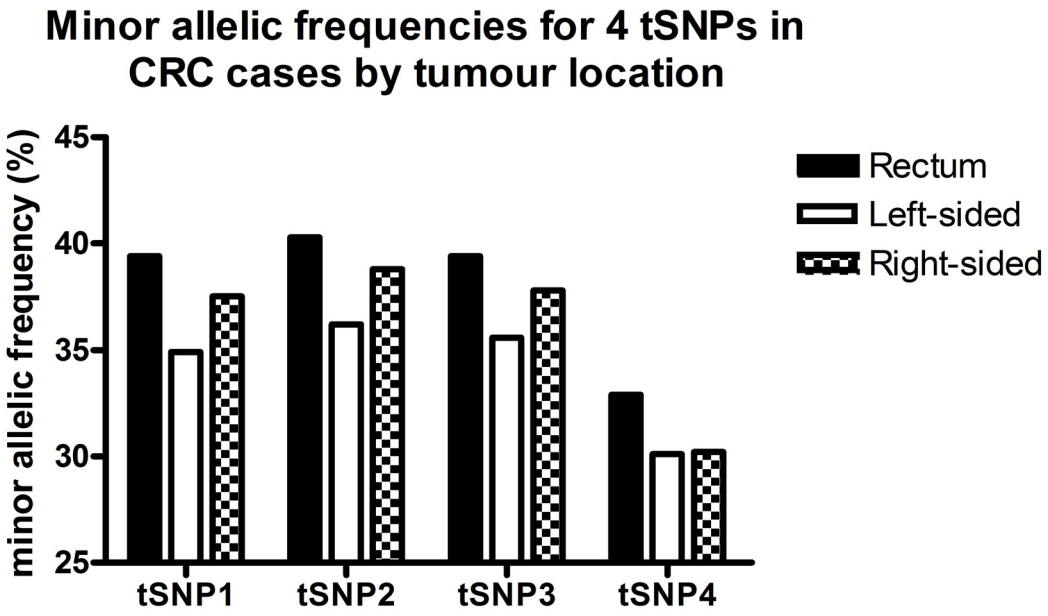

Figure 6-2 Minor allelic frequencies for 4 tSNPs by tumour location.

7 Gene-wide haplotype-tagging study of *IHH* in IBD.

7.1 *Abstract*

Introduction: It was demonstrated in Chapter 4, by analysis of HH pathway response elements (PTCH, GLI1 and HHIP) in IBD tissue, that HH pathway activity is decreased in established colonic inflammation in man. The expression of the ligand IHH was notable in that it was down-regulated in the colons of patients with UC compared with HC, regardless of inflammatory status. This raises the possibility that germ-line *IHH* variation may be associated with UC.

Aims: To determine whether *IHH* is associated with UC and CD in Scotland.

Methods: A gene-wide haplotype-tagging study was performed. tSNPs were identified using the aggressive tagging algorithm embedded in Tagger with r^2 threshold set at > 0.8. 5 tSNPs (rs1378640, rs394452, rs3731881, rs427587 and rs3100776) were genotyped in a cohort of Scottish patients with IBD (448 CD, 549 UC) and HC (n=428). Overall contribution of *IHH* to IBD susceptibility was assessed by log-likelihood analysis with 1000 permutations. Haplotype analysis was performed in Haploview. Chi^2 analysis was performed to assess the contribution of individual tSNPs to IBD susceptibility.

Results: On log-likelihood testing there was no overall association between *IHH* and IBD, CD or UC. The 5 tSNPs described 5 haplotypes with frequency >1% in HC population. There was no haplotype or tSNP association with IBD. However, there was weak association of haplotype C with UC (OR 1.39, p=0.042), haplotype E with CD (OR 3.60, p=0.0072) and protection of haplotype D with UC (OR 0.58, p=0.041). Individual tSNP analysis demonstrated association at rs3100776 with UC (OR 0.58, p=0.043) and at rs427857 with CD (OR 3.57, p=0.0075). The MAFs of both these SNPs is low in the Scottish HC population (0.039 and 0.006). There were no significant associations on genotype-phenotype testing.

Conclusions: In the present study there is no evidence of association at *IHH* with IBD. The borderline associations seen on haplotype and tSNP analysis in UC and CD are inconsistent and in relatively uncommon haplotypes and SNPs and are therefore of doubtful significance. Alternative explanations (e.g. epigenetics) should therefore be sought to explain the expression differences seen for IHH in UC.

7.2 *Introduction*

The detailed series of HH signalling pathway expression analyses presented in Chapter 4 demonstrated that PTCH and GLI1 are decreased in colonic inflammation in man. These pathway response elements are direct transcriptional targets of the pathway and as such predict pathway activity. Whilst no consistent expression changes were noted for SHH, IHH (the other major HH ligand in the intestine) was down-regulated in UC. Moreover, the altered expression profile demonstrated down-regulation in both inflamed and non-inflamed UC samples when compared with non-inflamed HCs. This suggests either that down-regulation of IHH in established UC persists despite resolution of inflammation, or that germ-line variation in *IHH* leads to lower expression of IHH in the gut predisposing patients to develop UC.

The published literature suggest that IHH is the major HH ligand in the mammalian intestine. Van den Brink and colleagues demonstrated IHH to be critical to normal colonic enterocyte differentiation and that IHH expression is lost where WNT pathway is constitutively overexpressed in the context of FAP, as discussed in more detail in Chapter 1.[318] This is further supported by the data presented in Chapter 4. Given that the expression pattern of SHH in the healthy colon does not mirror that of the HH pathway response genes (i.e. no gradient of increasing expression from proximal to distal colon) it is highly likely that changes in IHH expression in terminally differentiated enterocytes and not SHH is driving the expression changes in PTCH, GLI1 and HHIP. Taken together, these observations argue for a major role for IHH in the pathogenesis of UC.

To examine the role of *IHH* variation in IBD susceptibility, a gene-wide haplotype-tagging study was performed in Scottish IBD.

7.3 *Methods*

7.3.1 *IHH* genotyping

IHH is a small gene (6.0kb) located in a 194.7kb haplotype block on chromosome 2q35 (**Figure 7-1a-b**). 5 other genes are located within this block: *CCDC108* (coiled-

coil domain containing 108), *NHEJ1* (XRCC4-like factor), *SLC23A3* (solute carrier family 23 (nucleobase transporters), member 3), *C2orf24* (chromosome 2 open reading frame 24) and *C2orf17* (chromosome 2 open reading frame 17).

Tagger was used to select *IHH* tSNPs. 10kb of HapMap data (phase II CEU population) was imported into Haploview (version 4.2). Of 13 HapMap SNPs in this region, 9 satisfied criteria (HWE >0.01, MAF >0.01). The aggressive tagging algorithm embedded in Tagger was used with 2- and 3-marker haplotypes and r^2 threshold set at a minimum of 0.8. 6 tSNPs were selected: rs1378640, rs394452, rs3731881, rs3731878, rs427587 and rs3100776. rs3731878 failed manufacture by ABI. The remaining 5 tSNPs were genotyped by TaqMan in the Scottish IBD population consisting 448 patients with CD, 549 with UC and 428 HC. This represents the same cohort genotyped for *GLI1* in Chapter 4, with the exception of the reduced number of controls. A total of 95 samples were successfully genotyped in duplicate, giving 99.6% concordance across all SNPs. All SNPs were in HWE in the HC population (p>0.1). Genotypes were called in 94.7% (rs1378640), 87.8% (rs394452), 97.0% (rs3731881), 99.3% (rs427587) and 99.3% (rs3100776) of the total population (n=1425).

7.4 *Results*

7.4.1 Log-likelihood analysis

There was no overall association with *IHH* variation and IBD (p=0.13), UC (p=0.16) or CD (p=0.23) on log-likelihood analysis (model-free analysis) with 1000 permutations and 31 degrees of freedom.

7.4.2 Haplotype analysis

In the HC population there were 5 common haplotypes (frequency >1.0%), designated A-E (**Table 7-1**). The LD between the 5 tSNPs is demonstrated by D′ and r^2 values in **Figure 7-1**. There was no difference in the frequency of haplotypes A-E between IBD and HC populations (**Table 7-1**). Analysis of haplotype frequencies in UC demonstrated association with haplotype C (10.4% vs. 7.7%, p=0.042, OR 1.39, C.I. 1.01-1.92) and protection with haplotype D (2.3% vs. 3.9%, p=0.041, OR 0.58, C.I.

0.34-0.98). Haplotype E was associated with CD (2.1% vs. 0.6%, p=0.0072, OR 3.60, C.I. 1.33-9.73) but not with UC.

7.4.3 tSNP analysis

Analysis of the 5 individual tSNPs by chi^2, demonstrated no association with $tSNPs_{1-3}$ (**Table 7-2**). $tSNP_4$ (rs427857) was associated with CD (MAFs 0.021 CD vs. 0.006 HC, p=0.0075, OR 3.57, C.I. 1.32-9.66) but not with IBD or UC. No patients or controls were homozygous mutant at rs427857. $tSNP_5$ (rs3100776) was associated with UC (MAFs 0.023 UC vs. 0.039 HC, p=0.043, OR 0.58, C.I. 0.34-0.99). However, only one UC patient and one HC was homozygous for the $tSNP_5$ mutation.

7.4.4 Genotype-phenotype analysis

There were no significant genotype-phenotype associations when log-likelihood, haplotype and tSNP analysis was stratified for age at diagnosis, disease location or disease behaviour. Specifically, given the borderline associations with CD and UC for haplotypes C-E and $tSNPs_{4-5}$, analysis of colonic IBD was performed. Case-control analysis (colonic IBD vs. HC) demonstrated borderline association with $tSNP_5$ only (0.024 vs. 0.039, p=0.046, OR 0.61, C.I. 0.37-1.00). There was, however, no association on a within cases analysis (colonic IBD vs. not colonic IBD) for this (p=0.21) or other SNPs.

7.5 *Discussion*

In the present study a gene-wide haplotype-tagging approach was taken to determine whether variation in *IHH* is associated with IBD in the Scottish population. Overall gene variation, analysed by the log-likelihood test, was not associated with IBD, UC or CD. There was no associated with IBD on analysis of individual haplotypes or genotypes. However, given the down-regulated IHH mRNA expression documented in UC biopsies, it was of particular interest to determine whether or not there was any evidence of germ-line *IHH* variation association with UC. Of note, association of borderline significance (p=0.04) was shown for haplotype C (11211) as well as protection with haplotype D (21212). This association may in part be related to the protective association documented with tSNP5 (p=0.043). The LD between this intronic SNP and the other tSNPs is 1.0 by D′ but 0.0 – 0.13 by r^2. The low r^2 values would explain association at $tSNP_5$ in UC in the absence of association at $tSNPs_{1-4}$,

assuming this is real. It would, however, critically require replication in an independent UC population.

The associations seen for CD are of doubtful significance. Both haplotype E and $tSNP_4$ are present at very low frequencies in the HC population (0.6% for each). In the absence of a CD replication cohort in this study it is therefore best to interpret these findings with caution.

Overall, particularly given the negative log-likelihood analysis (the primary analysis designed for this type of gene-wide haplotype-tagging strategy), it is concluded that there is no substantial evidence for association of *IHH* with IBD in the Scottish population. Alternative explanations for the decreased mRNA IHH expression in UC should therefore be offered, including the possibility of epigenetic modification.

Haplotypes		**IBD**	**CD**	**UC**	**HC**	**IBD vs. HC** p-value OR (C.I.)	**CD vs. HC** p-value OR (C.I.)	**UC vs. HC** p-value OR (C.I.)
A	**11111**	65.5%	65.9%	67.0%	68.4%	0.32 OR 0.92 (0.77-1.09)	0.27 OR 0.89 (0.73-1.09)	0.52 OR 0.94 (0.77-1.14)
B	**22211**	19.9%	20.6%	19.3%	19.2%	0.67 OR 1.04 (0.85-1.28)	0.47 OR 1.09 (0.86-1.39)	0.98 OR 1.00 (0.80-1.26)
C	**11211**	9.3%	7.9%	10.4%	7.7%	0.19 OR 1.22 (0.91-1.64)	0.93 OR 1.02 (0.71-1.45)	0.042 OR 1.39 (1.01-1.92)
D	**21212**	2.8%	3.3%	2.3%	3.9%	0.11 OR 0.70 (0.45-1.10)	0.47 OR 0.83 (0.50-1.39)	0.041 OR 0.58 (0.34-0.98)
E	**22221**	1.4%	2.1%	0.9%	0.6%	0.06 OR 2.41 (0.93-6.29)	0.0072 OR 3.60 (1.33-9.73)	0.50 OR 1.46 (0.49-4.36)

Table 7-1 Haplotype structure of *IHH* tSNPs.
All haplotypes with frequency >1% are shown (A-E). Case-controlled analysis of haplotype frequency is demonstrated for IBD vs. HC, CD vs. HC, UC vs. HC (Chi^2 analysis of haplotype frequencies estimated by Haploview; 2-tailed p-values given with odds ratios and 95% confidence intervals).

	IBD MAF	**CD MAF**	**UC MAF**	**HC MAF**	**IBD vs. HC** p-value OR (C.I.)	**CD vs. HC** p-value OR (C.I.)	**UC vs. HC** p-value OR (C.I.)
rs1378640	0.237	0.256	0.221	0.238	0.97 OR 1.00 (0.82-1.21)	0.40 OR 1.10 (0.88-1.38)	0.42 OR 0.91 (0.73-1.14)
rs394452	0.199	0.206	0.194	0.185	0.41 OR 1.10 (0.88-1.37)	0.31 OR 1.14 (0.88-1.48)	0.65 OR 1.06 (0.83-1.36)
rs3731881	0.326	0.342	0.331	0.319	0.38 OR 1.08 (0.91-1.29)	0.32 OR 1.11 (0.90-1.36)	0.57 OR 1.06 (0.87-1.29)
rs427857	0.014	0.021	0.009	0.006	0.064 OR 2.40 (0.92-6.26)	0.0075 OR 3.57 (1.32-9.66)	0.50 OR 1.45 (0.49-4.35)
rs3100776	0.027	0.032	0.023	0.039	0.092 OR 0.68 (0.44-1.07)	0.42 OR 0.81 (0.48-1.36)	0.043 OR 0.58 (0.34-0.99)

Table 7-2 Minor allelic frequencies (MAF) of individual *IHH* tSNPs (1-5) in IBD, CD, UC and HC.
Case-controlled analysis of allelic frequency is demonstrated for IBD vs. HC, CD vs. HC, UC vs. HC (Chi^2 analysis, 2-tailed p-values given with odds ratios and 95% confidence intervals).

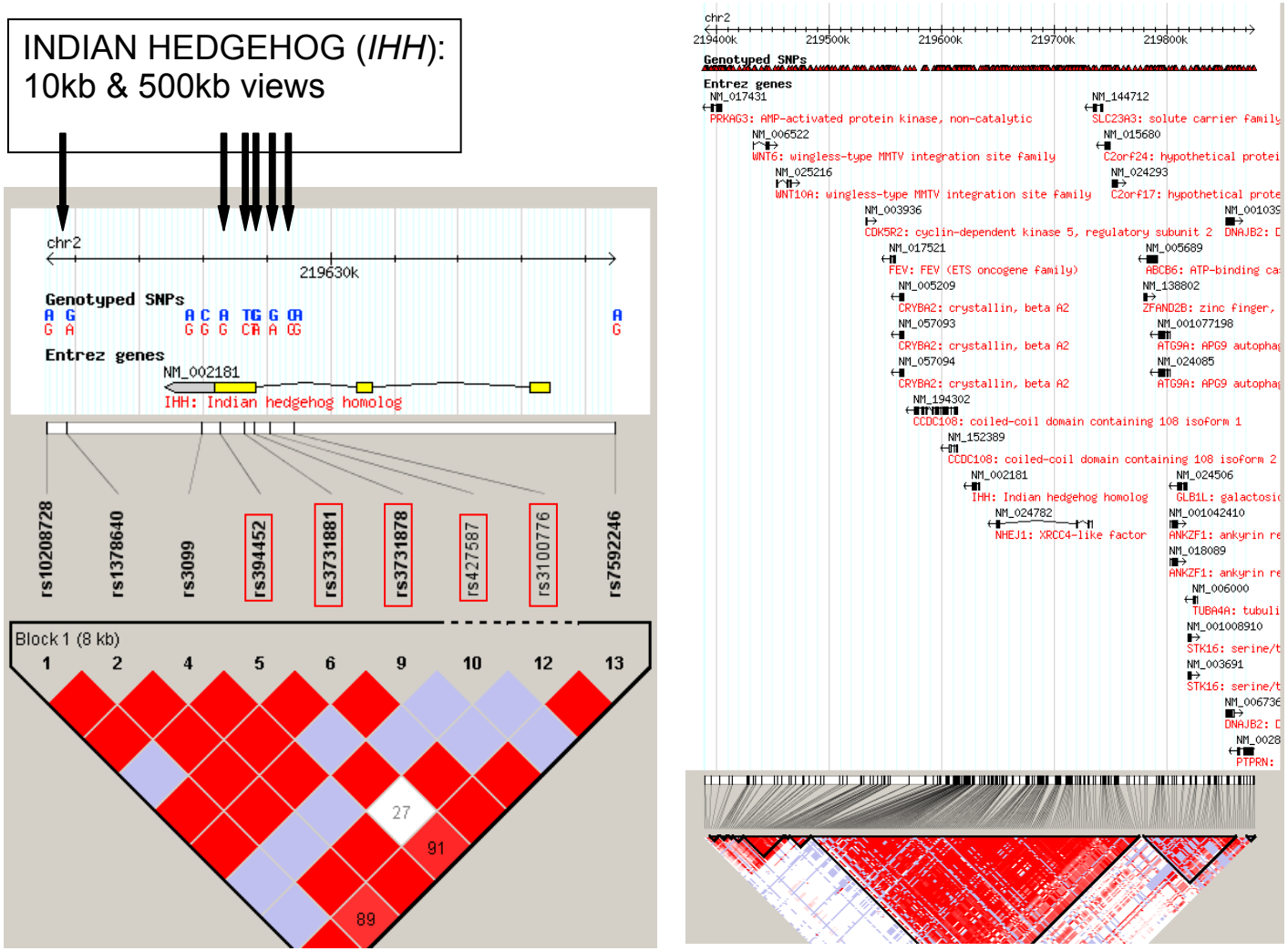

Figure 7-1 Representation of *IHH* haplotype structure from HapMap CEPH data in Haploview.

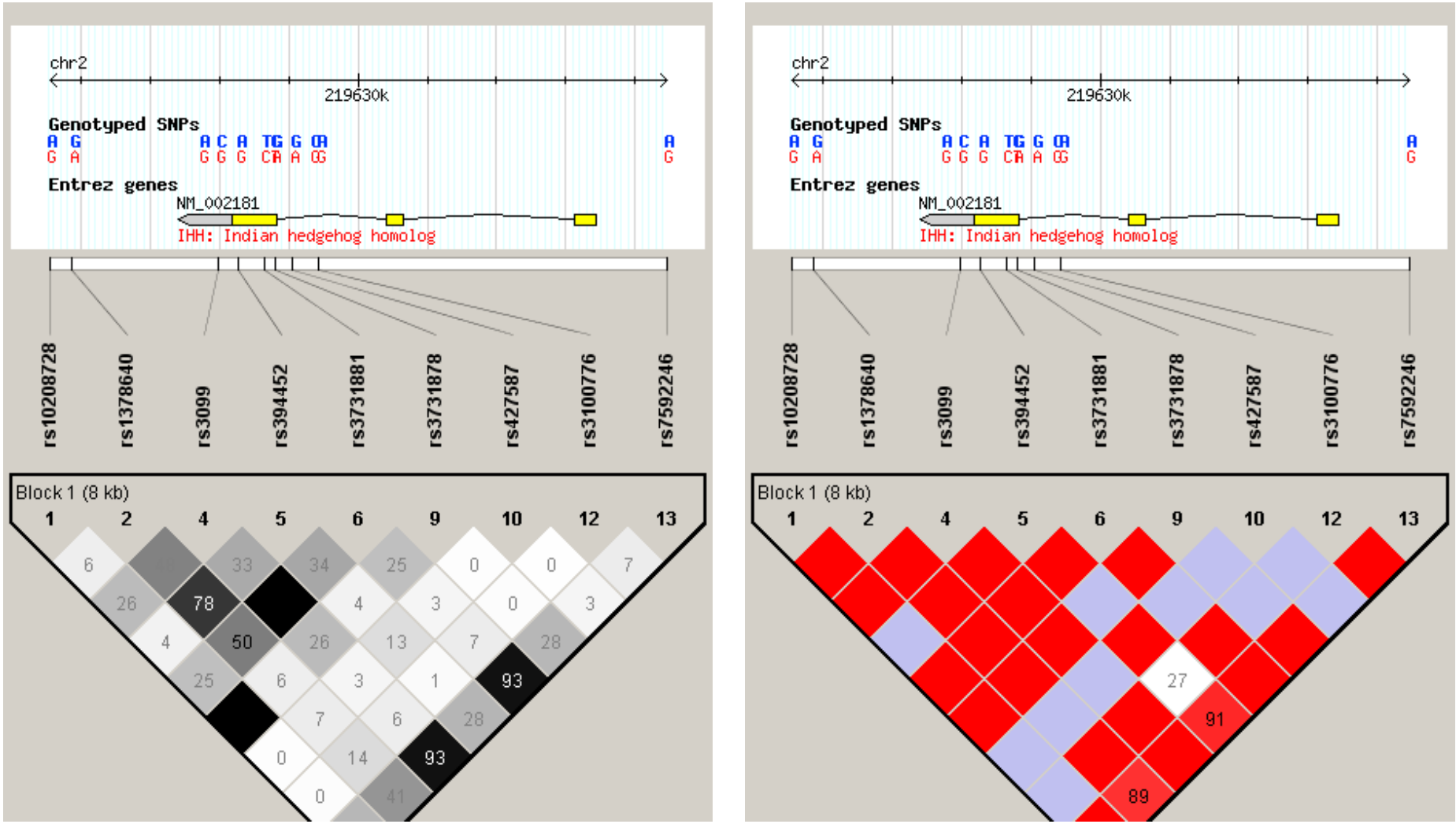

Figure 7-2 Linkage disequilibrium by r^2 (Left panel) and D′ (right panel) across *IHH* (HapMap CEPH data).

8 Germ-line variation in 13 HH and 27 WNT signalling genes in Crohn's disease: analysis of WTCCC GWAS data.

8.1 *Abstract*

Background. It has been demonstrated that the expression of key HH response genes, notably *GLI1* and *PTCH*, is dysregulated not only in UC, but also CD (Chapter 4). This is one of several lines of evidence that point to a potential role for HH signalling in CD pathogenesis. In addition, several WNT pathway genes showed differential expression in CD, and there was further evidence of interaction between these pathways in the intestine.

Aims. The aim of the present study was to analyse WTCCC data for 13 HH and 27 WNT signalling pathway genes for evidence of association with CD.

Methods. The WTCCC CD dataset was interrogated for all SNPs present within HH genes (*SHH, IHH, DHH, DISP1, PTCH, PTCH2, SMO, HHIP, SUFU, GLI1, GLI2, GLI3 AND KCTD11*), and those WNT signalling genes with known intestinal expression (*WNT2B, WNT3, WNT4, WNT5A, WNT5B, WNT6, WNT9B, FZD4, FZD5, FZD6, FZD7, LRP5, LRP6, sFRP1, sFRP5, DKK3, DKK4, TCF1, TCF2, TCF3, TCF4, APC, CTNNB1, AXIN1, ATOH, GSK3α,* and *GSK3β*). A total of 1Mb of data was analysed at every gene location. Two major analyses were performed for every SNP: CD (n=2000) vs. CC (consisting of non-CD WTCCC cases, n=12,000; analysis 1) and CD vs. HC (WTCCC controls, n=3000; analysis 2). A series of automated functions were generated in Microsoft Excel to process this data and generate chi^2 statistics with 2-tailed p-values and odds ratios with 95% confidence intervals. A Bonferroni correction was applied separately to each pathway analysis giving α-values of 0.00028 for HH pathway and 0.00019 for WNT pathway genes. Confirmatory data for SNPs significant in analysis 2 ($\alpha = 0.05$) was sought from the meta-analysis of WTCCC, French / Belgium and N American GWAS.

Results. 179 WTCCC SNPs were present in 10/13 HH genes (no SNPs in *IHH*, *DHH* and *KCTD11*) and 257 SNPs in 21/27 WNT genes (no SNPs in *FZD5, FZD7, sFRP5, TCF3, ATOH* and *GSK3α*). 24/179 (13.4%) HH pathway SNPs were significant at $p < 0.05$ on analysis 1; 17/24 remained significant on analysis 2 with no new signals detected. Only one HH pathway gene was significant after Bonferonni correction: rs147172020, a rare SNP (MAF 0.0001) in *GLI3* ($p=9.4 \times 10^{-6}$, OR 15.8) not present on the meta-analysis; 43 other SNPs in this gene were not associated with CD. 4 *GLI2* SNPs forming a single haplotype block in an intronic region were associated with CD (OR 1.09, p=0.043 on analysis 1), but lost significance in the meta-analysis

(p=0.23). 10 adjacent SNPs located within a large block of tight LD in *SUFU* were associated with CD (OR 0.92); this haplotype block extended into neighbouring genes. All of these SNPs were significant on analysis of meta-analysis data (p 0.0047 – 0.0093). Of 36/257 WNT pathway SNPs associated with CD, only 2 remained significant after Bonferonni correction: rs2083798 (OR 1.20, p=0.00015) and rs2083797 (OR 1.20, p=0.00018) in two adjacent genes – *WNT3* and *WNT9b* respectively. Whilst these SNPs remained significant in the meta-analysis, there was evidence of convergene (p=0.010) due to lack of signal in the ileal only N American GWAS dataset.

Conclusions. This analysis of WTCCC data has provided preliminary evidence for association at *SUFU* and at the *WNT3 / WNT9b* locus. *SUFU* is noteworthy as it is a direct binding partner of *GLI1*, but it is presently not possible to ascribe the association directly to *SUFU* as it is in a very large block of tight LD. The *WNT3 / WNT9b* locus is of considerable interest as both of these genes are highly expressed in Paneth cells. However, given the absence of a signal in the purely ileal cohort it is difficult to argue biological plausibility for this association. Both of these datasets would require urgent replication in large independent cohorts, with additional fine-mapping studies, before firm conclusions can be drawn.

8.2 *Introduction*

In Chapter 4 it was demonstrated that key HH pathway components are differently expressed in both CD and UC when compared with the healthy colon (**Figures 4-4** and **4-6**). Whilst there was no consistent expression changes between CD and HC for SHH and IHH, the HH pathway response elements PTCH and GLI1 were decreased in colonic CD compared with comparable HC biopsies (**Figure 4-6**). The observed decrease in HH pathway activity in CD, coupled with the published report of an Ashkenazi pedigree with significant overlap of CD and basal cell naevus syndrome (mutant *PTCH*)[414, 415] and the purported role of HH in Paneth cell differentiation,[332] suggest that HH may play an important role in CD pathogenesis.

Whislt the WNT pathway has central roles in the maintenance of intestinal stem cells and in colon cancer pathogenesis, relatively little is known about its function in response to acute or chronic inflammatory challenge in the intestine. The intestinal expression profiles of WNT pathway components have been carefully described.[398] In Chapter 4, relatively modest expression changes were noted in several of the intestinally expressed WNTs in UC and CD (**Table 4-3** and **Figure 4-11**).

The data from the WTCCC GWAS, discussed in detail in Chapter 1 (**1.5.2.3**), has been made publically available.[81] In keeping with one of the stated aims of the study, the summary genotype statistics are free for download to anyone with an internet connection; the full raw datasets available on application to qualified researchers. As such, data are available on the controls samples (1958 British Birth Cohort and UK Blood Service) and the diseases studied in phase I (CD, rheumatoid arthritis, coronary heart disease, hypertension, type I and type II diabetes mellitus, and bipolar disorder).

Given the potential role of HH signalling in CD pathogenesis, and the documented role of germline *GLI1* variation in IBD pathogenesis (Chapter 5), it was hypothesised that germ-line variation in other HH pathway genes was associated with CD. The aim of this study was therefore to mine the available WTCCC data for *SHH, IHH, DHH, DISP1, PTCH, PTCH2, SMO, HHIP, GLI1, GLI2, GLI3, SUFU* and *KCTD11* and perform a case-controlled study for all available SNPs. Furthermore, given the likely cross-talk between HH and WNT signalling in the healthy and inflamed intestine

(Chapters 1, 2 and 4), all 27 WNT signalling genes with known intestinal expression signatures were analysed for association in the same manner. For both datasets, the meta-analysis of U.K., European and N American CD GWAS data was used for initial replication of positive associations.[103]

8.3 *Methods*

8.3.1 Analysis of WTCCC data

In the main WTCCC study, 17,000 samples (14,000 cases and 3,000 shared healthy controls) were genotyped on GeneChip 500K Mapping Array Set (Affymetrix chip), comprising 500,568 SNPs.[81] Genotypes for each SNP for every individual were called using a new, custom-designed programme CHIAMO, with only 0.16% discordance on analysis of 10^8 duplicate genotypes. 93.8% of SNPs passed WTCCC quality control; of these, 392,575 had MAF >1%, 45,106 MAF <0.1%.[81]

The summary genotype statistics for CD from the WTCCC study were downloaded from http://www.wtccc.org.uk/info/summary_stats.shtml. These text files provide the following data for each SNP: SNP identification (Affymetrix id and rs number), chromosomal position and alleles, the average maximum posterior call (a measure of the certainty of the calls), genotypes for cases and controls (actual numbers given for 11, 12 and 22 at each SNP), as well as p-values (calculated by both an additive and a general model) and an assessment of clustering (good or bad). For CD, 2 main control datasets are available: 1) all other WTCCC cases combined (n=12,000; designated CC) and 2) the main WTCCC control cohort (n=3000; HC). Replication of all SNPs that attained significance ($p<0.05$) on analysis 2 (CD vs. HC) was performed by analysis of CD vs. HC data from the meta-analysis of U.K., European and N American GWAS (full description of the meta-analysis and methods is detailed in **1.5.3**), performed by Jeffrey Barrett.[103]

The chromosomal location of all HH and WNT pathway genes was determined by HapMap / Ensembl, and the number / position of WTCCC SNPs within each gene noted (**Tables 8-1 and 8-3 and Figures 8-1 – 8-10**). The haplotype structure for each gene was downloaded from HapMap (Data Release 23a/phaseII March 08, on NCBI B36 assembly, db SNP b126) and visualised in Haploview (version 4.2) with the

position of WTCCC SNPs plotted. 1Mb of WTCCC data around each gene was extracted into a Microsoft Excel spreadsheet for analysis. Given the large amount of data requiring analysis, a series of functions were automated within Excel to calculate allelic call rates and frequencies, chi^2 analysis (2-tailed, with and without Yates correction), odds ratio and 95% confidence intervals. Data for all genes were processed through this algorithm for two primary analyses: CD vs. CC (analysis 1) and CD vs. HC (analysis 2). A significance threshold was initially set at the conventional alpha level of < 0.05; given the candidate gene/pathway – based analysis performed, genome-wide levels of significance are not appropriate for this study. In addition, a Bonferroni correction was made to each pathway dataset based on the number of SNPs analysed: $p < 0.00028$ for HH (179 SNPs) and $p < 0.00019$ for WNT (257 SNPs). Further to identifying those SNPs within HH and WNT genes that reached significance in both analyses, SNPs in neighbouring regions were scanned to detect, for example, promoter polymorphisms that achieved significance.

8.4 *Results*

8.4.1 Analysis of WTCCC data for HH signalling components in CD

Out of the 13 HH pathway genes analysed, no WTCCC SNPs were present in *IHH, DHH* and *KCTD11* and there was only one SNP in *SHH* (**Table 8-1**). The position of all WTCCC SNPs in the 10/13 genes with genotype data available are demonstrated on Haploview gene maps for *GLI3* (**Figure 8-1**), *SUFU* (**Figure 8-2 and 8-3**), *GLI2* (**Figure 8-4, 8-5 and 8-6**), *DISP1* (**Figure 8-7**), *HHIP* (**Figure 8-8**), *SHH* (**Figure 8-9**), *PTCH* (**Figure 8-10**), *PTCH2* (**Figure 8-11**), *SMO* (**Figure 8-12**) and *GLI1* (**Figure 8-13**).

In total, 24 SNPs of 179 (13.4%) analysed within HH pathway genes attained significance on chi^2 analysis of CD vs. CC (analysis 1): 1 in *GLI3*, 13 in *SUFU*, 8 in *GLI2*, 1 in *DISP1* and 1 in *HHIP* (significant SNPs denoted with red arrows in **Figures 8-1 – 8-8**). However, none of these SNPs remained significant after Bonferroni correction. On analysis of CD vs. HC (analysis 2), 17/24 SNPs remained significant: 1 in *GLI3*, 10 in *SUFU*, 4 in *GLI2*, 1 in *DISP1* and 1 in *HHIP* (**Table 8-2**). No additional SNPs were significant on analysis of CD vs. HC.

The most significant SNP was rs147172020 in *GLI3* ($p = 9.4 \times 10^{-6}$ and 0.024 on analyses 1 and 2 respectively) with OR 15.79. However, this was a very rare polymorphism with MAF of 0.0017 in the CD population and no mutant homozygotes in either population. Furthermore, an additional 43 *GLI3* SNPs analysed did not reach significance. Meta-analysis data for rs147172020 was not available.

10/22 SNPs in *SUFU* were significant on both analyses (**Table 8-2**). There is a very large area of tight LD extending throughout and beyond this 129kb gene (**Figures 8-2 and 8-3**). As such the significance level for all of these genes was roughly equivalent ($p = 0.016 – 0.036$ on analysis 1) with a consistent OR of 0.92. All these SNPs achieved greater levels of significance for association with CD on analysis of GWAS meta-analysis data ($p = 0.0047 – 0.0082$). Furthermore, the haplotype block extends into neighbouring genes, including *ACTR1A* (ARP1 actin-related protein 1 homolog A), *TRIM8* (tripartite motif-containing 8) and *NFKB2* (nuclear factor of kappa light polypeptide gene).

8/47 SNPs in *GLI2* reached significance on analysis 1, although 4/8 dropped out in analysis 2. Of particular note was a cluster of 4 adjacent SNPs in an intronic region of *GLI2*, corresponding to a discrete block of LD (block 3 on **Figure 8-5**). These 4 SNPs were significant on both analysis 1 and 2 ($p=0.02$ and 0.04 respectively) with OR 1.09. The other 4 *GLI2* SNPs were only significant on analysis 1; 3 of these SNPs had MAF of < 0.01. However, these 4 SNPs were no longer significant in the GWAS meta-analysis (**Table 8-2**).

rs6668249, the one significant SNP in *DISP1* out of 31 analysed, had a p-value of 0.025 in both analyses with OR 0.92. Of note, this SNP was in a haplotype block spanning a large intronic region in which the other 9 SNPs analysed did not attain significance. rs6668249 was also not significant in the meta-analysis (**Table 8-2**). The *HHIP* SNP, rs973796, was located within the promoter region of the gene; 16 other SNPs within the same large block of LD spanning *HHIP* were not associated with CD.

8.4.2 Analysis of WTCCC data for WNT signalling components in CD

27 genes from the WNT / β-catenin signalling pathway were chosen for analysis based on detailed published healthy adult expression patterns in the murine intestine.[398] Of these, 21 had WTCCC SNPs within the gene itself (**Table 8-3**). As the HH analysis of CD vs. CC generated 7/24 2(9.2%) false positives but no false negative, only CD vs. HC analysis was performed for these WNT genes. 36/257 (14.0%) SNPs were associated with CD; 2 were significant at the Bonferroni corrected threshold of p<0.00019 (in *WNT3* and *WNT9b*), 4 at p<0.01 (in *FZD6*, *sFRP1* and *TCF4*) and a further 30 at <0.05 (*WNT6*, *LRP5*, *LRP6*, and *TCF-2*).

The two most significant SNPs were located at the *WNT3* and *WNT9b* locus on chromosome 17: rs2083798 (p=0.00015, OR 1.20 (C.I. 1.09-1.32) and rs2083797 (p=0.00018, OR 1.20, C.I. 1.09-1.32) (**Figure 11a**). Whilst rs2083798 and rs2083797 remained significant in the meta-analysis, there was evidence of convergence for both SNPs (p = 0.010).

8.5 *Discussion*

In the present study, preliminary evidence is provided from publically available WTCCC data for association of 5 HH pathway genes *GLI3*, *SUFU*, *GLI2*, *DISP1,* and *HHIP* with CD. Only one SNP each in *GLI3*, *DISP1* and *HHIP* was associated, and these may be of less significance. In contrast, multiple adjacent SNPs were associated in *SUFU* and *GLI2*, and these should be prioritised for further analysis. Along with *GLI1* (Chapter 4), all 3 of these genes are involved in the cellular transduction of the HH signal. It is noteworthy that SUFU binds directly to both GLI1 and GLI2, thus hinting at a potential functional mechanism to link with the association of these 3 HH pathway genes. However, after analysing meta-analysis data from WTCCC, French/Belgium and N American GWAS studies, only *SUFU* remained as a potential CD susceptibility gene in the HH pathway, with evidence for a mild protective effect in 10 adjacent SNPs (OR 0.92). This region requires further independent replication before reaching any firm conclusions. However, given the strength of association observed any replication cohort will need to be suitably powered, such as the WTCCC CD replication cohort (n = 1,182 CD and 2,024 HC) within the UK.[89]

Were this association to hold up after further replication, fine-mapping studies would be required. As *SUFU* is located in an area of extensive tight LD containing several other genes (*ACTR1A, TRIM8* and *NFKB2*), it is presently not possible to attribute the association to this gene. It is also of some interest that *SUFU* is located only 24.9Mb from *DLG5* which, despite initial promise as a CD gene, has failed to be consistently replicated across multiple populations.[78] Determination of causal variants in this region would require a very large study, with hundreds of SNPs genotyped in several thousand patients and controls. Even then, it may be impossible to determine causality given the small effect size and tight LD. For example, it has not yet been possible over the past 7 years to ascribe causality to an individual gene, let alone individual mutations in the IBD5 region.

The design of the present study used two different control populations for comparison of SNPs within HH genes with the CD cohort. The use of the combined cases as a control cohort provided increased power to detect small association signals, but none of the 12,000 samples are true HCs. Thus, there is the possibility of both false positive and false negative results. To counter this, all regions were additionally analysed against the true HC population. No additional SNPs were associated with CD on analysis 2. However, 29.2% of SNPs associated on analysis 1 dropped out at this stage. The allelic frequencies were very similar between both 'control' populations with no significant differences between the populations.

Because of this, on analysing 27 WNT pathway genes, only analysis 2 was performed. Whilst there were several genes with modest evidence for association with CD, the *WNT3 / WNT9b* region is notable for a number of reasons. Firstly, the two mostly significant SNPs in this region satisfied stringent correction for multiple testing by Bonferroni method. Secondly, the associated SNPs span two adjacent genes within the WNT pathway. Thirdly, these genes are of significant biological interest as they are highly expressed in murine Paneth cells.[398] It was therefore anticipated, that if the association at this loci were real, the significance level would increase in the meta-analysis containing the N American cohort with ileal only disease. However, the meta-analysis at the two most significant SNPs demonstrated substantial convergence. There are two possible explanations for this observation. Either, the observed

association at this locus is not real, or the expression of WNT3 and WNT9b is misleading and these genes are associated with colonic disease. The microarray analysis in Chapter 4 demonstrated up-regulation of WNT9b in inflamed UC (p=0.022), but no alterations for either gene in colonic or terminal ileal CD.

Of the other WNT genes with significant associations described here, the expression profiles of WNT6, LRP6, TCF2 and TCF4 are of interest (**Table 4-3**). WNT6 was decreased in colonic CD but, of note, only with non-inflamed samples (p=0.0003). LRP6 and TCF2 were both decreased in inflamed colonic CD (p=0.001 and 0.0025 respectively). TCF4 expression was higher in UC (p=0.021) and colonic CD (p=0.0038), but dysregulated expressed was not seen in terminal ileal CD. This is in contrast to the report by Wehkamp and Stange, who have described decreased TCF4 expressoin in terminal ileal CD, irrespective of inflammatory status and independent of *NOD2* genotype.[421]

This study additionally illustrates the variation in tagging density of certain genomic regions on the Affymetrix 500k chip (**Figures 8-1 – 8-15**). As demonstrated, it is particularly poor at tagging small genes. In addition to performing replication of the associated SNPs from this preliminary analysis, a detailed gene-wide haplotype-tagging strategy should be employed for all of these genes before finally concluding lack of association.

The use of the publically available WTCCC data has provided potentially important insights into the role of germ-line variation of HH signalling components (*SUFU* in addition to the previous evidence for association at *GLI1*) and WNT signalling genes (*WNT3* / *WNT9b*) in CD pathogenesis. All of these genes are critical to the normal development of the gastrointestinal tract, as is the homeodomain gene *NKX2.3*, a HH target and confirmed IBD susceptibility gene. It would be of great interest to extend this approach to other developmental signalling pathways of established homeostatic importance in the GI tractment (e.g. Notch, BMP and hox and sox genes). Given all of this data, a pathway-based analysis such as that recently reported for axon-guidance pathway genes from GWAS data in Parkinson's disease,[464] could then be performed.

GENE	Chr.	Region (bp)	Gene size (kb)	No. WTCCC SNPs in gene
Sonic hedgehog (SHH)	7q36.3	155288319 - 155297728	9.4	1
Indian hedgehog (IHH)	2q35	219627391 – 219633433	6.0	0
Desert hedgehog (DHH)	12q13.1	47769475- 47774869	5.4	0
Dispatched (DISP1)	1q41	221168406 – 221245960	77.6	31
Patched (PTCH)	9q22.3	97245083 – 97318923	73.8	6
Patched-2 (PTCH2)	134.1	45060675 – 45081203	20.5	3
Smoothened (SMO)	7q32.1	128615949 – 128640617	24.7	4
Hedgehog-interacting Protein (HHIP)	4q31.2	145786623 – 145879337	92.7	17
GLI1	12q13.3	56140201 – 56152312	12.1	3
GLI2	2q14.2	121266327 – 121466321	200.0	47
GLI3	7p14.1	41967075 – 42243137	276.1	44
Suppressor of Fused (SUFU)	10q24	104253734 – 104383282	129.5	22
KCTD11	17p13.1	7196205 – 7198984	2.8	0

Table 8-1 HH pathway genes.
Details are given of the major HH pathway genes, including genomic location, and the number of SNPs from the WTCCC GWAS present within each gene.

GENE	rs Number	Position	1/2	CD MAF	HC1 MAF	HC2 MAF	Analysis 1: CD vs CC: p	Analysis 2: CD vs HC: p	Meta-analysis: p	OR (CD vs HC1)
GLI3	rs17172020	41980236	A/T	0.0017	0.0001	0.0002	$9.4x10^{-6}$	0.024	n/t	15.79 (3.19-78.3)
SUFU	rs10786680	104307182	G/A	0.4118	0.4316	0.4322	0.028	0.051	0.0093	0.92 (0.86-0.99)
SUFU	rs10786682	104308312	A/C	0.4107	0.4316	0.4324	0.020	0.037	0.0068	0.92 (0.85-0.99)
SUFU	rs10883736	104310019	T/G	0.4109	0.4317	0.4325	0.021	0.039	0.0070	0.92 (0.85-0.99)
SUFU	rs10786684	104313018	A/C	0.4103	0.4316	0.4325	0.018	0.034	0.0058	0.92 (0.85-0.99)
SUFU	rs10786685	104313093	T/C	0.4103	0.4299	0.4315	0.029	0.043	0.0072	0.92 (0.86-0.99)
SUFU	rs729023	104331877	A/G	0.4110	0.4325	0.4332	0.017	0.034	0.0047	0.92 (0.85-0.98)
SUFU	rs11191347	104341607	T/C	0.3355	0.3537	0.3524	0.036	0.094	0.0064	0.92 (0.86-1.00)
SUFU	rs4917976	104350508	A/G	0.4101	0.4316	0.4323	0.016	0.033	0.0053	0.92 (0.85-0.98)
SUFU	rs7075269	104355714	T/C	0.4103	0.4314	0.4323	0.019	0.036	0.0058	0.92 (0.85-0.99)
SUFU	rs4919665	104356032	C/T	0.4109	0.4320	0.4324	0.019	0.040	0.0064	0.92 (0.85-0.99)
SUFU	rs4917978	104356343	C/T	0.4109	0.4317	0.4321	0.020	0.043	0.0069	0.92 (0.85-0.99)
SUFU	rs7907417	104360652	A/G	0.4106	0.4316	0.4322	0.019	0.038	0.0062	0.92 (0.85-0.99)
SUFU	rs12414407	104377009	A/G	0.3446	0.3637	0.3609	0.029	0.105	0.0082	0.92 (0.85-0.99)
GLI2	---	121277641	A/G	0.4048	0.3839	0.3836	0.020	0.044	n/t	1.09 (1.01-1.17)
GLI2	rs12475334	121278033	T/C	0.4036	0.3834	0.3823	0.024	0.043	0.233	1.09 (1.01-1.17)
GLI2	rs746345	121278448	C/A	0.4043	0.3836	0.3832	0.022	0.047	0.217	1.09 (1.01-1.17)
GLI2	rs4848632	121278825	T/C	0.4042	0.3834	0.3833	0.020	0.047	0.207	1.09 (1.01-1.17)
GLI2	rs4378815	121419363	G/C	0.0140	0.0099	0.0104	0.031	0.138	n/t	1.44 (1.05-1.97)
GLI2	rs2311803	121419853	T/C	0.0137	0.0098	0.0102	0.042	0.149	n/t	1.42 (1.03-1.94)
GLI2	rs17005499	121425671	A/G	0.0034	0.0014	0.0019	0.014	0.210	n/t	2.49 (1.27-4.86)
GLI2	rs12711535	121425766	G/A	0.0607	0.0514	0.0518	0.025	0.074	0.058	1.20 (1.03-1.39)
DISP1	rs6668249	221205568	A/G	0.4456	0.4658	0.4693	0.025	0.025	0.100	0.92 (0.86-0.99)
HHIP	rs973796	145781573	C/A	0.3670	0.3869	0.3927	0.024	0.013	0.042	0.92 (0.85-0.99)

Table 8-2 WTCCC data analysis on HH pathway genes.
Minor allelic frequencies (MAF) are given for WTCCC CD population, all non-CD WTCCC cases combined (CC) and WTCCC healthy controls (HC). All SNPs within hedgehog pathway genes (detailed in Table 1) from WTCCC GWAS that reached significance ($p < 0.05$) on analysis 1 (CD vs. CC) are listed, followed by analysis 2 (CD vs. HC) along with odds ratio and 95% confidence intervals derived from analysis 1. The meta-analysis p-values is presented for those SNPs present in the meta-analysis of WTCCC, French / Belgium and N American GWAS studies (data provided by Jeffrey Barrett). Those SNPs not tested in the meta-analysis are indicated by 'n/t'.

GENE	Chr.	Position (bp)	Gene Size (kb)	No. WTCCC SNPs in gene
WNT2b	1	112,811,563 – 112,865,428	53.9	4 (0)
WNT3	17	42,196,859 – 42,251,081	54.2	6 (1+) ***
WNT4	1	22,318,787 – 22,342,197	23.4	7 (0)
WNT5a	3	55,474,784 – 55,496,371	21.6	3 (0)
WNT5b	12	1,596,483 – 1,626,638	30.2	10 (0)
WNT6	2	219,432,790 – 219,447,198	14.4	3 (1) *
WNT9b	17	42,283,967 – 42,309,436	25.5	5 (2+) ***
FZD4	11	86,334,370 – 86,344,081	9.71	1 (0)
FZD5	2	208,335,557 – 208,342,388	6.83	0 (0)
FZD6	8	104,380,276 – 104,414,268	34.0	7 (3) **
FZD7	2	202,607,555 – 202,611,405	3.85	0 (0)
LRP5	11	67,836,684 – 67,973,319	136.6	12 (3) *
LRP6	12	12,164,958 – 12,311,013	146.1	11 (3) *
sFRP1	8	41,238,636 – 41,286,137	47.5	12 (4) **
sFRP5	10	99,516,508 – 99,521,760	5.25	0 (0)
DKK3	11	11,941,121 – 11,986,762	45.6	19 (0)
DKK4	8	42,350,744 – 42,353,831	3.09	1 (0)
TCF1	12	119,900,932 – 119,924,695	23.8	5 (0)
TCF2	17	33,120,548 – 33,179,182	58.6	20 (2) *
TCF3	19	1,560,295 – 1,601,277	41.0	0 (0)
TCF4	18	51,045,967 – 51,406,441	360.5	55 (10) **
APC	5	112,101,483 – 112,209,834	108.4	6 (0)
CTNNB1	3	41,216,016 – 41,256,938	40.9	3 (0)
AXIN1	16	277,441 – 342,465	65.0	10 (0)
ATOH	4	94,969,101 – 94,970,165	1.07	0 (0)
GSK3α	19	47,426,178 – 47,438,576	12.4	0 (0)
GSK3β	3	121,028,238 – 121,295,203	360.0	18 (0)

Table 8-3 WNT pathway genes.
Details of the WNT genes included in analysis of WTCCC data, including chromosomal position, gene size, number of WTCCC SNPs in each gene (plus number of significant SNPs ($p<0.05$) in each gene). * $p<0.05$, ** $p<0.01$, *** $p<0.00019$ (Bonferroni correct significance threshold).

GENE	rs number	Position	Allele 1/2	CD MAF	HC MAF	CD vs. HC:p	Meta-analysis: p	OR (C.I.)
WNT3	rs199515	42211804	C/G	0.193	0.223	0.0005	0.00051	0.83 (0.75-0.92)
WNT3	rs2083798	42276896	C/T	0.289	0.253	0.0001	0.010	1.20 (1.09-1.32)
WNT9b	rs2083797	42276928	G/A	0.288	0.253	0.0002	0.010	1.20 (1.09-1.32)
WNT9b	rs12602434	42283052	C/G	0.157	0.139	0.0149	0.050	1.16 (1.03-1.30)
WNT9b	rs4968279	42285344	C/T	0.158	0.139	0.0129	0.056	1.16 (1.03-1.31)
TCF2	rs11263756	33123933	A/G	0.027	0.036	0.0187	0.054	0.75 (0.59-0.96)
TCF2	rs2189303	33134218	A/G	0.251	0.233	0.0442	n/t	1.11 (1.00-1.22)
LRP6	rs12313200	12175857	A/G	0.155	0.174	0.0164	0.21	0.87 (0.78-0.98)
LRP6	rs12422312	12239176	A/G	0.146	0.163	0.0269	n/t	0.88 (0.78-0.99)
LRP6	rs12425946	12302231	T/C	0.145	0.163	0.0174	n/t	0.87 (0.77-0.98)
WNT6	rs2059717	219425821	T/A	0.474	0.495	0.0465	n/t	0.92 (0.84-1.00)
WNT6	rs608047	219441415	A/C	0.474	0.495	0.0463	n/t	0.92 (0.84-1.00)
LRP5	rs312786	67876553	A/C	0.293	0.314	0.0333	n/t	0.91 (0.83-0.99)
LRP5	rs576118	67934284	C/T	0.241	0.260	0.0404	n/t	0.90 (0.82-1.00)
LRP5	rs3736228	67957871	A/G	0.133	0.151	0.0170	n/t	0.86 (0.76-0.97)
LRP5	rs3867143	67979650	G/A	0.138	0.156	0.0158	n/t	0.86 (0.77-0.97)
FZD6	rs827536	104392202	G/T	0.524	0.558	0.0020	0.043	0.87 (0.80-0.95)
FZD6	rs703810	104403014	C/T	0.120	0.106	0.0454	n/t	1.14 (1.00-1.30)
FZD6	rs3808554	104406272	C/T	0.452	0.480	0.0106	0.054	0.90 (0.82-0.97)
sFRP1	rs4736959	41244318	G/A	0.432	0.408	0.0227	0.044	1.10 (1.01-1.20)
sFRP1	rs7843510	41254926	G/A	0.418	0.393	0.0170	0.042	1.11 (1.02-1.21)
sFRP1	rs7834090	41261852	T/C	0.043	0.032	0.0058	0.022	1.36 (1.09-1.69)
sFRP1	rs7833518	41262183	T/G	0.414	0.388	0.0124	0.035	1.12 (1.02-1.22)
TCF4	rs1942265	51066565	G/C	0.304	0.282	0.0297	n/t	1.11 (1.01-1.21)
TCF4	rs4800988	51070904	A/T	0.303	0.283	0.0399	n/t	1.10 (1.00-1.21)
TCF4	rs1539951	51072671	C/T	0.304	0.283	0.0318	n/t	1.11 (1.01-1.21)
TCF4	rs1788027	51093825	T/C	0.470	0.448	0.0364	n/t	1.09 (1.01-1.19)
TCF4	rs1660241	51186648	A/G	0.310	0.282	0.0042	0.021	1.14 (1.04-1.25)
TCF4	rs1660242	51189674	T/A	0.437	0.414	0.0301	n/t	1.10 (1.01-1.20)
TCF4	rs9320010	51204895	G/A	0.423	0.395	0.0156	n/t	1.11 (1.02-1.21)

TCF4	rs7235757	51218952	T/C	0.345	0.320	0.0138	0.028	1.12 (1.02-1.22)
TCF4	rs2919451	51235298	T/C	0.420	0.395	0.0182	n/t	1.11 (1.02-1.21)
TCF4	rs2958163	51236410	C/T	0.481	0.457	0.0234	n/t	1.10 (1.01-1.20)

Table 8-4 WTCCC data analysis on WNT pathway genes.
WTCCC SNPs from WNT pathway genes reaching significance ($p<0.05$) on analysis of CD vs. WTCCC HC. SNPs with MAF < 1.0% excluded from analysis. The meta-analysis p-values is presented for those SNPs present in the meta-analysis of WTCCC, French / Belgium and N American GWAS studies (data provided by Jeffrey Barrett). Those SNPs not tested in the meta-analysis are indicated by 'n/t'.

rs number	position	alleles	CD MAF	HC MAF	CD vs. HC p-value	OR
rs199535	42177829	G/A	0.193	0.224	0.0004	0.83 (1.09-1.34
rs13341140	42181223	G/C	0.188	0.222	7.38×10^{-5}	0.81 (0.73-0.90)
rs199530	42191820	C/T	0.220	0.254	0.0002	0.83 (0.75-0.92)
rs199515	42211804	C/G	0.193	0.223	0.0005	0.83 (0.75-0.92)
rs199496	42221887	A/G	0.084	0.092	0.2137	0.91 (0.79-1.06)
rs12452064	42223353	A/G	0.439	0.433	0.5945	1.02 (0.94-1.11)
rs199495	42223596	A/G	0.422	0.424	0.8507	0.99 (0.91-1.08)
rs199494	42224229	G/A	0.428	0.432	0.7261	0.98 (0.90-1.07)
rs7207916	42234514	T/C	0.418	0.420	0.8258	0.99 (0.91-1.08)
rs2083798	42276896	C/T	0.289	0.253	0.0001	1.20 (1.09-1.32)
rs2083797	42276928	G/A	0.288	0.253	0.0002	1.20 (1.09-1.32)
rs12602434	42283052	C/G	0.157	0.139	0.0149	1.16 (1.03-1.30)
rs4968279	42285344	C/T	0.158	0.139	0.0129	1.16 (1.03-1.31)
rs8082211	42296034	T/C	0.001	0.001	0.2776	2.24 (0.50-10.0)
rs12952746	42298412	A/G	0.112	0.124	0.1014	0.90 (0.79-1.02)
rs1530364	42306776	A/G	0.295	0.295	0.9603	1.00 (0.91-1.09)
rs17603901	42313021	A/G	0.101	0.107	0.3407	0.94 (0.82-1.07)

Table 8-5 Data analysis of WTCCC SNPs in region of WNT3 and WNT 9b on chromosome 17.

GLI3: 276kb view

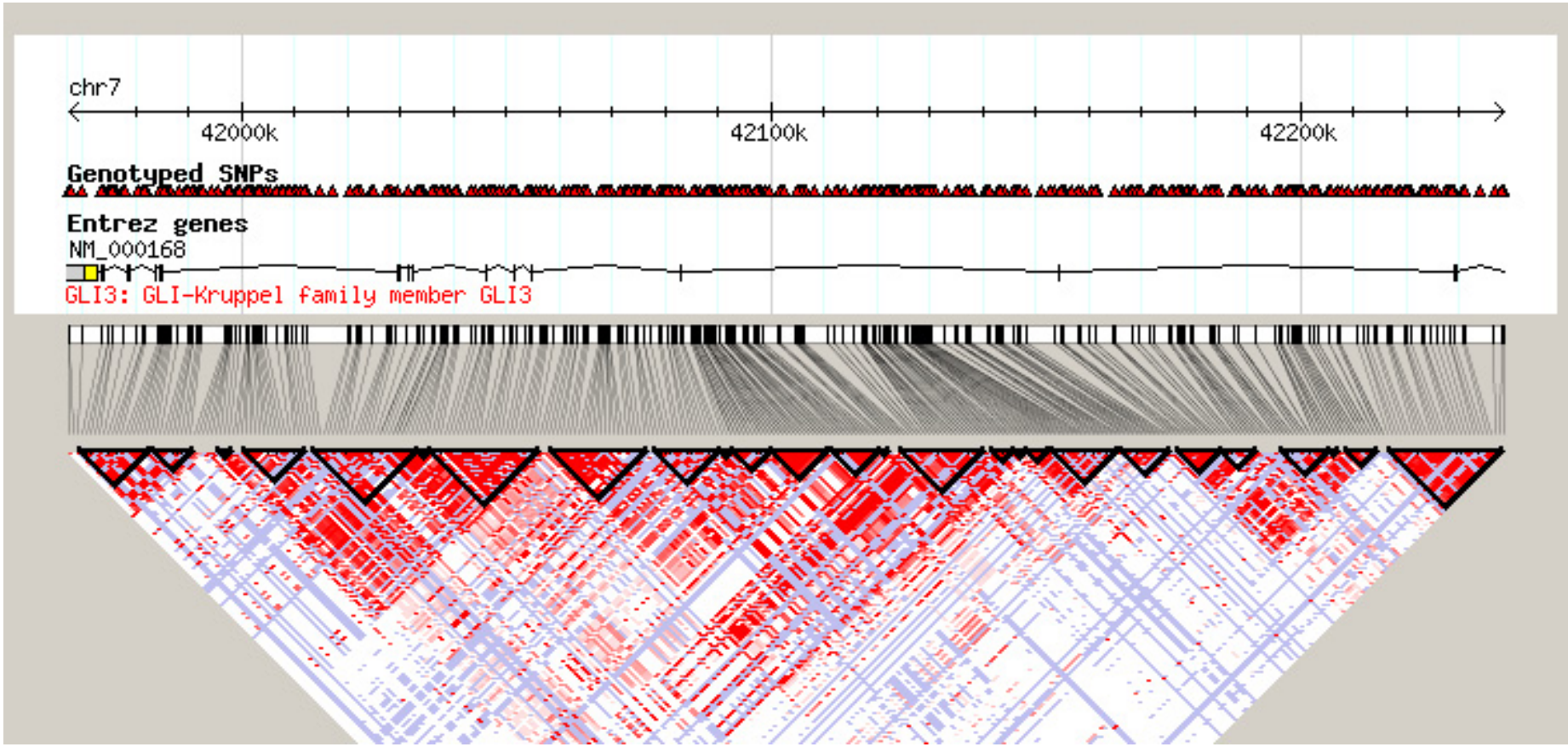

Figure 8-1 Haplotype structure of *GLI3*; 276kb view.
This, and all subsequent figures in the chapter, are generated in Haploview vers. 4.2, and derived from phase II CEPH (U.S. residents in 1980 of N and W European ancestry). HapMap data. LD values are calculated by D (red squares indicate tight LD).

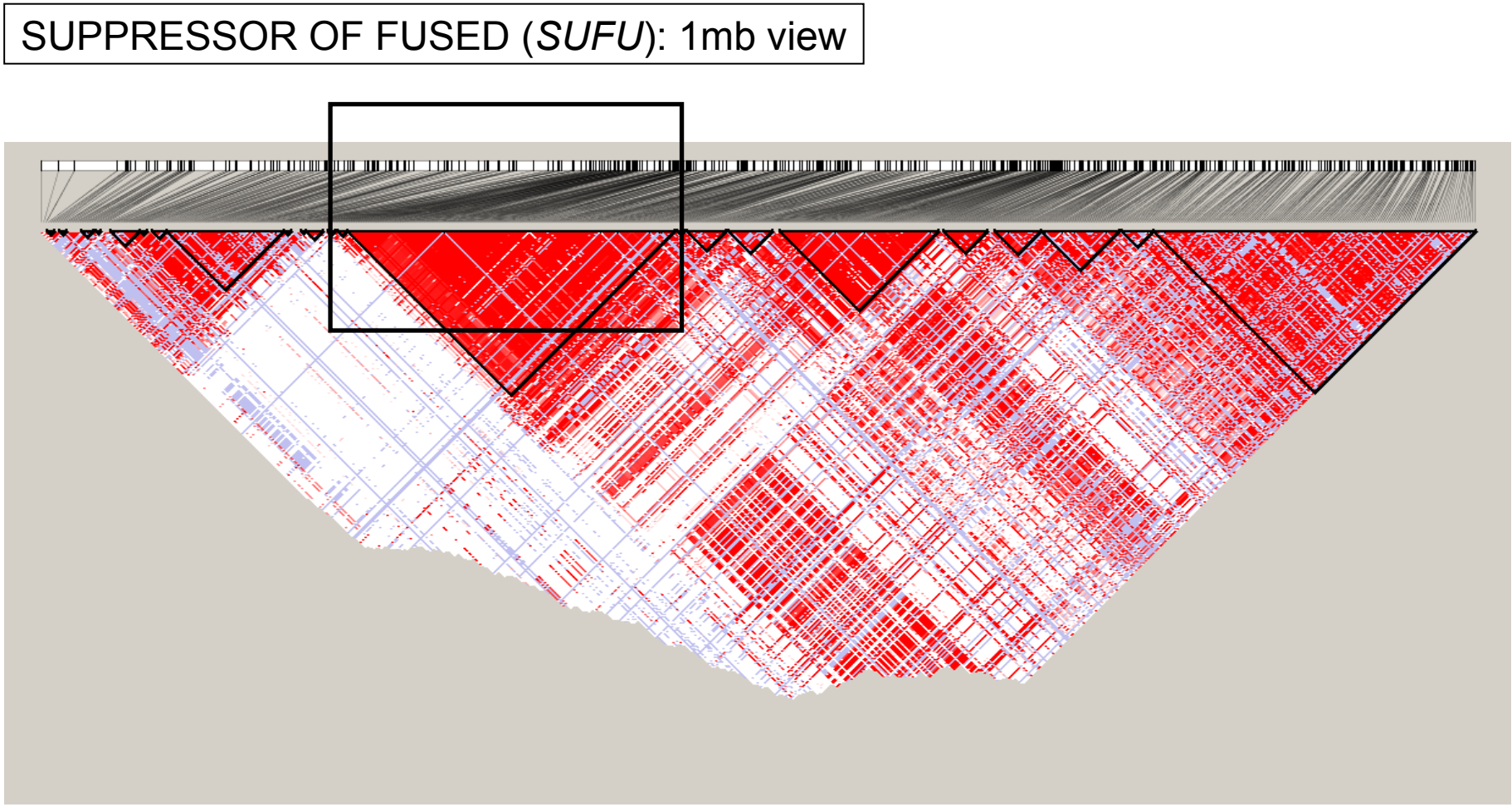

Figure 8-2 Haplotype structure of *SUFU*; 1mb view.
This figure demonstrates the extensive LD across *SUFU*: the 200kb area in the box is viewed in greater detail in Figure 8-3.

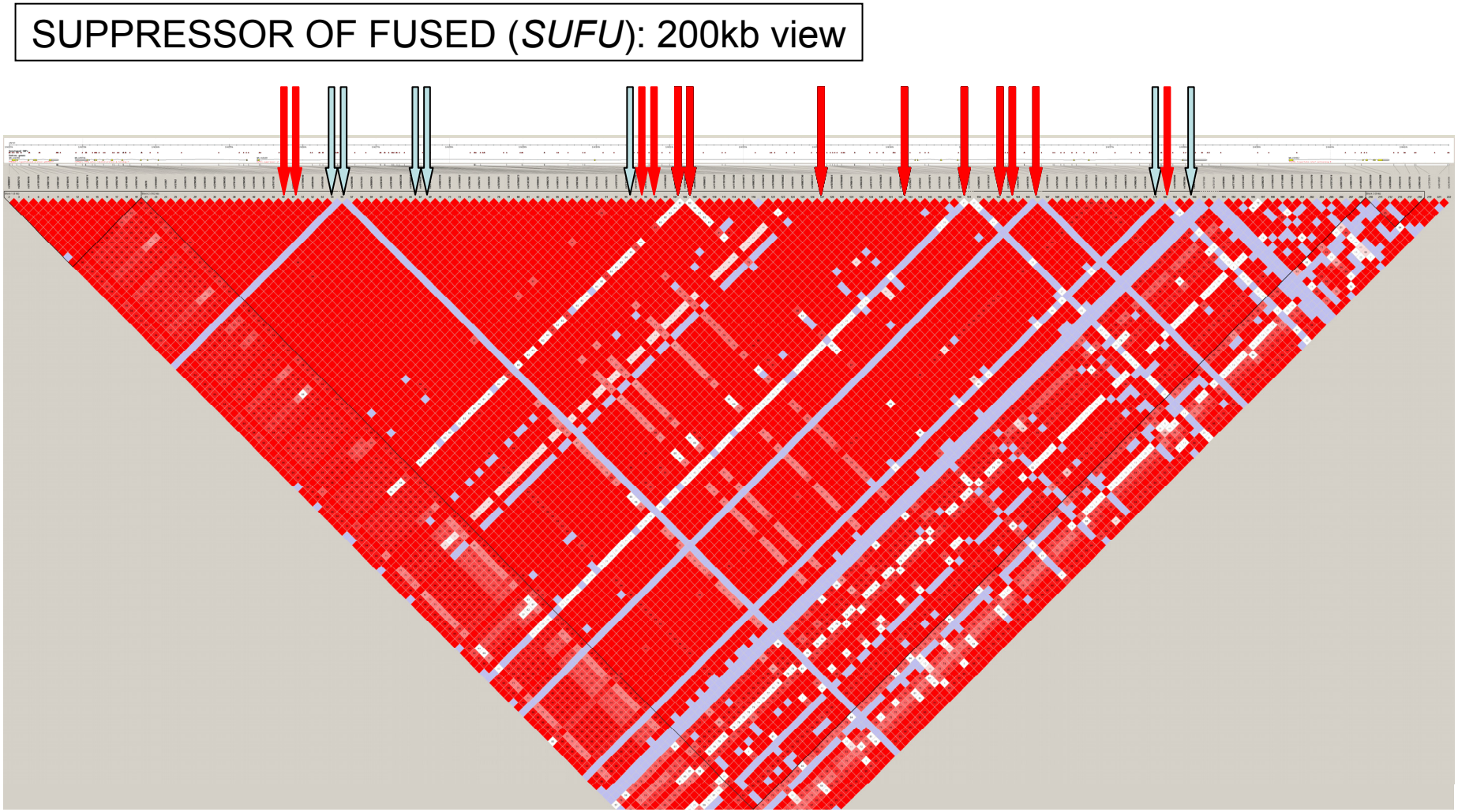

Figure 8-3 Haplotype structure of *SUFU*; 200kb view.
Expanded view of box in Figure 8-2. The position of WTCCC SNPs is demonstrated by arrows. The red arrows are those that reached significance ($p<0.05$) on analysis 1.

GLI2: 200kb view

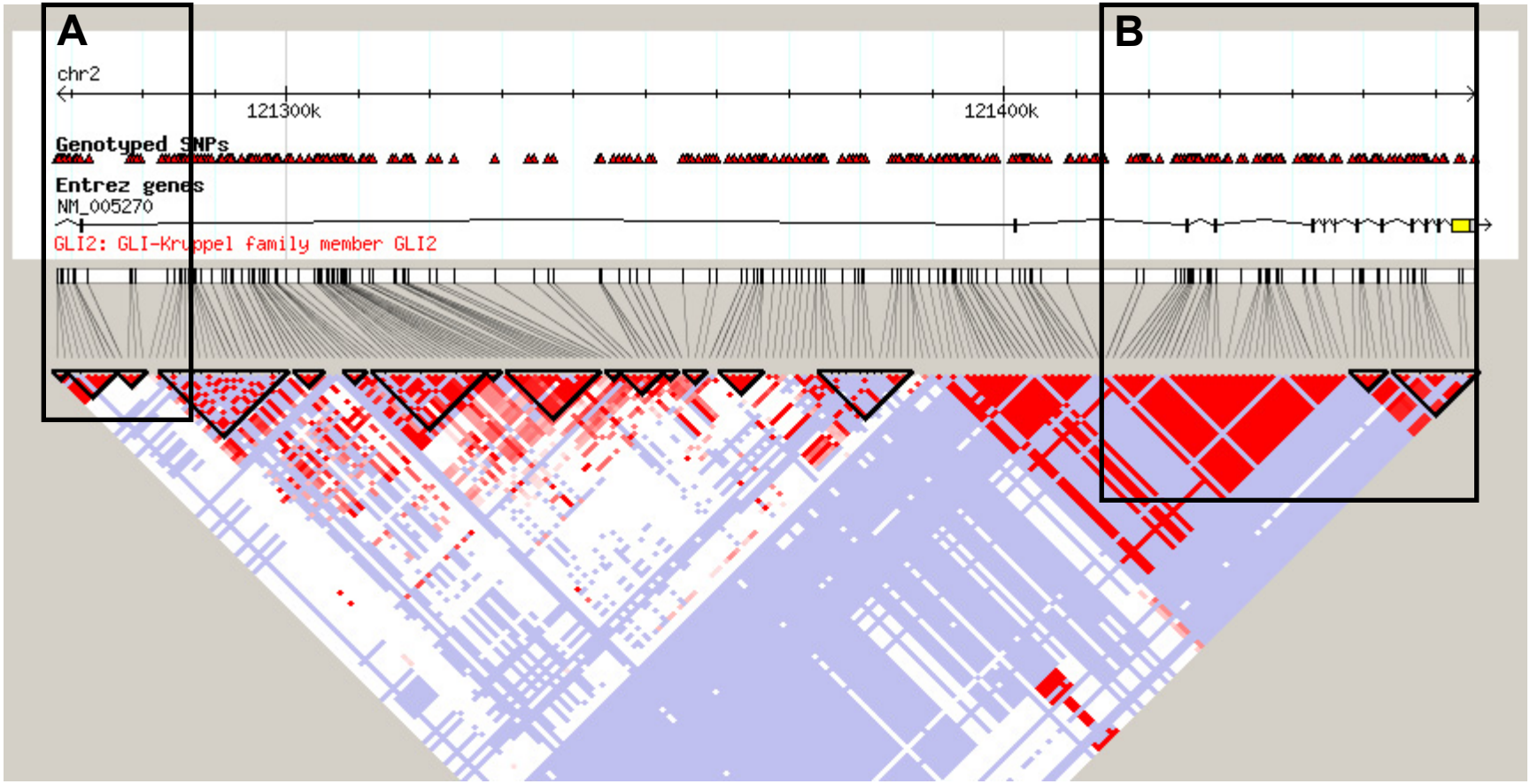

Figure 8-4 Haplotype structure of *GLI2*; 200kb view.
Magnified views of boxes A and B are depicted in Figures 8-5 and 8-6 respectively.

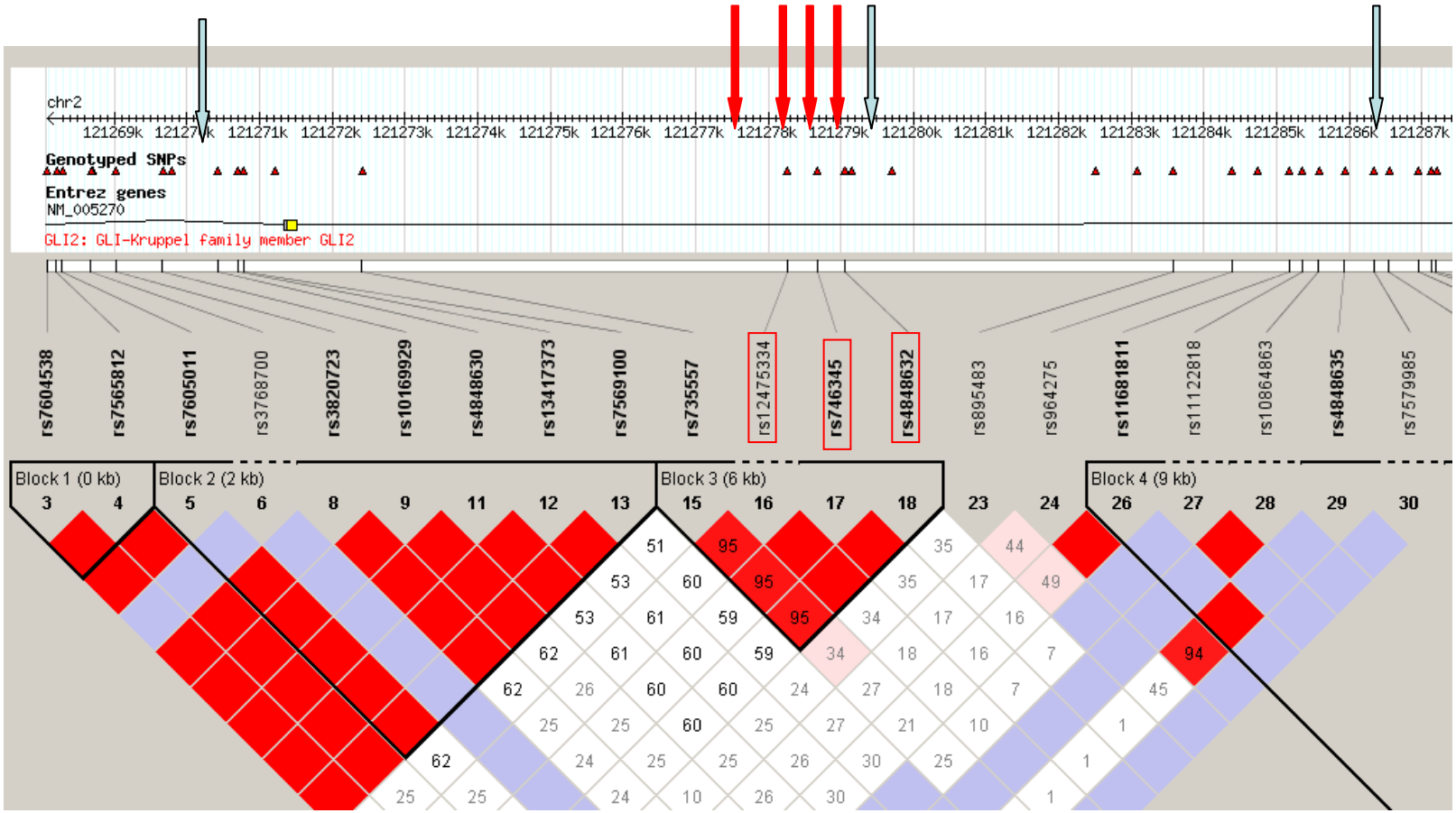

Figure 8-5 Haplotype structure of *GLI2* – fragment A from 8-4; 18kb view.
The position of WTCCC SNPs is demonstrated by arrows. The red arrows are those that reached significance ($p<0.05$) on analysis 1. Those HapMap SNPs that were typed in the WTCCC study are outlined in red.

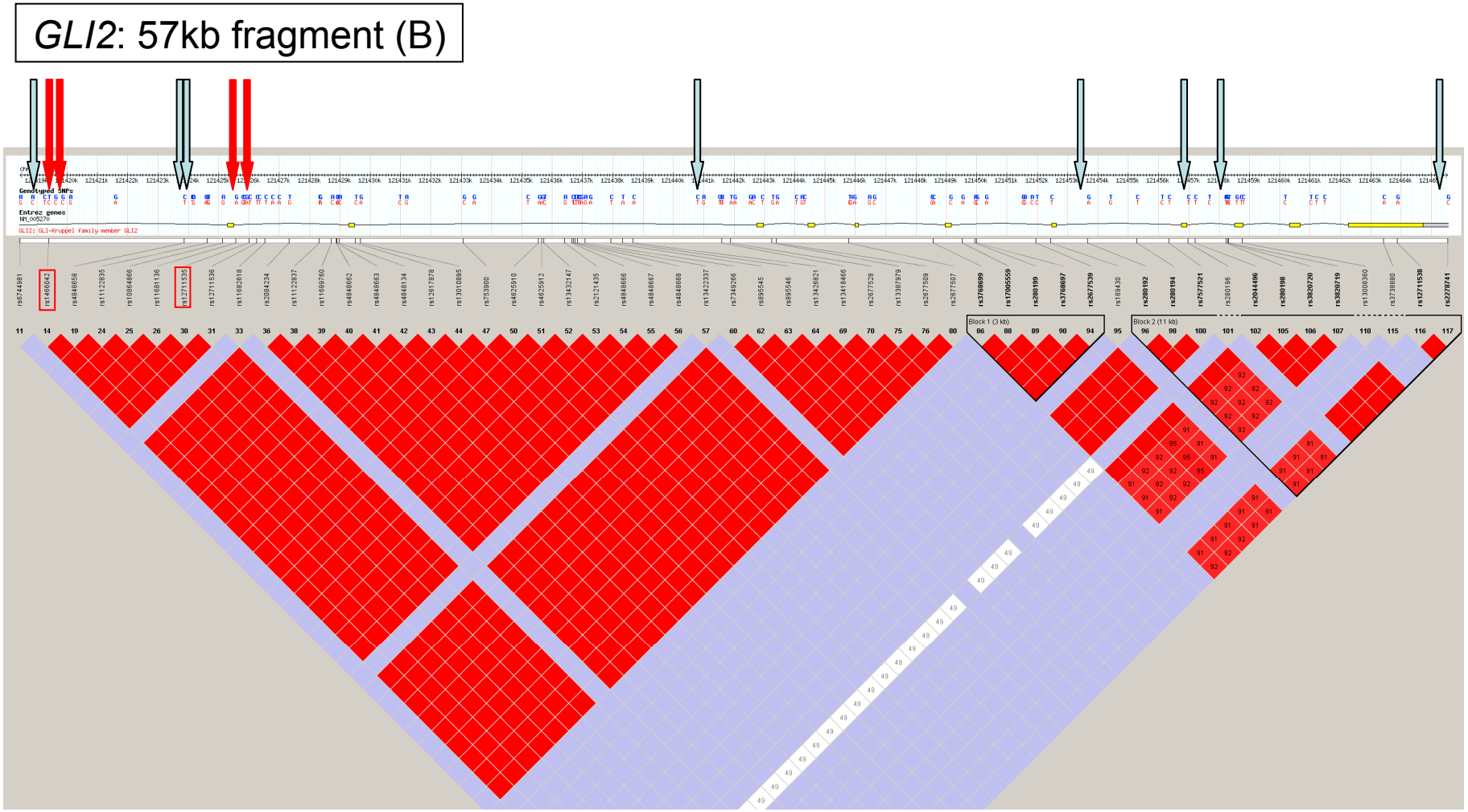

Figure 8-6 Haplotype structure of *GLI2* –fragment B from 8-4; 57kb view.
The position of WTCCC SNPs is demonstrated by arrows. The red arrows are those that reached significance ($p<0.05$) on analysis 1. Those HapMap SNPs that were typed in the WTCCC study are outlined in red.

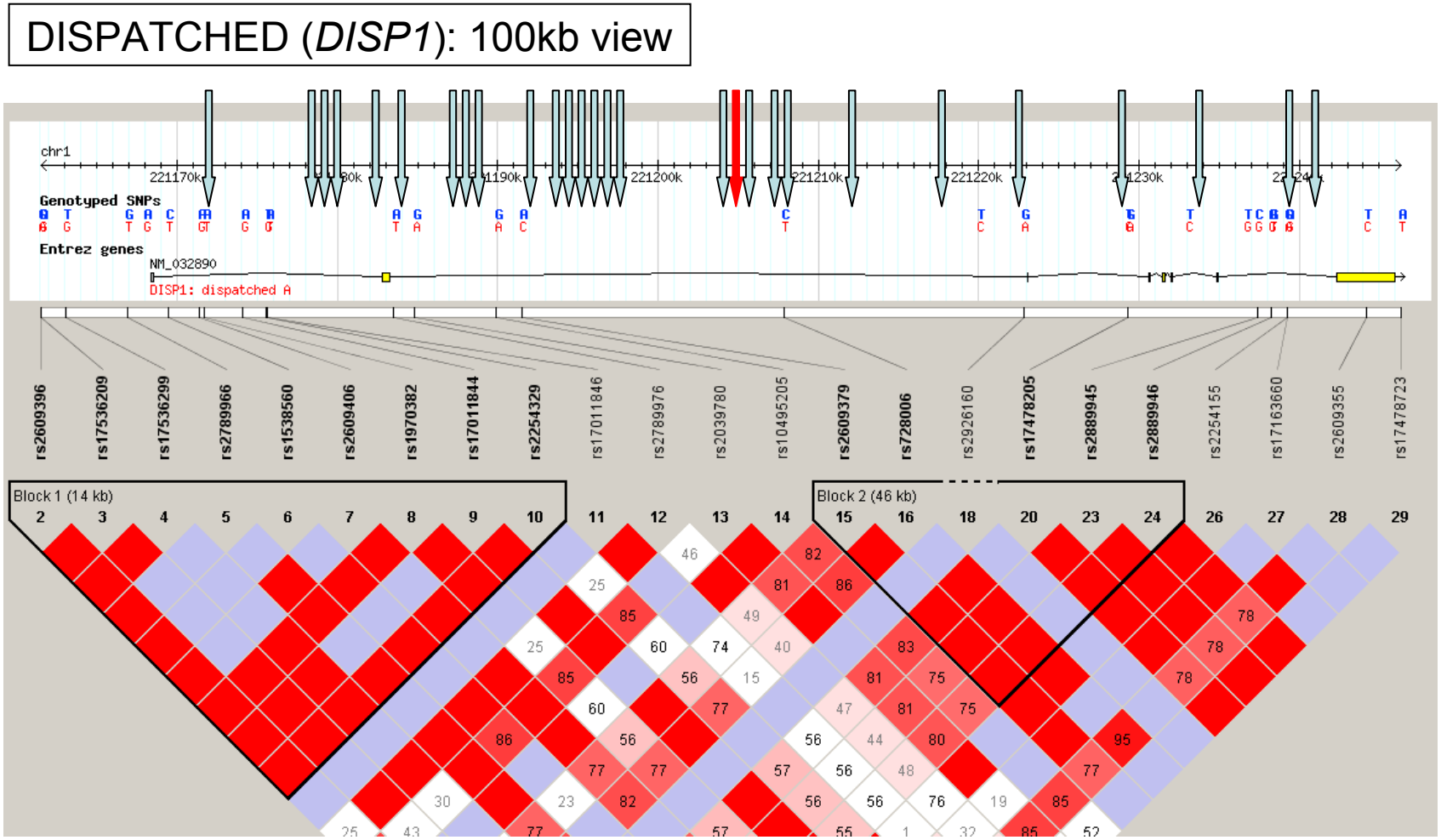

Figure 8-7 Haplotype structure of *DISP1*; 100kb view.
The position of WTCCC SNPs is demonstrated by arrows. The red arrows are those that reached significance ($p<0.05$) on analysis 1.

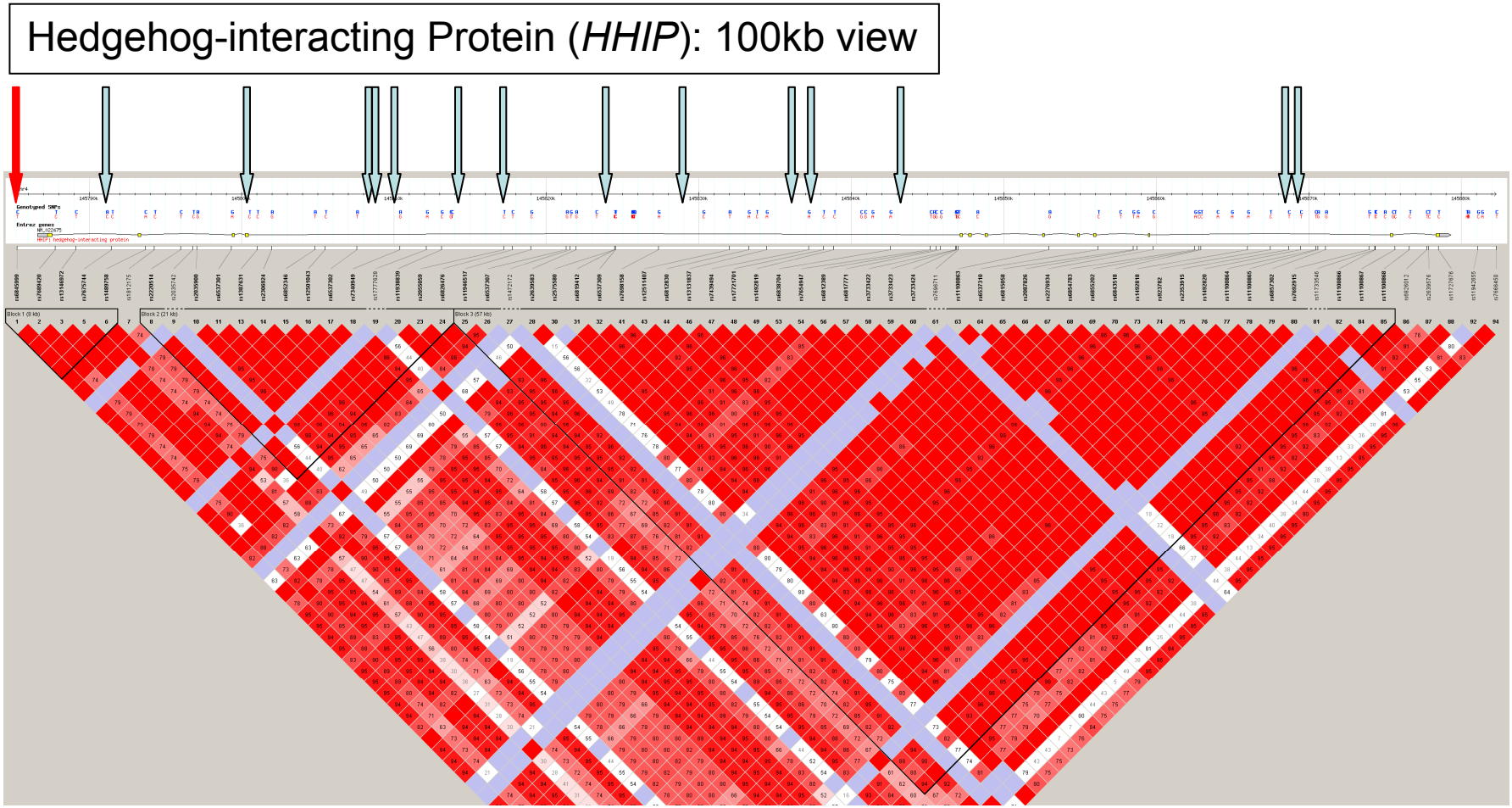

Figure 8-8 Haplotype structure of *HHIP*; 100kb view.
The position of WTCCC SNPs is demonstrated by arrows. The red arrows are those that reached significance ($p<0.05$) on analysis 1.

SONIC HEDGEHOG (*SHH*): 40kb view

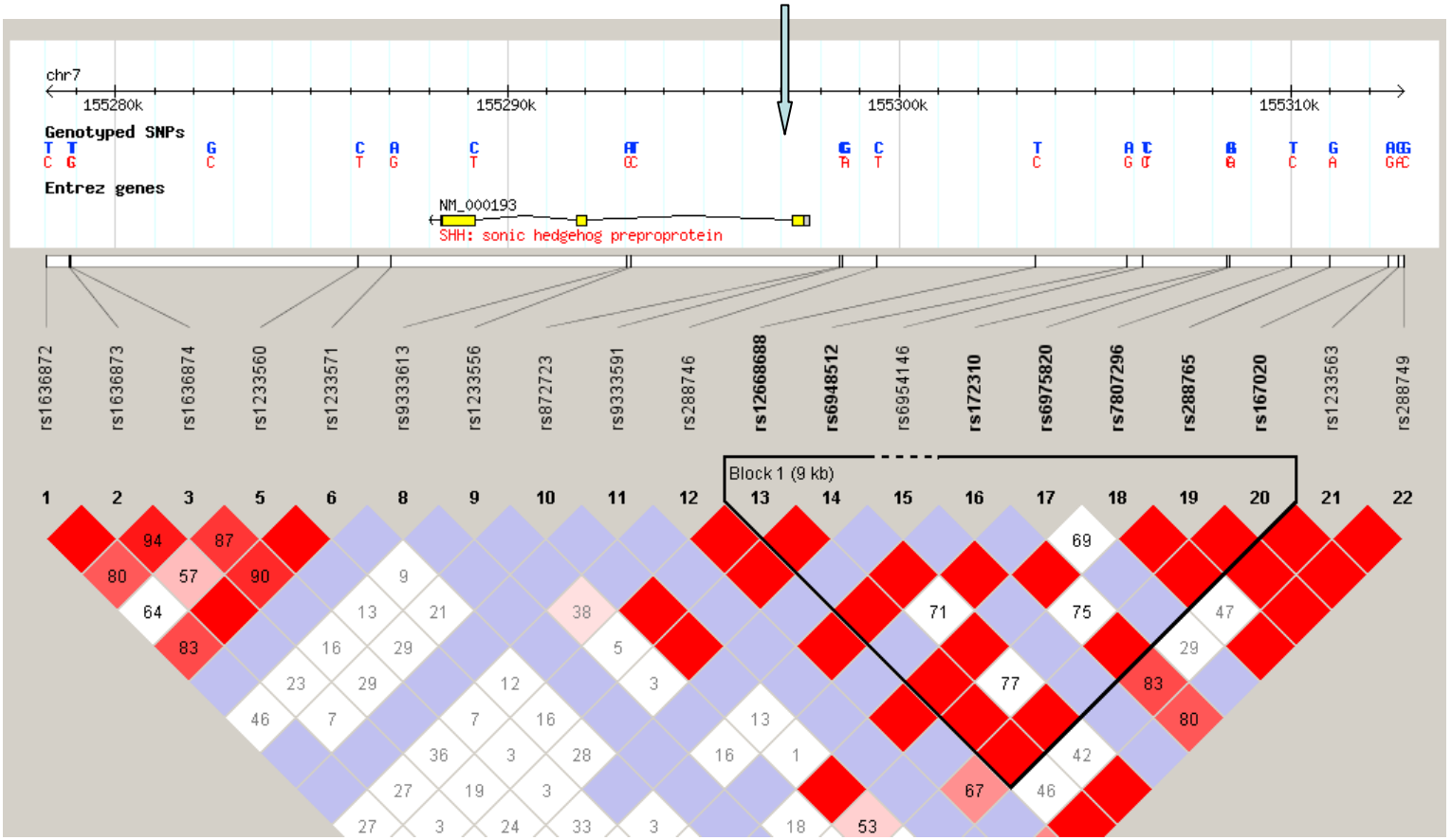

Figure 8-9 Haplotype structure of *SHH*; 40kb view.
The position of the only WTCCC SNP is demonstrated by blue arrow.

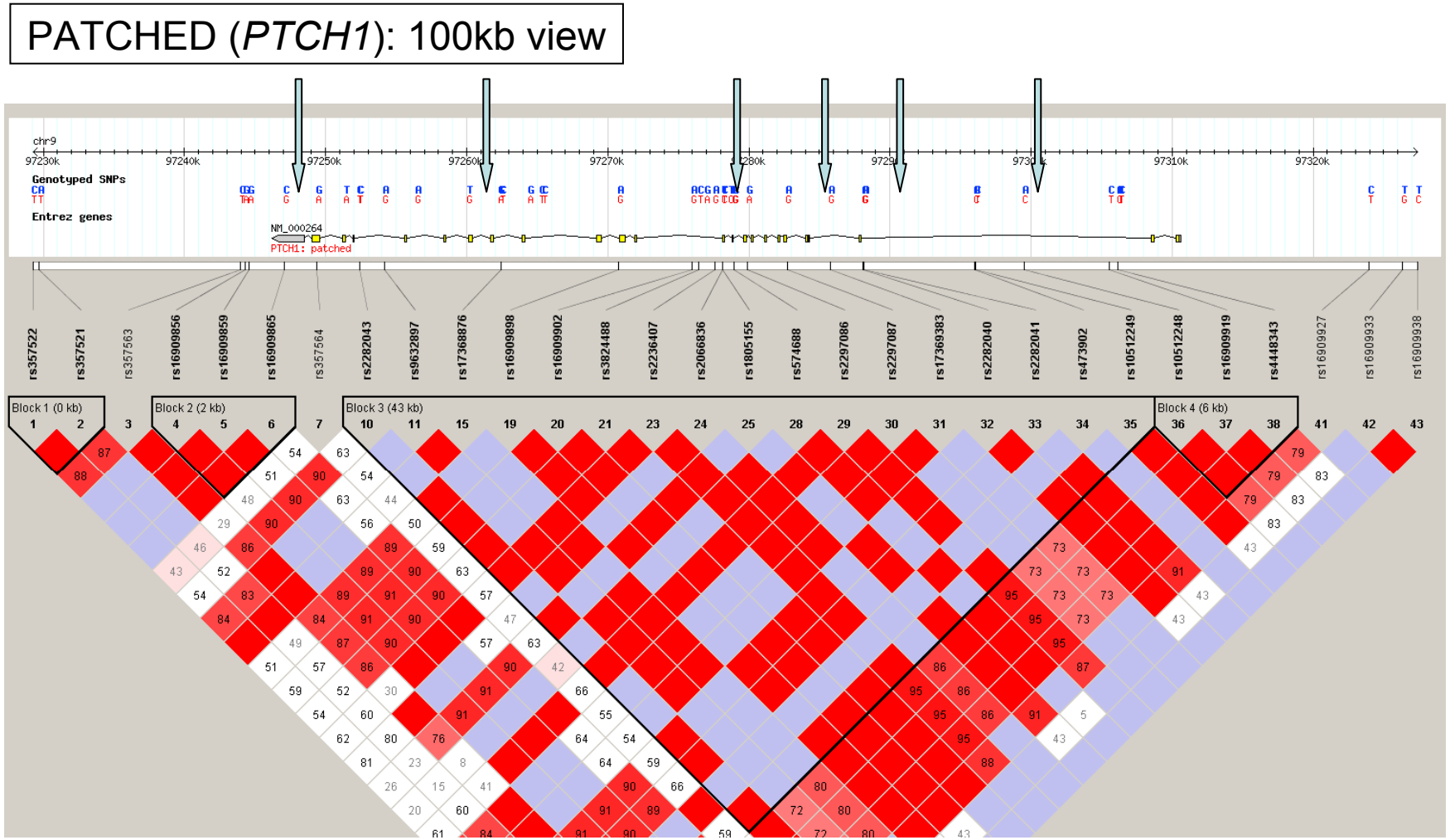

Figure 8-10 Haplotype structure of *PTCH*; 100kb view.
The position of WTCCC SNPs is demonstrated by blue arrows.

PATCHED-2 (*PTCH2*): 40kb view

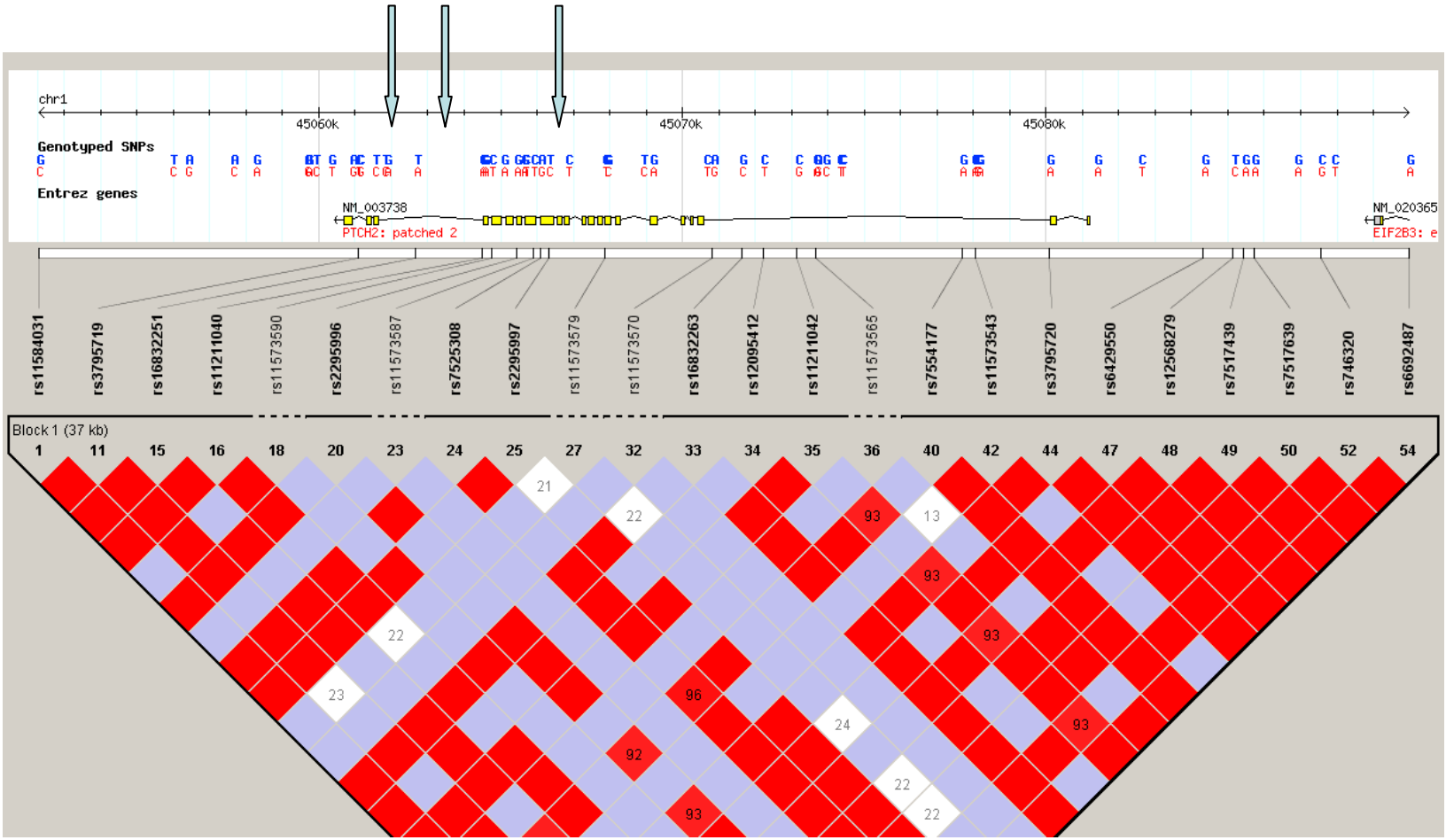

Figure 8-11 Haplotype structure of *PTCH2*; 40kb view.
The position of WTCCC SNPs is demonstrated by blue arrows.

SMOOTHENED (*SMO*): 40kb view

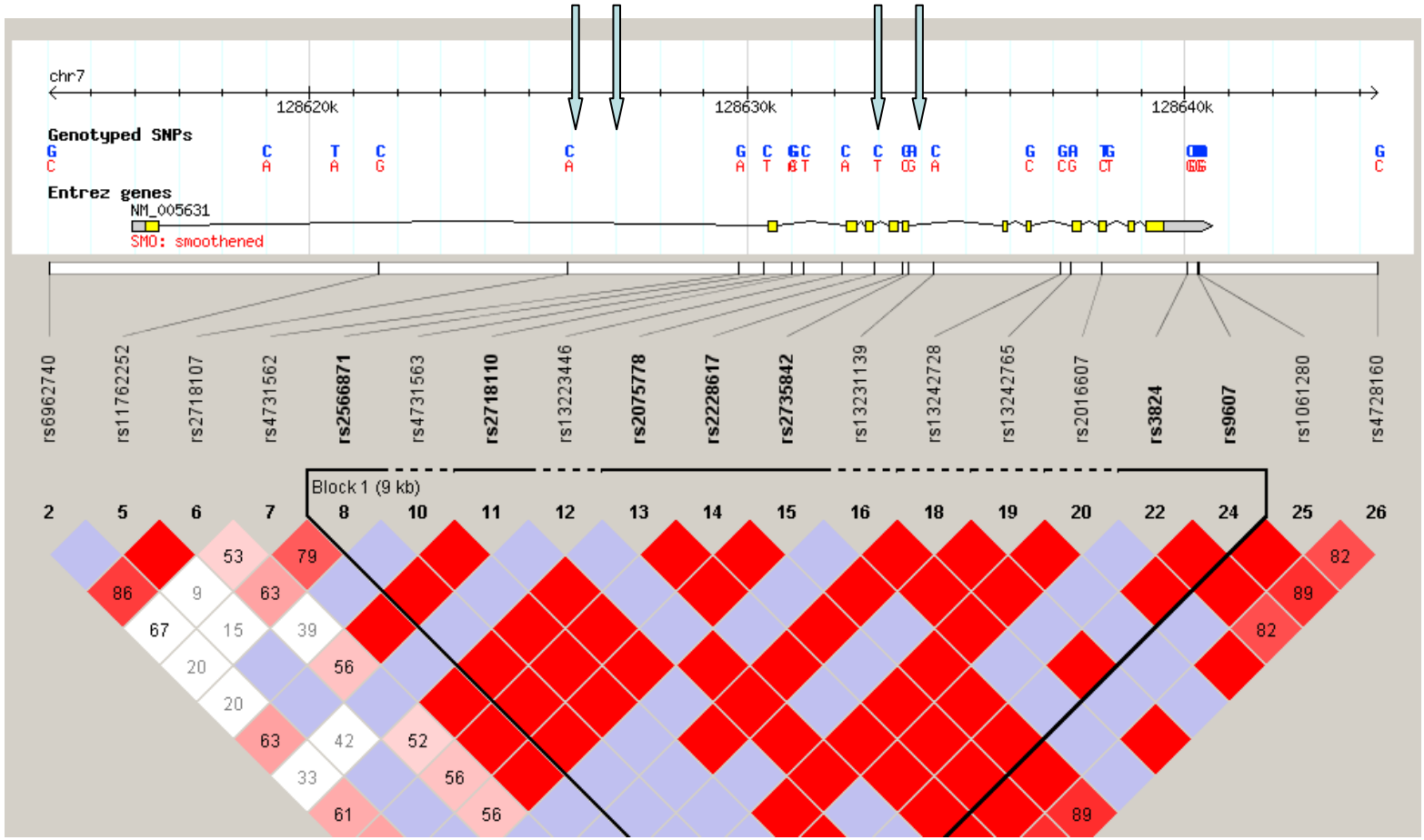

Figure 8-12 Haplotype structure of *SMO*; 40kb view.
The position of WTCCC SNPs is demonstrated by blue arrows.

GLI1: 16kb view

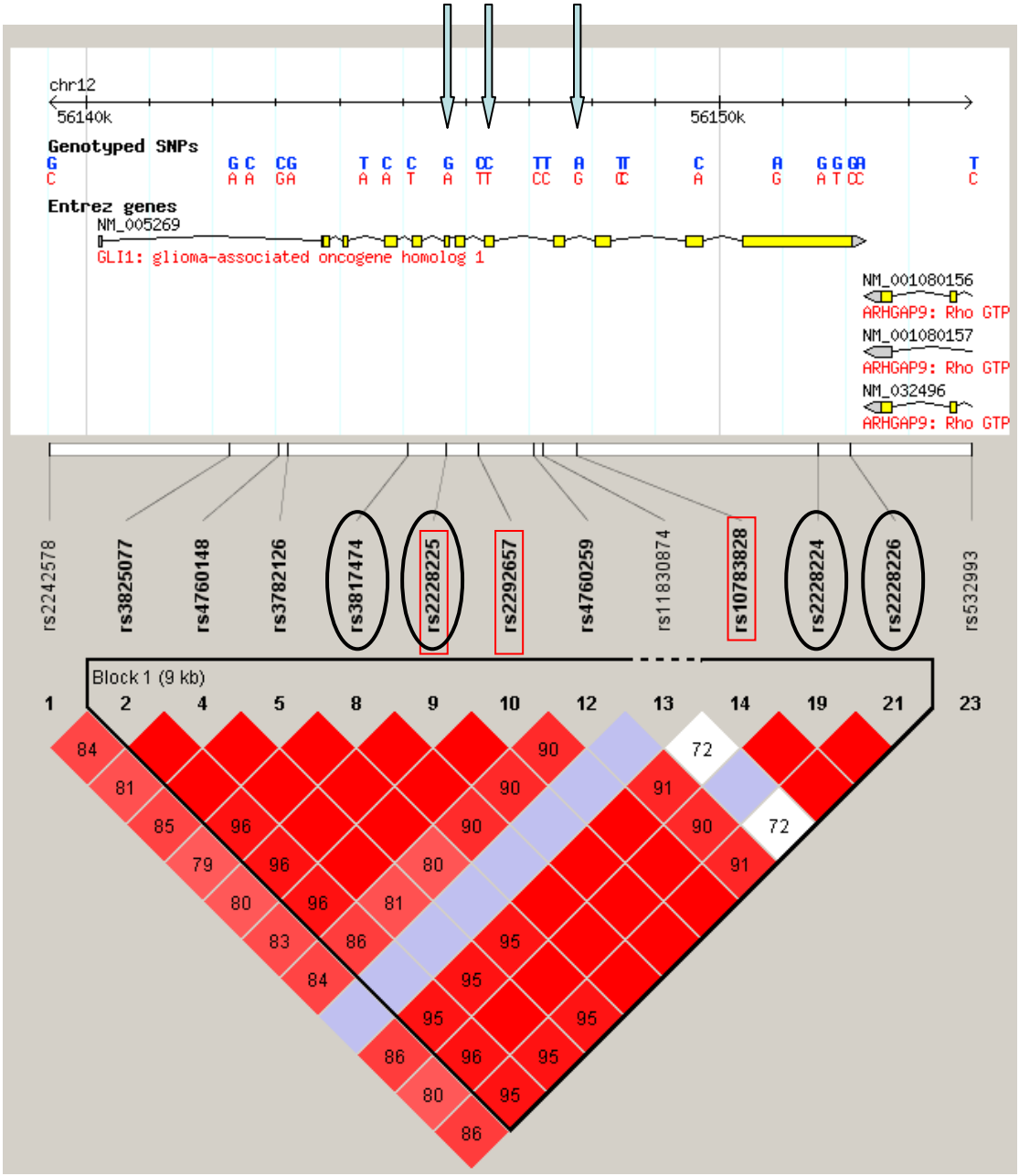

Figure 8-13 Haplotype structure of *GLI1*; 16kb view.
The position of WTCCC SNPs is demonstrated by arrows. Those HapMap SNPs that were typed in the WTCCC study are outlined in red. The four tSNPs genotyped in the gene-wide haplotype-tagging study in IBD (Chapter 5) and colo-rectal cancer (Chapter 6) are outlined by black ovals.

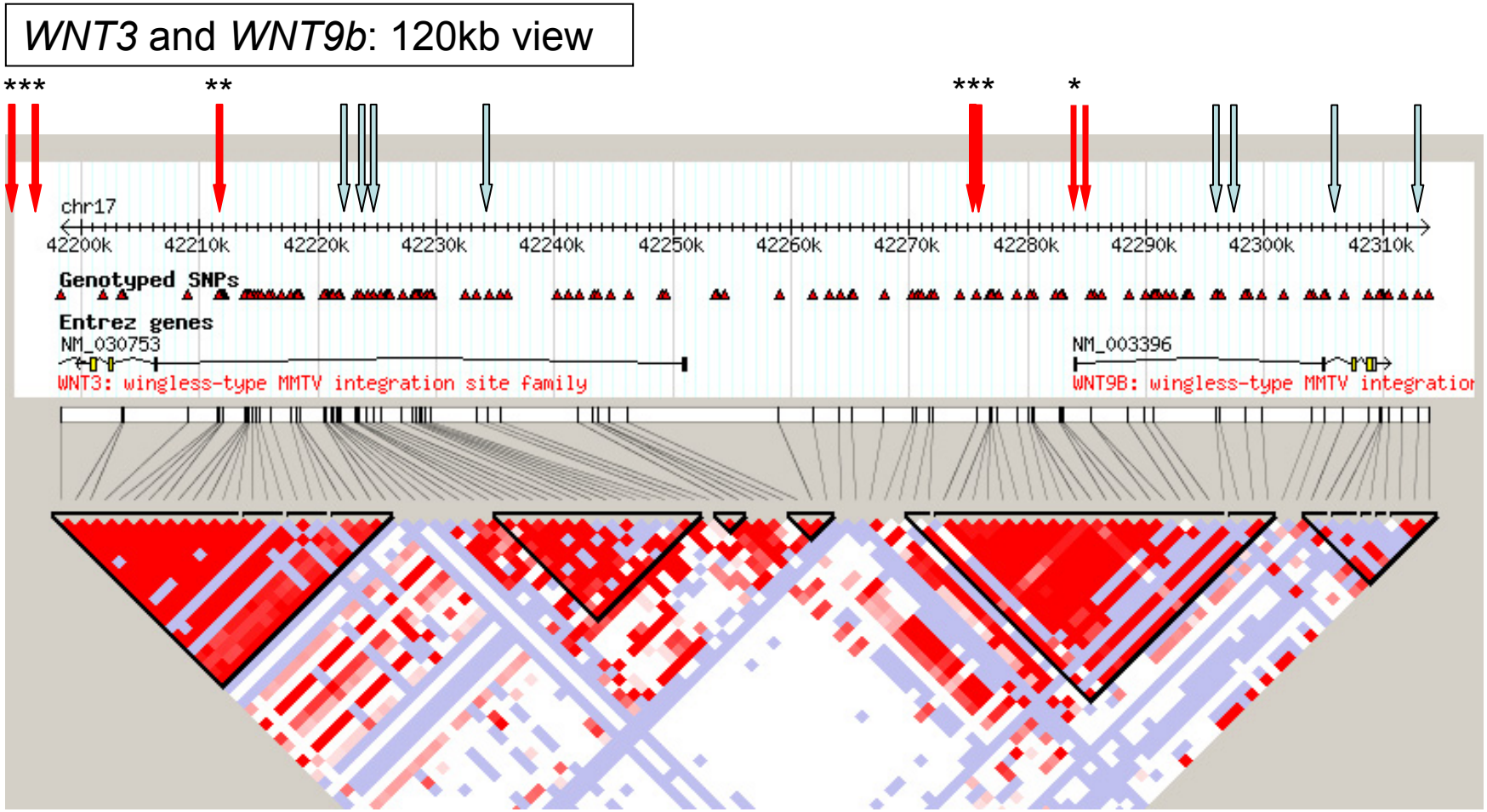

Figure 8-14 Haplotype structure of locus spanning *WNT3* and *WNT9b*; 120kb view.
The position of WTCCC SNPs is demonstrated by arrows. The red arrows are those that reached significance (*** $p<0.00019$, ** $p<0.01$, * $p<0.05$).

WNT3 and *WNT9b*: 700kb view

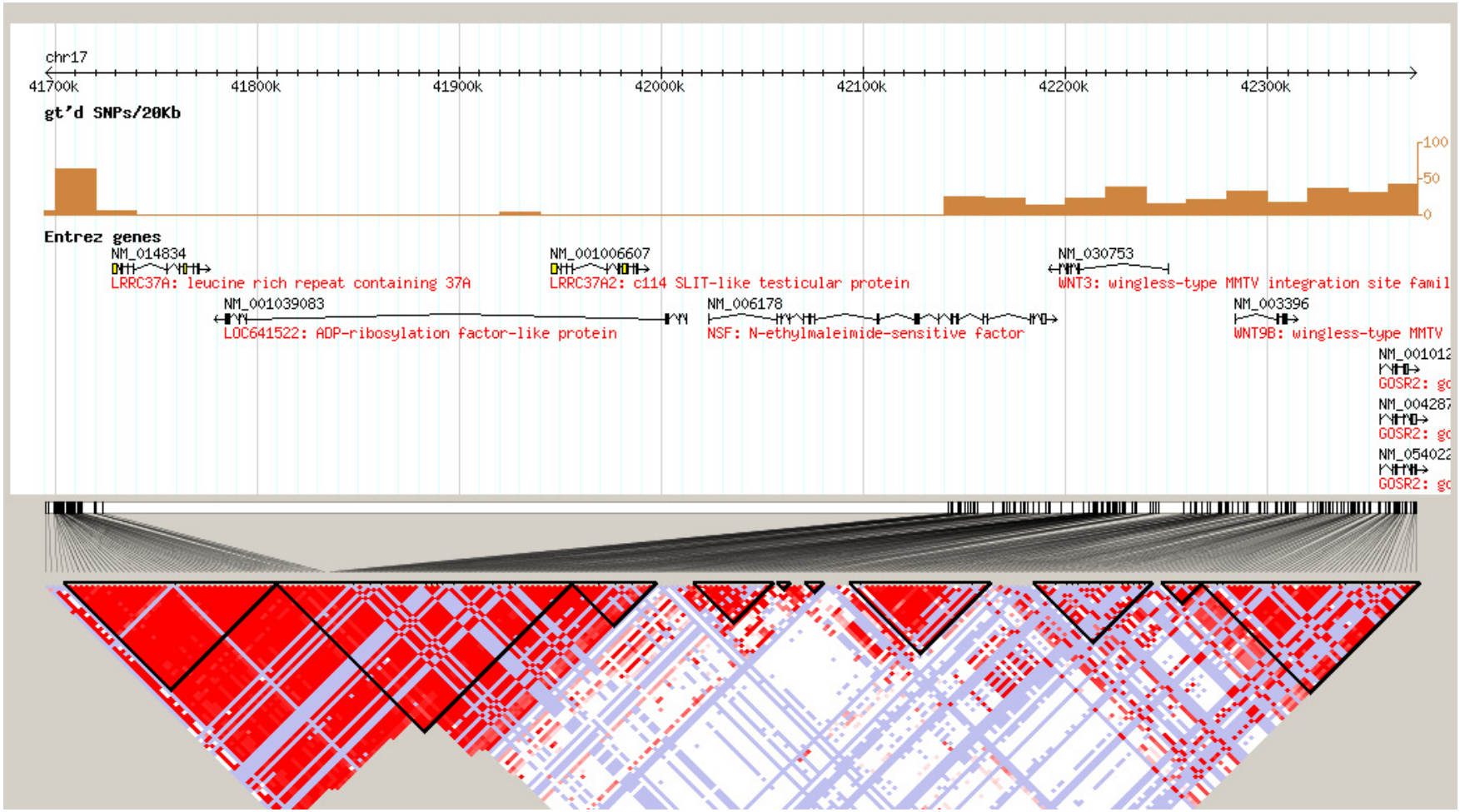

Figure 8-15 Haplotype structure of locus spanning *WNT3* and *WNT9b*; 700kb view.

9 *The effect of PAMPs on HH pathway expression, and of HH agonists / angtagonists on NFκB activity and cytokine profiles in vitro*.

9.1 *Abstract*

Introduction. As shown in the expression studies in Chapter 4, HH pathway activity increases along the length of the healthy adult colon, mirroring the increased bacterial burden in the colonic lumen. It was therefore hypothesised that HH pathway expression might be increased by pathogen-associated molecular patterns (PAMPs) in colonic epithelial cells *in vitro*. Furthermore, NFκB activity has been shown to modulate SHH expression in keratinocytes and pancreatic cells *in vitro*. In addition, HH pathway agonists and antagonists are known to modulate cytokine synthesis and secretion in peripheral CD4 T cells. The aims of the present study were therefore to assess the affect of PAMPs on HH pathway expression, and of HH agonists and antagonists on NFκB activity and cytokine synthesis and secretion *in vitro*.

Methods. Colonic epithelial adenocarcinoma cells (SW480) were stimulated *in vitro* with varying concentrations of LPS and MDP at different timepoints (up to 24 hours). The expression of SHH and PTCH mRNA was measured by Q-PCR (18s RNA was used as an internal control) and of nuclear GLI1 and p65 protein expression by immunohistochemistry. NFκB luciferase reporter assays were performed in SW480 cells treated with cyclopamine, rSHH and LPS for 24 hours. SW480 cells and PBMCs from patients with acute active UC were treated with rSHH or the HH agonist purmporphomine for 24 hours. Cytokine profiles (CCL20, IL8, IL1ß, IL10 and TNFα) were measured in culture supernatant by ELISA.

Results. SW480 cells express SHH, PTCH and GLI1 mRNA, NOD2 protein (cytoplasmic) and p65 protein (cytoplasmic and nuclear). 2.0μg/ml of LPS induced a modest increase in SHH at 2 hours ($p<0.05$). 2.0μg/ml of MDP resulted in a significant decrease in PTCH mRNA after 1 hour, peaking at 4 hours ($p<0.05$). The increase in nuclear p65 staining induced by 2.0μg/ml LPS was abrogated with the HH antagonist cyclopamine. However, neither rSHH nor cyclopamine significantly altered NFκB activity at 24 hours. rSHH, but not purmorphamine, led to a robust increase in CCL20 in PBMCs from patients with active UC. However, purmorphomine, but not rSHH, caused decreased CCL20 in SW480 cells.

Conclusions. Whilst some interesting effects were observed with respect PAMP stimulation and purmorphomine treatement of SW480 cells, no effect was seen on NFκB activity. Recent advances highlighting the paracrine nature of HH signalling in the small intestine and colon have subsequently questioned the validity of SW480

cells as a model for studies of HH signalling and inflammation. However, the studies of PBMCs with rSHH in patients with active UC have demonstrated a clear response in CCL20, suggesting exciting avenues for further exploration.

9.2 *Introduction*

As discussed in Chapter 1, different groups have reported variable expression of HH signalling components within different colonic cell lines. Between 2003 and 2006 (when the present *in vitro* studies were conducted), there was hot debate within the published literature (*in vivo* and *in vitro* expression data) as to whether autocrine HH signalling occurred in colonic epithelial cells, particularly in colonic adenocarcinoma cell lines. This has only very recently been conclusively disproven by the intricate and detailed series of experiments published by Yauch and colleagues in *Nature* in 2008 (see **1.9.1.2**),[361] although some HH signalling experts still question these findings.[362]

However, the historic studies were consistent at the time with autocrine HH signalling in at least some colonic adenocarcinoma cell lines.[330, 346, 359] SW480 cells from our lab, shown to express SHH, PTCH and GLI1, were therefore used for a series of *in vitro* experiments.

The gradients of HH pathway activity in the healthy adult human colon demonstrated in Chapter 4 (**Figure 4-3**) were striking for two reasons. Not only does this mirror the clincial distrubtion of UC, but it also potentially reflects the increasing bacterial burden in the distal, compared with the proximal colon. In addition, the linear regression analysis of GLI1 and potential HH target genes in Chapter 4 demonstrated close correlation with the peptidoglycan recognition proteins (PGLYRP1-4). Firstly, therefore, the effect of PAMP (LPS and MDP) stimulation of SW480 cells on SHH, PTCH and GLI1 expression was assessed. Secondly, given reports linking HH and NFκB signalling,[373] the effect of HH agonists / antagonists on NFκB activity was assessed.

Thirdly, given the historic reports linking HH signalling and inflammation, the effect of HH agonists on cytokine synthesis and secretion in SW480 cells was tested. In addition, due to the published studies by the Howie group on HH pathway activity within peripheral CD4 T cells,[365, 366] PBMCs from patients with active UC were used for these latter studies.

9.3 *Methods*

9.3.1 SW480 cell stimulation

LPS (Escherichia coli 026:B6, lyophilised, sterilised by γ-irradiation, prepared by TCA precipitation and gel filtration chromatography; Sigma) and MDP (N-acetylmuramyl-L-alanyl-D-isoglutamine hydrate; ≥98%; Sigma) were reconstituted in 1 ml of sterile Dulbecco's modified eagle's medium nutrient mixture (15mM HEPES, $NaHCO_3$, pyridoxine; Sigma.) Cells were treated in $25cm^2$ flasks at 70-80% confluence over 24 hours, except where indicated otherwise. For timecourse experiments, all flasks were removed from the incubator at each treatment point (-24h, -8h, -4h, -2h, -1h, -0.5h, -0.25h) and all harvested simultaneously at timepoint 0 (24 hours).

9.3.2 Patients for PBMC analysis

5 patients (3 female) with acute severe UC were recruited within 24 hours of admission to hospital for iv corticosteroid therapy. All were white Caucasian (4 Scottish, 1 Polish). For 3/5 this was the first presentation of UC. The other 2 patients had both been diagnosed 3 years previously. 2 patients were completely treatment naïve at the point of PBMC extraction. 1 had been treated with iv steroid for <12 hours. 2 patients were on long-term oral 5-ASA therapy and had taken 1 and 5 days of oral prednisolone. 3 patients were ex-smokers; 2 had never smoked. Informed, written consent was obtained from all patients.

9.3.3 Treatment of PBMCs

The methods for extraction of PBMCs over a Ficoll gradient are described in Chapter 3 (**3.9.2**).

9.3.4 RT-PCR, Q-PCR, immunohistochemistry, ELISAs and NFκB reporter assay

Full details on methods are given in chapter 3 (**3.9.4** – **3.9.11**).

9.3.5 Statistical analyses

The Kruskal Wallis one-way analysis of variance test followed by Dunn's Multiple Comparison post-test was used on the non-parametric Q-PCR data, NFκB reporter assays and cytokine analysis. Mann-Whitney U analysis was used to analyse proportions of cells with nuclear staining (GLI1 and p65). 2-tailed p-values are given throughout ($\alpha = 0.05$).

9.4 *Results*

9.4.1 Expression of HH signalling components in SW480 cells

RT-PCR for SHH, PTCH and GLI1 on mRNA from unstimulated SW480 cells identified expression of transcripts for all 3 components of HH signalling (**Figure 9-1**). In addition, protein expression of NOD2 (cytoplasmic) and p65 (predominantly cytoplasmic with occasional nuclear expression) was characterised by immunofluoresence in unstimulated cells (**Figure 9-2**).

9.4.2 HH expression in SW480 cells stimulated with LPS and MDP

SW480 cells were first stimulated for 24 hours with 2.0 μg/ml LPS (**Figure 9-3a**). As the greatest increase in SHH expression was noted at 2 hours ($p<0.05$), cells were then stimulated for 2 hours with a concentration gradient of LPS from 0.25 – 5.0 μg/ml (**Figure 9-3b**). No significant differences were noted in SHH or PTCH expression at LPS concentrations of 0.25 – 2.0 μg/ml. However, PTCH expression was significantly decreased at 5.0μg/ml ($p<0.05$) suggesting down-regulation of HH pathway activity with this high concentration of LPS at 2 hours. Recognising previous reports that commercial LPS is contaminated with MDP and that this effect might be magnified with higher concentrations, cells were stimulated with 2.0μg/ml of MDP over a 4 hour timecourse (**Figure 9-3c**). Significant down-regulation of PTCH expression was noted from 1 hour (Kruskall-Wallis $p=0.065$). This effect was sustained to a peak at 4 hours ($p<0.05$) at the termination of the timecourse.

To further assess the effect of PAMPs on HH pathway activity cells, GLI1 protein cellular location was assessed in cells stimulated for 2 hours with 2μg/ml of LPS (**Figure 9-4a-b**). There was no effect on the percentage of cells with nuclear GLI1 staining with LPS (25.2±3.2 vs. 23.0±1.7). In addition there was evidence of persistent background staining with omission of primary antibody, although this was equivalent in both groups (**Figure 9-4c-d**).

9.4.3 NFκB studies

2.0μg/ml of LPS induced an increase in nuclear p65 accumulation in SW480 cells after 2 hours (% cells with nuclear p65 staining: 21.5±2.9 vs. 11.8±1.8, $p<0.01$) (**Figure 9-4e-f**). This effect was abrogated with the addition of 2.0μg/ml cyclopamine

simultaneous to LPS treatment (13.5±1.1) (**Figure 9-4g-h**). This suggested that HH may have an effect on NFκB activity.

A NFκB luciferase reporter assay was optimised, with both LPS (Kruskall-Wallis p=0.0061) and TNFα ($p<0.05$) inducing robust increases in NFκB activity (**Figure 9-5a**). Whilst MDP had no effect on NFκB activity, it appeared to have a synergistic effect when added to TNFα. Cyclopamine had no effect on baseline NFκB activity at doses of 10μM, 5μM and 1μM when compared with medium only and tomatadine (10μM) controls (**Figure 9-5b**). Similarly, there was no effect with SHH at 100ng/ml. As basal NFκB activity in SW480 cells was low, cells were treated with cyclopamine 30 minutes prior to LPS. LPS was used at a physiological dose of 1μg/ml to avoid saturating NFκB activity; this produced a mean 1.70-fold increase in NFκB activity. Addition of cyclopmaine had no effect on LPS induced NFκB activity at 10μm, 5μm or 1μm concentrations (**Figure 9-3c**).

9.4.4 Effect of HH agonism on cytokine profiles of SW480 cells and PBMCs from patients with UC

PBMCs from 5 patients with acute active UC were treated with 50ng/ml and 500ng/ml rSHH and 2.0μM and 10.0μM purmorphomine with 4μl DMSO as control for purmophromine and 1μg/ml MDP, 50ng/ml TNF as positive controls. Culture supernatant, assayed for cytokine conentrations by ELISA, demonstrated significant increases in CCL20, IL1β, IL10 and TNFα with SHH 500ng/ml (complete cytokine profile for one representative patient depicted in **Figure 9-6**). CCL20 profiles were examined in more detail as this was the most novel finding of greatest potential relevance to the colonic epithelial cells under study.

CCL20 concentrations were significantly greater in PBMCs treated with rSHH (**Figure 9-7a**). A dose-dependent response was observed (Kruskall-Wallis p=0.0086) with the maximum concentration at 1000ng/ml ($p<0.05$). Purmorphomine had no effect on CCL20 (**Figure 9-7b**). However, in SW480 cells the opposite effect was observed. rSHH had no effect (**Figure 9-7c**), whilst purmorphomine decreased CCL20 (Kruskall-Wallis p=0.024) with maximal effect at 10.0mM ($p<0.05$) (**Figure 9-7d**). As basal secretion of CCL20 in SW480 cells was low (mean 62.4 ±13.3pg/ml)), MDP and TNFα were used to increase baseline secretion. MDP had no

effect, but TNFα increased CCL20 (Kruskall-Wallis p=0.0090) with maximal effect at 50ng/ml ($p<0.01$) (**Figure 9-7e**). Purmorphomine (2μM and 10μM), but not rSHH, co-treatment with TNFα decreased CCL20 back to baseline levels (purmorphomine 2μM and 10μM vs. TNFα 50ng/ml p=0.024).

9.5 *Discussion*

In the present study it has been shown that the NOD2 ligand MDP suppresses PTCH expression, that HH inhibition has no consistent effect on NFκB activity, and that exogenous SHH induces significant expression of CCL20 in PBMCs in patients with active UC.

This series of *in vitro* studies has a number of problems that have largely come to light since 2006 as the published literature has substantially clarified the role of HH signalling in both the intestine and immune system. Of greatest importance, the debate between autocrine and paracrine HH signalling in the intestine has now been resolved, thanks largely to carefully controlled *in vivo* experiments in WT and genetically modified mice. HH signalling is exclusively paracrine in the intestine as has been demonstrated throughout mammalian development,[303] homeostasis (Kolterud A et al, personal communication), inflammation (Lees *et al*, *PLoS Medicine* 2008)[423] and cancer (**1.9.1.2**).[361]

As a result of these developments, it is now clear for a number of reasons that colonic adenocarcinoma cell lines, or indeed any intestinal epithelial cells in monoculture, are an inappropriate model to address the questions posed here. Firstly, and most importantly, the system is too simple to address a complex and dynamic pathway like HH, that signals from epithelium (SHH and IHH) to its response network (PTCH, SMO, GLI1) in mesenchymal cells. Co-culture models (e.g. SW480 cells and Jurkat T cells) could go some way to address this; although, a better solution would be to stimulate isolated mesenchyme with HH agonists / antagonists (as described in murine tissue by Zacharias and colleagues; personal communication). The optimal models are genetically modified mice, as we have subsequently shown with *Gli1*$^{+/LacZ}$ mice (future plans are discussed in detail in Chapter 12).[423]

Secondly, the purtubations in HH signalling in colonic cancer bring into doubt the use of an adenocarcinoma cell line to address questions primarily pertaining to inflammation.[318, 330, 346, 359, 361] This is a big potential flaw in a large body of the published IBD literature that has made use of a variety of different colonic adenocarcinoma lines over many years. Thirdly, it is clear that the same cell lines (by name at least) vary widely not only between different laboratories but also between different passage numbers within the same laboratory. This latter point, in part, contributes to some of the wide error bars plotted for some of the datasets in this chapter. SW480 cells, themselves, are now notorious for their chromosomal duplications and high mutation rates with extreme genotyping differences between different groups. Whilst our SW480 cells are similar in their expression of HH components to those studied by Qualtrough,[359] the Chatel study found no expression in their SW480s.[360]

These factors considered, the fact that PTCH and GLI1 are expressed in the SW480 cell lines in our lab means that it is worth analysing the outcomes of these experiments in a little more depth. The downregulation of PTCH with MDP is noteworthy for a couple of reasons. Firstly, this seems to be the explanation for the change in PTCH expression with higher doses of LPS. It is now increasingly recognised that a large number of standard LPS preparations (like the one used here from Sigma) are contaminated with endotoxin and MDP. Ultra-pure preparations of LPS are now commercially available and these could be used to test this theory. Secondly, the identification of 2 copies of the GLI1 consensus motif upstream of *NOD2* further supports the notion that HH modulates NOD2 expression.[411] Whether HH is up- or down-stream of NOD2 (with or without a feedback loop, which the PTCH data may indicate) is not clear from the present data, and merits further testing in non-cancerous HH responsive cells (e.g. lamina propria myeloid cells)[423] in the first instance.

The studies in PBMCs from patients with active UC follow on firstly from data generated by the Howie group demonstrating the effects of exogenous rSHH on the secretion of cytokines in peripheral CD4 T cells *in vitro* (IL2, IL10 and IFNγ),[366] and secondly from the dysregulated expression of HH signalling demonstrated in UC patients in Chapter 4. The effects on CCL20 are novel and particularly noteworthy

given the important role this chemokine has been shown to play in IBD pathogenesis, as will be discussed in detail in the next chapter (where the contribution of CCL20 promoter polymorphisms to IBD pathogenesis is explored). The present cytokine studies are relatively preliminary and merit further testing in patients with quiescent UC, active and quiescent CD and a panel of controls. In addition, the data we have recently generated identifying myeloid APCs as targets of HH signalling in reponse to inflammation in *Gli1*$^{+/LacZ}$ mice, suggest a series of future studies examining the effects of HH agonists on specific subsets of these cells *in vitro* (see Chapter 12).[423]

Finally, the largely negative NFκB studies presented here do not exclude the possibility that HH and NFκB signalling interact *in vivo* for many of the reasons outlined above. It has been demonstrated that SHH is downstream of NFκB in epidermal keratinocytes and the pancreas.[371, 372] Most recently, it has been shown that SHH is directly regulated by NFκB.[373]

In summary, the data presented herein provide interesting avenues for exploration particularly with regard the effect of PAMPs on HH pathway activity and the effect of HH agonists on cytokine synthesis and secretion in patients with active disease. However, as discussed, the significant limitations of using the cell culture models described warrant the detailed study of the potential mechanisms *in vivo*. These avenues will be explored in more depth in Chapter 12.

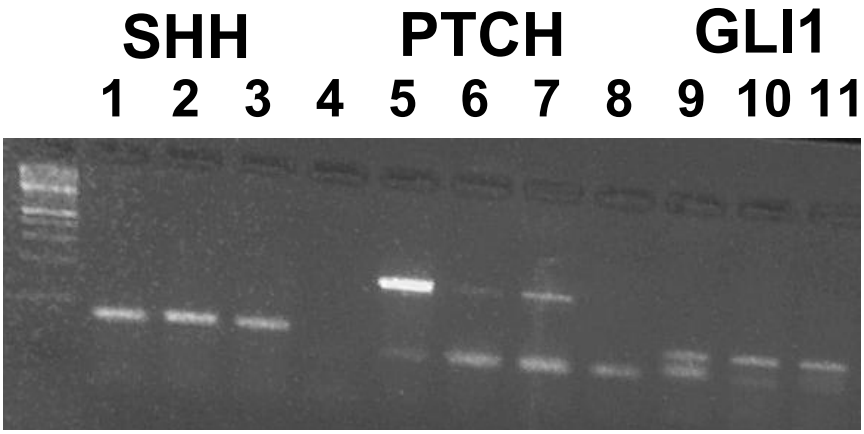

Figure 9-1 Expression of SHH, PTCH and GLI1 mRNA in untreated SW480 cells.
RT-PCR analysis of the expression of SHH (211bp product; lanes 1-3), PTCH (462bp product; lanes 5-7) and GLI1 (244bp product; lanes 9-11). Lanes 4, 8 and 12 contained water as a negative control.

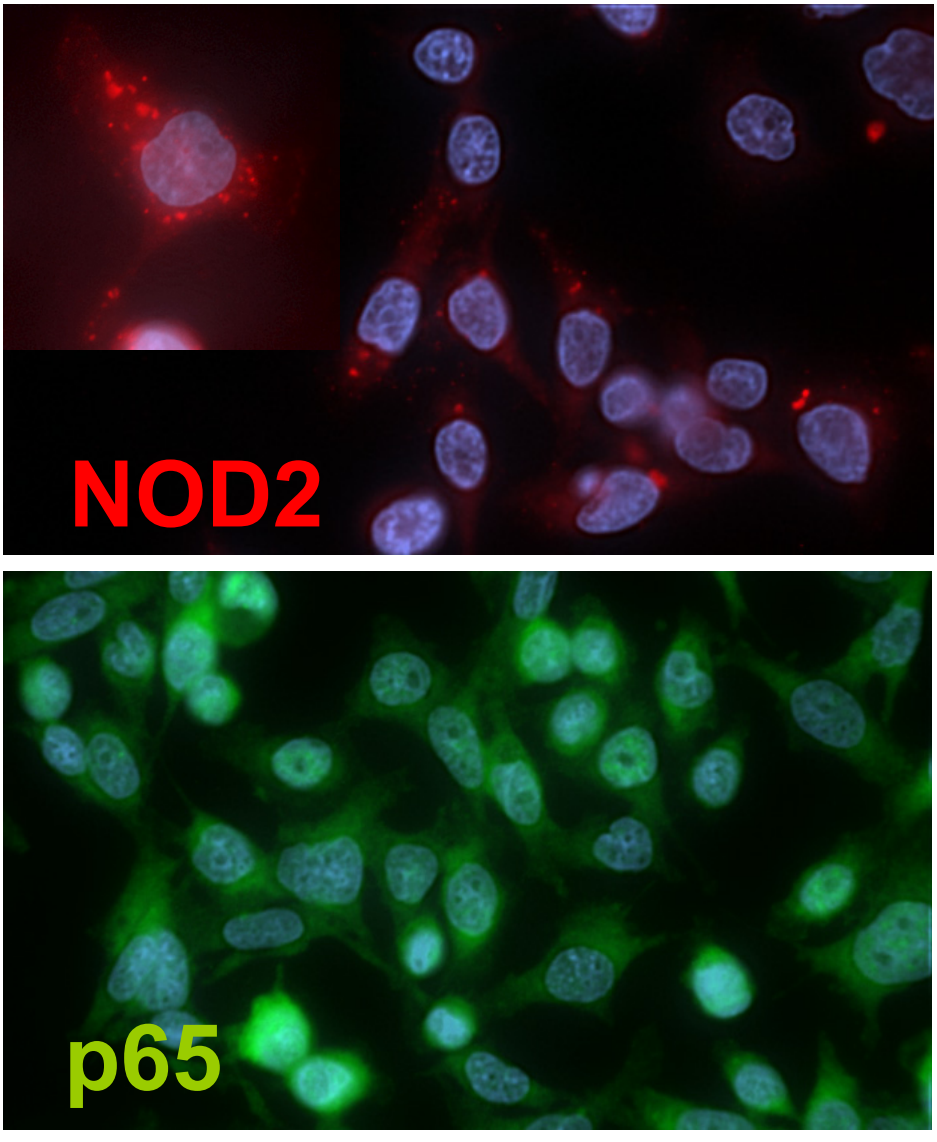

Figure 9-2 Immunofluoresence staining of NOD2 and p65 in untreated SW480 cells.

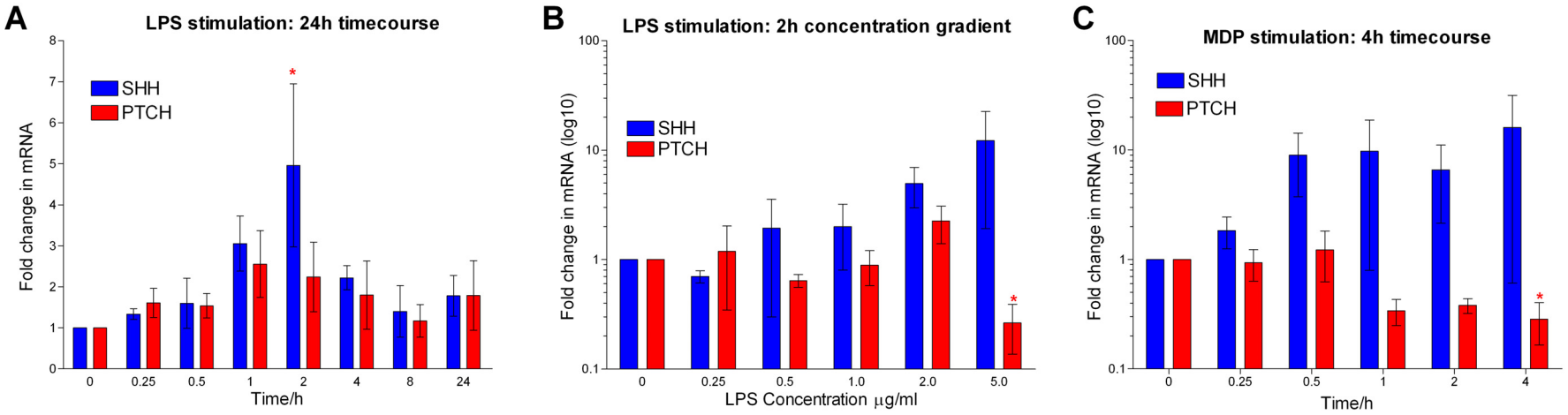

Figure 9-3 Fold changes in SHH and PTCH mRNA expression in SW480 cells treated with LPS and MDP.
A. SW480 cells were stimulated with 2.0µg/ml of LPS for 0.25h, 0.5h, 1h, 2h, 4h, 8h and 24h. FAM and VIC values obtained from real-time amplication of cDNA sequences were normalised to unstimulated mRNA controls, assigned values of 1. For 0.25h cultures n=4; 0.5h cultures n=4; 1h cultures n=5; 2h cultures n=7; 4h cultures n=4; 8h cultures n=3; and 24h cultures n=6. **B.** SW480 cells were stimulated with 0.25µg/ml, 0.5µg/ml, 1.0µg/ml, 2.0µg/ml and 5.0µg/ml LPS for 2 hours. N=3 (except for 2.0µg/ml where n=8). Relative expression plotted on a $1og_{10}$ scale. PTCH Kruskall-Wallis p=0.31; 5µg/ml vs. control $p<0.05$. SHH Kruskall-Wallis p=0.23, **C.** Cells were stimulated with 2.0µg/ml MDP for 0.25h, 0.5h, 1h, 2h, and 4h. N=3. Relative expression plotted on a $1og_{10}$ scale. PTCH Kruskall-Wallis p=0.065; 4h vs. 0h $p<0.05$. SHH Kruskall-Wallis p=0.96.
For all datasets, a Kruskal-Wallis test was performed followed by Dunn's Multiple Comparison test, comparing untreated cells with treated cells at different timepoints and concentrations of LPS and MDP. (*) denotes $p< 0.05$. (**) denotes $p<0.01$. Each bar indicates the mean with error bars for ± SEM.

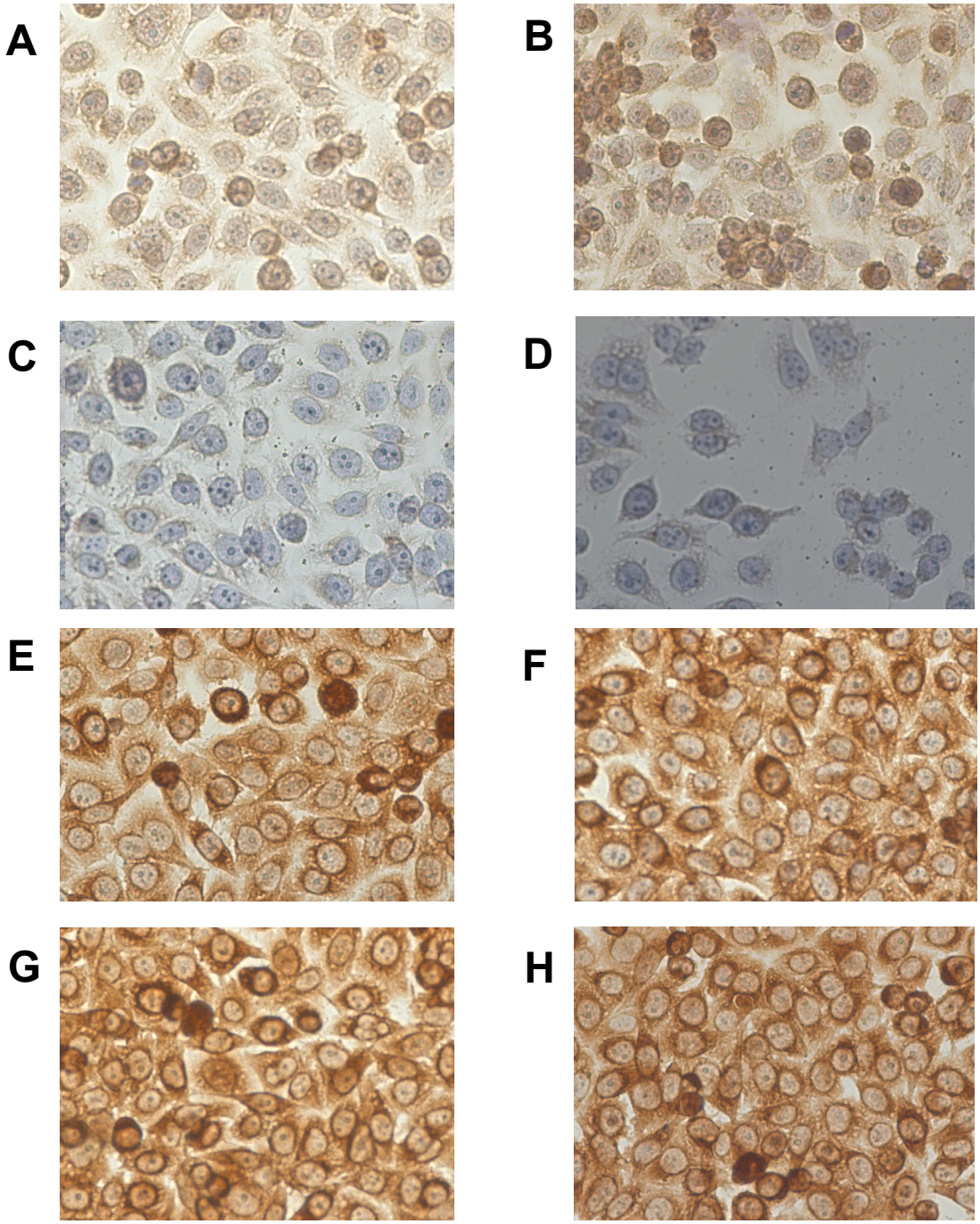

Figure 9-4 GLI1 and p65 immunohistochemistry on SW480 cells treated with LPS and cyclopamine.
Cells were stained for GLI1 protein following stimulation with 2.0µg/ml LPS (**A**) or control (**B**) for 2 hours. Negative controls were included for each experiment with omission of primary antibody (with (**C**) and without (**D**) LPS). Cells were stained for p65 protein following stimulation with 2.0µg/ml LPS only (**E**), control (**F**), 2.0µg/ml LPS plus 2.0µg/ml cyclopmaine (**G**) or 2.0µg/ml cyclopamine only (**H**). Original magnification x400.

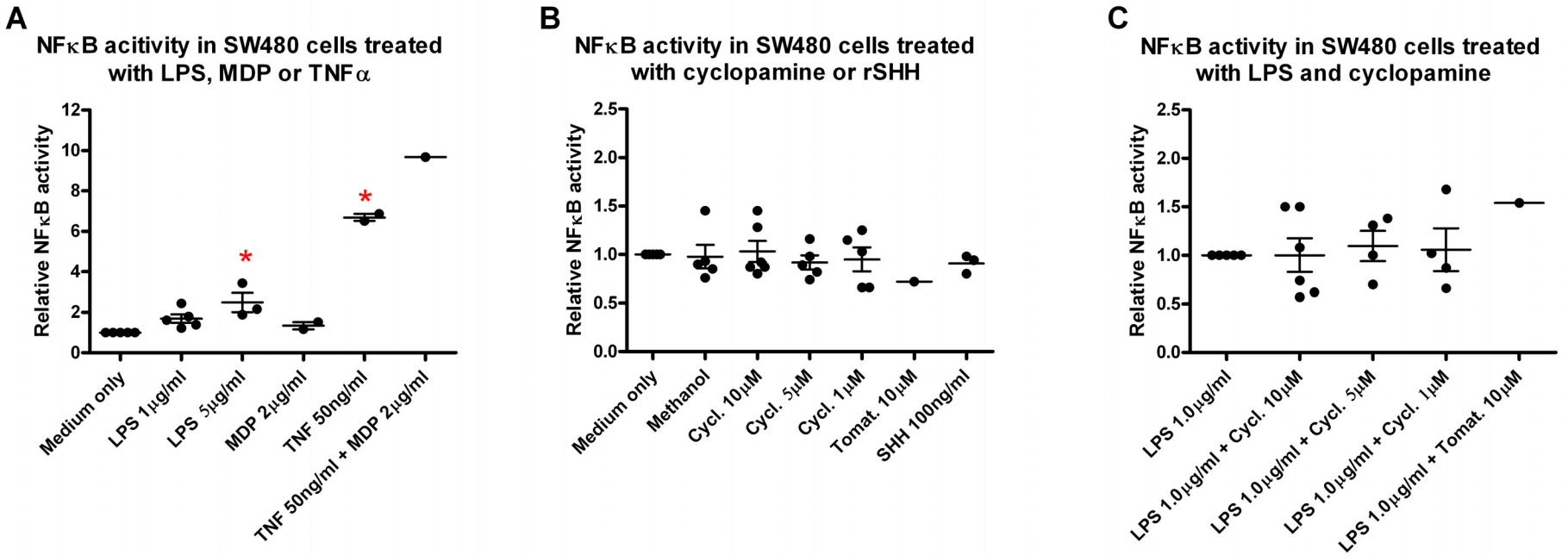

Figure 9-5 NFκB luciferase reporter assay in SW480 cells treated with PAMPs, TNFα, cyclopamine and rSHH.
A. SW480 cells were treated with 1.0μg/ml or 5.0μg/ml LPS, 2.0μg/ml MDP, 50ng/ml TNFα, or 2.0μg/ml MDP + 50ng/ml TNFα. NFκB activity was normalised to medium only controls, designated 1. Kruskall-Wallis (LPS 5μg/ml, LPS 1μg/ml and control) p=0.0061; LPS 5μg/ml vs. control p<0.01. **B.** SW480 cells treated with 10μM, 5μM, 1μm cyclopamine (cycl.), 10μM tomatadine (tomat.), and 100ng/ml SHH. NFκB activity normalised to medium only controls. Methanol was included as an additional control. **C.** SW480 cells treated with LPS 1.0μg/ml ± 10μM, 5μM or 1μM cyclopamine or 10μM tomatadine. NFκB activity normalised to LPS 1.0μg/ml treated cells. Individual data points are plotted with mean ± SEM. All treatment were for 24 hours. Statistical analysis was performed using Kruskall-Wallis test with Dunn's multiple comparison test on selected columns. Comparisons that remained significant using this test are indicated by * (p<0.01) and ** (p<0.01).

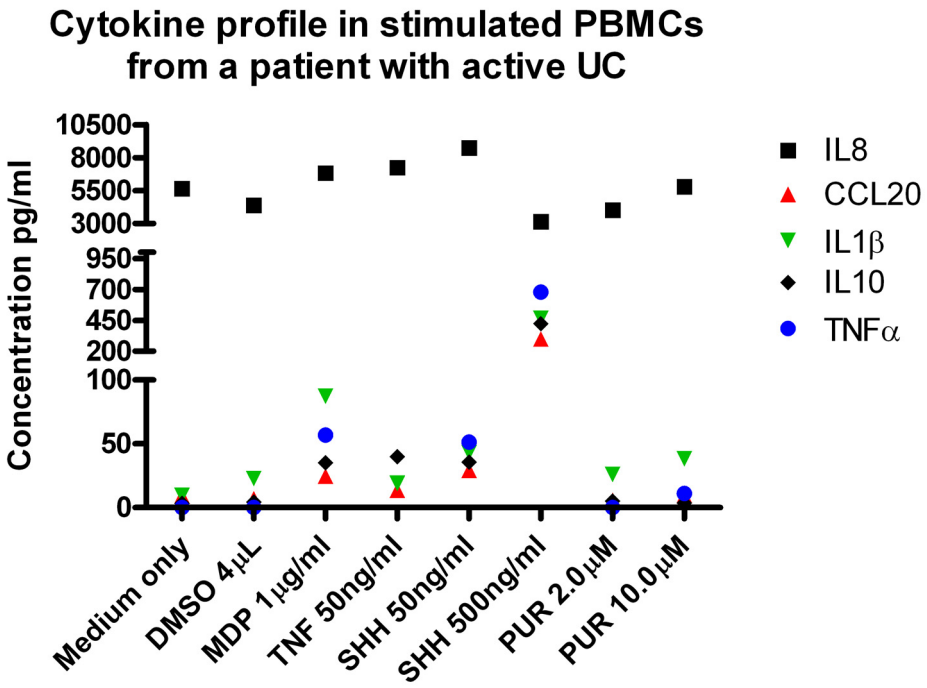

Figure 9-6 Cytokine profile after SHH and purmorphamine treatment of PBMCs from a patient with acute, severe UC.
Cytokine profile (IL8, CCL20, IL1β, IL10 and TNFα) measured by ELISA in culture supernatant from a patient with severe active UC after 24 hours stimulation as shown. Note that the y axis is split to accommodate range of values (all in pg/ml).

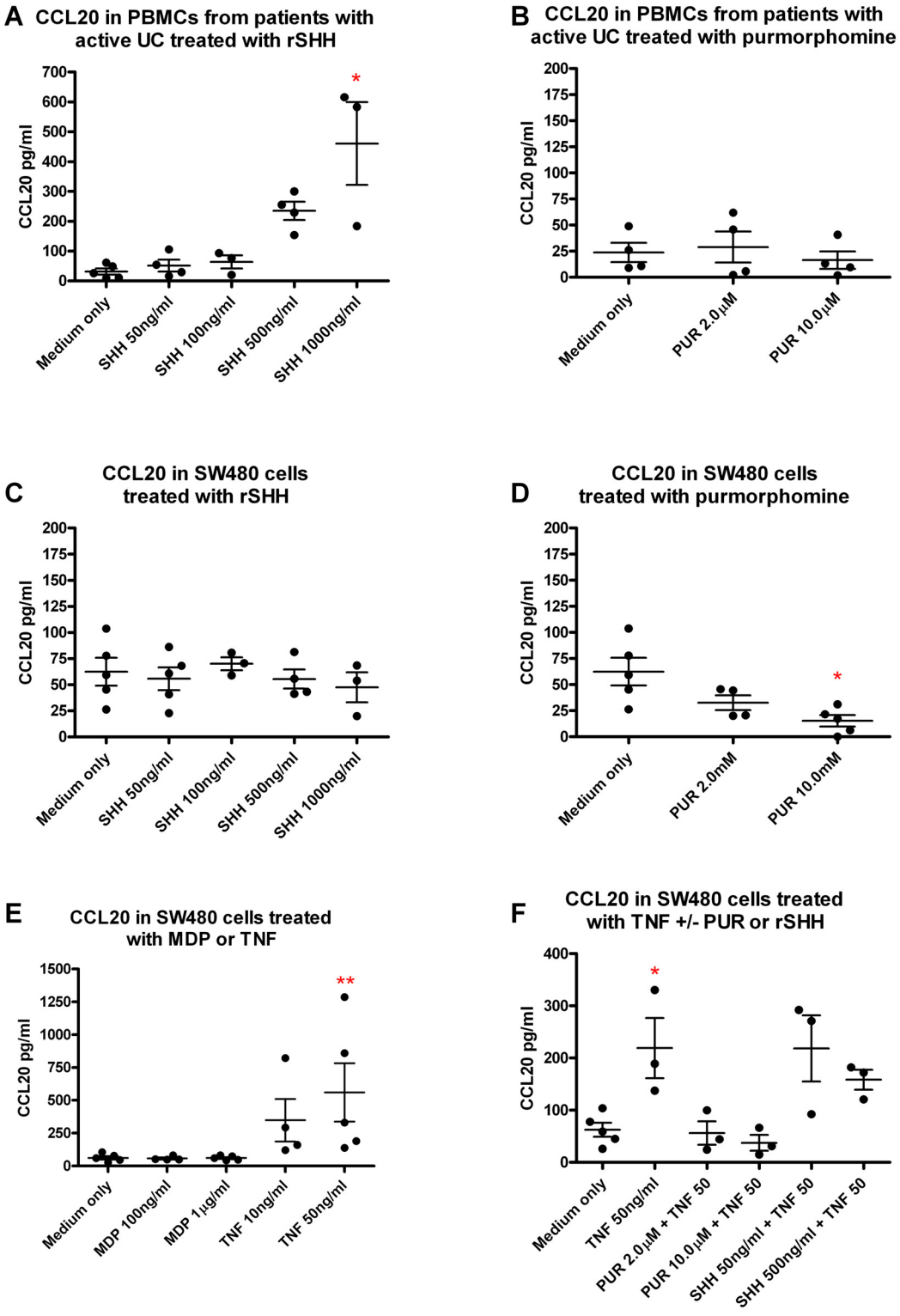

Figure 9-7 CCL20 in treated PBMCs and SW480 cells.
CCL20 concentration (ELISA) in culture supernatant of PBMCs from patients with active UC and SW480 cells treated as shown. **A.** PBMCs from patients with active UC treated with rSHH at increasing concentrations from 50 – 1000ng/ml. Kruskall-Wallis p=0.0086; 1000ng/ml rSHH vs. medium only $p<0.05$. **B.** PBMCs from patients with active UC treated with 2.0 or 10.0mM purmorphomine (PUR). **C.** SW480 cells treated with rSHH at increasing concentrations from 50 – 1000ng/ml. **D.** SW480 cells treated with 2.0 or 10.0mM purmorphomine (PUR). Kruskall-Wallis p=0.024; 10.0mM PUR vs. medium only $p<0.05$. **E.** SW480 cells treated with 100ng/ml and 1μg/ml

MDP and 10ng/ml and 50ng/ml TNFα. Kruskall-Wallis test (medium only, 10ng/ml and 50ng/ml TNFα) $p=0.0090$; 50ng/ml TNFα vs. medium only $p<0.01$. **F.** SW480 cells treated with 50ng/ml TNF +/- PUR or rSHH. Purmorphomine co-treatment decreased CCL20 back to baseline levels compared with TNFα treatment alone (2μM and 10μM + TNFα vs. TNFα $p=0.024$, Mann-Whitney U test).
All treatments were for 24 hours. Statistical analysis was performed using Kruskall-Wallis test with Dunn's multiple comparison test on selected columns. Comparisons that remained significant using this test are indicated by * ($p<0.01$) and ** ($p<0.01$).

10 Analysis of *CCL20* promoter variants in Japan, Sweden and the U.K. provides further evidence for genetic heterogeneity in IBD.

10.1 *Abstract*

Introduction. It was shown in Chapter 9 that rSHH induces expression of CCL20 in PBMCs from patients with acute severe UC. *CCL20* is a good candidate gene for IBD susceptibility based on positional (2q33-37), functional and expression studies. The *CCL20* gene encodes an inducible chemokine involved in maturation and recruitment of dendritic cells and has antimicrobial activity. Expression in intestinal epithelial cells and PBMCs is induced under inflammatory conditions. *CCL20* promoter polymorphisms were recently implicated in susceptibility to UC in a South Korean population.

Aims and Methods. The aim of the present study was to analyse the contribution of these variants to disease susceptibility in both Eastern and Western populations. Mutation screening revealed two SNPs (1706G→A and rs6749704T→C) in the *CCL20* promoter that were genotyped in 1531 IBD patients and 833 HC, from two Northern European (Scotland and Sweden, 2492 individuals) and one Far Eastern population (Japan, 395 individuals).

Results. The 1706G→A variant frequencies were strikingly different in Japanese HC compared with N European populations (38.4% vs. 13.3%, $p<0.00001$). The 1706G→A variant was associated with disease susceptibility in Japanese patients with UC (47.3% UC vs. 38.4% HC, $p=0.018$, OR 1.44, C.I. 1.07-1.92) but not in the combined N European cohort (by case-control, $p=0.09$ and transmission disequilibrium testing, $p>0.5$). The Japanese population-attributable risk of the 1706G→A variant was 14.3%. No association was seen in any of the populations for the rs6749704T→C SNP. Genotype-phenotype analysis, revealed no significant associations in any population.

Conclusions. These data complement studies of *NOD2*, HLA and IBD5, providing compelling evidence for genetic heterogeneity in IBD susceptibility between Eastern and Western populations.

10.2 *Introduction*

IBD was previously very uncommon in Far Eastern countries, but it is of considerable interest that incidence has increased rapidly in the past two decades as these countries adopt a more westernised lifestyle.[1, 465-469]

Chemokines have been recognised to play important roles as mediators of intestinal homeostasis. The CC chemokine CCL20 (macrophage inflammatory protein-3alpha [MIP-3α], liver and activation-regulated chemokine [LARC], Exodus) is noteworthy for its genomic location, protein structure, and its function.[470] The *CCL20* gene was identified independently by three groups in 1997 following a bioinformatics-based approach.[471-473] *CCL20* has 20-31% sequence similarity with other human CC chemokines. It is expressed mainly in mucosal and lymphoid tissues, including adult small intestine and colon,[471, 474-476] and PBMCs.[472, 473, 477] CCR6 was identified as the CCL20 receptor in 1997.[478] CCL20 is the only CCR6 triggering chemokine and CCL20 is unable to elicit a biological response through other known chemokine receptors. This specificity contrasts with the general promiscuity of chemokines and their receptors.

It is of great interest that *CCR6* was recently confirmed as a CD susceptibility locus.[103] Whilst it did not reach genome-wide levels of significance in the individual GWAS reports, association was confirmed at rs2301436 ($p = 3.30 \times 10^{-7}$) on meta-analysis of UK, French/Belgium and N American studies and subsequent replication ($p_{replication} = 3.26 \times 10^{-7}$; $p_{TDT} = 0.0057$; $p_{combined} = 5.22 \times 10^{-13}$, $OR_{case\text{-}control}$ 1.21, OR_{TDT} 1.16), as well as at an adjacent SNP rs7749278 ($p = 1.65 \times 10^{-7}$, $p_{replication} = 0.0012$, $p_{combined} = 1.58 \times 10^{-9}$).[103] Fine-mapping studies are required to confirm that this association is limited to *CCR6* as the presently defined block contains two other genes.

Together, CCL20 and CCR6 are involved in the maturation of DCs.[470] They play a role in the recruitment of immature DCs and their precursors to sites of potential antigen-entry. The *Ccr6*$^{-/-}$ mouse lacks any gross abnormalities in any major organ.[479, 480] However, it has underdeveloped Peyer's patches and CD11b+ myeloid

DCs, known for their functional CCR6 expression, are absent from the subepithelial dome. *Ccr6*$^{-/-}$ mice also have increased subpopulations of T cells in the intestinal mucosa. CCL20 expression, constitutively at low levels in human intestinal epithelium, is induced in inflamed colonic tissue from patients with IBD,[476, 481] as well as in PBMCs of patients with active UC compared with HCs.[477] CCL20 has been demonstrated to have antibacterial activity comparable to β-defensins.[482] Indeed, human β-defensins have been reported as non-chemokine ligands for CCR6, mediating the *in vitro* chemotactic activity of β-defensins for immature DCs and T cells.

The *CCL20* gene maps to 2q33-37, away from the major chemokine cluster at 17q11.2, suggesting that the gene was generated prior to amplification of the chromosome 17 chemokine cluster.[470] Two original genome-wide linkage studies (UK and US populations)[99, 150] and two meta-analyses in IBD[97, 102] suggest a region of linkage on the long-arm of chromosome 2 (**Figure 10-1** and **Figure 1-4**). CC chemokines genes are structurally similar with 3 exons and 2 introns. In contrast, the *CCL20* gene has 4 exons and 3 introns. In addition, two potential splice sites at the intron 1 / exon 2 boundary have been identified resulting in the deletion of three bases. The 96 amino acid precursor protein is cleaved between two alanine residues, giving rise to a mature protein of 70 amino acids. Several cDNA clones lack one alanine residue in the precursor, leading to an alternative mature protein of 69 amino acids. The *CCL20* promoter is of some interest, containing possible binding sites for different transcription factors; NFκB, activator protein-1 (AP1) and AP2, CAAT/enhancer-binding protein (C-EBP), stimulating protein-1 (Sp1) and the epithelium-specific Ets nuclear factor ESE-1.[483-486] In various different cell types NFκB has been demonstrated to bind and activate the *CCL20* promoter.[475, 483, 486, 487] Of further relevance to IBD pathogenesis, IL23 and IL1β induce the development of Th17 cells expressing CCL20.[488] Finally, IL21 stimulation of intestinal epithelial cells induces synthesis of CCL20 and treatment of IBD mucosal biopsies in organ culture anti anti-IL21 decreases synthesis of CCL20.[489]

Preliminary data, not yet published as a full paper,[490] demonstrated a highly significant association between *CCL20* promoter polymorphisms and UC in a South Korean population. We have set out to assess the contribution of *CCL20* promoter

polymorphisms in four complementary and contrasting populations of IBD: Japanese UC and HC, Scottish CD, UC and HC, Scottish early-onset CD and UC, and Swedish UC and HC. This has allowed not only substantial replication of the preliminary South Korean data, but in addition provides a thorough exploration of genetic heterogeneity between Far Eastern and Northern European populations in IBD susceptibility.

10.3 *Subjects and Methods*

10.3.1 Scottish subjects

The Scottish population consisted of 1022 IBD cases (454 CD, 540 UC and 27 IBDU) and 370 HCs (**Table 10-1**). This was made up of an adult IBD cohort (age 17 years or over at diagnosis), and an early-onset IBD cohort (age ≤16 years at diagnosis), collected prospectively and independently.

Adult-onset subjects: This cohort of 708 patients consisted of 249 with CD and 459 with UC. The population was 98.5% white, non-Jewish Caucasian. The median age at diagnosis was 32.8 (IQR 25.0 – 47.5).

Early-onset cohort: There were 205 CD patients, 81 with UC and 27 with IBDU, recruited from across Scotland. In addition, blood samples were collected from 523 parents (293 mothers; 230 fathers) resulting in 69% complete family trios.

Healthy controls: The control group comprised 370 individuals; blood donors (n = 260) and healthy control subjects (n = 110) all residing in SE Scotland. There were 187 males and 183 females; median age was 38.0 (IQR, 27.0-50.0).

10.3.2 Swedish subjects

296 UC patients were recruited from various hospitals in Stockholm County (**Table 10-1**). 281 Swedish HC were healthy volunteers from the Karolinska University Hospital staff (n=170) and orthopaedic day-case surgery patients (n=111) with no previous medical conditions.

10.3.3 Japanese subjects

Japanese subjects were recruited from Osaka and Tokyo. The population consisted of 213 patients with UC and 182 Japanese HC (**Table 10-1**).

10.3.4 Genotyping

Custom Taqman SNP Genotyping Assays for 1706G→A and rs6749704T→C SNPs were obtained from ABI by submitting the context sequence to the Assay By Design service. Genotypes were derived by a combination of the Taqman system and direct sequencing.

10.3.5 Data analysis

Disease susceptibility was determined by case-control study of genotype frequency for each SNP in all populations described (Scottish adult-onset IBD, CD and UC; Scottish early-onset IBD, CD and UC; Swedish UC; and Japanese UC). Minitab statistical software version 13 (Minitab Ltd, Coventry, UK) was used to analyse genotype-phenotype associations using chi^2 test or Fisher exact test where appropriate for univariate analysis. TDT analysis was performed on the early-onset trios using TRANSMIT version 2.5.[491]

10.3.6 Initial Mutation Screening

The promoter region of the *CCL20* gene was sequenced in the Scottish adult population in 12 CD and 12 UC patients. Of 4 SNPs previously identified in Choi's study,[490] only 1706G→A and rs6749704T→C were present in the Scottish population. These 2 SNPs were analysed in the Scottish adult and early-onset IBD populations and HC, Swedish UC and HC, and Japanese UC and HC.

10.3.7 Genotyping success rates

Genotypes were derived in >96% of Scottish, Swedish and Japanese subjects for 1706G→A and rs6749704T→C. Genotyping failures were consistent for both assays, representing poor DNA quality in these samples and not affecting overall validity of the assays. The validity of the assays were further verified by direct sequencing of 10-20% of all HC's, with 100% concordance with Taqman data. All data derived from Scottish and Japanese patients and controls were consistent with HWE. The Swedish HC population deviated significantly from HWE for rs6749704T→C ($p<0.01$); this most likely represents uneven sex distribution in this cohort (predominantly females) and not genotyping error. The genotype-based analysis performed in this cohort does not assume HWE.

10.4 *Results*

10.4.1 Expression of CCL20 and CCR6 in IBD tissue

Expression analysis of microarray data demonstrated increased expression of CCL20 in both UC and CD compared with HC, but this was limited to the inflamed group of biopsies ($p<0.0001$) with no difference between non-inflamed HC and non-inflamed UC or CD expression levels (**Figure 10-2d-e**). There was no change in expression of CCR6 in UC ($p=0.096$) or CD ($p=0.18$) compared with HC (**Figure 10-2c**).

10.4.2 1706G→A variant

10.4.2.1 Japanese

Analysis of the 1706G→A variant allele demonstrated a significant association with UC in Japanese patients when compared with Japanese HC (47.3% UC vs. 38.4% HC, $p=0.018$, OR 1.44, C.I. 1.07-1.92). Genotype frequencies were significantly different between cases and controls (AA vs GG, $p=0.027$, OR 1.89, CI 1.07-3.33) (**Table 10-2a**). Analysis of genotypes as a 2x3 contingency table approached significance when tested for independence ($p=0.057$); however, chi^2 testing for trend was significant ($p=0.0196$). The population-attributable risk for the 1706G→A variant was 14.3%.

10.4.2.2 Northern Europeans

Analysis of the 1706G→A genotypes in the combined N European IBD cohort did not demonstrate any significant difference between cases and controls (GA vs. GG, $p=0.09$, OR 1.22, CI 0.97-1.54; AA vs. GG, $p=0.22$, OR 0.68, CI 0.36-1.28) (**Table 10-2a**). There was no association with this SNP in these populations when analysed independently (Scottish IBD, UC, CD vs. Scottish HC or Swedish UC vs. Swedish HC; full genotypic data available in **Table 10-3**) for differences in genotype frequencies by chi^2 testing. TDT of the Scottish early-onset IBD families revealed no significant deviation from expected transmission rates for all IBD cases, and for CD and UC independently (**Table 10-4a-c**).

10.4.2.3 Genetic heterogeneity

In the Japanese HC population 1706G→A variant allelic frequency was 38.4%, significantly higher than the N European frequencies (12.3% Scottish HC, $p<0.00001$;

14.8% Swedish HC, p<0.00001; combined 13.3%, p<0.00001). Japanese UC allelic frequency was 47.3%, compared with 14.6% in N European UC (p<0.0001).

There were no differences between the Scottish and Swedish datasets, with similar HC allelic frequencies for 1706G→A (14.8% vs. 12.3%, p=0.17). rs6749704T→C HC allelic frequencies were similar in all three populations (Japanese HC 31.7%, Scottish HC 29.2%, Swedish HC 26.3%, p=0.22).

10.4.2.4 Comparison with South Korean dataset

In the previously reported S Korean dataset,[490] HC 1706G→A variant allelic frequencies were 39.1%, similar to our Japanese allelic frequency (38.4%, p>0.5). In the S Korean UC cases the 1706G→A variant allelic frequency was 60.7%, compared to 47.3% in Japanese UC patients (p=0.0007). This result was significant comparing cases and controls in the S Korean population (60.7% vs. 39.1%, p<0.0001). Variant homozygosity for this SNP was 38.6% in the S Korean UC cases, compared with 24% in Japanese cases (p<0.001), whilst HC frequencies were similar for both populations (16.3% vs 17.5%, p>0.5). These data demonstrate a much stronger contribution to disease susceptibility in S Korea than detected in the Japanese population.

10.4.3 rs6749704T→C variant

There was no association with the rs6749704T→C SNP when genotype frequencies were compared between Japanese patients with UC and Japanese HC (TC vs TT, p=0.57, OR 0.88, CI 0.58-1.35; CC vs TT, p=0.88, OR 0.95, CI 0.49-1.85; chi^2 test for independence, p=0.85) (**Table 10-2b**). However, in the combined N European cohort there was a significant difference between genotype frequencies (CC vs. TT, p=0.016, OR 0.65, CI 0.46-0.92; chi^2 test for independence, p=0.0022) (**Table 10-2b**; genotype frequencies for individual populations available in **Table 10-3**). This finding is of uncertain significance given the very similar allelic frequencies in the N European populations, and the negative TDT analysis in the Scottish early-onset IBD cohort for this SNP (**Table 10-4**).

10.4.3.1 Genotype-phenotype analysis

There were no significant associations on univariate genotype-phenotype analysis (variant allelic frequency, carriage rate and homozygosity), when age of onset, disease

location, disease behaviour, surgery, family history, smoking, or history of joint problems were assessed separately in adult and early-onset Scottish IBD populations (data not shown). No genotype-phenotype associations of significance were identified in any of the populations, once corrections for multiple comparisons were made in analysis of the rs6749704T→C variant (data not shown).

10.5 *Discussion*

Following initial data from S Korea, now only available in abstract form,[490] this is the first study to report formally on *CCL20* promoter polymorphisms and susceptibility to IBD. It has shown a positive association with the 1706G→A variant and inherited susceptibility to UC in a Japanese population. However, in two independent N European populations no association was found with IBD, CD or UC for this variant by case-control analysis or TDT. It is therefore of great interest that the only other group to report an association with this variant was from a similar geographical location to this Japanese population.[490] However, the very striking positive association reported in preliminary data by a S Korean group ($p<0.0001$, OR 2.4) are in contrast to the relatively modest effect this variant has in determining susceptibility to UC in the Japanese population ($p=0.018$, OR 1.44). Whilst variant allelic frequencies for 1706G→A were very similar in both S Korean and Japanese HC populations (39.1% and 38.4% respectively), the cohort of patients with UC from S Korea had significantly higher variant frequencies than those from Japan (60.7% vs. 47.3%, $p<0.0001$). There is a clear precedent for the strength of a genetic association being strongest in the index study than in subsequent replication cohorts.[492] In IBD this was clearly demonstrated for *NOD2*.[79, 80, 106] However, it is somewhat curious that over 3 years since the S Korean data was presented in abstract form, a full paper has yet to be published.

Combined, these data present further compelling evidence for global genetic heterogeneity in the inherited susceptibility to IBD. CD and UC are recognised to be diseases of the developed world, with highest incidence rates reported in Western Europe and N America. Within Scotland, the incidence of early-onset CD has increased dramatically in recent decades.[493] Intriguingly, there are latitudinal gradients of disease incidence reported, with higher rates documented not only in the north of Europe compared with the South,[494] but also within Scotland,[495] and within

France.[496] Different genetic backgrounds may play a role in these geographical variations, although these remain to be elucidated. *NOD2* variants cannot explain this, as frequencies decrease with more northern latitudes in Europe,[106, 107] where the PAR ranges from 7.9% (early-onset Scottish CD) to 11.4% (Swedish CD). Temporal trends are clearly not genetic, occurring over no more than a couple of generations. Environmental factors, such as increased hygiene, or refrigeration, have been postulated as aetiopathogenic mechanisms.[497] It is clear that increased rates of IBD in the Far East have lagged behind those in the West, and have been largely attributed to these countries adopting a more Westernised lifestyle.

Many dozens of centres worldwide now have established DNA banks from patients and controls, allowing the contribution of genetic variants to disease susceptibility to be observed in multiple, geographically and ethnically diverse populations. In the Far East this has been catalysed by the increased incidence of IBD over the past two to three decades.[1] In Japanese,[110] Chinese,[111] and S Korean[112] populations the three common *NOD2* mutations are absent (**Figure 1-5**). A similar phenomenon has been noted with the MHC on chromosome 6 (**1.5.4.1**) and the IBD5 locus (**1.5.1.2**). Common variants identified in the IBD5 risk haplotype in European and N American populations are very rare in the Japanese cohorts studied.[127, 498, 499] A degree of commonality in disease association was demonstrated with the *TNFSF15* gene (9q32), identified by a limited GWAS in a Japanese cohort with replication in two UK cohorts,[500] and subsequently confirmed as a CD susceptibility gene in the Barrett meta-analysis.[103]

There are also interesting parallels with the chemokine receptor gene *CCR5*. A 32bp deletion (denoted *ΔCCR5*) was demonstrated to prevent cell invasion by the primary transmitting strain of HIV1.[501-503] Consequently, despite repeat exposure to HIV1, homozygotes appeared to be resistant to infection.[502, 504] As demonstrated here for *CCL20*, there is significant variation in *ΔCCR5* allele frequencies in different populations globally. Within Europe, *ΔCCR5* frequency is approximately 10%; however, a North-South gradient is evident with frequencies of 14.7% documented in Iceland compared with 2.4% in Greece.[505] Further heterogeneity is demonstrated outside Europe; the variant is present in less than 1% of people of African origin, 2-5% of people from the Middle East and the Indian subcontinent, and absent in the Far

East.[505] *CCR5* is located at 3p21, coinciding with an IBD susceptibility locus identified by genome-wide scanning.[99] Several association studies in IBD and primary sclerosing cholangitis have failed to demonstrate a consistent association with *ΔCCR5*.[506-512]

Our data support a role for *CCL20* polymorphisms in susceptibility to UC in Eastern populations, but not in 2 independent N European cohorts. The functionality of the 1706G→A variant is unknown, but it may be in part due to altered binding of transcription factors such as NFκB to the *CCL20* promoter. The subtle genotypic association of rs6749704T→C between IBD and cases in the N European group is of uncertain significance given the flat allelic analysis and the negative TDT, but it may indicate that both this SNP and 1706G→A are in LD with a presently unidentified causal variant. Data from phase II of the Hapmap project describe an area of low LD around the *CCL20* gene (**Figure 10-3**). The gene itself is very small (3.68kb) and is located in an area of relatively low gene-density. We would suggest that further mutation screening across the entire *CCL20* gene in Eastern populations is required to locate any true functional variant. Additionally, before an established role for germline *CCL20* variation in susceptibility to UC in Eastern populations is ascribed, functional data describing the effect of these mutations is required.

	Scottish				Swedish		Japanese	
	All patients	**Adult-onset IBD**	**Early-onset IBD**	**Healthy controls**	**UC Cases**	**Healthy controls**	**UC Cases**	**Healthy controls**
Total number	1022	708	314	370	296	281	213	182
Male Sex (%)	513 (50.2%)	341 (48.2%)	172 (54.8%)	187 (50.5%)	163 (57.4%)	122 (44.4%)	102 (47.9%)	
Median Age at diagnosis (cases) / recruitment (controls) years (IQR)	25.9 (13.4 – 40.7)	32.8 (25.0 – 47.5)	11.1 (8.6 – 12.9)	38.0 (27.0 – 50.0)	28.2 (20.3 – 41.2)	48.4 (37.9 – 60.4)	34.0 (25.6 – 46.6)	
Non-Jewish Caucasian (%)	983/1001 (98.2%)	677/687 (98.5%)	306/314 (97.5%)	97%	275 (98.9%)	273 (99.3%)	N/A	N/A

Table 10-1 Demographic data of cohorts genotyped for *CCL20* promoter polymorphisms
Demographic details of Scottish IBD patients (adult-onset and early-onset disease) and controls, Swedish UC patients and controls, and Japanese UC patients and controls.

A.	EASTERN				NORTHERN EUROPEAN			
	JAPANESE				COMBINED			
	HC N=182	UC N=213	GA vs GG p-value OR CI	AA vs GG p-value OR CI	HC N=648	IBD N=1316	GA vs GG p-value OR CI	AA vs GG p-value OR CI
	5/182 (2.7%)	*13/213 (6.1%)*			31/648 (4.8%)	31/1316 (2.4%)		
GG	72/177 (40.7%)	59/200 (29.5%)	0.068 1.53 0.97-2.43	0.027 1.89 1.07-3.33	470 (76.1%)	941 (73.2%)	0.09 1.22 0.97-1.54	0.22 0.68 0.36-1.28
GA	74/177 (41.8%)	93/200 (46.5%)			130 (21.1%)	318 (24.7%)		
AA	31/177 (17.5%)	48/200 (24.0%)			17 (2.8%)	23 (1.8%)		
	*2x3 chi^2 for independence 0.057				*2x3 chi^2 for independence 0.094			

B.	EASTERN				NORTHERN EUROPEAN			
	JAPANESE				COMBINED			
	HC N=182	UC N=213	TC vs TT p-value OR CI	CC vs TT p-value OR CI	HC N=648	IBD N=1316	TC vs TT p-value OR CI	CC vs TT p-value OR CI
	2/182 (1.1%)	*3/209 (1.4%)*			10/648 (1.5%)	73/1316 (5.5%)		
TT	86/180 (47.8%)	104/206 (50.5%)	0.57 0.88 0.58-1.35	0.88 0.95 0.49-1.85	348 (54.5%)	653 (52.5%)	0.06 1.21 0.99-1.49	0.016 0.65 0.46-0.92
TC	74/180 (41.1%)	79/206 (38.3%)			223 (35.0%)	508 (40.9%)		
CC	20/180 (11.1)	23/206 (11.2%)			67 (10.5%)	82 (6.6%)		
	*2x3 chi^2 for independence 0.85				*2x3 chi^2 for independence 0.0022			

Table 10-2 Genotype data of *CCL20* promoter polymorphisms in Eastern and N European populations.

Genotype frequencies for **A.** 1706G→A and **B.** rs6749704T→C SNPs in Eastern (Japanese) and combined N European (Scottish and Swedish) IBD and HC populations. P-values, OR and 95% confidence intervals (CI) are given for chi^2 analysis of genotype frequencies using 2x2 contingency tables (12vs11 and 22vs11). *P-value for 2x3 contingency table analysis.

	Scottish sub-populations										Swedish	
A.	**HC**	**All IBD**	**IBD >17yrs**	**IBD <17yrs**	**All CD**	**CD >17yrs**	**CD <17yrs**	**All UC**	**UC >17yrs**	**UC <17yrs**	**HC**	**UC**
Number	**368**	**1022**	**708**	**314**	**454**	**249**	**205**	**540**	**459**	**81**	**280**	**294**
Failures	*5 (1.4%)*	*29 (2.8%)*	*19 (2.7%)*	*10 (3.2%)*	*10 (2.2%)*	*5 (2.0%)*	*5 (2.4%)*	*17 (3.1%)*	*14 (3.1%)*	*3 (3.7%)*	*26 (9.3%)*	*2 (0.7%)*
GG	283 (78.0%)	740 (75.4%)	509 (73.9%)	231 (76%)	334 (75.1%)	182 (74.6%)	152 (75.6%)	388 (74.2%)	327 (73.5%)	61 (78.2%)	187/254 (73.6%)	201/292 (69.9%)
GA	71 (19.6%)	235 (23.7%)	167 (24.2%)	68 (22.4%)	103 (23.1%)	58 (23.8%)	45 (22.4%)	125 (23.9%)	109 (24.5%)	16 (20.5%)	59/254 (23.2%)	83/292 (28.4%)
AA	9 (2.5%)	18 (1.8%)	13 (1.9%)	5 (1.6%)	8 (1.8%)	4 (1.6%)	4 (2.0%)	10 (1.9%)	9 (2.0%)	1 (1.3%)	8/254 (3.1%)	5/292 (1.7%)

B.	**HC**	**All IBD**	**IBD >17yrs**	**IBD <17yrs**	**All CD**	**CD >17yrs**	**CD <17yrs**	**All UC**	**UC >17yrs**	**UC <17yrs**	**HC**	**UC**
Failures	*8(2.2%)*	*68 (6.7%)*	*50 (7.1%)*	*18(5.7%)*	*31 (6.7%)*	*23 (9.2%)*	*8 (3.9%)*	*35 (6.5%)*	*27 (5.9%)*	*8 (9.9%)*	*3 (1.1%)*	*7 (2.4%)*
TT	188 (52.2%)	495 (51.9%)	335 (50.9%)	160 (54.1%)	226 (53.4%)	118 (52.2%)	108 (54.8%)	255 (50.5%)	217 (50.2%)	38 (52.1%)	160/278 (57.6%)	158/289 (54.7%)
TC	134 (37.2%)	396 (41.5%)	280 (42.6%)	116 (39.2%)	169 (40.0%)	94 (41.6%)	75 (38.1%)	215 (42.6%)	186 (43.1%)	29 (39.7%)	89/278 (32.0%)	112/289 (38.8%)
CC	38 (10.6%)	63 (6.6%)	43 (6.5%)	20 (6.8%)	28 (6.6%)	14 (6.2%)	14 (7.1%)	35 (6.9%)	29 (6.7%)	6 (8.2%)	29/278 (10.4%)	19/289 (6.6%)

Table 10-3 Detailed genotype analysis of *CCL20* promoter polymorphisms in Scottish and Swedish IBD.
Genotypes for Scottish IBD, CD, UC and HC in all patients, adult-onset (>17 years at diagnosis) and early-onset (<17 years at diagnosis) cohorts, Swedish UC and HC for **A.** 1706G→A and **B.** rs6749704T→C SNPs.

A. Outcome = IBD		**1706G→A**	**rs6749704T→C**
Number of Families		309	301
Number of Affected Offspring		303	296
Allele Transmission (Observed/ Expected)	Allele 1	529 / 531.1	435 / 438.0
	Allele2	77 / 74.9	157 / 154.0
Global Chi^2(1 df)		0.146	0.180
Global p-value for transmission		0.722	0.691

B. Outcome = CD		**1706G→A**	**rs6749704T→C**
Number of Families		309	301
Number of Affected Offspring		199	196
Allele Transmission (Observed/ Expected)	Allele 1	345 / 347.8	290 / 289.6
	Allele2	53 / 50.2	102 / 102.4
Global Chi^2 (1 df)		0.391	0.005
Global p-value for transmission		0.530	0.941

C. Outcome = UC		**1706G→A**	**rs6749704T→C**
Number of Families		309	301
Number of Affected Offspring		78	73
Allele Transmission (Observed/ Expected)	Allele 1	138 / 138.2	105 / 108.0
	Allele2	18 / 17.8	41 / 38.0
Global Chi^2 (1 df)		0.008	0.664
Global p-value for transmission		0.922	0.373

Table 10-4 TDT analysis of *CCL20* promoter polymorphisms in Scottish early-onset IBD cohort. Transmission disequilibrium testing for 1706 and rs6749704 SNPs in Scottish early-onset IBD families. Analysis carried out with TRANSMIT. The three different outcomes were defined as follows – **A.** IBD cases = all occurrences of CD, UC, indeterminate and non-specific colitis; **B.** CD cases = only definite CD cases: UC individuals were classed as being unaffected, while indeterminate/ non-specific diagnoses were classed as unknown; **C.** UC cases = only definite UC cases: CD individuals were classed as being unaffected, while indeterminate/ non-specific diagnoses were classed as unknown.

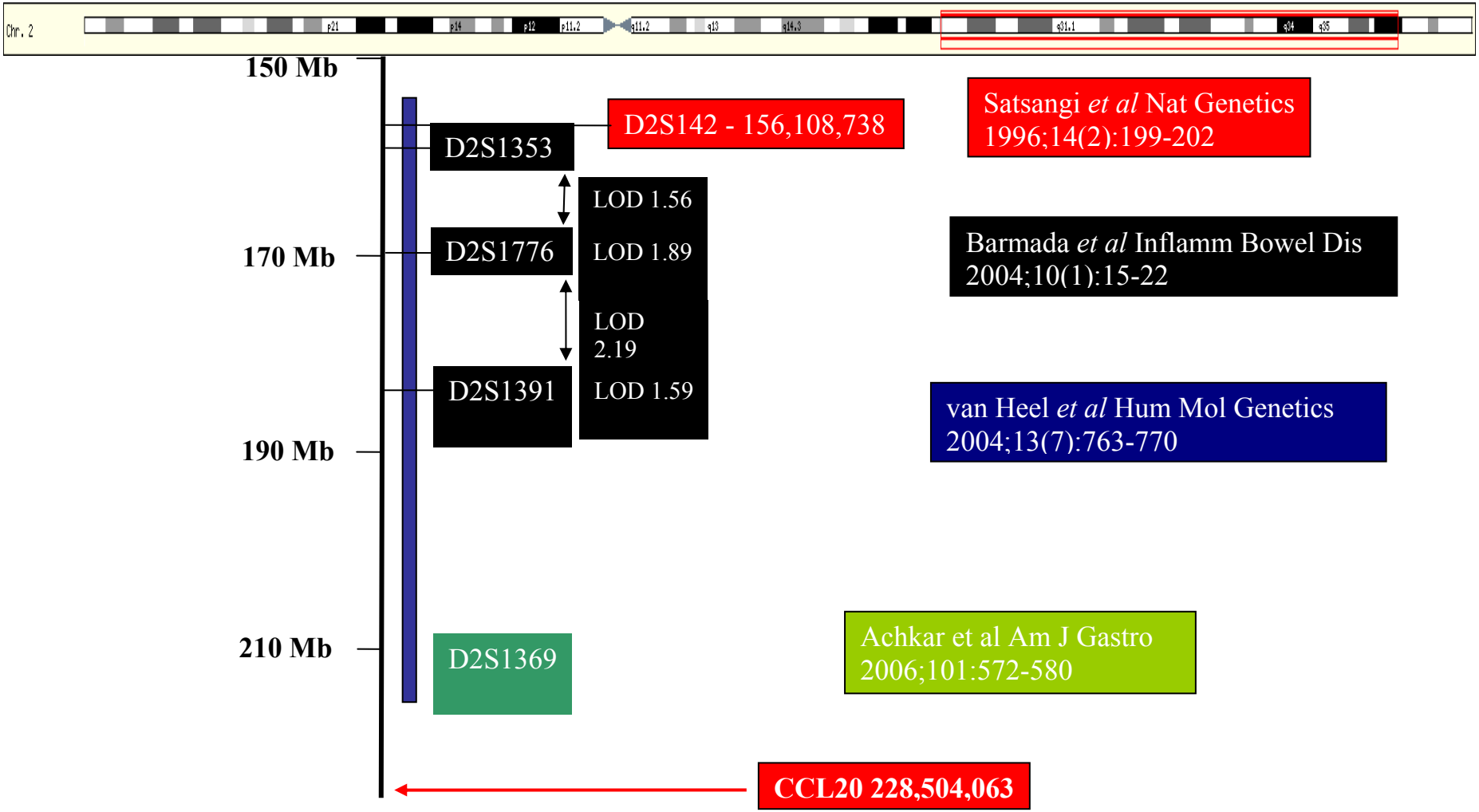

Figure 10-1 Evidence for linkage at chromosme 2q31.
Position of markers and relevant LOD scores from genome-wide linkage studies in UK[99] and US[150] populations, and from two different meta-analyses of linkage data.[97, 102] The genomic location of makers and *CCL20* are according to ensemble version 35.

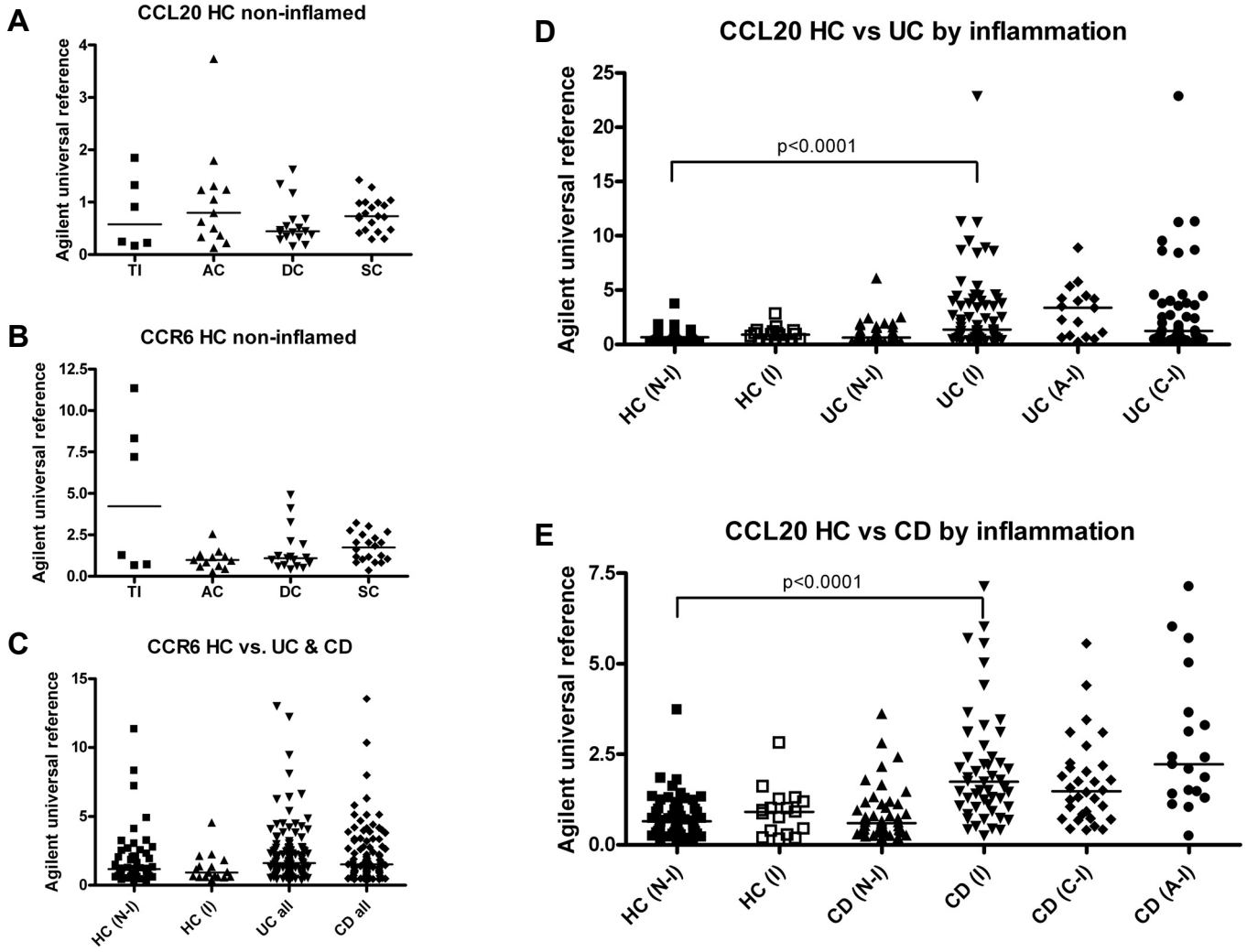

Figure 10-2 Microarray expression data for CCL20 and CCR6 in CD, UC and HC.

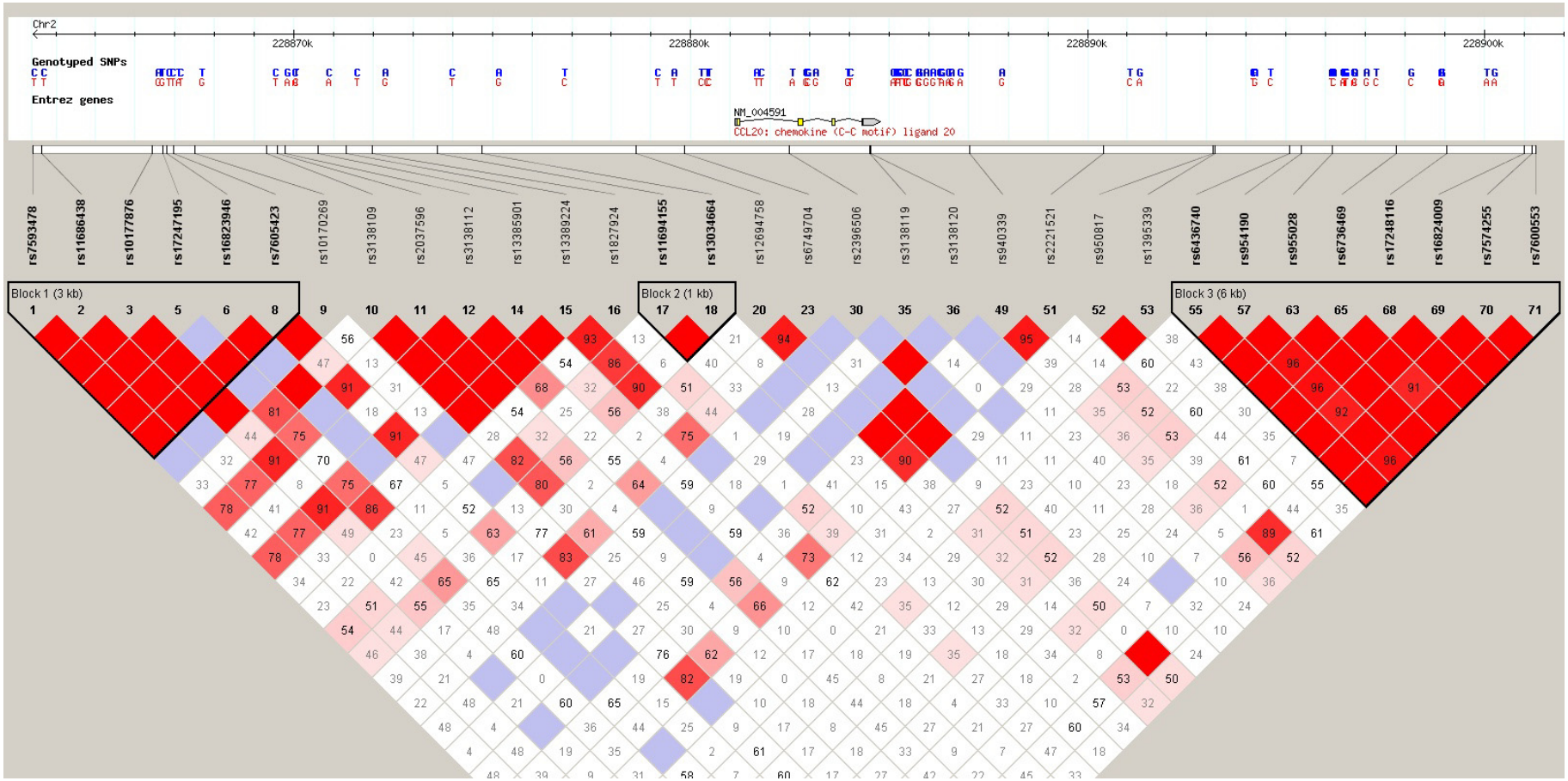

Figure 10-3 Haplotype structure of *CCL20*; 40kb view.
This map was generated using Haploview ver3.12 with phase II Hapmap data (based on NCBI B35 assembly) from 30 CEPH trios.

11 Haplotype-tagging SNP analysis of germ-line variation and re-sequencing of *NKX2.3*, an IBD susceptibility gene and target of HH signalling.

11.1 *Abstract*

Background. The homeodomain-containing gene *NKX2.3* (NK2 transcription factor related, locus 3) is a confirmed susceptibility gene for IBD (both CD and UC). The knock-out mouse has a lethal post-natal phenotype with abnormalities in both the intestine and spleen and significantly down-regulated MadCAM-1 expression, suggesting a critical role for NKX2.3 in lymphocyte homing. Furthermore, recent unpublished data indicate that NKX2.3 is a down-stream target of HH-GLI1 signalling.

Aims. The aims of the present study were to ascertain the contribution of *NKX2.3* germ-line variation to IBD susceptibility in the Scottish population, re-sequence the gene and determine whether adequate power existed in this cohort for a formal fine-mapping study.

Methods. The haplotype block containing *NKX2.3* (and no other genes) associated with CD on the WTCCC study was tagged by 7 tSNPs and genotyped in a Scottish population of adult and early-onset IBD consisting of 1386 patients (688 CD, 633 UC) and 363 HC. Log-likelihood, haplotype and individual tSNP chi^2 analysis was performed in this cohort. Additional analysis was performed on the 5 SNPs genotyped as part of the WTCCC study using a HC cohort comprising 3000 WTCCC controls in addition to the Scottish controls. 1kb of the promoter region along with the 2 major *NKX2.3* exons were re-sequenced in 96 CD patients. *In silico* expression analysis of microarray data was performed for NKX2.3 in colonic biopsies from CD and UC patients and HC.

Results. *NKX2.3* was not associated with IBD, CD or UC in the primary analysis in the Scottish adult-onset, early-onset or combined cohort versus the small panel of Scottish HC. However, on subsequent analysis with the addition of WTCCC HCs, *NKX2.3* was confirmed as an IBD susceptibility gene in Scotland with association noted at rs7095491, rs7081330 and rs1088371. The strongest association was at rs1088371 (IBD vs. HC: $p = 2.81 \times 10^{-5}$, OR 1.22). In the Scottish population, the association was notably stronger and more consistent for UC than CD. There was no clear association specifically with colonic IBD on genotype-phenotype analysis. On re-sequencing, 8 SNPs were detected within *NKX2.3* including one novel synonymous SNP. Expression analysis demonstrated up-regulation for one Agilent probe in CD regardless of inflammatory status, and for another probe with UC but only in inflamed

samples. The different expression profiles for these two probes was thought to correspond to an alternative shortened *NKX2.3* transcript missing exon 2.

Conclusion. *NKX2.3* is associated with IBD and UC and to a lesser extent CD in the Scottish population, but this cohort is underpowered for fine-mapping studies. *NKX2.3* is an excellent biological candidate for IBD and likely to play a major role in disease pathogenesis. As such, re-sequencing and subsequent fine-mapping has been prioritised as part of the WTCCC follow-up studies and these data are eagerly awaited. Finally, in the context of the other studies presented in this thesis, the links with HH-GLI1 signalling are very notable, and inform future studies.

11.2 *Introduction*

Although it was previously demonstrated that expression patterns of NKX2.3 do not alter in the *Shh*$^{-/-}$ mouse during embryonic development,[513] and that Shh and Ptch expression is unaltered in *Nkx2.3*$^{-/-}$ mice,[226] multiple lines of recent evidence suggest that Nkx2.3 is in fact a HH target gene. Firstly, it is notable that the expression pattern of NKX2.3 in the intestinal lamina propria (see Figure 3 in Wang *et al*, 2000)[226] mirrors that of GLI1 expression in human (**Table 4-2**) and mouse (Lees, Zacharias *et al*, 2008).[423] Secondly, microarray analysis of isolated murine gut mesenchyme stimulated with exogenous HH demonstrates robust upregulation of NKX2.3 gene expression (Zacharias W, personal communication).

As discussed in Chapter 1 (**1.5.2.5**), *NKX2.3* was identified in the WTCCC GWAS and subsequent replication studies as a novel CD susceptibility gene (OR ~1.18).[81, 89] More recently, *NKX2.3* has also been confirmed as a UC susceptibility gene in studies from the U.K. and Germany.[84, 85] *NKX2.3* is a small (3.6kb) gene at 10q24. Whilst it has two small exons, there appear to be two transcripts in the mouse at least;[225] the truncated transcript is predicted to lack exon 2.

The aim of the present study was therefore to describe the specific contribution of the WTCCC-associated haplotype block spanning *NKX2.3* to IBD susceptibility in the Scottish population, re-sequence the promoter and coding regions of the gene and then fine-map the associated region if suitably powered to do so in this cohort.

11.3 *Methods*

11.3.1 Genotyping

The extent of the association around *NKX2.3* from the WTCCC study was from 101264048 (rs10883359) to 101317841 (rs7091572) on chromosome 10 (**Table 11-1**). This maps to a 52.6kb haplotype block from 101261779 - 101314426 in the CEU HapMap (release 22/phase II April 07 on NCBI B36 assembly, dbSNP b126) population (**Figure 11-1 and 11-2**). tSNPs were derived in HapMap version 3.2 to tag SNPs in this block with MAF >0.1 and detecting haplotype frequencies of >5%. 6 tSNPs were identified: rs7095491, rs7081330, rs4919345, rs10786560, rs2902288 and

rs7091572. These 6 tSNPs plus rs10883371 (the second most significant WTCCC SNP in this block, but immediately upstream of the gene) were genotyped by TaqMan in both the Scottish adult and Scottish early-onset IBD populations and HCs. The former cohort has been described in Chapters 5, 7 and 10, the latter in Chapter 10.

11.3.2 Analysis

Log-likelihood analysis (EH/PM platform as described in Chapter 3), haplotype analysis (Haploview version 4.0) and chi^2 analysis on individual SNPs was performed for both cohorts of IBD, CD and UC patients separately and in combination. WTCCC data was available for 5/7 SNPs genotyped: rs7095491, rs7081330, rs10883371, rs2902288 and rs7091572. The summary genotype data was extracted for each of these SNPs. A combined genotype count was made between WTCCC HC and Scottish HC subjects and was analysed by chi^2 against the combined Scottish IBD cohort.

11.3.3 Sequencing

Primers were designed to sequence across the 2 *NKX2.3* exons and 1kb of the promoter region to cover all promoter and exonic SNPs listed in dbSNP. Sequencing was performed in 96 patients with CD from the main adult Scottish cohort. Further deep re-sequencing including all intronic regions is scheduled as part of the WTCCC follow-on studies (details of primer pairs in **Table 3-2**).

11.4 *Results*

11.4.1 Analysis of *NKX2.3* tSNPs in Scottish IBD and control populations

In the Scottish adult population no association was seen with IBD, CD or UC on log-likelihood testing (127 degrees of freedom, 1000 permutations, p=1.00), haplotype analysis (**Table 11-2**) or chi^2 testing of individual tSNPs (**Table 11-3**). Despite this lack of association, it was notable that a number of SNPs trended towards significance in the UC population, whereas the CD MAFs more closely mirrored the control population (**Table 11-3**). Comparison of the Scottish early-onset population with Scottish HC again demonstrated no evidence for association on either analysis (**Table 11-4**), and TDT analysis was negative (**Table 11-5**). A combined analysis of all

Scottish IBD, CD and UC patients (adult plus early-onset) demonstrated no association but a trend was again noted with UC (**Table 11-6**).

11.4.2 Analysis of *NKX2.3* tSNPs in Scottish IBD populations utilising WTCCC control data.

Comparison of allelic frequencies for 5 tSNPs in the present study with WTCCC data demonstrated broadly comparable MAFs for HC and CD populations. It was therefore postulated that the lack of association in the Scottish study was primarily due to insufficient power in this smaller population. The complete cohort of Scottish IBD patients (adult plus early-onset) was analysed for association against a combination of WTCCC and Scottish HC (**Table 11-7**). Strong evidence for association was noted for rs7095491, rs7081330 and rs1088371 in IBD and UC, with modest association for CD at rs7095491 and rs1088371. The most significantly associated SNP was rs1088371 (IBD vs. HC $p=2.81x10^{-5}$, OR 1.22, C.I. 1.11-1.33). The effect size was greater in UC than in CD (OR 1.28 and 1.17 respectively). There was no evidence for association at rs2902288 and rs7091572, despite both these SNPs achieving significance in the WTCCC study (p=0.0005 and p=0.0003).

11.4.3 Genotype-phenotype analysis

As a stronger association with *NKX2.3* was noted for UC than CD in the Scottish population, a genotype-phenotype analysis was performed at rs1088371 (the most significant tSNP) stratified for disease location. On analysis of all colonic IBD (UC plus CD colitis [L2]) versus the combined HC cohort, a similar association was noted as for UC alone (0.535 vs. 0.480, p=0.0001, OR 1.25, C.I. 1.12-1.40). There was no association with colonic CD (0.517 vs. 0.480, p=0.17, OR 1.16, C.I. 0.74-1.43) or ileal CD (L1 plus L4 plus L1+L4) (0.509 vs. 0.480, p=0.30, OR 1.13, C.I. 0.90-1.41). A within-cases analysis in CD for purely colonic disease (L2) versus ileal disease (L1 plus L4 plus L1+L4) demonstrated no significant difference between MAFs (0.517 vs. 0.509, p=0.85, OR 1.03, C.I. 0.76-1.39). No other genotype-phenotype associations were noted.

11.4.4 Re-sequencing of *NKX2.3*

Re-sequencing of *NKX2.3* in 96 patients with CD from the Scottish cohort confirmed the presence of 8 SNPs from dbSNP: rs11190141, rs10883371, rs10883372,

rs10883373, rs7908704, rs7893840, rs7911680 and rs888208 (**Table 11-8**). In addition, one novel synonymous SNP (C/A) was noted at position 101,283,025 (**Figure 11-3**). This SNP has subsequently been added to dbSNP128 and denoted rs41290504.

11.4.5 Expression analysis by microarray

Two probes from the Agilent Whole Genome Array microarray map to NKX2.3. A_23_P52425 (probe 'A') maps to exon 1 and A_24_P38702 (probe 'B') to exon 2. It is noteworthy that both EST evidence and the published literature identify a short transcriptional form that stops short of exon 2 (and is therefore not detected by probe B). Expression analysis demonstrated up-regulation of NKX2.3 (A) in CD compared with HC ($p=0.0003$), regardless of inflammatory status, but no difference between UC and HC (**Figure 11-4**). In contrast, NKX2.3 (B) was increased in inflamed UC samples versus HC ($p<0.0001$) but not non-inflamed UC ($p=0.17$) and only very minimally in CD ($p=0.04$).

11.5 *Discussion*

The present study describes further data on the association at *NKX2.3* in IBD, specifically in the Scottish population, re-sequencing of the promoter and 2 exons, and expression analysis from microarray data. The major issue highlighted is the lack of power inherent in the Scottish cohort. This is a major limiting factor in achieving the aim set out to fine-map the region. As such, further genotyping was not performed in this cohort. Rather, this must be performed in the larger cohorts described in the WTCCC GWAS and replication study for CD,[81, 89] and the UK consortium's UC studies.[84] Indeed, CD has been prioritised for re-sequencing and subsequent fine-mapping in the WTCCC and *NKX2.3* re-sequencing is underway as part of this first phase.

The combined Scottish cohorts of adult and early-onset IBD consists of a total of 1386 patients, comprising 688 patients with CD and 633 with UC, but only 363 HCss. Case-controlled analysis in this cohort did not achieve significance on any analysis, but trends were noted towards significance for a number of SNPs in IBD and UC. Similarly, the case-controlled and TDT analysis in the early-onset IBD population demonstrated lack of association, but this cohort is not powered to detect odds ratios

of <1.20 (most significant OR in WTCCC 1.18). Extracting the WTCCC HC data and combining this with the Scottish HC population confirmed that *NKX2.3* is associated with IBD, CD and UC in Scotland. Again, the effect was much greater for UC than for CD. A genotype-phenotype analysis stratified for colonic IBD (colonic CD plus UC) did not clearly demonstrate any added significance, but this may again be related to power. No genotype-phenotype data is currently available for *NKX2.3* as this was not presented in the WTCCC CD studies.[81, 89]

The re-sequencing data presented here has identified 9 SNPs within *NKX2.3* (or in the promoter or immediately downstream) that should be included in future fine-mapping studies (**Table 11-8**). The WTCCC re-sequencing data are eagerly awaited to determine if further variants should be added to this list. It is noteworthy that at present this list does not include any SNPs conferring a change in the amino acid structure of either of the *NKX2.3* transcripts. However, as recently demonstrated for *IRGM* (where a 20kb deletion immediately up-stream of the gene is not only associated with CD but also confers decreased IRGM function),[213] non-coding SNPs may be functional. All confirmed plus any novel *NKX2.3* SNPs will then be genotyped in the WTCCC CD and HC cohort (± replication cohorts) in an effort to identify the causal variant.

Following on from the next stage of the genetic studies, functional experiments will be a priority to ascertain the function of the causal SNP(s). Little is known about the expression and function of NKX2.3 in man. The expression data presented here demonstrate an effect in both CD and UC, but are complicated by the different expression profiles for the two probes present on the Agilent whole-genome chip. The two typical reasons for a gene's probes to offer distinct, significant expression profiles is cross-hybridization to another locus, which these don't, or hybridization to parts of the gene that are related differently from each other to the gene's alternative splicing, which is unlikely in the present situation but hard to disprove. A more intriguing possibility here is that these two probes are detecting the two different transcripts of Nkx2.3 first reported by Pabst and colleagues in the mouse in 1997.[225] This requires urgent evaluation by RT-PCR in a cohort genotyped for the *NKX2.3* variants as the alternative splicing required to truncate the protein short of exon 2

could go someway to explaining the underlying mechanism of *NKX2.3* mutation in IBD.

There is, however, presently a large body of published literature documenting the role of Nkx2.3 in the developing murine intestine and spleen with a striking phenotype noted for the knock-out mouse, as described above. This animal model deserves re-evaluation in the light of the association of *NKX2.3* with CD and UC before a future model is generated based on cloning the IBD-associated mutation(s) into new mice.

The apparently normal phenotype of the heterozygous *Nkx2.3* mouse has striking parallels to the $Gli^{+/LacZ}$ mice studied.[423] It would therefore seem logical to stress these animals with DSS to firstly ascertain their susceptibility to chemically-induced colitis and secondly use the LacZ reporter to describe the Nkx2.3 positive populations of mesenchymal signals responding to the acute inflammatory insult. Little is known about the upstream regulation of Nkx2.3 so it is of great interest that its expression appears to be regulated by HH in the intestine. Given the associations for both *GLI1* and *NKX2.3* described with IBD, a detailed assessment of the cross-talk of both these pathways in both intestinal homeostasis and inflammation is an urgent priority for follow-up.

In conclusion, *NKX2.3* is associated with IBD, UC and to a lesser extent CD in the Scottish population but this cohort is under-powered to allow further fine-mapping of the region with the ultimate aim of identifying the causative mutation(s) in IBD. This work is underway as part of the WTCCC follow-up studies and the data are eagerly awaited. There are many intriguing aspects to Nkx2.3 expression and function in the mouse that make an important role in IBD pathogenesis seem very plausible indeed, not least its regulation of MadCAM-1 expression. Finally, there is much to suggest a complex, but direct interaction between NKX2.3 and HH/GLI1 signalling in the response to acute inflammatory challenge in mouse that will form the major basis of future studies in man and mouse.

WTCCC SNP	Position			CD	HC	p-value	OR	95% lower	95% upper
rs4595481	101254152	C	T	72.6%	74.0%	0.16	0.93	0.85	1.03
rs10883359	101264023	C	T	25.7%	28.6%	0.0022	0.86	0.78	0.95
rs7095491	101264048	A	G	52.7%	47.0%	5.79×10^{-08}	1.26	1.16	1.37
rs7078219	101264355	A	G	33.9%	38.7%	2.85×10^{-06}	0.81	0.74	0.89
rs7081330	101264455	A	G	66.1%	61.2%	2.27×10^{-06}	1.24	1.13	1.35
rs10883365	101277754	C	T	53.7%	47.7%	1.65×10^{-08}	1.27	1.17	1.38
rs11190137	101277934	G	T	25.4%	28.6%	0.00072	0.85	0.77	0.93
rs10883367	101277980	A	G	53.7%	47.7%	2.66×10^{-08}	1.27	1.17	1.38
rs11190139	101278089	C	G	75.4%	72.2%	0.00090	1.18	1.07	1.29
rs7899176	101279113	C	T	21.5%	24.5%	0.0011	0.85	0.77	0.94
rs1548964	101279643	C	G	46.7%	52.7%	2.21×10^{-08}	0.79	0.72	0.85
rs1548962	101279725	C	G	46.3%	52.3%	1.82×10^{-08}	0.79	0.72	0.85
rs6584283	101280291	A	G	52.0%	46.2%	4.35×10^{-08}	1.26	1.16	1.37
rs10883371	101282445	G	T	53.7%	47.7%	1.61×10^{-08}	1.27	1.17	1.38
rs884144	101283846	A	G	13.3%	14.3%	0.19	0.92	0.82	1.04
rs7903232	101288792	C	G	86.8%	85.7%	0.15	1.09	0.97	1.23
rs7901565	101300855	C	T	100.0%	100.0%	0.71	0.59	0.04	9.51
rs989979	101312413	C	G	59.2%	55.0%	7.80×10^{-05}	1.19	1.09	1.29
rs2902288	101314355	C	G	10.5%	13.0%	0.00055	0.79	0.69	0.90
rs1332102	101314426	A	T	77.9%	73.1%	3.45×10^{-07}	1.29	1.17	1.43
rs1360522	101315787	A	G	23.7%	27.5%	5.32×10^{-05}	0.82	0.74	0.90
rs11190152	101315905	C	T	2.1%	2.1%	1.00	1.00	0.75	1.35
rs7091572	101317841	C	T	24.5%	27.9%	0.00030	0.84	0.76	0.92
rs12268645	101318008	G	T	71.3%	70.9%	0.66	1.02	0.93	1.12

Table 11-1 WTCCC SNPs associated with CD around *NKX2.3*.
The limits of the associated region are denoted by the solid black line (101264048 – 101317841). SNPs genotyped in the Scottish population are highlighted in yellow. The most significant SNP in the WTCCC study was rs10883371, and on the GWAS meta-analysis rs7078219.

	IBD (adult)	CD (adult)	UC (adult)	HC	IBD vs. HC p-value	CD vs. HC p-value	UC vs. HC p-value
1111111	0.280	0.276	0.284	0.275	0.78	0.90	0.69
1111211	0.190	0.176	0.200	0.168	0.19	0.65	0.08
2222112	0.124	0.126	0.122	0.119	0.74	0.67	0.84
2221111	0.110	0.119	0.103	0.123	0.35	0.80	0.19
2122111	0.099	0.092	0.106	0.105	0.67	0.42	0.93
2222122	0.074	0.082	0.067	0.079	0.64	0.85	0.33

Table 11-2 Haplotype structure of *NKX2.3* tSNPs.
The frequencies of the 6 haplotypes made by the 7 *NKX2.3* tSNPs are demonstrated for the adult IBD population and Scottish HCs. P-values for association were generated in Haploview version 4.2.

	IBD (adult) MAF	**CD (adult) MAF**	**UC (adult) MAF**	**HC MAF**	**IBD vs. HC** p-value OR (C.I.)	**CD vs. HC** p-value OR (C.I.)	**UC vs. HC** p-value OR (C.I.)
rs7095491	0.517	0.502	0.529	0.494	0.30 OR 1.10 (0.92-1.31)	0.75 OR 1.03 (0.84-1.37)	0.16 OR 1.15 (0.95-1.40)
rs7081330	0.348	0.368	0.332	0.368	0.35 OR 0.92 (0.76-1.10)	0.98 OR 1.00 (0.81-1.20)	0.13 OR 0.85 (0.70-1.05)
rs1088371	0.472	0.489	0.458	0.494	0.33 OR 1.09 (0.91-1.30)	0.84 OR 1.02 (0.83-1.25)	0.15 OR 1.15 (0.95-1.41)
rs4919345	0.367	0.374	0.362	0.366	0.95 OR 1.01 (0.84-1.21)	0.74 OR 1.04 (0.84-1.28)	0.86 OR 0.98 (0.80-1.20)
rs10786560	0.235	0.217	0.249	0.222	0.50 OR 1.08 (0.87-1.33)	0.81 OR 0.97 (0.76-1.24)	0.19 OR 1.17 (0.92-1.47)
rs2902288	0.122	0.141	0.106	0.129	0.63 OR 0.94 (0.72-1.22)	0.50 OR 1.11 (0.82-1.49)	0.15 OR 0.80 (0.59-1.08)
rs7091572	0.268	0.286	0.254	0.277	0.65 OR 0.96 (0.78-1.16)	0.71 OR 1.04 (0.83-1.31)	0.28 OR 0.89 (0.71-1.11)

Table 11-3 *NKX2.3* MAFs in adult-onset Scottish IBD population and HCs.
P-values, odds ratios (OR) and 95% confidence intervals are given for chi^2 analysis of IBD, CD and UC versus HC.

	IBD (paed.) MAF	**CD (paed.) MAF**	**UC (paed.) MAF**	**HC MAF**	**IBD (paed.) vs. HC** p-value OR (C.I.)	**CD (paed.) vs. HC** p-value OR (C.I.)	**UC (paed.) vs. HC** p-value OR (C.I.)
rs7095491	0.525	0.523	0.532	0.494	0.24 OR 1.13 (0.92-1.40)	0.35 OR 1.12 (0.88-1.42)	0.36 OR 1.16 (0.84-1.61)
rs7081330	0.340	0.351	0.314	0.368	0.26 OR 0.88 (0.71-1.10)	0.56 OR 0.93 (0.72-1.19)	0.17 OR 0.78 (0.561.11)
rs1088371	0.531	0.532	0.538	0.506	0.35 OR 1.11 (0.90-1.36)	0.40 OR 1.11 (0.87-1.41)	0.44 OR 1.14 (0.82-1.58)
rs4919345	0.359	0.360	0.362	0.366	0.77 OR 0.97 (0.78-1.20)	0.83 OR 0.97 (0.76-1.25)	0.91 OR 0.98 (0.70-1.37)
rs10786560	0.233	0.247	0.242	0.222	0.60 OR 1.07 (0.83-1.37)	0.33 OR 1.15 (0.87-1.52)	0.56 OR 1.12 (0.76-1.64)
rs2902288	0.135	0.143	0.141	0.129	0.73 OR 1.06 (0.78-1.43)	0.49 OR 1.13 (0.80-1.59)	0.66 OR 1.11 (0.69-1.78)
rs7091572	0.243	0.248	0.250	0.277	0.14 OR 0.84 (0.66-1.06)	0.27 OR 0.86 (0.65-1.13)	0.46 OR 0.87 (0.60-1.26)

Table 11-4 *NKX2.3* MAFs in early-onset Scottish IBD population (paed.) and HCs.
P-values, odds ratios (OR) and 95% confidence intervals are given for chi^2 analysis of IBD, CD and UC versus HC.

SNP	Transmitted	Untransmitted	p-value
rs7095491	111:99	99	0.4076
rs7081330	113:106	106	0.6362
rs1088371	104:97	97	0.6215
rs4919345	102:88	88	0.3098

Table 11-5 TDT analysis of *NKX2.3* tSNPs in early-onset Scottish IBD population.
TDT analysis was performing using the standard algorithm embedded in Haploview (version 4.2).
tSNPs 5,6&7 have mendelian errors (4, 2 and 3 respectively)

	IBD (all) MAF	**CD (all) MAF**	**UC (all) MAF**	**HC MAF**	**IBD (all) vs. HC** p-value OR (C.I.)	**CD (all) vs. HC** p-value OR (C.I.)	**UC (all) vs. HC** p-value OR (C.I.)
rs7095491	0.520	0.510	0.530	0.494	0.24 OR 1.11 (0.94-1.31)	0.52 OR 1.06 (0.88-1.28)	0.14 OR 1.15 (0.96-1.39)
rs7081330	0.346	0.362	0.329	0.368	0.27 OR 0.91 (0.76-1.08)	0.77 OR 0.97 (0.80-1.18)	0.09 OR 0.84 (0.69-1.03)
rs1088371	0.529	0.518	0.541	0.506	0.29 OR 1.10 (0.94-1.30)	0.60 OR 1.05 (0.87-1.27)	0.15 OR 1.15 (0.95-1.39)
rs4919345	0.365	0.369	0.362	0.366	0.95 OR 0.99 (0.83-1.19)	0.89 OR 1.01 (0.83-1.23)	0.85 OR 0.98 (0.81-1.20)
rs10786560	0.234	0.227	0.248	0.222	0.50 OR 1.07 (0.88-1.32)	0.78 OR 1.03 (0.82-1.29)	0.20 OR 1.16 (0.96-1.45)
rs2902288	0.126	0.142	0.111	0.129	0.82 OR 0.97 (0.75-1.25)	0.43 OR 1.12 (0.85-1.47)	0.26 OR 0.85 (0.64-1.13)
rs7091572	0.261	0.272	0.253	0.277	0.39 OR 0.92 (0.76-1.11)	0.82 OR 0.98 (0.79-1.20)	0.25 OR 0.88 (0.71-1.09)

Table 11-6 *NKX2.3* MAFs in combined Scottish IBD population (adult plus early-onset) and HCs.
P-values, odds ratios (OR) and 95% confidence intervals are given for chi^2 analysis of IBD, CD and UC versus HC.

	IBD (ALL) MAF	**CD (ALL) MAF**	**UC (ALL) MAF**	**ALL HC MAF**	**IBD (ALL) vs. HC (ALL)** p-value OR (C.I.)	**CD (ALL) vs. HC (ALL)** p-value OR (C.I.)	**UC (ALL) vs. HC (ALL)** p-value OR (C.I.)
rs7095491	0.520	0.510	0.530	0.472	4.59×10^{-5} OR 1.21 (1.10-1.33)	0.015 OR 1.16 (1.03-1.31)	0.0003 OR 1.26 (1.11-1.43)
rs7081330	0.346	0.362	0.329	0.386	0.0004 OR 0.84 (0.77-0.93)	0.11 OR 0.90 (0.80-1.02)	0.0002 OR 0.78 (0.69-0.89)
rs1088371	0.529	0.518	0.541	0.480	2.81×10^{-5} OR 1.22 (1.11-1.33)	0.012 OR 1.17 (1.03-1.32)	0.0001 OR 1.28 (1.13-1.45)
rs2902288	0.126	0.142	0.111	0.130	0.60 OR 0.96 (0.84-1.11)	0.25 OR 1.11 (0.93-1.32)	0.082 OR 0.84 (0.69-1.02)
rs7091572	0.261	0.272	0.253	0.279	0.079 OR 0.91 (0.82-1.01)	0.62 OR 0.97 (0.84-1.11)	0.065 OR 0.87 (0.76-1.01)

Table 11-7 *NKX2.3* MAFs in combined Scottish IBD population (adult plus early-onset) and combined HC (Scotland and WTCCC).
WTCCC data were available for 5 of the 7 tSNPs genotyped. P-values, odds ratios (OR) and 95% confidence intervals are given for chi^2 analysis of IBD, CD and UC versus HC.

SNP	Position	Alleles	Type	WTCCC SNP	Validated on re-sequencing
rs11190141	101,282,382	**C/T**	Promoter	N	Y
rs10883371	101,282,445	**A/C**	Promoter	Y	Y
rs10883372	101,282,473	**T/A**	Promoter	N	Y
rs10883373	101,282,474	**G/A**	Promoter	N	Y
rs41290504	101,283,025	**C/A**	Synonymous (Exon1)	N	Y
rs11340647	101,283,246	**G/-**	Ins/Del	N	N
rs7908704	101,283,308	**G/A**	Intronic	N	Y
rs7893840	101,283,331	**T/C**	Intronic	N	Y
rs7911680	101,283,458	**A/C**	Intronic	N	Y
rs884144	101,283,846	**C/T**	Intronic	Y	N
rs12247368	101,284,620	**A/C**	Intronic	N	N
rs10082511	101,285,216	**C/A**	Synonymous (Exon 2)	N	N
rs888208	101,285,853	**A/G**	Downstream	N	Y

Table 11-8 Position and location of all known SNPs within *NKX2.3*.
Includes 1kb of promoter region. Those SNPs validated on re-sequencing indicated. rs41290504 was a novel SNP at time of sequencing (2007) but is now listed on dbSNP128 (ENSEMBL:Watson). Two of these SNPs were genotyped in the WTCCC study (rs1088371 and rs884144).

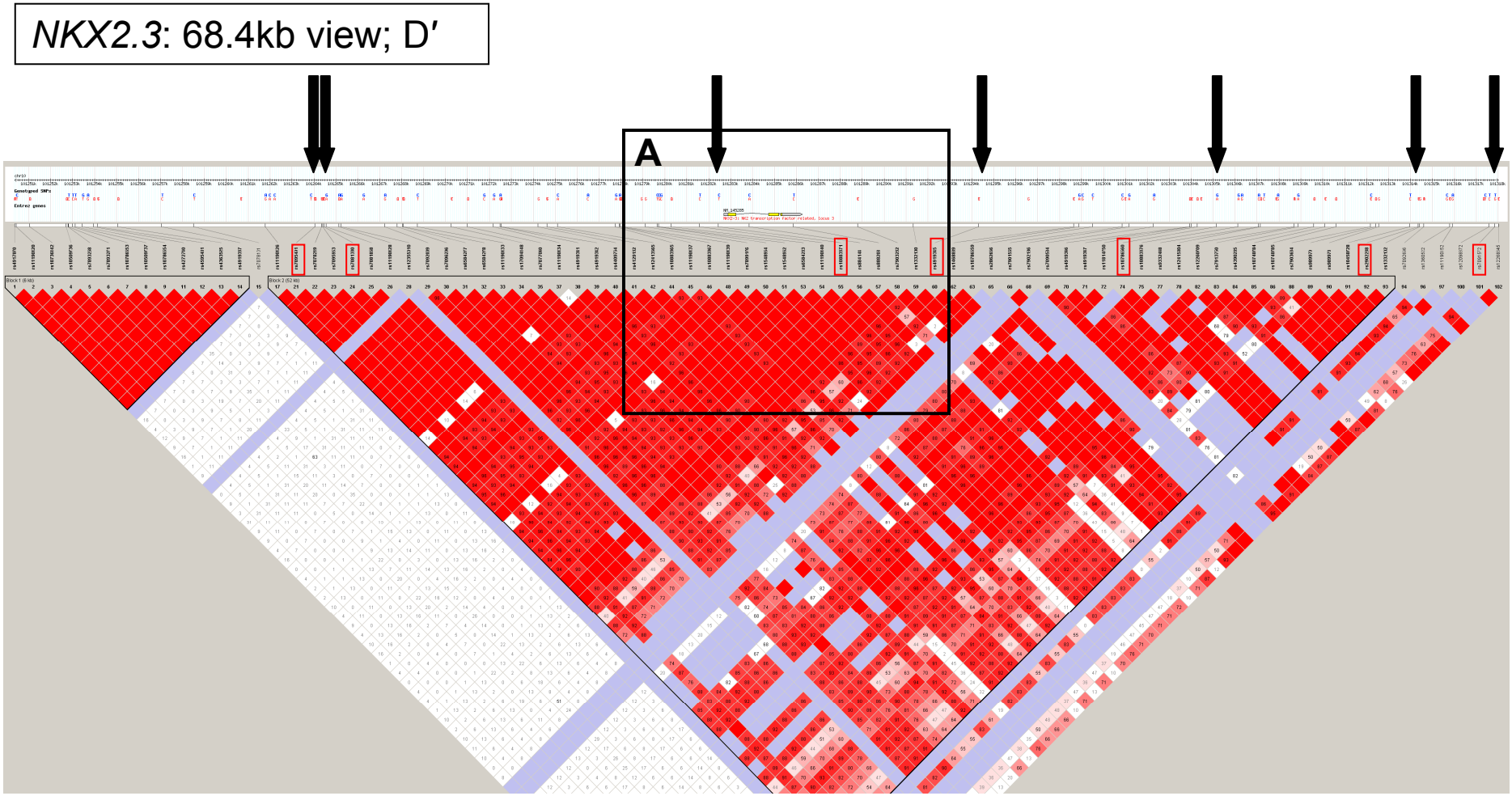

Figure 11-1 Haplotype structure of *NKX2.3* (68.4 view, D′).
HapMap data visualised on HaploView (version 4.2). 68.4kb region spanning *NKX2.3* including position of tSNPs.

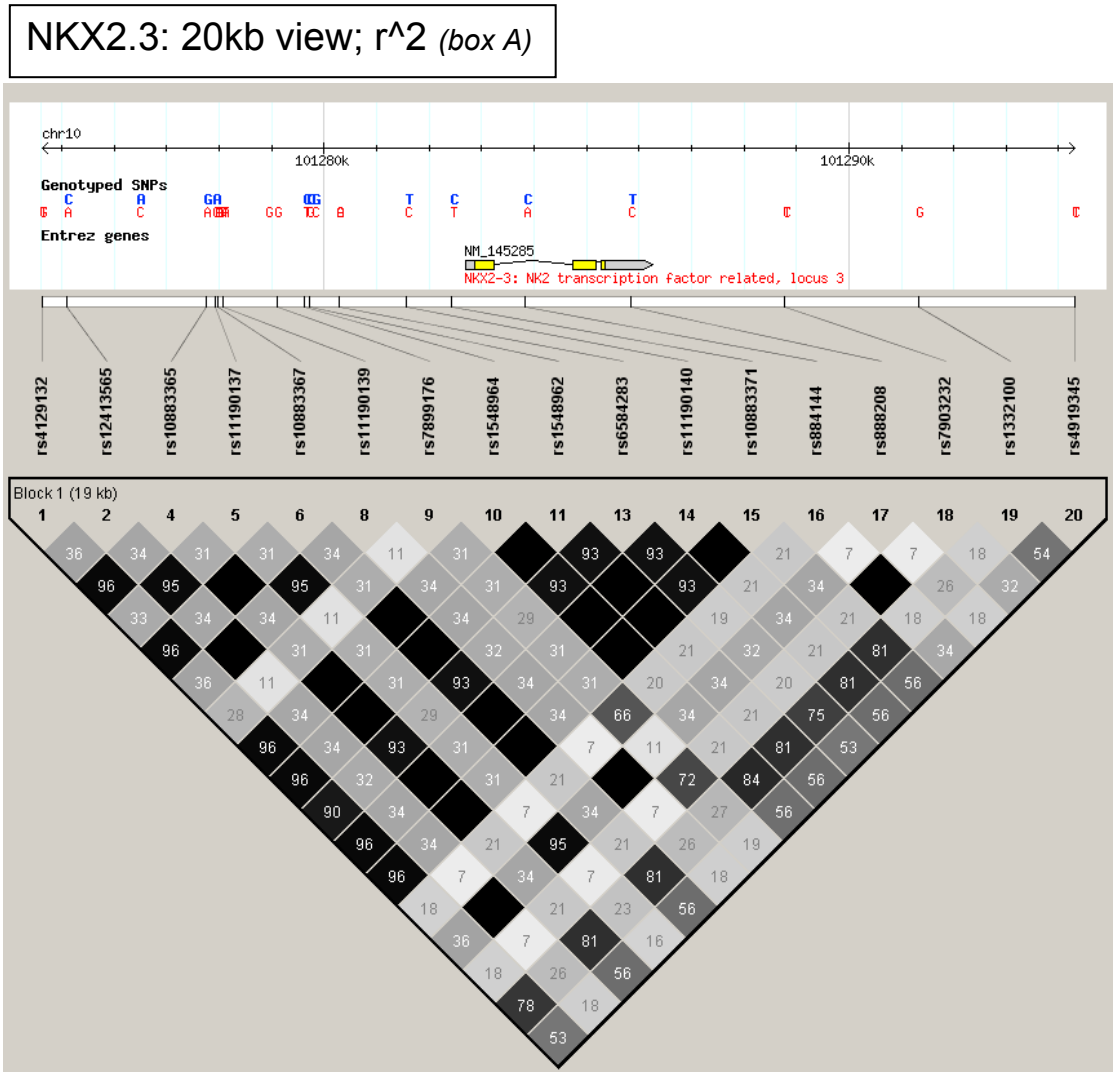

Figure 11-2 Haplotype structure of *NKX2.3* (42.6kb view, r^2).
42.6kb segment (box A in Figure 1A) with LD structure denoted by r^2 values demonstrates decreased LD around *NKX2.3*

101282880
TGATGTTACCAAGCCCGGTCACCTCCACCCCTTTCTCAGTCAAAGACATTTTGAATCTGG
AGCAGCAGCACCAGCACTTCCATGGTGCGCACTTGCAGGCGGACTTGGAGCACCACTTCC
ACTCTGCGCCCTGCATGCTGGCCGCC(C/A)GCTGAGGGGACGCAATTTTCTGACGGAGGGGAGG
AGGACGAGGAAGACGAGGGCGAGAAATTGTCCTATTTGAACTCACTAGCCGCAGCAGACG
GCCACGGGGATTCAGGGCTGTGTCCCCAGGGCTATGTCCACACGGTCCTGCGAGACTCGT
Position Chr 10: 101,283,025 C/A

Figure 11-3 Position of novel synonymous SNP from re-sequencing of *NKX2.3*.

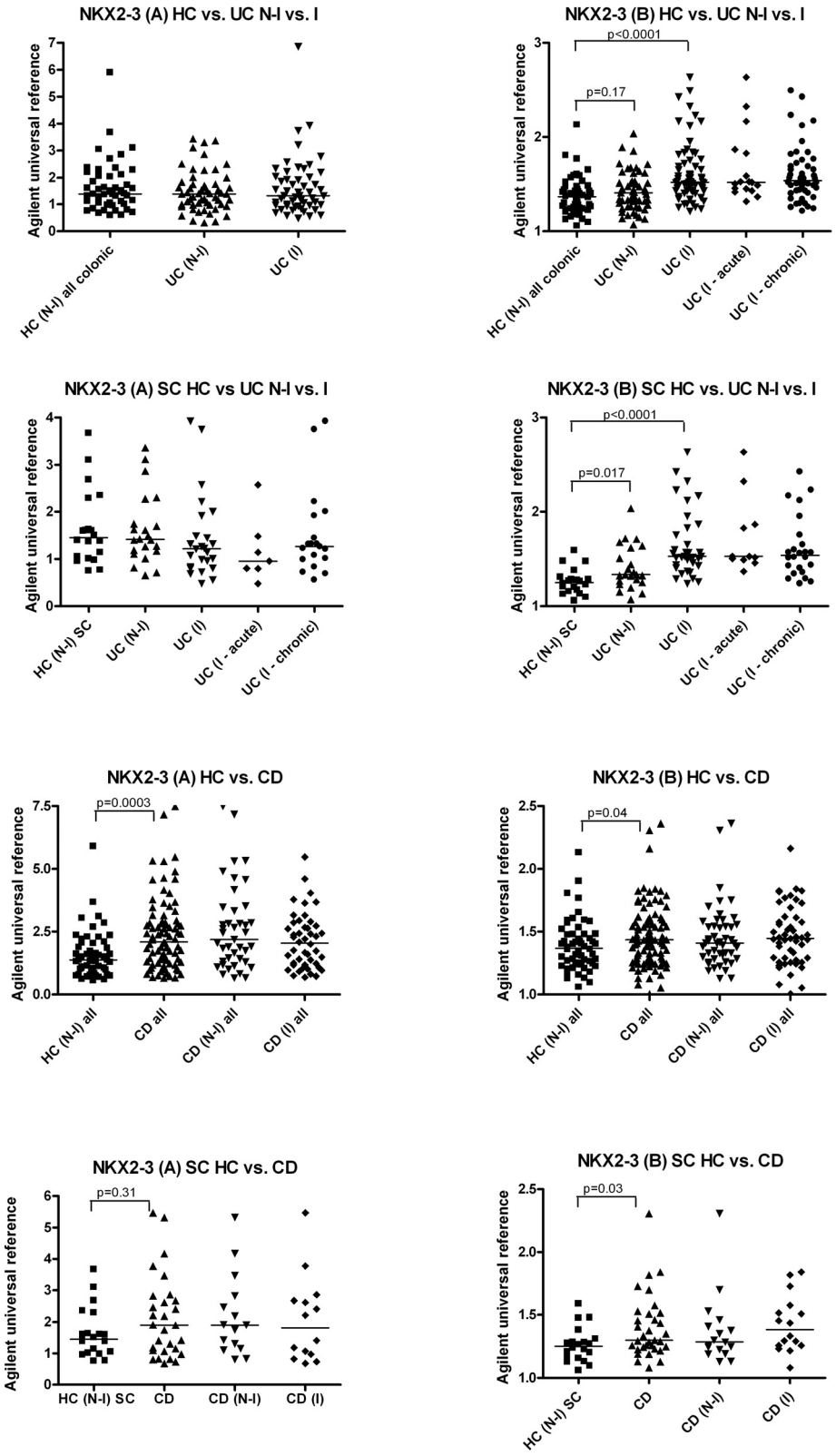

Figure 11-4 mRNA expression of *NKX2.3* in UC and CD by microarray.
Quantitative analysis of mRNA levels by microarray of *NKX2.3* in UC and CD compared with non-inflamed HC. For UC and CD, sub-categorisation is made for inflamed (I) and non-inflamed (N-I) samples. Individual datapoints are plotted with horizontal lines representing medians of each dataset. P-values, plotted where significant, are derived by Mann-Whitney U test.

12 Implications and future work

12.1 *Implications*

In the past 2 years, I have initiated a series of collaborative studies with William Zacharias and Professor Deborah Gumucio (Department of Cell and Developmental Biology, Ann Arbor, MI). Exploiting their expertise in gastrointestinal development and animal modelling, we have further explored the role of HH in IBD pathogenesis using the *Gli1*$^{+/LacZ}$ mouse. Whilst these experiments were performed in Ann Arbor, I conceived the hypotheses and study aims, and all aspects of study design and analysis were worked on in conjunction with Zacharias and Gumucio. The results of these studies, recently published (Lees, Zacharias *et al*, *PLoS Medicine* 2008; attached as appendix),[423] will be discussed in conjunction with the body of work presented in this thesis when considering the important implications herein. This lays the groundwork for the exciting future studies planned in Edinburgh and Ann Arbor.

As discussed in Chapter 2, multiple lines of evidence now indicate that HH plays an important role in inflammation. The data presented in the present thesis, along with recent data generated with the Gumucio lab now present a compelling argument that down-regulation of Hh ligand in the intestinal epithelium, with associated decreased pathway activity in the lamina propria, is associated with a pro-inflammatory state:

- Mice engineered in the Gumucio lab to have reduced intestinal Hh signalling (over-expression of the pan-Hh inhibitor Hhip [Villin-Hhip]) have spontaneous ileal inflammation[303]
- Microarray studies performed in Villin-Hhip ileum[303] (**Table 13-1**) and in isolated ileal mesenchyme treated with recombinant Hh demonstrate:
    - CX3CR1, LTC4S, CD28 and NKX2.3 are all increased by recombinant Hh and decreased in villin-Hhip;
    - IL1β, IL6, TLR2, MARCO and CXCL2 are decreased by recombinant Hh and increased in villin-Hhip.
- ***GLI1 expression is decreased in colonic inflammation.*** GLI1 mRNA (and by inference Hh pathway activity) increased along the length of the healthy adult colon in man (mirroring the increasing luminal bacterial load in the distal colon) and was reduced in all forms of colonic inflammation studied (UC, colonic CD and non-IBD inflammation).

- ***Germ-line variation in GLI1* (within IBD2 on 12q13) *is associated with IBD.*** This was demonstrated in 3 independent N European populations (Scotland, England, Sweden).
    - A non-synonymous SNP (rs2228226) in a highly conserved region of *GLI1*, next to a known transactivation domain, is associated with IBD (pooled odds ratio 1.194, C.I. 1.09-1.31, p=0.0002) and is functionally defective *in vitro*.
- ***Gli1*$^{+/LacZ}$ *develop early, severe colitis*.[423]** *Gli1*$^{-/-}$ mice develop normally, in contrast to lines deficient in all other Hh components. We have shown, in collaboration with the Gumucio lab, that stressing mice with a 50% reduction in functional Gli1 (*Gli1*$^{+/LacZ}$) with 3% DSS for 6 days leads to early, severe colitis with substantial mortality compared to WT littermates. To date, this is the only described phenotype for reduced Gli1 function *in vivo*.
- ***Myeloid CD11b and CD11c positive cells are direct targets of Hh signalling in the lamina propria.*[423]** The poor specificity of commercially available Hh pathway antibodies is now well documented.[334] Given that *Gli1*$^{+/LacZ}$ animals contain a copy of bacterial β-galactosidase under control of the Gli1 promoter, they represent an excellent model to identify cell types that are Hh-responsive prior to and during inflammatory disease. Initial analysis has demonstrated that whilst CD3 lymphocytes, and CD11b and CD11c myeloid cells co-localize LacZ in the steady state, only myeloid APCs are responding to Hh signals in response to acute DSS challenge. This was accompanied by significant up-regulation of IL12, IL17 and IL23 on analysis of whole intestinal mRNA. No co-localisation was seen with CD19-positive B lymphocytes at any stage.

12.2 *Current working hypothesis*

On the basis of this body of evidence, I have generated the hypothesis that *paracrine Hh signalling from epithelium to myeloid antigen presenting cells (APCs) is critical to the maintenance of intestinal mucosal immune homeostasis (**Figure 12-1**). In the steady state, soluble Hh protein in the intestinal epithelium instructs resident myeloid APCs to adopt a tolerogenic phenotype, which in turn leads to expansion of the regulatory T cell pool. In response to inflammatory challenge, down-regulated epithelial Hh, acting as a danger signal, results in a pro-inflammatory myeloid cell*

phenotype and expansion of Th17 lymphocytes in the lamina propria. Dysregulation of this homeostatic hedgehog signalling can result in IBD.

Planned future studies will now aim to:

1. Characterise the aberrant inflammatory signalling in lamina propria myeloid APCs downstream of the Hh signal in *Gli1*$^{+/LacZ}$ and *Gli1*$^{-/-}$ mice;
2. Elucidate the upstream regulators of epithelial Indian (Ihh) and Sonic hedgehog (Shh) gene expression; and
3. Examine whether Hh agonists represent a novel therapeutic modality in IBD.

12.3 *Future HH studies*

12.3.1 Characterise aberrant inflammatory signalling in *Gli1*$^{-/-}$ mice

1a) DSS challenge of *Gli1*$^{-/-}$ mice. I will first stress *Gli1*$^{-/-}$ mice and WT littermate controls (bred for at least 3 generations onto a C57BL/6 background) with 3% DSS administered in drinking water for 1-6 days. Animals will be monitored daily for signs of diarrhoea, bloody stool and weight loss. To establish the early mechanisms in this acute model of colonic inflammation, mice will be sacrificed at days 1, 2, 3 and 6 after initiation of DSS.

1b) Molecular and cellular targets of Hh. mRNA analysis and immunohistochemistry for Gli1 LacZ will be performed on whole colonic tissue sections from *Gli1*$^{-/-}$ mice and WT at baseline and at various timepoints after inflammatory challenge. Q-PCR will be performed for a variety of cytokines (incl. IL1β, IL4, IL6, IL8, IL10, IL12, IL17, IL23, IFNγ, & TNFα), chemokines (incl. CX3CR1, CCL20), murine cryptdins, TGFβ1, NKX2.3, and NOD2. To extend the LacZ observations in *Gli1*$^{+/LacZ}$ mice 4 days after DSS, I will characterise the cellular targets of Hh signalling early (days 1, 2 and 3) in *Gli1*$^{-/-}$ mice with antibodies to CD3, CD11b, CD11c, CD103 and MHCII (co-staining with β-galactosidase).

1c) *In vitro* characterisation of myeloid cell phenotypes. Whist the experiments already performed in *Gli1*$^{+/LacZ}$ mice demonstrate lamina propria myeloid cells responding to epithelial Hh signals and up-regulation of Th17 cytokines in total colonic mRNA, they do not directly connect these two observations. It is therefore not presently clear which cell type is responsible for the altered cytokine profile and precisely how defective Gli1 expression contributes to this. I will therefore isolate colonic lamina propria DCs and macrophages from *Gli1*$^{-/-}$ and WT mice at baseline

and 1, 2 and 3 days following DSS challenge. The cytokine profiles (IL6, IL10, IL23, IFNγ, CX3CR1, TGFβ) of these cells will be determined by ELISA in culture supernatant.

1d) Analysis of regulatory T cell (T_{REG}) populations. It is anticipated that the *in vitro* experiments in 1c will demonstrate increased secretion of pro-Th17 cytokines in *Gli1*$^{-/-}$ DCs and macrophages. In my working hypothesis the ultimate effect of these Hh conditioned myeloid cells is the determination of CD4 T cell phenotype (**Figure 12-1**). In order to ascertain the relative T_{REG} subset in the lamina propria of *Gli1*$^{-/-}$ mice, I will perform immunohistochemistry for Foxp3 and quantify relative to number of CD3 +ve cells. To further characterise the T_{REG} population in these animals I will isolate immune cells from spleen, mLNs and colon and measure T_{REG} subset by FACS (CD3+, FoxP3+), both with and without prior sorting for LacZ status and examine this for signs of Hh activity. If these initial approaches demonstrate a difference in the T_{REG} population, I will perform co-culture experiments to formally test this aspect of the hypothesis. Lamina propria DCs and/or macrophages (depending on the results from 1c) from WT and *Gli1*$^{-/-}$ mice will be co-cultured with naïve CD4 T cells to see if they are instructed to a Th17 or T_{REG} phenotype.

1e) *Il10*$^{-/-}$ mice. I will first establish the level of Hh signalling in *Il10*$^{-/-}$ mice in SPF conditions by Q-PCR and in situ hybridisation for Gli1 at different timepoints. We will then inter-breed *Il10*$^{-/-}$ and *Gli1*$^{-/-}$ mice and utilise clinical and histological parameters (as above) to determine whether *Il10*$^{-/-}$*;Gli1*$^{+/-}$ mice develop more severe colitis when compared with *Il10*$^{-/-}$ littermates. Furthermore, the role of the Hh signalling pathway will be established in this model by elucidating cellular (LacZ immunostaining) and molecular (Q-PCR) targets of Hh as outlined for the DSS experiments above.

12.3.2 Elucidation of upstream regulators of HH gene expression in the epithelium

Shh and Ihh ligands are exclusively expressed in the epithelium in the intestine. Very little remains known about the upstream regulation of ligand expression. I plan to address this using Shh & Ihh reporter mice (Shh-CRE-GFP and Ihh-LacZ) in 3 settings: i) response to DSS inflammatory challenge ii) antibiotic treatment to obliterate commensal bacterial flora and iii) following obliteration of intestinal epithelial cell NFκB expression. Mice heterozygous for the Shh-GFP insertion allele

(*Shh*$^{+/GFP}$) will be used as they develop normally, are available on B6 background and are indistinguishable from WT littermates.

2a) DSS challenge of *Shh*$^{+/GFP}$ and *Ihh*$^{+/LacZ}$ mice. 3% DSS will be administered in the drinking water to *Shh*$^{+/GFP}$, *Ihh*$^{+/LacZ}$ and WT mice, and compared with control groups of both reporter strains and WT mice given normal drinking water. DSS treated animals will be sacrificed at days 1, 2, 3 and 6, and Shh and Ihh expression patterns and levels will be determined on frozen sections by confocal microscopy of GFP & β-galactosidase staining for Shh & Ihh respectively.

2b) Antibiotic treatment. The increasing gradient of Hh ligand expression and Hh pathway activity demonstrated in healthy human colon reflects the increasing bacterial burden of the commensal flora in the distal colon. To ascertain whether commensal bacteria directly affects expression of Hh ligand in the epithelium *Shh*$^{+/GFP}$ and *Ihh*$^{+/LacZ}$ mice raised in SPF conditions, will be treated with vancomycin (1mg/L), neomycin sulphate (500mg/L) and metronidazole (1g/L) in drinking water for 7 days. 7 days following antibiotic treatment, these mice will be sacrificed and the expression pattern of GFP / LacZ and Gli1 mRNA levels in ileum and colon will be compared with control mice (not antibiotic treated). If a significant alteration in Shh and/or Ihh expression is determined with antibiotic treatment I will then proceed to administer antibiotics to *Gli1*$^{+/LacZ}$ mice and monitor Hh pathway activity by analysing LacZ expression and QPCR for Gli1.

2c) Intestinal epithelial cell specific knock-down of NFκB activity. Gumucio and colleagues have previously demonstrated that intestinal epithelial cell specific knock-down of NEMO (leading to absent epithelial NFκB signalling) results in spontaneous intestinal inflammation.[514] However, the level of Hh signalling activity has not yet been determined in this model. Despite the largely negative studies presented in Chapter 9, it is anticipated that Hh ligand expression will be decreased as there is good evidence that Shh is a direct target of NFκB in different tissues (*although not yet tested in the gut or with Ihh*). I will therefore analyse Hh ligand expression and Hh pathway activity by QPCR for Shh, Ihh, and Gli1 isolated from whole ileal and colonic tissue at 1 week (prior to the onset of inflammation), 3 weeks (onset of colonic inflammation) and 6 weeks (severe colitis). As these mice do not develop ileal inflammation it will be of interest to compare Hh expression in ileum as well as colon in this model.

12.3.3 HH as a therapeutic target: do HH agonists modulate inflammatory pathways?

The human genetic and expression data along with the murine data generated to date all supports down-regulated Hh signalling in response to mammalian intestinal inflammation. I will therefore examine, both *in vitro* & *in vivo*, whether the use of Hh agonists modulates intestinal inflammatory pathways to a tolerogenic phenotype and is therefore a potential therapeutic strategy in IBD.

3a) *In vitro* approach. I will first analyse the effects of Hh pre-treatment on the phenotype of human DCs and macrophages in culture, specifically the cytokine profile of treated DCs and their ability to instruct T cells. Initially, PBMCs from non-smoking healthy volunteers will be used. DCs (and macrophages if indicated by 1c) and CD4 T cells will be isolated using negative selection affinity columns. $5x10^5$ DCs (CD11c+ MHCII+) will be pre-treated with exogenous recombinant Shh or Ihh. The cytokine profiles (IL6, IL10, IL23, IFNγ, CX3CR1, TGFβ) of treated & control DCs will be determined by ELISA of culture supernatants. I will next co-culture DCs with $5x10^5$ unstimulated, allogeneic CD4 T cells, and enumerate T_{REGs} & Th17 cells by FACS. If this approach proves informative, I will proceed to repeat these experiments in patients with quiescent UC and CD (non-smokers with no recent (3 months) exposure to corticosteroids, immunosuppressants or biological therapy). IBD patients will be genotyped for the causative mutation (rs2228226) in *GLI1*.

3b) *In vivo* approach. I will subsequently assess whether the administration of Hh agonists can ameliorate acute colitis induced by 3% DSS. Commercially available agonists include agents from Curis Inc. (Cur199567) and Novartis. I will administer Hh agonist and PBS as a control intra-peritoneally both before and 6 days after the induction of colitis to see if a) increased Hh levels protect against colitis and b) whether established colitis can be treated by increasing Hh signalling activity. Clinical parameters will then be monitored and histological analysis performed as described above

12.4 *HH and MDR1 (ABCB1)*

Given the roles described for germline variation for both *GLI1* (presently)[423] and *MDR1* (Ho and colleagues in Edinburgh, see **1.5.4.2**)[260, 515-517] in UC pathogenesis, it is of great interest to note two papers in the literature that describe links between the two. Noting that inhibition of HH signalling increased the reponsiveness of various

cancer cell lines to chemotherapy, Sims-Mourtada and colleagues demonstrated that HH also regulated the expression of MDR1.[518] siRNA knock-down of MDR1 partially reversed HH-induced chemoresistance. Most recently, it was demonstrated that inhibition of HH (SMO-acting HhAntag691) potently decreased the expression and function of MDR1.[519] It will therefore be of great interest to assess whether HH-GLI1 signalling and MDR1 interact in the context of acute and chronic colonic inflammation and / or corticosteroid resistance in IBD. Ho is presently further exploring the function of MDR1 in UC (MRC Clinician Scientist Award) so this hypothesis will be eminently testable in Edinburgh in the near future.

12.5 *IHH and WNT in colitis-associated cancer*

The role of recapitulated embryonic HH signalling in response to injury in a number of different organs / tissues was discussed in Chapter 1 (**1.9.3**). Furthermore, it was proposed that in these scenarios HH may play a role in cancers arising in regions of chronic inflammation in, for example, the hepato-biliary sytem and the pancreas. In the colon however the present data would suggest a different model that requires testing *in vivo*. Down-regulation of HH pathway activity was noted in all forms of colonic inflammation studied in man and mouse. In UC, down-regulated IHH appeared to be driving this decrease. Given the evidence linking HH and WNT (negative regulation) laid out in Chapter 2, it is proposed that down-regulated IHH – GLI1 signalling results in unrestricted WNT signalling in colonic inflammation. In the acute setting this is most likely an appropriate response to resolve the acute inflammatory / infectious insult. However, in the scenario of chronic colitis, this results in chronic inappropriate upregulation of WNT signalling and the risk of malignant transformation (+/- somatic mutation in WNT signalling components).

12.6 *Future studies into the genetic basis of IBD*

We are presently performing a full GWAS in UC as part of the on-going WTCCC project. It is anticipated that these data will be combined with German and N American UC GWAS (see **1.5.4.5** and **1.5.4.6**) to perform a meta-analysis, given the success of this approach in CD. The WTCCC has prioritised CD for follow-up of hits in the second phase of this study designed to fine-map associated loci following deep re-sequencing in a large number of CD patients and HCs. The HH target gene *NKX2.3* is amongst those presently being re-sequenced. Our unit is presently also

involved in a large, multi-national collaborative GWAS in paediatric IBD (CHOPSTICKS), due to report soon.

12.6.1 Follow-up of HH and WNT pathway genetic studies

The haplotype-tagging study of *GLI1* has provided robust evidence for association with IBD. In addition, it has been shown that a nsSNP, that we have shown to result in a mutant protein that is a subfunctional transactivator of *GLI1*, largely accounts for the variation observed. However, deep re-sequencing of the entire *GLI1* gene remains a priority. In addition, it will be of considerable interest to type the CNV that has been documented at the *GLI1* locus.

The data presented in Chapter 8 provide a limited survey of HH and WNT pathway genes for association with CD from pre-existing WTCCC GWAS data and the CD meta-analysis. This has produced some very interesting leads, notably the associations between SUFU and the WNT3/9b locus with CD that merit follow-up by fine-mapping studies, although tight LD in both regions will make such an approach difficult. Given the predominant association of *GLI1* with UC, it will be of great interest to analyse the forth-coming WTCCC UC GWAS data for associations with other HH (and WNT) pathway genes. However, a more robust and formal approach would be to perform large-scale haplotype-tagging studies of all HH and WNT pathway genes in UC and CD in a suitably powered Scottish or UK-wide cohort to avoid type I errors.

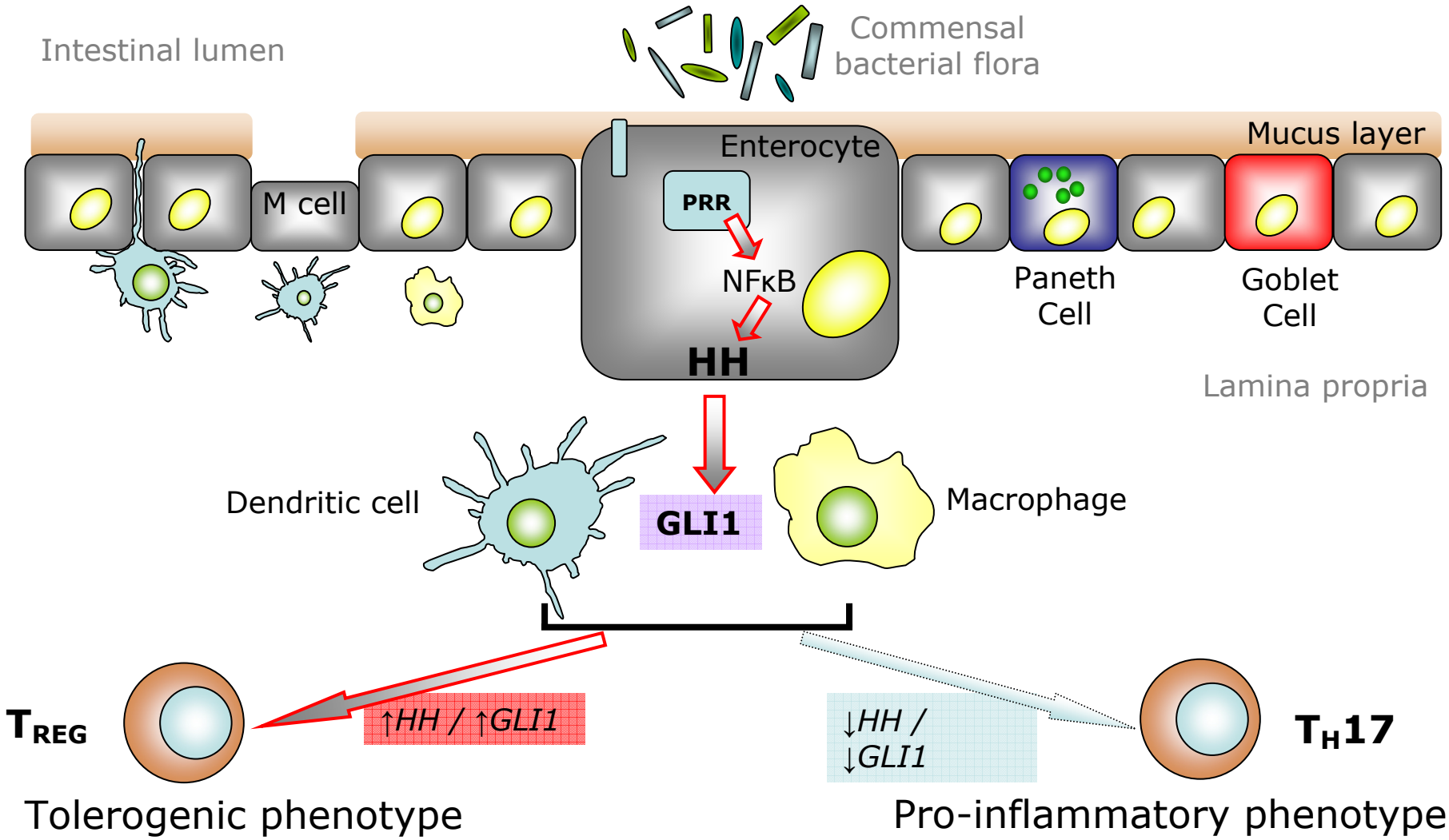

Figure 12-1 Current working hypothesis for the role of HH signalling pathway in the regulation of colonic inflammatory pathways

13 Appendices

Gene Symbol	Title	GO bio process	Fold Change_RMA	p-value_RMA	Fold Change_MAS5.0	p-value_MAS5.0	Fold Change_MAS4.0	p-value_MAS4.0
AI323594	small inducible cytokine A2	GO:6954; GO:7165; GO:6935; GO:6955	1.79	0.005	20.90	0.013	32.43	0.000
Pap	pancreatitis-associated protein	GO:6954; GO:6953;	1.40	0.020	10.78	0.001	29.40	0.006
AW546964	S100 calcium binding protein A9 (calgranulin B)		1.37	0.036	12.82	0.026	23.76	0.003
Gro1	GRO1 oncogene	GO:6954; GO:6955; GO:74; GO:8151	1.58	0.020	15.34	0.012	22.58	0.000
S100a8	S100 calcium binding protein A8 (calgranulin A)	GO:6935	2.33	0.002	7.80	0.025	10.57	0.001
Pglyrp	peptidoglycan recognition protein	GO:6955; GO:6915	2.51	0.002	3.90	0.025	5.27	0.001
Areg	amphiregulin		1.39	0.243	3.57	0.027	4.97	0.009
Scya7	small inducible cytokine A7	GO:6954; GO:7165; GO:6935; GO:6955	1.05	0.701	3.11	0.031	4.28	0.005
Icam1	intercellular adhesion molecule	GO:6952; GO:7155; GO:16337	-1.13	0.546	3.25	0.004	4.27	0.007
Scyb5	small inducible cytokine B, 5	GO:7165; GO:6955	-1.10	0.823	3.36	0.044	3.99	0.005
Serping1	serine proteinase inhibitorclade G, 1	GO:6956; GO:6958	2.33	0.032	2.58	0.038	3.58	0.002
AI385595	cytokine inducible SH2-containing protein	GO:7242; GO:1558	-1.07	0.634	2.30	0.038	3.38	0.037
S100a6	S100 calcium binding protein A6 (calcyclin)	GO:8283; GO:7049	3.67	0.005	2.97	0.086	2.96	0.011
Ifrd1	interferon-related developmental regulator 1		1.10	0.369	2.43	0.008	2.54	0.002
Slpi	secretory leukocyte protease inhibitor		2.18	0.028	2.32	0.036	2.32	0.140
Nrp	immediate early response 5		1.03	0.655	2.25	0.037	2.20	0.015

Il1r1	interleukin 1 receptortype I	GO:7166	-1.02	0.918	1.74	0.007	2.00	0.026
Scya25	small inducible cytokine A25	GO:6935; GO:6954; GO:7165; GO:6955	1.47	0.002	1.71	0.203	1.79	0.067
Cd24a	CD24a antigen	GO:6952	1.88	0.070	1.55	0.011	1.72	0.004
AW546738	B-cell translocation gene 1anti-proliferative		1.15	0.377	1.64	0.050	1.63	0.018
Cnlp	cathelin-like protein	GO:6952; GO:6805	-1.19	0.041	1.13	0.793	1.07	0.736
H2-T18	histocompatibility 2T region locus 3	GO:6952; GO:6955; GO:19883; GO:19885	-1.33	0.033	-1.31	0.557	-1.26	0.177
Bet3-pendi	Bet3 homolog (S. cerevisiae)		-1.22	0.047	-1.38	0.038	-1.33	0.085
Ifrg15-pen	interferon alpha responsive gene15 kDa		-1.15	0.025	-1.26	0.277	-1.37	0.194
AI528744	interleukin 10 receptorbeta	GO:7166	-1.45	0.018	-1.49	0.072	-1.47	0.025
H2-L	histocompatibility 2D region locus 1	GO:6952; GO:6955; GO:19883; GO:19885	-1.78	0.005	-1.60	0.109	-1.62	0.009
Enpep	glutamyl aminopeptidase	GO:6952; GO:6508	-1.95	0.020	-2.00	0.265	-1.64	0.079
H2-D1	histocompatibility 2D region locus 1	GO:6952; GO:6955; GO:19883; GO:19885	-1.14	0.331	-1.72	0.029	-1.69	0.002
H2-K	histocompatibility 2K region	GO:6952	-1.45	0.018	-1.65	0.007	-1.72	0.006
H2-Q2	histocompatibility 2Q region locus 2	GO:6952	-1.29	0.010	-1.77	0.091	-1.73	0.024
H2-T18	histocompatibility 2T region locus 18	GO:6952	-1.45	0.019	-1.66	0.010	-1.83	0.015
Ly64	lymphocyte antigen 64	GO:6952	-2.04	0.003	-1.72	0.002	-1.83	0.010
H2-L	histocompatibility 2L region	GO:6952	-1.41	0.022	-1.92	0.045	-1.84	0.003
Tcrb-V13	T-cell receptor germline beta-chain geneV-region promoter		-1.47	0.034	-1.81	0.027	-1.92	0.004
H2-Q1	histocompatibility 2Q region locus 1	GO:6952	-1.50	0.005	-2.26	0.017	-2.06	0.002
AI266795	S100 calcium binding		-1.32	0.018	-2.23	0.100	-2.26	0.002

	protein A1							
H2-Eb1	histocompatibility 2class II antigen E beta	GO:6952; GO:6955; GO:19884; GO:19886	-1.25	0.279	-2.56	0.094	-2.99	0.020
Ifit3	interferon-induced protein with tetratricopeptide repeats 3	GO:6955	-1.31	0.200	-2.65	0.042	-5.08	0.036
Ifit1	interferon-induced protein with tetratricopeptide repeats 1	GO:6955	-1.58	0.020	-4.45	0.001	-6.16	0.001

Table 13-1 Dysregulated inflammatory genes in microarray of villin-Hhip total small intestine vs. WT
Madison and Gumucio performed a microarray on total small intestine from villin-Hhip mice vs. WT at birth.[303] In this model, transgenic villin drives over-expression of the pan-Hh inhibitor, Hhip in the intestine from embryonic day 14-15, thus resulting in decreased Hh pathway activity. Professor Gumucio has provided this list of inflammatory genes (made by searching for relevant GO terms – see below) dysregulated in this model (unpublished data). The 3 columns that state fold-change relate to the 3 different microarray platforms used in this study: Affy 4.0, Affy 5.0 and RMA (Terry Speed's algorithm). The RMA is thought to be the most robust analysis. For example, in this model it is the only system that picked up a change in PTCH (a robust marker of altered HH pathway activity).

GO terms: 6954 inflammatory response; 7165 signal transduction; 6935 chemotaxis; 6955 immune response; 6953 acute-phase response; 74 cell cycle control; 8151 cell growth and/or maintenance; 6915 apoptosis; 6956 complement activation; 6958 complement activation, classical pathway; 7242 intracellular signaling cascade; 1558 regulation of cell growth; 8283 cell proliferation; 7049 cell cycle; 7166 cell surface receptor linked signal transduction; 6952 defense response; 6805 xenobiotic metabolism; 19883 antigen presentation, endogenous antigen via MHC class I; 6508 proteolysis and peptidolysis; 19884 antigen presentation, exogenous antigen via MHC class II

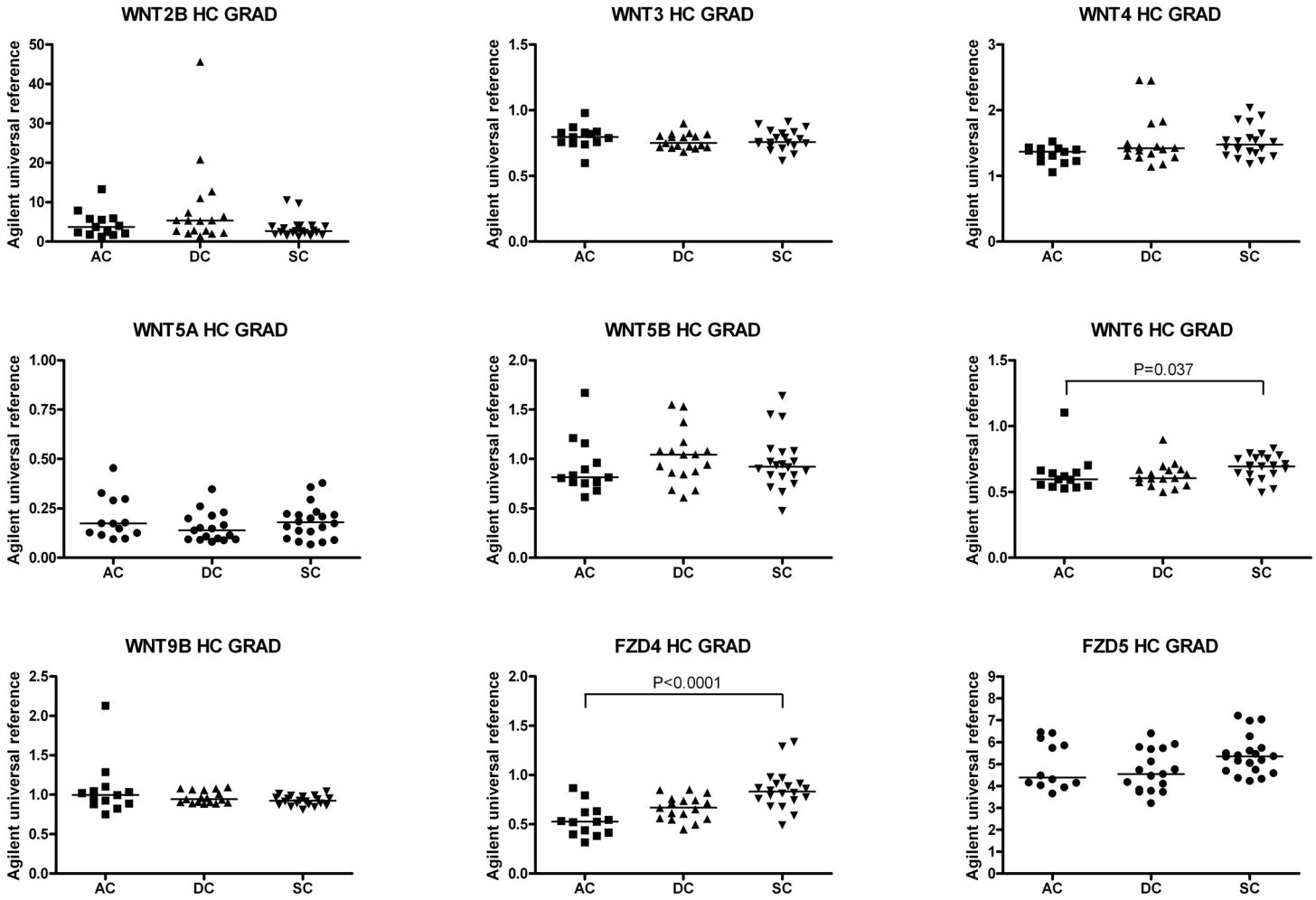

Figure 13-1 WNT signalling expression in healthy (non-inflamed) human colon – Part 1

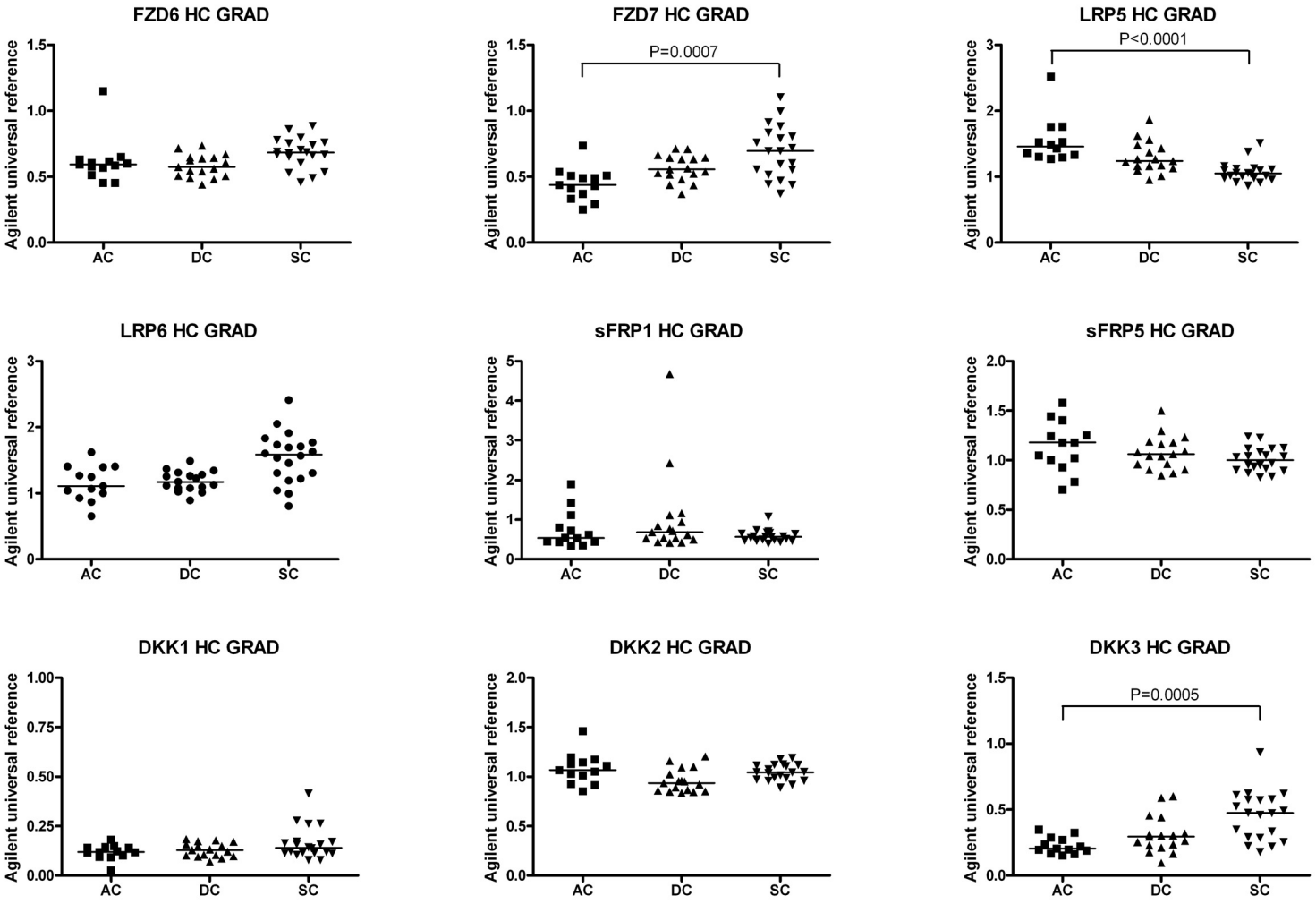

Figure 13-2 WNT signalling in healthy (non-inflamed) human intestine – Part 2

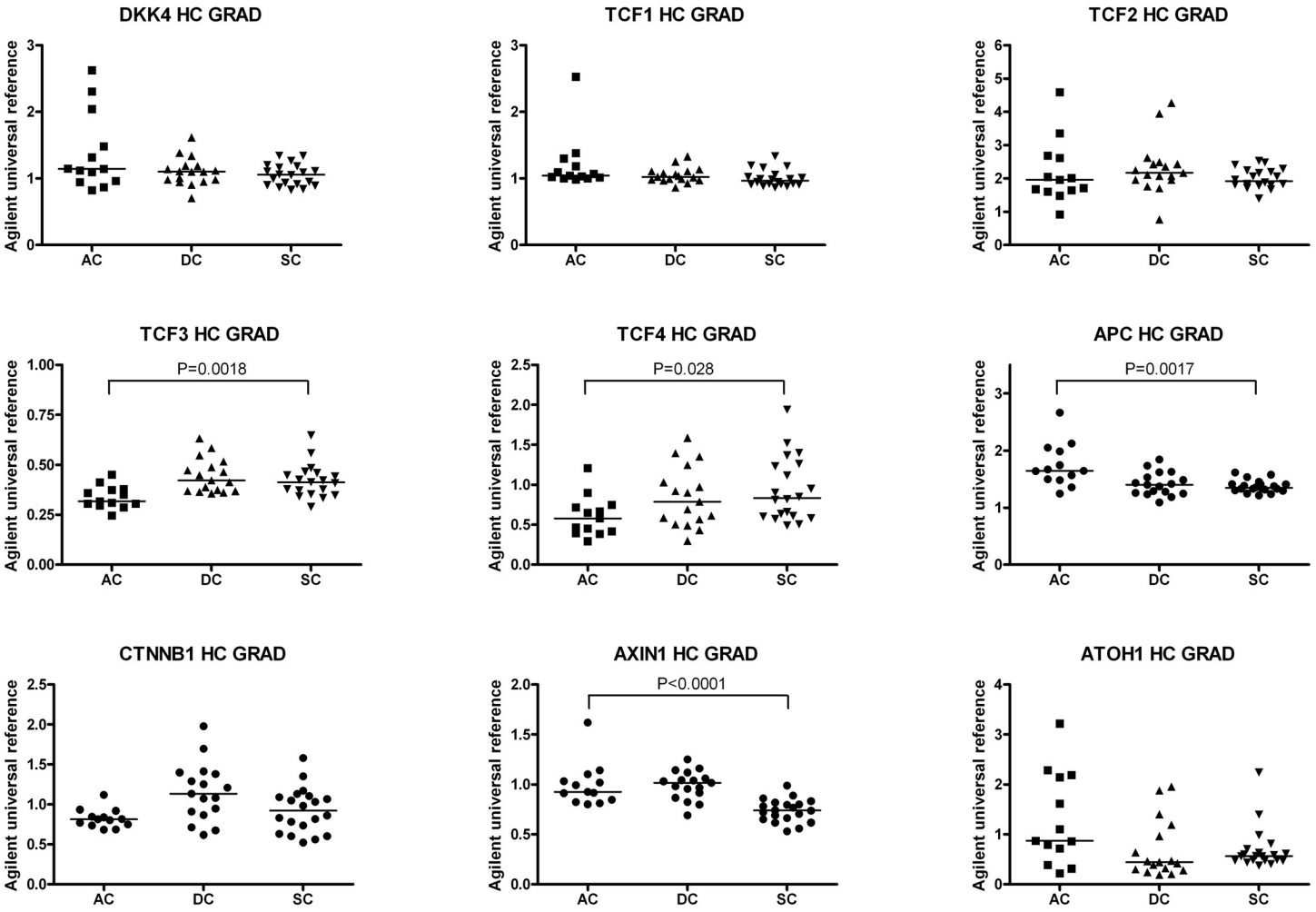

Figure 13-3 WNT signalling in healthy (non-inflamed) human colon – Part 3

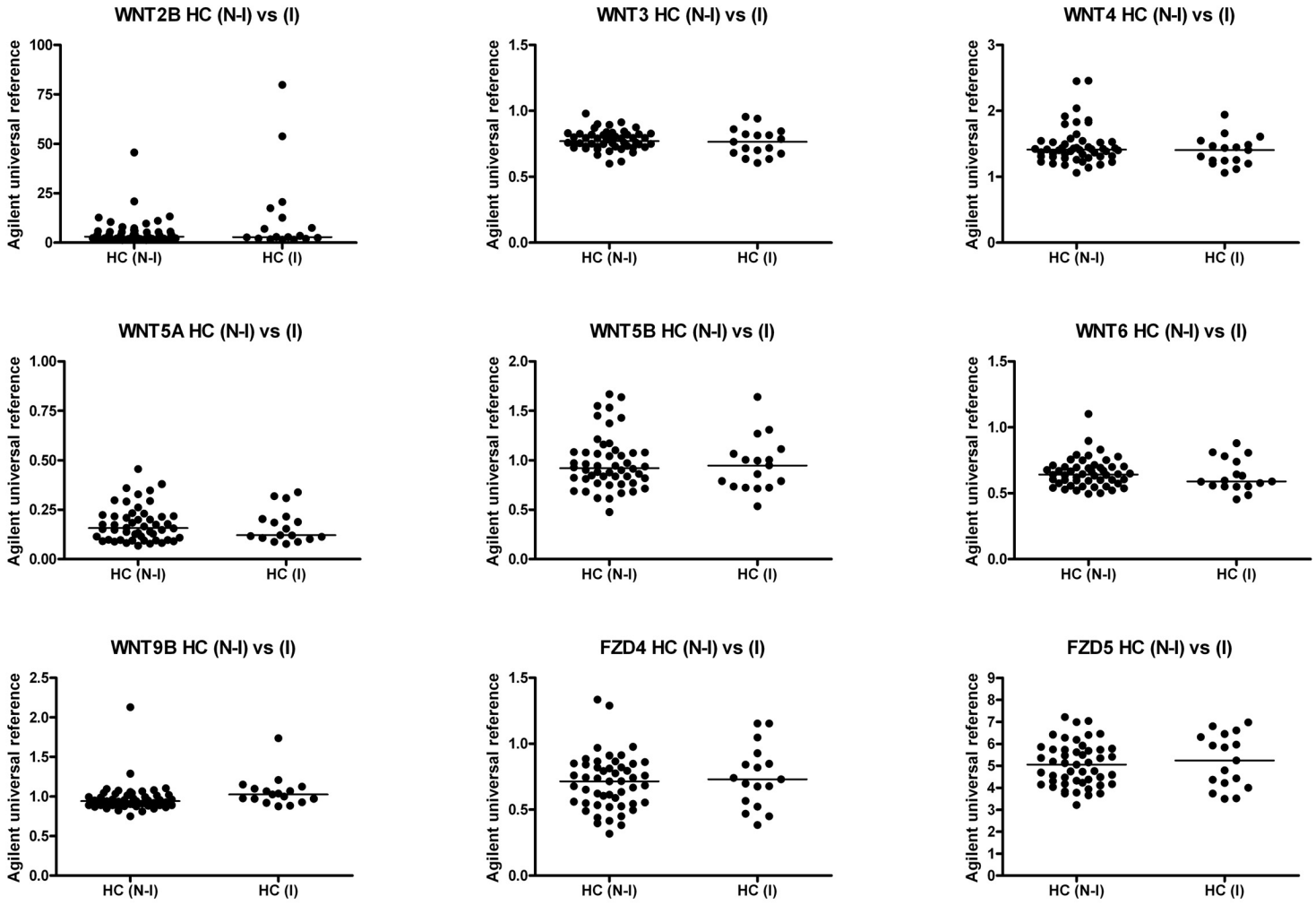

Figure 13-4 WNT signalling in non-IBD colonic inflammation – Part 1

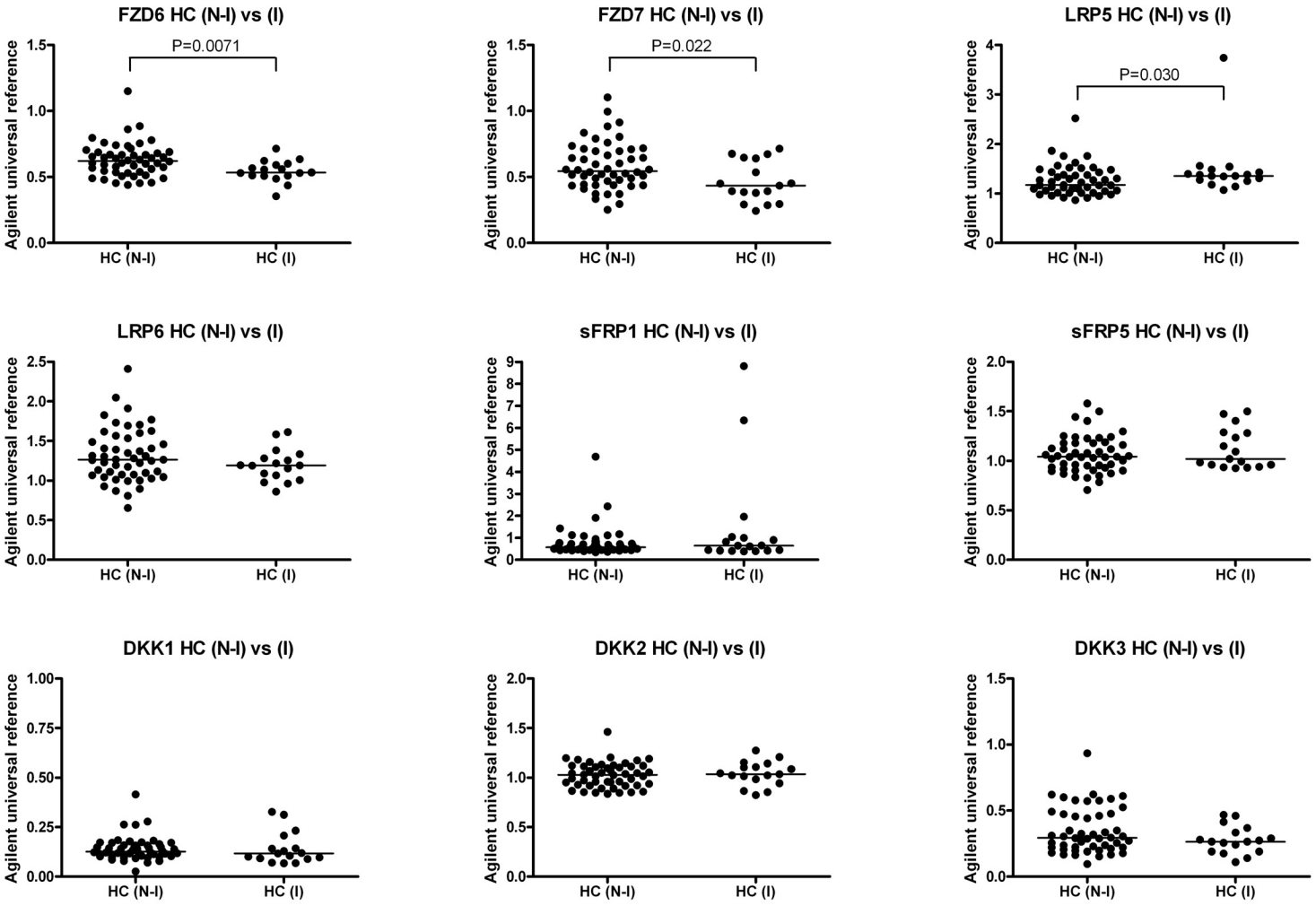

Figure 13-5 WNT signalling in non-IBD colonic inflammation – Part 2

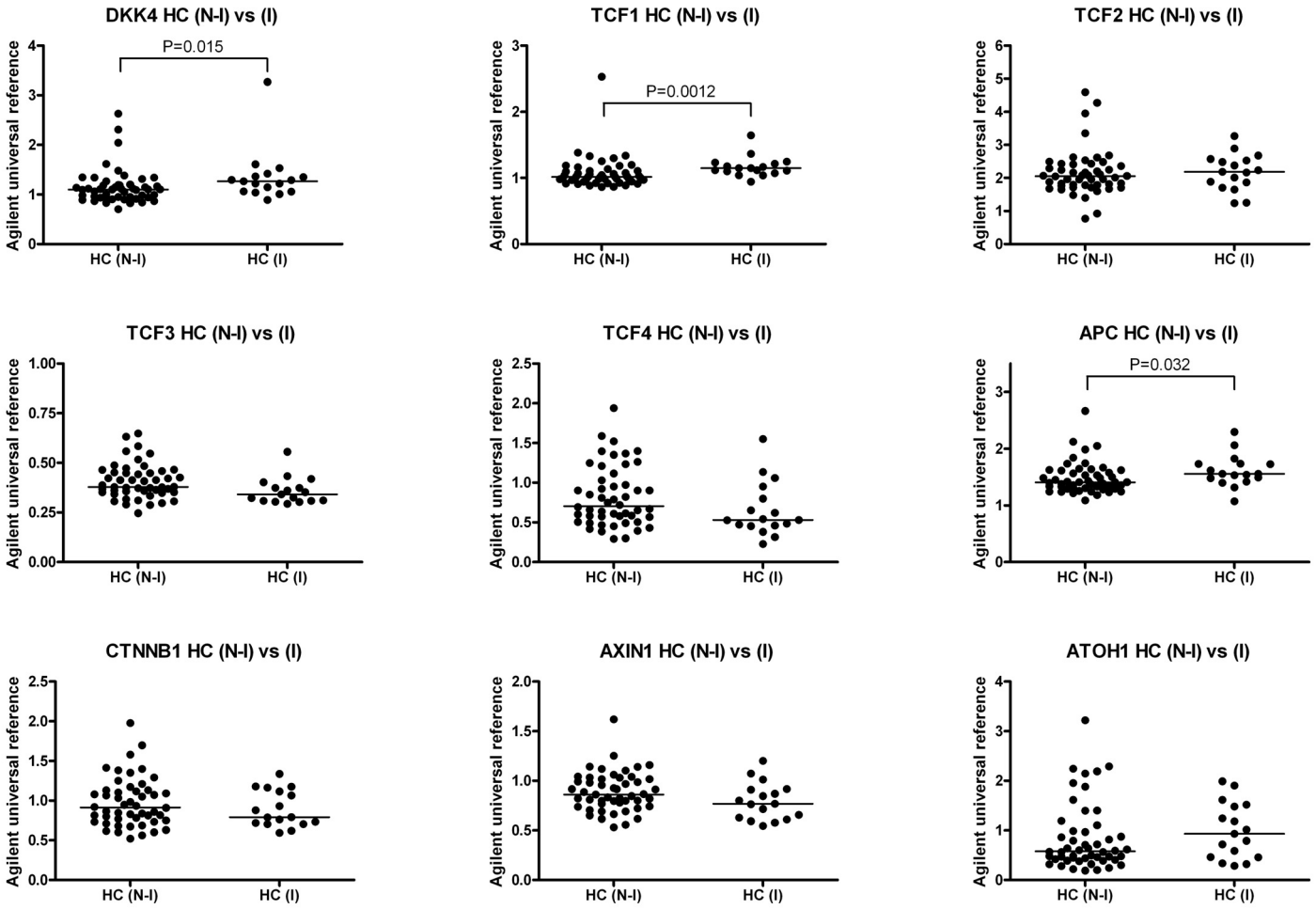

Figure 13-6 WNT signalling in non-IBD colonic inflammation – Part 3

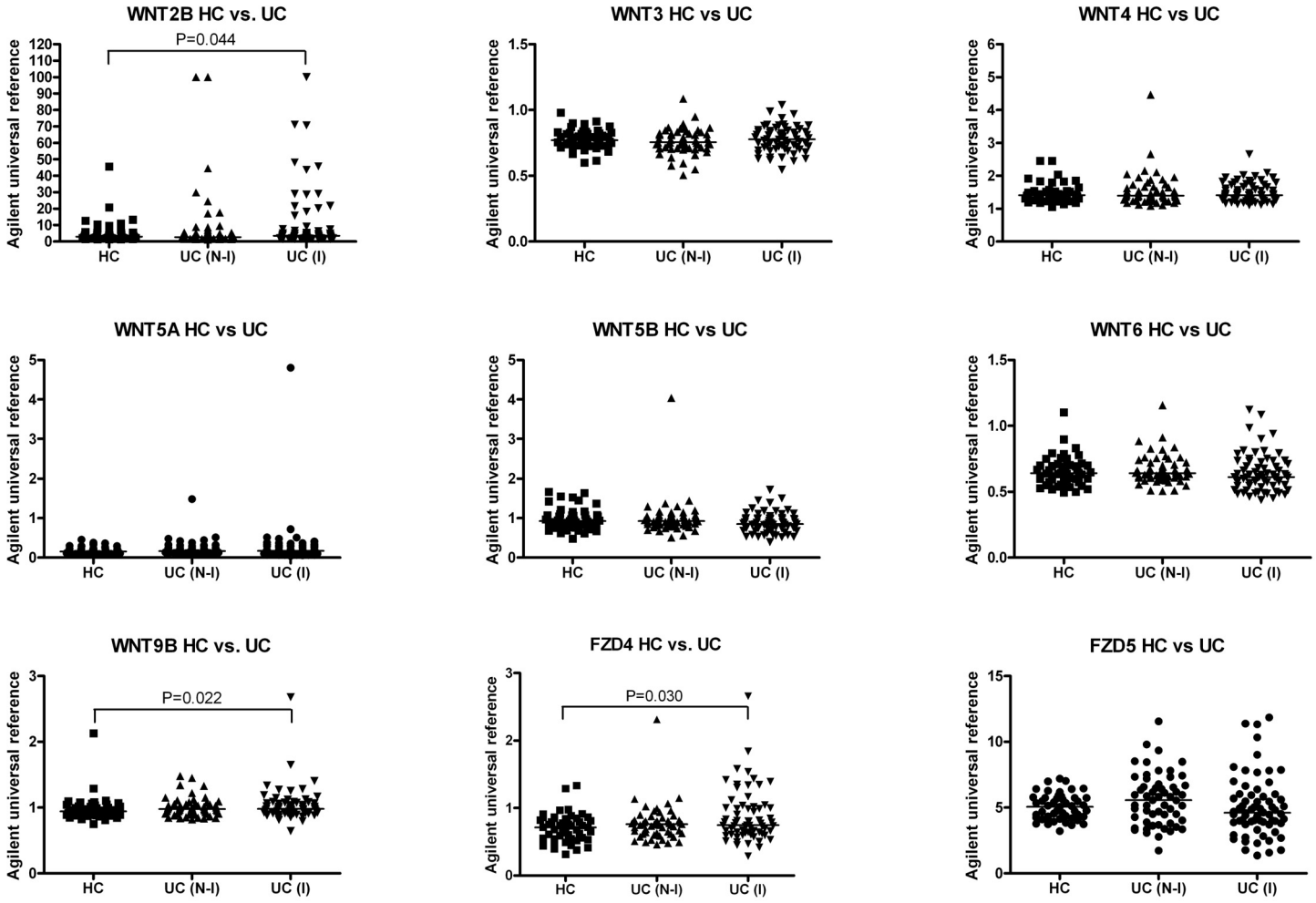

Figure 13-7 WNT signalling in UC versus HC – Part 1

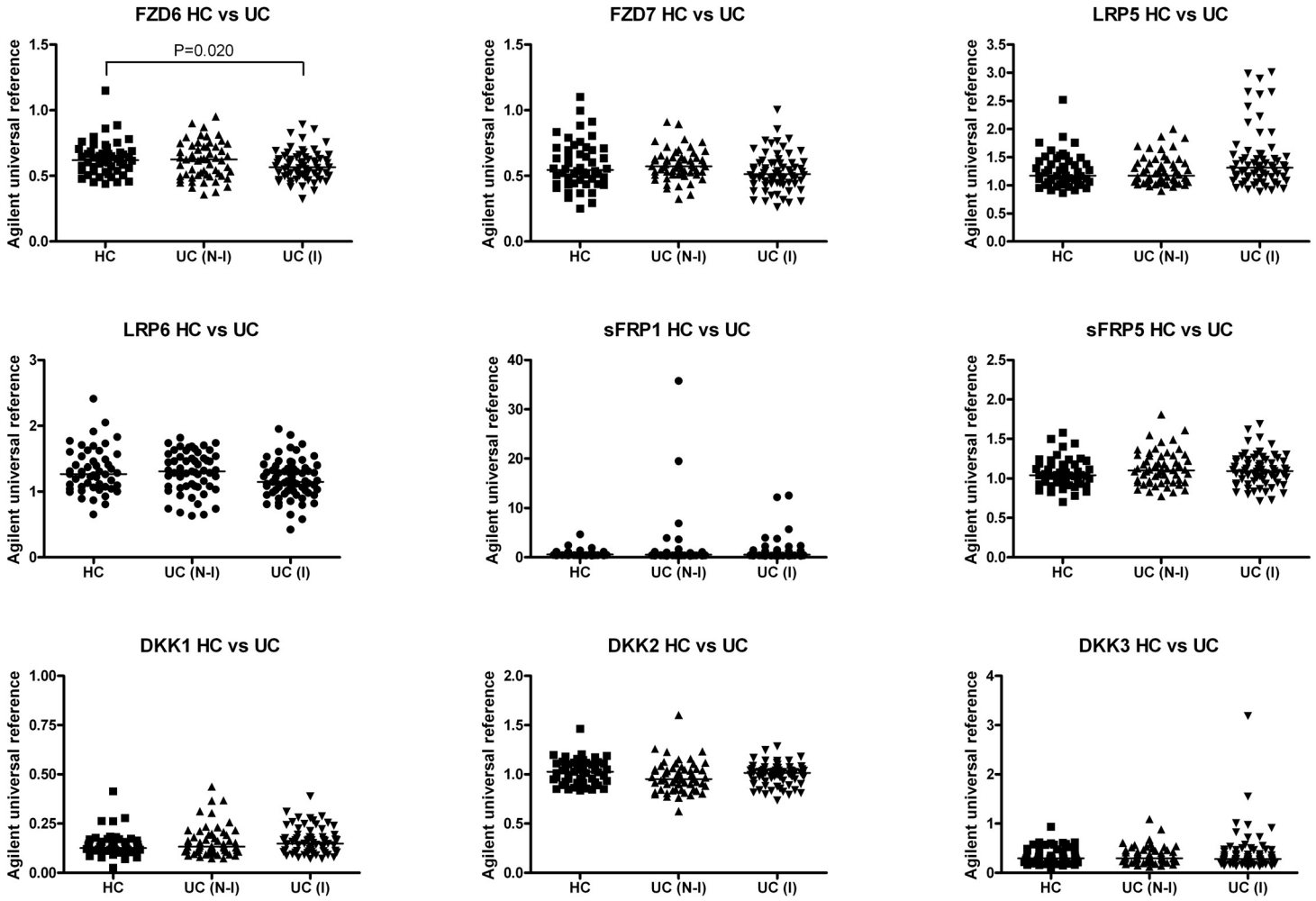

Figure 13-8 WNT signalling in UC versus HC – Part 2

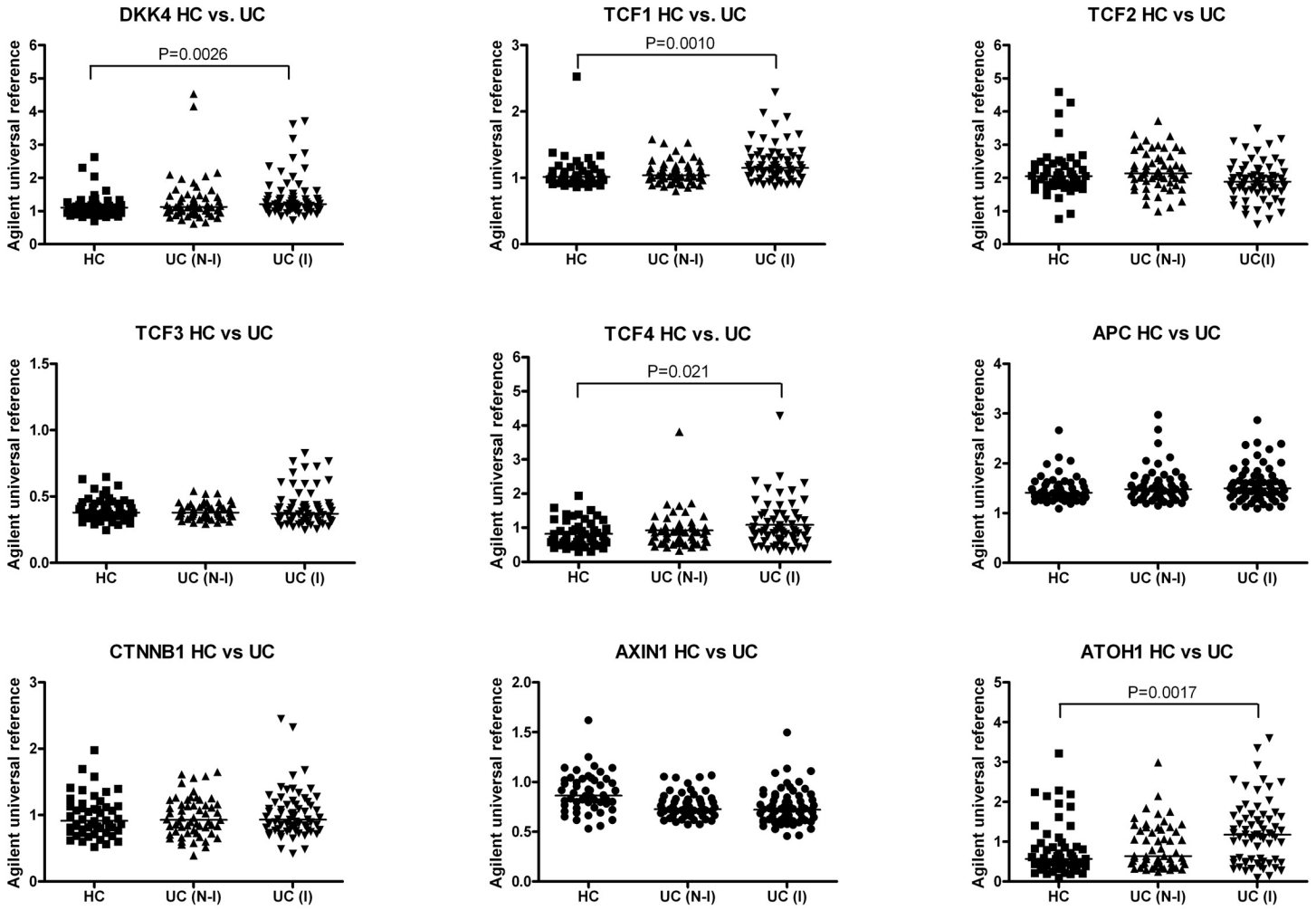

Figure 13-9 WNT signalling in UC versus HC – Part 3

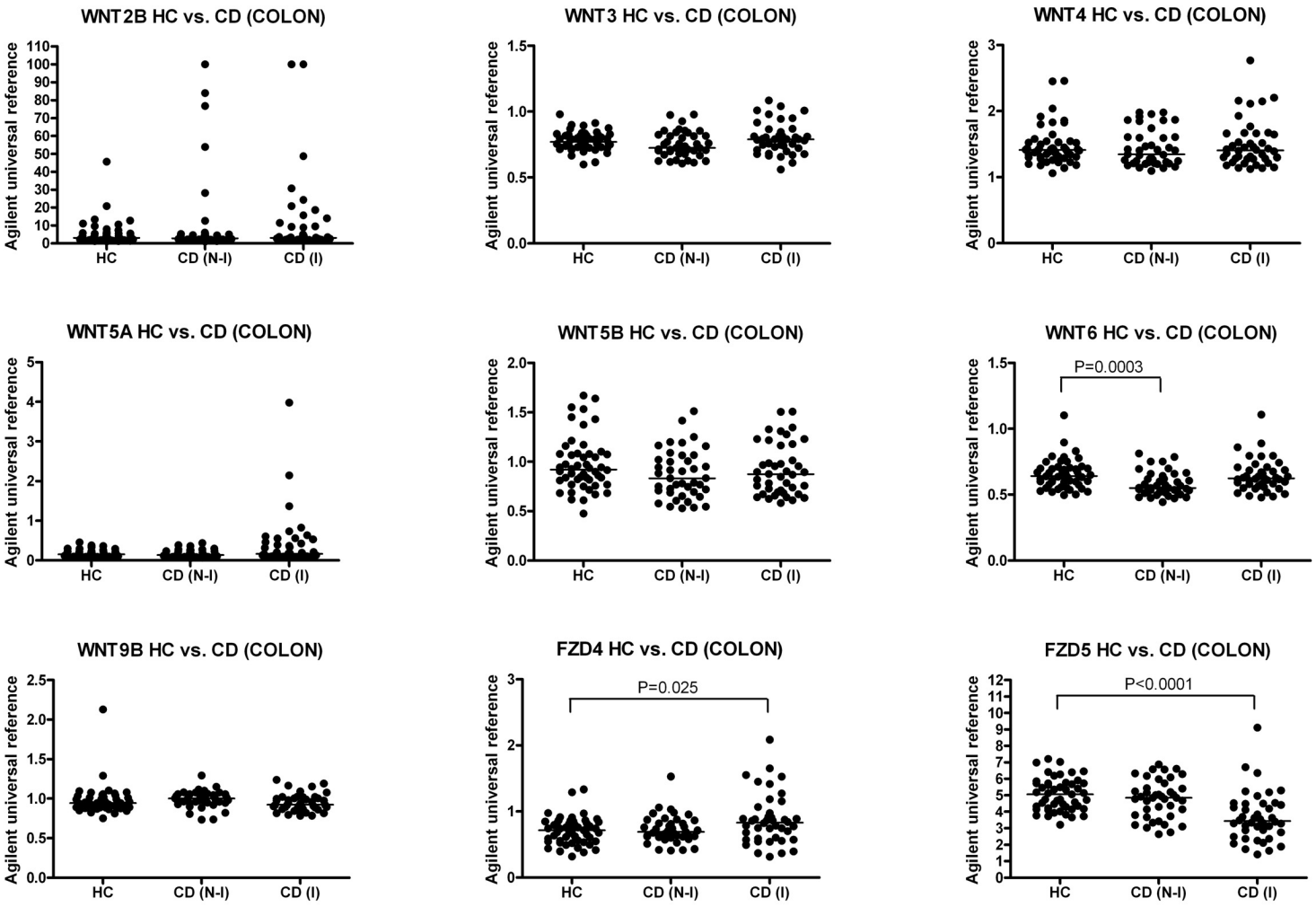

Figure 13-10 WNT signalling in colonic CD versus colonic HC – Part 1

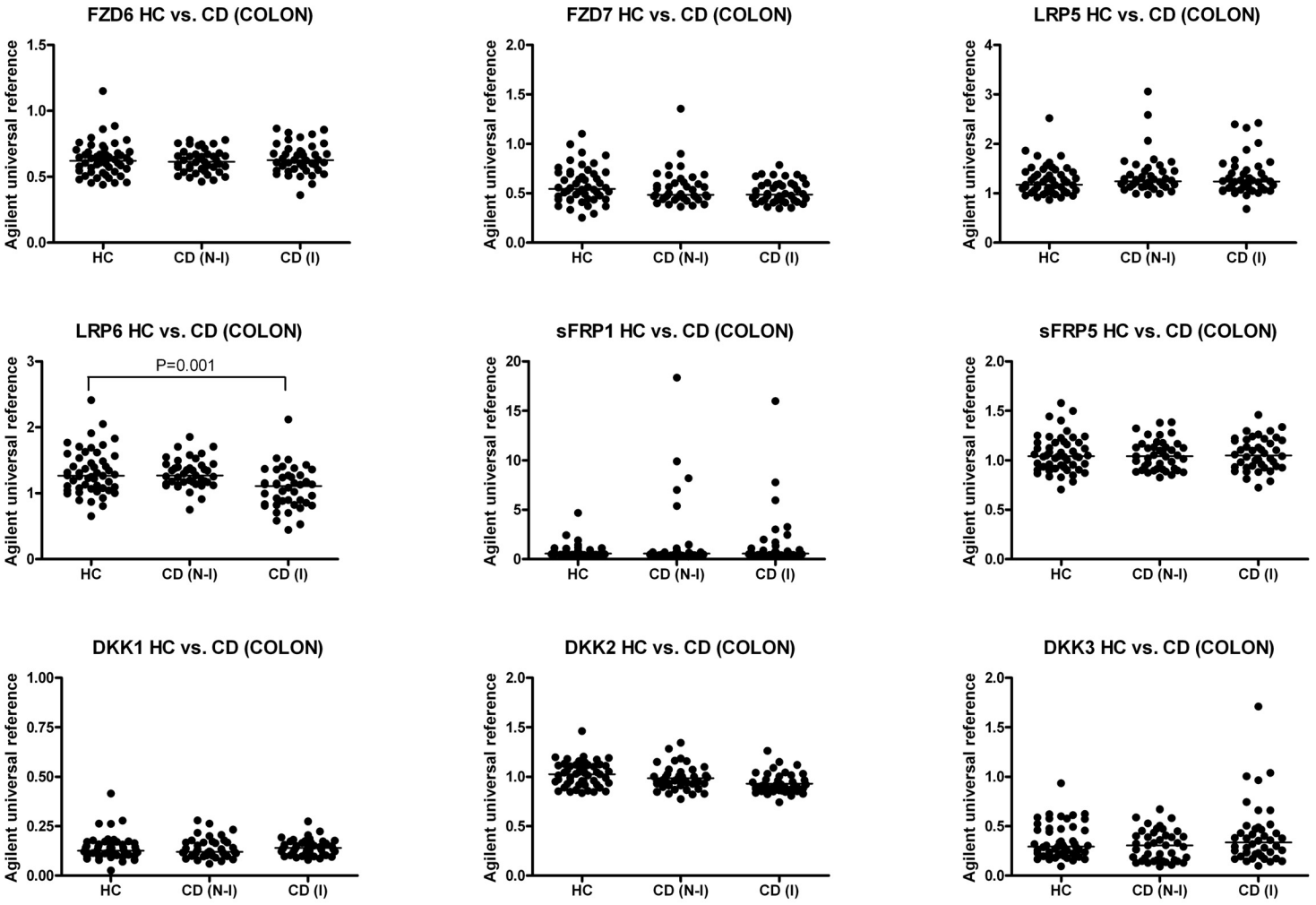

Figure 13-11 WNT signalling in colonic CD versus colonic HC – Part 2

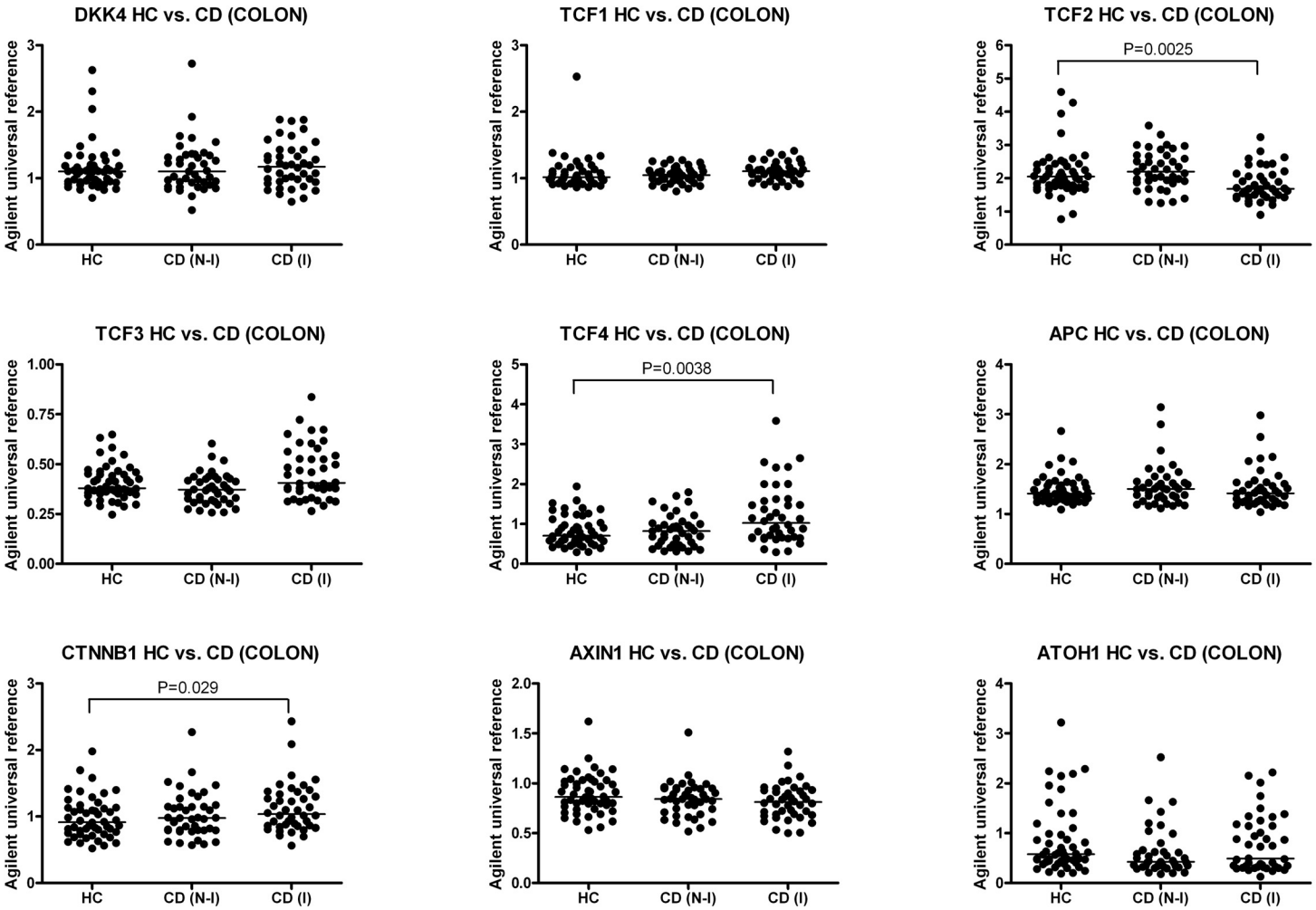

Figure 13-12 WNT signalling in colonic CD versus colonic HC – Part 3

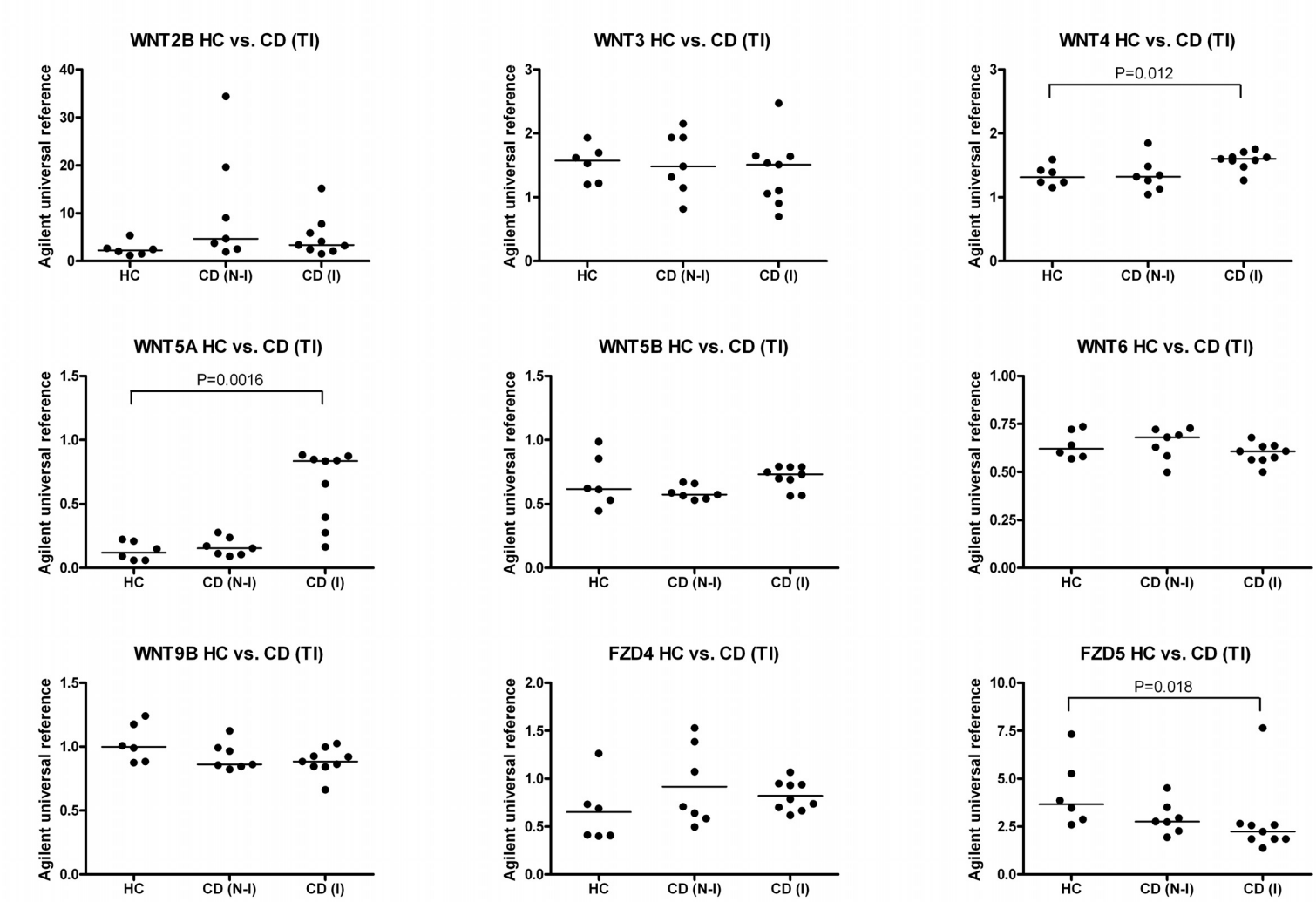

Figure 13-13 WNT signalling in ileal CD versus ileal HC – Part 1

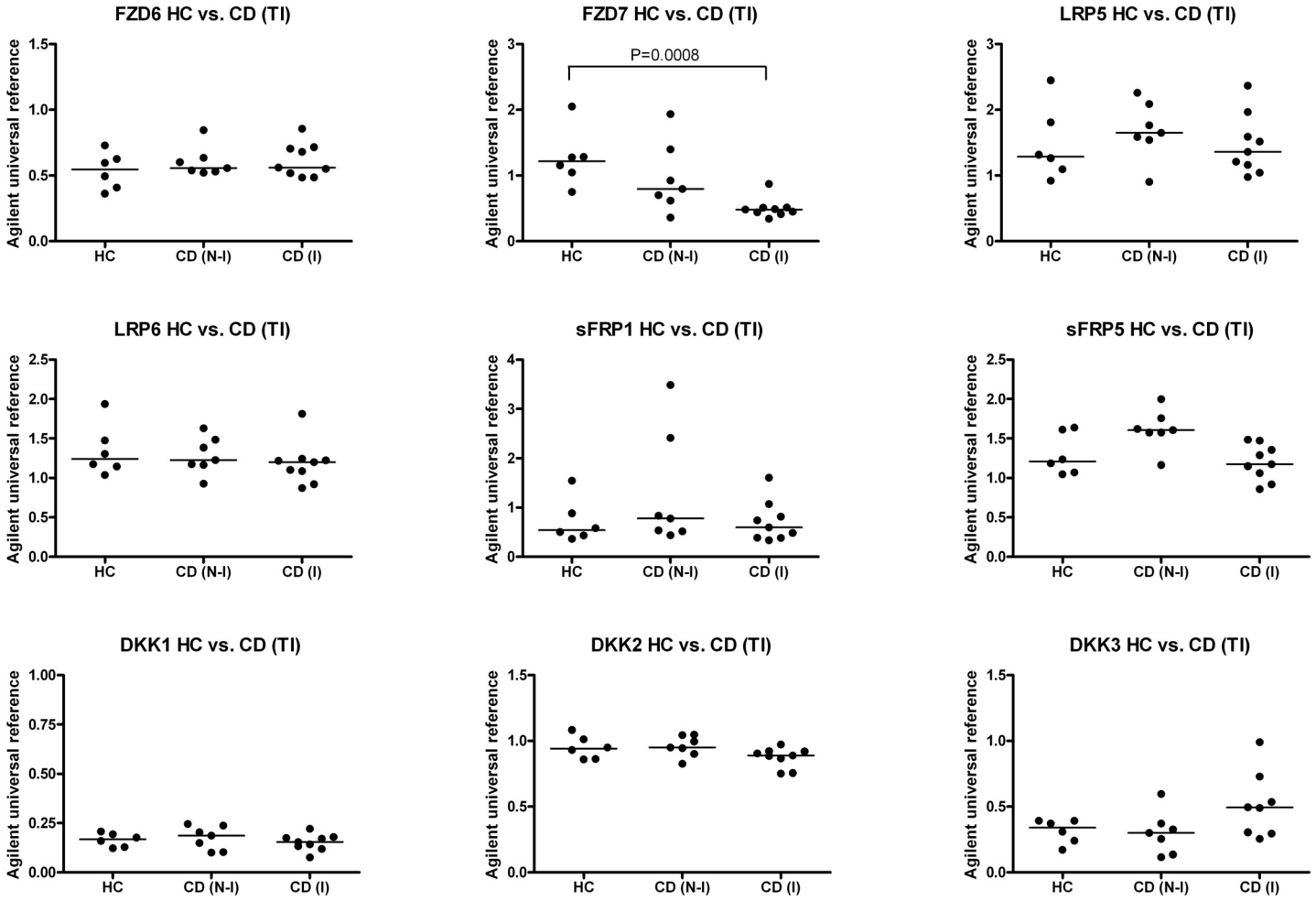

Figure 13-14 WNT signalling in ileal CD versus ileal HC – Part 2

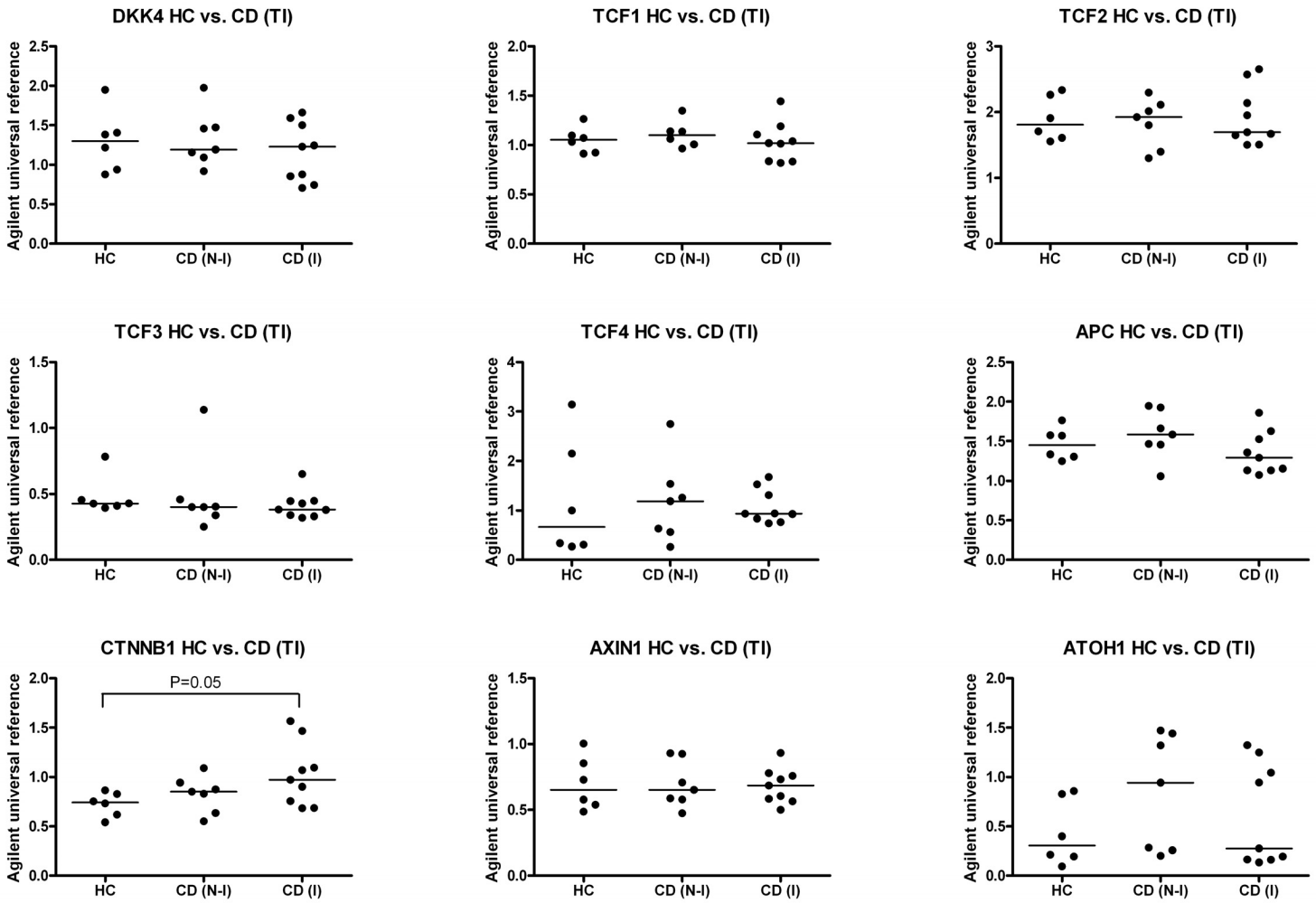

Figure 13-15 WNT signalling in ileal CD versus ileal HC – Part 3

14 References

Reference List

1. Loftus EV, Jr. Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. Gastroenterology 2004;126:1504-1517.

2. Baumgart DC, Carding SR. Inflammatory bowel disease: cause and immunobiology. Lancet 2007;369:1627-1640.

3. Truelove SC, WITTS LJ. Cortisone in ulcerative colitis; final report on a therapeutic trial. Br Med J 1955;2:1041-1048.

4. Faubion WA, Jr., Loftus EV, Jr., Harmsen WS, Zinsmeister AR, Sandborn WJ. The natural history of corticosteroid therapy for inflammatory bowel disease: a population-based study. Gastroenterology 2001;121:255-260.

5. Rutgeerts P, Sandborn WJ, Feagan BG, Reinisch W, Olson A, *et al.* Infliximab for induction and maintenance therapy for ulcerative colitis. N Engl J Med 2005;353:2462-2476.

6. Sandborn W, Sutherland L, Pearson D, May G, Modigliani R, Prantera C. Azathioprine or 6-mercaptopurine for inducing remission of Crohn's disease. Cochrane Database Syst Rev 2000;CD000545.

7. Pearson DC, May GR, Fick G, Sutherland LR. Azathioprine for maintaining remission of Crohn's disease. Cochrane Database Syst Rev 2000;CD000067.

8. Siegel CA, Sands BE. Review article: practical management of inflammatory bowel disease patients taking immunomodulators. Aliment Pharmacol Ther 2005;22:1-16.

9. Connell WR, Kamm MA, Dickson M, Balkwill AM, Ritchie JK, Lennard-Jones JE. Long-term neoplasia risk after azathioprine treatment in inflammatory bowel disease. Lancet 1994;343:1249-1252.

10. Kandiel A, Fraser AG, Korelitz BI, Brensinger C, Lewis JD. Increased risk of lymphoma among inflammatory bowel disease patients treated with azathioprine and 6-mercaptopurine. Gut 2005;54:1121-1125.

11. Lewis JD, Schwartz JS, Lichtenstein GR. Azathioprine for maintenance of remission in Crohn's disease: benefits outweigh the risk of lymphoma. Gastroenterology 2000;118:1018-1024.

12. Lees CW, Maan AK, Hansoti B, Satsangi J, Arnott ID. Tolerability and safety of mercaptopurine in azathioprine-intolerant patients with inflammatory bowel disease. Aliment Pharmacol Ther 2008;27:220-227.

13. Derijks LJ, Gilissen LP, Hooymans PM, Hommes DW. Review article: thiopurines in inflammatory bowel disease. Aliment Pharmacol Ther 2006;24:715-729.

14. Arnott ID, Watts D, Satsangi J. Azathioprine and anti-TNF alpha therapies in Crohn's disease: a review of pharmacology, clinical efficacy and safety. Pharmacol Res 2003;47:1-10.

15. Lennard L. The clinical pharmacology of 6-mercaptopurine. Eur J Clin Pharmacol 1992;43:329-339.

16. Lennard L, Lilleyman JS. Individualizing therapy with 6-mercaptopurine and 6-thioguanine related to the thiopurine methyltransferase genetic polymorphism. Ther Drug Monit 1996;18:328-334.

17. Poppe D, Tiede I, Fritz G, Becker C, Bartsch B, *et al.* Azathioprine suppresses ezrin-radixin-moesin-dependent T cell-APC conjugation through inhibition of Vav guanosine exchange activity on Rac proteins. J Immunol 2006;176:640-651.

18. Tiede I, Fritz G, Strand S, Poppe D, Dvorsky R, *et al.* CD28-dependent Rac1 activation is the molecular target of azathioprine in primary human CD4+ T lymphocytes. J Clin Invest 2003;111:1133-1145.

19. Schaeffeler E, Fischer C, Brockmeier D, Wernet D, Moerike K, *et al.* Comprehensive analysis of thiopurine S-methyltransferase phenotype-genotype correlation in a large population of German-Caucasians and identification of novel TPMT variants. Pharmacogenetics 2004;14:407-417.

20. Ansari A, Hassan C, Duley J, Marinaki A, Shobowale-Bakre EM, Seed P, Meenan J, Yim A, Sanderson J. Thiopurine methyltransferase activity and the use of azathioprine in inflammatory bowel disease. Aliment Pharmacol Ther 2002;16:1743-1750.

21. Gisbert JP, Nino P, Rodrigo L, Cara C, Guijarro LG. Thiopurine methyltransferase (TPMT) activity and adverse effects of azathioprine in inflammatory bowel disease: long-term follow-up study of 394 patients. Am J Gastroenterol 2006;101:2769-2776.

22. Gearry RB, Barclay ML, Burt MJ, Collett JA, Chapman BA, Roberts RL, Kennedy MA. Thiopurine S-methyltransferase (TPMT) genotype does not predict adverse drug reactions to thiopurine drugs in patients with inflammatory bowel disease. Aliment Pharmacol Ther 2003;18:395-400.

23. Hindorf U, Lindqvist M, Hildebrand H, Fagerberg U, Almer S. Adverse events leading to modification of therapy in a large cohort of patients with inflammatory bowel disease. Aliment Pharmacol Ther 2006;24:331-342.

24. Colombel JF, Sandborn WJ, Rutgeerts P, Enns R, Hanauer SB, *et al.* Adalimumab for maintenance of clinical response and remission in patients with Crohn's disease: the CHARM trial. Gastroenterology 2007;132:52-65.

25. Hanauer SB, Feagan BG, Lichtenstein GR, Mayer LF, Schreiber S, *et al.* Maintenance infliximab for Crohn's disease: the ACCENT I randomised trial. Lancet 2002;359:1541-1549.

26. Hanauer SB, Sandborn WJ, Rutgeerts P, Fedorak RN, Lukas M, *et al.* Human anti-tumor necrosis factor monoclonal antibody (adalimumab) in Crohn's disease: the CLASSIC-I trial. Gastroenterology 2006;130:323-333.

27. Sandborn WJ, Hanauer SB, Rutgeerts P, Fedorak RN, Lukas M, *et al.* Adalimumab for maintenance treatment of Crohn's disease: results of the CLASSIC II trial. Gut 2007;56:1232-1239.

28. Sands BE, Anderson FH, Bernstein CN, Chey WY, Feagan BG, *et al.* Infliximab maintenance therapy for fistulizing Crohn's disease. N Engl J Med 2004;350:876-885.

29. Sandborn WJ, Feagan BG, Stoinov S, Honiball PJ, Rutgeerts P, Mason D, Bloomfield R, Schreiber S. Certolizumab pegol for the treatment of Crohn's disease. N Engl J Med 2007;357:228-238.

30. Schreiber S, Rutgeerts P, Fedorak RN, Khaliq-Kareemi M, *et al.* A randomized, placebo-controlled trial of certolizumab pegol (CDP870) for treatment of Crohn's disease. Gastroenterology 2005;129:807-818.

31. Schreiber S, Khaliq-Kareemi M, Lawrance IC, Thomsen OO, Hanauer SB, McColm J, Bloomfield R, Sandborn WJ. Maintenance therapy with certolizumab pegol for Crohn's disease. N Engl J Med 2007;357:239-250.

32. Jarnerot G, Hertervig E, Friis-Liby I, Blomquist L, *et al.* Infliximab as rescue therapy in severe to moderately severe ulcerative colitis: a randomized, placebo-controlled study. Gastroenterology 2005;128:1805-1811.

33. Lees CW, Heys D, Ho GT, Noble CL, Shand AG, Mowat C, Boulton-Jones R, Williams A, Church N, Satsangi J, Arnott ID. A retrospective analysis of the efficacy and safety of infliximab as rescue therapy in acute severe ulcerative colitis. Aliment Pharmacol Ther 2007;26:411-419.

34. Hyams J, Crandall W, Kugathasan S, Griffiths A, *et al.* Induction and maintenance infliximab therapy for the treatment of moderate-to-severe Crohn's disease in children. Gastroenterology 2007;132:863-873.

35. Colombel JF, Loftus EV, Jr., Tremaine WJ, Egan LJ, Harmsen WS, Schleck CD, Zinsmeister AR, Sandborn WJ. The safety profile of infliximab in patients with Crohn's disease: the Mayo clinic experience in 500 patients. Gastroenterology 2004;126:19-31.

36. Ljung T, Karlen P, Schmidt D, Hellstrom PM, Lapidus A, Janczewska I, Sjoqvist U, Lofberg R. Infliximab in inflammatory bowel disease: clinical outcome in a population based cohort from Stockholm County. Gut 2004;53:849-853.

37. Keane J, Gershon S, Wise RP, Mirabile-Levens E, Kasznica J, Schwieterman WD, Siegel JN, Braun MM. Tuberculosis associated with infliximab, a tumor necrosis factor alpha-neutralizing agent. N Engl J Med 2001;345:1098-1104.

38. Toruner M, Loftus EV, Jr., Harmsen WS, Zinsmeister AR, Orenstein R, Sandborn WJ, Colombel JF, Egan LJ. Risk Factors for Opportunistic Infections in Patients With Inflammatory Bowel Disease. Gastroenterology 2008.

39. Warris A, Bjorneklett A, Gaustad P. Invasive pulmonary aspergillosis associated with infliximab therapy. N Engl J Med 2001;344:1099-1100.

40. Baert F, Noman M, Vermeire S, Van Assche G, D' Haens G, Carbonez A, Rutgeerts P. Influence of immunogenicity on the long-term efficacy of infliximab in Crohn's disease. N Engl J Med 2003;348:601-608.

41. Vermeire S, Noman M, Van Assche G, Baert F, D'Haens G, Rutgeerts P. Effectiveness of concomitant immunosuppressive therapy in suppressing the formation of antibodies to infliximab in Crohn's disease. Gut 2007;56:1226-1231.

42. Lees CW, Ali AI, Thompson AI, Ho GT, Forsythe RO, Marquez L, Cochrane CJ, Aitken S, Fennell J, Rogers P, Shand AG, Penman ID, Palmer KR, Wilson DC, Arnott ID, Satsangi J. The safety profile of anti-tumour necrosis factor therapy in inflammatory bowel disease in clinical practice: analysis of 620 patient-years follow-up. Aliment Pharmacol Ther 2009;29:286-297.

43. Mackey AC, Green L, Liang LC, Dinndorf P, Avigan M. Hepatosplenic T cell lymphoma associated with infliximab use in young patients treated for inflammatory bowel disease. J Pediatr Gastroenterol Nutr 2007;44:265-267.

44. Rennard SI, Fogarty C, Kelsen S, Long W, *et al.* The safety and efficacy of infliximab in moderate to severe chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2007;175:926-934.

45. Lees CW, Ironside J, Wallace WA, Satsangi J. Resolution of non-small-cell lung cancer after withdrawal of anti-TNF therapy. N Engl J Med 2008;359:320-321.

46. Koebel CM, Vermi W, Swann JB, Zerafa N, Rodig SJ, Old LJ, Smyth MJ, Schreiber RD. Adaptive immunity maintains occult cancer in an equilibrium state. Nature 2007;450:903-907.

47. D'Haens G, Baert F, Van Assche G, Caenepeel P, Vergauwe P, *et al.* Early combined immunosuppression or conventional management in patients with newly diagnosed Crohn's disease: an open randomised trial. Lancet 2008;371:660-667.

48. Lees CW, Satsangi J. Early combined immunosuppression in Crohn's disease. Lancet 2008;371:1995-1997.

49. Truelove SC, Willoughby CP, Lee EG, Kettlewell MG. Further experience in the treatment of severe attacks of ulcerative colitis. Lancet 1978;2:1086-1088.

50. Buckell NA, Lennard-Jones JE. How district hospitals see acute colitis. Lancet 1979;1:1226-1229.

51. Carbonnel F, Gargouri D, Lemann M, Beaugerie L, Cattan S, Cosnes J, Gendre JP. Predictive factors of outcome of intensive intravenous treatment for attacks of ulcerative colitis. Aliment Pharmacol Ther 2000;14:273-279.

52. Chakravarty BJ. Predictors and the rate of medical treatment failure in ulcerative colitis. Am J Gastroenterol 1993;88:852-855.

53. Ho GT, Mowat C, Goddard CJ, Fennell JM, Shah NB, Prescott RJ, Satsangi J. Predicting the outcome of severe ulcerative colitis: development of a novel risk score to aid early selection of patients for second-line medical therapy or surgery. Aliment Pharmacol Ther 2004;19:1079-1087.

54. Jarnerot G, Rolny P, Sandberg-Gertzen H. Intensive intravenous treatment of ulcerative colitis. Gastroenterology 1985;89:1005-1013.

55. Lindgren SC, Flood LM, Kilander AF, Lofberg R, Persson TB, Sjodahl RI. Early predictors of glucocorticosteroid treatment failure in severe and moderately severe attacks of ulcerative colitis. Eur J Gastroenterol Hepatol 1998;10:831-835.

56. Travis SP, Farrant JM, Ricketts C, Nolan DJ, Mortensen NM, Kettlewell MG, Jewell DP. Predicting outcome in severe ulcerative colitis. Gut 1996;38:905-910.

57. Cohen RD, Brodsky AL, Hanauer SB. A comparison of the quality of life in patients with severe ulcerative colitis after total colectomy versus medical treatment with intravenous cyclosporin. Inflamm Bowel Dis 1999;5:1-10.

58. Lichtenstein GR, Cohen R, Yamashita B, Diamond RH. Quality of life after proctocolectomy with ileoanal anastomosis for patients with ulcerative colitis. J Clin Gastroenterol 2006;40:669-677.

59. Waljee A, Waljee J, Morris AM, Higgins PD. Threefold increased risk of infertility: a meta-analysis of infertility after ileal pouch anal anastomosis in ulcerative colitis. Gut 2006;55:1575-1580.

60. Lichtiger S, Present DH, Kornbluth A, Gelernt I, Bauer J, Galler G, Michelassi F, Hanauer S. Cyclosporine in severe ulcerative colitis refractory to steroid therapy. N Engl J Med 1994;330:1841-1845.

61. D'Haens G, Lemmens L, Geboes K, Vandeputte L, Van AF, Mortelmans L, Peeters M, Vermeire S, Penninckx F, Nevens F, Hiele M, Rutgeerts P. Intravenous cyclosporine versus intravenous corticosteroids as single therapy for severe attacks of ulcerative colitis. Gastroenterology 2001;120:1323-1329.

62. Shibolet O, Regushevskaya E, Brezis M, Soares-Weiser K. Cyclosporine A for induction of remission in severe ulcerative colitis. Cochrane Database Syst Rev 2005;CD004277.

63. Probert CS, Hearing SD, Schreiber S, Kuhbacher T, Ghosh S, Arnott ID, Forbes A. Infliximab in moderately severe glucocorticoid resistant ulcerative colitis: a randomised controlled trial. Gut 2003;52:998-1002.

64. Armuzzi A, De PB, Lupascu A, Fedeli P, Leo D, Mentella MC, Vincenti F, Melina D, Gasbarrini G, Pola P, Gasbarrini A. Infliximab in the treatment of steroid-dependent ulcerative colitis. Eur Rev Med Pharmacol Sci 2004;8:231-233.

65. Chey WY, Hussain A, Ryan C, Potter GD, Shah A. Infliximab for refractory ulcerative colitis. Am J Gastroenterol 2001;96:2373-2381.

66. Eidelwein AP, Cuffari C, Abadom V, Oliva-Hemker M. Infliximab efficacy in pediatric ulcerative colitis. Inflamm Bowel Dis 2005;11:213-218.

67. Gornet JM, Couve S, Hassani Z, Delchier JC, Marteau P, Cosnes J, Bouhnik Y, Dupas JL, Modigliani R, Taillard F, Lemann M. Infliximab for refractory ulcerative colitis or indeterminate colitis: an open-label multicentre study. Aliment Pharmacol Ther 2003;18:175-181.

68. Ochsenkuhn T, Sackmann M, Goke B. Infliximab for acute, not steroid-refractory ulcerative colitis: a randomized pilot study. Eur J Gastroenterol Hepatol 2004;16:1167-1171.

69. Regueiro M, Curtis J, Plevy S. Infliximab for hospitalized patients with severe ulcerative colitis. J Clin Gastroenterol 2006;40:476-481.

70. Sands BE, Tremaine WJ, Sandborn WJ, Rutgeerts PJ, Hanauer SB, Mayer L, Targan SR, Podolsky DK. Infliximab in the treatment of severe, steroid-refractory ulcerative colitis: a pilot study. Inflamm Bowel Dis 2001;7:83-88.

71. Lees CW, Shand AG, Penman ID, Satsangi J, Arnott ID. Role of infliximab in ulcerative colitis: further questions. Inflamm Bowel Dis 2006;12:335-337.

72. Xavier RJ, Podolsky DK. Unravelling the pathogenesis of inflammatory bowel disease. Nature 2007;448:427-434.

73. Sartor RB. Mechanisms of disease: pathogenesis of Crohn's disease and ulcerative colitis. Nat Clin Pract Gastroenterol Hepatol 2006;3:390-407.

74. Cho JH. The genetics and immunopathogenesis of inflammatory bowel disease. Nat Rev Immunol 2008;8:458-466.

75. Russell RK, Farhadi R, Wilson M, Drummond H, Satsangi J, Wilson DC. Perinatal passive smoke exposure may be more important than childhood exposure in the risk of developing childhood IBD. Gut 2005;54:1500-1501.

76. Ingram JR, Thomas GA, Rhodes J, Green JT, Hawkes ND, Swift JL, Srivastava ED, Evans BK, Williams GT, Newcombe RG, Courtney E, Pillai S. A randomized trial of nicotine enemas for active ulcerative colitis. Clin Gastroenterol Hepatol 2005;3:1107-1114.

77. McGrath J, McDonald JW, Macdonald JK. Transdermal nicotine for induction of remission in ulcerative colitis. Cochrane Database Syst Rev 2004;CD004722.

78. Gaya DR, Russell RK, Nimmo ER, Satsangi J. New genes in inflammatory bowel disease: lessons for complex diseases? Lancet 2006;367:1271-1284.

79. Hugot JP, Chamaillard M, Zouali H, Lesage S, *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 2001;411:599-603.

80. Ogura Y, Bonen DK, Inohara N, Nicolae DL, *et al.* A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature 2001;411:603-606.

81. WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661-678.

82. Anderson CA, Massey DC, Barrett JC, Prescott NJ, Tremelling M, Fisher SA, Gwilliam R, Jacob J, Nimmo ER, Drummond H, Lees CW, Onnie CM, Hanson C, Blaszczyk K, Ravindrarajah R, Hunt S, Varma D, Hammond N, Lewis G, Attlesey H, Watkins N, Ouwehand W, Strachan D, McArdle W, Lewis CM, Lobo A, Sanderson J, Jewell DP, Deloukas P, Mansfield JC, Mathew CG, Satsangi J, Parkes M. Investigation of Crohn's Disease Risk Loci in Ulcerative Colitis Further Defines Their Molecular Relationship. Gastroenterology 2008.

83. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, *et al.* A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science 2006;314:1461-1463.

84. Fisher SA, Tremelling M, Anderson CA, Gwilliam R, Bumpstead S, Prescott NJ, Nimmo ER, Massey D, Berzuini C, Johnson C, Barrett JC, Cummings FR, Drummond H, Lees CW, Onnie CM, Hanson CE, Blaszczyk K, Inouye M, Ewels P, Ravindrarajah R, Keniry A, Hunt S, Carter M, Watkins N, Ouwehand W, Lewis CM, Cardon L, Lobo A, Forbes A, Sanderson J, Jewell DP, Mansfield JC, Deloukas P, Mathew CG, Parkes M, Satsangi J. Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease. Nat Genet 2008;40:710-712.

85. Franke A, Balschun T, Karlsen TH, Hedderich J, *et al.* Replication of signals from recent studies of Crohn's disease identifies previously unknown disease loci for ulcerative colitis. Nat Genet 2008;40:713-715.

86. Franke A, Balschun T, Karlsen TH, Sventoraityte J, *et al.* Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. Nat Genet 2008;40:1319-1323.

87. Hampe J, Franke A, Rosenstiel P, Till A, *et al.* A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. Nat Genet 2007;39:207-211.

88. Libioulle C, Louis E, Hansoul S, Sandor C, *et al.* Novel crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. PLoS Genet 2007;3:e58.

89. Parkes M, Barrett JC, Prescott NJ, Tremelling M, Anderson CA, Fisher SA, Roberts RG, Nimmo ER, Cummings FR, Soars D, Drummond H, Lees CW, Khawaja SA, Bagnall R, Burke DA, Todhunter CE, Ahmad T, Onnie CM, McArdle W, Strachan D, Bethel G, Bryan C, Lewis CM, Deloukas P, Forbes A, Sanderson J, Jewell DP, Satsangi J, Mansfield JC, Cardon L, Mathew CG. Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. Nat Genet 2007.

90. Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, *et al.* Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. Nat Genet 2007;39:596-604.

91. Kim SC, Tonkonogy SL, Albright CA, Tsang J, Balish EJ, Braun J, Huycke MM, Sartor RB. Variable phenotypes of enterocolitis in interleukin 10-deficient mice monoassociated with two different commensal bacteria. Gastroenterology 2005;128:891-906.

92. Orholm M, Binder V, Sorensen TI, Rasmussen LP, Kyvik KO. Concordance of inflammatory bowel disease among Danish twins. Results of a nationwide study. Scand J Gastroenterol 2000;35:1075-1081.

93. Thompson NP, Driscoll R, Pounder RE, Wakefield AJ. Genetics versus environment in inflammatory bowel disease: results of a British twin study. BMJ 1996;312:95-96.

94. Tysk C, Lindberg E, Jarnerot G, Floderus-Myrhed B. Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. Gut 1988;29:990-996.

95. Cavanaugh J. International collaboration provides convincing linkage replication in complex disease through analysis of a large pooled data set: Crohn disease and chromosome 16. Am J Hum Genet 2001;68:1165-1171.

96. Hugot JP, Laurent-Puig P, Gower-Rousseau C, Olson JM, *et al.* Mapping of a susceptibility locus for Crohn's disease on chromosome 16. Nature 1996;379:821-823.

97. Achkar JP, Dassopoulos T, Silverberg MS, Tuvlin JA, *et al.* Phenotype-stratified genetic linkage study demonstrates that IBD2 is an extensive ulcerative colitis locus. Am J Gastroenterol 2006;101:572-580.

98. Parkes M, Barmada MM, Satsangi J, Weeks DE, Jewell DP, Duerr RH. The IBD2 locus shows linkage heterogeneity between ulcerative colitis and Crohn disease. Am J Hum Genet 2000;67:1605-1610.

99. Satsangi J, Parkes M, Louis E, Hashimoto L, Kato N, Welsh K, Terwilliger JD, Lathrop GM, Bell JI, Jewell DP. Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. Nat Genet 1996;14:199-202.

100. Hampe J, Shaw SH, Saiz R, Leysens N, Lantermann A, *et al.* Linkage of inflammatory bowel disease to human chromosome 6p. Am J Hum Genet 1999;65:1647-1655.

101. Satsangi J, Welsh KI, Bunce M, Julier C, Farrant JM, Bell JI, Jewell DP. Contribution of genes of the major histocompatibility complex to susceptibility and disease phenotype in inflammatory bowel disease. Lancet 1996;347:1212-1217.

102. van Heel DA, Fisher SA, Kirby A, Daly MJ, Rioux JD, Lewis CM. Inflammatory bowel disease susceptibility loci defined by genome scan meta-analysis of 1952 affected relative pairs. Hum Mol Genet 2004;13:763-770.

103. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet 2008;40:955-962.

104. Hampe J, Cuthbert A, Croucher PJ, Mirza MM, *et al.* Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. Lancet 2001;357:1925-1928.

105. Lesage S, Zouali H, Cezard JP, Colombel JF, *et al.* CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. Am J Hum Genet 2002;70:845-857.

106. Arnott ID, Nimmo ER, Drummond HE, Fennell J, Smith BR, MacKinlay E, Morecroft J, Anderson N, Kelleher D, O'Sullivan M, McManus R, Satsangi J. NOD2/CARD15, TLR4 and CD14 mutations in Scottish and Irish Crohn's disease patients: evidence for genetic heterogeneity within Europe? Genes Immun 2004;5:417-425.

107. Russell RK, Drummond HE, Nimmo EE, Anderson N, Smith L, Wilson DC, Gillett PM, McGrogan P, Hassan K, Weaver LT, Bisset M, Mahdi G, Satsangi J. Genotype-phenotype analysis in childhood-onset Crohn's disease: NOD2/CARD15 variants consistently predict phenotypic characteristics of severe disease. Inflamm Bowel Dis 2005;11:955-964.

108. Helio T, Halme L, Lappalainen M, Fodstad H, Paavola-Sakki P, Turunen U, Farkkila M, Krusius T, Kontula K. CARD15/NOD2 gene variants are associated with familially occurring and complicated forms of Crohn's disease. Gut 2003;52:558-562.

109. Torkvist L, Noble CL, Lordal M, Sjoqvist U, Lindforss U, Nimmo ER, Russell RK, Lofberg R, Satsangi J. Contribution of CARD15 variants in determining susceptibility to Crohn's disease in Sweden. Scand J Gastroenterol 2006;41:700-705.

110. Inoue N, Tamura K, Kinouchi Y, Fukuda Y, *et al.* Lack of common NOD2 variants in Japanese patients with Crohn's disease. Gastroenterology 2002;123:86-91.

111. Leong RW, Armuzzi A, Ahmad T, Wong ML, Tse P, Jewell DP, Sung JJ. NOD2/CARD15 gene polymorphisms and Crohn's disease in the Chinese population. Aliment Pharmacol Ther 2003;17:1465-1470.

112. Croucher PJ, Mascheretti S, Hampe J, Huse K, *et al.* Haplotype structure and association to Crohn's disease of CARD15 mutations in two ethnically divergent populations. Eur J Hum Genet 2003;11:6-16.

113. Ahmad T, Armuzzi A, Bunce M, Mulcahy-Hawes K, *et al.* The molecular classification of the clinical manifestations of Crohn's disease. Gastroenterology 2002;122:854-866.

114. Cuthbert AP, Fisher SA, Mirza MM, King K, *et al.* The contribution of NOD2 gene mutations to the risk and site of disease in inflammatory bowel disease. Gastroenterology 2002;122:867-874.

115. Kobayashi KS, Chamaillard M, Ogura Y, Henegariu O, Inohara N, Nunez G, Flavell RA. Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. Science 2005;307:731-734.

116. Maeda S, Hsu LC, Liu H, Bankston LA, Iimura M, Kagnoff MF, Eckmann L, Karin M. Nod2 mutation in Crohn's disease potentiates NF-kappaB activity and IL-1beta processing. Science 2005;307:734-738.

117. Lala S, Ogura Y, Osborne C, Hor SY, Bromfield A, Davies S, Ogunbiyi O, Nunez G, Keshav S. Crohn's disease and the NOD2 gene: a role for paneth cells. Gastroenterology 2003;125:47-57.

118. Wehkamp J, Salzman NH, Porter E, Nuding S, *et al.* Reduced Paneth cell alpha-defensins in ileal Crohn's disease. Proc Natl Acad Sci U S A 2005;102:18129-18134.

119. Simms LA, Doecke JD, Walsh MD, Huang N, Fowler EV, Radford-Smith GL. Reduced alpha-defensin expression is associated with inflammation and not NOD2 mutation status in ileal Crohn's disease. Gut 2008;57:903-910.

120. van Heel DA, Ghosh S, Butler M, Hunt KA, *et al.* Muramyl dipeptide and toll-like receptor sensitivity in NOD2-associated Crohn's disease. Lancet 2005;365:1794-1796.

121. Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, *et al.* Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. Am J Hum Genet 2000;66:1863-1870.

122. Rioux JD, Daly MJ, Silverberg MS, Lindblad K, *et al.* Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. Nat Genet 2001;29:223-228.

123. Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. Nat Genet 2001;29:229-232.

124. Armuzzi A, Ahmad T, Ling KL, de SA, Cullen S, van HD, Orchard TR, Welsh KI, Marshall SE, Jewell DP. Genotype-phenotype analysis of the Crohn's disease susceptibility haplotype on chromosome 5q31. Gut 2003;52:1133-1139.

125. Giallourakis C, Stoll M, Miller K, Hampe J, Lander ES, Daly MJ, Schreiber S, Rioux JD. IBD5 is a general risk factor for inflammatory bowel disease: replication of association with Crohn disease and identification of a novel association with ulcerative colitis. Am J Hum Genet 2003;73:205-211.

126. Mirza MM, Fisher SA, King K, Cuthbert AP, *et al.* Genetic evidence for interaction of the 5q31 cytokine locus and the CARD15 gene in Crohn disease. Am J Hum Genet 2003;72:1018-1022.

127. Negoro K, McGovern DP, Kinouchi Y, Takahashi S, Lench NJ, Shimosegawa T, Carey A, Cardon LR, Jewell DP, van Heel DA. Analysis of the IBD5 locus and potential gene-gene interactions in Crohn's disease. Gut 2003;52:541-546.

128. Noble CL, Nimmo ER, Drummond H, Ho GT, Tenesa A, Smith L, Anderson N, Arnott ID, Satsangi J. The contribution of OCTN1/2 variants within the IBD5 locus to disease susceptibility and severity in Crohn's disease. Gastroenterology 2005;129:1854-1864.

129. Palmieri O, Latiano A, Valvano R, D'Inca R, *et al.* Variants of OCTN1-2 cation transporter genes are associated with both Crohn's disease and ulcerative colitis. Aliment Pharmacol Ther 2006;23:497-506.

130. Reinhard C, Rioux JD. Role of the IBD5 susceptibility locus in the inflammatory bowel diseases. Inflamm Bowel Dis 2006;12:227-238.

131. Russell RK, Drummond H, Nimmo E, Anderson N, Noble C, Wilson D, Gillett P, McGrogan P, Hassan K, Weaver L, Bisset M, Mahdi G, Satsangi J. Analysis of the influence of OCTN1/2 variants within the IBD5 locus on disease susceptibility and growth parameters in early-onset inflammatory bowel disease. Gut 2006.

132. Torkvist L, Noble CL, Lordal M, Sjoqvist U, Lindforss U, Nimmo ER, Lofberg R, Russell RK, Satsangi J. Contribution of the IBD5 locus to Crohn's disease in the Swedish population. Scand J Gastroenterol 2007;42:200-206.

133. Waller S, Tremelling M, Bredin F, Godfrey L, Howson J, Parkes M. Evidence for association of OCTN genes and IBD5 with ulcerative colitis. Gut 2006;55:809-814.

134. Peltekova VD, Wintle RF, Rubin LA, Amos CI, *et al.* Functional variants of OCTN cation transporter genes are associated with Crohn disease. Nat Genet 2004;36:471-475.

135. Silverberg MS, Duerr RH, Brant SR, Bromfield G, *et al.* Refined genomic localization and ethnic differences observed for the IBD5 association with Crohn's disease. Eur J Hum Genet 2007;15:328-335.

136. Chang M, Li Y, Yan C, Callis-Duffin KP, Matsunami N, *et al.* Variants in the 5q31 cytokine gene cluster are associated with psoriasis. Genes Immun 2008;9:176-181.

137. Friberg C, Bjorck K, Nilsson S, Inerot A, Wahlstrom J, Samuelsson L. Analysis of chromosome 5q31-32 and psoriasis: confirmation of a susceptibility locus but no association with SNPs within SLC22A4 and SLC22A5. J Invest Dermatol 2006;126:998-1002.

138. Li Y, Chang M, Schrodi SJ, Callis-Duffin KP, *et al.* The 5q31 variants associated with psoriasis and Crohn's disease are distinct. Hum Mol Genet 2008;17:2978-2985.

139. de Jong DJ, Franke B, Naber AH, Willemen JJ, *et al.* No evidence for involvement of IL-4R and CD11B from the IBD1 region and STAT6 in the IBD2 region in Crohn's disease. Eur J Hum Genet 2003;11:884-887.

140. Klein W, Tromm A, Folwaczny C, Hagedorn M, Duerig N, Epplen J, Schmiegel W, Griga T. The G2964A polymorphism of the STAT6 gene in inflammatory bowel disease. Dig Liver Dis 2005;37:159-161.

141. Tumer Z, Croucher PJ, Jensen LR, Hampe J, *et al.* Genomic structure, chromosome mapping and expression analysis of the human AVIL gene, and its exclusion as a candidate for locus for inflammatory bowel disease at 12q13-14 (IBD2). Gene 2002;288:179-185.

142. Silverberg MS, Cho JH, Rioux JD, McGovern DP, *et al.* Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. Nat Genet 2009.

143. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, *et al.* Global variation in copy number in the human genome. Nature 2006;444:444-454.

144. Fellermann K, Stange DE, Schaeffeler E, Schmalzl H, *et al.* A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. Am J Hum Genet 2006;79:439-448.

145. Agostini L, Martinon F, Burns K, McDermott MF, Hawkins PN, Tschopp J. NALP3 forms an IL-1beta-processing inflammasome with increased activity in Muckle-Wells autoinflammatory disorder. Immunity 2004;20:319-325.

146. Mariathasan S, Monack DM. Inflammasome adaptors and sensors: intracellular regulators of infection and inflammation. Nat Rev Immunol 2007;7:31-40.

147. Villani AC, Lemire M, Fortin G, Louis E, *et al.* Common variants in the NLRP3 region contribute to Crohn's disease susceptibility. Nat Genet 2009;41:71-76.

148. Kaser A, Lee AH, Franke A, Glickman JN, Zeissig S, Tilg H, Nieuwenhuis EE, Higgins DE, Schreiber S, Glimcher LH, Blumberg RS. XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease. Cell 2008;134:743-756.

149. Calfon M, Zeng H, Urano F, Till JH, Hubbard SR, Harding HP, Clark SG, Ron D. IRE1 couples endoplasmic reticulum load to secretory capacity by processing the XBP-1 mRNA. Nature 2002;415:92-96.

150. Barmada MM, Brant SR, Nicolae DL, Achkar JP, Panhuysen CI, Bayless TM, Cho JH, Duerr RH. A genome scan in 260 inflammatory bowel disease-affected relative pairs. Inflamm Bowel Dis 2004;10:513-520.

151. Hampe J, Schreiber S, Shaw SH, Lau KF, *et al.* A genomewide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. Am J Hum Genet 1999;64:808-816.

152. Vermeire S, Rutgeerts P, Van Steen K, Joossens S, Claessens G, Pierik M, Peeters M, Vlietinck R. Genome wide scan in a Flemish inflammatory bowel disease population: support for the IBD4 locus, population heterogeneity, and epistasis. Gut 2004;53:980-986.

153. Mathew CG. New links to the pathogenesis of Crohn disease provided by genome-wide association scans. Nat Rev Genet 2008;9:9-14.

154. Kugathasan S, Baldassano RN, Bradfield JP, Sleiman PM, *et al.* Loci on 20q13 and 21q22 are associated with pediatric-onset inflammatory bowel disease. Nat Genet 2008;40:1211-1215.

155. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 2008;9:356-369.

156. Cardon LR. Genetics. Delivering new disease genes. Science 2006;314:1403-1405.

157. Donnelly P. Progress and challenges in genome-wide association studies in humans. Nature 2008;456:728-731.

158. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, *et al.* Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet 2008;40:638-645.

159. Lettre G, Jackson AU, Gieger C, Schumacher FR, *et al.* Identification of ten loci associated with height highlights new biological pathways in human growth. Nat Genet 2008;40:584-591.

160. Weedon MN, Lango H, Lindgren CM, Wallace C, *et al.* Genome-wide association analysis identifies 20 loci that influence adult height. Nat Genet 2008;40:575-583.

161. Houlston RS, Webb E, Broderick P, Pittman AM, *et al.* Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat Genet 2008;40:1426-1435.

162. Tenesa A, Farrington SM, Prendergast JG, Porteous ME, *et al.* Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat Genet 2008;40:631-637.

163. Tomlinson IP, Webb E, Carvajal-Carmona L, Broderick P, *et al.* A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nat Genet 2008;40:623-630.

164. Eeles RA, Kote-Jarai Z, Giles GG, Olama AA, *et al.* Multiple newly identified loci associated with prostate cancer susceptibility. Nat Genet 2008;40:316-321.

165. Gudmundsson J, Sulem P, Steinthorsdottir V, Bergthorsson JT, *et al.* Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. Nat Genet 2007;39:977-983.

166. Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, *et al.* Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat Genet 2007;39:631-637.

167. Gudmundsson J, Sulem P, Rafnar T, Bergthorsson JT, *et al.* Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. Nat Genet 2008;40:281-283.

168. Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, *et al.* Multiple loci identified in a genome-wide association study of prostate cancer. Nat Genet 2008;40:310-315.

169. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, *et al.* Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet 2007;39:645-649.

170. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, *et al.* A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science 2007;316:889-894.

171. Loos RJ, Lindgren CM, Li S, Wheeler E, Zhao JH, *et al.* Common variants near MC4R are associated with fat mass, weight and risk of obesity. Nat Genet 2008;40:768-775.

172. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, *et al.* Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. Nat Genet 2009;41:18-24.

173. Willer CJ, Speliotes EK, Loos RJ, Li S, *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. Nat Genet 2009;41:25-34.

174. Smyth DJ, Plagnol V, Walker NM, Cooper JD, *et al.* Shared and distinct genetic variants in type 1 diabetes and celiac disease. N Engl J Med 2008;359:2767-2777.

175. Altshuler D, Daly M. Guilt beyond a reasonable doubt. Nat Genet 2007;39:813-815.

176. Houlston RS, Webb E, Broderick P, Pittman AM, *et al.* Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat Genet 2008;40:1426-1435.

177. Cummings JR, Ahmad T, Geremia A, Beckly J, Cooney R, Hancock L, Pathan S, Guo C, Cardon LR, Jewell DP. Contribution of the novel inflammatory bowel disease gene IL23R to disease susceptibility and phenotype. Inflamm Bowel Dis 2007;13:1063-1068.

178. Tremelling M, Cummings F, Fisher SA, Mansfield J, *et al.* IL23R variation determines susceptibility but not disease phenotype in inflammatory bowel disease. Gastroenterology 2007;132:1657-1664.

179. Van Limbergen J, Russell RK, Nimmo ER, Drummond HE, Smith L, Davies G, Anderson NH, Gillett PM, McGrogan P, Hassan K, Weaver L, Bisset WM, Mahdi G, Wilson DC, Satsangi J. IL23R Arg381Gln is associated with childhood onset inflammatory bowel disease in Scotland. Gut 2007;56:1173-1174.

180. Raelson JV, Little RD, Ruether A, Fournier H, *et al.* Genome-wide association study for Crohn's disease in the Quebec Founder Population identifies multiple validated disease loci. Proc Natl Acad Sci U S A 2007;104:14747-14752.

181. Borgiani P, Perricone C, Ciccacci C, Romano S, Novelli G, Biancone L, Petruzziello C, Pallone F. Interleukin-23R Arg381Gln is associated with susceptibility to Crohn's disease but not with phenotype in an Italian population. Gastroenterology 2007;133:1049-1051.

182. Latiano A, Palmieri O, Valvano MR, D'Inca R, *et al.* Replication of interleukin 23 receptor and autophagy-related 16-like 1 association in adult- and pediatric-onset inflammatory bowel disease in Italy. World J Gastroenterol 2008;14:4643-4651.

183. Amre DK, Mack D, Israel D, Morgan K, Lambrette P, *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. Am J Gastroenterol 2008;103:615-620.

184. Weersma RK, Stokkers PC, van Bodegraven AA, van Hogezand RA, *et al.* Molecular prediction of disease risk and severity in a large Dutch Crohn's disease cohort. Gut 2008.

185. Weersma RK, Zhernakova A, Nolte IM, Lefebvre C, *et al.* ATG16L1 and IL23R are associated with inflammatory bowel diseases but not with celiac disease in the Netherlands. Am J Gastroenterol 2008;103:621-627.

186. Marquez A, Mendoza JL, Taxonera C, az-Rubio M, De La Concha EG, Urcelay E, Martinez A. IL23R and IL12B polymorphisms in Spanish IBD patients: no evidence of interaction. Inflamm Bowel Dis 2008;14:1192-1196.

187. Lappalainen M, Halme L, Turunen U, Saavalainen P, Einarsdottir E, Farkkila M, Kontula K, Paavola-Sakki P. Association of IL23R, TNFRSF1A, and HLA-DRB1*0103 allele variants with inflammatory bowel disease phenotypes in the Finnish population. Inflamm Bowel Dis 2008;14:1118-1124.

188. Lakatos PL, Szamosi T, Szilvasi A, Molnar E, *et al.* ATG16L1 and IL23 receptor (IL23R) genes are associated with disease susceptibility in Hungarian CD patients. Dig Liver Dis 2008;40:867-873.

189. Baptista ML, Amarante H, Picheth G, Sdepanian VL, Peterson N, Babasukumar U, Lima HC, Kugathasan S. CARD15 and IL23R influences Crohn's disease susceptibility but not disease phenotype in a Brazilian population. Inflamm Bowel Dis 2008;14:674-679.

190. Dubinsky MC, Wang D, Picornell Y, Wrobel I, Katzir L, Quiros A, Dutridge D, Wahbeh G, Silber G, Bahar R, Mengesha E, Targan SR, Taylor KD, Rotter JI. IL-23 receptor (IL-23R) gene protects against pediatric Crohn's disease. Inflamm Bowel Dis 2007.

191. Van Limbergen JE, Russell RK, Nimmo ER, Drummond HE, Smith L, Anderson NH, Davies G, Gillett PM, McGrogan P, Hassan K, Weaver LT, Bisset MW, Mahdi G, Wilson DC, Satsangi J. IL23R Arg381Gln is associated with childhood onset inflammatory bowel disease in Scotland. Gut 2007.

192. Burton PR, Clayton DG, Cardon LR, Craddock N, *et al.* Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nat Genet 2007;39:1329-1337.

193. Rueda B, Orozco G, Raya E, Fernandez-Sueiro JL, Mulero J, Blanco FJ, Vilches C, Gonzalez-Gay MA, Martin J. The IL23R Arg381Gln non-synonymous polymorphism confers susceptibility to ankylosing spondylitis. Ann Rheum Dis 2008;67:1451-1454.

194. Cargill M, Schrodi SJ, Chang M, Garcia VE, *et al.* A large-scale genetic association study confirms IL12B and leads to the identification of IL23R as psoriasis-risk genes. Am J Hum Genet 2007;80:273-290.

195. Filer C, Ho P, Smith RL, Griffiths C, Young HS, Worthington J, Bruce IN, Barton A. Investigation of association of the IL12B and IL23R genes with psoriatic arthritis. Arthritis Rheum 2008;58:3705-3709.

196. Huffmeier U, Lascorz J, Bohm B, Lohmann J, *et al.* Genetic Variants of the IL-23R Pathway: Association with Psoriatic Arthritis and Psoriasis Vulgaris, but No Specific Risk Factor for Arthritis. J Invest Dermatol 2008.

197. Nair RP, Ruether A, Stuart PE, Jenisch S, Tejasvi T, *et al.* Polymorphisms of the IL12B and IL23R genes are associated with psoriasis. J Invest Dermatol 2008;128:1653-1661.

198. Chang M, Saiki RK, Cantanese JJ, Lew D, *et al.* The inflammatory disease-associated variants in IL12B and IL23R are not associated with rheumatoid arthritis. Arthritis Rheum 2008;58:1877-1881.

199. Park JH, Kim YJ, Park BL, Bae JS, Shin HD, Bae SC. Lack of association between interleukin 23 receptor gene polymorphisms and rheumatoid arthritis susceptibility. Rheumatol Int 2008.

200. Rueda B, Broen J, Torres O, Simeon C, Ortega-Centeno N, *et al.* The Interleukin 23 Receptor gene does not confer risk to systemic sclerosis and is not associated with SSc disease phenotype. Ann Rheum Dis 2008.

201. Hue S, Ahern P, Buonocore S, Kullberg MC, Cua DJ, McKenzie BS, Powrie F, Maloy KJ. Interleukin-23 drives innate and T cell-mediated intestinal inflammation. J Exp Med 2006;203:2473-2483.

202. Kullberg MC, Jankovic D, Feng CG, Hue S, Gorelick PL, McKenzie BS, Cua DJ, Powrie F, Cheever AW, Maloy KJ, Sher A. IL-23 plays a key role in Helicobacter hepaticus-induced T cell-dependent colitis. J Exp Med 2006;203:2485-2494.

203. Oppmann B, Lesley R, Blom B, Timans JC, Xu Y, *et al.* Novel p19 protein engages IL-12p40 to form a cytokine, IL-23, with biological activities similar as well as distinct from IL-12. Immunity 2000;13:715-725.

204. Wiekowski MT, Leach MW, Evans EW, Sullivan L, *et al.* Ubiquitous transgenic expression of the IL-23 subunit p19 induces multiorgan inflammation, runting, infertility, and premature death. J Immunol 2001;166:7563-7570.

205. Bettelli E, Kuchroo VK. IL-12- and IL-23-induced T helper cell subsets: birds of the same feather flock together. J Exp Med 2005;201:169-171.

206. Harrington LE, Hatton RD, Mangan PR, Turner H, Murphy TL, Murphy KM, Weaver CT. Interleukin 17-producing CD4+ effector T cells develop via a lineage distinct from the T helper type 1 and 2 lineages. Nat Immunol 2005;6:1123-1132.

207. McKenzie BS, Kastelein RA, Cua DJ. Understanding the IL-23-IL-17 immune pathway. Trends Immunol 2006;27:17-23.

208. Puccetti P, Belladonna ML, Grohmann U. Effects of IL-12 and IL-23 on antigen-presenting cells at the interface between innate and adaptive immunity. Crit Rev Immunol 2002;22:373-390.

209. Mannon PJ, Fuss IJ, Mayer L, Elson CO, Sandborn WJ, *et al.* Anti-interleukin-12 antibody for active Crohn's disease. N Engl J Med 2004;351:2069-2079.

210. Deretic V. Autophagy as an immune defense mechanism. Curr Opin Immunol 2006;18:375-382.

211. Cadwell K, Liu JY, Brown SL, Miyoshi H, Loh J, *et al.* A key role for autophagy and the autophagy gene Atg16l1 in mouse and human intestinal Paneth cells. Nature 2008.

212. Saitoh T, Fujita N, Jang MH, Uematsu S, Yang BG, *et al.* Loss of the autophagy protein Atg16L1 enhances endotoxin-induced IL-1beta production. Nature 2008.

213. McCarroll SA, Huett A, Kuballa P, Chilewski SD, *et al.* Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. Nat Genet 2008.

214. Taylor GA, Feng CG, Sher A. p47 GTPases: regulators of immunity to intracellular pathogens. Nat Rev Immunol 2004;4:100-109.

215. Bekpen C, Hunn JP, Rohde C, Parvanova I, Guethlein L, Dunn DM, Glowalla E, Leptin M, Howard JC. The interferon-inducible p47 (IRG) GTPases in vertebrates: loss of the cell autonomous resistance mechanism in the human lineage. Genome Biology 2005;6.

216. Singh SB, Davis AS, Taylor GA, Deretic V. Human IRGM induces autophagy to eliminate intracellular mycobacteria. Science 2006;313:1438-1441.

217. MacMicking JD, Taylor GA, McKinney JD. Immune control of tuberculosis by IFN-gamma-inducible LRG-47. Science 2003;302:654-659.

218. Feng CG, Collazo-Custodio CM, Eckhaus M, Hieny S, Belkaid Y, Elkins K, Jankovic D, Taylor GA, Sher A. Mice deficient in LRG-47 display increased susceptibility to mycobacterial infection associated with the induction of lymphopenia. J Immunol 2004;172:1163-1168.

219. Sanjuan MA, Dillon CP, Tait SW, Moshiach S, Dorsey F, Connell S, Komatsu M, Tanaka K, Cleveland JL, Withoff S, Green DR. Toll-like receptor signalling in macrophages links the autophagy pathway to phagocytosis. Nature 2007;450:1253-1257.

220. Suzuki T, Franchi L, Toma C, Ashida H, Ogawa M, Yoshikawa Y, Mimuro H, Inohara N, Sasakawa C, Nunez G. Differential regulation of caspase-1 activation, pyroptosis, and autophagy via Ipaf and ASC in Shigella-infected macrophages. PLoS Pathog 2007;3:e111.

221. Sartor RB. Does Mycobacterium avium subspecies paratuberculosis cause Crohn's disease? Gut 2005;54:896-898.

222. Massey DC, Bredin F, Parkes M. Use of sirolimus (rapamycin) to treat refractory Crohn's disease. Gut 2008;57:1294-1296.

223. Farkas S, Hornung M, Sattler C, Guba M, Steinbauer M, Anthuber M, Herfarth H, Schlitt HJ, Geissler EK. Rapamycin decreases leukocyte migration in vivo and effectively reduces experimentally induced chronic colitis. Int J Colorectal Dis 2006;21:747-753.

224. Yamazaki K, Takahashi A, Takazoe M, Kubo M, Onouchi Y, Fujino A, Kamatani N, Nakamura Y, Hata A. Positive association of genetic variants in the upstream region of NKX2-3 with Crohn's disease in Japanese patients. Gut 2009;58:228-232.

225. Pabst O, Schneider A, Brand T, Arnold HH. The mouse Nkx2-3 homeodomain gene is expressed in gut mesenchyme during pre- and postnatal mouse development. Dev Dyn 1997;209:29-35.

226. Wang CC, Biben C, Robb L, Nassir F, Barnett L, Davidson NO, Koentgen F, Tarlinton D, Harvey RP. Homeodomain factor Nkx2-3 controls regional expression of leukocyte homing coreceptor MAdCAM-1 in specialized endothelial cells of the viscera. Dev Biol 2000;224:152-167.

227. Pabst O, Zweigerdt R, Arnold HH. Targeted disruption of the homeobox transcription factor Nkx2-3 in mice results in postnatal lethality and abnormal development of small intestine and spleen. Development 1999;126:2215-2225.

228. Ibarra-Sanchez MJ, Simoncic PD, Nestel FR, Duplay P, Lapp WS, Tremblay ML. The T-cell protein tyrosine phosphatase. Semin Immunol 2000;12:379-386.

229. You-Ten KE, Muise ES, Itie A, Michaliszyn E, Wagner J, Jothy S, Lapp WS, Tremblay ML. Impaired bone marrow microenvironment and immune function in T cell protein tyrosine phosphatase-deficient mice. J Exp Med 1997;186:683-693.

230. Simoncic PD, Bourdeau A, Lee-Loy A, Rohrschneider LR, Tremblay ML, Stanley ER, McGlade CJ. T-cell protein tyrosine phosphatase (Tcptp) is a negative regulator of colony-stimulating factor 1 signaling and macrophage differentiation. Mol Cell Biol 2006;26:4149-4160.

231. Heinonen KM, Nestel FP, Newell EW, Charette G, Seemayer TA, Tremblay ML, Lapp WS. T-cell protein tyrosine phosphatase deletion results in progressive systemic inflammatory disease. Blood 2004;103:3457-3464.

232. Galic S, Klingler-Hoffmann M, Fodero-Tavoletti MT, Puryer MA, Meng TC, Tonks NK, Tiganis T. Regulation of insulin receptor signaling by the protein tyrosine phosphatase TCPTP. Mol Cell Biol 2003;23:2096-2108.

233. Tiganis T, Kemp BE, Tonks NK. The protein-tyrosine phosphatase TCPTP regulates epidermal growth factor receptor-mediated and phosphatidylinositol 3-kinase-dependent signaling. J Biol Chem 1999;274:27768-27775.

234. Simoncic PD, Lee-Loy A, Barber DL, Tremblay ML, McGlade CJ. The T cell protein tyrosine phosphatase is a negative regulator of janus family kinases 1 and 3. Curr Biol 2002;12:446-453.

235. ten HJ, de J, I, Fu Y, Zhu W, Tremblay M, David M, Shuai K. Identification of a nuclear Stat1 protein tyrosine phosphatase. Mol Cell Biol 2002;22:5662-5668.

236. Yamamoto T, Sekine Y, Kashima K, Kubota A, Sato N, Aoki N, Matsuda T. The nuclear isoform of protein-tyrosine phosphatase TC-PTP regulates interleukin-6-mediated signaling

pathway through STAT3 dephosphorylation. Biochem Biophys Res Commun 2002;297:811-817.

237. Lu X, Nechushtan H, Ding F, Rosado MF, Singal R, Alizadeh AA, Lossos IS. Distinct IL-4-induced gene expression, proliferation, and intracellular signaling in germinal center B-cell-like and activated B-cell-like diffuse large-cell lymphomas. Blood 2005;105:2924-2932.

238. Lu X, Chen J, Sasmono RT, Hsi ED, Sarosiek KA, Tiganis T, Lossos IS. TCPTP, Distinctively Expressed in ABC-Like Diffuse Large B-Cell Lymphomas, is the Nuclear Phosphatase of STAT6. Mol Cell Biol 2007.

239. van VC, Bukczynska PE, Puryer MA, Sadek CM, Shields BJ, Tremblay ML, Tiganis T. Selective regulation of tumor necrosis factor-induced Erk signaling by Src family kinases and the T cell protein tyrosine phosphatase. Nat Immunol 2005;6:253-260.

240. Klingler-Hoffmann M, Fodero-Tavoletti MT, Mishima K, *et al.* The protein tyrosine phosphatase TCPTP suppresses the tumorigenicity of glioblastoma cells expressing a mutant epidermal growth factor receptor. J Biol Chem 2001;276:46313-46318.

241. Nakazawa A, Dotan I, Brimnes J, Allez M, Shao L, Tsushima F, Azuma M, Mayer L. The expression and function of costimulatory molecules B7H and B7-H1 on colonic epithelial cells. Gastroenterology 2004;126:1347-1357.

242. Ito T, Yang M, Wang YH, Lande R, Gregorio J, Perng OA, Qin XF, Liu YJ, Gilliet M. Plasmacytoid dendritic cells prime IL-10-producing T regulatory cells by inducible costimulator ligand. J Exp Med 2007;204:105-115.

243. Komiya T, Tanigawa Y, Hirohashi S. Cloning of the novel gene intelectin, which is expressed in intestinal paneth cells in mice. Biochem Biophys Res Commun 1998;251:759-762.

244. Voehringer D, Stanley SA, Cox JS, Completo GC, Lowary TL, Locksley RM. Nippostrongylus brasiliensis: identification of intelectin-1 and -2 as Stat6-dependent genes expressed in lung and intestine during infection. Exp Parasitol 2007;116:458-466.

245. Carolan BJ, Harvey BG, De BP, Vanni H, Crystal RG. Decreased expression of intelectin 1 in the human airway epithelium of smokers compared to nonsmokers. J Immunol 2008;181:5760-5767.

246. Plowey ED, Cherra SJ, III, Liu YJ, Chu CT. Role of autophagy in G2019S-LRRK2-associated neurite shortening in differentiated SH-SY5Y cells. J Neurochem 2008;105:1048-1056.

247. Gianfrancesco F, Esposito T, Ombra MN, Forabosco P, *et al.* Identification of a novel gene and a common variant associated with uric acid nephrolithiasis in a Sardinian genetic isolate. Am J Hum Genet 2003;72:1479-1491.

248. Li Y, Liao W, Chang M, Schrodi SJ, Bui N, Catanese JJ, *et al.* Further Genetic Evidence for Three Psoriasis-Risk Genes: ADAM33, CDKAL1, and PTPN22. J Invest Dermatol 2008.

249. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 2007;316:1331-1336.

250. Satsangi J, Welsh KI, Bunce M, Julier C, Farrant JM, Bell JI, Jewell DP. Contribution of genes of the major histocompatibility complex to susceptibility and disease phenotype in inflammatory bowel disease. Lancet 1996;347:1212-1217.

251. Bouma G, Crusius JB, Garcia-Gonzalez MA, Meijer BU, Hellemans HP, Hakvoort RJ, Schreuder GM, Kostense PJ, Meuwissen SG, Pena AS. Genetic markers in clinically well defined patients with ulcerative colitis (UC). Clin Exp Immunol JID - 0057202 1999;115:294-300.

252. Roussomoustakaki M, Satsangi J, Welsh K, Louis E, Fanning G, Targan S, Landers C, Jewell DP. Genetic markers may predict disease behavior in patients with ulcerative colitis. Gastroenterology JID - 0374630 1997;112:1845-1853.

253. Duerr RH, Chesny IJ. Associations between HLA-DR alleles and subsets of ulcerative colitis defined by extent of colitis. 1997.

254. Toyoda H, Wang SJ, Yang HY, Redford A, Magalong D, Tyan D, McElree CK, Pressman SR, Shanahan F, Targan SR. Distinct associations of HLA class II genes with inflammatory bowel disease. Gastroenterology JID - 0374630 1993;104:741-748.

255. Futami S, Aoyama N, Honsako Y, Tamura T, Morimoto S, Nakashima T, Ohmoto A, Okano H, Miyamoto M, Inaba H. HLA-DRB1*1502 allele, subtype of DR15, is associated with susceptibility to ulcerative colitis and its progression. Dig Dis Sci JID - 7902782 1995;40:814-818.

256. Sugimura K, Asakura H, Mizuki N, Inoue M, Hibi T, Yagita A, Tsuji K, Inoko H. Analysis of genes within the HLA region affecting susceptibility to ulcerative colitis. Hum Immunol JID - 8010936 1993;36:112-118.

257. Stokkers PC, Reitsma PH, Tytgat GN, van Deventer SJ. HLA-DR and -DQ phenotypes in inflammatory bowel disease: a meta-analysis. Gut 1999;45:395-401.

258. Panwala CM, Jones JC, Viney JL. A novel model of inflammatory bowel disease: mice deficient for the multiple drug resistance gene, mdr1a, spontaneously develop colitis. J Immunol 1998;161:5733-5744.

259. Langmann T, Moehle C, Mauerer R, Scharl M, Liebisch G, Zahn A, Stremmel W, Schmitz G. Loss of detoxification in inflammatory bowel disease: dysregulation of pregnane X receptor target genes. Gastroenterology 2004;127:26-40.

260. Ho GT, Soranzo N, Nimmo ER, Tenesa A, Goldstein DB, Satsangi J. ABCB1/MDR1 gene determines susceptibility and phenotype in ulcerative colitis: discrimination of critical variants using a gene-wide haplotype tagging approach. Hum Mol Genet 2006;15:797-805.

261. van Bodegraven AA, Curley CR, Hunt KA, Monsuur AJ, *et al.* Genetic Variation in Myosin IXB Is Associated With Ulcerative Colitis. Gastroenterology 2006;131:1768-1774.

262. Latiano A, Palmieri O, Valvano MR, D'Inca R, Caprilli R, Cucchiara S, Sturniolo GC, Bossa F, Andriulli A, Annese V. The association of MYO9B gene in Italian patients with inflammatory bowel diseases. Aliment Pharmacol Ther 2008;27:241-248.

263. Monsuur AJ, de Bakker PI, Alizadeh BZ, Zhernakova A, *et al.* Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect. Nat Genet 2005;37:1341-1344.

264. Schreiber S, Heinig T, Thiele HG, Raedler A. Immunoregulatory role of interleukin 10 in patients with inflammatory bowel disease. Gastroenterology 1995;108:1434-1444.

265. Schreiber S, Fedorak RN, Nielsen OH, Wild G, Williams NC, Jacyna M, Lashner BA, Cohard M, Kilian A, LeBeaut A, Hanauer SB. A safety and efficacy study of recombinant human interleukin-10 (rHuIL-10) treatment in 329 patients with chronic active Crohn's disease (CACD). Gastroenterology 1998;114:A1080.

266. Fedorak RN, Gangl A, Elson CO, Rutgeerts P, Schreiber S, *et al.* Recombinant human interleukin 10 in the treatment of patients with mild to moderately active Crohn's disease. The Interleukin 10 Inflammatory Bowel Disease Cooperative Study Group. Gastroenterology 2000;119:1473-1482.

267. Schreiber S, Fedorak RN, Nielsen OH, Wild G, *et al.* Safety and efficacy of recombinant human interleukin 10 in chronic active Crohn's disease. Crohn's Disease IL-10 Cooperative Study Group. Gastroenterology 2000;119:1461-1472.

268. Steidler L, Hans W, Schotte L, Neirynck S, Obermeier F, Falk W, Fiers W, Remaut E. Treatment of murine colitis by Lactococcus lactis secreting interleukin-10. Science 2000;289:1352-1355.

269. Aujla SJ, Dubin PJ, Kolls JK. Th17 cells and mucosal host defense. Semin Immunol 2007;19:377-382.

270. Lees C, Howie S, Sartor RB, Satsangi J. The hedgehog signalling pathway in the gastrointestinal tract: implications for development, homeostasis, and disease. Gastroenterology 2005;129:1696-1710.

271. Nusslein-Volhard C, Wieschaus E. Mutations affecting segment number and polarity in Drosophila. Nature 1980;287:795-801.

272. Ingham PW, McMahon AP. Hedgehog signaling in animal development: paradigms and principles. Genes Dev 2001;15:3059-3087.

273. Pathi S, Pagan-Westphal S, Baker DP, Garber EA, Rayhorn P, Bumcrot D, Tabin CJ, Blake PR, Williams KP. Comparative biological responses to human Sonic, Indian, and Desert hedgehog. Mech Dev 2001;106:107-117.

274. Bellaiche Y, The I, Perrimon N. Tout-velu is a Drosophila homologue of the putative tumour suppressor EXT-1 and is needed for Hh diffusion. Nature 1998;394:85-88.

275. Han C, Belenkaya TY, Wang B, Lin X. Drosophila glypicans control the cell-to-cell movement of Hedgehog by a dynamin-independent process. Development 2004;131:601-611.

276. Mann RK, Beachy PA. Novel lipid modifications of secreted protein signals. Annu Rev Biochem 2004;73:891-923.

277. Carpenter D, Stone DM, Brush J, Ryan A, Armanini M, Frantz G, Rosenthal A, De Sauvage FJ. Characterization of two patched receptors for the vertebrate hedgehog protein family. Proc Natl Acad Sci U S A 1998;95:13630-13634.

278. Motoyama J, Takabatake T, Takeshima K, Hui C. Ptch2, a second mouse Patched gene is co-expressed with Sonic hedgehog. Nat Genet 1998;18:104-106.

279. Lum L, Beachy PA. The Hedgehog Response Network: Sensors, Switches, and Routers. Science 2004;304:1755-1759.

280. Bumcrot DA, Takada R, McMahon AP. Proteolytic processing yields two secreted forms of sonic hedgehog. Mol Cell Biol 1995;15:2294-2303.

281. Lee JJ, Ekker SC, von Kessler DP, Porter JA, Sun BI, Beachy PA. Autoproteolysis in hedgehog protein biogenesis. Science 1994;266:1528-1537.

282. Porter JA, Young KE, Beachy PA. Cholesterol modification of hedgehog signaling proteins in animal development. Science 1996;274:255-259.

283. Pepinsky RB, Zeng C, Wen D, Rayhorn P, Baker DP, *et al.* Identification of a palmitic acid-modified form of human Sonic hedgehog. J Biol Chem 1998;273:14037-14045.

284. Burke R, Nellen D, Bellotto M, Hafen E, Senti KA, Dickson BJ, Basler K. Dispatched, a novel sterol-sensing domain protein dedicated to the release of cholesterol-modified hedgehog from signaling cells. Cell 1999;99:803-815.

285. Tian H, Jeong J, Harfe BD, Tabin CJ, McMahon AP. Mouse Disp1 is required in sonic hedgehog-expressing cells for paracrine activity of the cholesterol-modified ligand. Development 2005;132:133-142.

286. Bijlsma MF, Spek CA, Peppelenbosch MP. Hedgehog: an unusual signal transducer. Bioessays 2004;26:387-394.

287. Kalderon D. Similarities between the Hedgehog and Wnt signaling pathways. Trends Cell Biol 2002;12:523-531.

288. Nusse R. Wnts and Hedgehogs: lipid-modified proteins and similarities in signaling mechanisms at the cell surface. Development 2003;130:5297-5305.

289. Chuang PT, McMahon AP. Vertebrate Hedgehog signalling modulated by induction of a Hedgehog-binding protein. Nature 1999;397:617-621.

290. Chuang PT, Kawcak T, McMahon AP. Feedback control of mammalian Hedgehog signaling by the Hedgehog-binding protein, Hip1, modulates Fgf signaling during branching morphogenesis of the lung. Genes Dev 2003;17:342-347.

291. Kawahira H, Ma NH, Tzanakakis ES, McMahon AP, Chuang PT, Hebrok M. Combined activities of hedgehog signaling inhibitors regulate pancreas development. Development 2003;130:4871-4879.

292. McCarthy RA, Barth JL, Chintalapudi MR, Knaak C, Argraves WS. Megalin functions as an endocytic sonic hedgehog receptor. J Biol Chem 2002;277:25660-25667.

293. McCarthy RA, Argraves WS. Megalin and the neurodevelopmental biology of sonic hedgehog and retinol. J Cell Sci 2003;116:955-960.

294. Jobin C, Sartor RB. The I kappa B/NF-kappa B system: a key determinant of mucosalinflammation and protection. Am J Physiol Cell Physiol 2000;278:C451-C462.

295. De Santa BP, Van Den Brink GR, Roberts DJ. Development and differentiation of the intestinal epithelium. Cell Mol Life Sci 2003;60:1322-1332.

296. Roberts DJ. Molecular mechanisms of development of the gastrointestinal tract. Dev Dyn 2000;219:109-120.

297. Harmon EB, Ko AH, Kim SK. Hedgehog signaling in gastrointestinal development and disease. Curr Mol Med 2002;2:67-82.

298. Litingtung Y, Lei L, Westphal H, Chiang C. Sonic hedgehog is essential to foregut development. Nat Genet 1998;20:58-61.

299. Ramalho-Santos M, Melton DA, McMahon AP. Hedgehog signals regulate multiple aspects of gastrointestinal development. Development 2000;127:2763-2772.

300. Roberts DJ, Johnson RL, Burke AC, Nelson CE, Morgan BA, Tabin C. Sonic hedgehog is an endodermal signal inducing Bmp-4 and Hox genes during induction and regionalization of the chick hindgut. Development 1995;121:3163-3174.

301. Zhang J, Rosenthal A, De Sauvage FJ, Shivdasani RA. Downregulation of Hedgehog signaling is required for organogenesis of the small intestine in Xenopus. Dev Biol 2001;229:188-202.

302. Wang LC, Nassir F, Liu ZY, Ling L, Kuo F, Crowell T, Olson D, Davidson NO, Burkly LC. Disruption of hedgehog signaling reveals a novel role in intestinal morphogenesis and intestinal-specific lipid metabolism in mice. Gastroenterology 2002;122:469-482.

303. Madison BB, Braunstein K, Kuizon E, Portman K, Qiao XT, Gumucio DL. Epithelial hedgehog signals pattern the intestinal crypt-villus axis. Development 2005;132:279-289.

304. Hebrok M, Kim SK, St Jacques B, McMahon AP, Melton DA. Regulation of pancreas development by hedgehog signaling. Development 2000;127:4905-4913.

305. Thomas MK, Rastalsky N, Lee JH, Habener JF. Hedgehog signaling regulation of insulin production by pancreatic beta-cells. Diabetes 2000;49:2039-2047.

306. Heemskerk J, DiNardo S. Drosophila hedgehog acts as a morphogen in cellular patterning. Cell 1994;76:449-460.

307. McMahon AP, Ingham PW, Tabin CJ. Developmental roles and clinical significance of hedgehog signaling. Curr Top Dev Biol 2003;53:1-114.

308. Bitgood MJ, McMahon AP. Hedgehog and Bmp genes are coexpressed at many diverse sites of cell-cell interaction in the mouse embryo. Dev Biol 1995;172:126-138.

309. Hebrok M, Kim SK, Melton DA. Notochord repression of endodermal Sonic hedgehog permits pancreas development. Genes Dev 1998;12:1705-1713.

310. Bitgood MJ, Shen L, McMahon AP. Sertoli cell signaling by Desert hedgehog regulates the male germline. Curr Biol 1996;6:298-304.

311. Mo R, Kim JH, Zhang J, Chiang C, Hui CC, Kim PC. Anorectal malformations caused by defects in sonic hedgehog signaling. Am J Pathol 2001;159:765-774.

312. Bose J, Grotewold L, Ruther U. Pallister-Hall syndrome phenotype in mice mutant for Gli3. Hum Mol Genet 2002;11:1129-1135.

313. Park HL, Bai C, Platt KA, Matise MP, Beeghly A, Hui CC, Nakashima M, Joyner AL. Mouse Gli1 mutants are viable but have defects in SHH signaling in combination with a Gli2 mutation. Development 2000;127:1593-1605.

314. Sukegawa A, Narita T, Kameda T, Saitoh K, Nohno T, Iba H, Yasugi S, Fukuda K. The concentric structure of the developing gut is regulated by Sonic hedgehog derived from endodermal epithelium. Development 2000;127:1971-1980.

315. Fu M, Lui VC, Sham MH, Pachnis V, Tam PK. Sonic hedgehog regulates the proliferation, differentiation, and migration of enteric neural crest cells in gut. J Cell Biol 2004;166:673-684.

316. Garcia-Barcelo MM, Lee WS, Sham MH, Lui VC, Tam PK. Is there a role for the IHH gene in Hirschsprung's disease? Neurogastroenterol Motil 2003;15:663-668.

317. Wright NA. Epithelial stem cell repertoire in the gut: clues to the origin of cell lineages, proliferative units and cancer. Int J Exp Pathol 2000;81:117-143.

318. Van Den Brink GR, Bleuming SA, Hardwick JC, Schepman BL, *et al.* Indian Hedgehog is an antagonist of Wnt signaling in colonic epithelial cell differentiation. Nat Genet 2004;36:277-282.

319. Buhman KK, Wang LC, Tang Y, Swietlicki EA, Kennedy S, *et al.* Inhibition of hedgehog signaling protects adult mice from diet-induced weight gain. J Nutr 2004;134:2979-2984.

320. Motoyama J, Liu J, Mo R, Ding Q, Post M, Hui CC. Essential function of Gli2 and Gli3 in the formation of lung, trachea and oesophagus. Nat Genet 1998;20:54-57.

321. Arsic D, Qi BQ, Beasley SW. Hedgehog in the human: a possible explanation for the VATER association. J Paediatr Child Health 2002;38:117-121.

322. Arsic D, Keenan J, Quan QB, Beasley S. Differences in the levels of Sonic hedgehog protein during early foregut development caused by exposure to Adriamycin give clues to the role of the Shh gene in oesophageal atresia. Pediatr Surg Int 2003;19:463-466.

323. Ioannides AS, Henderson DJ, Spitz L, Copp AJ. Role of Sonic hedgehog in the development of the trachea and oesophagus. J Pediatr Surg 2003;38:29-36.

324. Spilde TL, Bhatia AM, Mehta S, Ostlie DJ, Hembree MJ, Preuett BL, Prasadan K, Li Z, Snyder CL, Gittes GK. Defective sonic hedgehog signaling in esophageal atresia with tracheoesophageal fistula. Surgery 2003;134:345-350.

325. Spilde T, Bhatia A, Ostlie D, Marosky J, Holcomb G, III, Snyder C, Gittes G. A role for sonic hedgehog signaling in the pathogenesis of human tracheoesophageal fistula. J Pediatr Surg 2003;38:465-468.

326. Koebbe MJ, Golden JA, Bennett G, Finnell RH, Mackler SA. Effects of prenatal cocaine exposure on embryonic expression of sonic hedgehog. Teratology 1999;59:12-19.

327. Van Den Brink GR, Hardwick JC, Nielsen C, Xu C, *et al.* Sonic hedgehog expression correlates with fundic gland differentiation in the adult gastrointestinal tract. Gut 2002;51:628-633.

328. Dimmler A, Brabletz T, Hlubek F, Hafner M, Rau T, Kirchner T, Faller G. Transcription of sonic hedgehog, a potential factor for gastric morphogenesis and gastric mucosa maintenance, is up-regulated in acidic conditions. Lab Invest 2003;83:1829-1837.

329. Nielsen CM, Williams J, Van Den Brink GR, Lauwers GY, Roberts DJ. Hh pathway expression in human gut tissues and in inflammatory gut diseases. Lab Invest 2004;84:1631-1642.

330. Oniscu A, James RM, Morris RG, Bader S, Malcomson RD, Harrison DJ. Expression of Sonic hedgehog pathway genes is altered in colonic neoplasia. J Pathol 2004;203:909-917.

331. Van Den Brink GR, Hardwick JC, Tytgat GN, Brink MA, Ten Kate FJ, Van Deventer SJ, Peppelenbosch MP. Sonic hedgehog regulates gastric gland morphogenesis in man and mouse. Gastroenterology 2001;121:317-328.

332. Varnat F, Heggeler BB, Grisel P, Boucard N, Corthesy-Theulaz I, *et al.* PPARbeta/delta Regulates Paneth Cell Differentiation Via Controlling the Hedgehog Signaling Pathway. Gastroenterology 2006;131:538-553.

333. Lees CW, Satsangi J. Hedgehog, paneth cells, and colon cancer: a cautionary note for the use of systemic agonists/antagonists. Gastroenterology 2006;131:1657-1658.

334. van den Brink GR. Hedgehog signaling in development and homeostasis of the gastrointestinal tract. Physiol Rev 2007;87:1343-1375.

335. Lai K, Robertson MJ, Schaffer DV. The sonic hedgehog signaling system as a bistable genetic switch. Biophys J 2004;86:2748-2757.

336. Duman-Scheel M, Weng L, Xin S, Du W. Hedgehog regulates cell growth and proliferation by inducing Cyclin D and Cyclin E. Nature 2002;417:299-304.

337. Mill P, Mo R, Fu H, Grachtchouk M, Kim PC, Dlugosz AA, Hui CC. Sonic hedgehog-dependent activation of Gli2 is essential for embryonic hair follicle development. Genes Dev 2003;17:282-294.

338. Barnes EA, Kong M, Ollendorff V, Donoghue DJ. Patched1 interacts with cyclin B1 to regulate cell cycle progression. EMBO J 2001;20:2214-2223.

339. Fan H, Khavari PA. Sonic hedgehog opposes epithelial cell cycle arrest. J Cell Biol 1999;147:71-76.

340. Mehlen P, Rabizadeh S, Snipas SJ, Assa-Munt N, Salvesen GS, Bredesen DE. The DCC gene product induces apoptosis by a mechanism requiring receptor proteolysis. Nature 1998;395:801-804.

341. Thibert C, Teillet MA, Lapointe F, Mazelin L, Le Douarin NM, Mehlen P. Inhibition of neuroepithelial patched-induced apoptosis by sonic hedgehog. Science 2003;301:843-846.

342. Mehlen P, Thibert C. Dependence receptors: between life and death. Cell Mol Life Sci 2004;61:1854-1866.

343. Ruiz IA, Palma V, Dahmane N. Hedgehog-Gli signalling and the growth of the brain. Nat Rev Neurosci 2002;3:24-33.

344. Altaba A, Stecca B, Sanchez P. Hedgehog--Gli signaling in brain tumors: stem cells and paradevelopmental programs in cancer. Cancer Lett 2004;204:145-157.

345. Berman DM, Karhadkar SS, Hallahan AR, Pritchard JI, Eberhart CG, *et al.* Medulloblastoma growth inhibition by hedgehog pathway blockade. Science 2002;297:1559-1561.

346. Berman DM, Karhadkar SS, Maitra A, Montes DO, Gerstenblith MR, *et al.* Widespread requirement for Hedgehog ligand stimulation in growth of digestive tract tumours. Nature 2003;425:846-851.

347. Mazzola CA, Pollack IF. Medulloblastoma. Curr Treat Options Neurol 2003;5:189-198.

348. Thayer SP, Di Magliano MP, Heiser PW, Nielsen CM, *et al.* Hedgehog is an early and late mediator of pancreatic cancer tumorigenesis. Nature 2003;425:851-856.

349. Watkins DN, Berman DM, Burkholder SG, Wang B, Beachy PA, Baylin SB. Hedgehog signalling within airway epithelial progenitors and in small-cell lung cancer. Nature 2003;422:313-317.

350. Hahn H, Wicking C, Zaphiropoulous PG, Gailani MR, *et al.* Mutations of the human homolog of Drosophila patched in the nevoid basal cell carcinoma syndrome. Cell 1996;85:841-851.

351. Hahn H, Wojnowski L, Zimmer AM, Hall J, Miller G, Zimmer A. Rhabdomyosarcomas and radiation hypersensitivity in a mouse model of Gorlin syndrome. Nat Med 1998;4:619-622.

352. Kappler R, Calzada-Wack J, Schnitzbauer U, Koleva M, *et al.* Molecular characterization of Patched-associated rhabdomyosarcoma. J Pathol 2003;200:348-356.

353. Aboulkassim TO, LaRue H, Lemieux P, Rousseau F, Fradet Y. Alteration of the PATCHED locus in superficial bladder cancer. Oncogene 2003;22:2967-2971.

354. Maesawa C, Tamura G, Iwaya T, Ogasawara S, Ishida K, Sato N, *et al.* Mutations in the human homologue of the Drosophila patched gene in esophageal squamous cell carcinoma. Genes Chromosomes Cancer 1998;21:276-279.

355. Reifenberger J, Wolter M, Weber RG, Megahed M, Ruzicka T, Lichter P, Reifenberger G. Missense mutations in SMOH in sporadic basal cell carcinomas of the skin and primitive neuroectodermal tumors of the central nervous system. Cancer Res 1998;58:1798-1803.

356. Taylor MD, Liu L, Raffel C, Hui CC, Mainprize TG, *et al.* Mutations in SUFU predispose to medulloblastoma. Nat Genet 2002;31:306-310.

357. Kinzler KW, Bigner SH, Bigner DD, Trent JM, Law ML, O'Brien SJ, Wong AJ, Vogelstein B. Identification of an amplified, highly expressed gene in a human glioma. Science 1987;236:70-73.

358. Karhadkar SS, Bova GS, Abdallah N, Dhara S, Gardner D, Maitra A, Isaacs JT, Berman DM, Beachy PA. Hedgehog signalling in prostate regeneration, neoplasia and metastasis. Nature 2004;431:707-712.

359. Qualtrough D, Buda A, Gaffield W, Williams AC, Paraskeva C. Hedgehog signalling in colorectal tumour cells: induction of apoptosis with cyclopamine treatment. Int J Cancer 2004;110:831-837.

360. Chatel G, Ganeff C, Boussif N, Delacroix L, Briquet A, Nolens G, Winkler R. Hedgehog signaling pathway is inactive in colorectal cancer cell lines. Int J Cancer 2007;121:2622-2627.

361. Yauch RL, Gould SE, Scales SJ, Tang T, Tian H, Ahn CP, *et al.* A paracrine requirement for hedgehog signalling in cancer. Nature 2008;455:406-410.

362. Altaba A. Therapeutic inhibition of Hedgehog-GLI signaling in cancer: epithelial, stromal, or stem cell targets? Cancer Cell 2008;14:281-283.

363. Crompton T, Outram SV, Hager-Theodorides AL. Sonic hedgehog signalling in T-cell development and activation. Nat Rev Immunol 2007;7:726-735.

364. Outram SV, Hager-Theodorides AL, Shah DK, Rowbotham NJ, Drakopoulou E, Ross SE, Lanske B, Dessens JT, Crompton T. Indian hedgehog (Ihh) both promotes and restricts thymocyte differentiation. Blood 2008.

365. Lowrey JA, Stewart GA, Lindey S, Hoyne GF, Dallman MJ, Howie SE, Lamb JR. Sonic hedgehog promotes cell cycle progression in activated peripheral CD4(+) T lymphocytes. J Immunol 2002;169:1869-1875.

366. Stewart GA, Lowrey JA, Wakelin SJ, Fitch PM, Lindey S, Dallman MJ, Lamb JR, Howie SE. Sonic hedgehog signaling modulates activation of and cytokine production by human peripheral CD4+ T cells. J Immunol 2002;169:5451-5457.

367. Stewart GA, Hoyne GF, Ahmad SA, Jarman E, Wallace WA, Harrison DJ, Haslett C, Lamb JR, Howie SE. Expression of the developmental Sonic hedgehog (Shh) signalling pathway is up-regulated in chronic lung fibrosis and the Shh receptor patched 1 is present in circulating T lymphocytes. J Pathol 2003;199:488-495.

368. Benson RA, Lowrey JA, Lamb JR, Howie SE. The Notch and Sonic hedgehog signalling pathways in immunity. Mol Immunol 2004;41:715-725.

369. Wang J, Pham-Mitchell N, Schindler C, Campbell IL. Dysregulated Sonic hedgehog signaling and medulloblastoma consequent to IFN-alpha-stimulated STAT2-independent production of IFN-gamma in the brain. J Clin Invest 2003;112:535-543.

370. Varas A, Hernandez-Lopez C, Valencia J, Mattavelli S, *et al.* Survival and function of human thymic dendritic cells are dependent on autocrine Hedgehog signaling. J Leukoc Biol 2008;83:1476-1483.

371. Nakashima H, Nakamura M, Yamaguchi H, Yamanaka N, *et al.* Nuclear Factor-{kappa}B Contributes to Hedgehog Signaling Pathway Activation through Sonic Hedgehog Induction in Pancreatic Cancer. Cancer Res 2006;66:7041-7049.

372. Schmidt-Ullrich R, Tobin DJ, Lenhard D, Schneider P, Paus R, Scheidereit C. NF-kappaB transmits Eda A1/EdaR signalling to activate Shh and cyclin D1 expression, and controls post-initiation hair placode down growth. Development 2006;133:1045-1057.

373. Kasperczyk H, Baumann B, Debatin KM, Fulda S. Characterization of sonic hedgehog as a novel NF-{kappa}B target gene that promotes NF-{kappa}B-mediated apoptosis resistance and tumor growth in vivo. FASEB J 2008.

374. Pola R, Ling LE, Aprahamian TR, Barban E, Bosch-Marce M, Curry C, Corbley M, Kearney M, Isner JM, Losordo DW. Postnatal recapitulation of embryonic hedgehog pathway in response to skeletal muscle ischemia. Circulation 2003;108:479-485.

375. Miyaji T, Nakase T, Iwasaki M, Kuriyama K, Tamai N, Higuchi C, Myoui A, Tomita T, Yoshikawa H. Expression and distribution of transcripts for sonic hedgehog in the early phase of fracture repair. Histochem Cell Biol 2003;119:233-237.

376. Wang Y, Imitola J, Rasmussen S, O'Connor KC, Khoury SJ. Paradoxical dysregulation of the neural stem cell pathway sonic hedgehog-Gli1 in autoimmune encephalomyelitis and multiple sclerosis. Ann Neurol 2008;64:417-427.

377. Fendrich V, Esni F, Garay MV, Feldmann G, Habbe N, Jensen JN, Dor Y, Stoffers D, Jensen J, Leach SD, Maitra A. Hedgehog signaling is required for effective regeneration of exocrine pancreas. Gastroenterology 2008;135:621-631.

378. Kayed H, Kleeff J, Keleg S, Buchler MW, Friess H. Distribution of Indian hedgehog and its receptors patched and smoothened in human chronic pancreatitis. J Endocrinol 2003;178:467-478.

379. Jung Y, Brown KD, Witek RP, Omenetti A, Yang L, Vandongen M, Milton RJ, Hines IN, Rippe RA, Spahr L, Rubbia-Brandt L, Diehl AM. Accumulation of hedgehog-responsive progenitors parallels alcoholic liver disease severity in mice and humans. Gastroenterology 2008;134:1532-1543.

380. Omenetti A, Popov Y, Jung Y, Choi SS, Witek RP, Yang L, Brown KD, Schuppan D, Diehl AM. The Hedgehog Pathway Regulates Remodeling Responses to Biliary Obstruction in Rats. Gut 2008.

381. Oswari J, Matthay MA, Margulies SS. Keratinocyte growth factor reduces alveolar epithelial susceptibility to in vitro mechanical deformation. Am J Physiol Lung Cell Mol Physiol 2001;281:L1068-L1077.

382. Upadhyay D, Bundesmann M, Panduri V, Correa-Meyer E, Kamp DW. Fibroblast Growth Factor-10 Attenuates H2O2-Induced Alveolar Epithelial Cell DNA Damage: Role of MAPK Activation and DNA Repair. Am J Respir Cell Mol Biol 2004;31:107-113.

383. Gudjonsson JE, Aphale A, Grachtchouk M, Ding J, Nair RP, Wang T, Voorhees JJ, Dlugosz AA, Elder JT. Lack of Evidence for Activation of the Hedgehog Pathway in Psoriasis. J Invest Dermatol 2008.

384. Endo H, Momota Y, Oikawa A, Shinkai H. Psoriatic skin expresses the transcription factor Gli1: possible contribution of decreased neurofibromin expression. Br J Dermatol 2006;154:619-623.

385. Tas S, Avci O. Rapid clearance of psoriatic skin lesions induced by topical cyclopamine. A preliminary proof of concept study. Dermatology 2004;209:126-131.

386. Fleig SV, Choi SS, Yang L, Jung Y, Omenetti A, Vandongen HM, Huang J, Sicklick JK, Diehl AM. Hepatic accumulation of Hedgehog-reactive progenitors increases with severity of fatty liver damage in mice. Lab Invest 2007;87:1227-1239.

387. Greenbaum LE. Hedgehog signaling in biliary fibrosis. J Clin Invest 2008;118:3263-3265.

388. Jung Y, Brown KD, Witek RP, Omenetti A, Yang L, Vandongen M, Milton RJ, Hines IN, Rippe RA, Spahr L, Rubbia-Brandt L, Diehl AM. Accumulation of hedgehog-responsive progenitors parallels alcoholic liver disease severity in mice and humans. Gastroenterology 2008;134:1532-1543.

389. Omenetti A, Yang L, Li YX, McCall SJ, Jung Y, Sicklick JK, Huang J, Choi S, Suzuki A, Diehl AM. Hedgehog-mediated mesenchymal-epithelial interactions modulate hepatic response to bile duct ligation. Lab Invest 2007;87:499-514.

390. Omenetti A, Porrello A, Jung Y, Yang L, Popov Y, Choi SS, *et al.* Hedgehog signaling regulates epithelial-mesenchymal transition during biliary fibrosis in rodents and humans. J Clin Invest 2008;118:3331-3342.

391. Omenetti A, Popov Y, Jung Y, Choi SS, Witek RP, Yang L, Brown KD, Schuppan D, Diehl AM. The hedgehog pathway regulates remodelling responses to biliary obstruction in rats. Gut 2008;57:1275-1282.

392. Witek RP, Yang L, Liu R, Jung Y, Omenetti A, Syn WK, Choi SS, Cheong Y, Fearing CM, Agboola KM, Chen W, Diehl AM. Liver Cell-Derived Microparticles Activate Hedgehog Signaling and Alter Gene Expression in Hepatic Endothelial Cells. Gastroenterology 2008.

393. Beachy PA, Karhadkar SS, Berman DM. Mending and malignancy. Nature 2004;431:402.

394. Nusse R, Varmus HE. Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. Cell 1982;31:99-109.

395. Nusse R, Brown A, Papkoff J, Scambler P, Shackleford G, McMahon A, Moon R, Varmus H. A new nomenclature for int-1 and related genes: the Wnt gene family. Cell 1991;64:231.

396. Clevers H. Wnt/beta-catenin signaling in development and disease. Cell 2006;127:469-480.

397. Logan CY, Nusse R. The Wnt signaling pathway in development and disease. Annu Rev Cell Dev Biol 2004;20:781-810.

398. Gregorieff A, Pinto D, Begthel H, Destree O, Kielman M, Clevers H. Expression pattern of Wnt signaling components in the adult intestine. Gastroenterology 2005;129:626-638.

399. van Es JH, Jay P, Gregorieff A, Van Gijn ME, Jonkheer S, Hatzis P, Thiele A, van den BM, Begthel H, Brabletz T, Taketo MM, Clevers H. Wnt signalling induces maturation of Paneth cells in intestinal crypts. Nat Cell Biol 2005;7:381-386.

400. Korinek V, Barker N, Moerer P, van DE, Huls G, Peters PJ, Clevers H. Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. Nat Genet 1998;19:379-383.

401. Reya T, Clevers H. Wnt signalling in stem cells and cancer. Nature 2005;434:843-850.

402. Staal FJ, Luis TC, Tiemessen MM. WNT signalling in the immune system: WNT is spreading its wings. Nat Rev Immunol 2008;8:581-593.

403. Jeannet G, Scheller M, Scarpellino L, Duboux S, Gardiol N, Back J, Kuttler F, Malanchi I, Birchmeier W, Leutz A, Huelsken J, Held W. Long-term, multilineage hematopoiesis occurs in the combined absence of beta-catenin and gamma-catenin. Blood 2008;111:142-149.

404. Ding Y, Shen S, Lino AC, Curotto de Lafaille MA, Lafaille JJ. Beta-catenin stabilization extends regulatory T cell survival and induces anergy in nonregulatory T cells. Nat Med 2008;14:162-169.

405. Blumenthal A, Ehlers S, Lauber J, Buer J, Lange C, Goldmann T, Heine H, Brandt E, Reiling N. The Wingless homolog WNT5A and its receptor Frizzled-5 regulate inflammatory responses of human mononuclear cells induced by microbial stimulation. Blood 2006;108:965-973.

406. Lehtonen A, Ahlfors H, Veckman V, Miettinen M, Lahesmaa R, Julkunen I. Gene expression profiling during differentiation of human monocytes to macrophages or dendritic cells. J Leukoc Biol 2007;82:710-720.

407. Apelqvist A, Ahlgren U, Edlund H. Sonic hedgehog directs specialised mesoderm differentiation in the intestine and pancreas. Curr Biol 1997;7:801-804.

408. Goodrich LV, Milenkovic L, Higgins KM, Scott MP. Altered neural cell fates and medulloblastoma in mouse patched mutants. Science 1997;277:1109-1113.

409. Vortkamp A, Pathi S, Peretti GM, Caruso EM, Zaleske DJ, Tabin CJ. Recapitulation of signals regulating embryonic bone formation during postnatal growth and in fracture repair. Mech Dev 1998;71:65-76.

410. El Andaloussi A, Graves S, Meng F, Mandal M, Mashayekhi M, Aifantis I. Hedgehog signaling controls thymocyte progenitor homeostasis and differentiation in the thymus. Nat Immunol 2006;7:418-426.

411. Hallikas O, Palin K, Sinjushina N, Rautiainen R, Partanen J, Ukkonen E, Taipale J. Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. Cell 2006;124:47-59.

412. Yamamoto-Furusho JK, Barnich N, Xavier R, Hisamatsu T, Podolsky DK. Centaurin Beta 1 down-regulates NOD1 and NOD2-dependent NF-kB activation. J Biol Chem 2006.

413. Di ML, Ferretti E, De SE, Argenti B, Mincione C, *et al.* REN(KCTD11) is a suppressor of Hedgehog signaling and is deleted in human medulloblastoma. Proc Natl Acad Sci U S A 2004;101:10833-10838.

414. Katz J, Savin R, Spiro HM. The basal cell nevus syndrome and inflammatory disease of the bowel. Am J Med 1968;44:483-488.

415. Panhuysen CI, Karban A, Knodle MA, Bayless TM, Duerr RH, Bailey-Wilson JE, Epstein EH, Jr., Brant SR. Identification of genetic loci for basal cell nevus syndrome and inflammatory bowel disease in a single large pedigree. Hum Genet 2006;120:31-41.

416. Fujii K, Miyashita T, Omata T, Kobayashi K, Takanashi J, Kouchi K, Yamada M, Kohno Y. Gorlin syndrome with ulcerative colitis in a Japanese girl. Am J Med Genet 2003;121A:65-68.

417. Akiyoshi T, Nakamura M, Koga K, Nakashima H, Yao T, Tsuneyoshi M, Tanaka M, Katano M. Gli1, downregulated in colorectal cancers, inhibits proliferation of colon cancer cells involving Wnt signalling activation. Gut 2006;55:991-999.

418. Hoseong YS, Andl T, Grachtchouk V, Wang A, Liu J, Syu LJ, Ferris J, Wang TS, Glick AB, Millar SE, Dlugosz AA. Pathological responses to oncogenic Hedgehog signaling in skin are dependent on canonical Wnt/beta-catenin signaling. Nat Genet 2008.

419. Katoh Y, Katoh M. WNT antagonist, SFRP1, is Hedgehog signaling target. Int J Mol Med 2006;17:171-175.

420. He J, Sheng T, Stelter AA, Li C, Zhang X, Sinha M, Luxon BA, Xie J. Suppressing Wnt signaling by the hedgehog pathway through sFRP-1. J Biol Chem 2006;281:35598-35602.

421. Wehkamp J, Wang G, Kubler I, Nuding S, Gregorieff A, Schnabel A, Kays RJ, Fellermann K, Burk O, Schwab M, Clevers H, Bevins CL, Stange EF. The Paneth cell alpha-defensin deficiency of ileal Crohn's disease is linked to Wnt/Tcf-4. J Immunol 2007;179:3109-3118.

422. Chan VS, Chau SY, Tian L, Chen Y, Kwong SK, Quackenbush J, Dallman M, Lamb J, Tam PK. Sonic hedgehog promotes CD4+ T lymphocyte proliferation and modulates the expression of a subset of CD28-targeted genes. Int Immunol 2006.

423. Lees CW, Zacharias WJ, Tremelling M, Noble CL, Nimmo ER, Tenesa A, Cornelius J, Torkvist L, Kao J, Farrington S, Drummond HE, Ho GT, Arnott ID, Appelman HD, Diehl L, Campbell H, Dunlop MG, Parkes M, Howie SE, Gumucio DL, Satsangi J. Analysis of germline GLI1 variation implicates hedgehog signalling in the regulation of intestinal inflammatory pathways. PLoS Med 2008;5:e239.

424. Lennard-Jones JE. Classification of inflammatory bowel disease. Scand J Gastroenterol Suppl 1989;170:2-6.

425. Gasche C, Scholmerich J, Brynskov J, D'Haens G, *et al.* A simple classification of Crohn's disease: report of the Working Party for the World Congresses of Gastroenterology, Vienna 1998. Inflamm Bowel Dis 2000;6:8-15.

426. Silverberg MS, Satsangi J, Ahmad T, Arnott ID, *et al.* Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: Report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. Can J Gastroenterol 2005;19 Suppl A:5-36.

427. Van Den Brink GR, Peppelenbosch MP. Expression of hedgehog pathway components in the adult colon. Gastroenterology 2006;130:619.

428. Noble CL, Abbas AR, Cornelius J, Lees CW, Ho GT, Toy K, Modrusan Z, Pal N, Zhong F, Chalasani S, Clark H, Arnott ID, Penman ID, Satsangi J, Diehl L. Regional variation in gene expression in the healthy colon is dysregulated in ulcerative colitis. Gut 2008;57:1398-1405.

429. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 1988;16:1215.

430. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. Nat Genet 2005;37:1217-1223.

431. A haplotype map of the human genome. Nature 2005;437:1299-1320.

432. Denman DW, III, Schlesselman JJ. Interval estimation of the attributable risk for multiple exposure levels in case-control studies. Biometrics 1983;39:185-192.

433. Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J Natl Cancer Inst 2004;96:434-442.

434. Newton-Cheh C, Hirschhorn JN. Genetic association studies of complex traits: design and analysis issues. Mutat Res 2005;573:54-69.

435. Birkenkamp-Demtroder K, Olesen SH, Sorensen FB, Laurberg S, Laiho P, Aaltonen LA, Orntoft TF. Differential gene expression in colon cancer of the caecum versus the sigmoid and rectosigmoid. Gut 2005;54:374-384.

436. Zhu Y, James RM, Peter A, Lomas C, Cheung F, Harrison DJ, Bader SA. Functional Smoothened is required for expression of GLI3 in colorectal carcinoma cells. Cancer Lett 2004;207:205-214.

437. Van Den Brink GR, Hardwick JC. Hedgehog Wnteraction in colorectal cancer. Gut 2006;55:912-914.

438. Doble BW, Patel S, Wood GA, Kockeritz LK, Woodgett JR. Functional redundancy of GSK-3alpha and GSK-3beta in Wnt/beta-catenin signaling shown by using an allelic series of embryonic stem cell lines. Dev Cell 2007;12:957-971.

439. Sacedon R, Diez B, Nunez V, Hernandez-Lopez C, Gutierrez-Frias C, Cejalvo T, Outram SV, Crompton T, Zapata AG, Vicente A, Varas A. Sonic hedgehog is produced by follicular dendritic cells and protects germinal center B cells from apoptosis. J Immunol 2005;174:1456-1461.

440. Koslowski MJ, Wang GX, Kuebler IE, Gersemann M, Fellermann K, Stange E, Wehkamp J. Selective influence of Tcf-4 mediated wnt signaling on intestinal innate and adaptive immunity of ileal Crohn's disease. Gastroenterology 2008;134:A514.

441. You J, Nguyen AV, Albers CG, Lin F, Holcombe RF. Wnt pathway-related gene expression in inflammatory bowel disease. Dig Dis Sci 2008;53:1013-1019.

442. Vitart V, Carothers AD, Hayward C, Teague P, Hastie ND, Campbell H, Wright AF. Increased level of linkage disequilibrium in rural compared with urban communities: a factor to consider in association-study design. Am J Hum Genet 2005;76:763-772.

443. Wacholder S, Rothman N, Caporaso N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. J Natl Cancer Inst 2000;92:1151-1158.

444. Thomas DC, Clayton DG. Betting odds and genetic associations. J Natl Cancer Inst 2004;96:421-423.

445. Hunt KA, Zhernakova A, Turner G, Heap GA, Franke L, *et al.* Newly identified genetic risk variants for celiac disease related to the immune response. Nat Genet 2008;40:395-402.

446. Vermeire S, Peeters M, Vlietinck R, Parkes M, Satsangi J, Jewell D, Rutgeerts P. Exclusion of linkage of Crohn's disease to previously reported regions on chromosomes 12, 7, and 3 in the Belgian population indicates genetic heterogeneity. Inflamm Bowel Dis 2000;6:165-170.

447. Yoon JW, Liu CZ, Yang JT, Swart R, Iannaccone P, Walterhouse D. GLI activates transcription through a herpes simplex viral protein 16-like activation domain. J Biol Chem 1998;273:3496-3501.

448. Huntzicker EG, Estay IS, Zhen H, Lokteva LA, Jackson PK, Oro AE. Dual degradation signals control Gli protein stability and tumor formation. Genes Dev 2006;20:276-281.

449. Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV, Gottesman MM. A "silent" polymorphism in the MDR1 gene changes substrate specificity. Science 2007;315:525-528.

450. Belloni E, Muenke M, Roessler E, Traverso G, Siegel-Bartelt J, *et al.* Identification of Sonic hedgehog as a candidate gene responsible for holoprosencephaly. Nat Genet 1996;14:353-356.

451. Roessler E, Belloni E, Gaudenz K, Jay P, Berta P, Scherer SW, Tsui LC, Muenke M. Mutations in the human Sonic Hedgehog gene cause holoprosencephaly. Nat Genet 1996;14:357-360.

452. Roessler E, Du YZ, Mullor JL, Casas E, Allen WP, Gillessen-Kaesbach G, Roeder ER, Ming JE, Ruiz IA, Muenke M. Loss-of-function mutations in the human GLI2 gene are associated with pituitary anomalies and holoprosencephaly-like features. Proc Natl Acad Sci U S A 2003.

453. Kang S, Graham JM, Jr., Olney AH, Biesecker LG. GLI3 frameshift mutations cause autosomal dominant Pallister-Hall syndrome. Nat Genet 1997;15:266-268.

454. Vortkamp A, Gessler M, Grzeschik KH. GLI3 zinc-finger gene interrupted by translocations in Greig syndrome families. Nature 1991;352:539-540.

455. Senee V, Chelala C, Duchatelet S, Feng D, Blanc H, Cossec JC, Charon C, Nicolino M, Boileau P, Cavener DR, Bougneres P, Taha D, Julier C. Mutations in GLIS3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism. Nat Genet 2006;38:682-687.

456. St-Jacques B, Hammerschmidt M, McMahon AP. Indian hedgehog signaling regulates proliferation and differentiation of chondrocytes and is essential for bone formation. Genes Dev 1999;13:2072-2086.

457. Kinzler KW, Ruppert JM, Bigner SH, Vogelstein B. The GLI gene is a member of the Kruppel family of zinc finger proteins. Nature 1988;332:371-374.

458. Kogerman P, Grimm T, Kogerman L, Krause D, Unden AB, Sandstedt B, Toftgard R, Zaphiropoulos PG. Mammalian suppressor-of-fused modulates nuclear-cytoplasmic shuttling of Gli-1. Nat Cell Biol 1999;1:312-319.

459. Jemal A, Siegel R, Ward E, Murray T, Xu J, Smigal C, Thun MJ. Cancer statistics, 2006. CA Cancer J Clin 2006;56:106-130.

460. Barker N, van Es JH, Kuipers J, Kujala P, van den BM, Cozijnsen M, Haegebarth A, Korving J, Begthel H, Peters PJ, Clevers H. Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature 2007;449:1003-1007.

461. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 2000;343:78-85.

462. Barnetson RA, Cartwright N, van VA, Haq N, Drew K, Farrington S, Williams N, Warner J, Campbell H, Porteous ME, Dunlop MG. Classification of ambiguous mutations in DNA mismatch repair genes identified in a population-based study of colorectal cancer. Hum Mutat 2008;29:367-374.

463. Tenesa A, Campbell H, Barnetson R, Porteous M, Dunlop M, Farrington SM. Association of MUTYH and colorectal cancer. Br J Cancer 2006;95:239-242.

464. Lesnick TG, Papapetropoulos S, Mash DC, Ffrench-Mullen J, Shehadeh L, de Andrade M, Henley JR, Rocca WA, Ahlskog JE, Maraganore DM. A genomic pathway approach to a complex disease: axon guidance and Parkinson disease. PLoS Genet 2007;3:e98.

465. Jiang L, Xia B, Li J, Ye M, Yan W, Deng C, Ding Y, Luo H, Hou W, Zhao Q, Liu N, Ren H, Hou X, Xu H. Retrospective survey of 452 patients with inflammatory bowel disease in Wuhan city, central China. Inflamm Bowel Dis 2006;12:212-217.

466. Aghazadeh R, Zali MR, Bahari A, Amin K, Ghahghaie F, Firouzi F. Inflammatory bowel disease in Iran: a review of 457 cases. J Gastroenterol Hepatol 2005;20:1691-1695.

467. Morita N, Toki S, Hirohashi T, Minoda T, Ogawa K, Kono S, Tamakoshi A, Ohno Y, Sawada T, Muto T. Incidence and prevalence of inflammatory bowel disease in Japan: nationwide epidemiological survey during the year 1991. J Gastroenterol 1995;30 Suppl 8:1-4.

468. Yang SK, Hong WS, Min YI, Kim HY, Yoo JY, Rhee PL, Rhee JC, Chang DK, Song IS, Jung SA, Park EB, Yoo HM, Lee DK, Kim YK. Incidence and prevalence of ulcerative colitis in

the Songpa-Kangdong District, Seoul, Korea, 1986-1997. J Gastroenterol Hepatol 2000;15:1037-1042.

469. Yang SK, Loftus EV, Jr., Sandborn WJ. Epidemiology of inflammatory bowel disease in Asia. Inflamm Bowel Dis 2001;7:260-270.

470. Schutyser E, Struyf S, Van Damme J. The CC chemokine CCL20 and its receptor CCR6. Cytokine Growth Factor Rev 2003;14:409-426.

471. Hieshima K, Imai T, Opdenakker G, Van Damme J, *et al.* Molecular cloning of a novel human CC chemokine liver and activation-regulated chemokine (LARC) expressed in liver. Chemotactic activity for lymphocytes and gene localization on chromosome 2. J Biol Chem 1997;272:5846-5853.

472. Hromas R, Gray PW, Chantry D, Godiska R, Krathwohl M, Fife K, Bell GI, Takeda J, Aronica S, Gordon M, Cooper S, Broxmeyer HE, Klemsz MJ. Cloning and characterization of exodus, a novel beta-chemokine. Blood 1997;89:3315-3322.

473. Rossi DL, Vicari AP, Franz-Bacon K, McClanahan TK, Zlotnik A. Identification through bioinformatics of two new macrophage proinflammatory human chemokines: MIP-3alpha and MIP-3beta. J Immunol 1997;158:1033-1036.

474. Anderle P, Rumbo M, Sierro F, Mansourian R, Michetti P, Roberts MA, Kraehenbuhl JP. Novel markers of the human follicle-associated epithelium identified by genomic profiling and microdissection. Gastroenterology 2005;129:321-327.

475. Izadpanah A, Dwinell MB, Eckmann L, Varki NM, Kagnoff MF. Regulated MIP-3alpha/CCL20 production by human intestinal epithelium: mechanism for modulating mucosal immunity. Am J Physiol Gastrointest Liver Physiol 2001;280:G710-G719.

476. Kwon JH, Keates S, Bassani L, Mayer LF, Keates AC. Colonic epithelial cells are a major site of macrophage inflammatory protein 3alpha (MIP-3alpha) production in normal colon and inflammatory bowel disease. Gut 2002;51:818-826.

477. Lee HJ, Choi SC, Lee MH, Oh HM, Choi EY, Choi EJ, *et al.* Increased expression of MIP-3alpha/CCL20 in peripheral blood mononuclear cells from patients with ulcerative colitis and its down-regulation by sulfasalazine and glucocorticoid treatment. Inflamm Bowel Dis 2005;11:1070-1079.

478. Baba M, Imai T, Nishimura M, Kakizaki M, Takagi S, Hieshima K, Nomiyama H, Yoshie O. Identification of CCR6, the specific receptor for a novel lymphocyte-directed CC chemokine LARC. J Biol Chem 1997;272:14893-14898.

479. Cook DN, Prosser DM, Forster R, Zhang J, Kuklin NA, *et al.* CCR6 mediates dendritic cell localization, lymphocyte homeostasis, and immune responses in mucosal tissue. Immunity 2000;12:495-503.

480. Varona R, Villares R, Carramolino L, Goya I, Zaballos A, Gutierrez J, Torres M, Martinez A, Marquez G. CCR6-deficient mice have impaired leukocyte homeostasis and altered contact hypersensitivity and delayed-type hypersensitivity responses. J Clin Invest 2001;107:R37-R45.

481. Kaser A, Ludwiczek O, Holzmann S, Moschen AR, Weiss G, Enrich B, Graziadei I, Dunzendorfer S, Wiedermann CJ, Murzl E, Grasl E, Jasarevic Z, Romani N, Offner FA, Tilg H. Increased expression of CCL20 in human inflammatory bowel disease. J Clin Immunol 2004;24:74-85.

482. Hoover DM, Boulegue C, Yang D, Oppenheim JJ, Tucker K, Lu W, Lubkowski J. The structure of human macrophage inflammatory protein-3alpha /CCL20. Linking antimicrobial

and CC chemokine receptor-6-binding activities with human beta-defensins. J Biol Chem 2002;277:37647-37654.

483. Harant H, Eldershaw SA, Lindley IJ. Human macrophage inflammatory protein-3alpha/CCL20/LARC/Exodus/SCYA20 is transcriptionally upregulated by tumor necrosis factor-alpha via a non-standard NF-kappaB site. FEBS Lett 2001;509:439-445.

484. Kwon JH, Keates S, Simeonidis S, Grall F, Libermann TA, Keates AC. ESE-1, an enterocyte-specific Ets transcription factor, regulates MIP-3alpha gene expression in Caco-2 human colonic epithelial cells. J Biol Chem 2003;278:875-884.

485. Nelson RT, Boyd J, Gladue RP, Paradis T, Thomas R, *et al.* Genomic organization of the CC chemokine mip-3alpha/CCL20/larc/exodus/SCYA20, showing gene structure, splice variants, and chromosome localization. Genomics 2001;73:28-37.

486. Sugita S, Kohno T, Yamamoto K, Imaizumi Y, Nakajima H, Ishimaru T, Matsuyama T. Induction of macrophage-inflammatory protein-3alpha gene expression by TNF-dependent NF-kappaB activation. J Immunol 2002;168:5621-5628.

487. Imaizumi Y, Sugita S, Yamamoto K, Imanishi D, Kohno T, Tomonaga M, Matsuyama T. Human T cell leukemia virus type-I Tax activates human macrophage inflammatory protein-3 alpha/CCL20 gene transcription via the NF-kappa B pathway. Int Immunol 2002;14:147-155.

488. Wilson NJ, Boniface K, Chan JR, McKenzie BS, Blumenschein WM, *et al.* Development, cytokine profile and function of human interleukin 17-producing helper T cells. Nat Immunol 2007;8:950-957.

489. Caruso R, Fina D, Peluso I, Stolfi C, Fantini MC, Gioia V, *et al.* A functional role for interleukin-21 in promoting the synthesis of the T-cell chemoattractant, MIP-3alpha, by gut epithelial cells. Gastroenterology 2007;132:166-175.

490. Choi SC, Seo GS, Lee EK, Jun CD, Kim TH, Nah YH, *et al.* Molecular variations in the promoter region of Mip-3a/ccl20 gene and relationship to its mRNA expression in patients with ulcerative colitis. Gastroenterology 2005;128:A137.

491. Clayton D. A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. Am J Hum Genet 1999;65:1170-1177.

492. Cardon LR, Bell JI. Association study designs for complex diseases. Nat Rev Genet 2001;2:91-99.

493. Armitage E, Drummond H, Ghosh S, Ferguson A. Incidence of juvenile-onset Crohn's disease in Scotland. Lancet 1999;353:1496-1497.

494. Shivananda S, Lennard-Jones J, Logan R, Fear N, Price A, Carpenter L, van Blankenstein M. Incidence of inflammatory bowel disease across Europe: is there a difference between north and south? Results of the European Collaborative Study on Inflammatory Bowel Disease (EC-IBD). Gut 1996;39:690-697.

495. Armitage EL, Aldhous MC, Anderson N, Drummond HE, Riemersma RA, Ghosh S, Satsangi J. Incidence of juvenile-onset Crohn's disease in Scotland: association with northern latitude and affluence. Gastroenterology 2004;127:1051-1057.

496. Nerich V, Monnet E, Etienne A, Louafi S, Ramee C, Rican S, Weill A, Vallier N, Vanbockstael V, Auleley GR, Allemand H, Carbonnel F. Geographical variations of inflammatory bowel disease in France: a study based on national health insurance data. Inflamm Bowel Dis 2006;12:218-226.

497. Hugot JP, Alberti C, Berrebi D, Bingen E, Cezard JP. Crohn's disease: the cold chain hypothesis. Lancet 2003;362:2012-2015.

498. Tosa M, Negoro K, Kinouchi Y, Abe H, Nomura E, *et al.* Lack of association between IBD5 and Crohn's disease in Japanese patients demonstrates population-specific differences in inflammatory bowel disease. Scand J Gastroenterol 2006;41:48-53.

499. Yamazaki K, Takazoe M, Tanaka T, Ichimori T, Saito S, Iida A, Onouchi Y, Hata A, Nakamura Y. Association analysis of SLC22A4, SLC22A5 and DLG5 in Japanese patients with Crohn disease. J Hum Genet 2004;49:664-668.

500. Yamazaki K, McGovern D, Ragoussis J, Paolucci M, *et al.* Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn's disease. Hum Mol Genet 2005;14:3499-3506.

501. Dean M, Carrington M, Winkler C, Huttley GA, Smith MW, *et al.* Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. Science 1996;273:1856-1862.

502. Liu R, Paxton WA, Choe S, Ceradini D, Martin SR, Horuk R, MacDonald ME, Stuhlmann H, Koup RA, Landau NR. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. Cell 1996;86:367-377.

503. Samson M, Libert F, Doranz BJ, Rucker J, Liesnard C, Farber CM, *et al.* Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature 1996;382:722-725.

504. Paxton WA, Martin SR, Tse D, O'Brien TR, Skurnick J, VanDevanter NL, Padian N, Braun JF, Kotler DP, Wolinsky SM, Koup RA. Relative resistance to HIV-1 infection of CD4 lymphocytes from persons who remain uninfected despite multiple high-risk sexual exposure. Nat Med 1996;2:412-417.

505. Martinson JJ, Chapman NH, Rees DC, Liu YT, Clegg JB. Global distribution of the CCR5 gene 32-basepair deletion. Nat Genet 1997;16:100-103.

506. Craggs A, Welfare M, Donaldson PT, Mansfield JC. The CC chemokine receptor 5 delta32 mutation is not associated with inflammatory bowel disease (IBD) in NE England. Genes Immun 2001;2:114-116.

507. Eri R, Jonsson JR, Pandeya N, Purdie DM, Clouston AD, *et al.* CCR5-Delta32 mutation is strongly associated with primary sclerosing cholangitis. Genes Immun 2004;5:444-450.

508. Hampe J, Lynch NJ, Daniels S, Bridger S, Macpherson AJ, *et al.* Fine mapping of the chromosome 3p susceptibility locus in inflammatory bowel disease. Gut 2001;48:191-197.

509. Henckaerts L, Fevery J, Van Steenbergen W, Verslype C, *et al.* CC-type chemokine receptor 5-Delta32 mutation protects against primary sclerosing cholangitis. Inflamm Bowel Dis 2006;12:272-277.

510. Paavola P, Helio T, Kiuru M, Halme L, Turunen U, *et al.* Genetic analysis in Finnish families with inflammatory bowel disease supports linkage to chromosome 3p21. Eur J Hum Genet 2001;9:328-334.

511. Rector A, Vermeire S, Thoelen I, Keyaerts E, Struyf F, Vlietinck R, Rutgeerts P, Van Ranst M. Analysis of the CC chemokine receptor 5 (CCR5) delta-32 polymorphism in inflammatory bowel disease. Hum Genet 2001;108:190-193.

512. Satsangi J, Simmons J, Marshall S, Mitchell S, Chapman RW, Welsh KI, Jewell DP. CCR5 Delta 32 polymorphism in inflammatory bowel disease: Further association with ulcerative colitis and with primary sclerosing cholangitis. Gastroenterology 2000;118:A337.

513. Pabst O, Herbrand H, Takuma N, Arnold HH. NKX2 gene expression in neuroectoderm but not in mesendodermally derived structures depends on sonic hedgehog in mouse embryos. Dev Genes Evol 2000;210:47-50.

514. Nenci A, Becker C, Wullaert A, Gareus R, van LG, Danese S, Huth M, Nikolaev A, Neufert C, Madison B, Gumucio D, Neurath MF, Pasparakis M. Epithelial NEMO links innate immunity to chronic intestinal inflammation. Nature 2007;446:557-561.

515. Ho GT, Moodie FM, Satsangi J. Multidrug resistance 1 gene (P-glycoprotein 170): an important determinant in gastrointestinal disease? Gut 2003;52:759-766.

516. Ho GT, Gaya DR, Satsangi J. Multidrug resistance (MDR1) gene in inflammatory bowel disease: a key player? Inflamm Bowel Dis 2005;11:1013-1019.

517. Ho GT, Nimmo ER, Tenesa A, Fennell J, Drummond H, Mowat C, Arnott ID, Satsangi J. Allelic variations of the multidrug resistance gene determine susceptibility and disease behavior in ulcerative colitis. Gastroenterology 2005;128:288-296.

518. Sims-Mourtada J, Izzo JG, Ajani J, Chao KS. Sonic Hedgehog promotes multiple drug resistance by regulation of drug transport. Oncogene 2007;26:5674-5679.

519. Zhang Y, Laterra J, Pomper MG. Hedgehog pathway inhibitor HhAntag691 is a potent inhibitor of ABCG2/BCRP and ABCB1/Pgp. Neoplasia 2009;11:96-101.