

# **MICROARRAY PROFILING OF INFLAMMATORY BOWEL DISEASE**

This thesis has been submitted for the degree of Doctor of Philosophy at the  
University of Leicester

by

**Varsha Kumari Khodiyar, B.Sc**

November 2002

Department of Pathology,  
University of Leicester,  
Leicester, UK

UMI Number: U170425

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U170425

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.  
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against  
unauthorized copying under Title 17, United States Code.



ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

# **ABSTRACT**

## **Microarray Profiling of Inflammatory Bowel Disease**

Varsha Kumari Khodiyar

In this study of inflammatory bowel diseases (IBD, i.e. Crohn's disease and ulcerative colitis), the gene transcription profile of colonic IBD resection specimens were analysed by oligonucleotide microarray analysis. A total of 33,625 genes were profiled across 23 colonic mucosa samples; 5 involved Crohn's disease, 4 uninvolved Crohn's disease, 5 involved ulcerative colitis, 3 uninvolved ulcerative colitis and 6 samples from macroscopically normal areas of colorectal cancer resections (controls).

A number of data-mining tools, encompassing clustering (e.g. hierarchical & K-means) and matrix-based methods were evaluated for the analysis of this microarray data. Mining strategies were formulated, tested and then applied to the data set to identify genes showing interesting and novel expression patterns across the samples. The application of these tools to the data set resulted in the generation of gene expression profiles for Crohn's disease and ulcerative colitis. Genes of interest were annotated using publicly accessible sequence and literature databases. Potential links by previous research in the inflammatory bowel disease field were analysed for selected genes.

Common pathways emerging from the annotation effort and potentially linking together several of the genes of interest were investigated. Specifically, the energy deficiency hypothesis proposed by Roediger in 1980 and the relevance of potential cancer and apoptosis related genes were reviewed with regard to the findings.

## **DEDICATION**

This thesis is dedicated to my parents Kiritbhai and Hansaben Jethwa for their support and encouragement throughout my education and their continued belief in me.

## ACKNOWLEDGMENTS

To my academic supervisor Professor Peter Furness, thank you for your unerring guidance and encouragement throughout this project. Your continued faith in me provided much needed support during times of self-doubt.

To Mike Thomas at the Leicester General Hospital, thanks for your encouragement, support and the clinical samples, without which this project would not have been possible.

I am indebted to AstraZeneca R&D Charnwood for sponsoring this project. Thanks to my industrial supervisors Mark McHale and Philip Shelton at AstraZeneca, your support and scientific input during the experimental phase of the project was invaluable.

To all at the Leicester General Hospital histopathology laboratory, the kindness and openness with which you taught me the art of cutting and staining tissues is very much appreciated.

To Carl, Caroline and the rest of the bioinformatics team at AstraZeneca, thank you for your assistance with the data handling aspects of the project. Thanks to all in the Molecular Biology department at AstraZeneca for your friendship, and for allowing me to invade your lab and freezer space.

I would like to express my gratitude to my family, Dad, Mum, Sudhir, Sonia and Tejal, your support and understanding throughout this project is much appreciated.

Finally to Gitesh, my friend at the start of the project, my husband by the end, your love and support have been vital to the successful completion of this thesis. Your willingness (usually) to share the PC and your exceptional IT skills in recovering documents from black holes were invaluable.

# CONTENTS

<b>Abstract</b> .....	<b>I</b>
<b>Dedication</b> .....	<b>II</b>
<b>Acknowledgments</b> .....	<b>III</b>
<b>Contents</b> .....	<b>IV</b>
<b>List of Figures</b> .....	<b>X</b>
<b>List of Tables</b> .....	<b>XIII</b>
<b>Abbreviations</b> .....	<b>XIV</b>
<b>Chapter One - Introduction</b> .....	<b>1</b>
1.1 Inflammatory Bowel Disease .....	2
1.1.1 Historical Overview .....	2
1.1.2 Clinical Features.....	2
1.1.2.1 Crohn's disease .....	2
1.1.2.2 Ulcerative colitis .....	2
1.1.2.3 Similarities .....	2
1.1.3 Pathological Features .....	3
1.1.3.1 Crohn's disease .....	3
1.1.3.2 Ulcerative colitis .....	5
1.1.3.3 Indeterminate colitis .....	8
1.1.4 Complications .....	8
1.1.4.1 Perforation.....	8
1.1.4.2 Toxic megacolon .....	9
1.1.4.3 Haemorrhage .....	9
1.1.4.4 Cancer .....	9
1.1.4.5 Extra-intestinal complications .....	9
1.2 Therapies .....	10
1.2.1 Nutritional support .....	10
1.2.2 Medication.....	10
1.2.2.1 Sulphasalazine.....	10
1.2.2.2 Steroids.....	10
1.2.2.3 Other anti-inflammatory agents indicated in IBD .....	10
1.2.2.4 Biologic therapies.....	11
1.2.3 Surgery .....	11
1.2.3.1 Crohn's disease .....	11
1.2.3.2 Ulcerative colitis .....	11
1.3 Epidemiology .....	13
1.3.1 Incidence .....	13
1.3.1.1 Crohn's disease .....	13
1.3.1.2 Ulcerative colitis .....	14
1.3.2 Environmental factors .....	14
1.3.2.1 Geography .....	14
1.3.2.2 Sociology.....	14
1.3.3 Lifestyle factors.....	14
1.3.3.1 Western lifestyle.....	14
1.3.3.2 Smoking .....	14
1.3.3.3 Other factors .....	14
1.3.4 Genetic factors .....	15

1.3.4.1 Race.....	15
1.3.4.2 Familial Inheritance.....	15
1.3.4.3 IBD loci.....	15
1.4 The Gastrointestinal Mucosal Immune System in IBD.....	17
1.4.1 Anatomy of the colon wall.....	17
1.4.1.1 The epithelial layer.....	18
1.4.1.2 The lamina propria.....	18
1.4.2 Anatomy of the GALT.....	19
1.4.2.1 M cells.....	19
1.4.2.2 A typical immune response <sup>119</sup> .....	20
1.4.2.3 T lymphocytes.....	21
1.4.2.4 B lymphocytes.....	22
1.4.2.5 Other immune cells.....	22
1.5 Aetiology.....	23
1.5.1 Persistent Infection.....	23
1.5.1.1 Mycobacterium avium subspecies paratuberculosis.....	23
1.5.1.2 Measles.....	24
1.5.3 Genetic predisposition.....	24
1.5.2.1 Defective mucosal barrier.....	24
1.5.2.2 NOD2.....	25
1.5.3 Inappropriate immune response.....	25
1.5.3.1 Autoantibodies.....	26
1.5.3.2 Immunoglobulins <sup>118</sup> .....	26
1.5.3.3 Response to normal gut flora.....	26
1.5.3.4 Immune response suppression.....	26
1.5.3.5 Epithelial cell antigen presentation.....	26
1.6 Aims & Objectives.....	28
1.6.1 Previous gene expression studies.....	29
1.7 Investigating Gene Expression.....	31
1.7.1 Classical methods.....	31
1.7.2 Massively parallel methods.....	31
1.7.2.1 Serial analysis of gene expression.....	31
1.7.2.2 Microarrays.....	31
<b>Chapter Two – Materials &amp; Methods.....</b>	<b>33</b>
2.1 Collection of Tissue Samples.....	34
2.1.1 Materials.....	34
2.1.2 Method.....	34
2.1.3 Clinical history of patients studied.....	34
2.1.4 Pathological features of samples used in microarray analysis.....	35
2.2 Microarray Experiments.....	39
2.2.1 Method Overview.....	39
2.2.2 RNA Isolation.....	40
2.2.2.1 Materials.....	40
2.2.2.2 Total RNA Isolation.....	40
2.2.2.3 Organic extraction of total RNA.....	40
2.2.2.4 mRNA isolation.....	41
2.2.2.5 Organic extraction of mRNA.....	41
2.2.2.6 Measuring RNA quality and quantity.....	41
2.2.3 cDNA Synthesis from mRNA.....	42
2.2.3.1 Materials.....	42
2.2.3.2 1 <sup>st</sup> strand cDNA synthesis.....	43
2.2.3.3 2 <sup>nd</sup> strand synthesis cDNA synthesis.....	43
2.2.3.4 Organic extraction of cDNA.....	44
2.2.4 Removal of mRNA isolation step.....	44
2.2.4.1 Introduction.....	44

2.2.4.2 Materials.....	44
2.2.4.3 cDNA Synthesis from total RNA .....	45
2.2.4.4 Generation of target RNA .....	45
2.2.4.7 Results .....	46
2.2.4.8 Conclusions .....	47
2.2.5 Modification of total RNA isolation method.....	47
2.2.5.1 Materials.....	47
2.2.5.2 Total RNA isolation .....	47
2.2.5.3 RNeasy cleanup of total RNA .....	48
2.2.5.4 Organic extraction of total RNA .....	48
2.2.6 <i>In vitro</i> transcription (IVT) .....	49
2.2.6.1 Materials.....	49
2.2.6.2 Method .....	49
2.2.7 cRNA Purification.....	49
2.2.7.1 Materials.....	50
2.2.7.2 Method .....	50
2.2.7.3 Results & Discussion.....	50
2.2.8 cRNA Fragmentation .....	51
2.2.8.1 Reagents .....	51
2.2.8.2 Method .....	51
2.2.9 Hybridisation to Microarrays .....	52
2.2.9.1 Materials.....	52
2.2.9.2 Reagents .....	53
2.2.9.3 Preparation of hybridisation cocktail.....	53
2.2.9.4 Hybridisation to probe arrays .....	53
2.2.9.5 Reuse of hybridisation cocktail .....	53
2.2.10 Staining the Microarrays .....	54
2.2.10.1 Materials.....	54
2.2.10.2 Reagents .....	54
2.2.10.3 Standard wash method.....	54
2.2.10.4 Antibody amplification wash method.....	54
2.2.11 Generation & Analysis of the Microarray Image .....	55
2.2.11.1 Scanning the arrays .....	55
2.2.11.2 Image Analysis by Affymetrix GeneChip Software.....	56
2.2.11.3 The Average Difference and the Absolute Call.....	57
2.2.12 Normalisation .....	57
2.2.12.1 Normalisation during target preparation.....	57
2.2.12.2 Normalisation to Housekeeping Genes .....	57
2.2.12.3 Global Normalisation .....	58
2.2.13 Verification of the microarray data .....	58
2.3 Immunocytochemistry.....	60
2.3.1 Method Overview.....	60
2.3.2 Tissue Section Origins .....	60
2.3.3 Antibody Staining .....	61
2.3.3.1 Materials.....	61
2.3.3.2 Pre-treatment methods.....	62
2.3.3.3 Method .....	62
2.3.3.4 Counter-stain removal .....	63
2.3.4 Haematoxylin & Eosin (H&E) Staining.....	63
2.3.5 Volume fraction quantification .....	64
2.3.5.1 Antibody stained sections.....	64
2.3.5.2 H&E stained sections .....	65
2.3.6 Normalisation of array data to ICC data.....	65
2.3.7 ICC results.....	66
<b>Chapter Three – Data Mining.....</b>	<b>68</b>
3.1 Chapter Introduction .....	69
3.2 GeneChip software.....	70

3.2.1 Introduction .....	70
3.2.2 The Gene Expression Matrix.....	70
3.2.3 Querying the Data with DMT .....	70
3.2.4 Viewing the Data with DMT.....	71
3.2.4.1 Scatter Graph.....	71
3.2.4.2 Fold Change Graph .....	71
3.2.4.3 The Bar Graph.....	71
3.2.4.4 The Histogram.....	74
3.2.5 Discussion .....	74
3.3 Matrix Based Mining .....	75
3.3.1 Introduction .....	75
3.3.2 Raw Data.....	75
3.3.2.1 Accession number (column B) .....	75
3.3.2.2 Absolute call & average difference value (columns K to U).....	75
3.3.3 Pooled Sample Data .....	77
3.3.3.1 Mean average difference (column BU) .....	77
3.3.3.2 Absolute call count (columns BL to BM) .....	77
3.3.3.3 Standard Deviation .....	77
3.3.3.4 Standard Error .....	77
3.3.4 Array Comparison Data .....	77
3.3.4.1 Fold Change .....	77
3.3.4.2 Normalisation Method.....	78
3.3.4.3 T-test values (columns CE to CK).....	78
3.3.5 Annotation Data .....	79
3.3.5.1 Incyte Description (column G).....	79
3.3.5.2 Incyte Functional Hierarchy (columns H to I).....	79
3.3.5.3 My Annotation (column A).....	80
3.3.6 Discussion .....	80
3.4 Similarity Metrics (Clustering).....	81
3.4.1 Hierarchical Clustering .....	81
3.4.1.1 Definition .....	81
3.4.1.2 Computational method .....	81
3.4.1.3 Output.....	82
3.4.2 K-means .....	84
3.4.2.1 Definition .....	84
3.4.2.2 Computational method .....	84
3.4.2.3 Output.....	84
3.4.3 Self-Organising Maps .....	85
3.4.3.1 Definition .....	85
3.4.3.2 Computational method .....	85
3.4.3.3 Output.....	85
3.4.4 Principle Component Analysis (PCA).....	85
3.4.4.1 Definition .....	85
3.4.4.2 Computational method .....	85
3.4.4.3 Output.....	85
3.4.5 Evaluation of similarity metric methods .....	86
<b>Chapter Four – Data Mining Results .....</b>	<b>87</b>
4.1 Chapter Introduction .....	88
4.2 Which genes are differentially expressed in the involved IBD samples?.....	91
4.2.1 Rationale .....	91
4.2.2 Methods.....	91
4.2.3 Results & Section summary .....	91
4.3 Identification of an abnormal involved ulcerative colitis sample.....	93
4.3.1 Rationale .....	93
4.3.2 Method & Results .....	93

4.3.2.1 Hierarchical clustering .....	93
4.3.2.2 K-means clustering.....	93
4.3.3 Section summary .....	94
4.4 Which genes are differentially expressed in the involved IBD samples excluding patient UC_5?.....	96
4.4.1 Rationale .....	96
4.4.2 Method .....	96
4.4.3 Over expressed in CDi .....	97
4.4.4 Over expressed in UCi .....	98
4.4.5 Section summary .....	98
4.5 Which genes are specifically differentially regulated in involved IBD?.....	99
4.5.1 Rationale and Method .....	99
4.5.2 CDi specific genes.....	99
4.5.2.1 Opposing expression patterns in CDi and UCi.....	100
4.5.2.2 Highly over expressed CDi genes .....	102
4.5.3 UCi specific genes.....	107
4.5.3.1 Similar expression patterns in UCi and CDi samples.....	108
4.5.3.2 Highly over expressed UCi genes .....	109
4.5.3.3 Highly under expressed UCi genes .....	109
4.5.4 Further annotation of involved IBD specific genes.....	111
4.5.4.1 Signal transduction group.....	111
4.5.4.2 Nucleic acid binding group .....	113
4.5.4.3 Immunity related group.....	115
4.5.4.4 Miscellaneous protein function group .....	118
4.5.4.5 Correlation of IBD loci and chromosomal origin of identified genes .....	121
4.5.5 Section summary .....	121
4.6 Which genes are differentially regulated in the involved IBD tissues as a group?.....	123
4.6.1 Rationale and Method .....	123
4.6.2 Results.....	123
4.6.2.1 Genes with decreased expression in IBDi .....	123
4.6.2.2 Genes with increased expression in IBDi.....	128
4.7 Which genes are specifically differentially regulated between uninvolved IBD and controls? .....	133
4.7.1 Rationale and Method .....	133
4.7.2 Genes with a lower expression level in IBDi tissues (Cluster 7).....	137
4.7.3 Genes with a higher expression level in IBDi tissues (Cluster 18).....	139
4.7.4 Section summary .....	141

**Chapter Five – Discussion of Microarray Data .....** 142

5.1 General Discussion - Study Limitations.....	143
5.1.1 Sample Issues .....	143
5.1.1.1 Collection .....	143
5.1.1.2 Variability.....	143
5.1.2 Microarray Issues .....	144
5.1.2.1 Specificity.....	144
5.1.2.2 Sensitivity.....	144
5.1.2.3 Reproducibility.....	145
5.1.2.4 Variability.....	145
5.1.3 Mining Issues .....	145
5.1.4 Issues specific to this Study.....	146
5.1.4.1 Sample size.....	146
5.1.4.2 Type II errors.....	146
5.1.4.3 Use of steroid refractory samples .....	146
5.1.4.4 Cellular make-up of samples .....	147
5.1.4.5 Gene expression to protein function.....	147
5.2 General Discussion – Microarray Data .....	148
5.2.1 Ulcerative Colitis Energy Deficiency Theory .....	149

5.2.2 Colonic Butyrate .....	149
5.2.2.1 Butyrate Oxidation in Ulcerative Colitis .....	150
5.2.3 Hydrogen Sulphide as an Inhibitor of Butyrate Oxidation.....	150
5.2.4 Sulphate Reducing Bacteria .....	152
5.2.5 Down-regulation of Sulphate Transporters in IBD .....	152
5.2.6 The Proposed Mechanism .....	153
5.2.6.1 5-ASA .....	153
5.2.6.2 The mucus layer .....	154
5.2.6.3 Epithelial shedding .....	155
5.2.7 Discussion of Sulphate Transporter Genes and Energy Deficiency Hypothesis .....	155
5.2.8 Cancer and Regeneration in Inflammatory Bowel Disease .....	157
5.2.8.1 The microarray data.....	157
5.2.8.2 Discussion of cancer related genes.....	158
5.2.9 Apoptosis and Inflammatory Bowel Disease .....	159
5.2.9.1 Apoptotic pathways.....	159
5.2.9.2 Apoptosis Induction .....	160
5.2.9.3 The microarray data.....	161
5.2.9.4 Discussion of apoptosis related genes .....	163
5.3 Achievements .....	164
5.3.1 Objectives reached .....	164
5.3.2 Study design improvements .....	164
5.3.3 Future studies .....	164
<b>Concluding Remarks.....</b>	<b>166</b>
<b>References .....</b>	<b>167</b>
<b>Appendix .....</b>	<b>193</b>

## LIST OF FIGURES

Figure 1.1 – Skip lesions in Crohn's disease (modified from Burkitt et al, 1997 <sup>12</sup> ).....	3
Figure 1.2 – Cobblestone appearance of Crohn's disease involved colonic mucosa (from Burkitt et al, 1997 <sup>12</sup> ).....	4
Figure 1.3 – Colon affected by Crohn's disease (modified from Burkitt et al, 1997 <sup>12</sup> ).....	4
Figure 1.4 – Mucosal fissuring in Crohn's disease (from Brooke et al, 1977 <sup>5</sup> ) .....	5
Figure 1.5 – Granuloma in Crohn's disease (courtesy of Dr E.H. MacKay, Leicester General Hospital) .....	6
Figure 1.6 – Continuous mucosal involvement in ulcerative colitis (from Burkitt et al, 1997 <sup>12</sup> ) .....	6
Figure 1.7 -- Two crypt abscesses in an involved section from ulcerative colitis patient UC_2 .....	7
Figure 1.8 – Anatomy of colonic mucosa (from Brooke et al, 1977 <sup>5</sup> ).....	17
Figure 1.9 – Invagination of the M cell on the lamina propria face (from MacDermott, 1994 <sup>118</sup> ).....	19
Figure 1.10 – IBD aetiology theories.....	23
Figure 2.1 – Relative measure of pathological markers of inflammation in IBD tissues.....	38
Figure 2.2 – Outline of microarray procedure.....	39
Figure 2.3 – Good quality total RNA (left) & cDNA on 1% EtBr gels .....	41
Figure 2.4 – Schematic of the two methods being compared.....	44
Figure 2.5 - Target hybridisation flowchart .....	46
Figure 2.6 - cRNA yield as a percentage of amount placed on column .....	50
Figure 2.7 – Gel showing ds cDNA (lane 2), purified cRNA (lane 3) and fragmented cRNA (lane 4) fragmentation .....	51
Figure 2.8 – A typical microarray image .....	56
Figure 2.9 – Arrangement of probes on microarray.....	56
Figure 2.10 – Comparison of microarray and Q-PCR gene expression data.....	59
Figure 2.11 – H & E section of sample NI_5 mucosa.....	60
Figure 2.12– Staining of differentiated epithelial cells only with anti-CAM 5.2 antibody.....	64
Figure 3.1 – Basic gene expression matrix .....	70
Figure 3.2 - Gene expression matrix adapted for Affymetrix expression data.....	70
Figure 3.3a – Scatter plot showing all genes.....	72
Figure 3.3b – Scatter plot showing genes with a different absolute call in the two samples .....	72
Figure 3.4 – DMT fold change graph.....	73
Figure 3.5 – The bar graph.....	73
Figure 3.6 – The Histogram .....	74
Figure 3.7 – Typical view of Microsoft Excel spreadsheet for matrix-based mining .....	76
Figure 3.8 – Comparison of Tissue Groups .....	78
Figure 3.9 – Normal distribution of average difference values in a larger sample set .....	79
Figure 3.10 – View of a dendrogram created in Stanford's Cluster program and viewed in TreeView .....	83
Figure 3.11 – K-means clustering using J-express.....	84
Figure 3.12 – PCA output in J-express .....	86
Figure 4.1 – Annotation columns within the spreadsheet .....	90
Figure 4.2 – Expression of 'over expressed in CDi' kinase / phosphatase genes, in sample UCi_5 with minimal expression in rest of UCi samples .....	92
Figure 4.3 – Hierarchical clustering analysis .....	93
Figure 4.4 – K-means clustering of data into decreasing numbers of nodes.....	94
Figure 4.5 – Generalised protein function and pathway groups for CDi over expressed genes.....	98
Figure 4.6 – CDi specific differentially expressed genes shown by fold change from NI samples .....	100
Figure 4.7 – Transcript expression pattern of iNOS .....	101
Figure 4.8 – The action of nitric oxide synthase .....	101
Figure 4.9 – Transcript expression pattern of SAS .....	102
Figure 4.10 – Transcript expression pattern of hGSTT2.....	103
Figure 4.11 – Transcript expression pattern of EXLM1 .....	103
Figure 4.12 – Transcript expression pattern of hJag2 .....	104
Figure 4.13 – Transcript expression pattern of ERF .....	104
Figure 4.14 - Transcript expression pattern of Ets-2.....	105

Figure 4.15 – Transcript expression pattern of EST similar to oxalyl-CoA decarboxylase.....	105
Figure 4.16 - Transcript expression pattern of GBP-2.....	106
Figure 4.17 – UCi specific differentially expressed genes shown by fold change from NI samples.....	107
Figure 4.18 – Transcript expression pattern of Igλ heavy chain.....	108
Figure 4.19 - Transcript expression pattern of Snn.....	108
Figure 4.20 – BdrC3 transcript expression pattern.....	109
Figure 4.21 – Dynactin transcript expression pattern.....	110
Figure 4.22 – MRP-1 transcript expression pattern.....	110
Figure 4.23 – CED-6 transcript expression pattern.....	111
Figure 4.24 – SGK1 transcript expression pattern.....	112
Figure 4.25 – GAK transcript expression by tissue group.....	112
Figure 4.26 – PCTAIRE1 transcript expression by tissue group.....	113
Figure 4.27 – RNase K6 transcript expression by tissue group.....	114
Figure 4.28 – TRRAP transcript expression by tissue group.....	114
Figure 4.29 – RBM8A transcript expression pattern.....	115
Figure 4.30 – β2-microglobulin transcript expression by tissue group.....	116
Figure 4.31 - Human IgD transcript expression pattern.....	116
Figure 4.32 – SCYB10 transcript expression pattern.....	116
Figure 4.33 – DDT transcript expression by tissue group.....	117
Figure 4.34 – MDG1 transcript expression pattern by tissue group.....	118
Figure 4.35 - SEC14L transcript expression pattern.....	118
Figure 4.36 – Gp96 transcript expression pattern by tissue group.....	119
Figure 4.37 – Transcript expression of the gp96 receptor, CD91.....	119
Figure 4.38 - MSE55 / Borg5 / CEP1 transcript expression pattern.....	120
Figure 4.39 – Trophinin transcript expression pattern on Hu35K_subA.....	120
Figure 4.40 – Trophinin transcript expression pattern on HuFL.....	121
Figure 4.41 – Correlation of chromosomal location of identified genes and numbers of published references....	121
Figure 4.42 – Transcript expression pattern of PAF65B.....	124
Figure 4.43 – HS2ST1 transcript expression by tissue group.....	124
Figure 4.44 – Pex14 transcript expression pattern.....	125
Figure 4.45 – ITPKA transcript expression by tissue group.....	125
Figure 4.46 – ITPKA transcript expression pattern.....	125
Figure 4.47 – DTDST transcript expression by tissue group.....	126
Figure 4.48 – DRA transcript expression by sample.....	127
Figure 4.49 – DRA expression by pooled tissue group.....	127
Figure 4.50 – Genes that are over expressed in IBDi by fold change from NI samples.....	128
Figure 4.51 - Comparing B lymphocyte ICC data to HLA-DOβ antigen transcript expression.....	129
Figure 4.52 – CD22 transcript expression pattern.....	129
Figure 4.53 – Eotaxin precursor transcript expression pattern.....	130
Figure 4.54 – CD2 transcript expression pattern.....	130
Figure 4.55 – CD2 expression after normalisation with T cell ICC data.....	131
Figure 4.56 – HPK1 transcript expression pattern.....	131
Figure 4.57 – BAM32 transcript expression pattern.....	132
Figure 4.58 – Humig transcript expression pattern.....	132
Figure 4.59 – K-means clustering of IBDu genes into 24 nodes.....	134
Figure 4.60 – NR5A2 transcript expression pattern.....	137
Figure 4.61 – PRL-1 transcript expression pattern.....	138
Figure 4.62 – ZNF137 transcript expression pattern.....	138
Figure 4.63 – REG1α transcript expression.....	139
Figure 4.64 – REG1β transcript expression.....	140
Figure 4.65 – GBP2 transcript expression.....	140
Figure 4.66 – ICB-1 transcript expression pattern.....	141
Figure 4.67 – SIRT2 transcript expression pattern.....	141
Figure 5.1 - Fold Change in COx gene expression in IBDi compared to NI samples.....	151
Figure 5.2 - Fold Change in Acyl-CoA dehydrogenase gene expression in IBDi compared to NI samples.....	151

Figure 5.3 – Sulphate transporters' gene expression normalised to proportion of epithelial cells in sample.....	153
Figure 5.4a – Sulphate metabolism in the healthy colon .....	154
Figure 5.4b - Reduced Uptake of Sulphate Results in Energy Deficiency.....	154
Figure 5.5 – Fold Change of Apoptosis-related genes from NI samples .....	161
Figure 5.6 – Expression pattern of caspases-1, -2, -4, -5 and -9 across all samples.....	163

## LIST OF TABLES

Table 1.1 – Summary of the pathological differences between Crohn’s disease and ulcerative colitis (modified from Ross et al, 1995 <sup>18</sup> ).....	8
Table 1.2 – Association of HLA types in IBD (modified from Fiocchi 1998 <sup>25</sup> ).....	27
HLA type.....	27
Ulcerative colitis .....	27
Crohn's disease .....	27
Table 1.3 – Previously determined gene expression patterns in IBD.....	29
Table 2.1 – Weights of samples used in microarray analysis.....	34
Table 2.2 - Clinical features of participating patients.....	37
Table 2.3 - Reagent volumes in 1 <sup>st</sup> strand cDNA reaction .....	43
Table 2.4 - Reagents for 2 <sup>nd</sup> strand cDNA synthesis .....	43
Table 2.5 - Amounts of reverse transcriptase in 1 <sup>st</sup> strand cDNA reaction .....	45
Table 2.6 - Re-suspension volumes for double stranded cDNA.....	45
Table 2.7 - Targets generated for method comparison.....	45
Table 2.8 - Summary of comparison analyses.....	46
Table 2.9 - Description of purification columns tested .....	50
Table 2.10 - Example cRNA fragmentation mix.....	52
Table 2.11 - Hybridisation cocktail for 50 µg fragmented cRNA.....	53
Table 2.12 - Fluidics program (EukGE-WS1) for standard wash method .....	54
Table 2.13 - Fluidics program (EukGE-WS2) for antibody wash method .....	55
Table 2.14 – Primary Antibodies .....	62
Table 3.1 – Matching the UC and NI patients on basis of sex and age .....	77
Table 4.1 – Databases accessed through SRS for gene annotation in the current study.....	89
Table 4.2 – Criteria for the query ‘Over expressed in CDi compared to UCi’ .....	91
Table 4.3 – Criteria for the query ‘Over expressed in UCi compared to CDi’ .....	91
Table 4.4– Criteria for the query ‘Over expressed in CDi compared to UCi’ .....	96
Table 4.5 – Criteria for the query ‘Over expressed in UCi compared to CDi’ .....	96
Table 4.6 – Kinase / phosphatase genes over expressed in CDi.....	96
Table 4.7 – UCi over expressed genes .....	98
Table 4.8 – Generation of gene expression profiles for CDi & UCi samples .....	99
Table 4.9 – Identification of genes differentially expressed in IBDi samples.....	123
Table 4.10 – Genes that are called present in IBDi samples and called absent in NI samples.....	128
Table 4.11 – Identification of genes differentially expressed in IBDu samples.....	133
Table 4.12 – Descriptive summary of IBDu K-means clustering (figure 4.59).....	136
Table 4.13 – Inflammatory markers from cluster 18 of figure 4.59 .....	139
Table 5.1 – Cancer-related genes that are over expressed in IBDi tissues .....	157
Table 5.2 – Protein function of differentially regulated apoptosis related genes .....	162

## ABBREVIATIONS

As per convention, gene names are *italicised*, and their protein products are in normal script. Not all gene name abbreviations used, are listed here and these are explained where they occur within the text.

5-ASA	5- aminosalicylic acid
APC	antigen presenting cell
CD	Crohn's disease
CDi	involved Crohn's disease
CDi_10	involved sample from Crohn's disease patient 10
CDu	uninvolved Crohn's disease
CDu_10	uninvolved sample from Crohn's disease patient 10
contig	contiguous nucleotide sequence
CRC	colorectal cancer
db	database
DNA	deoxyribonucleic acid
DNBS	dinitrobenzenesulfonic acid
EMBL	European Molecular Biology Laboratory
GALT	gut associated lymphoid tissue
GI	gastrointestinal tract
HGD	high-grade dysplasia
IBD	inflammatory bowel disease
IEL	intraepithelial lymphocyte
Ig	immunoglobulin
IL	interleukin
LGD	low-grade dysplasia
M cells	mucosal cells
MALT	mucosa associated lymphoid tissue
MHC	major histocompatibility complex
ORF	open reading frame
PBS	phosphate buffered saline
RNA	ribonucleic acid
RT-PCR	reverse transcriptase polymerase chain reaction
TCR	T cell receptor
TFP	trefoil peptide
TNF	tumour necrosis factor
UC	ulcerative colitis
UCi	involved ulcerative colitis
UCi_10	involved sample from ulcerative colitis patient 10
UCu	uninvolved ulcerative colitis
UCu_10	uninvolved sample from ulcerative colitis patient 10

**CHAPTER ONE - INTRODUCTION**

## **1.1 Inflammatory Bowel Disease**

---

### **1.1.1 Historical Overview**

'Inflammatory bowel disease' (IBD) includes a number of conditions that result in inflammation of the small and/or large bowel. In particular, the term is used for Crohn's disease (CD) and ulcerative colitis (UC) collectively; two distinct but possibly related conditions. The aetiology of the two diseases is likely to be different but the endpoint is the same, namely inflammation of the bowel.

Clinical descriptions of what is now known as inflammatory bowel disease have been recorded for many thousands of years<sup>1</sup>. The Index Medicus contains clearly recognisable descriptions of ulcerative colitis since its inception in 1840\*. The clinical symptoms of IBD are similar to intestinal tuberculosis; thus, the prevalence of tuberculosis in the 19<sup>th</sup> and early 20<sup>th</sup> centuries may have masked the true diagnosis of IBD in many cases<sup>2, 3</sup>. The term ulcerative colitis was first used in 1888<sup>1</sup>, but Crohn's disease was not recognised as a distinct entity until 1932<sup>2</sup>. What is now commonly known as Crohn's disease was initially named 'non-specific enteritis'<sup>2</sup> and the terms regional or granulomatous enteritis are also sometimes used<sup>4</sup>.

### **1.1.2 Clinical Features**

#### *1.1.2.1 Crohn's disease*

Diarrhoea is the most common clinical feature of Crohn's disease<sup>5</sup>. Abdominal pain is a general feature of IBD caused by affected tissues obstructing the intestine. The pain may be chronic or acute depending on the amount of intestine involved<sup>5</sup>. Compared to ulcerative colitis diarrhoea in Crohn's disease tends to be slightly less severe but is more frequently associated with abdominal pain<sup>4</sup>. Fatigue, malnutrition and weight loss are also major symptoms of Crohn's disease<sup>4-6</sup>. The symptoms of Crohn's disease may also depend on the age of the patient; for example the most obvious symptom in a child or adolescent may be under-average growth and a delay in the development of secondary sexual characteristics<sup>7</sup>.

#### *1.1.2.2 Ulcerative colitis*

The main symptom of ulcerative colitis is severe diarrhoea; frequently passed with blood and mucus<sup>8</sup>. Tenesmus is often strong and urgent<sup>6</sup>, although some patients suffer tenesmus with constipation<sup>8</sup>.

#### *1.1.2.3 Similarities*

Three clinical patterns of colitis\* have been identified, the most common being chronic relapsing colitis<sup>6</sup>. The course of chronic relapsing ulcerative colitis is cyclic, with periods of severe ill health being followed by spontaneous remission - a period of mild or no symptoms before the disease flares up again. After each flare up, the patient does not return to full health and gradually becomes increasingly ill after each attack. Chronic

---

\* Personal observation

continuing colitis is of moderate severity, but with no symptom-free periods. The most acute form of the disease is active fulminating colitis, characterised by a rapid and sometimes fatal deterioration<sup>8</sup>. Ulcerative colitis very rarely remits completely and even when in apparent remission, patients remain at risk of further attacks<sup>6</sup>. Crohn's disease very rarely progresses to active fulminating colitis<sup>5</sup>.

### 1.1.3 Pathological Features

#### 1.1.3.1 Crohn's disease

##### *Morphology*

In their seminal description Crohn, Ginzburg and Oppenheimer reported that Crohn's disease was specific to the terminal ileum<sup>2</sup>. The colonic form was recognized in 1960<sup>9</sup> and it is now known that Crohn's disease can occur in varying pathological forms anywhere along the gastrointestinal tract from mouth to anus<sup>4, 10</sup>. However, the terminal ileum is the most common site of Crohn's disease<sup>11</sup>.

In colonic Crohn's disease, a common pattern of involvement is the terminal ileum to the descending colon, with sigmoid colon and rectal sparing<sup>4</sup>. In both the ileum and the colon, Crohn's disease affects the intestine in discontinuous patches known as skip lesions, which are a distinctive feature of Crohn's disease (figure 1.1).



**Figure 1.1 – Skip lesions in Crohn's disease (modified from Burkitt et al, 1997<sup>12</sup>)**

*The presence of apparently normal mucosa can be seen between the involved areas (arrows). These uninvolved areas have been 'skipped' over.*

##### *Gross Pathology*

On examination of the gross specimen, involved mucosa typically shows a 'cobblestone' pattern (figure 1.2), due to the oedematous nature of the mucosa<sup>13</sup>.

\* Colitis describes colonic inflammation regardless of cause and thus describes both CD and UC caused colonic inflammation.

**Figure 1.2 – Cobblestone appearance of Crohn's disease involved colonic mucosa (from Burkitt et al, 1997<sup>12</sup>)**



In another common manifestation of Crohn's disease, the mucosal surface loses its normal folded structure, becoming a smooth sheet. This is accompanied by thickening of the bowel wall (fibrosis) and narrowing of the lumen (stricturing)<sup>13</sup> (figure 1.3). The stricturing may be severe enough to block the intestines completely. Strictures can be detected radiologically and are commonly used to diagnose Crohn's disease<sup>14</sup>. Whether stricturing is a feature or a cause of the disease process is controversial. It was previously assumed that the strictures arose as a direct result of the chronic inflammation and fibrosis. However, there is anecdotal evidence suggesting that balloon dilation of ileal strictures not only results in resolution of the lesion as expected, but also improves the associated inflammation and ulceration<sup>14</sup>.



**Figure 1.3 – Colon affected by Crohn's disease (modified from Burkitt et al, 1997<sup>12</sup>)**

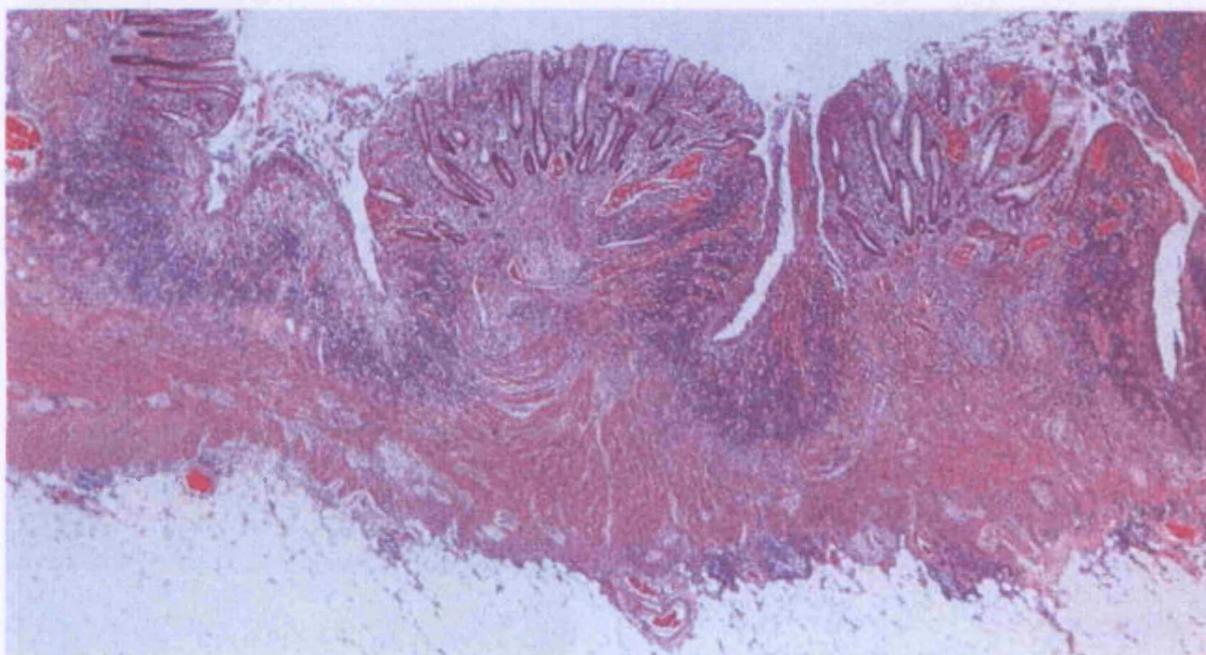
A stricture is evident in the transverse colon at the top of the figure (arrow). The abnormal thickness of the bowel walls can be seen. The 'spared' rectum is also apparent at the bottom right.

The ulceration in Crohn's disease creates fissures that extend throughout the bowel wall, (i.e. transmurally), often into the serosa<sup>4, 15</sup>. This sometimes results in a fibrotic attachment between the intestines and surrounding organs<sup>4</sup>.

#### *Histology<sup>15-17</sup>*

The presence of transmural lymphoid aggregates is indicative of Crohn's disease. As a result, the mucosa becomes ulcerated, progressing either laterally or penetrating into the deeper layers of the bowel wall (figure 1.4). The linear ulceration proceeds along the length of the bowel, whilst deeply penetrating fissures can form through the bowel wall. Both of these lesions are typical in Crohn's disease. Granulomas can form in all affected tissue layers from the mouth to the rectum and are seen in about half of all Crohn's disease cases (figure 1.5). They do not form in ulcerative colitis; however, the absence of granulomas does not rule out Crohn's disease, as they are frequently undetectable even in major resections of severe active disease.

**Figure 1.4 – Mucosal fissuring in Crohn's disease (from Brooke et al, 1977<sup>5</sup>)**



#### **1.1.3.2 Ulcerative colitis**

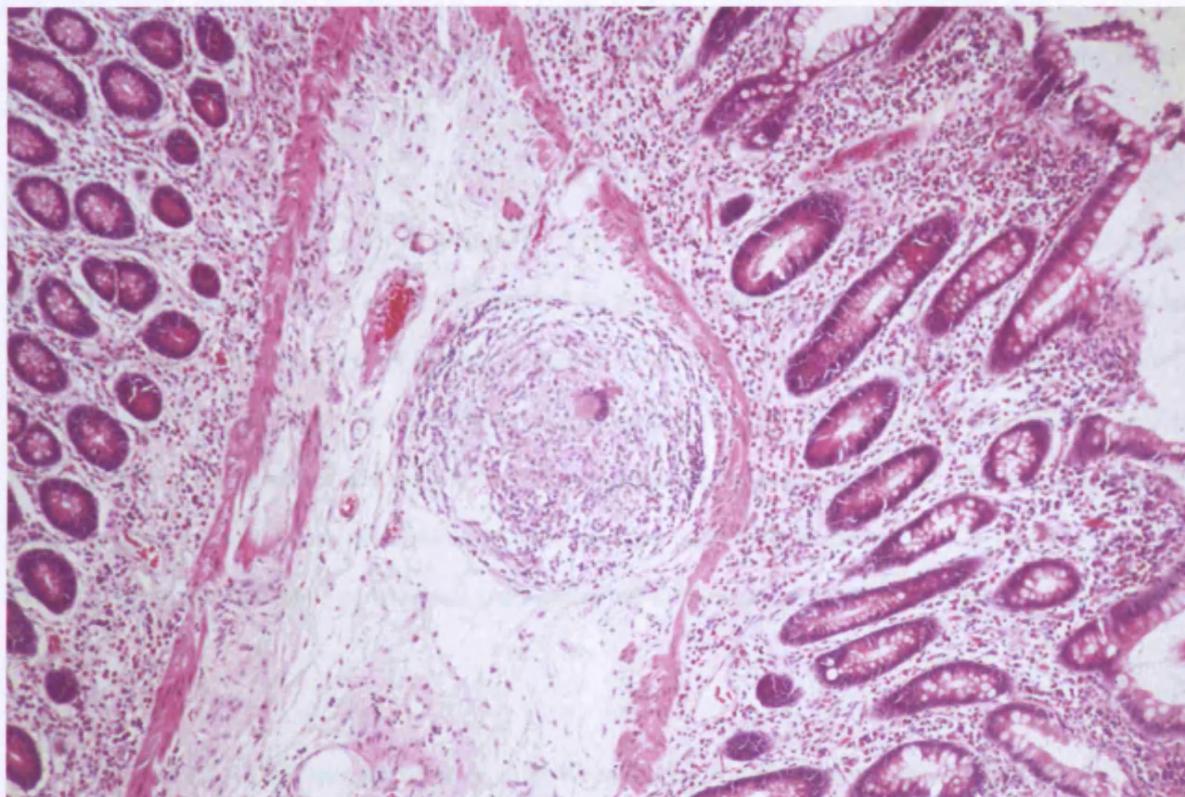
##### *Morphology<sup>8, 13</sup>*

Ulcerative colitis is limited to the colon and rectum, i.e. the large bowel\*. Its pattern of distribution within the colon is different to that of Crohn's disease, as it usually involves the rectum and sigmoid. The inflammation may involve the rest of the colon, appearing to spread in a continuous pattern, towards the caecum, but total colonic involvement is not inevitable and the proximal limit of involvement can be at any point. Unlike Crohn's

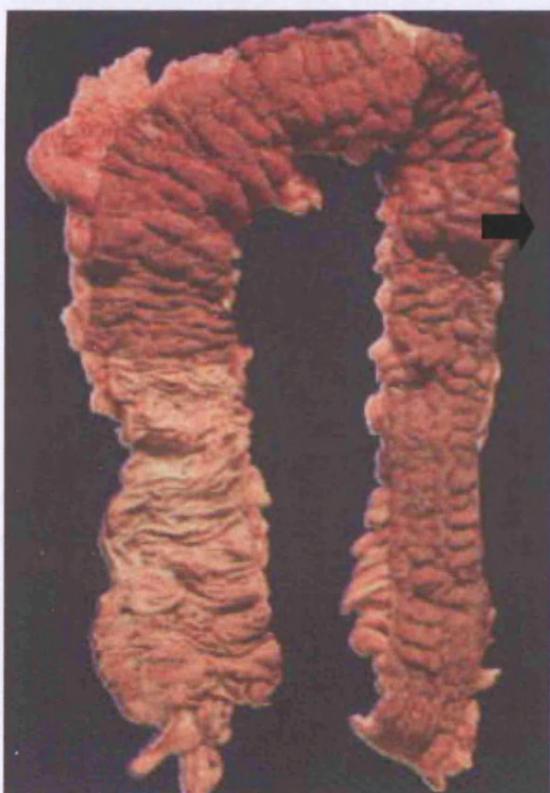
\* Very rarely, the phenomenon of 'backwash ileitis' is observed in patients with total colonic involvement.

disease, discrete patches of inflammation do not occur in ulcerative colitis; all areas between the proximal and distal limits of the diseased tissue is always involved (figure 1.6).

**Figure 1.5 – Granuloma in Crohn's disease (courtesy of Dr E.H. MacKay, Leicester General Hospital)**



*In this case of Crohn's disease, the granuloma is located in the lamina propria and contains a giant cell is at its centre.*



**Figure 1.6 – Continuous mucosal involvement in ulcerative colitis (from Burkitt et al, 1997<sup>12</sup>)**

*The colon is involved from the rectum to descending colon. All the mucosa between the distal and proximal limits are involved. There is a sharp demarcation between the diseased mucosa and the apparently normal mucosa.*

### Gross Pathology

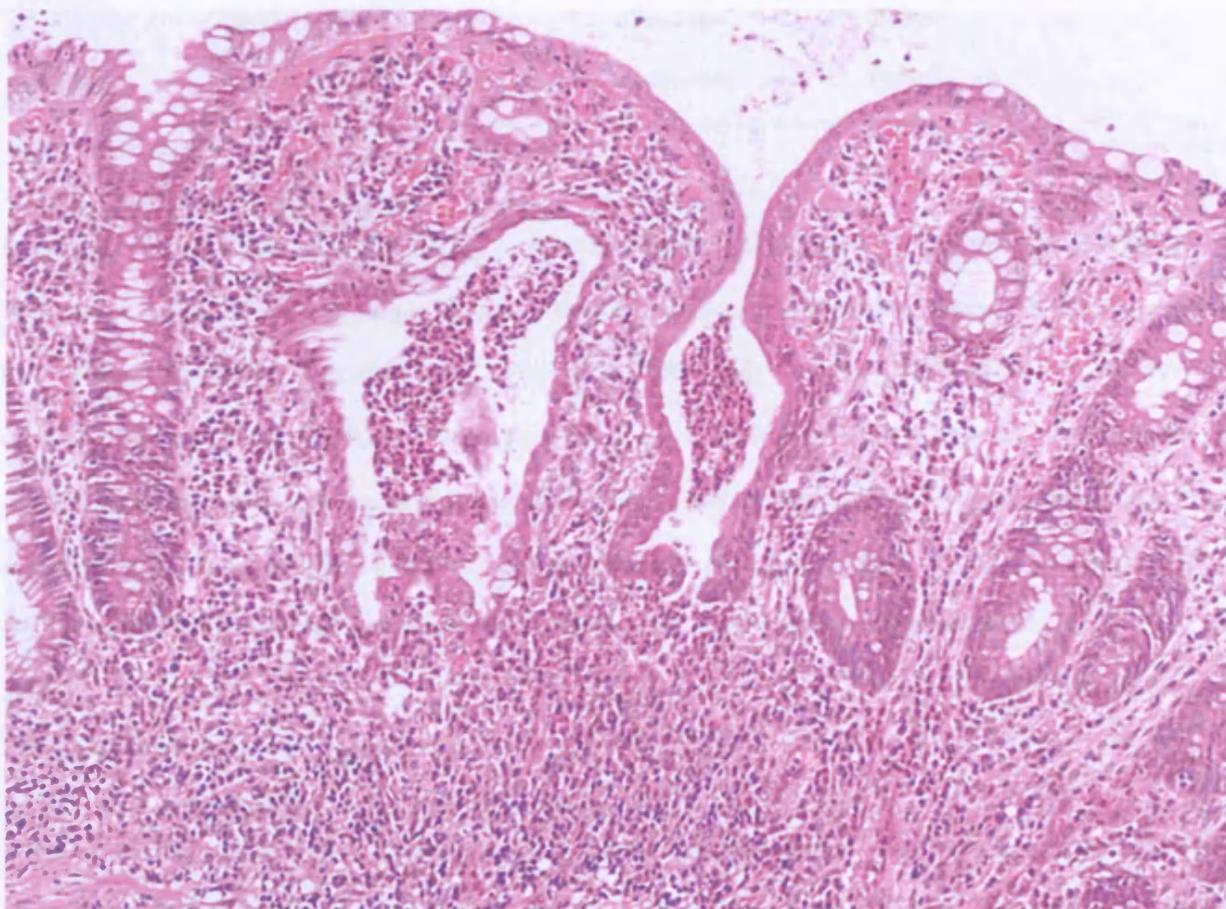
Ulcerative colitis may result in the loss of mucosa leaving the submucosa directly exposed to the luminal contents<sup>8, 13</sup>. When present the mucosa may show a cobblestone appearance (as in figure 1.5) or numerous inflammatory polyps may be present. Unlike in Crohn's disease mural thickening does not occur<sup>16</sup> and fibrosis is minimal.

### Histology<sup>15-17</sup>

Despite the name ulcers are not very common<sup>13</sup>. When they do occur they tend to be more superficial than those in Crohn's disease, affecting only the mucosal layer and the inflammatory response in general does not extend beyond the submucosa (except in toxic megacolon, described below). The earliest histological feature of ulcerative colitis is focal neutrophilic infiltration into the epithelial layer. The neutrophils migrate into isolated crypts initially creating a crypt abscess (figure 1.7) but ultimately destroying the crypt. This crypt destruction leads to progressive atrophy and the architecture of the mucosa is therefore destroyed.

Table 1.1 summarises the main pathological and histological differences between Crohn's disease and ulcerative colitis.

**Figure 1.7 – Two crypt abscesses in an involved section from ulcerative colitis patient UC\_2**



**Table 1.1 – Summary of the pathological differences between Crohn’s disease and ulcerative colitis (modified from Ross et al, 1995<sup>18</sup>)**

<i>Lesion</i>	<i>Crohn’s disease</i>	<i>Ulcerative colitis</i>
<b>Pathological</b>		
Thickened bowel wall	Typical	Uncommon
Luminal narrowing	Typical	Uncommon
Skip lesions	Common	Absent
Right colonic predominance	Common	Absent
Fissures and fistulas	Common	Absent
Linear ulcers	Common	Absent
Inflammatory Pseudopolyps	Absent	Common
<b>Histological</b>		
Transmural inflammation	Typical	Uncommon
Submucosal fibrosis	Typical	Absent
Serosal fibrosis	Typical	Absent
Fissures	Typical	Rare
Granulomas	Common	Absent
Crypt abscesses	Uncommon	Typical

### **1.1.3.3 Indeterminate colitis**

It is highly likely that in the early 20<sup>th</sup> century, many cases of colonic Crohn's disease were mistakenly diagnosed as ulcerative colitis<sup>9</sup>. The recognition of colonic Crohn's disease has reduced this, but in about 5% of IBD cases, the clinical and pathological evidence does not point to conclusively to either Crohn's disease or ulcerative colitis<sup>19</sup>. These cases are termed indeterminate colitis, with some features of both diseases. Indeterminate colitis is generally regarded as a temporary diagnosis and not a true pathological entity<sup>20-22</sup> and clinicians aim to eventually give a definite diagnosis of either Crohn's disease or ulcerative colitis<sup>19</sup>. However, the possibility that Crohn's disease and ulcerative colitis represent two ends of a single disease spectrum remains.

## **1.1.4 Complications**

### **1.1.4.1 Perforation**

When an ulcer penetrates the full thickness of the bowel wall, the muscle underneath is exposed to the inflammatory process happening in the mucosa, resulting in destruction of muscle tissue. This results in a thin wall, which could dilate with the gut contents and most dangerously, is liable to become perforated<sup>5</sup>. Due to the

tendency of Crohn's disease to form fissured ulcers (and fibrotic adhesions), the bowel wall tends to thicken rather than dilate. Furthermore, adhesions form with surrounding tissues making the perforation rare in Crohn's disease<sup>5</sup>. However, perforation is possible without accompanying dilation<sup>8</sup> and inflammatory destruction passing through serosal adhesions can sometimes create fistulae.

#### ***1.1.4.2 Toxic megacolon***

A serious complication of ulcerative colitis is toxic megacolon. This is an extreme dilation of the bowel wall, which carries the risk of perforation. Once diagnosed, the patient usually undergoes emergency surgery to remove the entire colon (total colectomy)<sup>8</sup>.

#### ***1.1.4.3 Haemorrhage***

Another complication of ulcerative colitis is haemorrhage. Rectal bleeding is commonly associated with diarrhoea in ulcerative colitis and if severe enough the patient may require a blood transfusion<sup>8</sup>. If the haemorrhage is severe and sudden it could prove fatal through massive blood loss.

#### ***1.1.4.4 Cancer***

Colon cancer has been associated with ulcerative colitis for many years and there is a large literature dedicated to the subject. The risk of colon cancer after ten years of colitis is 23 fold greater than that of the general population, increasing to 32 fold after twenty years of colitis<sup>15</sup>. The risk is greatest in those patients who have total, or near total, colonic involvement<sup>15, 23, 24</sup>. The incidence of colon cancer in ulcerative colitis is 3-5%<sup>15</sup>. A number of studies have shown that the incidence of cancer is increased for those whose colitic symptoms began in adolescence<sup>24</sup>, but this may be the result of a longer follow-up period, rather than a true risk factor. The association between Crohn's disease and colon cancer is not as strong as that for ulcerative colitis. In the past, Crohn's disease was not believed to be a risk factor for cancer<sup>25</sup>, however this is now changing<sup>26, 27</sup>. It is possible that the increasing incidence of cancer in Crohn's disease is related to the increase in colonic Crohn's disease<sup>28</sup>. The colon may be more susceptible to cancer than the small bowel and as ulcerative colitis only affects the colon, this would be seen as a significant correlation. As the incidence of colonic Crohn's disease increased, colorectal cancer in Crohn's patients would also increase.

#### ***1.1.4.5 Extra-intestinal complications***

Patients with IBD, especially Crohn's disease, occasionally present with inflammatory symptoms elsewhere in the body. These include inflammation of the joints (e.g. peripheral arthritis, axial arthropathy), inflammation of parts of the eye (e.g. uveitis) and dermatological manifestations (e.g. erythema nodosum)<sup>6, 29</sup>. It is believed that these are secondary complications, as they often seem to be cured by removal of the colon in ulcerative colitis<sup>8</sup>. There are also reports of autoimmune disorders<sup>30</sup> and airway disease<sup>31</sup> being associated with IBD.

## **1.2 Therapies**

---

### **1.2.1 Nutritional support**

This is an important aspect of therapy, especially in Crohn's disease, as the CD patients are often malnourished. The carbohydrate, protein and vitamin intake of patients can be supplemented in a number of ways. Enteral feeding may involve (i) changing the patients diet and / or adding nutritional supplements, (ii) feeding through a nasogastric tube, (iii) or feeding through a gastrostomy or a jejunostomy. The aim is to make use of the patient's own GI tract as much as possible. However, in some cases the patients' own digestive tract is not able to take food in any form and in these cases parenteral therapy is indicated, i.e. nutrients are delivered intravenously. Parenteral therapy is very expensive and has many potential complications; it is used only as a last resort when enteral therapy has failed.

### **1.2.2 Medication**

#### ***1.2.2.1 Sulphasalazine***

Sulphasalazine is indicated in mild to moderate ulcerative colitis and for active colonic Crohn's disease, and has been the main method of treating long-term ulcerative colitis for many years<sup>32</sup>. Sulphasalazine breaks down into two constituent chemical moieties, sulphonamide and 5-aminosalicylic acid (5-ASA). 5-ASA acts by inhibiting the enzyme complex prostaglandin-endoperoxide synthase. Both moieties inhibit natural killer cell activity and lymphocyte activation. 5-ASA is believed to be the active moiety in sulphasalazine against IBD. Its generic name is mesalazine and there are numerous compounds on the market based on its chemical structure, e.g. olsalazine. The alternative forms of 5-ASA are no more or less efficacious than each other and are indicated in different patients based on individual tolerance<sup>33</sup>.

#### ***1.2.2.2 Steroids***

Corticosteroids, such as prednisolone and prednisone, are indicated in acute colitis for their anti-inflammatory and immunosuppressive effects<sup>32</sup>. Part of their immunosuppressive action comes from the inhibition of neutrophil and monocyte adherence to capillary endothelium. Prednisolone and prednisone also inhibit phospholipase A2 activity, thereby decreasing the formation of pro-inflammatory prostaglandins and leukotrienes. Osteoporosis, increased risk of infection and adrenal suppression are all potential side effects of long-term steroid treatment, although some of the newer steroids such as budesonide can reduce these effects. (Reviewed in Barnes, 1998<sup>34</sup>.)

#### ***1.2.2.3 Other anti-inflammatory agents indicated in IBD***

Another drawback of steroid treatment is the fact that the initial good response often deteriorates over time and in these cases the general immunosuppressant azothioprine may be used<sup>35</sup>. Azothioprine affects both the specific and non-specific arms of the immune system, reducing IgG and IgM synthesis and diminishing non-

specific immunity<sup>32</sup>. Methotrexate (for CD) and cyclosporin A (for UC) are also indicated for severe steroid refractory disease<sup>33</sup>. Methotrexate is a folic acid antagonist and has an anti-metabolic action<sup>36</sup>. Cyclosporin A suppresses the immune system by inhibiting both the action of T helper cells<sup>37</sup> and the production of macrophage specific lymphokines<sup>36</sup>.

#### **1.2.2.4 Biologic therapies**

Most of the drugs discussed above are primarily indicated in ulcerative colitis. However, a new class of drugs, tumour necrosis factor alpha (TNF $\alpha$ ) inhibitors, are indicated in Crohn's disease<sup>38, 39</sup>. The first TNF $\alpha$  inhibitor to be launched was an anti-TNF $\alpha$  monoclonal antibody, Infliximab. Since its approval by the American FDA in 1998, it has been tested in a number of clinical trials for Crohn's disease<sup>40, 41</sup>. The exact mechanism by which the drug works is unknown. It may act by binding to TNF $\alpha$  expressed on cell membranes, neutralising secreted TNF $\alpha$ , or by inducing apoptosis in TNF $\alpha$  producing cells<sup>38</sup>.

Clinical trials of interleukin-10 (IL-10) were initially carried out in Crohn's disease with varying success rates<sup>42</sup>. IL-10 was also the subject of a clinical trial in ulcerative colitis<sup>43</sup> but has not shown the success seen with anti-TNF $\alpha$  antibodies in Crohn's disease.

### **1.2.3 Surgery**

#### **1.2.3.1 Crohn's disease**

As discussed above (section 1.1.3.1), Crohn's disease is a recurrent disease that can affect any part of the GI tract from the mouth to the anus. Therefore surgery is rarely curative and is indicated only in complications of the disease, e.g. if a fistula or an abscess has formed, or if the formation of a stricture has caused an intestinal obstruction. The type of surgery indicated depends on the nutritional state of the patient and the anatomical site of the diseased tissue. Where there is an abscess or a fistula segmental resection is indicated, with the remaining bowel being anastomosed or exteriorised as a stoma (an ileostomy or a colostomy). 'Strictureplasty' widens the intestinal lumen and is sometimes indicated in obstructive disease. However recurrent disease, especially at resection margins, is common<sup>44-46</sup> and therefore surgery remains a last resort in the treatment of Crohn's disease.

#### **1.2.3.2 Ulcerative colitis**

Ulcerative colitis is confined to the large bowel (section 1.1.3.2) and thus removal of the entire colon and rectum (colo-proctectomy) effectively 'cures' the patients of ulcerative colitis and removes the risk of colorectal cancer. This procedure is indicated as an elective procedure in (i) chronic relapsing disease, (ii) in cases where medical therapy has failed (including avoidance of steroid side effects), (iii) in cases of dysplasia or carcinoma. Colectomy may be indicated as an emergency procedure in cases of acute fulminate disease or toxic megacolon.

With acute disease, the clinicians may be unsure as whether the patient is suffering from ulcerative colitis or Crohn's disease. Therefore, a subtotal colectomy is indicated, which removes only the colon, leaving the rectum. An ileostomy is formed which may be reversed in a future procedure, depending on the patient's health. In elective surgery there are two routes that may be taken, a colectomy, or a colectomy with an ileoanal pouch. The first option has relatively few complications, but leaves the patient with a permanent stoma. The second procedure may be staged with a temporary ileostomy, but ultimately does not require a stoma. However possible complications include high frequency (6-8 times a day), 'pouchitis' and potentially pouch incontinence.

### 1.3 Epidemiology

---

The cause of IBD is both genetic and environmental in nature. The concordance rate in monozygotic twins has been measured at 58.3% for Crohn's disease and 18.2% for ulcerative colitis<sup>47</sup>, while the concordance rate between dizygotic twin pairs has been measured as 0-12% in Crohn's disease and 0-5% in ulcerative colitis; that is, equivalent to the concordance rate between other sibling pairs<sup>48, 49</sup>. Further evidence of the environmental influence on IBD comes from a case of monozygotic twins suffering discordant IBD; one had Crohn's disease and the other had ulcerative colitis<sup>50</sup>. This implies a role for both genes and environment in the aetiology of IBD, with a larger environmental influence on ulcerative colitis than Crohn's disease.

#### 1.3.1 Incidence

Ulcerative colitis is more prevalent than Crohn's disease, with an annual incidence of 4 to 7 per 100,000, compared to 0.5 to 5 per 100,000 for Crohn's disease<sup>17</sup>. In general, IBD is more prevalent in Western countries compared to developing countries. IBD can occur at any age, but both diseases tend to show the greatest incidence in the 15-20 age group<sup>16, 17</sup>. Ulcerative colitis shows a minor incidence peak in the 60's and Crohn's disease in the 50-70 age group<sup>51</sup>. IBD affects both sexes, with ulcerative colitis showing a slight male bias and Crohn's disease a slight female bias<sup>51</sup>. The incidence and prevalence of IBD has changed in the past for a number of reasons, discussed below.

##### 1.3.1.1 Crohn's disease

Since the initial paper by Crohn *et al* in 1932<sup>2</sup> there has been an increased awareness of Crohn's disease alongside an improvement of diagnostic methods and therefore, an apparent rise in the incidence of Crohn's disease since 1932 would be expected. A rise in the incidence of Crohn's disease was indeed observed in years following 1932<sup>52</sup>. However, the rise in incidence continues past the time when these factors are no longer relevant, implying that there has been a real increase in the incidence of Crohn's disease in the middle half of the 20<sup>th</sup> century<sup>52</sup>. Whether the incidence of Crohn's disease is still on the increase in Western countries is unclear, with some groups observing a continued rise<sup>26, 53-55</sup> and some reporting a plateau<sup>11, 56</sup>.

There is evidence that the nature of Crohn's disease may be changing. In a study investigating the incidence of Crohn's disease in Stockholm from 1955-1989 it was reported that colonic Crohn's disease is increasing while ileocaecal disease is decreasing<sup>28</sup>. The same study also reported that elderly Crohn's disease patients had more likelihood of small bowel disease compared to younger patients. However, this apparent rise in incidence of colonic Crohn's disease must be treated cautiously as it may simply be a result of improved differentiation between colonic Crohn's disease and ulcerative colitis. A more recent study by the same group adds that the age at which IBD is diagnosed seems to be increasing<sup>26</sup>.

### ***1.3.1.2 Ulcerative colitis***

The incidence of ulcerative colitis increased greatly in the first half of this century<sup>52</sup> and now seems to have reached a plateau<sup>56</sup>. However, some groups report an increase in incidence<sup>53</sup> whilst others report that the incidence of ulcerative colitis is stable, with a tendency towards decrease<sup>54</sup>. These apparently contradicting reports can be reconciled when the geographical population studied is taken into account. It seems that southern Europe has had a lower incidence of ulcerative colitis in the past, but the incidence in northern Europe was increasing. At present the incidence in northern Europe seems to be stabilising, while in southern Europe the incidence is increasing to approximate that of northern Europe<sup>55</sup>.

## **1.3.2 Environmental factors**

### ***1.3.2.1 Geography***

IBD is a disease of developed nations, affecting Western Europe and North America at a greater frequency than less developed countries. It seems that as countries become increasingly better developed the incidence of IBD is also increasing<sup>55</sup>. There have been reports that the incidence of IBD in urban areas is higher than in rural areas<sup>51, 57</sup>. However, this is not a universal finding<sup>58</sup>.

### ***1.3.2.2 Sociology***

Affluence is associated with IBD, especially with Crohn's disease. Those with a higher income<sup>57</sup> or higher education<sup>59</sup> are more likely to suffer from Crohn's disease. There is also some evidence to link a higher social standing to an increased risk of IBD<sup>56</sup>.

## **1.3.3 Lifestyle factors**

### ***1.3.3.1 Western lifestyle***

The Western lifestyle is linked to a higher incidence of IBD. The incidence of IBD in first generation eastern migrants to western countries is lower than that of the native western population<sup>60, 61</sup>. However, descendants of the migrant population show an IBD incidence equal to that of the native population by the second generation<sup>7, 62-64</sup>.

### ***1.3.3.2 Smoking***

Smoking is a well-established risk factor for Crohn's disease<sup>65, 66</sup>. Conversely, *not* smoking is well established as a risk factor in ulcerative colitis<sup>67, 68</sup> and ulcerative colitis patients may benefit from nicotine therapy<sup>69</sup>. This apparent difference between the two forms of IBD remains unexplained<sup>70-73</sup>.

### ***1.3.3.3 Other factors***

Not all studies universally agree on which other lifestyle factors increase the risk of IBD. For example, the use of oral contraceptives has been linked to a higher risk of IBD by some<sup>66, 74</sup>, but is disputed by others<sup>75</sup>. For

many years there has been anecdotal evidence suggesting that stressful events trigger relapses in pre-existing IBD patients<sup>8</sup>. This link between stress and colitis relapses was further supported by a recent animal study<sup>76</sup>. Colitis was induced in mice by dinitrobenzenesulfonic acid (DNBS), which resolved after six weeks. Colitis was subsequently re-activated by a sub-threshold dose of DNBS and stress, but could not be re-activated by DNBS alone.

The environment and lifestyle factors are likely to be interdependent. For example, those with higher incomes are more likely to be able to afford higher education, maybe delay beginning a family in favour of a career and thus be more likely to take the pill; i.e. it is very difficult to pin any one factor as directly causing IBD. The epidemiological literature can be summarised as indicating a higher incidence of IBD in young adults, those of higher social standing, who are affluent, live in an urban environment and follow a western lifestyle.

### **1.3.4 Genetic factors**

#### ***1.3.4.1 Race***

When different racial groups in the same geographical region are studied, the incidence of IBD is higher in Caucasians than Asian or African-Americans<sup>4</sup>. However, this effect tends to be limited to first generation migrants and subsequent generations have the same incidence as the native population (discussed above), indicating a larger environmental influence as opposed to a genetic one. However the increased risk of IBD in Jewish populations has been noted for a number of years<sup>77</sup>.

#### ***1.3.4.2 Familial Inheritance***

First-degree relatives of IBD patients are at a 10-15 fold greater risk of IBD relative to the general population<sup>78</sup>, whilst siblings are at a 30-fold greater risk<sup>79</sup>. It has been shown by a number of groups that the IBD-affected offspring of IBD affected parents have a significantly decreased age of onset<sup>80</sup>. This is especially evident in Crohn's disease where the difference in age at onset between parent and child may be as much as 20 years. It was initially believed that this might be due to genetic anticipation<sup>81, 82</sup>, whereby the age at disease onset decreases at each generation. This is seen in the monogenic Huntington's disease and in that disease is due to the inheritance of nucleotide triplets in the disease gene which increase in number over successive generations<sup>83</sup>. The phenomenon of genetic anticipation in IBD has been re-examined in light the discovery of the molecular mechanism in Huntington's disease. However, the methods used in studies supporting genetic anticipation have been criticised and recent studies do not support genetic anticipation in IBD<sup>78, 80, 84, 85</sup>.

#### ***1.3.4.3 IBD loci***

There have been numerous studies attempting to elucidate mutations in genetic loci that predispose to IBD. The association between Crohn's disease and a region on chromosome 16 is well established<sup>86-90</sup> and this region is

known as *IBD1*<sup>91</sup>. Another strongly IBD associated loci is *IBD2*, a region on chromosome 12q<sup>88, 89, 92</sup>. This locus is specifically associated with ulcerative colitis<sup>93, 94</sup>. A third IBD locus (*IBD3*) on chromosome 6p has recently been identified with specificity to Crohn's disease<sup>94, 95</sup>. Additionally, putative loci on many other chromosomes have been proposed, for example, chromosomes 1<sup>96, 3<sup>97</sup>, 14<sup>98</sup></sup> and X<sup>99</sup>.

With the identification of a specific predisposition gene in *IBD1*, *NOD2*<sup>100-102</sup>, the emphasis has shifted from *IBD1* as an epidemiological feature, to the role of *NOD2* in the aetiology of Crohn's disease (section 1.5.2.2).

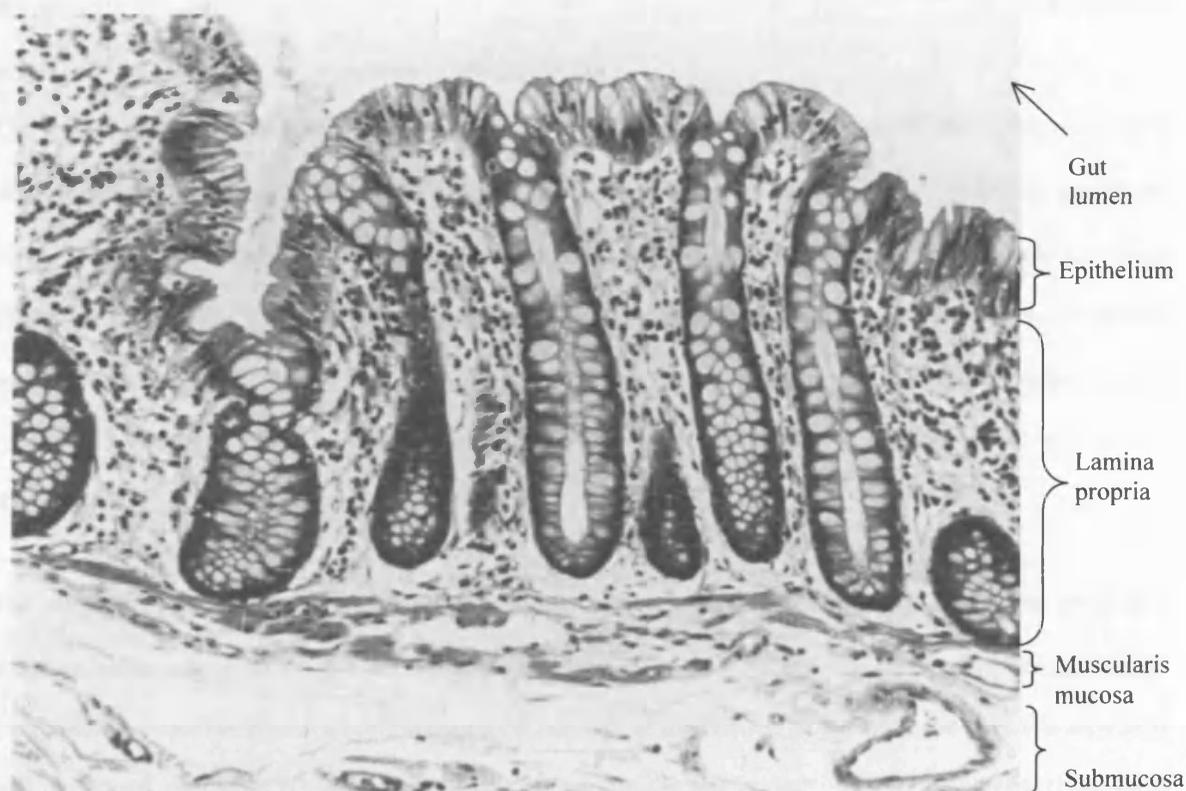
## 1.4 The Gastrointestinal Mucosal Immune System in IBD

As mentioned previously, the result of both Crohn's disease and ulcerative colitis is an intense inflammatory reaction in the bowel mucosa. Limited to the upper layers of the bowel wall in ulcerative colitis, the inflammation spreads throughout the bowel wall in Crohn's disease. In order to identify the cause of the inflammatory response (i.e. the aetiology), this section briefly describes the cellular anatomical framework within which the inflammation occurs.

### 1.4.1 Anatomy of the colon wall

There are four distinct functional layers running along the length of the gastrointestinal tract. The layer immediately in contact with the lumen is the mucosa; histologically divided into the epithelium, the lamina propria and the muscularis mucosae (figure 1.8). Below the mucosa is a layer of loose connective tissue, containing minor nerves, lymph and blood vessels, called the submucosa. The layer of smooth muscle surrounding the submucosa is the muscularis propria, the rhythmic contraction and relaxation of which is the basis of peristalsis. The outermost layer is the adventitia (or serosa where the gut wall is not attached to the abdominal wall), consisting of loose supporting tissue and conducting the major nerves and blood vessels to and from the gut wall.

Figure 1.8 – Anatomy of colonic mucosa (from Brooke et al, 1977<sup>5</sup>)



The mucosa is thought to be the point at which the inflammatory process leading to IBD begins and is therefore discussed in detail below. The discussion focuses on colonic mucosa, as that is the origin of all the samples used in this study.

#### **1.4.1.1 The epithelial layer**

The cells of the mucosal epithelium are in direct contact with the luminal contents and must serve two main functions; to absorb nutrients and to act as a barrier to potentially harmful foreign antigens. The presence of tight junctions between adjacent epithelial cells ensures the impermeable nature of healthy epithelium to foreign antigens and other large molecules.

A number of different cell types make up this one-cell thick lining. The majority of these are columnar absorptive cells (enterocytes), which in the colon absorb electrolytes and water via the sodium-potassium ion ( $\text{Na}^+\text{-K}^+$ ) activated ATPase-driven transport system<sup>18</sup>. They have numerous surface microvilli to increase the absorptive area.

Increasing in number from the caecum to the rectum, goblet cells secrete mucus to lubricate the passage of digestive matter. This layer of mucus is the first physical barrier to foreign antigens, as well as protecting the surface of the epithelium from variations in luminal pH<sup>103</sup>. The mucus is made up of mucin molecules, and Distinct subpopulations of goblet cells exist within the human gut mucosa, secreting distinct types of mucin glycoproteins<sup>104</sup>. In IBD affected mucosa the types of mucin produced are altered<sup>105-108</sup> and the numbers of goblet cells is reduced<sup>109, 110</sup>, compared to healthy mucosa.

The epithelium also contains large numbers of lymphocytes (intraepithelial lymphocytes; IELs), the majority of which express T lymphocyte markers<sup>111</sup>. Of these, about 80% are of the suppressor / cytotoxic phenotype (CD8+), with 10-20% expressing the helper phenotype (CD4+)<sup>112, 113</sup>. Intraepithelial lymphocytes have been quantified as typically 20 per 100 jejunal epithelial cells, 13 per 100 ileal epithelial cells and 5 per 100 colonic epithelial cells<sup>114, 115</sup>. Together with the immune cells located in the lamina propria these make up the immune system of the mucosa, the MALT (mucosa-associated lymphoid tissue). In the gastrointestinal tract, this is known as the gut-associated lymphoid tissue or GALT.

#### **1.4.1.2 The lamina propria**

The lamina propria is the layer beneath the basal membrane of the epithelial sheet. In the colon the lamina propria contains a layer of collagen just below the basement membrane of the epithelial sheet<sup>18</sup>. Throughout the gastrointestinal tract the lamina propria contains the majority of the GALT. In the colon the GALT is especially well developed, with large lymphoid nodules that extend into the submucosa\*. The epithelium overlying these lymphoid nodules is specialised to take advantage of the lymphocyte and immune cell population directly

---

\* In the ileum these are called Peyer's patches.

underneath. Along with an absence of crypts and villi, there is a localised decrease in the number of goblet cells, resulting in a thinner mucus layer. Follicle associated epithelium also contains specialised epithelial cells, called mucosal (M) cells. These factors act in concert to enable random immunological sampling of the luminal contents.

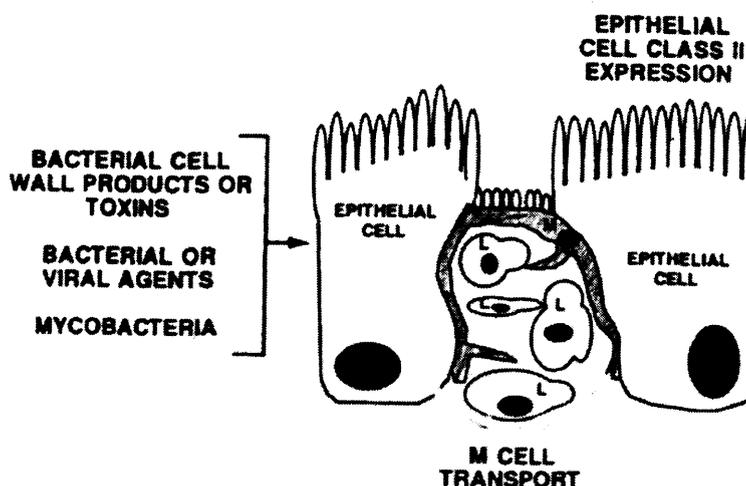
#### 1.4.2 Anatomy of the GALT

The mucosal surfaces of the body are the most vulnerable to foreign antigens. As a result the MALT encompasses the majority of the total lymphoid cell population of the body<sup>116</sup>. Of these, the gastrointestinal tract is the most exposed; feeding introduces large numbers of antigens directly and the gastrointestinal tract is home to millions of commensal bacterial cells (and their associated by-products). Therefore the GALT plays an important role in keeping the body free of disease and achieves this by containing a large number of immune cells. In fact, the volume of immune cells in the gut mucosa in the unchallenged state would represent chronic activation in the peripheral blood. This section briefly introduces the important features of the colonic GALT system with regard to IBD.

##### 1.4.2.1 M cells

The M cell is primarily a transport cell, transporting luminal antigens across the epithelial layer into the lymphoid tissue underneath. The antigens presented to the lymphoid tissue are identical to the antigens present in the lumen, as the M cell does not modify the antigens in any way<sup>117</sup>. M cells are deeply invaginated on the lamina propria face, forming deep pockets into which dendritic cells, macrophages, B and T lymphocytes can accumulate (figure 1.9)<sup>118</sup>. In this way a large random sample of luminal antigens are 'seen' by the GALT. M cells are not specific to the gut and are found throughout the MALT<sup>119</sup>. Unlike other intestinal enterocytes, M cells do not present antigen to CD4+ T lymphocytes, as they do not express the MHC II molecule<sup>118, 120</sup>.

*Figure 1.9 – Invagination of the M cell on the lamina propria face (from MacDermott, 1994<sup>118</sup>)*



### 1.4.2.2 A typical immune response<sup>119</sup>

Before detailing the specific features of the immune response in IBD, a brief description of a typical immune response is warranted. A very detailed picture of the molecules involved in the immune response has been elucidated over the years, but this discussion provides only a basic overview in order to understand the aetiology theories discussed in section 1.5.

There are two broad types of immune response, specific and non-specific. All the cells of the immune system, regardless of whether they respond to specific or non-specific antigen, communicate by producing and reacting to small peptides called cytokines. Cytokines act on and are secreted by, cells of the immune system, but can also be produced by cells that are not strictly part of the immune system; epithelial cells, for example<sup>121</sup>.

Non-specific immune responses are mediated by cells of the macrophage/monocyte line, which act in a phagocytic manner, absorbing foreign material from the intercellular environment and breaking it down within the cell. These phagocytic cells can respond to multiple foreign antigens; thus the response is non-specific.

The non-specific immune response can initiate the specific immune response. Phagocytes display MHC molecules and can therefore act as antigen presenting cells (APCs). The display of foreign antigen / MHC complex on the APC's surface activates the specific arm of the immune system, the lymphocytes.

The fragment of antigen displayed on an APC is termed the antigen's epitope. Each lymphocyte has the ability to recognise a specific epitope and will only become active after binding to it. There are two major types of lymphocytes, T and B. The T lymphocyte group is further divided, first into  $T_{\text{helper}}$  ( $T_{\text{h}}$ ), which are  $CD4^+$  and  $T_{\text{cytotoxic}}$  ( $T_{\text{c}}$ ), which are  $CD8^+$  cells. All the different types of lymphocyte have differing roles in the specific immune response.  $T_{\text{h}}$  cells are the first to be activated by APCs and these in turn activate both  $T_{\text{c}}$  and B cells. All types of T and B cell undergo clonal expansion. The  $T_{\text{c}}$  lymphocytes induce cells displaying their specific epitope to die and cells of the macrophage/monocyte line then phagocytose the remains of these cells.

B cells act by producing epitope specific antibodies. The antibodies attach to both cell surface associated and extracellular antigens. The attachment of antibody to epitopes on an infective agent (e.g. virus) inhibits its ability to infect further cells and also acts as a marker for phagocytic cells to phagocytose the antigen-antibody complex. The attachment of antibody to cell surface associated antigen induces the activation of complement, a cascade of proteins and enzymes that may result in the death of the cell exhibiting the foreign antigen.

Many features of the immune system also act to control the immune response. For example, immunoglobulins can suppress the immune response by masking antigen epitopes, or by clearing antigen from the environment. T lymphocytes can also suppress the immune system, although the exact mechanism by which they do so is not clear. It is thought that  $T_{\text{h}}$  lymphocytes become T suppressor ( $T_{\text{s}}$ ) cells by altering the cytokines they produce in response to unknown stimuli. The cytokines produced by the  $T_{\text{s}}$  cells then act to down-regulate the responses of the other immune cells in the local environment.

### 1.4.2.3 T lymphocytes

The T cell receptor (TCR) is the defining marker of a T lymphocyte. There are two types of TCR,  $\alpha\beta$  and  $\gamma\delta$ .  $\gamma\delta$  T lymphocytes are of particular importance in the gut mucosa as they make up 5-15% of small intestinal intraepithelial lymphocytes and up to 40% of colonic intraepithelial lymphocytes<sup>111</sup>, whilst accounting for only 5% of T lymphocytes in the peripheral tissues<sup>119</sup>. Therefore any functional differences between the two types of T lymphocyte may have a particular significance in the gastrointestinal tract.

#### *$\gamma\delta$ T lymphocytes*

The main functional difference between the two TCR types is that all  $\alpha\beta$  T cells are MHC restricted, whereas most  $\gamma\delta$  T cells are not<sup>122</sup>. Functionally this means that  $\alpha\beta$  T cells exhibit antigen specificity whereas  $\gamma\delta$  T cells do not. Thus  $\gamma\delta$  T cells seem to act as 'sentinels' in the lamina propria, 'mopping up' excess antigen so as to avert a full-blown immune response. The  $\gamma\delta$  TCR heterodimer is not always disulphide linked and the two chains may exist as monomers<sup>119</sup>; however, the functional significance of this is unclear.

The secretion of cytokines, chemokines and epithelial growth factors by epithelial  $\gamma\delta$  T cells, suggests a role for these cells in the repair of epithelial damage, such that seen in IBD. There are elevated numbers of  $\gamma\delta$  T cells at involved sites compared to uninvolved sites from the same IBD patient<sup>123</sup>, whilst untreated IBD patients have higher levels of  $\gamma\delta$  T cells in their peripheral blood than control patients<sup>124</sup>. However, the mechanism by which  $\gamma\delta$  T cells may contribute to IBD pathology remains unclear.

#### *Th1 / Th2 T lymphocyte profiles*

On the basis of cytokine profiles,  $\alpha\beta$  T<sub>h</sub> cells from the mouse can be divided into at least two groups, Th1 and Th2<sup>125</sup>. Th1 responses tend to promote cell-mediated reactions. Th1 cells promote B cell help to an extent, but if the concentration of Th1 cells is high, then B cells are suppressed. In contrast, Th2 responses tend to encourage a humoral response, promoting the production of antibodies (especially IgE) and eosinophil proliferation, i.e. a strong allergic type response (see Romagnani, 2000<sup>126</sup> for review).

Th1/Th2 polarity has been shown in a mouse model of IBD. Mice genetically engineered to be deficient in interleukin (IL) -12, interferon gamma (IFN $\gamma$ ) or IL-4 were treated with trinitrobenzene sulfonic acid (TNBS), which is a standard method of inducing experimental colitis. The mice developed distinct forms of colitis, with Th1 responses associated with a fatal, acute and transmural colonic lesions, whilst Th2 responses inducing diffuse atopic changes in crypts and the mucosal layer<sup>125</sup>. This Th1/Th2 distinction has been applied to human IBD on the basis of cytokine expression. T cells from Crohn's disease patients express IL-12, IFN $\gamma$  and lymphotoxin, and thus Crohn's disease is said to show a Th1-like profile. On the other hand, T cells in ulcerative colitis tend to express IL-4, -5 and -9, and ulcerative colitis is described as having a Th2-like profile<sup>25</sup>.

However, it must be noted that the Th1/ Th2 phenotype was originally described in the mouse and the cytokine profiles in humans are far less clear-cut. It is possible that a true Th1/Th2 phenotype may only ever be determined in mice, and caution must be applied when attempting to extrapolate mouse-model data to human diseases.

#### ***1.4.2.4 B lymphocytes***

The first antibody produced by all B lymphocytes, regardless of position within the body, is IgM, though this usually switches to another type as the response matures<sup>119</sup>. In the gut mucosa the majority of the B lymphocytes express IgA<sup>127</sup>. IgM is the second most common class of antibody in the gut, followed by IgG<sup>103</sup>. In the lamina propria, IgA is mostly found in its dimeric form; by contrast, in peripheral blood serum it is usually monomeric<sup>119</sup>. Dimeric IgA is secreted into the lamina propria by B lymphocytes and then transported into the gut lumen by enterocytes. A glycoprotein called 'secretory component', expressed by the enterocytes, acts as a receptor for the dimeric IgA, preventing bacterial cells from adhering to the enterocytes lining the lumen and thus inhibiting an immune response to the bacterial antigen<sup>128</sup>.

#### ***1.4.2.5 Other immune cells***

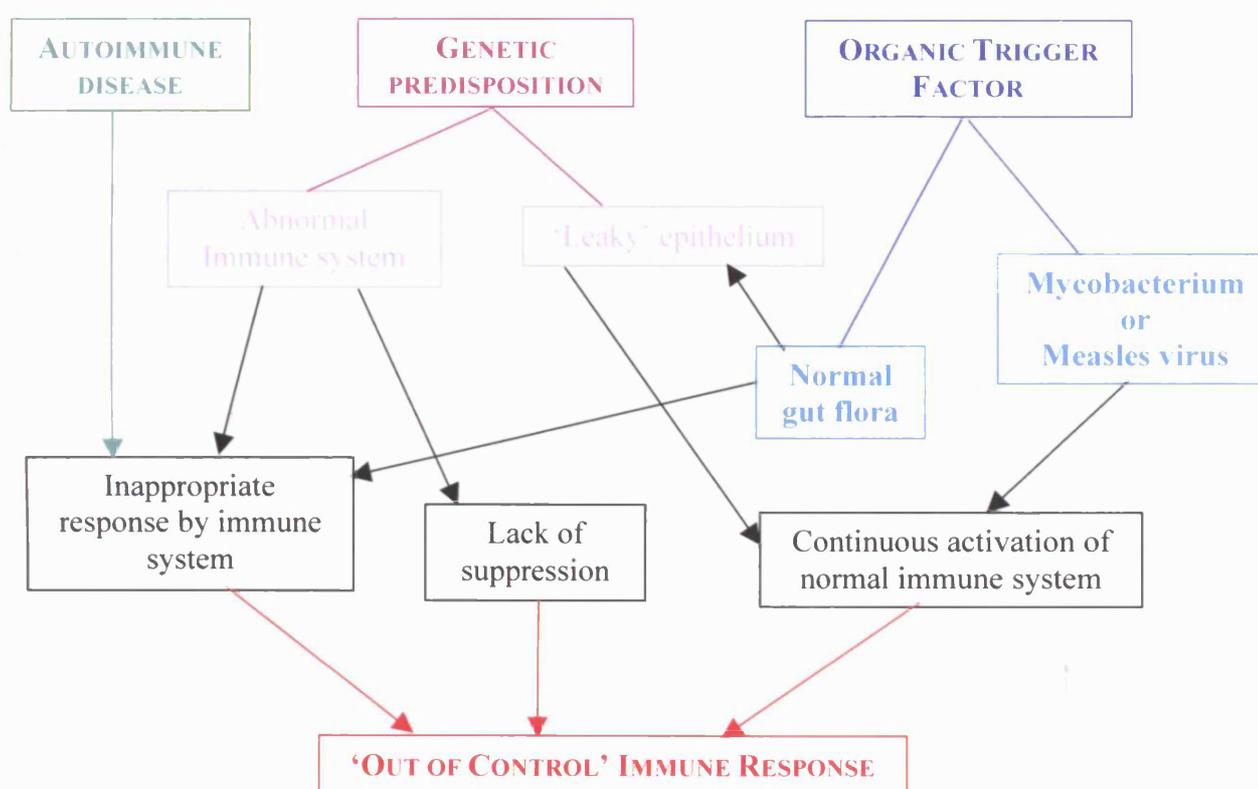
Neutrophils are the archetypal phagocytic cell of the non-specific acute inflammatory response. They play an important role in ulcerative colitis. Not present in normal mucosa, large numbers of neutrophils migrate into the crypts, forming crypt abscesses (figure 1.7). Ultimately these destroy the mucosal architecture and result in shedding of the epithelial layer.

Other important mucosal leukocytes include phagocytic cells, such as macrophages and so-called professional APCs, such as dendritic cells. As these present antigens to T<sub>h</sub> lymphocytes they are important in the initiation of the immune response.

## 1.5 Aetiology

A major source of difficulty in determining the aetiology of IBD has been sorting out the true primary causes from the secondary characteristics. IBD pathogenesis has both genetic and environmental components. Many of the immunological features of IBD have been determined, but it is unknown whether these represent causative or secondary features. There are three main lines of thought on the aetiology of IBD. The first is that IBD results from a normal immune response to a persistent infection, and section 1.5.1 discusses two of the candidates that have been investigated. The second theory is that IBD is the result of a genetic predisposition (section 1.5.2). The third theory is that IBD results when the immune system mounts an inappropriate response to normal bacterial or self-antigens (section 1.5.3). Figure 1.10 attempts to summarise these theories.

Figure 1.10 – IBD aetiology theories



### 1.5.1 Persistent Infection

#### 1.5.1.1 *Mycobacterium avium subspecies paratuberculosis*

Since colonic Crohn's disease was first distinguished from tuberculosis<sup>2</sup> an organic cause for the disease has been sought. A ruminant IBD similar to Crohn's disease in humans (Johne's disease) is caused by *Mycobacterium avium subspecies paratuberculosis* (MAP)<sup>129</sup>. The notion that MAP also causes human Crohn's disease has recurred frequently since 1932. However, it is difficult to culture or otherwise show the presence of mycobacteria and thus this theory has generally been disregarded<sup>51</sup>. However, the idea that Crohn's disease may have a mycobacterial aetiology has not disappeared and a review of the evidence in 1989<sup>130</sup> concluded that there was sufficient evidence to warrant further investigation.

Improvements in technology have allowed the presence or absence of *MAP* to be more accurately defined and this issue has recently been revisited by a number of groups. The *MAP* insertion sequence 'IS900'<sup>131</sup> and a serological immune response to *MAP* antigens has been demonstrated in Crohn's disease tissues<sup>132, 133</sup>. In another study *MAP* was cultured from 6 of 7 Crohn's disease resected tissues (86%), compared to 2 of 36 controls (5.6%)<sup>134</sup>. Thus the evidence supporting a role for *MAP* in Crohn's disease is mounting.

#### ***1.5.1.2 Measles***

The theory that exposure to the measles virus (either by infection or vaccination) could be involved in the aetiology of IBD has proved extremely controversial since it was first proposed in the mid-1990s. Measles virus particles, protein and RNA have been reported in Crohn's disease tissues<sup>135-137</sup>. However, these studies have attracted much criticism regarding the exposure of the control groups to the measles virus. Many studies have since been done to establish whether exposure to the measles virus does in fact increase the risk of Crohn's disease. A case-controlled study investigating measles virus exposure in a cohort of 140 IBD patients found no evidence to support the hypothesis<sup>138</sup>. The polymerase chain reaction (PCR) approach is widely regarded as the most reliable technique for detecting viral transcripts in host tissues. However, PCR studies by numerous groups including those which originally proposed the theory, have failed to detect measles virus in IBD tissues<sup>139, 140</sup>. The controversy over the measles, mumps and rubella (MMR) vaccine has led many parents to refuse the vaccine for their children. However, the current consensus is that exposure to measles does not cause Crohn's disease<sup>51, 141-144</sup>.

### **1.5.3 Genetic predisposition**

#### ***1.5.2.1 Defective mucosal barrier***

An increase in intestinal permeability results in the 'leakage' of antigens from the lumen into the lamina propria and submucosa. This increase in absorption may lead to an aggravated immune response, which could cause the tissue damage seen in IBD. In ulcerative colitis, it seems that the 'leaky gut' may result from abnormalities in the mucin layer. Gut mucin molecules are subject to post-translational modification prior to secretion. The mucin MUC2 must be sulphated prior to secretion into the mucous layer and there is evidence that this process is disrupted in ulcerative colitis<sup>145-147</sup>. Unsulphated MUC2 is not secreted, resulting in a decrease of MUC2 in the mucosal layer<sup>148, 149</sup>, resulting in an abnormal mucus layer, which could potentially result in increased permeability to luminal antigens.

The 'leaky gut' has been associated with Crohn's disease by a number of groups<sup>25, 150</sup>. Some groups have found that symptom-free, first-degree relatives of Crohn's disease patients also have an increased intestinal

permeability<sup>151, 152</sup>. Although this finding seems to depend on the methodology used<sup>25</sup> it is an important one. If the leaky gut is hereditary, it could represent an important factor predisposing patients to Crohn's disease. Although the nature of the 'leaky gut' is different in Crohn's disease and ulcerative colitis, the effect would be the same, i.e. the barrier function of the epithelium is impaired allowing the many inflammatory cells in the lamina propria uncontrolled contact with the normal colonic flora, thus initiating an inflammatory response.

#### **1.5.2.2 NOD2**

Of the many loci associated with IBD (section 1.3.4.3), to date only one has yielded a candidate gene. Three recent studies have linked mutations in the *NOD2* gene to Crohn's disease<sup>100-102</sup>. *NOD2* activates nuclear factor kappa B (NF- $\kappa$ B), is expressed exclusively by monocytes<sup>153</sup> and has been shown to confer responsiveness to bacterial lipopolysaccharides<sup>154</sup>. All three studies showed that the mutations in the *NOD2* gene resulted in truncation of the *NOD2* protein. In one of the studies embryonic kidney cell lines were transfected with wild-type and mutated *NOD2*<sup>102</sup>. When challenged with bacterial lipopolysaccharides there was a lower NF- $\kappa$ B activity in the mutant cell lines. This does not fit with the theory that NF- $\kappa$ B activation is perpetuated in active CD tissues<sup>155</sup>.

*NOD2* also contains two caspase recruitment domains (CARD) at the N-terminal. The CARD motif is structurally and functionally related to the death domain and the involvement of these protein domains in the apoptotic pathway implies a tentative link for the role of *NOD2* in apoptosis<sup>156</sup>.

It is plausible that the 20% of Crohn's disease cases that are familial<sup>157</sup> could be attributed to *NOD2* mutations; mutations in *NOD2* are claimed to be responsible for one in five of CD cases<sup>79</sup>. However, the genetic aetiology in the remaining 80% of Crohn's disease cases remains unknown, as does the precise mechanism of action of the truncated *NOD2* protein in the development of Crohn's disease.

#### **1.5.3 Inappropriate immune response**

It is not known whether the immune system abnormalities observed in IBD represent primary or secondary disease features. An inappropriate immune response could also be the result of a genetic predisposition to IBD, i.e. the protein product of an affected gene could malfunction either in isolation or in response to an environmental trigger. The end result however is the same, i.e. chronic inflammation. Therefore, the dysregulated immune system may well be the result of a genetic predisposition; however, until further studies have established the genes and mechanisms involved, the immunopathology of IBD is discussed in terms of the observable phenotype.

### ***1.5.3.1 Autoantibodies***

The observation that some ulcerative colitis patients produce anti-neutrophil autoantibodies<sup>158, 159</sup> has led to the idea that ulcerative colitis may be an autoimmune disease. Another potential autoantibody against an antigen designated P40 has also been investigated (reviewed by Radford-Smith, 1997<sup>128</sup>). This antigen has been detected in the goblet cells of the normal terminal ileum and colon, but not the rectum. The co-localisation of IgG1 and complement to P40 in inflamed ulcerative colitis tissue and the finding of P40 in the eye and joints supports the notion that this antigen may be important in the pathogenesis of ulcerative colitis, and that ulcerative colitis may be an autoimmune disease.

### ***1.5.3.2 Immunoglobulins<sup>118</sup>***

In the normal intestine there is a predominance of IgA-positive B cells. However, in IBD affected mucosa IgG-positive B cells predominate. IgA molecules protect the passively by aggregation and immune exclusion, rather than by activating an intense inflammatory response. However, IgG subclasses are associated with the activation and recruitment of non-antigen specific immune cells and activation of the complement pathway. The IgG subclasses thought to be most important in IBD include IgG<sub>2</sub> in Crohn's disease, IgG<sub>3</sub> in ulcerative colitis and IgG<sub>1</sub> in both.

### ***1.5.3.3 Response to normal gut flora***

Another aetiology theory is that there is a dysregulated immune response to normal gut bacterial antigens. Transgenic mice that are IL-2-deficient or express the human HLA-B27 gene normally develop colitis spontaneously, but when these mice are raised in sterile, germ-free conditions, colitis does not occur. In support of this hypothesis, introduction of a pathogen-free flora can then induce colitis in these mice<sup>160</sup>. However in human IBD, onset of disease does not correlate to bacterial colonisation of the gut, which occurs shortly after birth. There is a Th1 bias that has been noted in Crohn's disease, and this does support the notion of an abnormal immune reaction to bacterial antigens. However, lack of a Th1 bias in ulcerative colitis, means that this theory does not explain the chronic inflammation in that disease.

### ***1.5.3.4 Immune response suppression***

T cells isolated from IBD patients and stimulated with autologous monocytes and B cells showed defective suppression, whereas normal suppressor function was seen with controls<sup>166</sup>. In another study IBD derived epithelial cells failed to stimulate T<sub>s</sub> cells<sup>165</sup>. In both experiments the ability to activate T<sub>h</sub> cells was comparable to that seen with controls, implying a specific failure to suppress the immune response in IBD.

### ***1.5.3.5 Epithelial cell antigen presentation***

The ability of intestinal mucosal cells to act as antigen presentation cells has been shown by a number of groups<sup>120, 161-163</sup>. The differential association of different HLA types in IBD (table 1.2) has led to the theory

that this may lead to abnormal antigen presentation abilities, and thus result in abnormal T cell activation and / or regulation<sup>164, 165</sup>.

**Table 1.2 – Association of HLA types in IBD (modified from Fiocchi 1998<sup>25</sup>)**

<i>HLA type</i>	<i>Ulcerative colitis</i>	<i>Crohn's disease</i>
HLA-A3	Decreased	None
HLA-A9, HLA-B27		Decreased
HLA-A7, A11	Increased	None
HLA-B12, DR1, DQB1*0501		Increased
HLA-Bw52, Bw35, DQw1, DPB1*0901	Increased	
DR2, DRB1*1502	None, increased	None
DR4, DRw6	Decreased	None
DRB1*01, 07, 501, 1302, DRB3*0301		Increased
DRB1*03, DQB1*0602, *0603		Decreased

## 1.6 Aims & Objectives

---

The main aim of this study was to discover differences in the gene expression pattern between Crohn's disease and ulcerative colitis. It was anticipated that the study might also identify novel therapeutic targets, identify differences between the two for improved diagnosis and improve understanding of the pathology behind the two diseases. At the start of this project it was anticipated that numerous IBD samples would be collected and probed with the microarrays. However, only a few samples were applied to the microarrays, thus the study can only be considered as a pilot. The Bradford-Hill criteria<sup>167</sup> are widely used to establish causation, as opposed to association. These are discussed below in relation to this study:

*Strength & Consistency* - These factors depend very much on the number of samples investigated. In this study there were too few samples to reliably conclude anything about the strength or consistency of association.

*Specificity* - Specificity could be seen with some genes, where a gene may be present in all of one tissue group and absent in the other. However, the low numbers made it harder to see specificity in cases where the gene expression patterns were not so unambiguous.

*Temporality* - Theoretically, if multiple samples were taken and analysed from first presentation to resection in a each patient, specific early gene expression patterns which may impact on later gene expression, could be discovered. In this study all samples were from end-stage resected disease specimens; thus only late stage gene expression was measured.

*Biological gradient* -This criteria essentially relates to the quantitative aspect of gene expression levels. If a gene with higher expression levels in a sample with high disease activity has low expression in a sample with low disease activity, this could indicate a potential causative factor. In this study, the low number of samples combined with the large number of pathological disease features, meant that any potential gene expression gradients could not be established.

*Plausibility, Coherence & Analogy* - Gene expression patterns of interest were evaluated for these criteria through the annotation process. Information regarding previously determined gene function and expression patterns was discussed with regard to the current findings for plausibility, coherence and analogy.

*Experiment* - The microarray results of a few genes were confirmed by Q-PCR analysis, but the novel hypotheses generated in this study were not tested as part of this project.

As noted previously, the lack of samples remains a major limitation of this study. Thus, it is not possible to differentiate between association and causation of the gene expression patterns observed, and this study cannot then establish causation factors for IBD. However, this study retains the potential to generate multiple novel hypotheses, which could be tested in further experiments using different methods.

## 1.6.1 Previous gene expression studies

Table 1.3 – Previously determined gene expression patterns in IBD

<i>Gene</i>	<i>Expression Pattern</i>	<i>Method</i>
Connective tissue growth factor ( <i>CTGF</i> )	Over expressed in both UC and CD (surgical specimens)	RT-PCR <sup>168</sup>
Cyclooxygenase 2 ( <i>COX2</i> )	Over expressed in active IBD (colonic biopsies)	RT-PCR <sup>169</sup>
Down regulated in adenoma ( <i>DRA</i> )	Under expressed in IBD tissues	<i>In situ</i> hybridisation <sup>170</sup>
ENA-78	Over expressed in UC tissues	<i>In situ</i> hybridisation <sup>171</sup>
Eotaxin	Over expressed in IBD tissues	<i>In situ</i> hybridization <sup>172</sup>
IL-1 $\alpha$	Over expressed in IBD tissues	RT-PCR <sup>173</sup>
IL-6	Over expressed in CD, control levels in UC (mononuclear cells)	<i>In situ</i> hybridiation on colonic biopsies <sup>174</sup>
	Over expressed in IBD tissues	RT-PCR <sup>173</sup>
IL-6 receptor	Over expressed in CD (peripheral lymphocytes)	RT-PCR on peripheral lymphocytes <sup>175</sup>
IL-8	Over expressed in UC; moderately over expressed in CD (mononuclear cells)	<i>In situ</i> hybridiation on colonic biopsies <sup>174</sup>
	Over expressed in colonic biopsies from untreated UC and CD	RT-PCR <sup>176</sup>
	Over expressed in IBD tissues	RT-PCR <sup>173</sup>
Inducible nitric oxide synthase ( <i>iNOS</i> )	Over expressed in CD tissue sections	<i>In situ</i> RT-PCR <sup>177</sup>
	Over expressed in untreated IBD colonic biopsies	RT-PCR <sup>176</sup>
Interferon gamma (IFN $\gamma$ )	Over expressed in CD tissues	<i>In situ</i> hybridisation <sup>178</sup>
Rantes	Over expressed in CD tissues	<i>In situ</i> hybridisation <sup>178</sup>
Serum/glucocorticoid regulated kinase 1 ( <i>SGK1</i> )	Over expressed in small intestinal CD tissues	<i>In situ</i> hybridization <sup>179</sup>
Tumour necrosis factor alpha (TNF $\alpha$ )	Over expressed in IBD tissues	RT-PCR <sup>173</sup>

Table 1.3 details the results of some previous gene transcript expression studies in human IBD. The table is not intended to be exhaustive, but aims to give an indication of the types of genes that have been investigated previously and the methods used.

Most of these genes had been selected for investigation due to their actual or presumed involvement in the inflammatory process. Also, the methods used allowed the investigation of just a few genes\*. Thus, the expression of non-inflammation associated genes was historically being overlooked in IBD studies. These unknown gene expression patterns could hold vital clues for the three anticipated outcomes of this study. Thus, to allow the investigation of a large and unbiased group of genes it was decided to use microarray technology.

---

\* During the course of the project, two papers describing the use of oligonucleotide arrays to investigate IBD tissues were published<sup>180, 181</sup> and these are discussed further in chapters 4 and 5.

## **1.7 Investigating Gene Expression**

---

For the reasons explained above, a recently marketed commercial oligonucleotide microarray technology was used. However, there are a number of methods by which to measure gene transcript expression and this section briefly discusses these other methods, alongside an evaluation of the suitability of microarray technology for this study.

### **1.7.1 Classical methods**

There are a number of 'classical' methods for comparing the expression of genes between two tissue types. These include northern blotting<sup>182</sup>, S1 nuclease protection<sup>183</sup>, subtraction cloning<sup>184, 185</sup> and differential display<sup>186</sup>. These methods depend on the sequence of the gene being known and allow the comparison of only a few genes at one time. Also, the genes to be analysed must be decided upon at the beginning of the study making it difficult to identify novel gene expression patterns.

### **1.7.2 Massively parallel methods**

The progress of genome projects means that the sequences of many genes (and in some cases entire genomes) are known. Together with the development of new methods to study gene expression, hundreds or thousands of genes can be measured simultaneously.

#### ***1.7.2.1 Serial analysis of gene expression***

Serial analysis of gene expression (SAGE) was described in 1995<sup>187</sup>. SAGE involves the sequencing of short expressed sequence tags (ESTs) in a concatenated form to determine the gene expression profile of the tissue or cell the ESTs were originally isolated from. This method has been used to simultaneously investigate the expression of around 1000 genes<sup>187</sup>.

#### ***1.7.2.2 Microarrays***

The first microarray was described in 1993 by Uwe Maskos and Edwin Southern<sup>188, 189</sup>. Maskos and Southern described the synthesis of 72 oligonucleotide probes *in situ* on a glass microscope slide. From this 'hand-made' array described by Maskos and Southern, the production of microarrays began to be automated and high-speed robotic printing methods were first used for the generation of cDNA microarrays by Schena *et al*<sup>190, 191</sup>. As the generation of microarrays became automated, the accuracy of placing probes onto arrays increased. The ability to place hundreds and then thousands of probes on a single array enabled massively parallel gene expression analysis to become commonplace.

There are two kinds of microarray, cDNA and oligonucleotide, with probes of cDNA and oligonucleotide respectively. All microarrays consists of probe sequences immobilised on glass or nylon sheets. Labelled complementary RNA (cRNA) is created from the cell line or tissue sample of interest, to represent the mRNA transcripts in the original sample. The cRNA is labelled by either a radioactive or fluorescent moiety, and hybridised to the immobilised probes on the array. Any cRNA that binds to a probe sequence is then detected.

The position of each probe sequence on the array is known and so the sequence of the bound cRNA can be deduced.

#### *cDNA microarrays*

Sequenced cDNA is spotted onto glass microscope slides to create an array of cDNA clones. Two cRNA targets (control and experimental) are applied to each array, one labelled with a red dye and the other with a green dye. The array is then examined to determine spots where the only one dye is present, i.e. the represented gene is only expressed in one of the targets. The number of probes that can be spotted onto a single slide is increasing, but is limited by the accuracy of the spotting instrument, and the size of the spot.

#### *Oligonucleotide microarrays*

The main advantage of using a commercial oligonucleotide microarray as compared to a self-made cDNA microarray is ability to perform the expression analysis without having to create the array first. However, at the time this project was undertaken, the Affymetrix arrays did not represent the entire genome, and thus potentially significant gene expression patterns were not investigated. The representation of expressed sequence tags (ESTs) on the oligonucleotide arrays aimed to include as many genes as possible; however, unless an EST has been assigned to a full-length gene its expression pattern is not very useful. With the human genome-sequencing project now nearing completion this is less of a problem. The inclusion of probe sequences on a cDNA array do not depend on the progress of genome sequencing projects as the cDNA to be arrayed is isolated directly from the cell of interest. However, the sequence of each cDNA probe must be determined manually adding to the time it takes to do an experiment. The commercial array systems can be very expensive and cDNA arrays are much more economical. However, the main disadvantage of the cDNA array is the limited number of probes that can be placed onto a single array. With the Affymetrix™ GeneChip® microarray platform used in this study, up to 8,500 oligonucleotide probes are synthesized *in situ* on a single glass array.

In brief, the Affymetrix system involves the isolation of total RNA from the tissues or cells of interest. mRNA from the total RNA pool is used as the template for the synthesis of double stranded (ds) cDNA. The ds cDNA is in turn used as a template for cRNA synthesis. Biotinylated ribonucleotides are used for the synthesis of the cRNA resulting in a labelled target. The labelled cRNA target is applied to the immobilized oligonucleotide probes on the microarray and after a period of hybridisation, excess target removed. A phycoerythrin-streptavidin conjugate is then applied to the array. The streptavidin binds the biotin on the labelled ribonucleotides of the target. Thus the presence of the phycoerythrin molecule indicates the presence of the streptavidin-biotin complex and the target cRNA. The fluorescent phycoerythrin is detected by confocal scanning, and specialised software analyses the resulting image to generate the gene expression microarray data. This method is described in greater detail in section 2.2, and schematics of the method are shown in figures 2.2 and 2.4.

**CHAPTER TWO – MATERIALS & METHODS**

## 2.1 Collection of Tissue Samples

### 2.1.1 Materials\*

Phosphate buffered saline† (PBS) (Microgen Bioproducts #M34A)

Tenotomy scissors (autoclaved before each use)

21 gauge needles with 5 mL syringe

### 2.1.2 Method

The mucosal samples were collected directly from the operating theatre as soon as possible after resection, to minimise RNA degradation. With the assistance of the theatre staff, the author aimed to be present at least 5 minutes before removal of the colon from the patient. The colon was cut open and washed in tap water only if necessary to remove faecal debris. 20% phosphate buffered saline (PBS) was injected under the mucosa to separate the mucosal layer from the underlying submucosa. Strips of mucosa (approximately 1 cm by 10 cm) were cut away from the sample and immediately frozen in liquid nitrogen. If the specimen was from an IBD patient, a small piece from one end of the strip was cut (prior to freezing in liquid nitrogen), fixed in formalin and embedded in paraffin wax (see section 2.3). The frozen mucosa strips were weighed (table 2.1) and stored at  $-80^{\circ}\text{C}$ , until RNA isolation.

*Table 2.1 – Weights of samples used in microarray analysis*

<i>Pathological Feature</i>	<i>Crohn's disease</i>		<i>Ulcerative colitis</i>		<i>Controls</i>
	<i>Involved</i>	<i>Uninvolved</i>	<i>Involved</i>	<i>Uninvolved</i>	
n	5	4	5	3	6
Sample weight (g)					
Range	0.4 - 0.8	0.4 - 1.1	0.2 - 0.8	0.4 - 0.5	0.3 - 2.68
Mean	0.60	0.73	0.38	0.43	1.26

### 2.1.3 Clinical history of patients studied

151 samples were collected, from 11 Crohn's, 15 colitis and 24 colorectal cancer patients. 83 of these represent duplicate samples. Control samples were collected from patients undergoing resection for colorectal cancer. Macroscopically normal mucosa was collected at least 5 cm from the site of the tumour.

Only samples with good quality total RNA, and which resulted in a good quality target, as checked by hybridisation to a test array, were applied to the experimental microarrays. 23 samples were rejected due to poor quality total RNA or test array results. 15 samples were rejected as they had originated from the ileum or from defunctioned colon (which even if normal can develop 'diversion colitis', a condition outside the remit of this study). 7 samples were used at various points to practise the method, leaving 23 samples that were used in the

\* Subsequent materials sections only list materials and reagents that have not been described previously in this thesis.

† Diluted to 20% with sterile H<sub>2</sub>O

microarray experiments. Another 5 samples were collected from 3 other patients with familial adenomatous polyposis, an ileostomy reversal and indeterminate IBD; these samples were not used.

Table 2.2 summarises the clinical details of the participating patients. The mean age of the Crohn's disease group was 41.6 years, the ulcerative colitis patients had a mean age of 43.4 years and the control group had a mean age of 63.5 years. The average disease duration of the Crohn's disease group was 9.8 years and the average disease duration of the ulcerative colitis group\* was 5.8 years.

The indication for surgery was 'resistant disease' in all the IBD cases. Azathioprine had been prescribed only to the Crohn's disease patients; however, the steroid prescription was similar between the two IBD groups, with most patients having been on steroids for greater than 3 months prior to surgery.

#### **2.1.4 Pathological features of samples used in microarray analysis**

The relative proportion of microscopic inflammatory markers in each sample subject to microarray analysis was scored from an H&E section by an experienced consultant histopathologist (PNF). The IBD H&E sections were cut from a block adjacent to the microarray sample; therefore, these sections did not contain submucosa or muscularis propria. However, the control sections were cut from a normal block taken for routine histopathology and these were full bowel wall cross sections.

The scores were marked on a score sheet designed for the purpose and the observations recorded by the pathologist converted into numerical scores. The involved UC (UCi) samples tended to show more abnormal pathological features than the involved CD (CDi) samples. For example, the crypt architecture and distribution was more abnormal in the UCi samples compared to the other tissue groups. The lamina propria cellularity was increased in all of the involved IBD samples, however a much more marked increase was evident in the UCi compared to the CDi samples. The neutrophil infiltrate in the UCi samples was 3-10% on average, whereas the CDi samples showed an average neutrophil infiltrate of <2%. Giant cells and an increased proportion of intraepithelial lymphocytes were only seen in CDi samples. These observations are summarised in figure 2.1.

---

\* The clinical data for patients UC\_3, NI\_1, NI\_4 and NI\_6 was unobtainable, thus this figure does not include the data of patient UC\_3.

# Special Note

Page **36** missing from  
the original

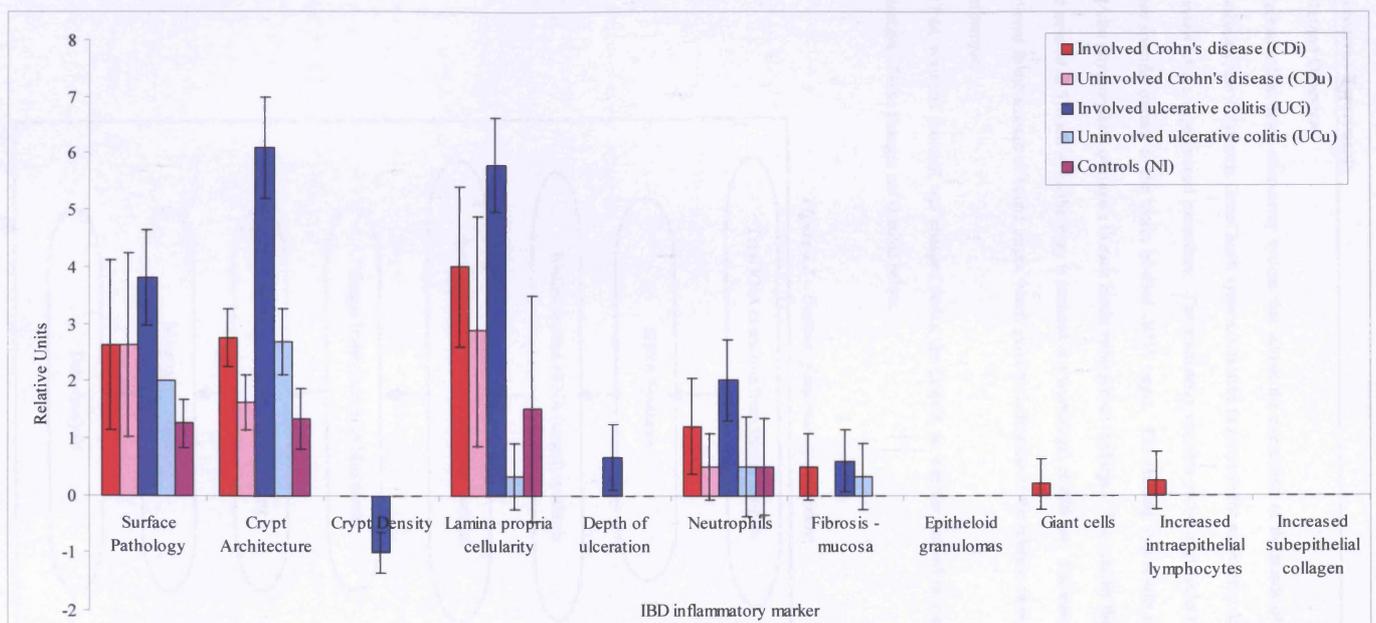
**Table 2.2 - Clinical features of participating patients**

<i>Patient*</i>	<i>Sex</i>	<i>Age on date of surgery</i>	<i>Disease Duration</i>	<i>Extent of disease</i>	<i>Current Therapy</i>	<i>Indication for surgery</i>
CD_1	F	68	3 years	Left colon	Azathioprine >3 months, Mesalazine >3 months	Resistant disease & fistula
CD_2	F	27	9 years	Terminal ileum & colon	Azathioprine >3 months, Mesalazine >3 months	Resistant disease
CD_3	F	37	15 years	Ileo-colonic anastomosis	Prednisolone <3 months	Resistant disease, stricture at ileo-colonic anastomosis
CD_4	F	45	12 years	Ileum & colon at site of previous anastomosis	Prednisolone >3 months, Hydrocortisone per op	Resistant disease & anastomotic stricture
CD_5	F	31	10 years	Right colon	Steroids >3 months, Azathioprine >3 months, Metronidazole per op, Hydrocortisone per op	Resistant disease & colo-duodenal fistula
UC_1	F	61	4 years	Total colon	Mesalazine >3 months	Resistant disease
UC_2	F	36	4 years	Rectum & Left colon	Prednisolone >3 months, Hydrocortisone per op	Resistant disease
UC_3**	M	30				
UC_4	F	54	10 years	Left colon	Prednisolone >3 months, Hydrocortisone per op	Resistant disease
UC_5	F	36	5 years	Total colon	Prednisolone <3 months, Hydrocortisone per op	Resistant disease
NI_1 **	F	72				
NI_2	F	44	<3 months	Rectal tumour		Cancer
NI_3	F	77	<3 months	Rectal tumour		Cancer
NI_4 **	F	65				
NI_5	F	72	<3 months	Ascending colon & rectum		Cancer
NI_6	M	51				

\* All patients were of white European ethnic origin, except for patient UC\_2 whose ethnic origin was south Asian.

\*\* Clinical data unavailable for these patients.

Figure 2.1 – Relative measure of pathological markers of inflammation in IBD tissues



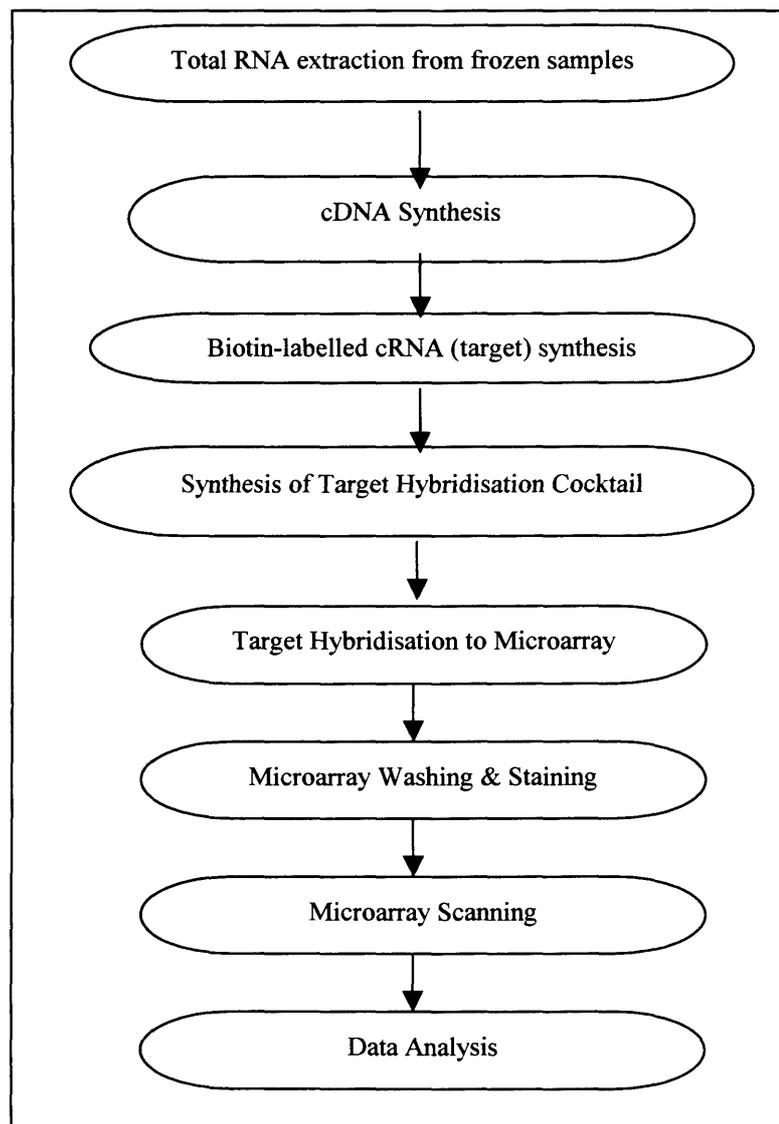
## 2.2 Microarray Experiments

### 2.2.1 Method Overview

An oligonucleotide based microarray system that allows the expression of thousands of genes to be studied simultaneously, the Affymetrix GeneChip® system, was used to generate the microarray data. Figure 2.2 shows an overview of the experimental procedure. The microarray contains probe oligonucleotides, which bind in a sequence specific manner to the biotin labelled cRNA target. The staining step results in the attachment of a streptavidin phycoerythrin conjugate to each biotin molecule on the target. The scanner then records the amount of light emitted at 570 nm when the array is scanned at a wavelength of 488 nm. The amount of light emitted is proportional to the amount of bound target, which gives an indication of the relative amount of transcript in the original sample.

The cDNA synthesis protocol was changed during the project, as was the method of staining the arrays post-hybridisation. These changes are detailed below.

*Figure 2.2 – Outline of microarray procedure*



### 2.2.2 RNA Isolation

#### 2.2.2.1 Materials

TRIzol Reagent (GibcoBRL #15596-018)

Electric homogeniser (IKA Labortechnik Ultra-turrax T8)

Chloroform (Fisher Reagents #BPE1145-1)

Isopropanol (Fisher Chemicals #P/7500/PB17)

Phenol/chloroform/isoamyl alcohol [25:24:1] (Ambion #9732)

DEPC-treated water (Ambion #9920)

Absolute Ethanol (Fisher Chemicals # E/0650DF/P17)

Oligotex Direct mRNA midi kit (Qiagen #70042)

Wash buffer OW1

Elution buffer

3M Sodium Acetate (Sigma #S-7899)

1% ethidium bromide gel (appendix A2.1)

Pellet paint (Novagen #69049-3)

#### 2.2.2.2 Total RNA Isolation

Frozen samples (stored at  $-80^{\circ}\text{C}$ ) were placed on dry ice immediately before homogenisation to minimise RNA degradation and homogenised in TRIzol (1 mL per 100 mg tissue) with an electric homogeniser in a class II hood. Care was taken to ensure that tissue particles were no longer visible to the naked eye. Once the tissue had been completely disrupted the sample was incubated at room temperature for 2 min to allow TRIzol mediated disruption at the cellular level. Chloroform (0.2 mL per 1 mL TRIzol) was added to allow dissociation of nucleoprotein complexes and the tube shaken vigorously for 15 s before incubating at room temperature for a further 2 min. The tissue / TRIzol / chloroform mixture was centrifuged at 4000 rpm for 15 min at  $4^{\circ}\text{C}$ . The mixture separated into a lower red layer containing protein and DNA, a middle interphase containing DNA and an upper aqueous layer containing total RNA. The upper phase was transferred into a second tube (taking care to avoid the interphase and lower layer) and an equal volume of phenol/chloroform/isoamyl alcohol was added for a second extraction. The reagents were mixed and then centrifuged at 4000 rpm for 10 min at  $4^{\circ}\text{C}$ . This resulted in the formation of 3 layers, with RNA residing exclusively in the upper aqueous layer as before.

#### 2.2.2.3 Organic extraction of total RNA

The upper layer was carefully transferred into a clean tube and isopropanol added (0.5 mL for every 1 mL TRIzol) to extract the total RNA. This mixture was gently agitated, aliquoted into 1.5 mL Eppendorf tubes and microcentrifuged at full speed for 30 min. The total RNA formed a small white pellet at the bottom of each tube. Each pellet was washed twice in 0.5  $\mu\text{l}$  80% ethanol to remove residual impurities and air-dried to remove all traces of ethanol. All RNA pellets from a single sample were pooled in 500  $\mu\text{l}$  DEPC-treated  $\text{H}_2\text{O}$ .

#### 2.2.2.4 mRNA isolation

The manufacturer's instructions were followed to isolate mRNA from 500-750  $\mu\text{g}$  of total RNA, using Oligotex mRNA mini columns. Briefly, latex oligo dT<sub>30</sub> beads bind the poly (A)<sup>+</sup> RNA molecules, whilst the remaining total RNA is removed. The poly (A)<sup>+</sup> RNA (mRNA) is then freed from the beads with elution buffer. All buffers are supplied with the kit.

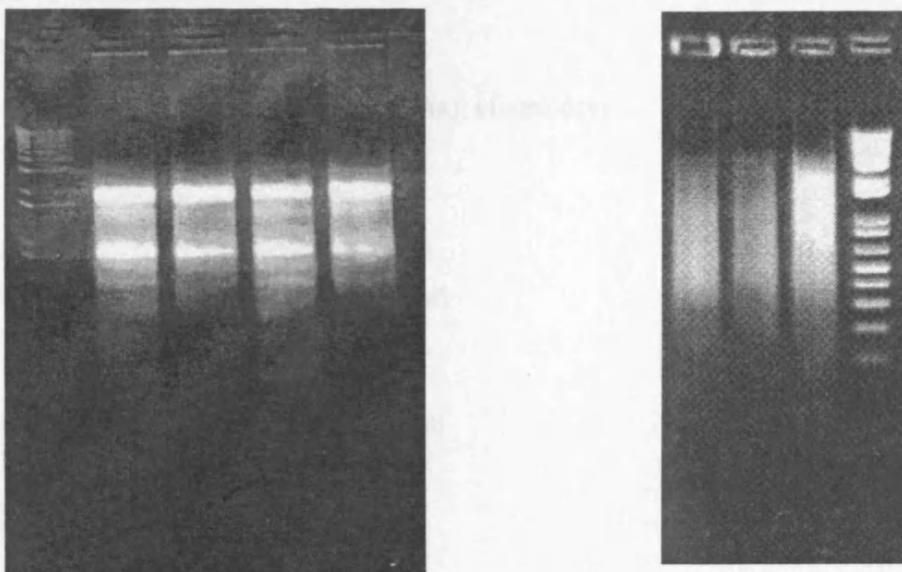
500  $\mu\text{l}$  2X binding buffer and 45  $\mu\text{l}$  Oligotex bead suspension were added to the total RNA (suspended in 500  $\mu\text{l}$  DEPC-treated H<sub>2</sub>O). The sample was gently mixed and incubated at 65°C for 3 min, to disrupt the secondary structure in the RNA. To allow hybridisation between the oligo dT<sub>30</sub> on the latex beads and the poly A tail of the mRNA molecules the sample was incubated at room temperature for 10 min. The samples were centrifuged at full speed for 2 min, to pellet the mRNA-containing resin. The supernatant was discarded and the poly (A)<sup>+</sup> RNA / resin pellet was washed twice with 400  $\mu\text{l}$  wash buffer OW2. The resin was re-suspended in the column with 100  $\mu\text{l}$  pre-heated elution buffer (70°C) and centrifuged at full speed for 30 s, to extract the poly (A)<sup>+</sup> RNA in the flow through. Fresh elution buffer was applied a further 2 times, resulting in the mRNA being eluted in a total of 300  $\mu\text{l}$  elution buffer.

#### 2.2.2.5 Organic extraction of mRNA

150  $\mu\text{l}$  sodium acetate, 900  $\mu\text{l}$  100% ethanol and 1  $\mu\text{l}$  pellet paint was added to the mRNA / elution buffer mixture, chilled in dry ice until the mixture became viscous and then centrifuged at full speed for 20 min. The supernatant was discarded and resulting pellet washed twice with 80% ethanol. After air-drying the mRNA pellet was re-suspended in 10  $\mu\text{l}$  DEPC-treated H<sub>2</sub>O.

#### 2.2.2.6 Measuring RNA quality and quantity

Figure 2.3 – Good quality total RNA (left) & cDNA on 1% EtBr gels



All RNA isolated (total RNA and mRNA) was subject to quality and quantity checks by gel electrophoresis and spectrophotometry. The quality of synthesised cDNA was checked by gel electrophoresis.

Ethidium bromide gels (see A 2.1) were used to visualise the samples. Good quality total RNA was determined by the presence of two distinct bands at ~2000 kB and ~900 kB, representing 28S and 18S ribosomal RNA. Intact mRNA populations also show these residual rRNA bands. Good quality cDNA showed a smear within the 1-5kB range of the DNA ladder run with each gel (figure 2.3).

A spectrophotometer was used to measure the quantity of RNA isolated. The expected  $A_{260}/A_{280}$  ratio for good quality total RNA was between 1.8 and 2.1.

### 2.2.3 cDNA Synthesis from mRNA

#### 2.2.3.1 Materials

##### *1<sup>st</sup> strand synthesis*

mRNA in DEPC-treated H<sub>2</sub>O

From Superscript Choice system (GibcoBRL #18090-019)

5X 1<sup>st</sup> strand buffer

0.1M DTT

10mM deoxyribonucleotides (dNTPs)

Superscript Reverse Transcriptase II (SSRT II) [200 U per  $\mu$ l]

HPLC purified primer [100 pmol/ $\mu$ l] (Sigma-Genosys Ltd)

Sequence 5' – GGCCAGTGAATTGTAATACGACTCACTATAGGGAGGCGG-(dT)<sub>24</sub> – 3'

##### *2<sup>nd</sup> strand synthesis*

EDTA

From Superscript Choice system (GibcoBRL #18090-019)

5X 2<sup>nd</sup> strand buffer

E. coli Ligase [10 U per  $\mu$ l]

E. coli Polymerase I [10 U per  $\mu$ l]

E. coli RNase H [2 U per  $\mu$ l]

T4 DNA Polymerase [5 U per  $\mu$ l]

##### *Organic extraction*

Phenol/chloroform/isoamyl alcohol [25:24:1] (Ambion #9732)

7.5 M Ammonium acetate (Sigma #A-2706)

### 2.2.3.2 1<sup>st</sup> strand cDNA synthesis

The kit manufacturer's method was followed, substituting the primer described in section 2.2.3.1 for the oligo (dT) and random primers supplied with the kit. The amount of reverse transcriptase (RT) was dependent on the starting amount of mRNA (4-5 µg), as detailed in table 2.3.

**Table 2.3 - Reagent volumes in 1<sup>st</sup> strand cDNA reaction**

	<i>Initial amount of mRNA</i>		<i>Final concentration or amount in reaction</i>
	<i>4 µg</i>	<i>5 µg</i>	
mRNA (in DEPC-treated H <sub>2</sub> O)	8 µl	7 µl	4-5 µg
T7-(dT) <sub>24</sub> primer (100 pmol/µl)	1 µl	1 µl	100 pmol
5X 1 <sup>st</sup> strand buffer	4 µl	4 µl	1X
0.1M DTT	2 µl	2 µl	10 mM
10 mM dNTP mix	1 µl	1 µl	500 µM each
Reverse Transcriptase (SSRT II)	4 µl	5 µl	800-1000 U
<b>Total reaction volume</b>	<i>20 µl</i>	<i>20 µl</i>	

After adding the primer to the mRNA, the reactions were incubated at 70°C for 10 min. At the end of the incubation they were rapidly spun and put on ice. The 1<sup>st</sup> strand buffer, DTT and dNTP were added to each tube in the order listed. The tubes were mixed and incubated at 37°C for 2 min, before adding the reverse transcriptase and incubating at 37°C for 1 h.

### 2.2.3.3 2<sup>nd</sup> strand synthesis cDNA synthesis

2<sup>nd</sup> strand synthesis was carried immediately after 1<sup>st</sup> strand synthesis. The 1<sup>st</sup> strand synthesis reactions were placed on ice and briefly centrifuged. The following reagents were added to each reaction in the order listed (table 2.4). The tubes were gently mixed, spun and incubated at 16°C for 2 h.

**Table 2.4 - Reagents for 2<sup>nd</sup> strand cDNA synthesis**

<i>Component</i>	<i>Volume (µl)</i>	<i>Final concentration or amount in reaction</i>
DEPC-treated H <sub>2</sub> O	91	
5X 2 <sup>nd</sup> strand buffer	30	1X
10mM dNTP mix	3	200 µM each
DNA Ligase	1	10 U
DNA Polymerase I	4	40 U
RNase H	1	2 U
<i>Final volume</i>	<i>150 µl</i>	

2 µl T4 DNA polymerase was then added to each tube and incubated at 16°C for a further 5 min to ensure that the cDNA termini were blunt. 10 µl EDTA was added to stop the reaction and the ds cDNA was either extracted immediately' as detailed in 2.2.3.4, or stored at -20°C for extraction at a later time.

#### 2.2.3.4 Organic extraction of cDNA

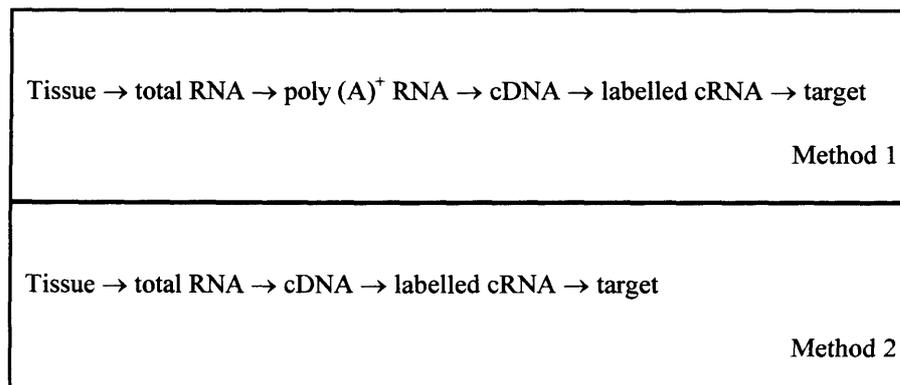
An equal volume of phenol:chloroform:isoamyl alcohol (162  $\mu$ l) was added to each cDNA reaction, mixed thoroughly and centrifuged at full speed for 5 min. This resulted in the formation of a lower milky layer and an upper clear layer, which contained the cDNA. The upper layer was carefully transferred to a clean 1.5 mL Eppendorf and 0.5 volumes of ammonium acetate, 2.5 volumes 100% ethanol and 0.5  $\mu$ l pellet paint added to it. This was mixed thoroughly and cooled in dry ice for  $\sim$ 10 min, until the mixture became viscous. The samples were centrifuged at full speed for 20 min at room temperature, resulting in the formation of a small pellet. The cDNA pellet was washed with 0.5 mL 80% ethanol and air-dried before re-suspending in DEPC-treated H<sub>2</sub>O (3  $\mu$ l per 1  $\mu$ g mRNA used in the reaction).

#### 2.2.4 Removal of mRNA isolation step

##### 2.2.4.1 Introduction

An alternative method that eliminated the need to isolate mRNA before synthesising cDNA was developed by Affymetrix and evaluated by the author. To compare the two methods, two samples were processed, an involved mucosal sample from an ulcerative colitis patient and a non-inflamed sample from a cancer patient (figure 2.4). Total RNA was isolated and extracted as described in sections 2.2.2.2 and 2.2.2.3. After checking the quality and quantity of the total RNA (section 2.2.2.6), poly (A)<sup>+</sup> RNA was isolated from 700  $\mu$ g total RNA (section 2.2.2.4) and used to synthesise cDNA (section 2.2.3), i.e. method 1 in figure 2.4. 40  $\mu$ g of total RNA was used to produce cDNA as detailed in section 2.2.4.3 below, i.e. method 2 in figure 2.4.

*Figure 2.4 – Schematic of the two methods being compared*



##### 2.2.4.2 Materials

Test Chip 1 (Affymetrix #510006)

Hu6800 subC (Affymetrix #510131)

Hu6800 subD (Affymetrix #510133)

### 2.2.4.3 cDNA Synthesis from total RNA

The basic reagents used for cDNA synthesis were the same as section 2.2.3, the only differences being volumes used and the incubation temperature of 1<sup>st</sup> strand synthesis. This modified technique allowed the synthesis of cDNA from 5-40 µg total RNA. The amount of reverse transcriptase used in the cDNA reaction depended on the amount of total RNA as detailed in table 2.5.

**Table 2.5 - Amounts of reverse transcriptase in 1<sup>st</sup> strand cDNA reaction**

<i>Total RNA (µg)</i>	<i>Superscript II reverse transcriptase 200 U/µl (µl)</i>
5.0 to 8.0	1.0
8.1 to 16.0	2.0
16.1 to 24.0	3.0
24.1 to 32.0	4.0
32.1 to 40.0	5.0

DEPC-treated H<sub>2</sub>O was added to the RNA to bring the volume of total RNA, SSRT II and water to 12 µl, although the enzyme was not added until the final step. The other reagents were added as detailed in table 2.3. The reactions were incubated at 42°C, for 2 min before the addition of SSRT II and for 1 h after its addition. The 2<sup>nd</sup> strand cDNA reaction and organic extractions were carried out as described above (sections 2.2.3.3 and 2.2.3.4 respectively). The double stranded cDNA pellet was dissolved in a volume of DEPC-treated H<sub>2</sub>O dependant on the amount of total RNA in the reaction (table 2.6).

**Table 2.6 - Re-suspension volumes for double stranded cDNA**

<i>Total RNA (µg)</i>	<i>Re-suspension volume (µl)</i>
5.0 to 8.0	1.5
8.1 to 16.0	3.0
16.1 to 24.0	4.5
24.1 to 32.0	6.0
32.1 to 40.0	7.5

### 2.2.4.4 Generation of target RNA

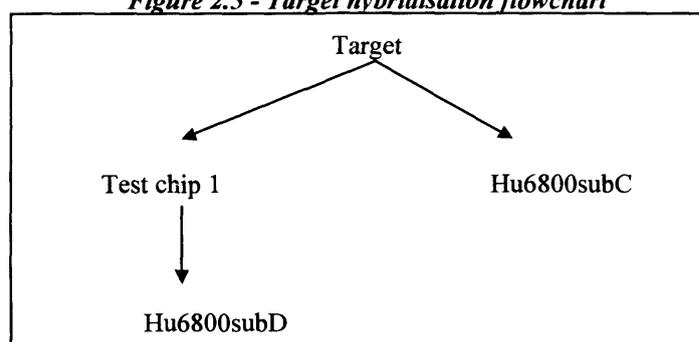
Table 2.7 details the four targets that were synthesised for the comparison of methods. 1.5 µl of cDNA made by each method was used to generate the hybridisation targets (sections 2.2.6 to 2.2.9).

**Table 2.7 - Targets generated for method comparison**

<i>Target Number</i>	<i>Method used to generate cDNA</i>	<i>Description</i>
1	Method 1 (poly (A) <sup>+</sup> RNA)	Normal mucosa
2	Method 2 (total RNA)	Normal mucosa
3	Method 1 (poly (A) <sup>+</sup> RNA)	Inflamed mucosa
4	Method 2 (total RNA)	Inflamed mucosa

The arrays were washed and stained with streptavidin phycoerythrin as described in section 2.2.10.3 and scanned as described in section 2.2.11. The targets were first hybridised to Test 1 arrays, to check the integrity of the cRNA. In order to establish the reusability of the targets synthesised by method 2, the target already used on the test chips was hybridised to Hu6800subD, whilst the Hu6800subC chips had fresh target hybridised to them (figure 2.5).

**Figure 2.5 - Target hybridisation flowchart**



#### 2.2.4.7 Results

The data was analysed using Affymetrix GeneChip software and the Affymetrix data-mining tool, DMT 1.0. The intensity scale factor for all four arrays was set to 75 to enable comparisons between the arrays. Targets 1 & 2 were hybridised to the Hu6800\_subC and Hu6800\_subD arrays, whereas targets 3 & 4 were hybridised to the Hu6800\_subC array only. The null hypothesis was tested; i.e. 'there is no significant difference between targets generated by the two methods'. This was tested by calculating the percentage of probe sets with a fold change of more than 3.0 (table 2.8). 3.0 was set as the cut-off point, as fold changes of less than three cannot be discriminated reliably from background noise. Fold changes of more than 5.0 were checked visually on the scanned images, using the probe cell viewing facility. For the target 1 C chip, it was found that four of the differences corresponded to extraneous dust particles on the image and these were disregarded in the final calculations (table 2.8).

**Table 2.8 - Summary of comparison analyses**

<i>Arrays Compared</i>	<i>SF value</i>	<i>% probe sets &gt; 3.0 fold change</i>	<i>% probe sets &gt; 3.0 fold change (minus dust aberrations)</i>
Normal Mucosa C chip	7.59	1.11	0.89
Inflamed Mucosa C chip	7.06	1.82	1.82
Normal Mucosa D chip	9.86	1.54	1.54

The targets made as per method 1 can be reused. To investigate the quality of the second use 'method 2' targets, the D chips had been hybridised to target that had been used once already. The intensity scale factor was again set at 75 to keep the SF parameter value less than 10. As before, method 1 chips were used as a baseline for

method 2 chips. Unfortunately, one of the inflamed tissue targets (target 3) hybridised poorly to the D chip and therefore the D chip could not be compared for the inflamed tissues.

#### **2.2.4.8 Conclusions**

Table 2 shows that over the three chips the percentage of probe sets with a fold change of greater than 3.0 are between 0.89% and 1.82%. A 1-2% difference would be predicted even between two duplicate samples, as minor variations in hybridisation are expected<sup>192</sup>. This study therefore supports the null hypothesis that there is no significant difference between the two methods of producing the hybridisation target. The results also show that hybridisation target synthesised from total RNA can be reused with as much confidence as that synthesised from mRNA.

There are two main benefits of synthesising the hybridisation target without mRNA isolation. Less starting material is required; with scheme 1, approximately 500 µg - 1mg of total RNA is required to ensure at least 1µg of mRNA, whilst with scheme 2, 5-40 µg of total RNA is sufficient. As an isolation step is missed out the new protocol is quicker to carry out. In all subsequent microarray experiments, total RNA was used as the template for cDNA synthesis. A cDNA synthesis kit was utilised for this step in the procedure (SuperScript Choice System – GibcoBRL #18090), using 20 µg of total RNA as the template.

#### **2.2.5 Modification of total RNA isolation method**

This section details further modifications made to the protocol, after it was decided to synthesise cDNA directly from total RNA (section 2.2.4). All subsequent discussion relates to data generated using the modified method described in this section.

##### **2.2.5.1 Materials**

Acid phenol (Sigma #P-4682)

RNeasy mini kit (Qiagen #74104)

Buffer RLT (10 µl β-mercaptoethanol per 1 mL buffer was added prior to use, as per manufacturer's instructions)

Buffer RPE (diluted with 4 volumes absolute ethanol as per manufacturer's instructions)

##### **2.2.5.2 Total RNA isolation**

Total RNA was isolated with the TRIzol reagent as described in section 2.2.2.2, but the extraction with the phenol/chloroform/isoamyl alcohol reagent was replaced with a phenol chloroform extraction using acid phenol. After the initial TRIzol / chloroform spin, the upper phase was transferred to a fresh tube and acid phenol added (1 mL per 1 mL TRIzol). This was incubated at room temperature for 5 min and chloroform added (0.2 mL per 1 mL acid phenol). This mixture was shaken hard for 15 s and then centrifuged at 4000 rpm for 15 min. This separated out into layers and as with the TRIzol extraction, the total RNA remained in the upper aqueous layer,

which was carefully transferred to a clean tube. If the aqueous layer looked cloudy, a second acid phenol extraction was performed, before organic extraction was carried as detailed in 2.2.2.3 above.

#### ***2.2.5.3 RNeasy cleanup of total RNA***

After quantification of the total RNA (as described in 2.2.2.6), it was further purified using the Qiagen RNeasy kit. 100 µg of total RNA (the limit of the RNeasy column) was purified from each tissue sample. Any remaining total RNA was kept at -80°C for long term storage. The RNeasy columns work by binding RNA, whilst impurities are washed away, then releasing RNA when treated with DEPC-treated H<sub>2</sub>O.

The method followed was as described by the RNeasy manufacturer with a few minor modifications. The RNA sample was adjusted to 100 µl with DEPC-treated H<sub>2</sub>O and 350 µl buffer RLT added and mixed. 250 µl absolute ethanol was added, mixed gently and the whole sample (700 µl) applied to a RNeasy mini spin column. This was centrifuged for 15 s, resulting in RNA binding to the membrane in the column. The flow-through was reapplied to the column and spun again for 15 s, to ensure binding of the complete RNA sample to the membrane. The flow-through from the second application was kept on ice and only discarded after RNA had been obtained from the procedure.

After transferring the column to a fresh 2 mL collection tube, 500 µl buffer RPE was added and the column centrifuged for 15 s. The flow-through was discarded and another 500 µl buffer RPE applied to the column. The column was centrifuged for 2 min, the flow-through discarded and the column spun for 30 s to ensure removal of any trace of buffer RPE from the RNA.

The column was transferred to a fresh 1.5 mL collection tube, taking care to avoid any ethanol traces on the side of the old tube. 100 µl DEPC-treated H<sub>2</sub>O was applied to the column, which was then incubated at room temperature for 1 min before being centrifuged at full speed for 1 min. The 'soaking' time allows the RNA to dissolve in the H<sub>2</sub>O and makes the procedure more efficient. The elution step was repeated a further two times into the same collection tube to ensure elution of as much RNA as possible.

#### ***2.2.5.4 Organic extraction of total RNA***

150 µl sodium acetate, 900 µl absolute ethanol and 0.5 µl pellet paint was added to the 300 µl of total RNA eluted from the RNeasy spin column. The samples were mixed thoroughly and incubated in dry ice until the solution became viscous, then microcentrifuged at full speed for 30 min. The resulting RNA pellet was washed twice in 0.5 mL 80% ethanol and air-dried before re-suspending in 20 µl DEPC-treated H<sub>2</sub>O. Quality and quantity were measured as described in 2.2.2.6 above. 20 µg of total RNA extracted in this way was used as the template for cDNA synthesis as described in section 2.2.4.3 above.

### 2.2.6 *In vitro* transcription (IVT)

The aim of this step was to create biotin labelled complementary RNA (cRNA), using double stranded cDNA as a template. The pool of synthesised cRNA is therefore representative of the pool of RNA transcript in the initial tissue sample.

#### 2.2.6.1 *Materials*

1.5µl double stranded cDNA

DEPC-treated H<sub>2</sub>O

BioArray High Yield RNA transcript labelling kit (Enzo #900182)

10X HY reaction buffer

10X Biotin labelled ribonucleotides

10X DTT

20X T7 RNA polymerase

10X RNase Inhibitor mix

#### 2.2.6.2 *Method*

20.5 µl DEPC-treated H<sub>2</sub>O was added to each cDNA sample (1.5 µl). 4 µl HY reaction buffer, 4 µl labelled ribonucleotides, 4 µl DTT, 4 µl RNase inhibitor mix and 2 µl RNA polymerase was then added to each reaction in the order stated. The samples were thoroughly mixed, centrifuged briefly to collect the reagents at the bottom of the tube and then incubated at 37°C for 5 h. During the incubation period the reactions were gently mixed and centrifuged every 45-60 min to encourage a high yield of IVT product. At the end of the incubation period the cRNA was either purified immediately (section 2.2.7) or stored at -20°C for later purification.

### 2.2.7 cRNA Purification

Affymetrix recommend purifying cRNA using the RNeasy column method used to purify total RNA (section 2.2.5.3). The IVT reaction appeared to be working extremely well, with high yields being measured prior to purification. However, after purification with the RNeasy column, too little of the cRNA was being recovered. As a trial, one target was generated from unpurified cRNA and hybridised to a test array. However, the quality of the target was very poor, making it clear that purification of the cRNA was necessary to remove any free nucleotides that have not been incorporated in the IVT reaction.

Using the Affymetrix-recommended columns was resulting in the loss of around 95% of the cRNA. To resolve this, the RNeasy column was tested with a slight modification to the protocol\*, alongside other commercial purification columns. None of the columns were manufactured to purify cRNA specifically; rather they were produced to purify either RNA or DNA.

## 2.2.7.1 Materials

Table 2.9 - Description of purification columns tested

<i>Manufacturer</i>	<i>Column name</i>	<i>Filtration method</i>
Amersham Pharmacia Biotech	Microspin G25	Holds back small molecules
Amersham Pharmacia Biotech	Autospin G50	Holds back small molecules
Cambridge Biosciences	Gel filtration	Holds back small molecules
Qiagen	Dye Ex Spin kit	Holds back small molecules
Qiagen	RNeasy mini kit	Binds RNA allowing impurities to flow-through, prior to elution

## 2.2.7.2 Method

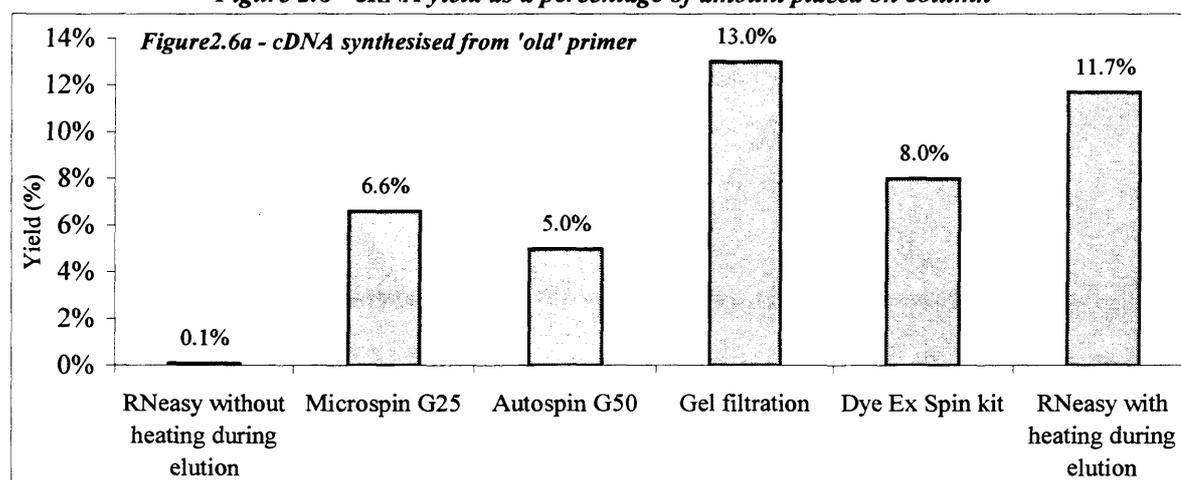
An IVT reaction was done specifically for this experiment and the amount of cRNA measured spectrophotometrically (section 2.2.2.6). Aliquots of this cRNA were purified through each column as per the manufacturer's instructions and re-quantified.

As the RNeasy columns had proved highly effective when purifying total RNA it was decided to change the primer used for the cDNA synthesis as a separate experiment and see whether that made any difference at the IVT purification stage. Two sets of the same oligo dT primer were ordered and tested using the RNeasy columns to purify the total IVT reaction.

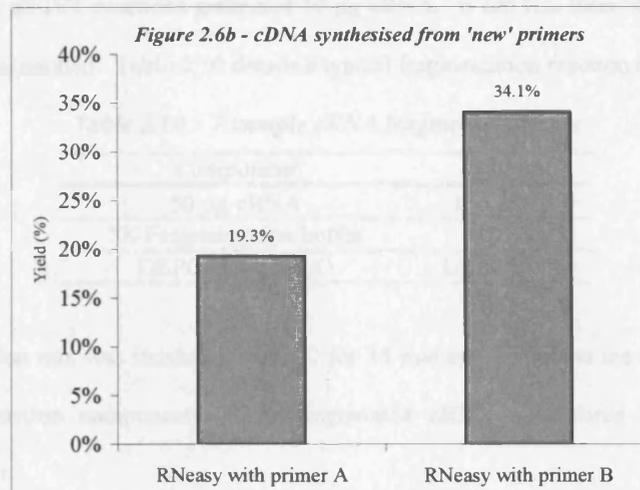
## 2.2.7.3 Results &amp; Discussion

Cambridge Biosciences' gel filtration column gave the best yield of all the columns tested with the old primer (figure 2.6a). Having changed the primer, the yields were much improved (figure 2.6b). It was therefore decided to use primer B and heat the RNeasy columns during elution to obtain the best yield of purified cRNA. The purification of total RNA using the RNeasy columns was yielding 80-95% of the amount placed on the column initially. The yields obtained with primer B were not considered ideal, but as the IVT reaction was generating large amounts of cRNA, the yield of purified cRNA represented a workable amount, typically 30-60 µg.

Figure 2.6 - cRNA yield as a percentage of amount placed on column



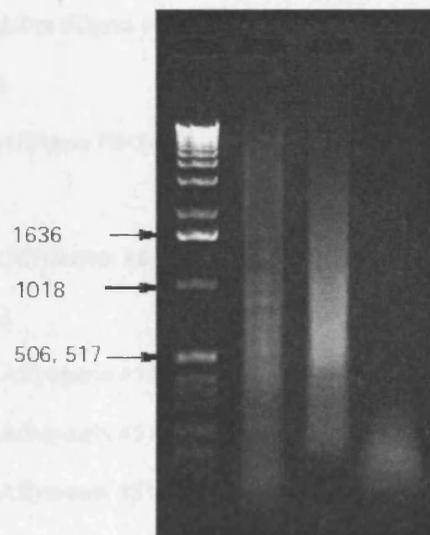
\* Method as in section 2.2.5.3, but column heated to 55°C during 1 min 'soaking' periods to ensure dissolution of cRNA in water.



### 2.2.8 cRNA Fragmentation

For better hybridisation of the labelled cRNA target to the probes on the arrays, the cRNA was fragmented using an Affymetrix-designed buffer. This fragmentation buffer acts on the labelled cRNA in a random manner, producing fragments of 35 to 200 bases long (figure 2.7). The shorter targets are less affected than longer ones by steric hindrance when binding to the 25-mer oligonucleotide probes. However, it is important to note that fragmenting the target does not affect the sequence specific binding of the target to the probe.

**Figure 2.7 – Gel showing ds cDNA (lane 2), purified cRNA (lane 3) and fragmented cRNA (lane 4)**



#### 2.2.8.1 Reagents

5X Fragmentation Buffer (appendix A2.2)

#### 2.2.8.2 Method

The cRNA was adjusted to a minimum concentration of 0.6  $\mu\text{g}/\mu\text{l}$  with DEPC-treated  $\text{H}_2\text{O}$ . The amount of fragmentation buffer used varied depending on the amount of cRNA. The final cRNA concentration in the fragmentation reaction could range between 0.5 – 2  $\mu\text{g}/\mu\text{l}$ . Most of the fragmentation reactions performed in this study had a cRNA concentration of approximately 1  $\mu\text{g}/\mu\text{l}$ . Affymetrix recommend fragmenting a minimum of 12  $\mu\text{g}$  for each array to which the target will be hybridised. 50  $\mu\text{g}$  of cRNA was fragmented for as many targets

as possible, although not all IVT reactions generated 50 µg cRNA. When less than 50 µg cRNA was generated the whole amount was fragmented. Table 2.10 details a typical fragmentation reaction mix.

**Table 2.10 - Example cRNA fragmentation mix**

<i>Component</i>	<i>Volume</i>
50 µg cRNA	1 to 40 µl
5X Fragmentation buffer	10 µl
DEPC-treated H <sub>2</sub> O	Up to 50 µl

The fragmentation reaction mix was incubated at 94°C for 35 min and placed on ice immediately afterwards to aid condensation of reaction components. The fragmented cRNA was stored at -20°C until used for hybridisation.

## 2.2.9 Hybridisation to Microarrays

### 2.2.9.1 Materials

Acetylated Bovine Serum Albumin (BSA) solution (50 mg/ml) (GibcoBRL #15561-020)

Herring sperm DNA (10 mg/ml) (Promega #D-1811)

GeneChip Eukaryotic Hybridisation control kit (Affymetrix #900299)

Control Oligo B2 3 nM (Affymetrix #900301)

5 M NaCl RNase-free DNase-free (Sigma #S-5150)

MES free acid monohydrate SigmaUltra (Sigma # M-5287)

MES sodium salt (Sigma #m-3885)

EDTA disodium salt 0.5 M solution (Sigma #E-7889)

10% Tween-20 (Pierce #28320)

GeneChip hybridisation oven 640 (Affymetrix #800139)

Test 1 array\* (Affymetrix #510006)

Hu6800 set\*                      Sub A (Affymetrix #510127)

   Sub B (Affymetrix #510129)

   Sub C (Affymetrix #510131)

   Sub D (Affymetrix #510133)

HuFL microarray\*\* (Affymetrix #900183)

Hu35K microarray set\*\*      Sub A (Affymetrix #900184)

   Sub B (Affymetrix #900185)

   Sub C (Affymetrix #900186)

   Sub D (Affymetrix #900187)

\* Feature size 50 µm

### 2.2.9.2 Reagents

12X MES Stock (appendix A2.3)

2X Hybridisation Buffer (appendix A2.4)

### 2.2.9.3 Preparation of hybridisation cocktail

The volumes of the reagents in the hybridisation mix were based on the amount of cRNA fragmented in section 2.2.8. Table 2.11 shows the reagent volumes for 50 µg fragmented cRNA. These volumes were adjusted based on the amount of cRNA fragmented, in order to retain the same final concentrations. The hybridisation cocktail was used immediately or stored at -20°C, until required.

**Table 2.11 - Hybridisation cocktail for 50 µg fragmented cRNA**

<i>Reagent</i>	<i>Amount or Volume</i>	<i>Final concentration</i>
Fragmented cRNA	50 µg	0.05 µg/µl
Control Oligonucleotide B2	17 µl	50 pM
Eukaryotic Hybridisation controls ( <i>BioB</i> , <i>BioC</i> , <i>BioD</i> , <i>cre</i> )	50 µl	1.5, 5, 25, 100 pM respectively
Herring sperm DNA	10 µl	0.1 mg/ml
Acetylated BSA	10 µl	0.5 mg/ml
2X Hybridisation buffer	500 µl	1X
H <sub>2</sub> O	to a final volume of 1000 µl	

### 2.2.9.4 Hybridisation to probe arrays

The probe arrays were equilibrated to room temperature immediately before use. If the hybridisation cocktail had been frozen it was first heated to 65°C for 5 min to completely re-suspend the cRNA. The hybridisation cocktail was then heated to 99°C for 5 min. The probe array was filled completely with 1X hybridisation buffer (approx. 250 µl) and rotated for 10 min at 45°C. The hybridisation cocktail was transferred from 99°C to 45°C and incubated for 5 min. The hybridisation cocktail was then centrifuged at full speed for 5 min to remove any insoluble materials. The buffer solution was removed from the array and 200 µl of the clarified hybridisation cocktail immediately applied, avoiding any pelleted insoluble material at the bottom of the tube. The arrays with the hybridisation cocktail applied, were rotated at 45°C for 16 h.

### 2.2.9.5 Reuse of hybridisation cocktail

The hybridisation cocktail from each sample was applied to each of the arrays in the same order, i.e. Test 1, HuFL, Hu35K\_subA, Hu35K\_subB, Hu35K\_subC and Hu35K\_subD. The products of the *in vitro* transcription reactions were insufficient to generate 1000 µl hybridisation cocktail in some cases. This meant that the corresponding hybridisation cocktail had to be reused on subsequent arrays. Affymetrix recommend that a hybridisation cocktail be used a maximum of five times, but in two cases where the hybridisation cocktail was

insufficient, hybridisation to Hu35K\_subD array represented the sixth hybridisation. Therefore, all the Hu35K\_subD array results have been excluded from the remainder of this thesis.

### 2.2.10 Staining the Microarrays

At the end of the 16 h incubation the hybridisation cocktail was removed from the array and stored at  $-20^{\circ}\text{C}$ . The arrays were filled with completely with non-stringent wash buffer and stored at  $4^{\circ}\text{C}$  for a maximum of 1 h, before proceeding with the washing and staining procedure.

#### 2.2.10.1 Materials

R-Phycoerythrin Streptavidin (Molecular Biology Probes #S-866)

PBS pH 7.2 (GibcoBRL #20012-027)

20X SSPE (GibcoBRL #15590-035)

10 mg/ml Goat IgG Stock (Sigma #I-5256)

Resuspended 50 mg in 5 mL PBS and stored at  $4^{\circ}\text{C}$ .

Biotinylated anti-streptavidin, goat antibody (Vector Laboratories #BA-0500)

Antifoam 0-30 (appendix A2.5)

GeneChip fluidics station 400 (Affymetrix #800101)

GeneArray scanner (Affymetrix #900156)

#### 2.2.10.2 Reagents

Wash buffer A - Non-Stringent Wash Buffer (appendix A2.6)

Wash buffer B - Stringent Wash Buffer (appendix A2.7)

2X Stain Buffer (appendix A2.8)

#### 2.2.10.3 Standard wash method

The arrays for the method evaluation experiments described in section 2.2.4 were stained using this wash method. For each array, 600  $\mu\text{l}$  SAPE stain solution (appendix A2.9) was prepared and aliquoted into a 1.5 mL Eppendorf tube.

**Table 2.12 - Fluidics program (EukGE-WS1) for standard wash method**

<i>Stage</i>	<i>EukGE-WS1</i>
Post hybridisation wash 1	10 cycles of 2 mixes per cycle with wash buffer A at $25^{\circ}\text{C}$
Post hybridisation wash 2	4 cycles of 15 mixes per cycle with wash buffer B at $50^{\circ}\text{C}$
Probe array staining	30 min at $25^{\circ}\text{C}$
Final wash	10 cycles of 4 mixes per cycle with wash buffer A at $25^{\circ}\text{C}$
Holding temperature	$25^{\circ}\text{C}$

#### 2.2.10.4 Antibody amplification wash method

All the data discussed in this thesis was generated using the antibody wash method to stain the arrays. 1.2 mL SAPE solution (appendix A2.9) and 600 µl antibody solution (appendix A2.10) was prepared for each array.

**Table 2.13 - Fluidics program (EukGE-WS2) for antibody wash method**

<i>Stage</i>	<i>EukGE-WS2</i>
Post hybridisation wash 1	10 cycles of 2 mixes per cycle with wash buffer A at 25°C
Post hybridisation wash 2	4 cycles of 15 mixes per cycle with wash buffer B at 50°C
Probe array staining with 1 <sup>st</sup> aliquot of SAPE solution	10 min at 25°C
Post Stain wash	10 cycles of 4 mixes per cycle with wash buffer A at 25°C
<i>Replace used SAPE solution with fresh antibody solution</i>	
Probe array staining with antibody solution	10 min at 25°C
<i>Replace used antibody solution with fresh SAPE solution</i>	
Probe array staining with 2 <sup>nd</sup> aliquot of SAPE solution	10 min at 25°C
Final wash	15 cycles of 4 mixes per cycle with wash buffer A at 30°C
Holding temperature	25°C

After washing was complete, the arrays were kept in the dark until ready for scanning. If the length of time between completion of washing and scanning was greater than 30 min, the arrays were stored at 4°C for a maximum of 1 h.

## 2.2.11 Generation & Analysis of the Microarray Image

### 2.2.11.1 Scanning the arrays

If the stained array had been stored at 4°C, it was left to equilibrate to room temperature before proceeding. The glass surface of each array was carefully wiped to remove any particles that may interfere with the image obtained. Each array was scanned twice, at a wavelength of 570 nm. An example of the resulting image is shown in figure 2.8.

The digital image must be analysed to obtain biologically meaningful data. The array is divided into 64,000 squares called features, each containing  $10^7$ -  $10^8$  copies of an identical 25-mer oligonucleotide probe sequence (figure 2.9). For each oligonucleotide probe matching the target mRNA exactly (perfect match; PM), there is a corresponding mismatch (MM) probe sequence that differs from the PM probe by a single base at the 13<sup>th</sup> position. This mismatch is enough to seriously inhibit the binding of the target RNA to the MM and therefore acts as a control for non-sequence specific binding. A PM feature with its corresponding MM feature is referred

to as a 'probe pair', with each gene on the array is represented by 16 to 20 different probe pairs, i.e. a 'probe set' (figure 2.9).

Figure 2.8 – A typical microarray image

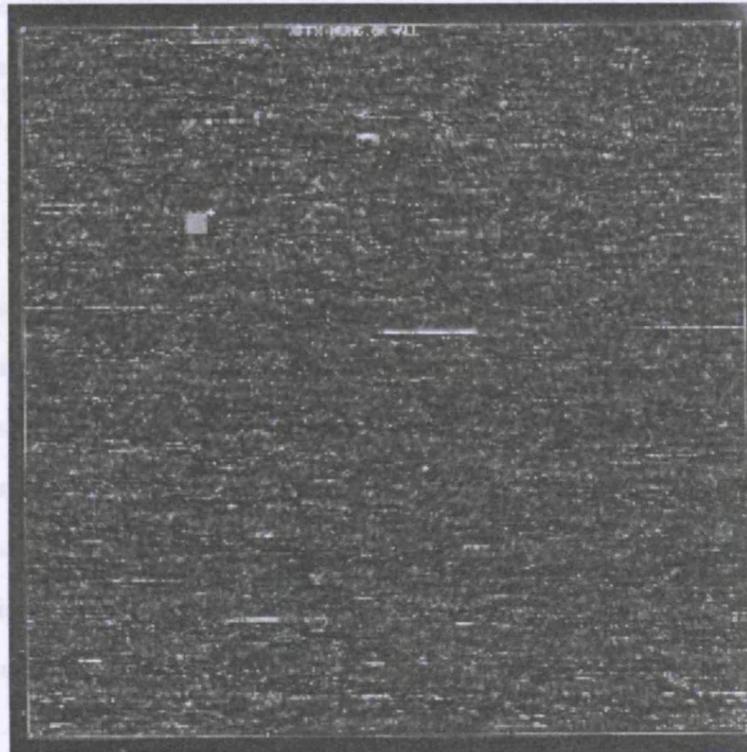
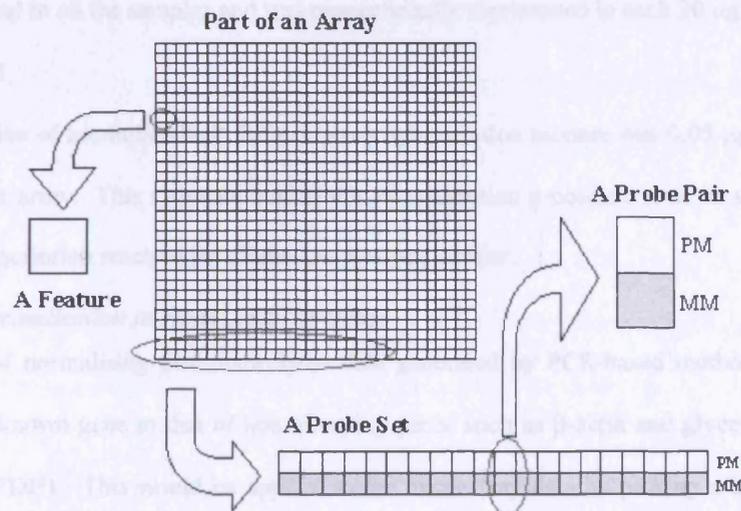


Figure 2.9 – Arrangement of probes on microarray



2.2.11.2 Image Analysis by Affymetrix GeneChip Software

Before any analysis of the data could be carried out, the raw pixel intensity values had to be converted into numerical values, in order to obtain information about the relative quantity of transcripts in the original samples. There are a number of parameters calculated by the software based on the measured expression of each transcript

on each array, independent of any other array. The calculation of these 'absolute parameters' is described in appendix A2.11.

### ***2.2.11.3 The Average Difference and the Absolute Call***

Of the absolute parameters calculated, the 'average difference' (avg diff) and the 'absolute call' (abs call) are the most important. The absolute call indicates whether a transcript was detected in a sample. If the transcript was detected, the average difference value then gives an indication of the relative level of expression of the transcript. Therefore these two parameters represent the 'raw data' upon which all subsequent data mining procedures were performed.

## **2.2.12 Normalisation**

115 microarrays were used to generate the data discussed in this thesis. Comparing the arrays to each other generated the expression difference data that was of interest. However, before the arrays could be compared, they had to be normalised. Each possible normalisation method is based on one or more assumptions. Whether a particular normalisation method was utilised or not depended how applicable the corresponding assumptions were to the target preparation method. These are discussed below.

### ***2.2.12.1 Normalisation during target preparation***

#### ***Total RNA***

The amount of total RNA used to synthesise cDNA was independent of the total yield from a sample, as a maximum of 20 µg total RNA was used in all cases (section 2.2.4.3). The method assumes that the total RNA to mRNA ratio was equal in all the samples and was proportionally represented in each 20 µg total RNA used.

#### ***Biotin labelled cRNA***

The final concentration of biotin labelled cRNA in each hybridisation mixture was 0.05 µg/ml and 200 µl of this mixture was used per array. This rationale behind this normalisation procedure rests on the assumption that the efficiency of the transcription reaction for all the samples was similar.

### ***2.2.12.2 Normalisation to Housekeeping Genes***

A popular method of normalising gene expression data generated by PCR-based methods is to normalise the expression of the unknown gene to that of housekeeping genes such as  $\beta$ -actin and glyceraldehyde-3-phosphate dehydrogenase (GAPDH). This would be applied to the microarray data by picking a housekeeping gene and dividing the average difference for each probe set by the average difference of the housekeeping gene probe set on the same array. This relies on the assumption that the regulation of these housekeeping genes does not vary significantly between the different tissues and samples. However, the 'housekeeping' gene may be involved in more than a simple housekeeping function<sup>193</sup>, or be regulated to a larger extent than previously supposed<sup>194-196</sup>. It was therefore decided not to apply this normalisation strategy.

### 2.2.12.3 Global Normalisation

This is another popular normalisation method. The expression measurement of each transcript (in the case of Affymetrix arrays, this is the average difference value) is divided by the mean of this measurement across all the transcripts on each array. Thus if the mean average difference value of an array is 300 and the average difference of a gene is 600, the normalised value of that transcript is 2. The assumption with this model is that the total number of transcript molecules per cell is constant.

However, the GeneChip software carries out its own form of global normalisation called scaling, which is based on bacterial RNA (the eukaryotic hybridisation control) added to the target cRNA as part of the hybridisation mix (table 2.11). The output (average difference) of each gene is multiplied by a factor (the scaling factor) to make its average intensity equal to an arbitrary target intensity defined manually and based on the bacterial RNA. For all the microarray experiments discussed in this thesis the target intensity was set to 250, (with the exception of the preliminary experiments detailed in section 2.2.4). All the microarray studies taking place in the lab where this work was carried out were scaled to a target intensity of 250 to facilitate comparisons between different projects.

### 2.2.13 Verification of the microarray data

In order to verify the microarray gene expression data, quantitative PCR (Q-PCR) was used to measure the gene expression of three of the genes identified (discussed in chapter 4). The TaqMan Q-PCR system was used. The TaqMan system has been described previously<sup>197, 198</sup>, and is a well established method of Q-PCR.

The total RNA that was isolated for the microarray experiment was used for the Q-PCR method. Probes and primers against the *DTDST*, *DRA* and *CD9* genes were designed using the Primer Express 1.0 program, and BLASTed against public databases to ensure specificity. The Q-PCR data were normalized to the 18S RNA for each sample. In order to allow a comparison between the Q-PCR and microarray data, the NI\_1 sample was arbitrarily selected to represent the base value. All other samples were then plotted as a fraction of the NI\_1 microarray and Q-PCR data. Figure 2.10 shows the results of these comparisons across all three genes. The Q-PCR and microarray data generally show a good correlation, as would be expected.

Figure 2.10 – Comparison of microarray and Q-PCR gene expression data

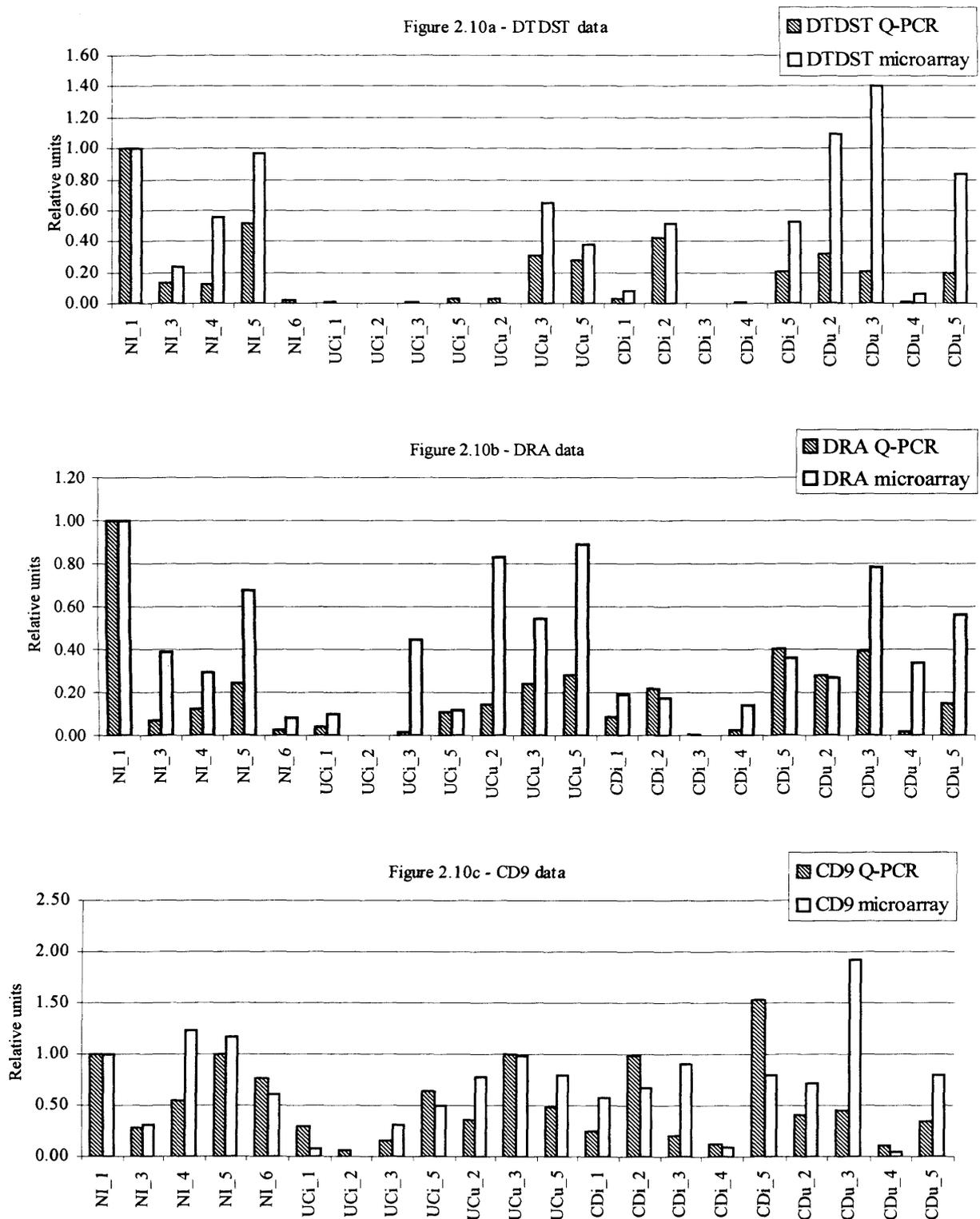


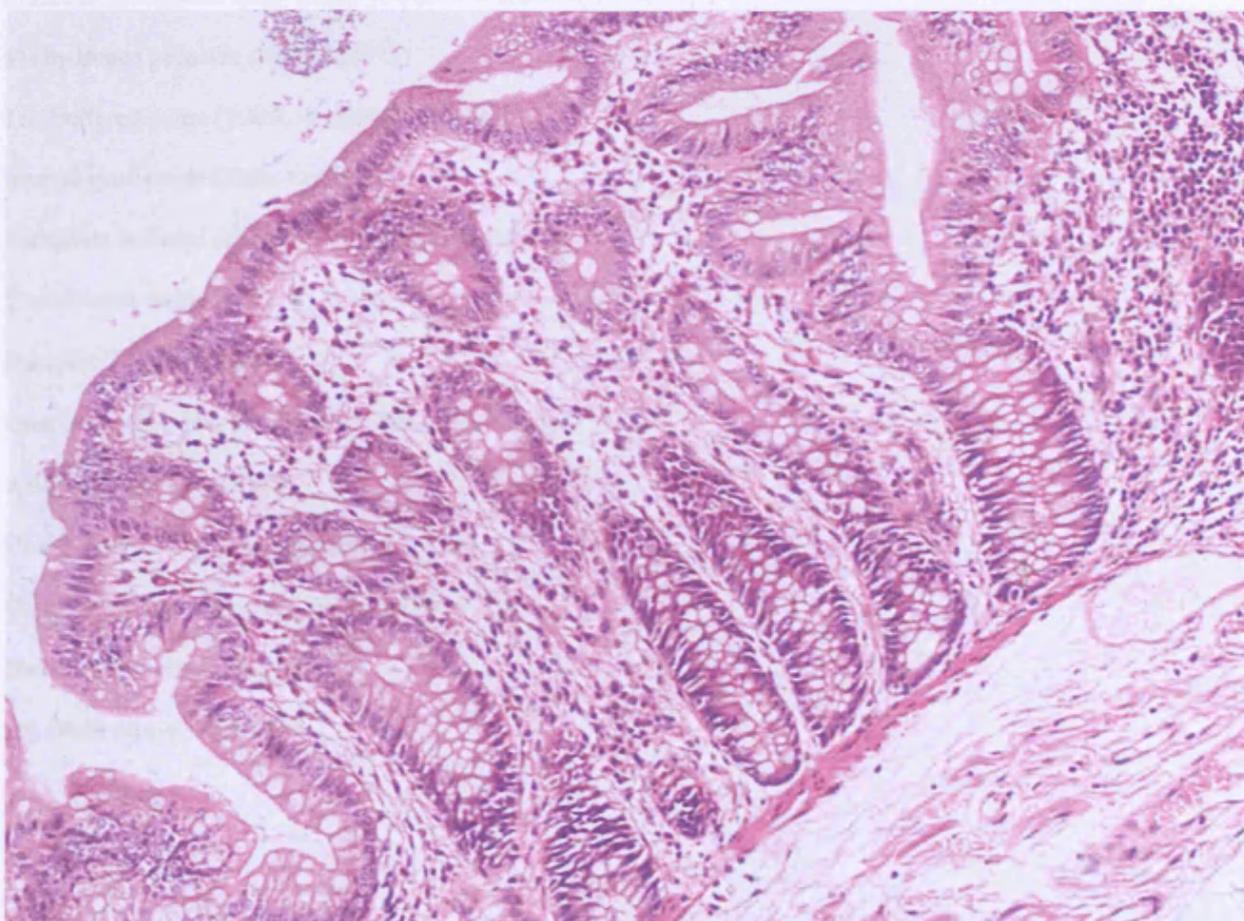
Figure 2.10 shows the transcript expression of DTDST (figure 2.10a), DRA (figure 2.10b) and CD9 (figure 2.10c) as measured by Q-PCR and microarray analyses and relative to the NI\_1 sample.

## 2.3 Immunocytochemistry

### 2.3.1 Method Overview

The mucosal samples collected for microarray analysis contain many different cell types (figure 2.11). If one cell type (e.g. neutrophils) is increased tenfold in IBD, then any gene which is exclusively expressed by neutrophils would appear to show a tenfold increase in expression even if its true level of expression was completely unchanged. The relative proportions of the different cell types in the samples were determined to aid in the analysis of the microarray data. Changes in the genes expressed by the epithelial cells are believed to lead to IBD; therefore, it was especially important to determine the proportion of epithelial cells in each of the samples. Immunocytochemistry (ICC) was used to identify epithelial cells, T lymphocytes, B lymphocytes, macrophages, mast cells and neutrophils. The relative volume fraction of mucosa consisting of a particular cell type was quantified and used to normalise the microarray data where possible.

*Figure 2.11 – H & E section of sample NI\_5 mucosa*



### 2.3.2 Tissue Section Origins

The formalin-fixed tissue blocks were embedded in paraffin wax by the histopathology department at Leicester General Hospital. Sections were taken from microscopically normal archival blocks for the control patients. The 'microscopically normal' archival blocks were part of a series of blocks taken from each resection specimen as standard diagnostic practice. Sections were cut by the author to 4  $\mu$ m on a microtome (Leica #RM2035)

situated in the Leicester General Hospital histology department. The sections were placed onto silicate-coated slides and left to dry in a 60°C oven.

### **2.3.3 Antibody Staining\***

The ICC method used was appropriate for primary antibodies raised in rabbits or mice. The secondary antibody was a goat  $\alpha$  mouse/rabbit immunoglobulin. The third antiserum included a biotinylated horseradish peroxidase and streptavidin, which reacts with the diaminobenzidine (DAB) reagent. This results in specific localisation of the DAB product to the cell targeted by the primary antibody. The insoluble product was visible on the section as a brown stain.

#### **2.3.3.1 Materials**

Xylene (Genta Medical #1330-20-7)

Industrial methylated spirit† (IMS) (Genta Medical #64-17-5)

6% hydrogen peroxide (BDH #28517)

Tris buffered saline (T.B.S.) (appendix A2.12)

Normal goat serum (Dako #X0907)

Phosphate buffered saline (P.B.S.) (appendix A2.13)

1° antibodies (table 2.14)

Duet Kit (Dako #K0492)

Goat  $\alpha$  mouse/rabbit immunoglobulins - 2° antibody

ABC complex - 3° antibody

Diaminobenzidine (DAB) (appendix A2.14)

Copper sulphate solution (BDH #36204)

Haematoxylin (appendix A2.15)

1% Eosin (appendix A2.16)

---

\* All antibody staining was carried out in the special histology department at Leicester Royal Infirmary  
† Diluted to 99%, 85% and 75% with distilled H<sub>2</sub>O

Table 2.14 – Primary Antibodies

<i>Target Cell</i>	<i>Antibody</i>	<i>Manufacturer (Product N<sup>o</sup>.)</i>	<i>Dilution (Pre-treatment)*</i>	<i>Positive Control</i>
B lymphocytes	Monoclonal mouse anti-human CD20 IgG2a, κ	Dako (#M-0755)	1:250 (M)	Tonsil
T lymphocytes	Affinity purified rabbit anti-human CD3	Dako (#A-0452)	1:300 (T)	Tonsil
Macrophages	Monoclonal mouse anti-human CD68 IgG3, κ	Dako (#M-0876)	1:50 (T)	Tonsil
Glandular epithelial cells	Monoclonal mouse anti-human CAM5.2 IgG2a	Becton Dickinson (# 349205)	1:10 (T)	Colon
Neutrophils	Monoclonal mouse anti-human neutrophil elastase IgG1, κ	Dako (#M-0752)	1:200 (N)	Tonsil
Mast cells	Monoclonal mouse anti-human mast cell tryptase IgG1, κ	Dako (#M-7052)	1:300 (T)	Tonsil

### 2.3.3.2 Pre-treatment methods

#### *Microwave treatment*

Slides are microwaved in citrate buffer (appendix A2.17) for 12 min at full power.

#### *Trypsin treatment*

0.30 g trypsin (Ditco #215230) and 0.36 g granular calcium chloride (BDH #27587) were dissolved in 300 mL distilled water at 37°C and pH adjusted to 7.8 with weak sodium hydroxide solution (BDH #19147). The slides were incubated in the trypsin solution at 37°C for 10 min.

#### *No pre-treatment*

Slides were kept in running tap water until proceeding to next stage.

### 2.3.3.3 Method

The tissue sections were dried for 2 min in a 40°C oven, before being rehydrated. The sections were rehydrated by taking them through 2 washes of xylene and 4 washes of decreasing IMS concentration. Endogenous peroxidases were blocked by incubating the sections in 6% hydrogen peroxide for 10 min, before washing in running tap water for 5 min. The sections were pre-treated as required (table 2.14 & section 2.3.3.2) and washed in running tap water for 2 min, followed by T.B.S.† for 5 min.

The sections were incubated with normal goat serum\* for 10 min to block non-specific binding of the goat antibody. This was drained and the section wiped around, before incubation with the primary antibody for 1 h (50-100 µl antisera was applied per section). The sections were then washed in P.B.S. for 20 min to remove any unbound primary antibody. The secondary antibody (goat α mouse/rabbit) was applied to the sections and

\* See section 2.3.3.2 - M = microwave; T = trypsin; N = no pre-treatment

† All washes were carried out in approximately 300 mL of buffer, in a dish on a magnetic stirrer.

incubated for 30 min. The slides were washed in P.B.S. for 20 min to remove unbound secondary antibody. ABC complex was incubated with the sections for 30 min, followed by two P.B.S. washes of 10 min each to remove any unbound ABC complex. The sections were treated with DAB solution for 5 min to stain the bound ABC complex. Excess DAB was removed by washing in running tap water and the DAB stain was enhanced by incubating the sections in copper sulphate solution for 5 min. Excess copper sulphate was removed by washing in running tap water for 5 min. The slides were counter-stained with haematoxylin for 30 s and washed in running tap water until the water ran clear. The slides were dehydrated by taking them back through the IMS and xylene washes in reverse order, i.e. through increasing concentrations of IMS and xylene. A standard cover-slipping machine was used to cover the slides.

#### **2.3.3.4 Counter-stain removal**

After the initial sections were stained, it was found that the staining quantification software (section 2.3.4) was more accurate when the counter-staining was weak. It was therefore decided to leave out the haematoxylin-staining step when staining further sections. For the sections already stained, the following method was used to remove the haematoxylin, without affecting the antibody staining.

The slides were soaked in acetone for 30 min to remove the coverslip, then rehydrated through xylene and decreasing IMS concentrations. The sections were then soaked in 1% acid alcohol for 2 h to remove the haematoxylin counterstain. The sections were dehydrated as before and fresh cover slips re-fixed by hand over the sections.

#### **2.3.4 Haematoxylin & Eosin (H&E) Staining†**

The CAM 5.2 antibody was initially used to identify epithelial cells, but was it found that this only stained the differentiated cells at the luminal edge of the crypts (figure 2.12). It was therefore decided to count epithelial cells from an H&E section manually, using a previously described graticule<sup>199</sup>.

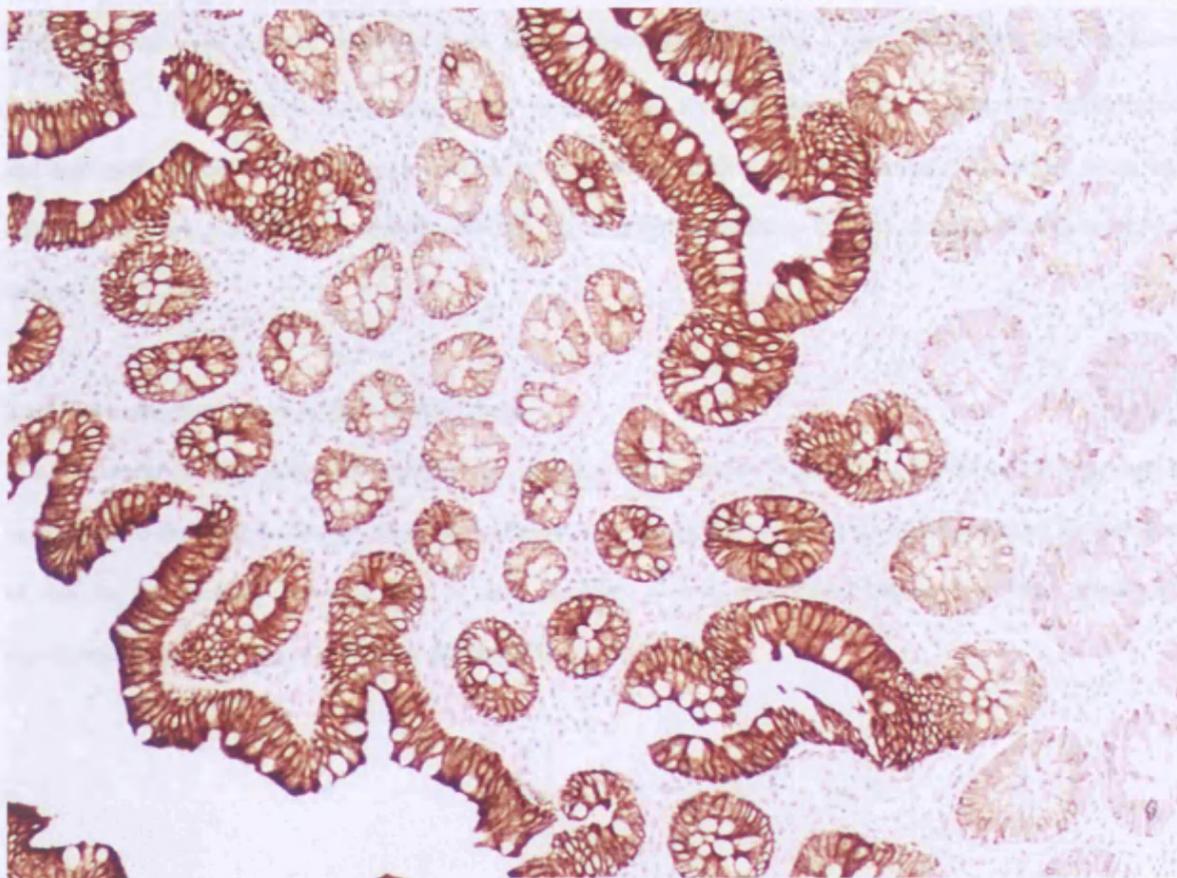
The H&E staining was carried out on a dedicated staining machine. This was a conveyor belt, which immersed the microscope slides into a number of solutions automatically. The slides were re-hydrated through xylene and decreasing concentrations of IMS before being stained in a series of haematoxylin pots. This was followed by a series of eosin stains, before being taken through water to remove the excess stain. The slides were finally dehydrated through increasing concentrations of IMS and two xylene washes. The slides were kept in xylene until manual cover slipping.

---

\* All antisera incubations were carried out at room temperature in moist slide chambers.

† All H&E staining was carried out in the histology laboratory at Leicester General Hospital.

**Figure 2.12– Staining of differentiated epithelial cells only with anti-CAM 5.2 antibody**



*Only differentiated epithelial cells were fully stained with the anti-CAM 5.2 antibody, as shown in this section from patient NI\_5. Therefore, the epithelial cells were counted manually from H & E sections.*

### **2.3.5 Volume fraction quantification**

The IBD sections were taken from the samples that were subsequently processed to form target cRNA and hybridised to the microarrays. Therefore, these sections consisted of mucosa without underlying muscle. The NI sections were taken from whole bowel wall blocks taken as part of routine histological analysis. For these sections, the author identified the mucosa and only the cells within it were quantified. The method of quantification depended on the staining method used.

#### **2.3.5.1 Antibody stained sections**

The proportion of mucosa stained was quantified on a Macintosh G3, using the public domain NIH Image program (developed at the US National Institutes of Health and available at <http://rsb.info.nih.gov/nih-image/>). The mucosal area was defined manually and the software calculated the proportion of defined area which was stained. On average, 17 fields were sampled for quantification in each section, unless there was a scarce amount of mucosa, in which case the entire mucosal area was sampled. The proportion of mucosa corresponding to each cell type was calculated by averaging the result across all sampled fields.

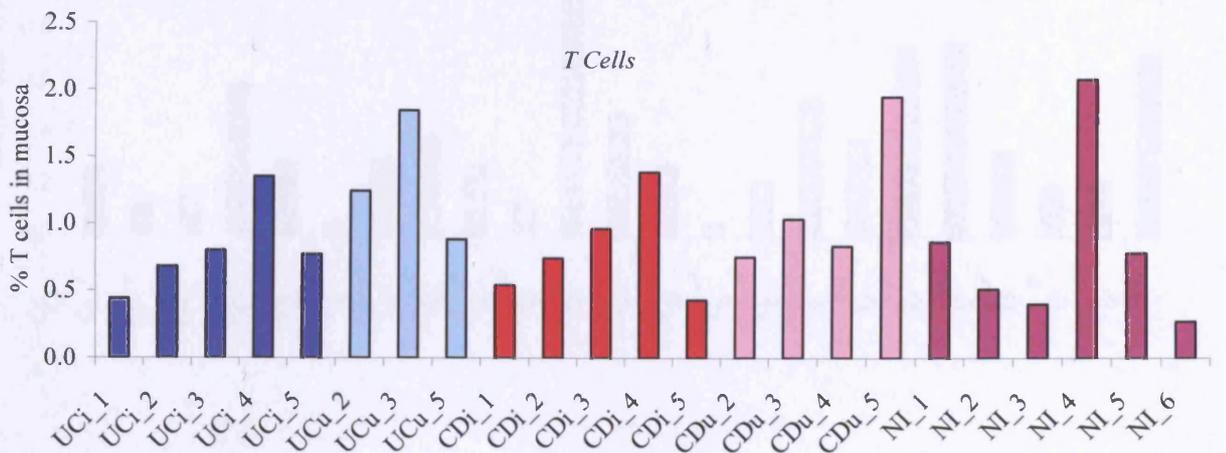
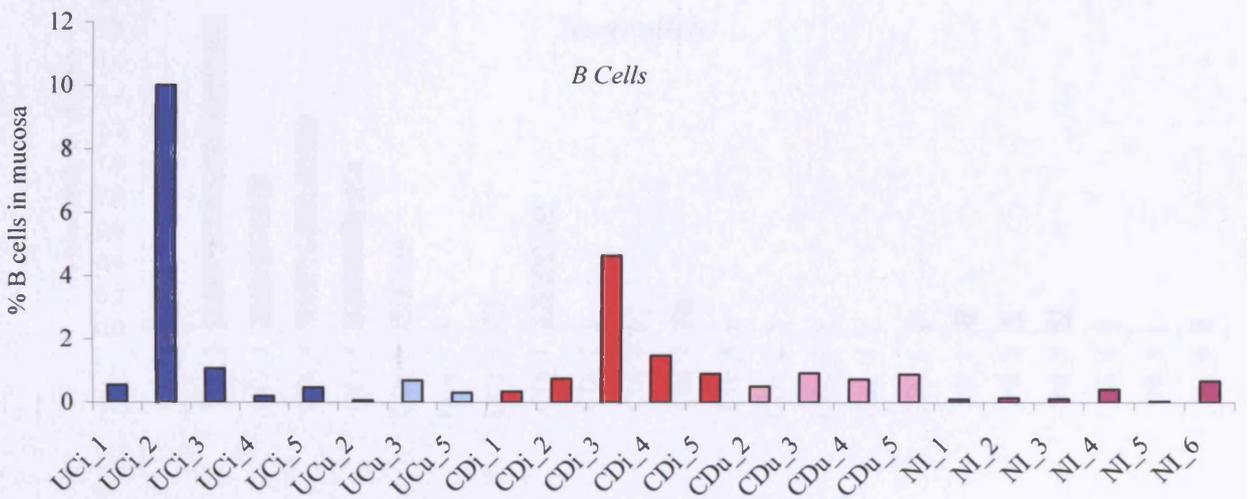
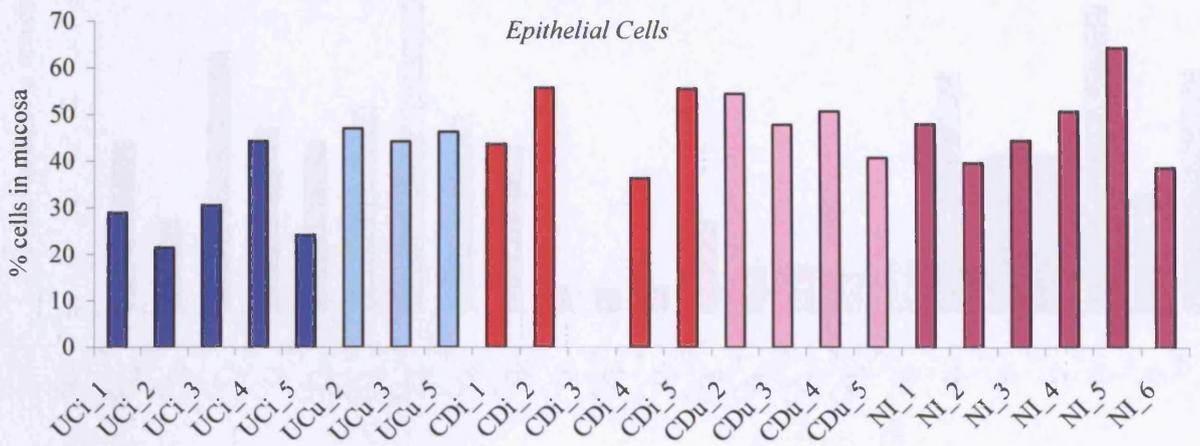
### **2.3.5.2 H&E stained sections**

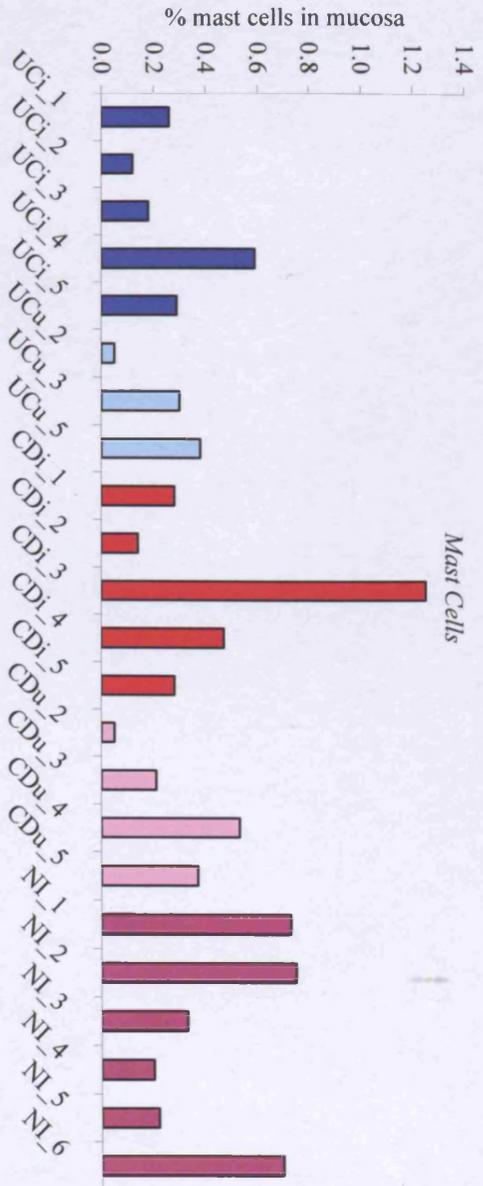
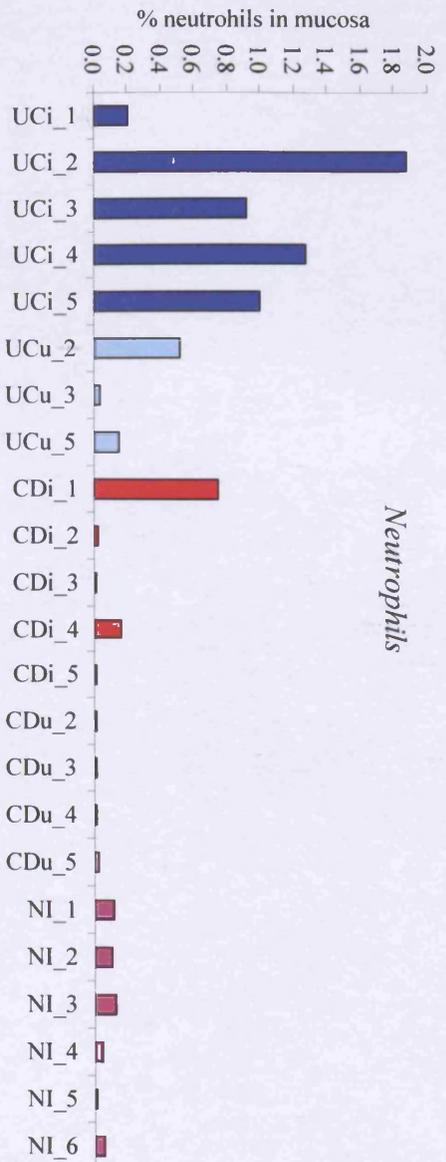
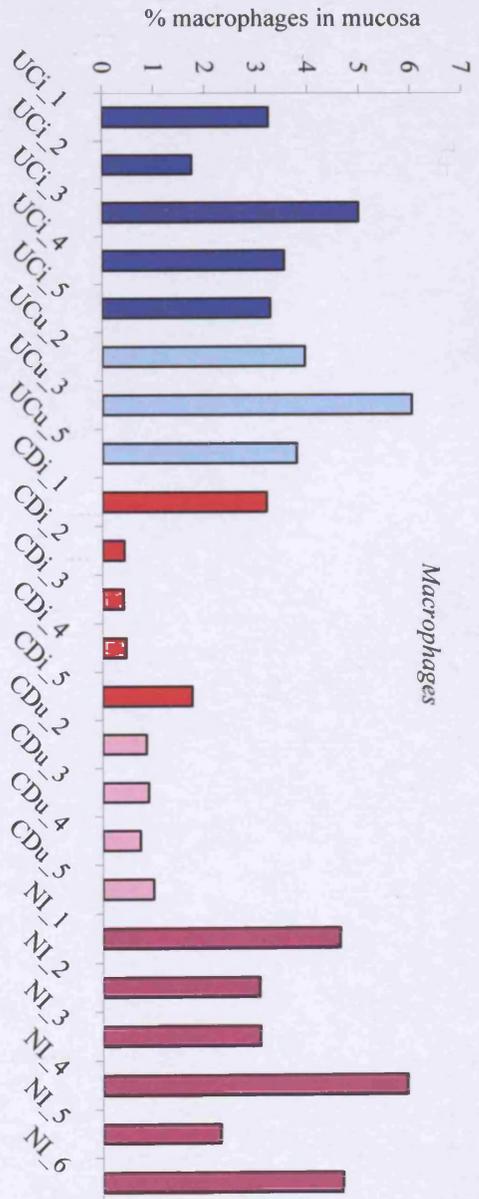
Epithelial cells were counted manually from H&E stained sections using a light microscope and the Lennox graticule<sup>199</sup>. 12 to 15 randomly selected fields were sampled for each section. The numbers of epithelial cells and non-epithelial cells in the mucosa were determined for each field. The percentage of mucosa taken up by epithelial cells was calculated by dividing the number of epithelial cells by the total number of cells counted and multiplying by 100.

### **2.3.6 Normalisation of array data to ICC data**

In cases where the expression of a gene could reliably be attributed to a single cell, the microarray data was normalised to the volume fraction data. The microarray average difference value was divided by the percentage of mucosa occupied by the cell type in each sample. These normalised values were then re-tested for significance with an unpaired T-test (as discussed in the next chapter).

2.3.7 ICC results





**CHAPTER THREE – DATA MINING**

**3.1 Chapter Introduction**

---

Due to the sheer volume of data generated, the analysis of microarray data represents a significant bottleneck in the experimental procedure. Obtaining biologically relevant information from data generated requires the development of specialist methods for handling and analysing (mining) the data. Any mining technique must be able to elucidate differential gene expression between arrays reliably and efficiently. A balance must be struck between generating false positives and missing potentially important, albeit quantitatively small, gene expression changes. The currently available mining techniques can be broadly categorised as those that are based on clustering and those that are not. Due to the novelty of the field, one of the directions that became apparent during the mining of the data was the need for an evaluation of different mining tools and protocols.

### 3.2 GeneChip software

#### 3.2.1 Introduction

The most obvious starting point for analysis of the microarray data was the software that has been specifically developed for use with the Affymetrix arrays, i.e. the 'Data Mining Tool' or DMT. The author had access to DMT version 1.2\*. DMT allows the gene expression data to be viewed in a number of different ways and enables some rudimentary comparative analyses to be performed.

#### 3.2.2 The Gene Expression Matrix

The gene expression matrix is a tabular form of viewing the data (figure 3.1). It is perhaps an obvious way to organise the data, but it is an important concept and is discussed further in section 3.3. It is mentioned here, as the expression matrix was also the starting point for data mining in DMT.

*Figure 3.1 – Basic gene expression matrix*

	Sample 1	Sample 2		Sample n
Gene 1				
Gene 2				
Gene n				

Average Difference values

Having selected the arrays to be analysed the expression matrix is created during the pivot operation. The pivot operation retrieves the previously saved microarray data from the database and organises it as shown. The average difference and absolute call parameters were selected for inclusion on the gene expression matrix, as these two parameters represent the 'raw data' upon which all data mining procedures are based, resulting in a slightly modified gene expression matrix (figure 3.2).

*Figure 3.2 - Gene expression matrix adapted for Affymetrix expression data*

	Sample 1		Sample 2				Sample n	
	Avg Diff	Abs Call	Avg Diff	Abs Call			Avg Diff	Abs Call
Gene 1								
Gene 2								
Gene n								

#### 3.2.3 Querying the Data with DMT

The gene expression matrix discussed above is the starting point for further analyses in DMT and is created by 'pivoting' the data using the average difference and absolute call parameters. The gene expression matrix can be limited to probe sets that behave in particular ways by 'querying' the data before the pivot operation. For

example, the query can be set to only return probe sets that have been called present. However, this facility was difficult to use and therefore it was abandoned early in the project in favour of other mining methods.

### 3.2.4 Viewing the Data with DMT

Although the query facility in DMT was not found to be useful, the graphical capabilities were. The software can display four different graphs and for each one a baseline array and an experimental array must be chosen. The baseline array is used as the comparative (comp) array, from which the differences in transcript expression of the experimental (exp) array can be calculated. To create the figures shown in this section the HuFL array data for NI\_1 (comp sample) and CDi\_1 (exp sample) were used. No query was applied to the data, so all probe sets on the HuFL array were returned.

#### 3.2.4.1 Scatter Graph

The scatter plot can be used to compare two different arrays, or two different groups of arrays. The genes shown on the plot can be selected (figure 3.3a) or deselected (figure 3.3b) depending on the change in absolute call between the two axes. In figure 3.3 the NI sample is assigned to the x-axis and the CDi sample to the y-axis. The lines going across the graphs represent fold changes of 2, 3, 10 and 30 in both directions. The points can be roped off to create smaller subsets of genes for annotation.

#### 3.2.4.2 Fold Change Graph

The fold change graph (figure 3.4) focuses on the fold change between two arrays, or two groups of arrays. The fold change parameter defines the change in transcript expression between arrays in a quantitative manner. The fold change algorithm used by the DMT software is described in appendix A3.1. As with the scatter plot, groups of genes can be roped off to create smaller subsets for annotation.

#### 3.2.4.3 The Bar Graph

The bar graph displays a number of different parameters per probe set for the pivoted samples. The parameter to be displayed is chosen manually; figure 3.5 shows the average difference values. The graph is very long in order to accommodate all the probe sets on the arrays and figure 3.5 shows a single screen shot only.

This graph was not considered useful, as finding genes with different expression patterns involves visually picking out genes that seem to show very different quantitative expression between the two samples. As is evident in figure 3.5 in order to accommodate the largest average difference values, the scale of the graph makes it difficult to identify significant differences in genes with numerically small average difference values.

Figure 3.3a – Scatter plot showing all genes

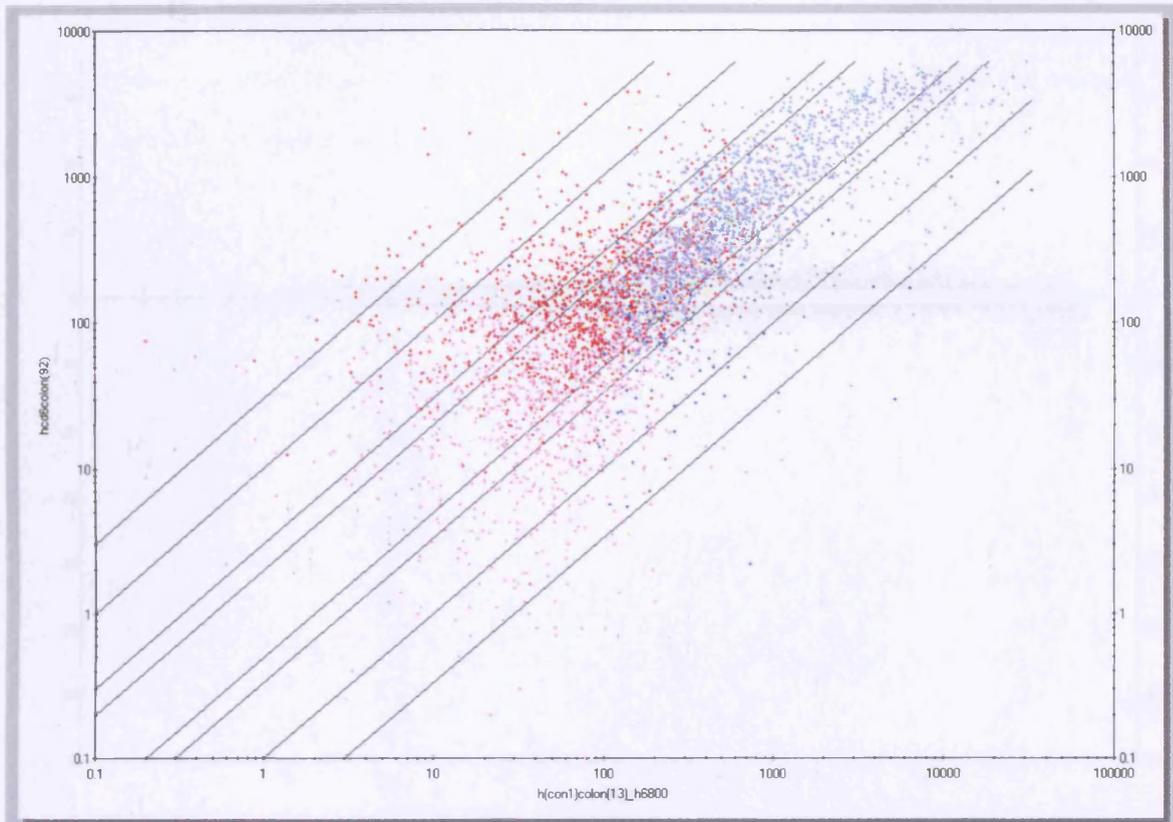


Figure 3.3b – Scatter plot showing genes with a different absolute call in the two samples

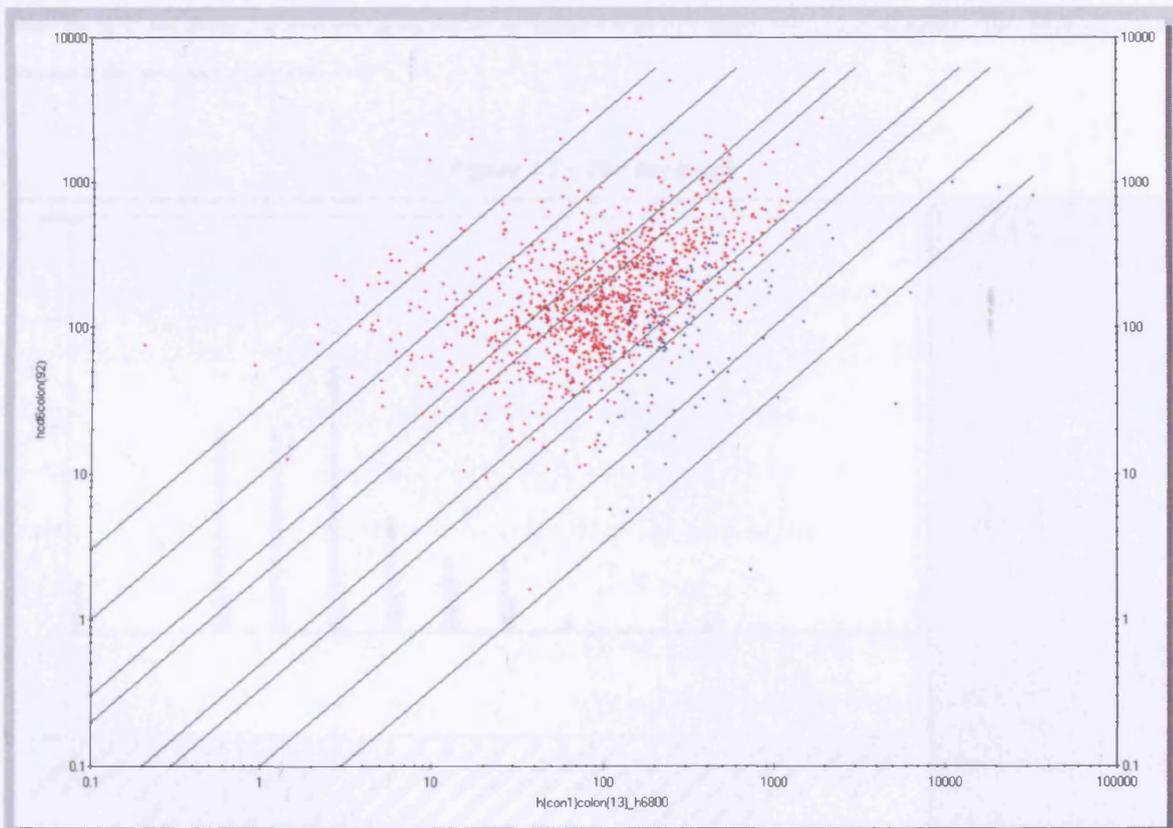
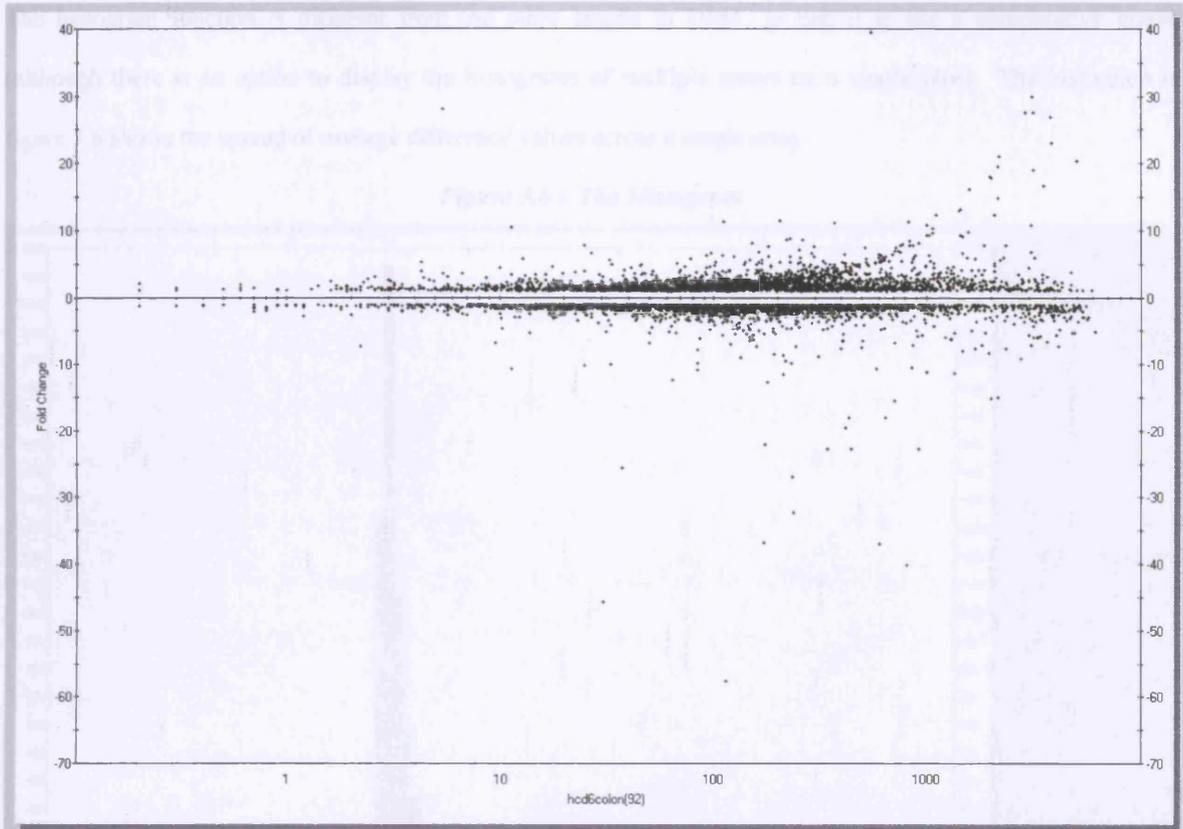


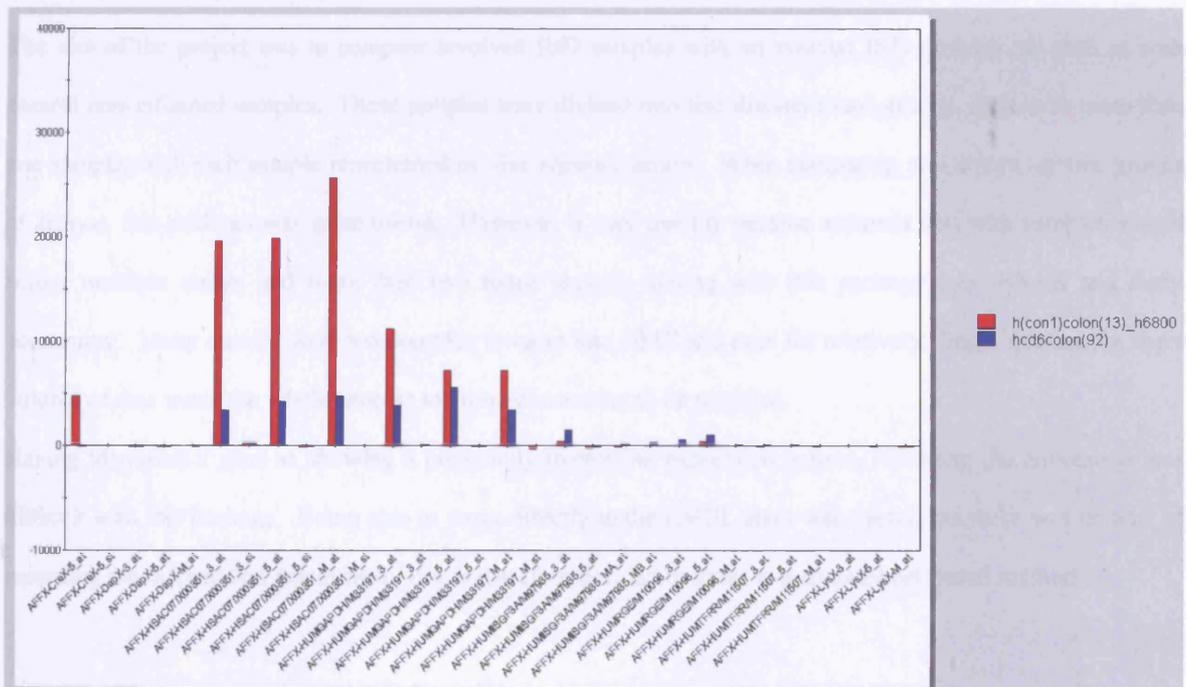
Figure 3.3a shows all the genes represented on the array. All absolute calls are represented: 'A to A' (pink), 'A to M' (pink), 'A to P' (red), 'M to A' (yellow), 'M to M' (yellow), 'M to P' (yellow), 'P to A' (light blue), 'P to M' (light blue) and 'P to P' (blue), in the NI sample to the CDi sample. Figure 3.3b shows only the genes whose absolute call changes from 'A to P' (red) and 'P to A' (blue).

Figure 3.4 – DMT fold change graph



The DMT fold change graph enables the easy identification of genes with large fold changes between the two samples. Dots appearing above the 'zero' fold change line represents genes that are up regulated in the CDi sample compared to the control. The converse is true for dots below the 'zero' fold change line.

Figure 3.5 – The bar graph

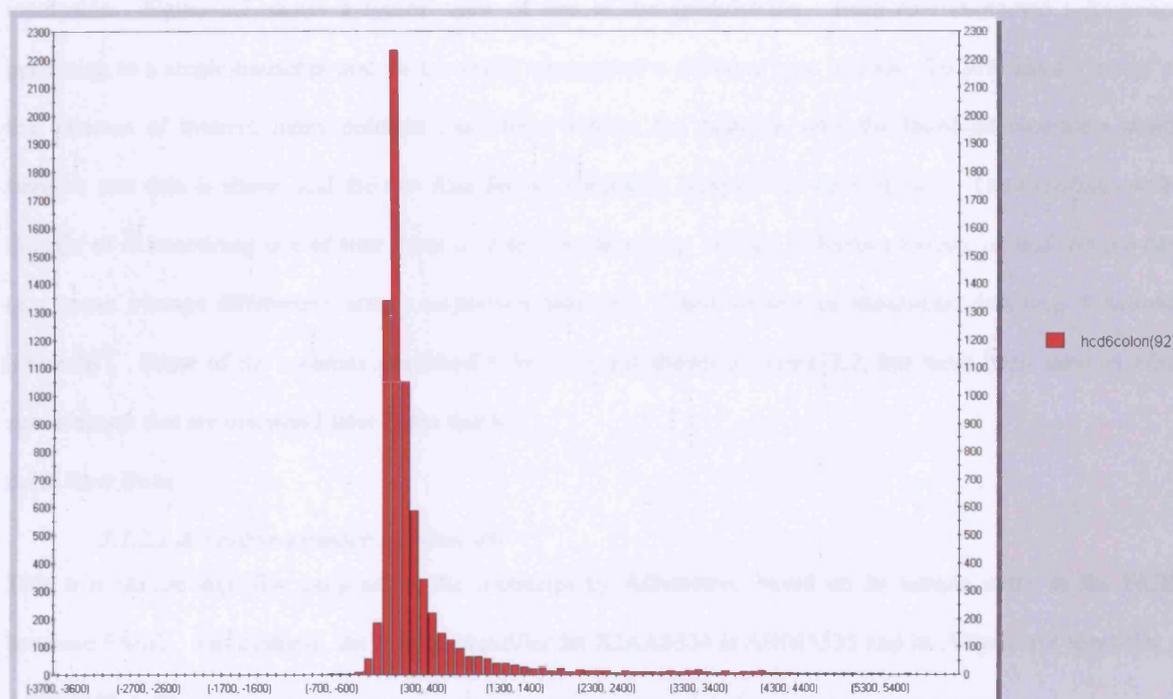


The average difference value is plotted on the x-axis, with individual probe sets on the y-axis. Large average difference values are displayed alongside numerically small values, making it difficult to identify potentially significant differences in expression.

### 3.2.4.4 The Histogram

The histogram function is different from the other graphs in DMT, in that it is not a comparative graph, (although there is an option to display the histograms of multiple arrays on a single plot). The histogram in figure 3.6 shows the spread of average difference values across a single array.

Figure 3.6 – The Histogram



The y-axis represents a range of average difference values, and the x-axis represents the number of probe sets in each range.

### 3.2.5 Discussion

The aim of the project was to compare involved IBD samples with uninvolved IBD samples, as well as with control non-inflamed samples. These samples were divided into five distinct tissue groups, each with more than one sample, with each sample represented on five separate arrays. When comparing two arrays (or two groups of arrays), this package was quite useful. However, it very quickly became apparent that with samples spread across multiple arrays and more than two tissue groups, mining with this package was tedious and time-consuming. Many queries were too complex to input into DMT and even for relatively simple queries the sheer volume of data made the whole process too time-consuming to be practical.

Having identified a gene as showing a potentially interesting expression pattern, recording the annotation was difficult with this package. Being able to move directly to the EMBL entry was useful, but there was no way of recording the information discovered. It was therefore decided to move to a spreadsheet-based method.

### 3.3 Matrix Based Mining

#### 3.3.1 Introduction

As discussed above, the arrangement of the raw data into a matrix made it possible to use a spreadsheet to mine the data. The features of Microsoft Excel were very useful for this application, with the filter facility being invaluable. Figure 3.7 shows a typical view of one of the spreadsheets. Each row contained information pertaining to a single transcript and each column represented a different type of data. To provide a view of all the columns of interest, many columns have been hidden; for example, only the involved ulcerative colitis samples raw data is shown and the raw data for the remaining samples has been hidden. The columns can be thought of as containing one of four types of data: raw data, (e.g. average difference value), pooled sample data (e.g. mean average difference), array comparison data, (e.g. T-test values), or annotation data (e.g. functional hierarchy). Some of the columns described below are not shown in figure 3.7, but have been used in other spreadsheets that are discussed later in the thesis.

#### 3.3.2 Raw Data

##### 3.3.2.1 Accession number (column B\*)

This is a unique identifier assigned to the transcript by Affymetrix, based on its unique entry in the NCBI database EMBL. For example, the EMBL identifier for KIAA0334 is AB005535 and its Affymetrix identifier is AB005535\_s\_at.

##### 3.3.2.2 Absolute call & average difference value (columns K to U)

The absolute call and average difference values were included in the spreadsheets, for each transcript on each sample. The samples were grouped according to the five tissue groups (UCi, UCu, CDi, CDu and NI) for ease of viewing the data. Figure 3.7 shows the involved ulcerative colitis columns expanded in order to view the raw data.

The absolute call value can be seen as a measure of confidence of the average difference value, which is a measure of the level of expression. Therefore, the absolute call should be considered before the average difference. Sometimes the average difference is high, but the transcript has been called absent. This is possible, as the two parameters are calculated independently of each other. As a single probe pair can skew the average difference value, the absolute call is more reliable. Therefore priority was given to the absolute call and for transcripts that had been called absent, the average difference was set to zero.

If the absolute call was present or marginal and the average difference was less than 20, the average difference was set to a baseline level of 20 as very low average difference values have been found to be unreliable indicators of expression level<sup>200</sup>.

---

\* The column letter refers to figure 3.7

Microsoft Excel - All IBD\_with UC13

File Edit View Insert Format Tools Data Section Budgets Window Help

Times New Roman 10 B I U

1 2 3 4	A	B	G	H	I	Involved UC					Inv UC		Mean	Two-tailed T-test Probability (less than 5 is sig at 6)											
						UCL1	UCL2	UCL3	UCL4	UCL5	Inv UC	Count		Avg Diff	control & uninv UC	inv UC & uninv UC	inv UC & control	inv UC & inv CD	inv CD & control	inv CD & uninv CD	con & ur C				
						Avg Dihbs	Cavg Dihbs	Dihbs	Cavg Dihbs	Cavg Dihbs	Cavg Dihbs	Call	A	P	Inv UC										
34915	W28161_st	mRNA, complete cds.				1307	P	0	A	1792	M	1252	P	2540	P	1	3	1394.4	87.15	28.97	19.76	99.56	7.67	8.66	64
34916	W28214_st	lcyte Unique				0	A	0	A	0	A	0	A	0	A	5	0	0.0	1	1	1	37.39	37.39	37.39	
34917	W28229_st	Human mRNA for centaurin beta2.				0	A	0	A	0	A	0	A	0	A	5	0	0.0	1	1	1	19.02	19.02	19.02	
34918	W28230_st	lcyte Unique				0	A	0	A	455	P	0	A	0	A	4	1	91.0	42.26	70.86	37.39	74.95	37.39	37.39	
34919	W28235_st	Human potential membrane protein C1orf11 mRNA, complete cds.				1175	P	197	P	523	P	737	P	1298	P	0	5	786.2	5.78	2.31	50.45	50.52	83.04	7.68	29
34923	W28275_st	Y60ASA.20				0	A	0	A	0	A	0	A	0	A	5	0	0.0	1	1	1	1	1	39.10	39
34924	W28360_st	cds.				0	A	0	A	0	A	0	A	0	A	5	0	0.0	86.58	42.26	36.32	18.19	30.61	69.50	8.
34925	W28362_st	partial cds.				446	P	0	A	0	A	237	P	0	A	3	2	136.5	60.85	39.11	59.79	60.22	6.70	28.91	50
34926	W28366_st	partial cds.				0	A	0	A	0	A	414	P	713	P	3	2	225.3	4.54	9.73	84.83	16.14	2.09	42.12	11
34928	W28391_st	homolog (MG4-1) mRNA, complete cds.				###	P	252	P	1221	P	1329	P	1329	P	0	5	1542.1	4.93	7.74	32.90	16.89	10.88	1.13	51
34929	W28404_st	lcyte Unique				649	P	0	A	0	A	896	P	312	P	2	3	369.4	61.87	71.52	62.75	82.50	70.82	59.80	94
34930	W28406_st	ADSE00627.				0	A	0	A	0	A	0	A	0	A	5	0	0.0	36.32	1	36.32	1	36.32	1	36
34931	W28414_st	bs86; (from clone DKF2p564O1716);				0	A	0	A	0	A	0	A	0	A	5	0	0.0	36.32	1	36.32	1	36.32	1	36
34933	W28450_st	complete cds.				0	A	0	A	0	A	125	P	253	P	3	2	75.6	18.65	20.93	6154	3107	39.64	22.88	18.
34935	W28498_st	ZB81C06.				0	A	0	A	221	P	101	P	256	P	2	3	115.7	49.91	11.82	25.11	30.54	74.10	53.85	44
						822	P	0	A	0	A	0	A	0	A	4	1	164.4	86.07	91.45	96.35	80.12	80.36	85.47	44
34937	W28516_st	putative transporter				0	A	0	A	0	A	0	A	0	A	5	0	0.0	36.32	1	36.32	1	36.32	1	36
34939	W28548_st	ZEDSA03.				0	A	0	A	0	A	0	A	0	A	5	0	0.0	36.32	1	36.32	1	36.32	1	36
34941	W28595_st	Human tetrahydrocoptide repeat protein (tpr2) mRNA, complete cds.				297	M	0	A	0	A	0	A	0	A	4	0	53.5	42.00	85.83	56.53	83.53	6194	27.27	23
34942	W28610_st	fragment, clone TM4, reverse read cpg7H4.rtd.				316	P	0	A	0	A	0	A	0	A	4	1	63.2	1	37.39	37.39	37.39	1	1	

All arrays with UC13 /

Figure 3.7 – Typical view of Microsoft Excel spreadsheet for matrix-based mining

### 3.3.3 Pooled Sample Data

#### 3.3.3.1 Mean average difference (column BU)

The mean of the average difference ( $\chi$ ) values for each sample in a tissue group was calculated and reported in the 'mean average difference' ( $\chi$ -bar) column. In figure 3.7, there is only space to show the involved ulcerative colitis 'mean average difference' column.

#### 3.3.3.2 Absolute call count (columns BL to BM)

There are two absolute call count columns for each tissue group. These indicate the number of samples in which, a transcript was called either absent or present. These columns enable the selection of genes that, for example, are called present in 100% of the involved ulcerative colitis samples.

#### 3.3.3.3 Standard Deviation

The standard deviation (SD) for each 'mean average difference' value was calculated using the formula function in Excel.

#### 3.3.3.4 Standard Error

The standard error (SE) for each 'mean average difference' value was calculated using the following formula. It was used as an estimate of error in some graphical representations of the data (mentioned when used).

$$SE = \frac{SD_{NI}}{\sqrt{(n_{NI} - 1)}}$$

Where: n = number of samples in the tissue group

### 3.3.4 Array Comparison Data

The data was initially compared in a pair wise manner and patient matching was carried out based on age at colectomy and sex. At this stage only ulcerative colitis samples were available and these were matched to the NI patients. However, the age at colectomy varied by a large amount between the two patient groups (table 3.1). It was therefore decided to abandon patient matching and the data was mined in a group wise manner.

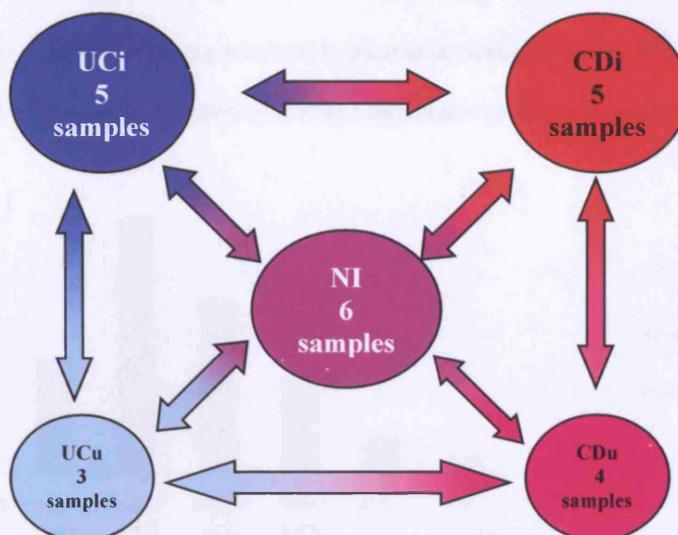
*Table 3.1 – Matching the UC and NI patients on basis of sex and age*

<i>UC patient (sex/age)</i>	<i>NI patient (sex/age)</i>
UC_1 (F / 61)	NI_4 (F / 65)
UC_2 (F / 36)	NI_5 (F / 72)
UC_3 (M / 30)	NI_6 (M / 51)
UC_4 (F / 54)	NI_1 (F / 72)
UC_5 (F / 36)	NI_3 (F / 77)

#### 3.3.4.1 Fold Change

The 'mean average difference' for each transcript across each tissue group (section 3.3.3.1) was used to calculate the fold change between comparative tissue groups (figure 3.8).

Figure 3.8 – Comparison of Tissue Groups



Two fold change values were calculated for each transcript. For example, the fold change between UCi and NI samples for the MHC HLA-Dw12 DQ-beta chain transcript (accession number AA258595) was calculated as follows:

$$\bar{\chi}_{UCi} / \bar{\chi}_{NI} = 2981 / 469 = 6.35$$

$$\bar{\chi}_{NI} / \bar{\chi}_{UCi} = 469 / 2981 = 0.16$$

∴ the transcript is expressed at a level 6 times greater in the UCi samples than the NI samples.

The fold change columns could therefore be used to select **all** the transcripts that were showing higher expression in one tissue group compared to another. However, this does not give any information about the statistical significance of the observations.

#### 3.3.4.2 Normalisation Method

This formula attempted to indicate whether the difference in mean average difference between two tissue groups was significant or not. If the value was calculated to be greater than 2, the difference in expression, (between the UCi and NI samples in this example), was considered significant.

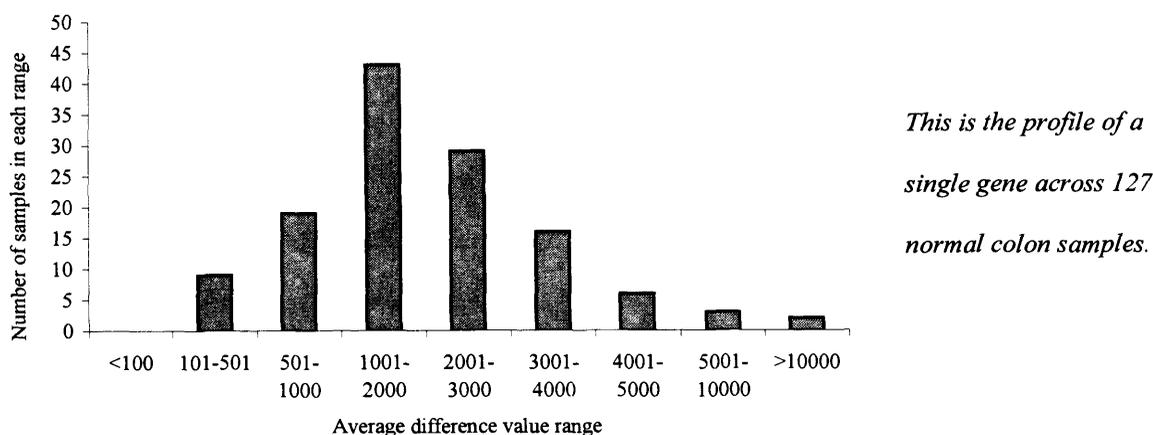
$$\text{Normalisation value} = \frac{\bar{\chi}_{UCi} - \bar{\chi}_{NI}}{SD_{NI}}$$

#### 3.3.4.3 T-test values (columns CE to CK)

The T-test can only be used on data that shows an approximately normal distribution, i.e. the expression of the gene within each tissue group must be normally distributed. At first glance, the microarray data does not show a normal distribution. However, the T-test is not a global test of significance and is limited to testing the significance of one gene across two different tissue groups. With a greater number of samples in a tissue group,

the expression data shows an approximately normal distribution (figure 3.9). The two-tailed T-test has been used to test the statistical significance of oligonucleotide microarray data previously <sup>200</sup>.

**Figure 3.9 – Normal distribution of average difference values in a larger sample set\***



However in this data set, the number of T-tests carried out (7 per gene) and the small number of samples meant that the T-test could not be used to ascertain statistical significance of the data; i.e. some of the 'significant' results would be expected by chance alone as the result of a 'type II error'. A Bonferroni correction could be applied to correct the problem, but with so many T-tests and so few samples, the correction value would be extremely small as has been noted previously <sup>201</sup>. Complex statistical treatments have been developed and applied to microarray data <sup>202, 203</sup>, but these were not applicable to the present data set due to the small number of samples. Therefore, the T-test was used as a method of selecting gene expression patterns of interest across the five tissue groups, as discussed in chapter four. Figure 3.8 illustrates the different comparisons done, with each arrow representing a T-test (all columns are not shown in figure 3.7 due to lack of space). The problem of statistical analysis is considered further in the final chapter.

### 3.3.5 Annotation Data

#### 3.3.5.1 Incyte Description (column G)

The sponsoring company subscribes to the Incyte database and although unable to mine the database directly, the author did have access to Incyte 'look up tables'. These correlate the Affymetrix identifier to the most recent description and functional hierarchy in the Incyte database. The description column provided the text description of the transcript.

#### 3.3.5.2 Incyte Functional Hierarchy (columns H to I)

The functional hierarchy is a gene classification system based on the function of the encoded protein. Using the filters in Excel, this is column increases the flexibility and functionality of the spreadsheet based mining method.

\* Figure 3.9 was created using data from the GeneLogic database, GeneExpress. This database details the results of the application of thousands of different samples to Affymetrix arrays. The author had access to the database through AstraZeneca for this figure.

### 3.3.5.3 My Annotation (column A)

This column was included to provide a space for personal observations and notes within the spreadsheet, although detailed notes on genes of interest were also maintained as separate Microsoft Word documents.

### 3.3.6 Discussion

The columns described above were used to varying extents as the spreadsheets evolved. As the number of columns increased, the flexibility of mining increased. However, the spreadsheets became increasingly unwieldy as the amount of data gathered increased. With only the HuFL array data (7129 probe sets) all the columns described above could be used. However, with the Hu35K set (4 arrays), the spreadsheet was too large if all the columns were incorporated. Such large spreadsheets slowed down the computer to such an extent that even opening the file took time. Also some columns were less useful than others, therefore the spreadsheet quickly evolved to include only those columns that were necessary and provided the most information.

The Hu35K set and the HuFL array data was combined on to a single spreadsheet to eliminate duplication of effort. All the pooled data columns were kept, except for the standard error, as it was felt that the standard deviation would be a more accurate error bar for graphs plotting the mean average difference value for each tissue group.

Of the different array comparison methods, it was felt that the T-test method was the best, as it added the dimension of statistical significance to the data. The normalisation method has not been used to establish significance in a study of this kind (to the author's knowledge), unlike the T-test method<sup>200, 201, 204</sup>. Selection of genes based on fold change has been very popular<sup>180, 181, 191, 205</sup>, but as it lacks the statistical significance dimension it was not employed here for this purpose. However, due to its popularity in previous publications, most of the figures created to show differential gene expression utilise the fold change data. Therefore, the fold change columns were retained for use in differential expression figures.

The annotation data columns were kept on the spreadsheet as they provide a method of keyword based searching. This is not comprehensive, as the search results depend on what the transcript has been named and while some gene names are functionally descriptive, others are not. Also the genes may have several functions within the cell and therefore the name may be descriptive only of the first studied function. However, this provides a starting point for further annotation of the genes of interest using public databases.

### 3.4 Similarity Metrics (Clustering)

The large amount of data generated by microarray analysis implicates the use of similarity metric based analysis, or clustering. Essentially, this is dividing the items of a set into related subsets based on a distance metric. In the case of the Affymetrix array data, the items can be genes or arrays and the distance metric is the numerical average difference value. There are many clustering methods applicable to microarray data; these differ by the algorithm the method is based on and therefore the output. The author experimented with three publicly available programs: Stanford's Cluster (available at <http://rana.lbl.gov/>), the Whitehead / MIT Institute's GeneCluster (available at <http://waldo.wi.mit.edu/MPR/>) and MolMine's J-express (available at <http://www.molmine.com>). Together these cover the most common clustering techniques; hierarchical clustering, K-means clustering, self-organising maps (SOMs) and principle component analysis (PCA).

#### 3.4.1 Hierarchical Clustering

##### 3.4.1.1 Definition

The aim behind this type of clustering is to create a dendrogram of the items to be clustered (either genes or arrays), where the length of the branches represent relative similarity. The beginning of each branch represents a divergence in similarity between the items and the further apart two items are, the less similar they are. Conversely, the closer together two items are, the more similar they are.

##### 3.4.1.2 Computational method

There are many similarity metrics that can be used for hierarchical cluster analysis. In this thesis, Stanford's 'Cluster' program was used to create hierarchical dendrograms. This program uses the Pearson correlation. Cluster provides the option of four different variations of the Pearson correlation. The standard textbook version (correlation centred) was used when clustering arrays, as it is a parametric test and the shape of the expression pattern is more important than the actual values.

The standard Pearson correlation coefficient between any two series of numbers;

$X = \{X_1, X_2, \dots, X_N\}$  and  $Y = \{Y_1, Y_2, \dots, Y_N\}$  is defined as

$$r = \frac{1}{N} \sum_{i=1, N} \left( \frac{X_i - X\text{-bar}}{\sigma_X} \right) \left( \frac{Y_i - Y\text{-bar}}{\sigma_Y} \right)$$

Where: X-bar is the average of values in X and  $\sigma_X$  is the standard deviation of these values.

However, this standard metric considers genes with perfectly opposite expression patterns as being very distant. In the cell, two genes with perfectly opposite expression patterns are likely to be part of the same pathway, with

one gene being expressed at lower levels, as the expression levels of the second gene rise, i.e. genes with opposite expression patterns may actually be closely related. Therefore, a modified version of the standard Pearson correlation metric was used when clustering genes. This modified version (correlation uncentred) is also a parametric test\* and is described below.

$$r = \frac{1}{N} \sum_{i=1}^N \left( \frac{X_i}{\sqrt{\frac{1}{N} \sum_{i=1}^N (X_i)^2}} \right) \left( \frac{Y_i}{\sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i)^2}} \right)$$

For both versions of the correlation the value of  $r$  is always between 1 and  $-1$ , with 1 indicating a perfect correlation, 0 indicating no correlation (i.e. completely independent) and  $-1$  indicating a perfectly opposite correlation.

### 3.4.1.3 Output

The program calculates the similarity between each item using the metric selected and then uses this information to cluster the items. The items to be clustered (i.e. the genes or arrays) can be thought of as true items and the forming dendrogram as a pseudo-item. A vector is assigned to each pseudo-item and this vector is used to calculate distances between this pseudo-item and true items. The vector assigned to the pseudo-item is the mean of all the true items represented by the pseudo-item.

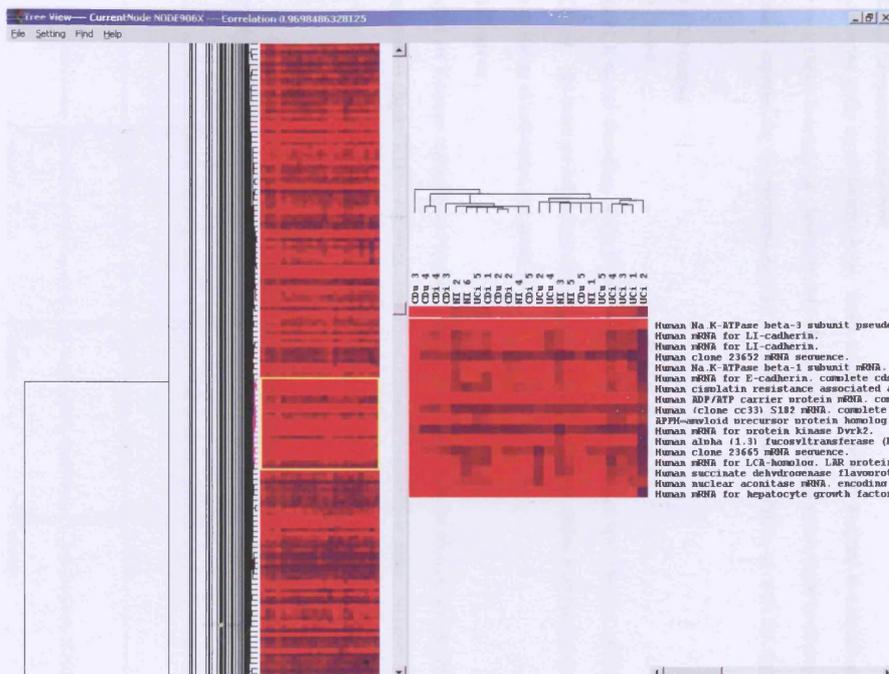
The hierarchical clustering process is common to all the clustering programs. The true items are organised into a list and the two most similar items are joined first. These joined items are removed from the list and replaced by a branch (or node) representing both of them (i.e. the pseudo-item). The distance between the pseudo-item and all other items is calculated and the next most similar true item added to the branch and removed from the list. In some cases a true item will be more similar to another true item than to the forming pseudo-item and in this case a second pseudo-item will be formed. The process is repeated until only a single pseudo-item remains, i.e. the fully formed dendrogram with all the true items represented.

The dendrogram created by Cluster is viewed in another program called TreeView (also available at <http://rana.lbl.gov>) and figure 3.10 shows a typical dendrogram as seen in TreeView†.

\* Descriptions of the non-parametric correlation metrics available in Stanford's Cluster program, can be found at [www.ulib.org/webRoot/Books/Numerical\\_Recipes/bookcpdf/c14-6.pdf](http://www.ulib.org/webRoot/Books/Numerical_Recipes/bookcpdf/c14-6.pdf)

† GeneCluster and J-express contain viewers within the same software package

Figure 3.10 – View of a dendrogram created in Stanford's Cluster program and viewed in TreeView



The genes showing patterns of interest (left) can be selected to appear in the zoomed version (right). In this example, both genes and arrays have been clustered.

### 3.4.2 K-means

#### 3.4.2.1 Definition

The K-means clustering method organises items into a defined number of groups/nodes, where items within a node are similar and items between nodes are dissimilar. The number of nodes is defined manually and, as in hierarchical clustering, both genes and arrays can be clustered.

#### 3.4.2.2 Computational method

K-means clustering works on an iterative basis. Items are initially randomly assigned to a cluster and the mean vector for each cluster is computed. Items are then reassigned to the cluster whose mean is closest to the item. This process is repeated for the maximum number of cycles (manually defined), or until the system reaches stability.

#### 3.4.2.3 Output

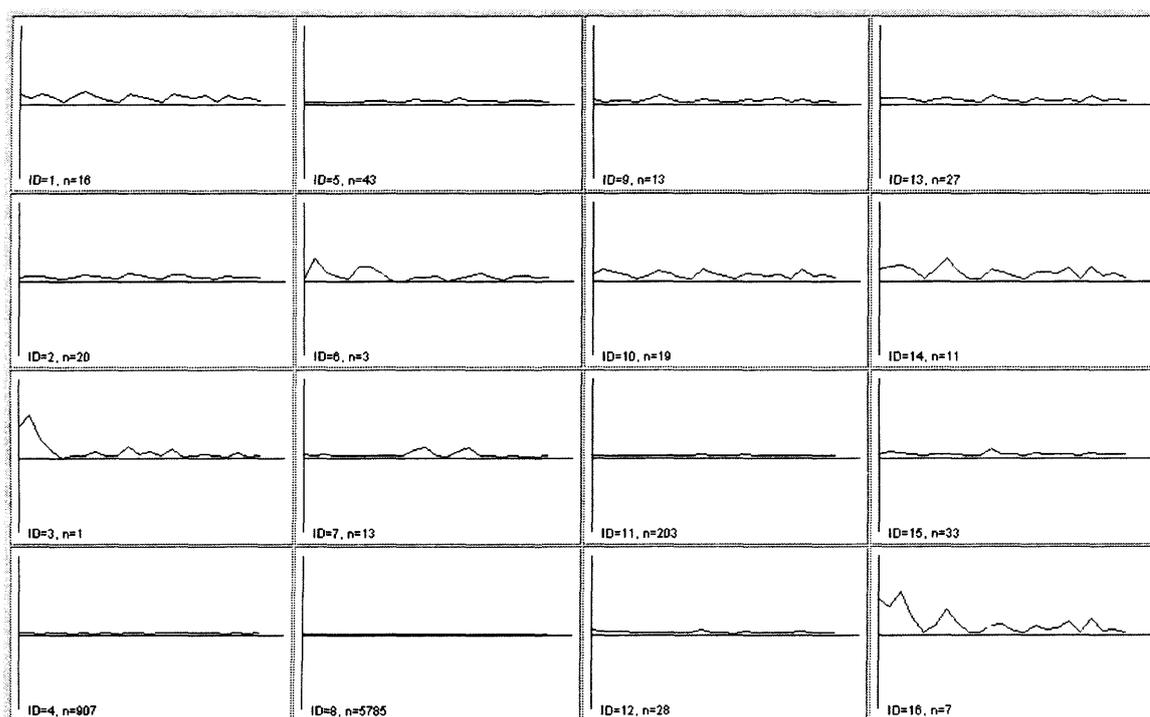
##### *Stanford's Cluster*

The output of K-means clustering in the Cluster program is a tab-delimited text file that can be viewed in Microsoft Excel. The items are listed in the top row, with a blank column to define individual clusters. Cluster allows the clustering of both arrays and genes.

##### *Molmine's J-express*

The output of the K-means clustering in J-express consists of a series of graphs showing the expression pattern of different clusters (figure 3.11). J-express only allows K-means clustering of genes, however the graphical output is very useful as it enables the easy identification of clusters of genes that show interesting expression patterns.

**Figure 3.11 – K-means clustering using J-express**



### 3.4.3 Self-Organising Maps

#### 3.4.3.1 Definition

Self-organising maps (SOMs) are similar to K-means cluster analysis in that they cluster items into a manually defined number of nodes. The main difference is that in SOMs the nodes are ordered, so that there is a progression of similarity between the nodes, i.e. in a SOM of 5 nodes, nodes 1 and 2 will be more similar than nodes 1 and 3 and nodes 1 and 5 will be the least similar of all.

#### 3.4.3.2 Computational method

Tamayo et al describe the application of the Whitehead / MIT Institute's GeneCluster program's SOM component to the analysis of haematopoietic differentiation<sup>206</sup>. This paper includes a good description of the computation of a SOM.

#### 3.4.3.3 Output

The output of the SOM clustering is identical to the output of the K-means clustering in each program. In the Cluster program, the tab-delimited text file output is not obviously ordered and it is impossible to see the progression order of the nodes. Both J-express and GeneCluster produce a view similar to that shown in figure 3.10. The difference is that there is an appreciable gradual change from the first to the last node.

### 3.4.4 Principle Component Analysis (PCA)

#### 3.4.4.1 Definition

K-means and SOMs allow the visualisation of the expression pattern of each cluster across the samples. PCA allows visualisation of the distance between clusters. In brief, PCA describes the relatedness of clustered items (genes or arrays) using  $k$  variables to represent the expression profiles of  $k$  genes. The PCA program aims to replace the  $k$  variables by a smaller number of principal components,  $p$ , which are plotted onto a graph for easy visualisation of relatedness.

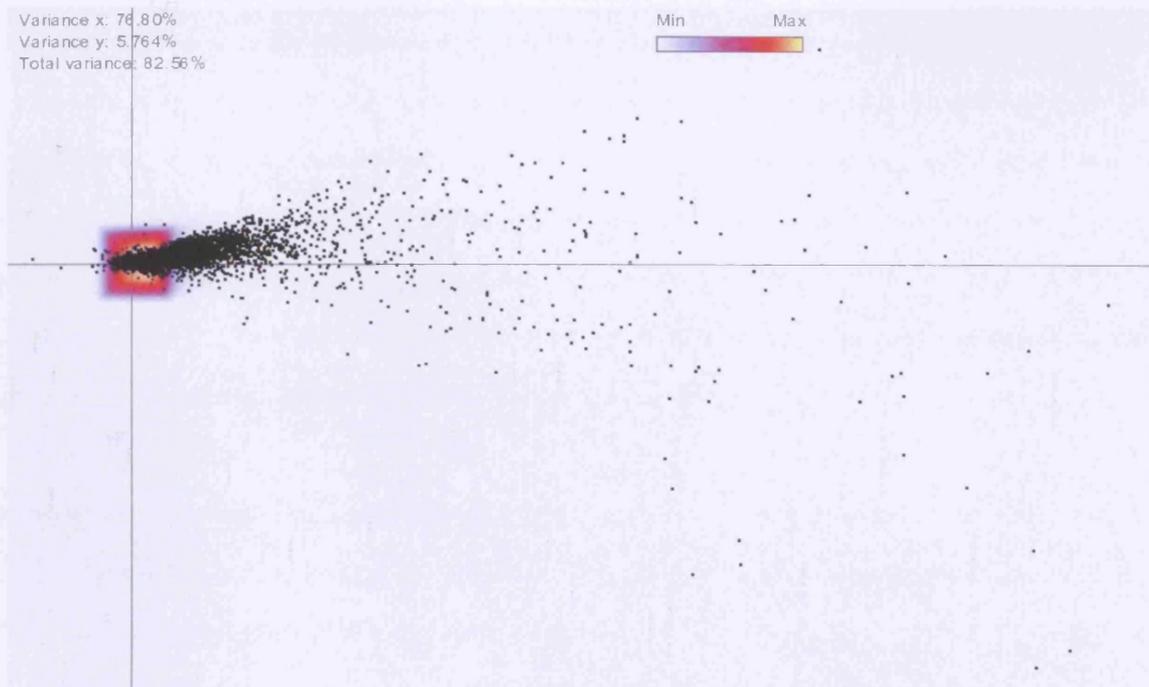
#### 3.4.4.2 Computational method

The computation of a PCA is described in detail in Lu *et al*<sup>207</sup>.

#### 3.4.4.3 Output

There is no PCA function in GeneCluster. The PCA output in Cluster is a text file, which the author feels negates the value of the PCA analysis. The output of a PCA analysis in J-express is shown in figure 3.12. In the 2D version, hovering over the spots gives the gene information; however this is not the case in the 3D view. Selecting a cluster of genes in the 2D view plots the expression pattern of the selected genes in a format similar to that seen with K-means and SOM analysis.

Figure 3.12 – PCA output in J-express



### 3.4.5 Evaluation of similarity metric methods

Clustering is a useful approach as it can be used to group together genes with similar expression patterns. The assumption is that genes that show a similar, or perfectly opposite expression pattern could be part of the same pathway. In this way novel relationships between distinct groups of genes may be discovered. Clustering on the basis of gene expression to look at clinical tissue group is also a useful method. If enough samples and genes are investigated, clustering analysis of microarray data could be used to diagnostically. A previous study showed the use of cluster analysis to distinguish between B lymphomas which responded to therapy and those which did not<sup>208</sup>. Clustering methods become more useful as the amount of data put into them increases. Therefore, they are ideal for data from multiple experiments<sup>209</sup>, e.g. from a public data repository.

Only the methods applied to the data generated in the current study have been detailed in this chapter. There are other methods available and these tend to be variations of the four basic methods described. The emerging application of neural networks to this type of data analysis holds great promise and there are a number of groups using unsupervised neural networks to perform hierarchical clustering on gene expression data<sup>210, 211</sup>. The author did not have access to this technology for the current study, but the potential of this approach is recognised.

**CHAPTER FOUR – DATA MINING RESULTS**

#### 4.1 Chapter Introduction

---

To increase the chances of finding gene expression patterns of interest, the data\* was mined using a combination of the methods described in chapter 3; for example the T-test columns were used in combination with the absolute call column within the spreadsheet. The exact method used depended on the question being asked and each section in this chapter discusses the method used and results of a single query. Similarity metrics were also used to cluster both genes and arrays. This strategy was instrumental in the identification of an atypical UCI sample (section 4.3). Typically, the spreadsheet was used to identify genes showing interesting expression patterns and then clustering methods were applied to the resulting data set.

Having identified and clustered genes showing an interesting expression pattern, a number of questions arose:

- 1) Are there any protein function properties that the members of the cluster have in common?
- 2) How do these functional properties relate to the present query?
- 3) Has this pattern of gene expression been reported in the published literature previously?
  - i) In IBD tissue?
  - ii) In inflamed tissues?
  - iii) In colonic or intestinal tissue?

These are annotation questions; the answers for which were sought by intensive use of a number of different databases. It was attempted to answer each of these questions for each gene of interest. Some of the 'interesting' genes were EST sequences with very little annotation and in these cases the EST sequence was BLASTed against public sequence data using an AstraZeneca internal portal. This internal portal (called sequence retrieval system or SRS) allowed access to a number of databases (table 4.1) containing both public and proprietary data. The author only had access to data in the public domain, but was able to take advantage of the search and retrieval facilities of the SRS portal. Two other databases that the author found particularly useful due to their speed are publicly available on the Internet; PubMed† and GeneCards‡. PubMed is similar to MedLine in that it is a bibliographic database, but PubMed is updated daily whereas MedLine is updated less frequently. Freely available to academic researchers, GeneCards contains data from many of the databases in table 4.1 on a single page per gene, eliminating the need to search through different databases for the same information.

---

\* from HuFL, Hu35K\_subA, Hu35K\_subB and Hu35K\_subC arrays only, as discussed in section 2.2.9.5.

† <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>

‡ <http://bioinfo.weizmann.ac.il/cards>

Table 4.1 – Databases accessed through SRS for gene annotation in the current study

<i>Database Name</i>	<i>No. of Entries*</i>	<i>Type of Data</i>	<i>Description of Database</i>
EMBL	12784552	Nucleotide	Maintained by EBI†. Nucleotide sequences collected from literature, patent applications and direct submission. Each entry equals one contig. Full release every 3 months
EMBLNEW	222950	Nucleotide	Data store in between full EMBL release. Updated 4 times a week
EMBLDAILY	3427	Nucleotide	Previous days submissions to EMBL.
UNIGENE	96768	Nucleotide	Created to provide novel, non-redundant mapping candidates from EST data. Contains nucleotide sequence clusters containing known genes and clusters of just ESTs.
SWISSPROT	86593	Protein	Translation of EMBL database, collected from literature, direct submissions. Very high quality annotation.
SWISSNEW	55432	Protein	Data store in between full SWISSPROT release
TREMBL	525264	Protein	EMBL translation sequences, not yet incorporated into SWISSPROT, awaiting annotation
TREMBLNEW	93677	Protein	Latest data, awaiting submission to main TREMBL database
OMIM	13381	Genetic Information	Cytogenetic map location of disease genes and brief overview of information from published literature.
OMIMALLELE	9343	Genetic Information	Alleles (mutations and polymorphisms) reported in OMIM
MEDLINE	10726231	Literature	NLM‡ bibliographic database covering all aspects of health and pre-clinical sciences.
PATHWAY	148	Pathways	List of pathway maps in KEGG§ system
Hu6K	6454	Gene Expression	Annotations for genes represented on Hu6K array which have an entry in EMBL
Hu35K	34775	Gene Expression	Annotations for genes represented on Hu35K array set which have an entry in EMBL
GENESEQN	786274	Patented Sequence	Covers all nucleic acids >10 bases long and all PCR primers and probes within patents. Updated fortnightly.
GENESEQP	441873	Patented Sequence	Covers all protein sequences >4 amino acids long and all PCR primers and probes within patents. Updated fortnightly

\* September 2001

† European Bioinformatics Institute

‡ United States National Library of Medicine

§ Kyoto Encyclopaedia of Genes and Genomes

Recording the annotation of the many genes investigated was a significant challenge and was solved in two ways. Initially all annotation was added directly to the spreadsheet. A column was added (figure 4.1, column A) to enable brief notes to be made during the mining process. This column could then be used to re-identify genes as and when needed. However, detailed notes could not be added in a single column, so further columns were introduced (figure 4.1 columns P to U). The rationale behind recording annotation data directly onto spreadsheet was to enable selection of genes based on annotation data, in addition to the numerical data. However, the subsequent availability of the functional hierarchy reduced the usefulness of this manual annotation effort. Also, the spreadsheet had to be re-saved after every addition to an annotation column. The ever-increasing size of the spreadsheets meant this took a significant amount of time and amplified the possibility of the file becoming corrupt.

Figure 4.1 – Annotation columns within the spreadsheet

A	B	D	O	P	Q	R	S	T	U	V
Annotation	Acc #	Affy Description (August 1997)	Gene function	Novel?	Subcellular Location	Extra info	Diseases	Imp in CD or UC?		My Annotation
	AF001548	815A9.1 gene (myosin heavy chain) extracted from Homo sapiens chromosome 16 BAC clone CIT987SK-815A9 complete sequence.	muscle							
	D00654_at	Human enteric smooth muscle gamma-actin gene, 5' flank and	muscle							
	D00749_s	Human T cell surface antigen CD7 gene - Also Represents: X06180	inflamm marker	expected			expressed by thymocytes & mature T cells			
metall(	D11139_at	Human gene for tissue inhibitor of metalloproteinases, partial sequence	<b>TIMP</b>							
	D11428_at	Human mRNA for PMP-22(PAS-II/SR13/Gas-3) of peripheral myelin, complete cds	nervous system ycc		transmem		charcot-marie tooth disease type 1A; C			
	D26129_at	Human mRNA for ribonuclease A (RNase A), complete cds	Rnasec							
	D26535_s	Human gene for dihydroipoamide succinyltransferase, complete cds (exon 1-15) - Also Represents: S72422	TCA cycle							
NOS	D29675_at	Human inducible nitric oxide synthase gene, promoter and exon 1 /gb=D29675 /ntype=DNA /annot=exon - Also Represents: X73029	NOS			iNOS?				
NOS	D29675_s	Human inducible nitric oxide synthase gene, promoter and exon 1 /gb=D29675 /ntype=DNA /annot=exon - Also Represents: X73029	NOS					CD cluster LH		
	D42047_at	Human mRNA for KIAA0089 gene, partial cds	?							
metall(	D45917_s	Human mRNA for TIMP-3, partial cds (C-terminus region) - Also Represents: U14394	<b>TIMP</b>							
	D49950_at	Human Liver mRNA for interferon-gamma inducing factor(IGIF), complete cds	IFN inducing				expressed by r. IL-18 family member; Th1 cytokine			

Therefore, a template to annotate the genes was created. The template headings were used as general reminders of searches to do, but were changed according to what was relevant to the gene in question. A separate Word document was created for each gene of interest and this was updated with new information as needed. This proved a useful method of recording gene annotations.

The specific mining parameters used to answer specific queries are detailed in this chapter. The exact strategy used to mine the data was dependant on the query being posed and the chapter is divided into sections detailing the strategy and results from a single query. Thus, the section headings are in the form of the question being addressed in that section.

## 4.2 Which genes are differentially expressed in the involved IBD samples?

### 4.2.1 Rationale

One of the main aims of this study was to identify genes that are differentially expressed between involved Crohn's disease and ulcerative colitis; therefore this was the first question asked of the data.

### 4.2.2 Methods

The first query was, 'over-expressed in CDi compared to UCi'. To answer this question, the T-test and absolute call columns within the spreadsheet were used (table 4.2).

*Table 4.2 – Criteria for the query 'Over expressed in CDi compared to UCi'*

<i>Criteria</i>	<i>Genes returned by individual query</i>	<i>Genes returned in cumulative query</i>
CDi vs. UCi T-test column = $P < 0.05$	1200	1200
Absent in $\geq 4$ UCi samples	21914	623
Present in $\geq 4$ CDi samples	9671	557

Hence this query returned genes that were called absent in 80% or more of the UCi samples, called present in 80% or more of the CDi samples and showed a significant difference in the mean average difference between the two groups. As table 4.2 shows, a large number of genes were returned. The selection method was repeated to answer the query 'Over expressed in UCi compared to CDi' (table 4.3).

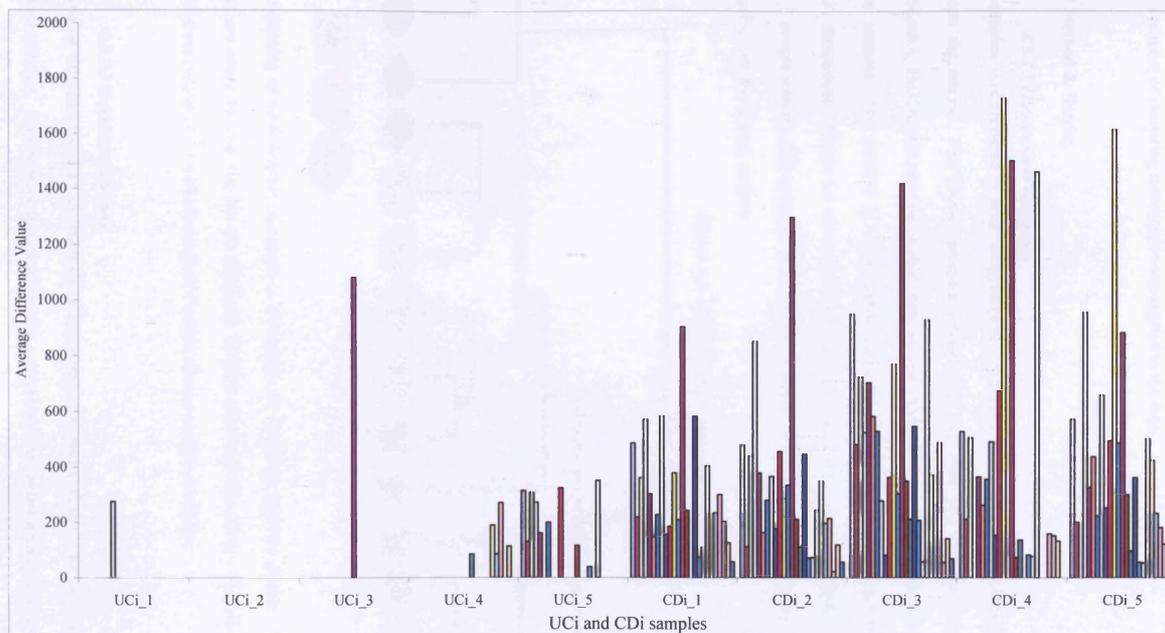
*Table 4.3 – Criteria for the query 'Over expressed in UCi compared to CDi'*

<i>Criteria</i>	<i>Genes returned by individual query</i>	<i>Genes returned in cumulative query</i>
CDi vs. UCi T-test column = $P < 0.05$	1200	1200
Present in $\geq 4$ UCi samples	21914	15
Absent in $\geq 4$ CDi samples	9671	10

### 4.2.3 Results & Section summary

As kinases and phosphatases tend to make good drug targets, these gene classes were selected from the 567 genes using Excel's custom filter with the criteria 'contains kinase OR phosphatase' in the gene description column. This returned 25 genes, all of which were over expressed in the CDi samples (appendix A4.1). One of the UCi samples (UCi\_5) was expressing many more of these genes than the other UCi samples (figure 4.2), which alerted the author to the possibility that this sample was showing atypical gene expression. Consequently this sample was subjected to further investigation.

Figure 4.2 – Expression of 'over expressed in CDi' kinase / phosphatase genes, in sample UCi\_5 with minimal expression in rest of UCi samples



N.B. The graph shows the low, but consistent expression of many of these genes in the UCi\_5 sample, whereas in the other UCi samples these genes tend to be scored absent. The names of the genes are irrelevant to this observation and so have not been given here, but are listed in appendix A4.1 in the same order as the bars of the graph.

### 4.3 Identification of an abnormal involved ulcerative colitis sample

#### 4.3.1 Rationale

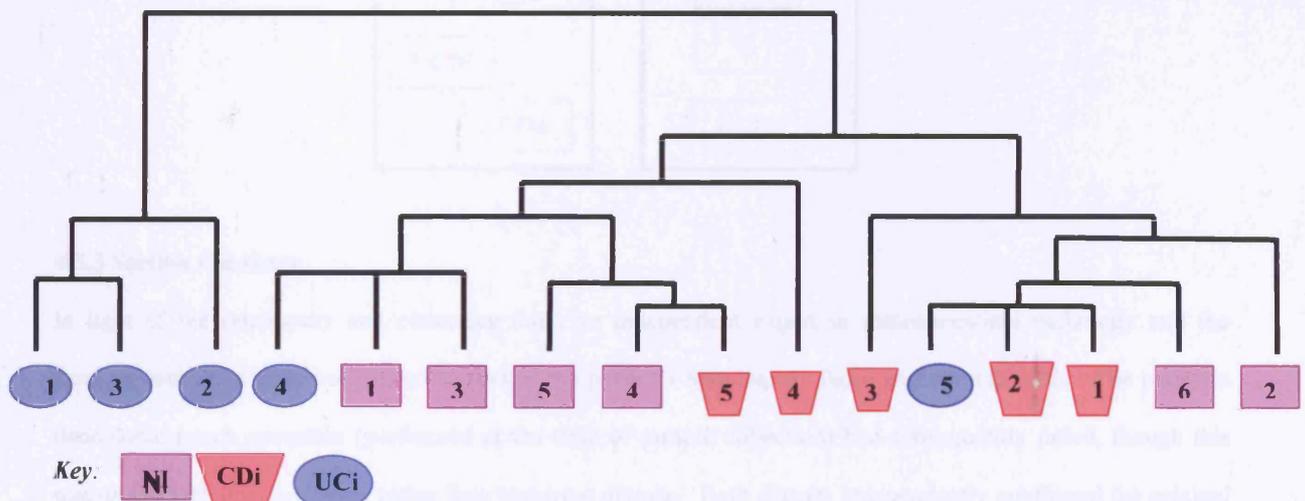
To provide further evidence of the abnormal gene expression pattern seen in sample UCi\_5 with the microarray data, a number of clustering techniques were applied to the data.

#### 4.3.2 Method & Results

##### 4.3.2.1 Hierarchical clustering

The samples were clustered by array using Stanford's Cluster program into a hierarchical tree to show similarity between the arrays. The Cluster program cannot deal with data sets that are too large, so the HuFL, Hu35KsubA, Hu35KsubB and Hu35KsubC data set was 'trimmed' to exclude genes that were scored absent in all the samples. This resulted in 23,131 genes remaining, i.e. the arrays were matched for similarity across 23,131 dimensions. Figure 4.3 shows the resulting hierarchical tree. It clearly shows the association of the UCi\_5 sample with the CDi samples. A second UCi sample, UCi\_4 also clustered away from the UCi samples, clustering with the control samples.

Figure 4.3 – Hierarchical clustering analysis

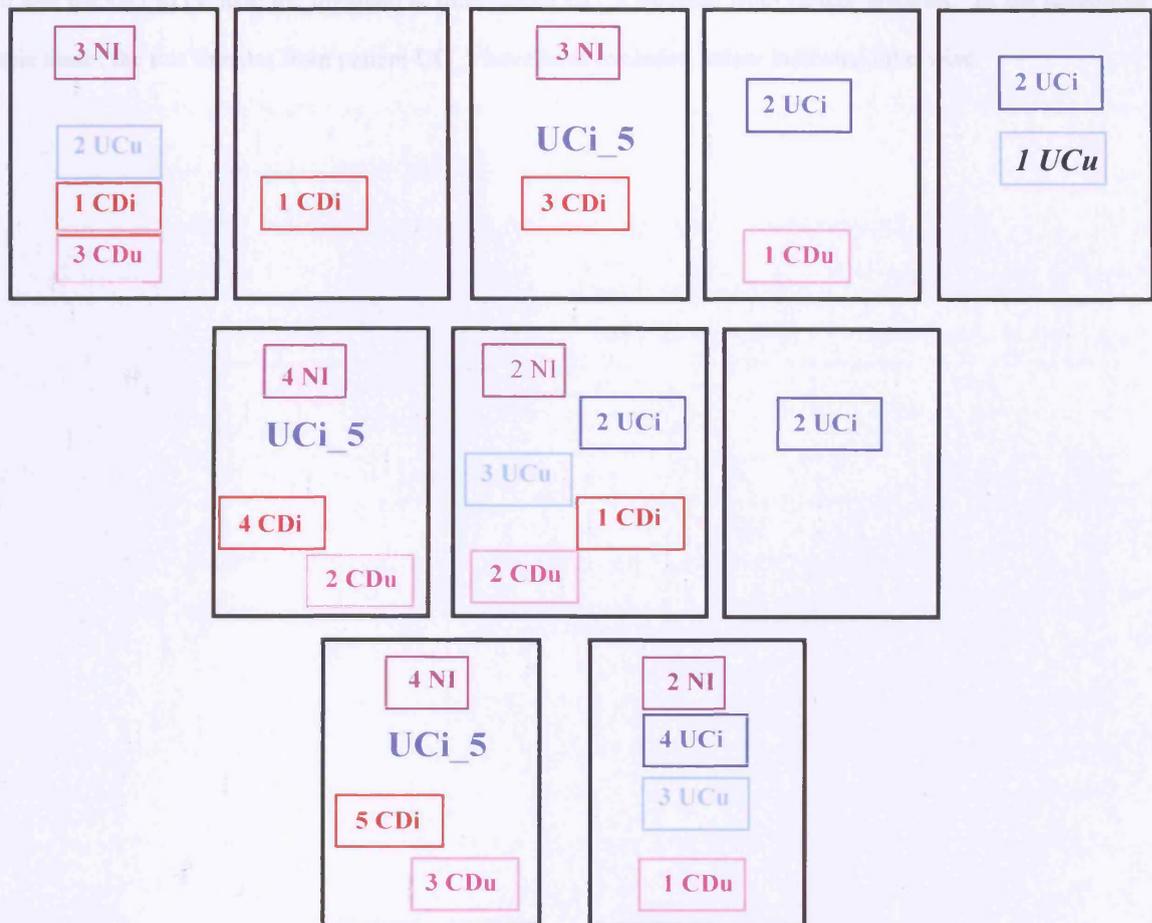


The beginning of each branch represents a divergence in similarity between arrays. The further apart two arrays are along the tree, the less the similarity between them and vice versa. The clustering of the involved sample from patient UC\_5 with the involved Crohn's disease samples is evident.

##### 4.3.2.2 K-means clustering

To confirm the outcome from the hierarchical clustering, a K-means cluster analysis was performed using the Stanford software. 5 nodes were specified initially, as there are five pathologically distinguishable tissue groups. Decreasing the number of nodes from 5 to 2 (figure 4.4) clearly showed the clustering of the involved UCi\_5 sample with involved Crohn's disease samples.

Figure 4.4 – K-means clustering of data into decreasing numbers of nodes



#### 4.3.3 Section summary

In light of the microarray and clustering data, an independent expert in gastrointestinal pathology and the surgical consultant were both asked to review the patient's histological slides and clinical notes. The patient's ileocolonic pouch operation (performed at the time of sample collection) had subsequently failed, though this was due to technical problems rather than recurrent disease. Both experts independently confirmed the original diagnosis of ulcerative colitis.

In the past, there have been theories that the two diseases represent a singular entity, with 'typical' Crohn's disease and ulcerative colitis at the opposite ends of a continuous scale. It is possible that patient UC\_5 represents an intermediary disease state between true Crohn's and true ulcerative colitis. Unfortunately, as the author only came across one such sample during the course of the study it is not possible to draw any conclusions. It would be interesting to compare the gene expression profiles of 'indeterminate colitis' samples that are subsequently re-classified as either CD or UC with the UC\_5 samples; however, this is beyond the scope of the present study.

Since the main aim of the present study was to investigate differences in gene expression between CDi and UCi, it was decided to exclude the involved & uninvolved UC\_5 samples from further analysis. In the remainder of this thesis, the two samples from patient UC\_5 have been excluded unless indicated otherwise.

#### 4.4 Which genes are differentially expressed in the involved IBD samples excluding patient UC\_5?

##### 4.4.1 Rationale

Having excluded the UC\_5 samples from the data set, the initial query was re-run. As detailed in section 4.2, the T-test and absolute call values were used to identify genes that showed differential expression between the two involved IBD tissue groups (tables 4.4 and 4.5).

**Table 4.4– Criteria for the query ‘Over expressed in CDi compared to UCi’**

<i>Criteria</i>	<i>Genes returned by individual query</i>	<i>Genes returned in cumulative query</i>
CDi vs. UCi T-test column = P<0.05	1718	1718
Absent in ≥ 3 UCi samples	24429	1216
Present in ≥ 4 CDi samples	9671	1143

**Table 4.5 – Criteria for the query ‘Over expressed in UCi compared to CDi’**

<i>Criteria</i>	<i>Genes returned by individual query</i>	<i>Genes returned in cumulative query</i>
CDi vs. UCi T-test column = P<0.05	1718	1718
Present in ≥ 3 UCi samples	5669	256
Absent in ≥ 4 CDi samples	18370	5

##### 4.4.2 Method

49 genes that had the words ‘kinase’ or ‘phosphatase’ in the gene description were identified from the 1143 genes over expressed in CDi. This list of 49 genes was refined further, by calculating the fold change between the mean average differences of the CDi samples and the NI samples. 10 genes that had a fold change of 5 or more are shown in table 4.6.

**Table 4.6 – Kinase / phosphatase genes over expressed in CDi**

<i>Acc #</i>	<i>Incyte Description May 2000</i>	<i>Fold Change*</i>	
		<i>UCi</i>	<i>CDi</i>
RC_AA417569_f_at	Human casein kinase I gamma 3L (CSNK1G3L) mRNA, complete cds.	2.1	12.9
RC_AA417569_i_at	Human casein kinase I gamma 3L (CSNK1G3L) mRNA, complete cds.	1.4	8.8
RC_AA621500_at	Human mRNA for 6-phosphofructo-2-kinase.	A in UCi	8.7
RC_AA431502_at	Human lok mRNA for protein kinase, complete cds.	-1.4	7.1
RC_AA358109_at	Human mRNA for AMP-activated protein kinase alpha-1, complete cds.	A in UCi	6.7
RC_AA452584_at	Human mRNA for inhibitor 2 of protein phosphatase 1.	-1.4	6.5
U68111_at	Human mRNA for inhibitor 2 of protein phosphatase 1.	1.2	5.8
U89896_at	Human casein kinase I gamma 2 mRNA, complete cds.	A in UCi	5.8
D25328_at	Human mRNA for platelet-type phosphofructokinase, complete cds.	-1.6	5.7
D15049_at	Human mRNA for protein tyrosine phosphatase precursor, complete cds.	A in UCi	5.0

\* Fold change cannot be calculated if a gene has been called absent in all the samples of one tissue group and these are indicated in the table.

The 'casein kinase I gamma 3L' and 'inhibitor 2 of protein phosphatase 1' genes are represented in the table more than once. In the case of the protein phosphatase 1 inhibitor, the first entry corresponds to an EST sequence, whilst the second entry represents the full-length gene. The casein kinase gene is represented twice on the Hu35K\_subA array.

Bioinformatic searches were carried out for the genes with the highest fold changes to determine whether these expression patterns had been observed in IBD tissues previously. To illustrate, the process is detailed for the first gene in table 4.6, human casein kinase I gamma 3L.

The accession numbers given in table 4.6 for casein kinase I $\gamma$ 3L (CSNK1G3L) correspond to an EST. A BLAST search was performed and the EST showed 100% identity to a region on chromosome 5. The name of the gene had been linked to the ESTs by AstraZeneca's bioinformatics department using a proprietary database. However, as the location of CSNK1G3L was confirmed as chromosome 5<sup>212</sup>, the gene identity was considered accurate. CSNK1G3 is part of the CKI $\gamma$  subfamily and is one of the many forms of this family found in mammals<sup>213</sup>. Casein kinases in general have not previously been linked to IBD<sup>214</sup> and besides general casein kinase information there was no CSNK1G3 specific functional information to indicate why this gene should be over expressed in CDi<sup>212</sup>.

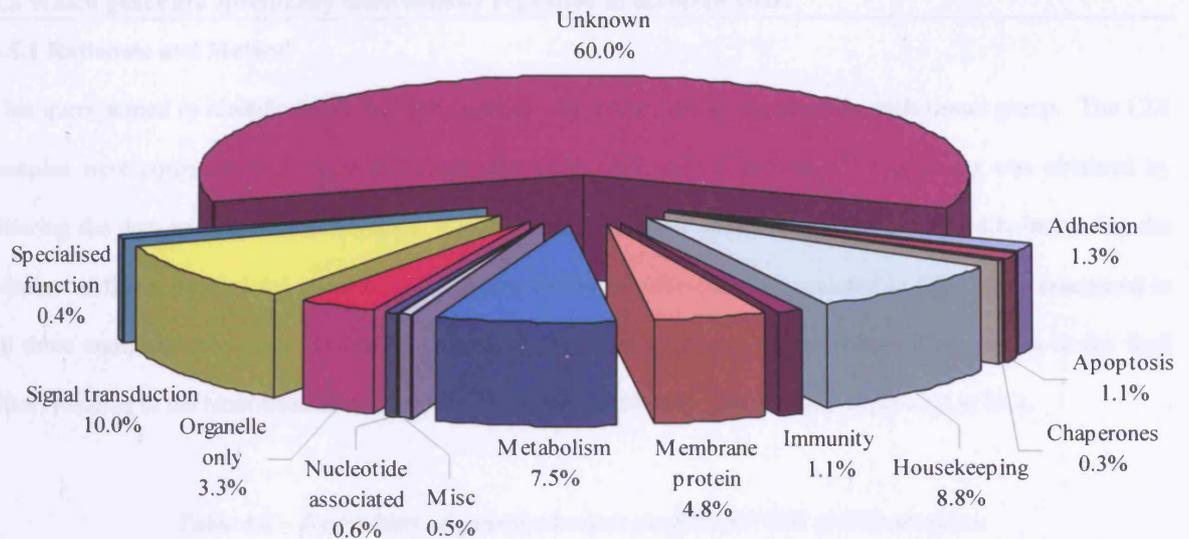
For all the subsequent queries in this thesis, the same method was followed on genes showing interesting expression patterns, to identify genes of particular interest. From this point, only genes of particular interest are discussed. BLAST searches were carried out on all ESTs showing an interesting expression pattern, but are only reported here if the BLAST search yielded something of significance.

#### 4.4.3 Over expressed in CDi

Based on the Incyte functional hierarchy, the 1,143 genes over expressed in the CDi samples were grouped into 13 groups in an attempt to summarise the data (figure 4.5). The functional hierarchy is not complete however and many of the genes remain wholly or partially uncharacterised.

The 'unknown' group contains EST sequences and full-length genes for which no function or intracellular location is known. The 'organelle only' group contains those genes for which only intracellular location of the protein is known, but a definitive function is not. The 'miscellaneous' group contains six genes that do not fall into any of the other groups. Eliminating the 'unknown' and 'location only' genes left 419 genes to consider; however this was still too many for manual annotation.

Figure 4.5 – Generalised protein function and pathway groups for CDi over expressed genes



#### 4.4.4 Over expressed in UCI

As only five genes were found to be over expressed in UCI, these are all detailed in table 4.7. The baseline for the fold change value was the mean average difference in the NI samples.

Table 4.7 – UCI over expressed genes

Acc #	Incyte Description May 2000	Fold Change	
		UCi	CDi
RC_AA425937_at	Incyte Unique	11.2	A in CDi
M34516_at	Human Ig-related 14.1 protein mRNA, complete cds.	6.0	-2.0
AA476967_at	Human CpG island DNA genomic MseI fragment, clone 151a12	2.9	-3.9
RC_R45355_at	Human mRNA; cDNA DKFZp434E202 (from clone DKFZp434E202); partial cds.	1.7	A in CDi
RC_AA147056_at	Human mRNA; cDNA DKFZp434K1013 (from clone DKFZp434K1013).	1.3	-6.5

The over expression of the immunoglobulin related protein (M34516\_at) is not surprising, as it is involved in B cell development and the UCI samples contained a high number of B-lymphocytes. When this gene was normalised to the number of B lymphocytes in each sample, the expression difference between the CDi and UCI samples was no longer significant. This gene was also noted to be increased in UC tissues in a previous microarray experiment<sup>181</sup>. All the other accession numbers in table 4.7 represent ESTs and BLAST searches for each of them yielded no significant information.

#### 4.4.5 Section summary

The 'over expressed in CDi' gene set was too large to mine manually and the over expressed in UCI too small. Also, the query posed does not take into account the NI and uninvolved IBD samples. Therefore a new query was formulated to take account of all the available samples.

## 4.5 Which genes are specifically differentially regulated in involved IBD?

### 4.5.1 Rationale and Method

This query aimed to identify genes that were specifically differentially regulated in each tissue group. The CDi samples were compared to 3 types of tissues, the CDu, UCi and NI groups. The gene set was obtained by filtering the data with the T-test columns, at a confidence level of 95%. As shown in table 4.8, increasing the number of filters enabled the selection of 36 genes that were differentially regulated in CDi, when compared to all three comparative groups. A similar approach for the UCi samples (substituting UCu samples in the final filter) resulted in the identification of 34 genes that were specifically differentially expressed in UCi.

*Table 4.8 – Generation of gene expression profiles for CDi & UCi samples*

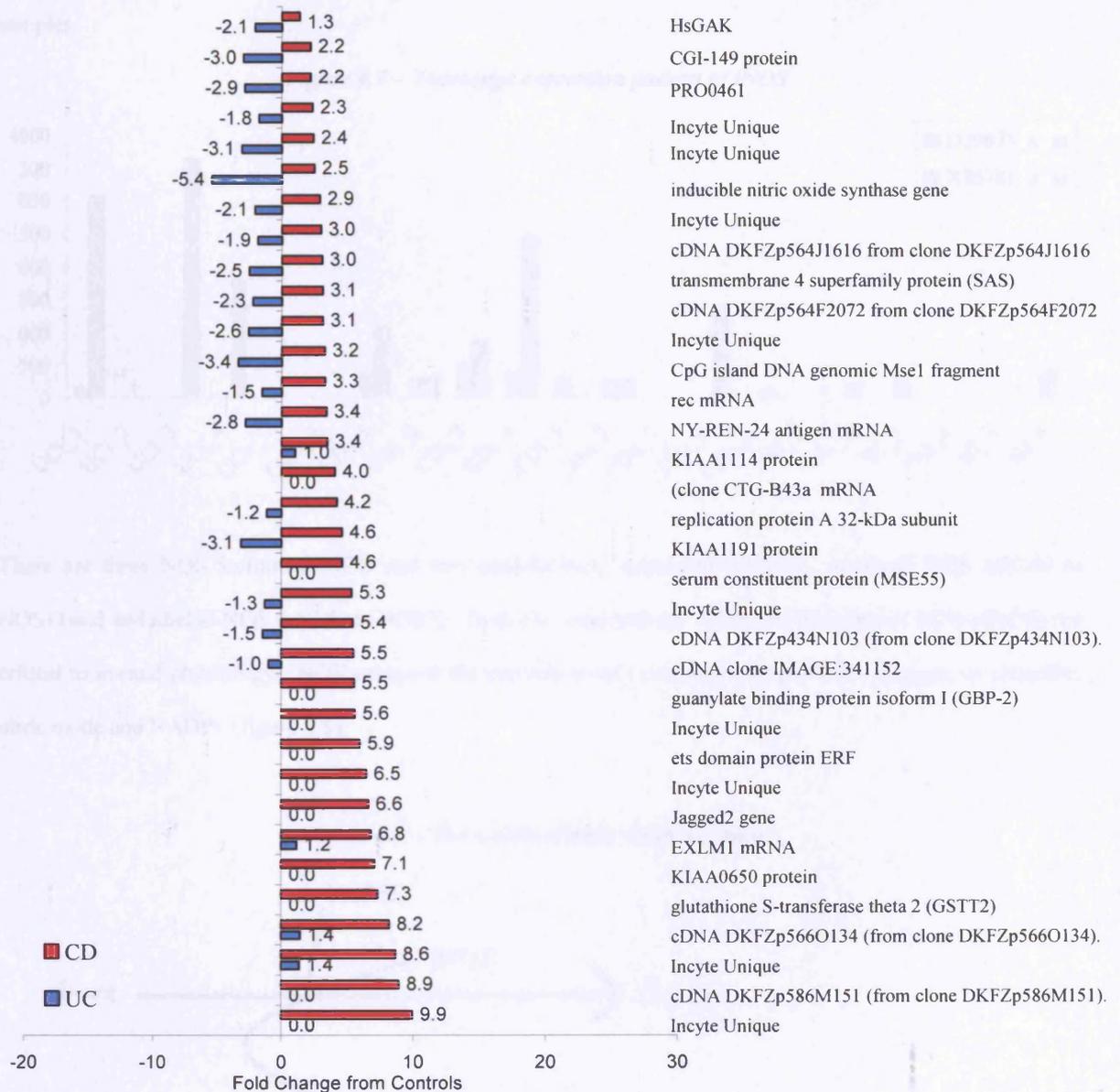
<i>Gene expression profile for:</i>	<i>Comparative group for T-test (P&lt;0.05)</i>	<i>Genes returned</i>	<i>Cumulative number of genes returned</i>
CDi	CDi vs. UCi	1718	1718
	CDi vs. Control	1369	585
	CDi vs. CDu	393	<b>36</b>
UCi	UCi vs. CDi	1718	1718
	UCi vs. Control	1012	329
	UCi vs. UCu	792	<b>34</b>

### 4.5.2 CDi specific genes

The 36 genes specifically differentially expressed in CDi are listed in appendix A4.2, with the full gene descriptions and accession number. Annotation of all 36 genes was attempted. However, the majority of these represent EST sequences. BLAST searches were carried out on all ESTs, but in 15 cases no full-length gene could be associated with the sequence.

All 36 genes are upregulated in the CDi samples compared to the NI samples. Some of the genes show similar expression levels in the UCi and NI samples, whilst others show an opposing expression pattern in CDi and UCi. The fold change of the genes from NI in the CDi and UCi samples is shown in figure 4.6. The remainder of this section details the results of the annotation effort.

Figure 4.6 – CDi specific differentially expressed genes shown by fold change from NI samples



#### 4.5.2.1 Opposing expression patterns in CDi and UCi

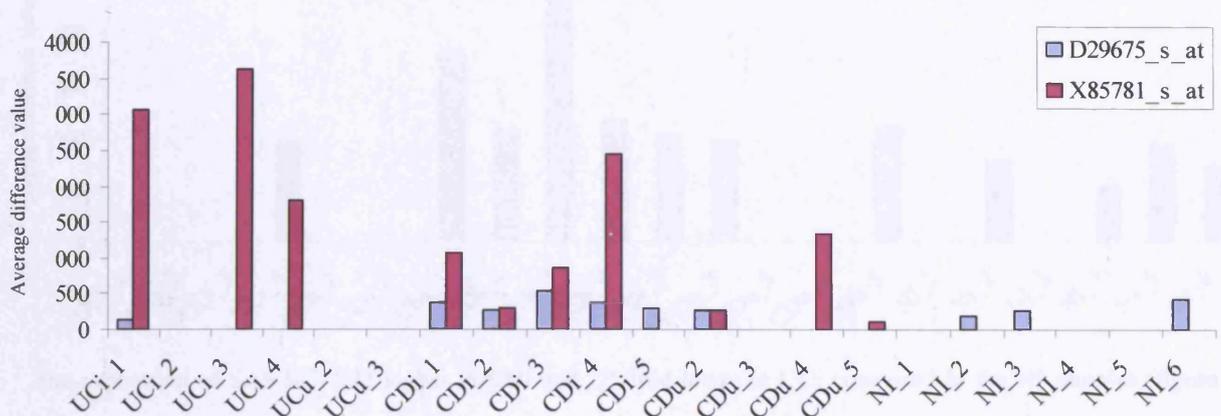
Eleven of the 36 CDi specific genes showed at least a 2-fold change from the NI samples, in opposite directions in CDi and UCi. All eleven show up regulation in CDi and down regulation in UCi compared to the NI samples. Nine of the probe sets represent an EST and only two of these genes could be annotated.

##### *Nitric oxide synthase (D29675\_s\_at)*

This probe set represents the 'Human inducible nitric oxide synthase gene, promoter and exon 1' according to the EMBL database. The over expression of inducible nitric oxide synthase (iNOS or NOS2) in IBD is known<sup>176, 177</sup>. However, the expression level of iNOS in involved ulcerative colitis is lower than the level in the NI samples (figure 4.7), which is contradictory to published reports. A second probe set to iNOS (X85781\_s\_at; EMBL entry '*H. sapiens* iNOS gene, exon 27') on the same array (HuFL), did follow the

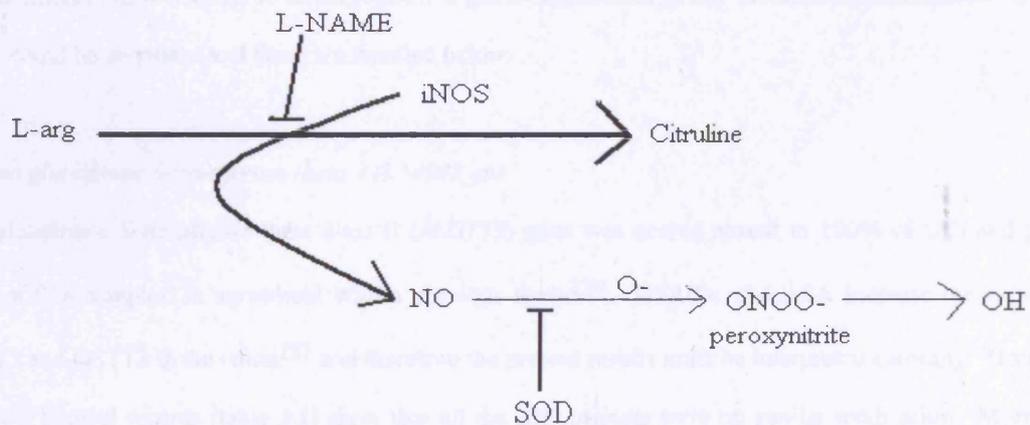
expected expression pattern, being scored absent in 100% of the controls and present in 75% of the involved UC samples.

**Figure 4.7 – Transcript expression pattern of iNOS**



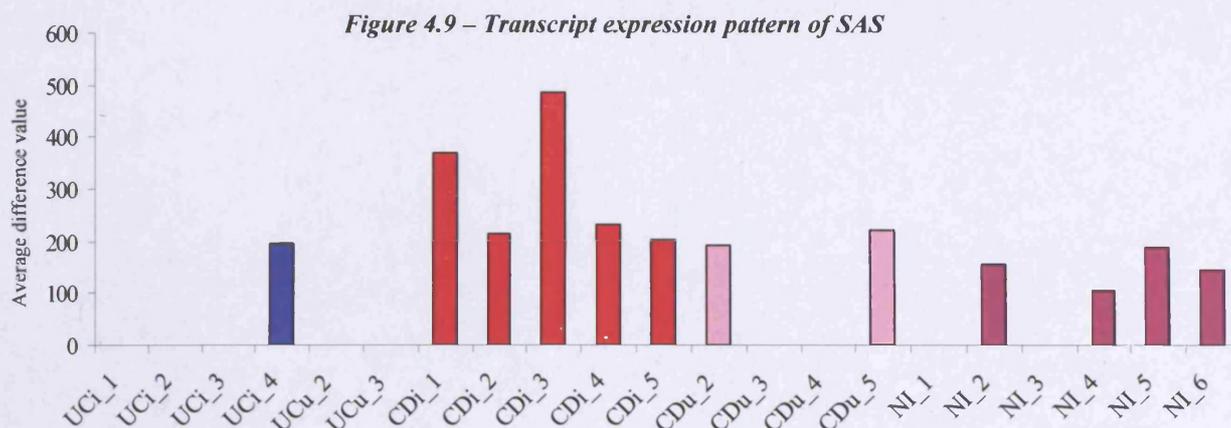
There are three NOS isoforms, iNOS and two constitutively expressed isoforms, neuronal NOS (nNOS or NOS1) and endothelial NOS (eNOS or NOS3). Both the constitutively expressed isoforms of NOS (cNOS) are critical to normal physiology. NOS catalyses the conversion of L-arginine, NADPH and oxygen, to citrulline, nitric oxide and NADP<sup>+</sup> (figure 4.8).

**Figure 4.8 – The action of nitric oxide synthase**



The present data indicates a higher expression level of iNOS in involved Crohn's disease compared to ulcerative colitis; a previous study noted a higher expression of cNOS in Crohn's disease compared to ulcerative colitis<sup>215</sup>. The over expression of NOS is likely to be a secondary feature of inflammation, as increased NO production has been noted in other inflammatory diseases<sup>216</sup>. However, an over expression of iNOS may not always result in mucosal damage. In some animal models of IBD the inhibition of iNOS increases injury, whilst in others iNOS inhibition is beneficial (reviewed in Kubes & McCafferty, 2000<sup>217</sup>).

*Transmembrane 4 superfamily protein (SAS) mRNA (U01160\_at)*



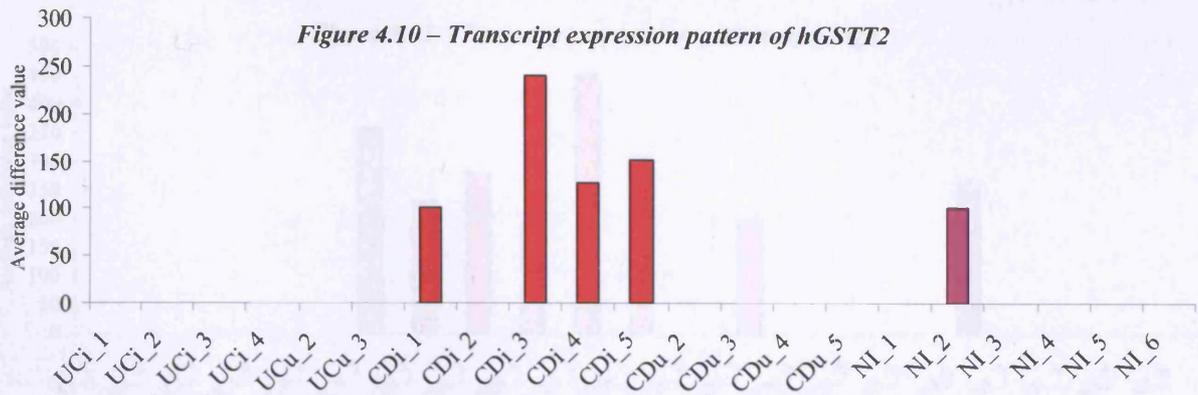
The expression of SAS is 3 fold higher in CDi and 2.5 fold lower in UCI compared to the NI samples (figure 4.9). Located on chromosome 12q13-14, SAS (sarcoma amplified sequence) is part of the tetraspanin family of integral membrane proteins. As the name suggests, it was first isolated from human sarcoma tissue<sup>218</sup>. However, apart from the fact that it has four hydrophobic membrane spanning domains<sup>219</sup>, no specific functional information has been published. The gene has not previously been linked to IBD.

#### 4.5.2.2 Highly over expressed CDi genes

15 genes showed a 5 fold or more increase in expression in CDi compared to NI. Over expressed genes are of special interest, as it is easier to down regulate a genes' expression in any therapeutic intervention. 6 of the 15 genes could be annotated and these are detailed below.

#### *Human glutathione S-transferase theta 2 (L38503\_at)*

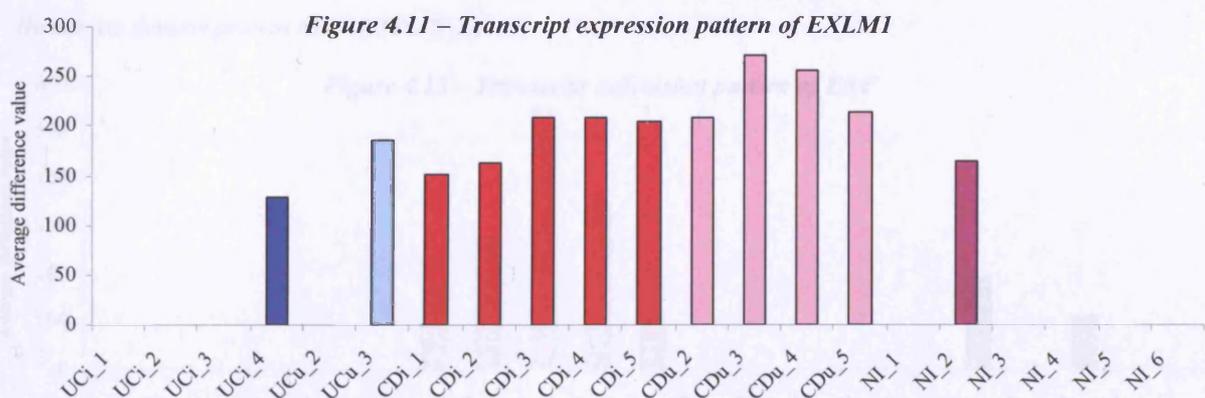
The glutathione S-transferase theta class II (*hGSTT2*) gene was scored absent in 100% of UCI and present in 80% of CDi samples, in agreement with a previous study<sup>220</sup>. NSAIDs and ASA increase the expression of GSTT1 and GSTT2 in the colon<sup>221</sup> and therefore the present results must be interpreted carefully. However, the patient's clinical records (table 2.1) show that all the IBD patients were on similar medication. Moreover, the current study also investigated uninvolved IBD tissues. If the over expression of *hGSTT2* in CDi were due to differences in medication between UC and CD patients, then a similar level of over expression would be expected in the CDu tissues, as both the involved and uninvolved samples were taken from the same patients at the same time. This was not seen and GSTT2 was scored absent in 100% of the CDu samples (figure 4.10).



The glutathione S-transferases (GSTs) are large family of detoxification enzymes and the theta GSTs are one of seven classes (reviewed by Landi, 2000<sup>222</sup>). There are two theta GSTs, (GSTT1 and GSTT2) with a 55% amino acid sequence homology. The GST theta class conjugate reduced glutathione to a wide number of exogenous and endogenous hydrophobic electrophiles and show sulphatase activity<sup>223</sup>. A lack of expression of hGSTT1 in null *hGSTT1* allele homozygotes increases the risk of bladder, gastrointestinal and tobacco related tumours<sup>222</sup>. Expression of GSTT2 is also believed to confer protection against colonic carcinoma<sup>224</sup>. The higher risk of colorectal carcinoma in ulcerative colitis as compared to Crohn's disease could be related to the current observation of *GSTT2* under-expression in ulcerative colitis.

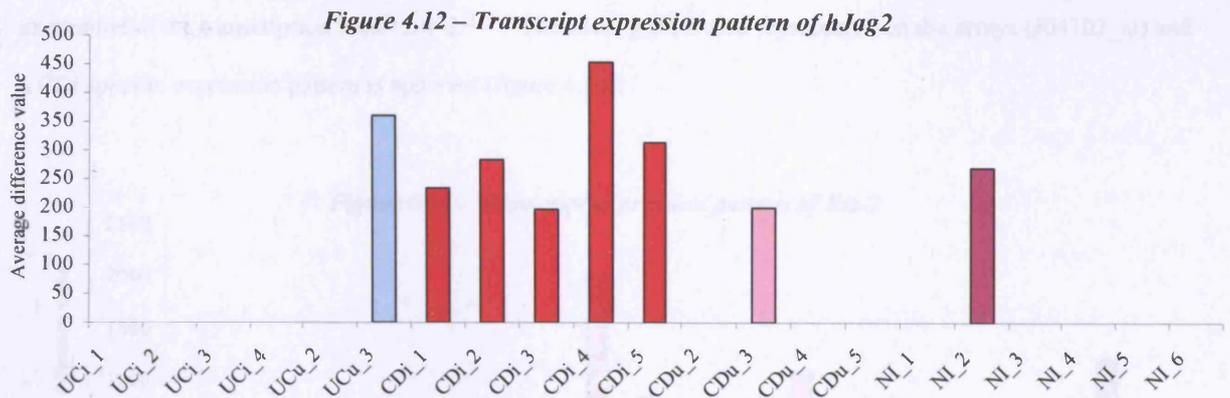
#### Human *EXLM1* mRNA (RC\_AA460683\_at)

*EXLM1* was expressed in involved and uninvolved Crohn's disease samples (figure 4.11) and is one of many co-factors required for activation of the transcription factor SP-1<sup>225</sup>. SP-1 is part of the transcription machinery and thus enhances the transcription of many genes. *EXLM1* has a number of aliases (CRSP2, TRAP170, CSRP, RGR1, CXORF4, CRSP150), but none have previously been linked to IBD. It remains unclear as to why this gene is specifically over expressed in the Crohn's disease samples.



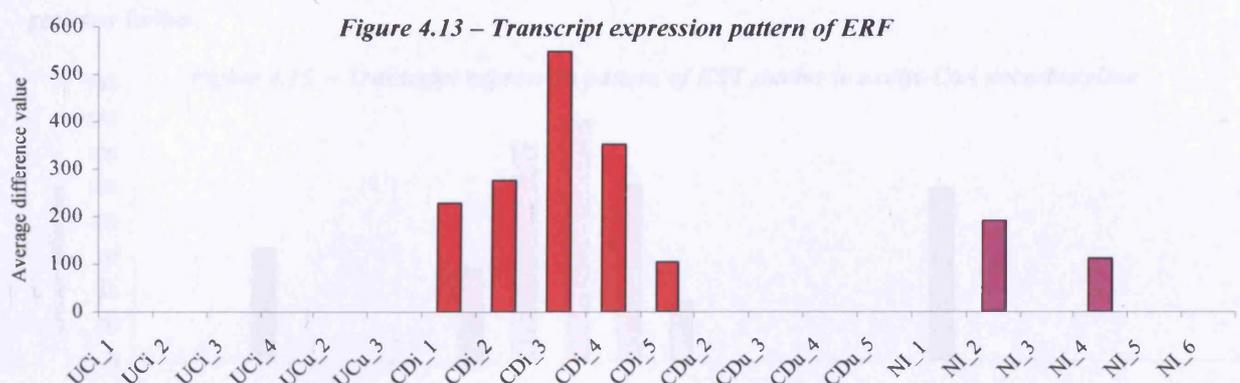
#### Human *Jagged 2* (W04732\_at)

The *Jagged 2* gene (*hJag2*) was over expressed in the CDi samples by a 6.6 fold compared to the NI samples. It was scored absent in 100% of the UCi samples (figure 4.12).



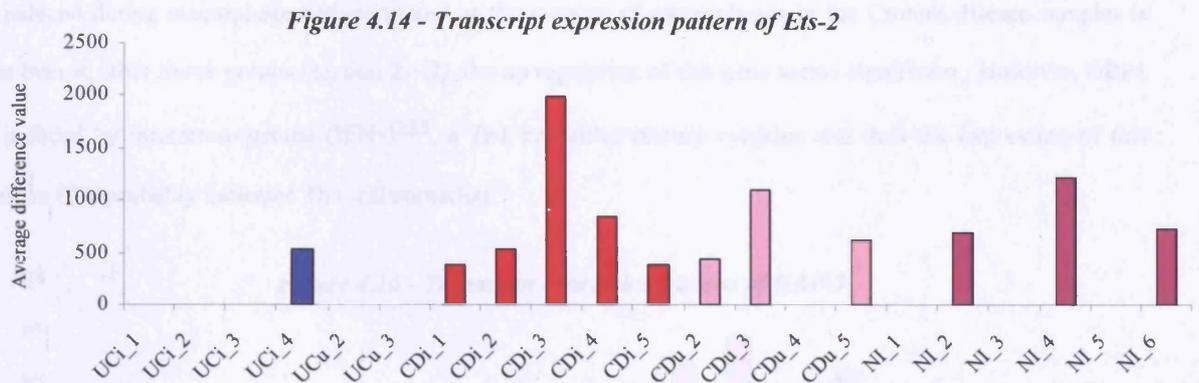
*hJag2* is localised to chromosome 14q32<sup>226</sup> and its promoter region has potential binding sites for a number of transcription factors, including NF- $\kappa$ B, Ets-1 and OCT-1<sup>227</sup>. *Jag2* is a ligand of the Notch1 receptor. The Notch receptor family and their ligands have been studied in *Drosophila* in developmental biology. *hJag2* and *hJag1* are the human homologues of the *Drosophila* gene *Serrate*. The Notch signalling pathway is critical in many developmental pathways and is well conserved in evolution with 89% amino acid homology between rat and human *Jag2*. The role of *Jag2* *in vivo* has been investigated by removing the receptor domain of the *Jag2* gene. Mice that were homozygous for this showed a number of abnormalities, including impaired differentiation in the development of  $\gamma\delta$  T cells<sup>228</sup>. The expression of Notch pathway components, including *Jag2*, is upregulated in injured rat vascular cells<sup>229</sup> and the expression of full length *Jagged2* by endothelial cells promotes the proliferation of haematopoietic progenitors *in vitro*, when the cells are in direct contact<sup>230</sup>. The over expression of *Jag2* in CDi and its absence in almost all the UC samples may be indicative of differential inflammatory processes in the two diseases. In terms of pathogenesis, it is more likely that *Jag2* is a secondary inflammation mediator.

*Human ets domain protein ERF (U15655\_at)*



The over expression of the Ets-2 repressor factor (*ERF*) is specific to CDi compared to UCi, CDu and NI samples (figure 4.13). Localised to chromosome 19q13.1, *ERF* is a transcriptional repressor which inhibits the

expression of the transcription factor *Ets-2*<sup>231</sup>. The *Ets-2* gene is also represented on the arrays (J04102\_at) and a CDi specific expression pattern is apparent (figure 4.14).

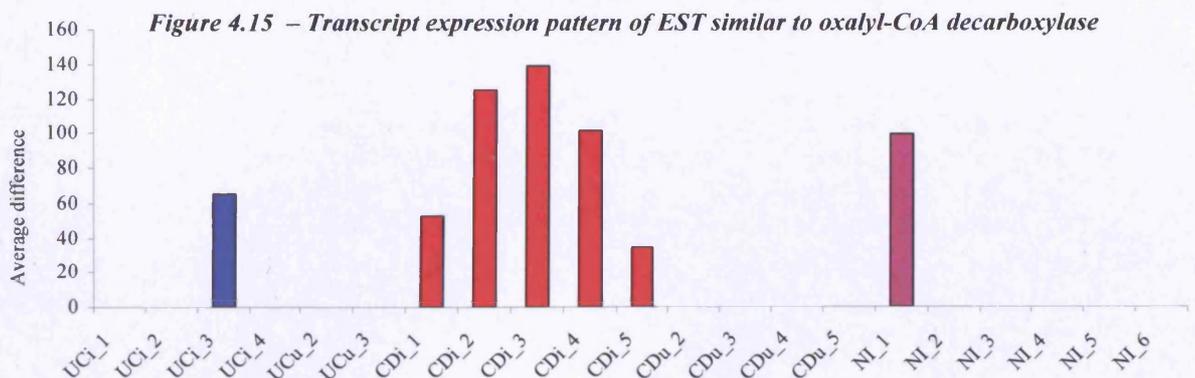


*Ets-2* is a proto-oncogene and its translocation from chromosome 21 to chromosome 8 has been implicated in some forms of acute myeloid leukaemia<sup>232</sup>. Increased *Ets-2* gene dosage, as seen in Down's syndrome, results in skeletal abnormalities in mice<sup>233</sup>.

The over expression of *ERF* in the CDi samples is likely to have been induced by the increased expression of the *Ets-2* gene. Neither *ERF* nor *Ets-2* have been linked to IBD previously. However, over expression of *ERF* has been noted in rheumatoid arthritis<sup>234</sup>. Like Crohn's disease, rheumatoid arthritis is a Th1 mediated disease, so it may be that the over-expression of *ERF* noted here is a feature of Th1 mediated inflammation and may prove to be a marker of Th1 diseases in general.

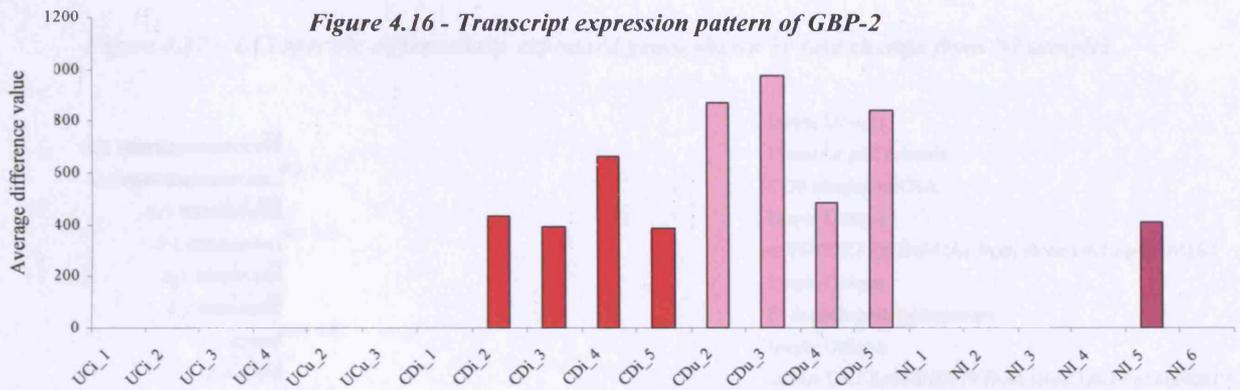
*cDNA clone IMAGE:341152 (RC\_W58611\_at)*

According to the EMBL database, this EST is similar to oxalyl-CoA decarboxylase. This gene does show CDi specific up regulation, however, the values are not very high (figure 4.15) and it was decided not to annotate this gene any further.



*Interferon-induced guanylate-binding protein 1 (R39374\_at)*

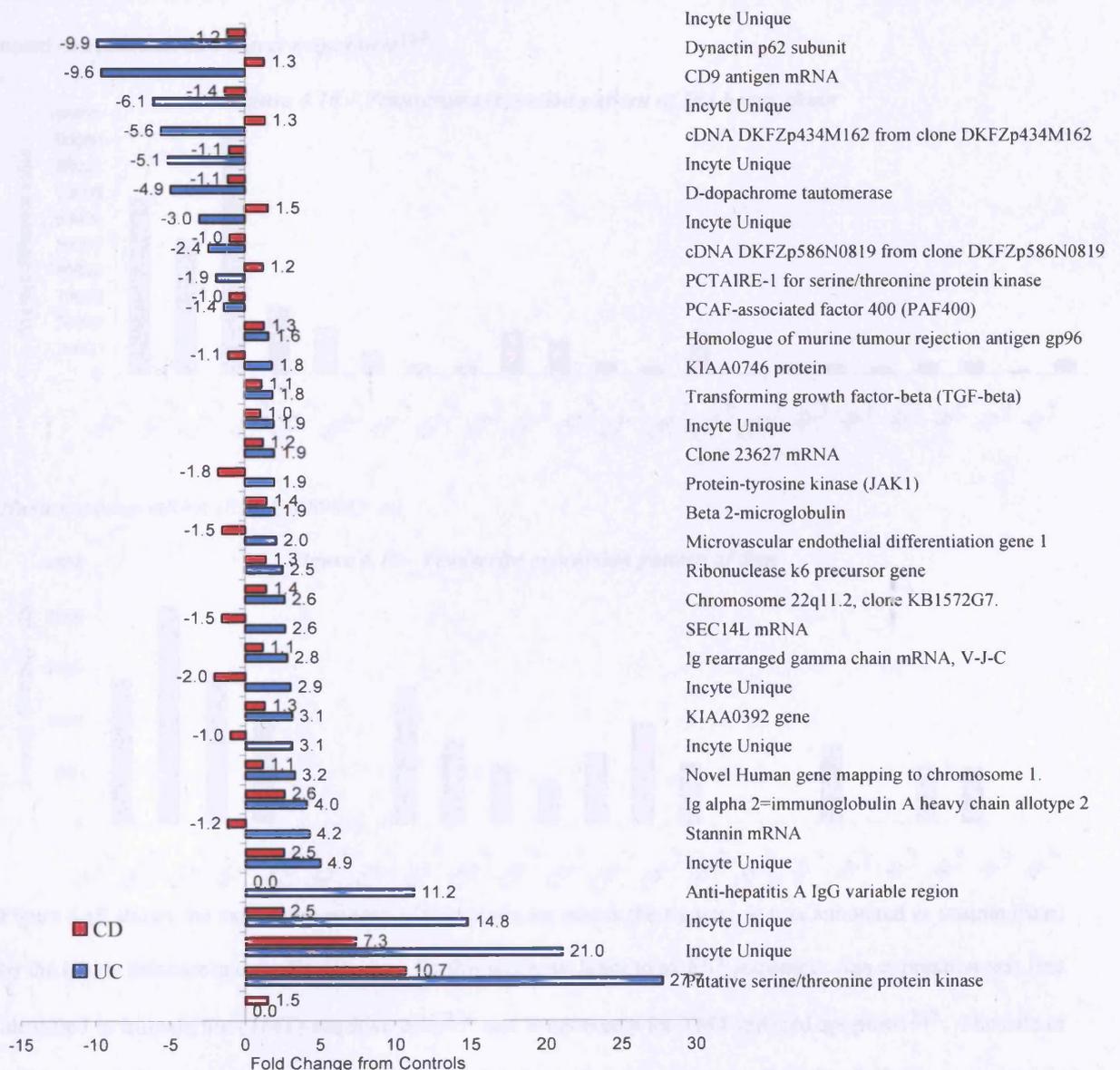
This gene (*GBP1*) is expressed almost exclusively in the Crohn's disease samples at a high level (figure 4.16). It is induced during macrophage activation and as the number of macrophages in the Crohn's disease samples is less than in other tissue groups (section 2.3.7), the up regulation of this gene seems significant. However, *GBP1* is induced by interferon-gamma ( $IFN\gamma$ )<sup>235</sup>, a Th1 pro-inflammatory cytokine and thus the expression of this gene in CDi probably indicates Th1 inflammation.



4.5.3 UCi specific genes

The 34 genes specifically differentially expressed in UCi are listed in appendix A4.3, with the full gene descriptions and accession number. As with the CDi specific genes, annotation of all 34 genes was attempted. A full-length gene could not be associated with 18 of the EST sequences. 10 genes showed a decreased expression level in the UCi samples compared to the NI samples, with the remaining 24 genes showing an increase in expression level (figure 4.17). As with the CDi specific genes, the UCi specific genes were annotated in different groups based on their expression patterns.

Figure 4.17 – UCi specific differentially expressed genes shown by fold change from NI samples

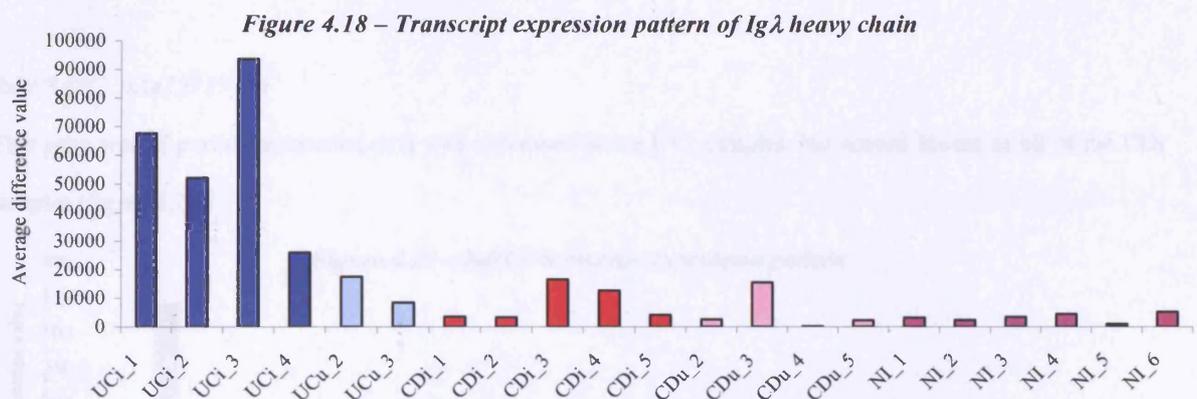


#### 4.5.3.1 Similar expression patterns in UCI and CDi samples

None of the UCI specific genes showed an increase of more than 2 fold in opposite directions in the CDi and UCI samples. However, 5 genes in the UCI specific gene set showed a greater than 2-fold change in the both the IBDi samples sets. Two of these could be associated with a full-length gene and were annotated.

##### Human mRNA for Ig lambda heavy chain. (M87789\_s\_at)

The over expression of immunoglobulin genes in the involved IBD tissues compared to the NI tissues and the specific over expression in UCI compared to CDi was expected. The number of B cells in the UCI tissues was higher than in other tissues (section 2.3.7) and is reflected by the expression pattern of this immunoglobulin gene (figure 4.18). The expression pattern for this gene, i.e. an 8-fold increase in UC compared to controls, was also noted in a previous microarray experiment<sup>180</sup>.



##### Human stannin mRNA (RC\_AA489045\_at)

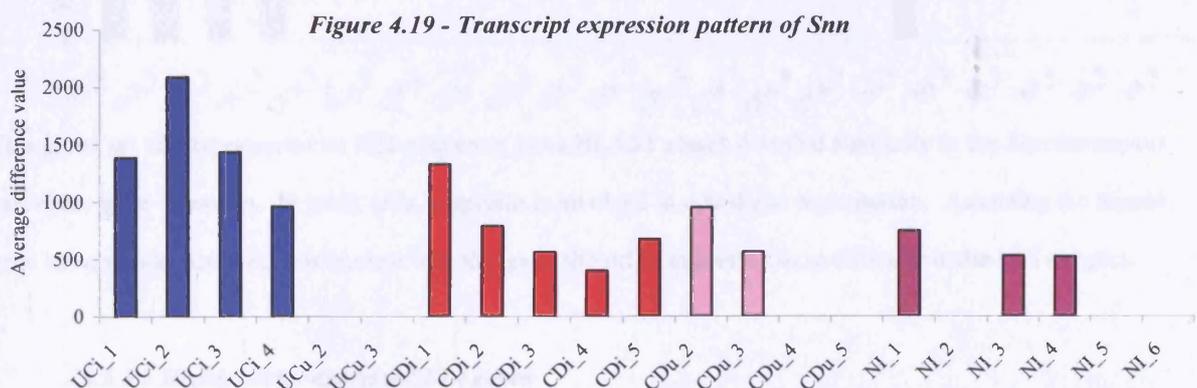


Figure 4.19 shows the expression pattern of this probe set across the tissues. It was annotated as stannin (Snn) by the Incyte database and the EMBL entry for this probe set leads to an EST sequence. *Snn* expression was first identified in trimethyltin (TMT) sensitive cells<sup>236</sup> and is necessary for TMT induced apoptosis<sup>237</sup>. The role of Snn in organotin toxicity has therefore been studied, but no other functions are alluded to in the literature and its expression has not previously been studied in IBD tissues. However, a BLAST search revealed that the EST sequence of 440 bp length showed 100% identity to human *Snn* along a length of only 59bp. The human stannin

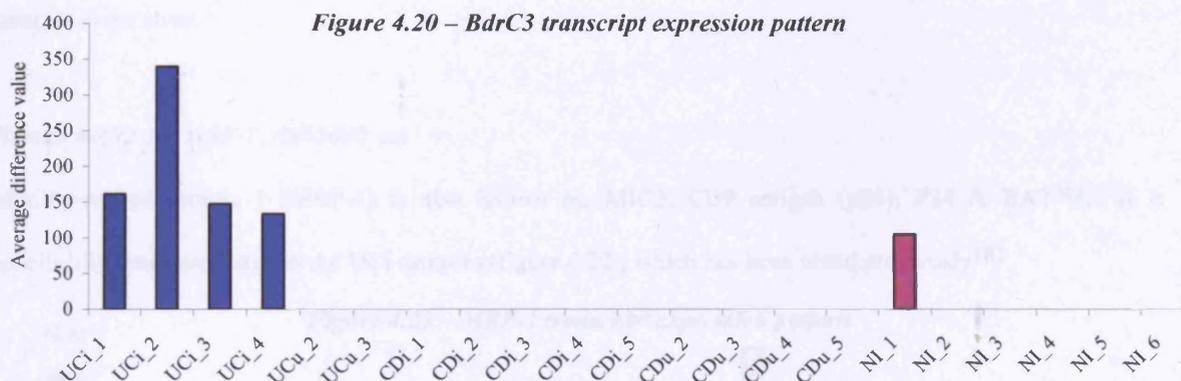
gene is 3295 bp long<sup>238</sup>. The best homology identified by the BLAST search was 99% identity of the entire EST to a patented sequence. From the title of the patent<sup>239</sup> it was ascertained that the gene is a thioredoxin family member, but there are no publications linking the family to IBD at the time of writing. As this probe set is over expressed in the involved IBD tissues specifically, it is likely that the represented gene plays a part in the inflammatory process.

#### 4.5.3.2 Highly over expressed UCi genes

4 genes showed an expression level of more than 5-fold compared to the NI samples. Two of these were also selected in the last query (section 4.5.3.1); an EST (RC\_N52440\_at) and the immunoglobulin gene (M87789\_s\_at). The third gene was also an EST (RC\_N69207\_at), but could not be annotated beyond the chromosome location. The remaining gene is described below.

##### *BdrC3* (RC\_AA425937\_at)

This gene was of particular interest as it was expressed in the UCi samples, but scored absent in all of the CDi samples (figure 4.20).



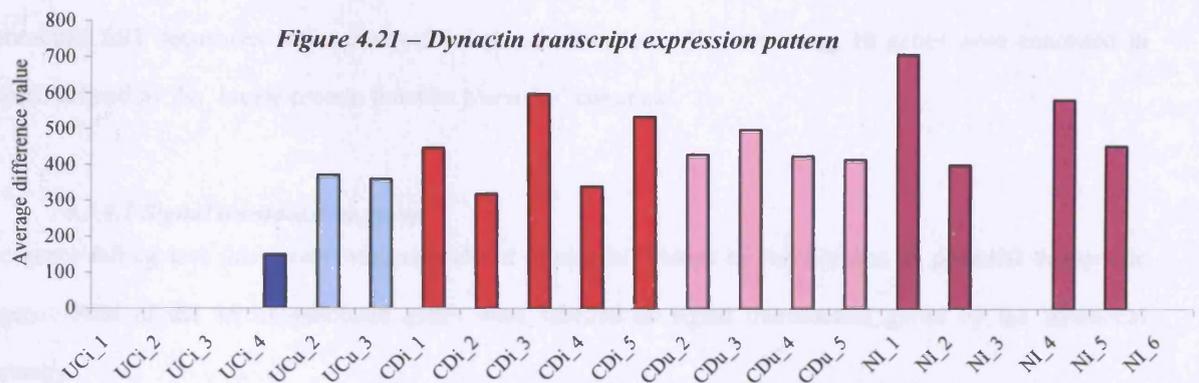
This probe set also represented an EST sequence, but a BLAST search revealed similarity to the *Saccharomyces cerevisiae* gene *verprolin*. In yeast cells, verprolin is involved in cytoskeletal organisation. Assuming the human gene has a similar function, it is unclear why this gene should be expressed so specifically in the UCi samples.

#### 4.5.3.3 Highly under expressed UCi genes

5 genes showed a down regulation of more than five fold compared to the NI samples. Two of these were EST sequences and no information other than chromosome location could be obtained by BLAST searching (RC\_H57166\_at & RC\_AA452245\_s\_at). Another of the genes was the product of a Japanese sequencing effort and represents a transcribed protein with no known function (AA401894\_at). The remaining two genes are described below.

*Dynactin p62 subunit mRNA (T34752\_s\_at)*

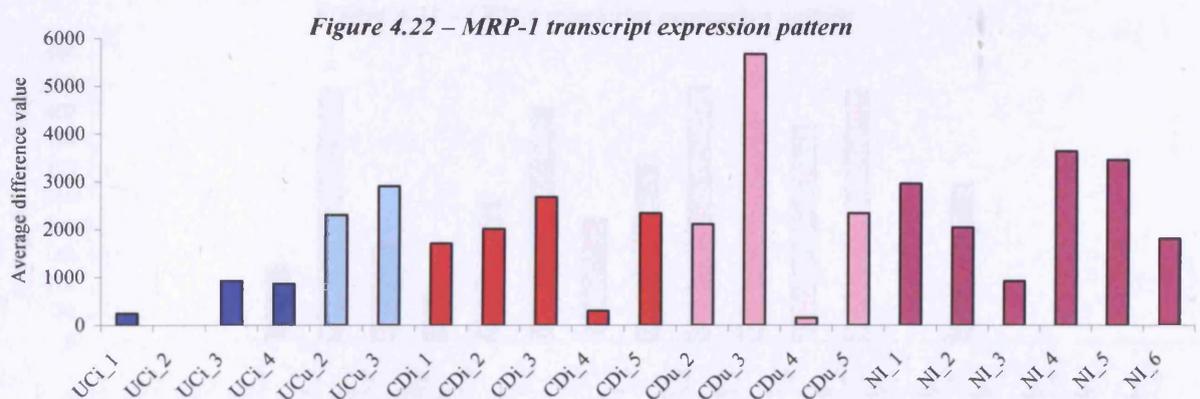
This gene is undetected in most of the UCI samples (figure 4.21). Dynactin is expressed by gut epithelial cells<sup>240</sup> and there tends to be fewer epithelial cells in UCI samples compared to the other samples (section 2.3.7). However, even after normalising the array data to the ICC data, the decrease of dynactin expression in UCI was still significant in comparison with UCU and CDi.



Dynactin is part of a microtubule-based motor complex and the yeast homologue has been subject to much investigation (reviewed by Karki & Holzbaur, 1999<sup>241</sup>). Why this gene should be down regulated in the UCI samples is not clear.

*Human mRNA for MRP-1. (M38690\_at)*

Motility-related protein 1 (MRP-1) is also known as; MIC3, CD9 antigen (p24), P24 & BA2<sup>242</sup>. It is specifically down regulated in the UCI samples (figure 4.22), which has been noted previously<sup>181</sup>.



The CD9 antigen is part of the tetraspannin membrane family. It is expressed by a variety of haematopoietic and epithelial cells<sup>243</sup>. A decreased expression (compared to controls) has been observed, or associated with a poor prognosis in a number of different cancers; endometrial<sup>244</sup>, oral squamous cell carcinoma<sup>245</sup>, diffuse non-Hodgkin's lymphoma<sup>246</sup>, pancreatic<sup>247</sup>, breast<sup>248</sup> and colon<sup>249</sup>. CD9 is thought to inhibit metastasis and the expression of CD9 in colorectal cancer cells *in vitro* inhibited tumour cell motility<sup>250</sup>. The potential value of

CD9 gene transfer therapy has also been investigated<sup>251</sup>. It is plausible that the decreased CD9 expression in the UCi samples may be associated with the increased risk of cancer in ulcerative colitis (section 1.1.4.4).

#### 4.5.4 Further annotation of involved IBD specific genes

This query represents the main focus of this thesis and annotation of 37 of the 70 CDi and UCi specific genes not discussed within the queries posed in sections 4.5.2 & 4.5.3, are detailed below. Of these, 21 probe sets represented EST sequences with no associated protein function. The remaining 16 genes were annotated in groups, defined by the 'Incyte protein function hierarchy' columns.

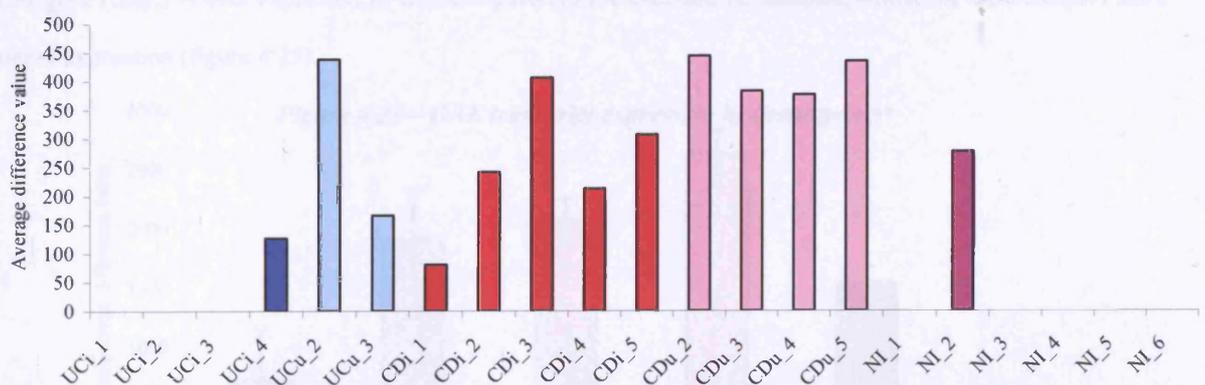
##### 4.5.4.1 Signal transduction group

The genes falling into this group were considered of special interest to AstraZeneca as potential therapeutic targets. Four of the 16 un-annotated genes were labelled as signal transduction genes by the functional hierarchy.

##### *cDNA DKFZp434N103 (RC\_AA446944\_at)*

Linking the EMBL entry of this EST to the UniGene database reveals that this EST represents the human homologue of the *Caenorhabditis elegans* gene, *CED-6*. In *C. elegans*, *CED-6* is a vital apoptotic protein, which is involved in the phagocytosis of apoptosed cells<sup>252</sup>. Human *CED-6* has been shown to have a similar role<sup>253</sup>. The over expression of *CED-6* in the Crohn's disease samples (figure 4.23), therefore implies an up regulation of apoptosis in Crohn's disease.

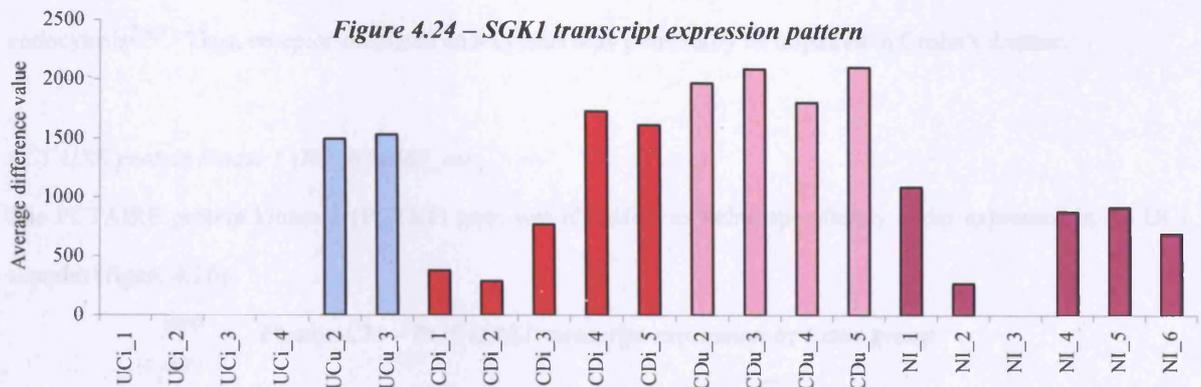
Figure 4.23 – *CED-6* transcript expression pattern



##### *Serine/threonine protein kinase (Y10032\_at)*

This gene is called serum/glucocorticoid regulated kinase 1 (*SGK1*) and has previously been observed as being up regulated in small intestinal Crohn's tissues<sup>179</sup>, but has not been investigated in colonic Crohn's or ulcerative

colitis. In the current study, *SGK1* is not detected (called absent) in 100% of the UCi samples. This is significant, as it is expressed at a high level in samples where it has been detected (figure 4.24).

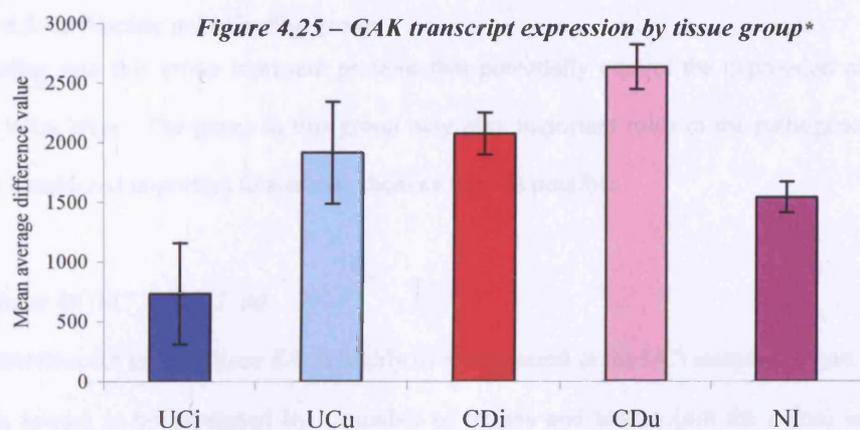


*SGK1* expression is induced by the corticosteroid aldosterone<sup>254-256</sup> and mediates the effect of aldosterone on sodium homeostasis. Specifically *SGK1* acts on the sodium channel ENaC (epithelial Na<sup>+</sup> channel). ENaC is composed of four homologous subunits ( $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\delta$ ) and mediates Na<sup>+</sup> absorption in the kidney, lung, exocrine glands and the distal colon<sup>255</sup>. *SGK1* stimulates ENaC action by activating pre-existing channels and inducing the transcription of ENaC subunits<sup>255, 257</sup>.

The down regulation of *SGK1* in the UCi samples therefore implies that the ability of UCi mucosa to absorb Na<sup>+</sup> is altered. The reduction of Na<sup>+</sup> absorption in UC affected mucosa is well documented and has been proposed as a cause of the diarrhoea in ulcerative colitis<sup>258</sup>. The decreased expression of *SGK1* in UCi provides a potential mechanism at the molecular level.

#### *Cyclin G associated kinase (RC\_AA287111\_at)*

This gene (*GAK*) is over expressed in CDi compared to the UCi and NI samples, whilst the CDu samples show higher expression (figure 4.25).

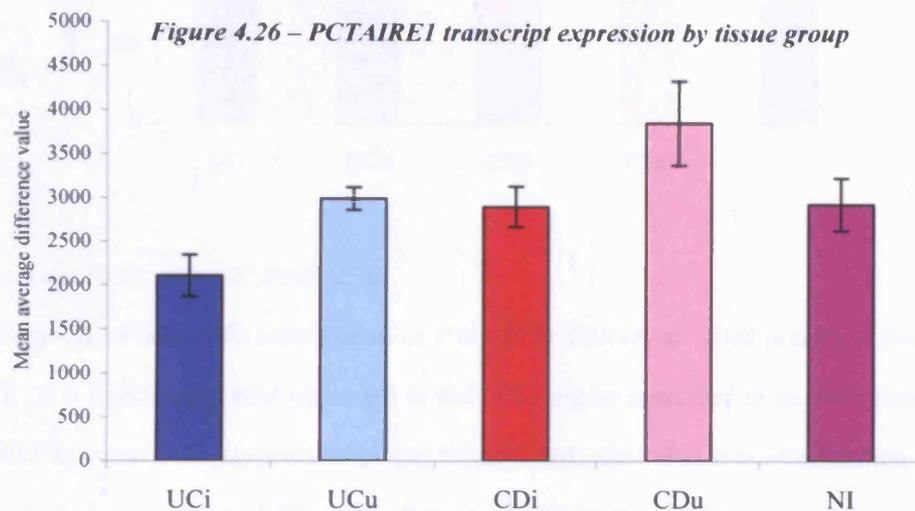


\* The error bars in the 'by tissue group' graphs represent the standard error (detailed in section 3.3.3.4).

GAK plays a vital role in the uncoating of clathrin-coated vesicles *in vitro*<sup>259</sup>, a crucial step in clathrin-mediated endocytosis and cells that transiently over express GAK *in vitro* show impaired receptor-mediated endocytosis<sup>260</sup>. Thus, receptor-mediated endocytosis may potentially be impaired in Crohn's disease.

#### *PCTAIRE protein kinase 1 (RC\_N58463\_at)*

The PCTAIRE protein kinase 1 (PCTK1) gene was identified as being specifically under expressed in the UCi samples (figure 4.26).



PCTK1 is one of three kinases in the PCTK sub-family of serine-threonine kinases<sup>261</sup>. PCTK1 is involved in regulation of the cell cycle and is ubiquitously expressed. It is part of the Cdc2 signalling cascade and while it has not been investigated in colonic or intestinal tissues, it is highly expressed in differentiated cells such as postmitotic neurons and spermatogenic cells<sup>262</sup>. Its expression level is cell cycle dependant<sup>263</sup>, but there is nothing in the literature to suggest why *PCTK1* should be specifically under expressed in the UCi tissues.

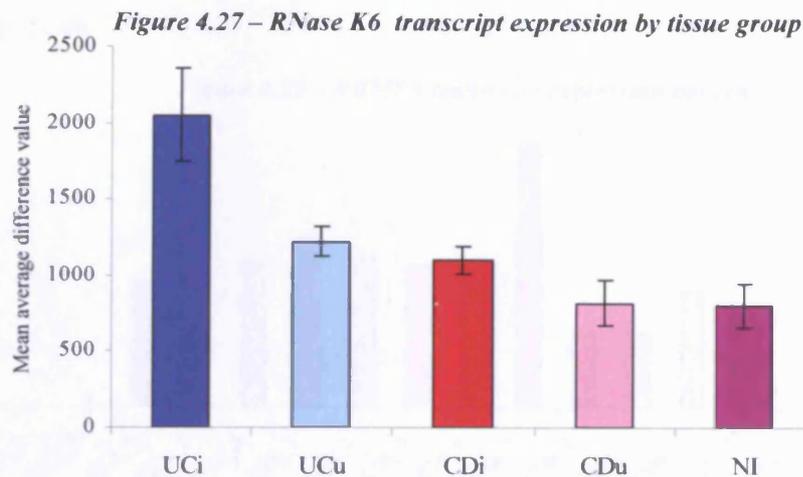
#### **4.5.4.2 Nucleic acid binding group**

Genes falling into this group represent proteins that potentially control the expression of other proteins at the DNA or RNA level. The genes in this group may play important roles in the pathogenesis of IBD and it was therefore considered important to annotate them as fully as possible.

#### *Ribonuclease k6 (RC\_N49002\_at)*

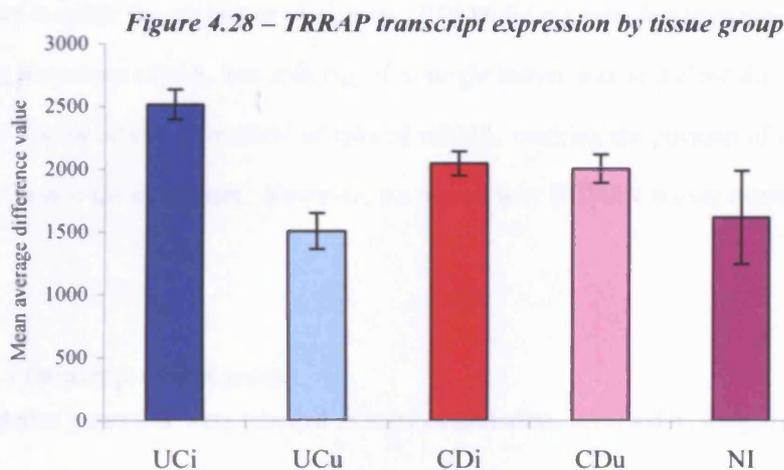
The ribonuclease k6 gene (*RNase K6*) is highly over expressed in the UCi samples (figure 4.27). The RNase K6 protein is known to be expressed by a number of organs and tissues (not the colon) and by monocytes and neutrophils<sup>264</sup>. The expression of RNase K6 by monocytes and neutrophils implies that *RNase K6* may be involved in host defence<sup>265</sup>. In addition, the over expression seen in the UCi samples probably reflects the higher number of neutrophils observed in the UCi samples (section 2.3.7). However, the *RNase K6* expression

data cannot be normalised to the ICC data, as the gene is expressed by both monocytes and neutrophils. The mechanism by which this gene may be involved in host defence is unclear.



*PCAF-associated factor 400 (RC\_R20732\_at)*

The PCAF-associated factor 400 (also known as *PAF400* or *TRRAP*) is called present in every sample, except one (NI\_5). It is significantly over expressed in the UCi samples compared to all other tissues groups (figure 4.28). *TRRAP* expression has not previously been linked to IBD, the colon or to inflammation in general.



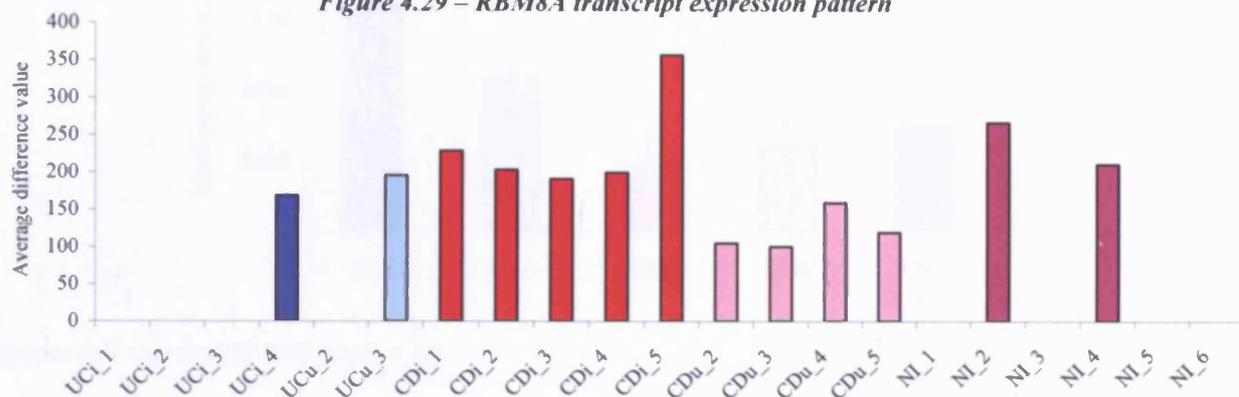
*TRRAP* deficient cells do not sustain mitotic arrest during cell proliferation *in vitro* and transgenic mice with a null mutation results in pre-implantation lethality due to the blocked proliferation of blastocytes<sup>266</sup>. Besides its apparent role in normal cell cycle progression, *TRRAP* is an essential cofactor in the c-myc and E1A/E2F oncogenic transcription factor pathways<sup>267</sup>. Thus the over expression of this gene, may be related to the increased risk of colorectal cancer in ulcerative colitis (section 1.1.4.4).

*cDNA DKFZp564J1616 from clone DKFZp564J1616 (RC\_AA450078\_at)*

This probe set represents an EST from the RNA binding motif protein 8A (*RBM8A*). It is over expressed in the Crohn's disease samples compared to all other tissues groups and over expressed in the CDi samples compared

to the CDu samples (figure 4.29). The overall expression level in the CDi samples is not very high (mean of 235 average difference units), but as the gene was called absent in the majority of the other samples, it was still considered significant.

**Figure 4.29 – RBM8A transcript expression pattern**



RBM8A is identical to another protein, RBM8B that is 16 amino acids shorter at the N-terminus of the protein. Both interact with OVCA1, a candidate tumour suppressor protein<sup>268</sup>. However the two proteins are located on separate chromosomes, 14q22-q23 and 5q13-q14 respectively<sup>269, 270</sup>. Collectively the two proteins are referred to as RBM8. A study by Kataoka *et al*<sup>271</sup> found that RBM8 associates preferentially with mRNA produced by splicing, located in either the nucleus or cytoplasm. RBM8 did not associate with pre-mRNA, introns, or mRNA produced from intronless cDNA, but splicing of a single intron was sufficient for RBM8 association. They concluded that RBM8 acts as a 'marker' of spliced mRNA, marking the position of the removed introns as the mRNA re-locates into the cytoplasm. However, the reason why RBM8A is over expressed in the CDi mucosa is not clear.

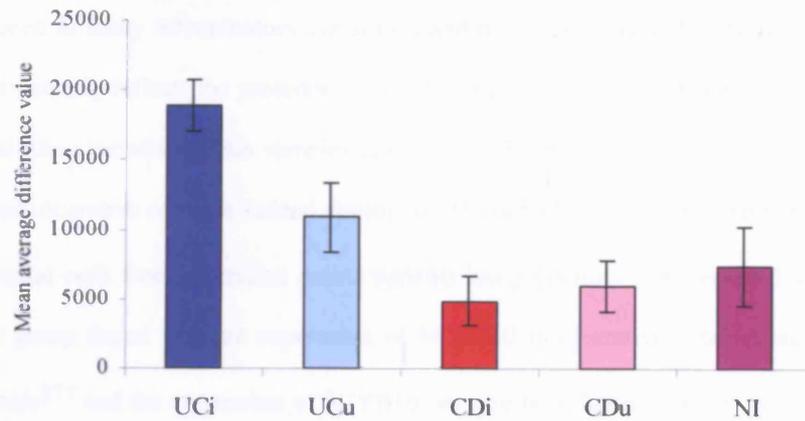
#### 4.5.4.3 Immunity related group

This group includes genes that were labelled as immunoglobulins, involved in antigen presentation, or cytokines. These genes are expected to be over-expressed in the IBDi tissues due to the colitis in IBD.

##### *β2-microglobulin (S82297\_at)*

$\beta$ 2-microglobulin ( $\beta$ 2m) is the small subunit of the MHC I molecule<sup>272</sup> and has previously been investigated in IBD. The over expression of  $\beta$ 2m in UCI (figure 4.30) is compatible with the report that HLA-B27 transgenic rats with a high copy number of the human  $\beta$ 2m gene develop a chronic colitis comparable to ulcerative colitis<sup>273</sup>. Elevated levels of  $\beta$ 2m protein are believed to reflect increased secretion from activated T cells and neutrophils<sup>274</sup>. The current data implies that  $\beta$ 2m may prove to be a specific marker of UCI.

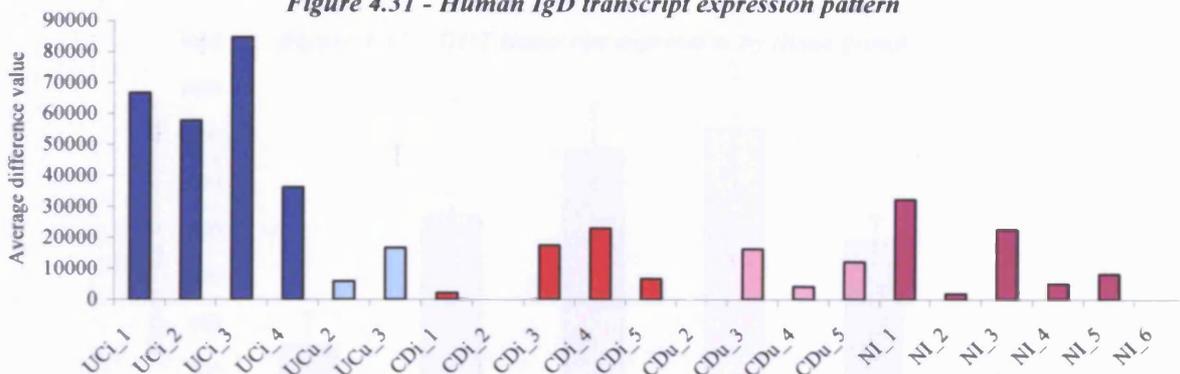
Figure 4.30 –  $\beta$ 2-microglobulin transcript expression by tissue group



Human IgD segment (S71043\_rnal\_s\_at)

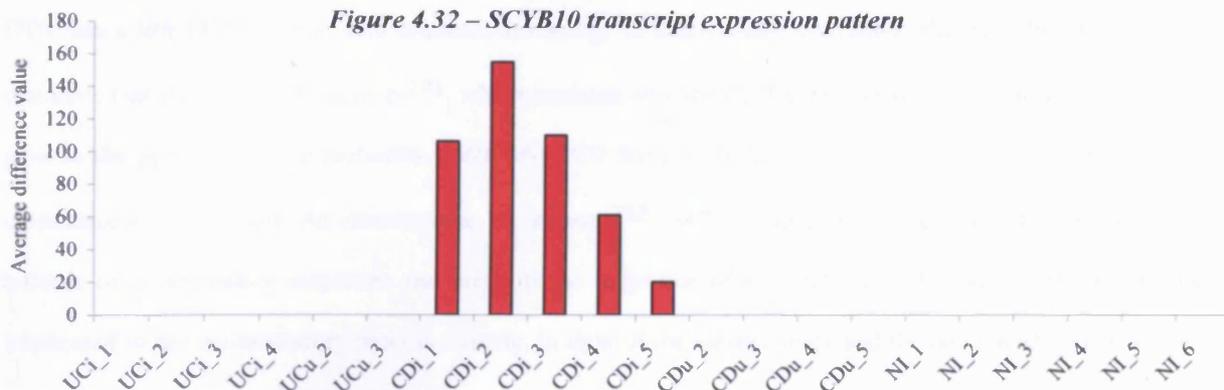
Due to the large number of activated B cells in ulcerative colitis, the over expression of this immunoglobulin related gene in the UCI samples, is expected (figure 4.31). The increased expression of this immunoglobulin gene in UCI tissues compared to controls (4-fold increase) was also noted in a previous microarray experiment<sup>181</sup>.

Figure 4.31 - Human IgD transcript expression pattern



$\gamma$ -interferon inducible early response gene (X02530\_at)\*

Figure 4.32 – SCYB10 transcript expression pattern



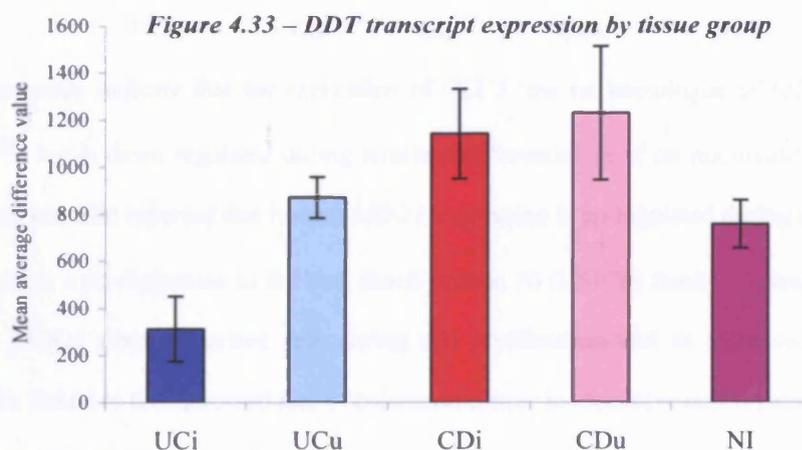
Although the level of expression of SCYB10 in CDi is low, the difference is significant as it is called absent in 100% of the other samples (figure 4.32). SCYB10 is a secreted CXC chemokine, chemotactic for monocytes

\* There are many alternative names for this gene including 'small inducible cytokine subfamily B', 'SCYB10', 'interferon- $\gamma$  inducible protein 10', 'IP10' and 'INP10'. It is referred to in this thesis as SCYB10.

and T cells and induced in a number of different cell types in response to IFN $\gamma$ <sup>275</sup>. The expression of SCYB10 has been investigated in many inflammatory conditions and it seems likely that the over expression observed in the current study simply reflects the presence of an inflammatory reaction. However, the fact that SCYB10 is called absent in all the ulcerative colitis samples (except UC\_5) implies that the inflammatory response in the Crohn's disease and ulcerative colitis is indeed dissimilar. In contradiction to the current study, the expression of SCYB10 by epithelial cells from ulcerative colitis patients has previously been reported as being increased<sup>276</sup>. However another group found that the expression of SCYB10 in ulcerative colitis tissues was not increased compared to controls<sup>277</sup> and the expression of SCYB10 has also been linked to a Th1 type immune response<sup>277, 278</sup>, supporting its induction by IFN $\gamma$  and the CDi specific expression seen here.

#### *D-dopachrome tautomerase (U49785\_at)*

D-dopachrome tautomerase (DDT) catalyses the conversion of D-dopachrome into 5,6-dihydroxyindole<sup>279</sup>. The current study shows a decrease of *DDT* expression by the UCi samples compared to the other tissues groups (figure 4.33); this has not been previously observed.



DDT has a low (33%) amino acid sequence homology to macrophage migration inhibitory factor (MIF)<sup>280</sup>, a cytokine, that itself has DDT activity<sup>281</sup>, which explains why the *DDT* gene was marked as an immunity related gene in the protein function hierarchy. *MIF* and *DDT* have a similar gene structure and are closely linked on chromosome 22 (human) and chromosome 10 (mouse)<sup>282</sup>. *MIF* is expressed by activated T-lymphocytes in a number of inflammatory responses and prevents the migration of macrophages. Although *DDT* has not been implicated in any inflammatory process directly, in light of the current study and the fact that this gene sequence has been patented as MIF-3<sup>283</sup> it is highly likely that *DDT* does indeed play a role in inflammation.

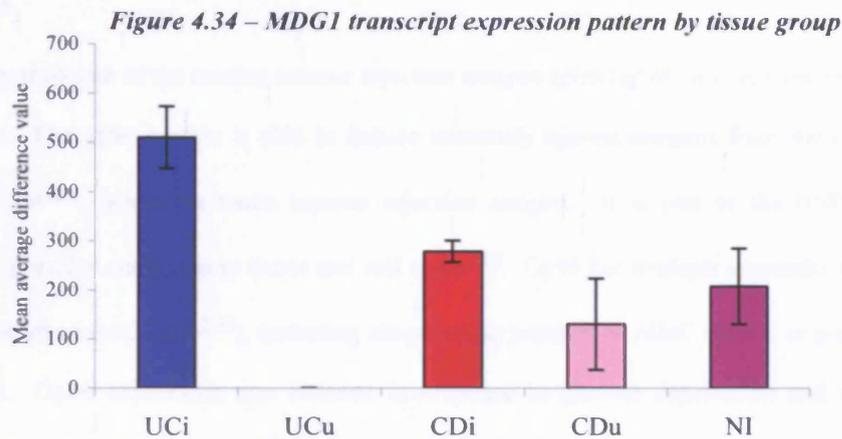
Interestingly, *DDT* and *GSTT2* (called absent in all ulcerative colitis samples; figure 4.10) lie on the same gene complex and are thought to be under the control of a single, bi-directional promoter<sup>224</sup>. Further study of this promoter may provide important clues to the pathogenesis of UC.

#### 4.5.4.4 Miscellaneous protein function group

This group contains 5 proteins that show a CDi or UCi specific expression pattern, but do not fall into any of the protein function categories discussed above.

##### *Microvascular endothelial differentiation gene 1(RC\_AA045793\_s\_at)*

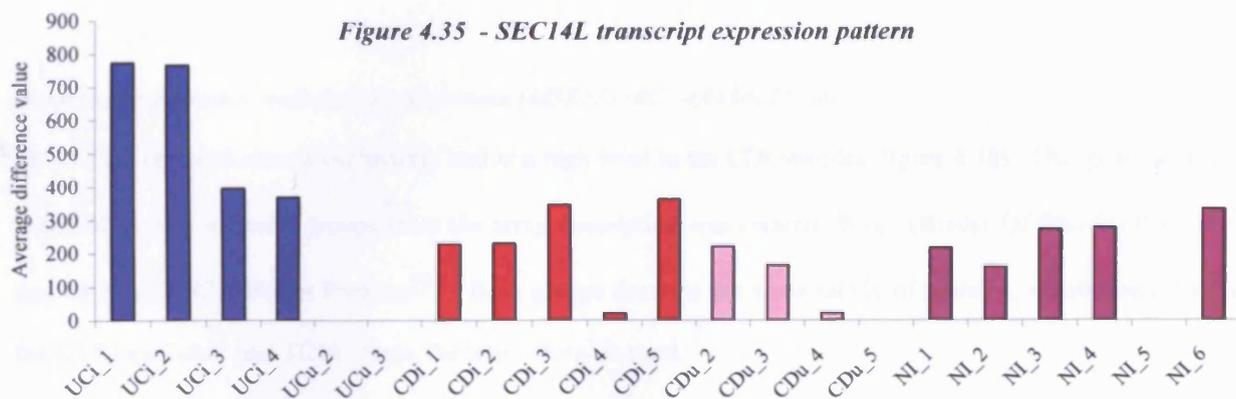
The microvascular endothelial differentiation gene (*MDG1*) shows higher expression in the UCi samples compared to the other tissue groups (figure 4.34). It is expressed by adult endothelial and epithelial cells<sup>284</sup>, which makes the current data even more significant, as the UCi samples contained the least proportion of epithelial cells of all the tissue groups (section 2.3.7).



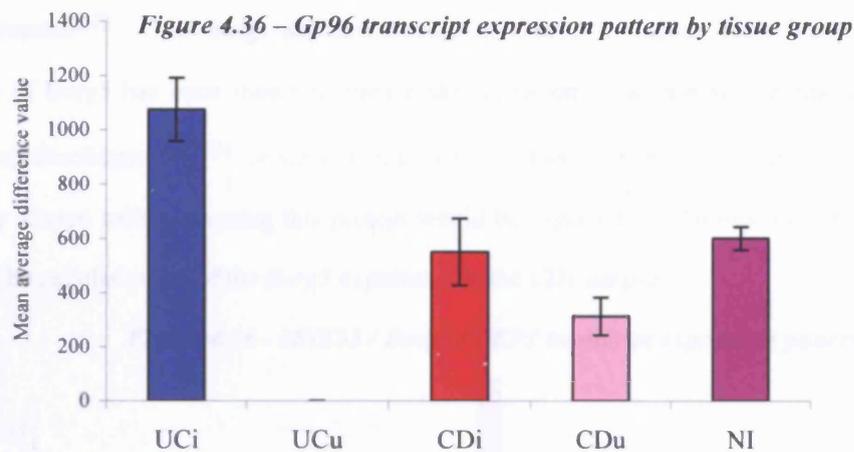
Studies by one group indicate that the expression of *CEC5* (the rat homologue of *MDG1*) is increased during proliferation<sup>285</sup>, but is down regulated during terminal differentiation of rat microvascular endothelial cells<sup>286</sup>. The same group has also reported that human *MDG1* expression is up regulated during angiogenesis *in vitro* and that *MDG1* acts as a co-chaperone to the heat shock protein 70 (HSP70) family of protein chaperones<sup>284</sup>. This indicates that *MDG1* plays an active role during cell proliferation and its increased expression in the UCi samples may be linked to the increased rate of colorectal cancer in ulcerative colitis patients.

##### *Human SEC14L (RC\_AA122317\_s\_at)*

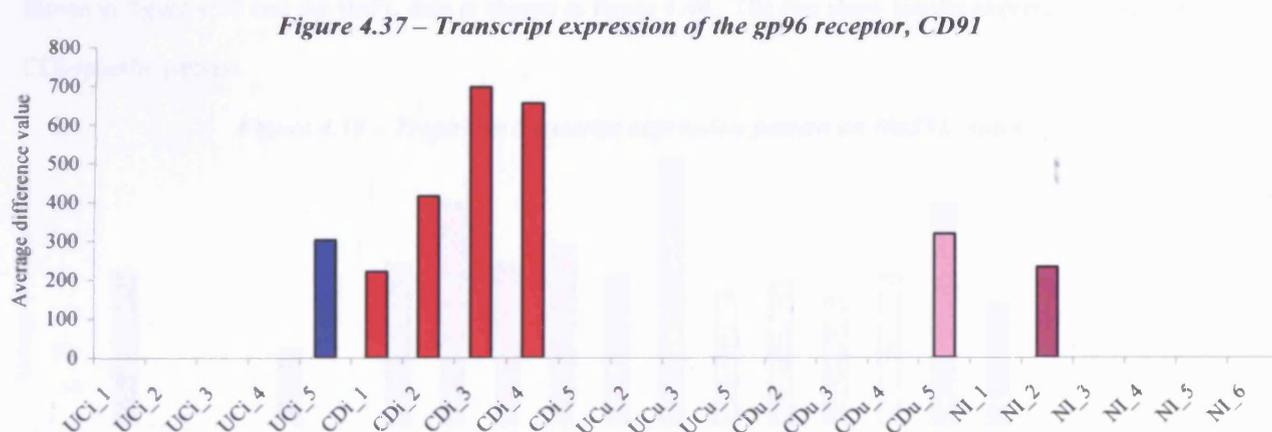
This gene is over expressed in the UCi samples compared to the other tissues (figure 4.35). Human *SEC14L* is homologous to a *C. elegans* protein and a yeast protein called *SEC14*, which is putatively involved in intracellular transport<sup>287</sup>. The function of the human *SEC14L* is unknown.



## Human homologue of murine tumour rejection antigen gp96. (RC\_R20669\_f\_at)



The human homologue of the murine tumour rejection antigen gp96 (*gp96*) is over expressed in the UCi samples (figure 4.36). The gp96 protein is able to induce immunity against antigens from the cell from which it was initially derived<sup>288</sup>, hence the name tumour rejection antigen. It is part of the HSP90 family and shows ubiquitous expression across many tissue and cell types<sup>289</sup>. Gp96 has multiple immunity related roles (reviewed in Schild & Rammensee, 2000<sup>290</sup>), including chaperoning peptides to MHC class I, in preference to MHC class II molecules. Gp96 expression was induced in response to glucose deprivation and its expression was up regulated in colon cancer cell lines<sup>288</sup>. The gp96 receptor, CD91 (also known as  $\alpha$ 2-macroglobulin receptor; X13916\_at)<sup>291</sup> is present on the HuFL arrays, but unexpectedly was called absent in all the ulcerative colitis samples except UCi\_5 (figure 4.37).

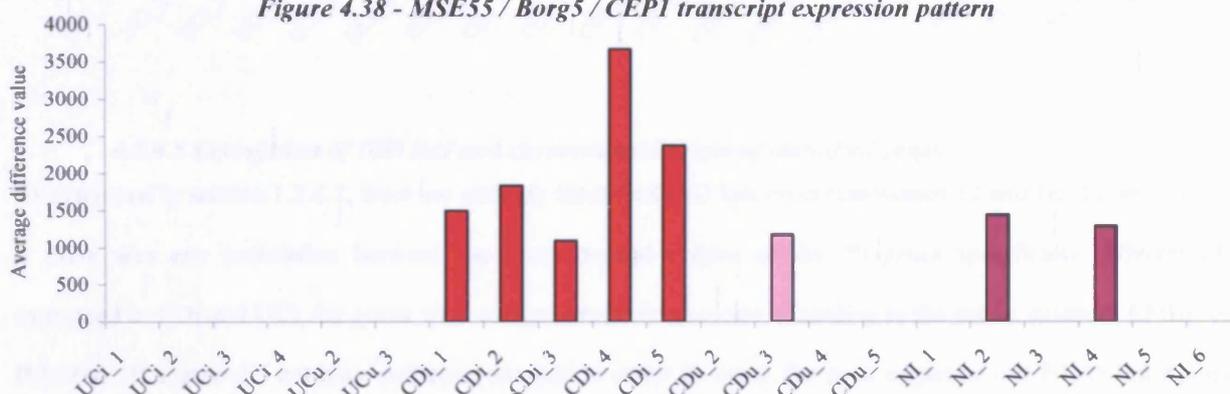


## Bone marrow stromal/endothelial cell protein (MSE55) (RC\_AA434225\_at)

MSE55 is expressed almost exclusively and at a high level in the CDi samples (figure 4.38). This gene has been renamed by two different groups since the array description was created; *Borg5* (Binder Of Rho-GTPases)<sup>292</sup> and CEPI (Cdc42 Effector Protein)<sup>293</sup>. Both groups describe the same family of proteins, which interact with the GTPases Cdc42 and TC10. Here, the term 'Borg' is used.

Borg5 expression is limited to endothelium and bone marrow, whereas other members of the Borg family show wider expression<sup>292</sup>. The Borgs act downstream of Cdc42 to induce actin filament assembly and over expression of Borg5 has been shown to induce the formation of actin-based cellular extensions in epithelial cells<sup>293</sup> and fibroblasts<sup>292, 294</sup> *in vitro*. If the *in vivo* effect of *Borg5* over expression were the same, then abnormally shaped cells expressing this protein would be expected in CDi mucosa. It would be interesting to determine the cellular origin of the *Borg5* expression in the CDi samples.

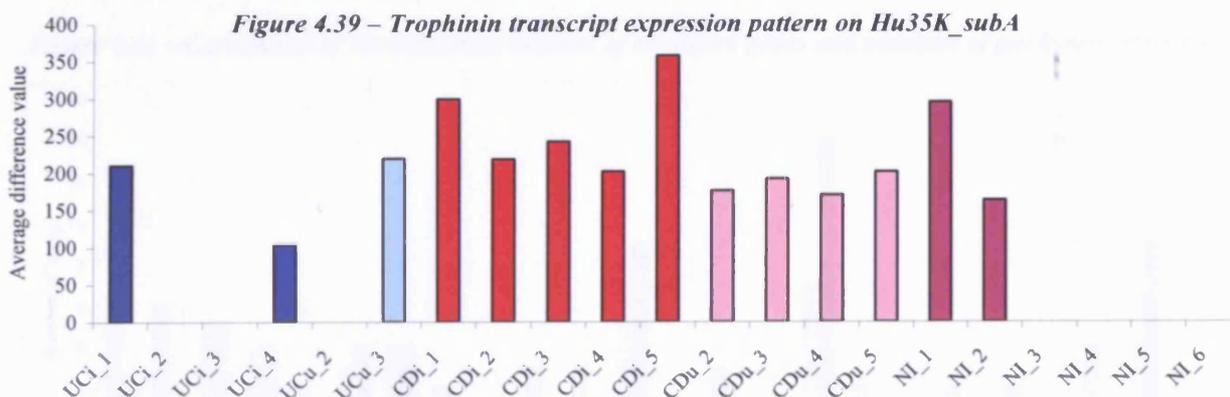
**Figure 4.38 - MSE55 / Borg5 / CEPI transcript expression pattern**



#### *KIAA1114* protein (U04811\_at)

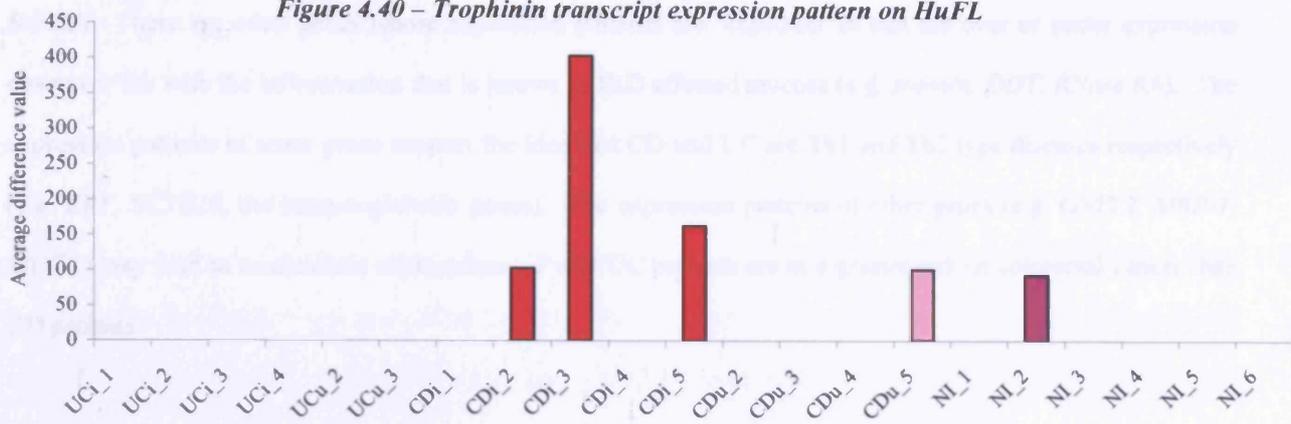
The accession number for this gene links to the EMBL entry for 'trophinin' and the array description has been updated since this 1997 description. Trophinin is specifically over expressed in the CDi samples (figure 4.39) and is represented on two of the arrays used, HuFL and Hu35K\_subA. The results from the Hu35K\_subA are shown in figure 4.39 and the HuFL data is shown in figure 4.40. The two show similar expression levels, with a CDi specific pattern.

**Figure 4.39 – Trophinin transcript expression pattern on Hu35K\_subA**



Trophinin is expressed by the embryo and, with the tastin protein, it forms the trophoblast adhesion complex<sup>295</sup>. The tastin gene is also represented on two arrays (HuFL & Hu35K\_subC), but is called absent in all the samples on both arrays. Therefore, the trophoblast complex is not being formed in the CDi samples and trophinin may have another, as yet unidentified role to play, independent of the tastin protein.

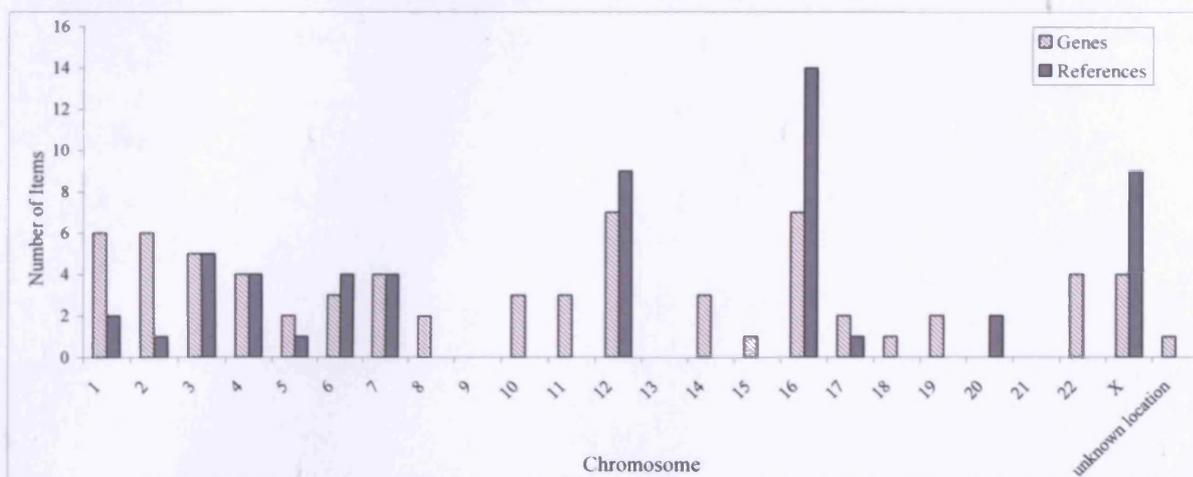
Figure 4.40 – Trophinin transcript expression pattern on HuFL



#### 4.5.4.5 Correlation of IBD loci and chromosomal origin of identified genes

As discussed in section 1.3.4.3, there are strongly identified IBD loci on chromosomes 12 and 16. In order to see if there was any correlation between the chromosomal origins of the 70 genes specifically differentially expressed in CDi and UCI, the genes were assigned to a chromosome according to the public database EMBL on PubMed. If a gene did not have sufficient annotation under its entry, the gene sequence was BLASTed against the public database, to find the closest match. In one case, the sequence did not match to any one chromosome with sufficient specificity to allow assignment. Figure 4.41 shows the results of this, and it was interesting to note that chromosomes 12 and 16 were well represented. However, the microarray data measures gene expression, whilst the IBD loci are associated with mutations in the genome. Also, the genes were not located to specific loci, and may not originate from the specific IBD associated areas. Whether the genes identified do in fact originate from the IBD specific loci remains a question for future studies, as the further refinement of the human genome allows the mapping of genes with increasing precision.

Figure 4.41 – Correlation of chromosomal location of identified genes and numbers of published references



#### 4.5.5 Section summary

The genes discussed above have roles in a wide range of cellular functions. Some of the expression patterns discussed cannot be explained from the published literature in terms of a known IBD hypothesis (e.g. *dynactin*,

*BdrC3*). There are other genes whose expression patterns are 'expected' in that the over or under expression observed, fits with the inflammation that is known in IBD affected mucosa (e.g. *stannin*, *DDT*, *RNase K6*). The expression patterns of some genes support the idea that CD and UC are Th1 and Th2 type diseases respectively (e.g. *ERF*, *SCYB10*, the immunoglobulin genes). The expression patterns of other genes (e.g. *GSTT2*, *MRP-1*, *MDG1*) may lead to mechanistic explanations of why UC patients are at a greater risk of colorectal cancer than CD patients.

#### 4.6 Which genes are differentially regulated in the involved IBD tissues as a group?

##### 4.6.1 Rationale and Method

The aim of this query was to identify genes that tend to follow similar patterns in the involved IBD (IBDi) samples. A few of these genes were identified in section 4.5, but the main difference in this query was that the CDi and UCi samples were treated as a single group to compare against the uninvolved IBD and control samples. As the IBD samples were being pooled, the UCi\_5 samples were included. Thus, a new spreadsheet was created specifically to answer this query, the 'pooled data' spreadsheet. The stringency of the queries was adjusted to give the optimal number of genes for annotation (detailed in table 4.9).

*Table 4.9 – Identification of genes differentially expressed in IBDi samples*

<i>IBDi expression pattern</i>	<i>Criteria</i>	<i>Genes in individual query</i>	<i>Genes returned in cumulative query</i>
	IBDi vs. NI T-test column = $P < 0.05$	22171	22171
	IBDi vs. IBDu T-test column = $P < 0.05$	22164	21202
Decreased	Absent in $\geq 7$ IBDi samples	20373	7952
	Present in $\geq 5$ IBDu samples	7049	98
	Present in $\geq 4$ NI samples	7199	14
	IBDi vs. NI T-test column = $P < 0.05$	22171	22171
	IBDi vs. IBDu T-test column = $P < 0.05$	22164	21202
Increased	Present in $\geq 7$ IBDi samples	7708	7708
	Absent in $\geq 5$ IBDu samples	21283	200
	Absent in $\geq 5$ NI samples	20200	29

##### 4.6.2 Results

###### 4.6.2.1 Genes with decreased expression in IBDi

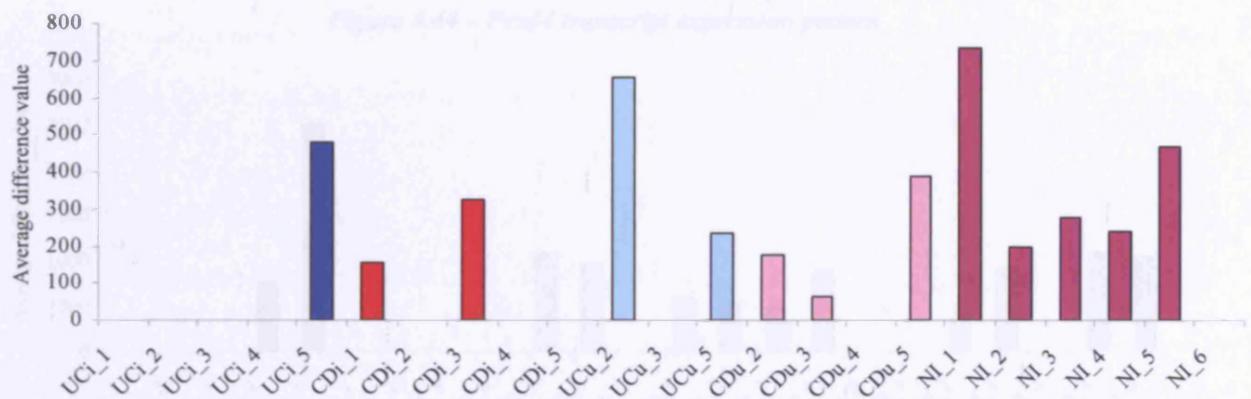
The 14 genes that show decreased expression in IBDi compared to NI are listed in appendix A4.4. Eight of these are EST sequences and BLAST searching revealed the chromosomal location only. Another of the genes represents a protein of unknown function (KIAA0872). The remaining five genes are annotated below.

###### *Putative transcription factor (RC\_AA055932\_at)*

This gene is labelled as an EST in the EMBL database and as a putative transcription factor (PCAF associated factor 65 beta; PAF65 $\beta$ ) in the UniGene database. PAF65 $\beta$  shows medium expression levels in 100% and 71% of the NI and IBDu samples respectively. By contrast, it is called absent in 70% of the IBDi samples (figure 4.42).

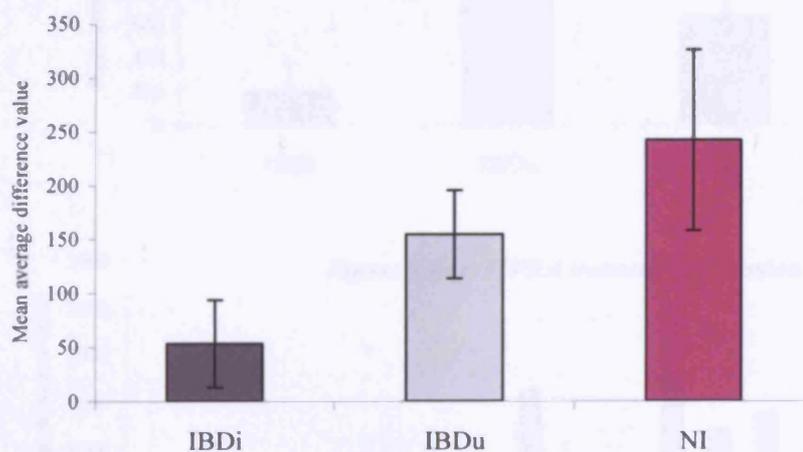
PAF65 $\beta$  has not previously been linked to IBD, inflammation or the colon. The PAF65 $\beta$  protein is part of the PCAF histone acetylase complex, which is involved in transcription, cell cycle progression and differentiation<sup>296</sup>.

**Figure 4.42 – Transcript expression pattern of PAF65B**



*KIAA0448 protein (RC\_AA283046\_at)*

The KIAA0448 protein is also known as heparan sulphate 2-O-sulfotransferase (*Hs2st1*)<sup>297</sup>, a heparan sulphate biosynthetic enzyme. *Hs2st1* is part of a family of enzymes that act in concert to generate multiple distinct heparan sulphate structures. The generation of the correct heparan sulphate group on target proteins is vital in heparan dependant protein-ligand interactions<sup>298</sup>. *Hs2st1* transfers sulphate to the second carbon (C-2) of the iduronic acid residue of heparan sulfate<sup>299</sup>. The observed decrease in the expression of *Hs2st1* (figure 4.43) indicates that proteins that are dependant on sulphated heparan side chains may exhibit altered functional capability in the IBD affected mucosa.

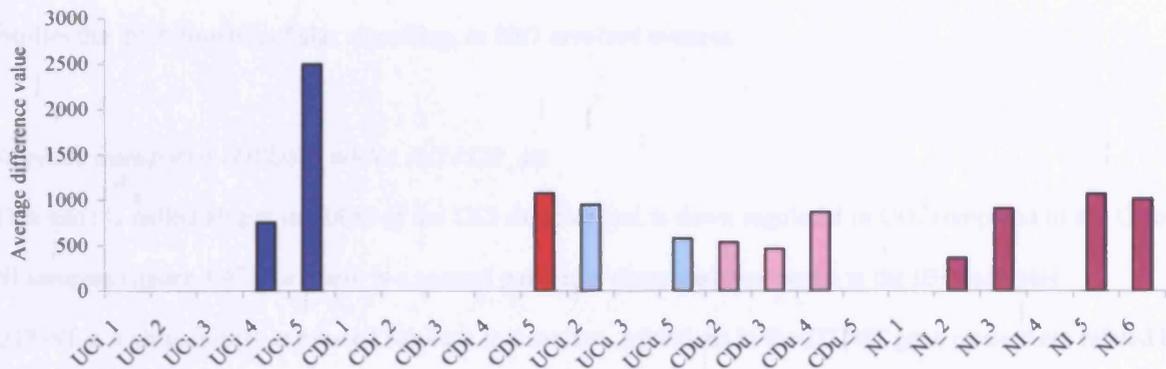


**Figure 4.43 – HS2ST1 transcript expression by tissue group**

*Human peroxisomal membrane anchor protein (RC\_AA450302\_s\_at)*

This gene, (also known as peroxisomal biogenesis factor 14 or Pex14), acts as a docking protein in the peroxisome<sup>300</sup>. However, this gene is not clearly down regulated in the IBDi samples, with UCi\_5 showing very high expression levels (figure 4.44). Also, not all of the IBDu and NI samples express this gene and in light of this expression pattern, this gene expression pattern was not considered to be of interest.

Figure 4.44 – Pex14 transcript expression pattern



*Inositol 1,4,5-triphosphate 3-kinase A (RC\_T16315\_s\_at)*

The down regulation of *inositol 1,4,5-triphosphate 3-kinase A* (or *ITPKA*) in the IBDi samples is evident from the array data (figure 4.45). *ITPKA* is called present in only 2 of the IBDi samples (20%), in 6 of the IBDu (86%) and 4 of the NI samples (67%) (figure 4.46).

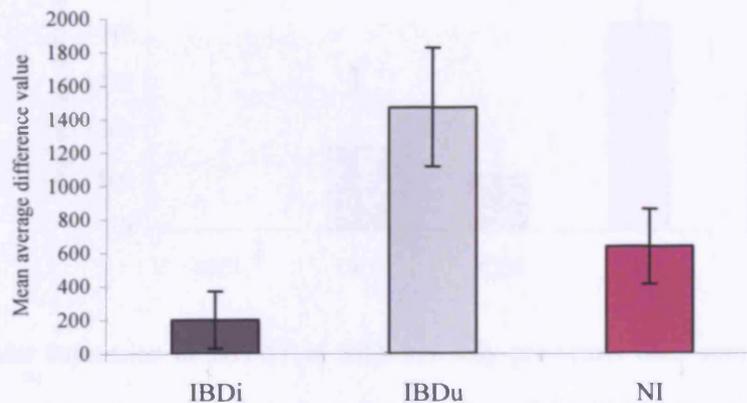


Figure 4.45 – ITPKA transcript expression by tissue group

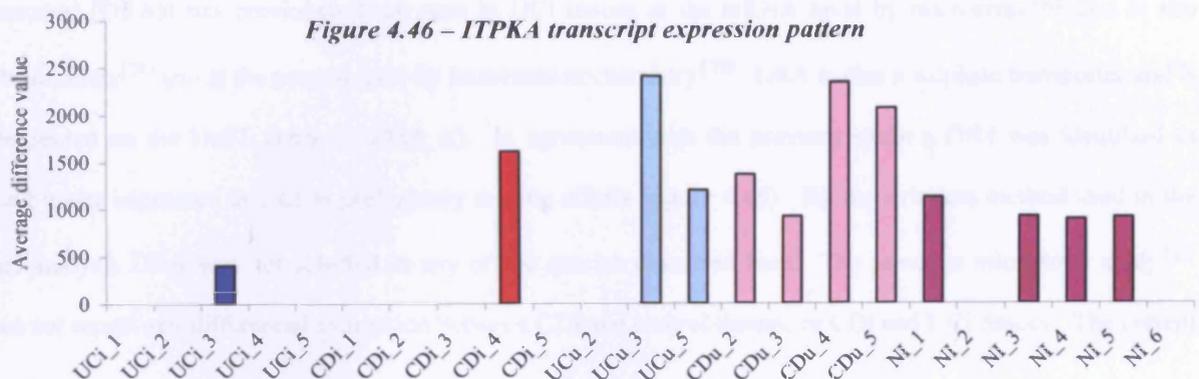


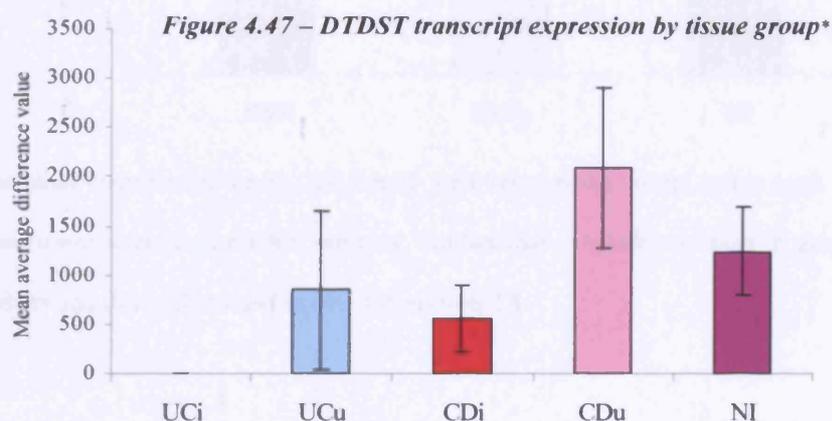
Figure 4.46 – ITPKA transcript expression pattern

ITPKA is one of at least three differentially expressed isoforms of inositol 1,4,5-triphosphate 3-kinase (ITPK), which are regulated by calmodulin and calcium and in turn regulate the level of D-myo-inositol 1,4,5-trisphosphate (IP<sub>3</sub>)<sup>301</sup>. IP<sub>3</sub> is a second messenger molecule that is used as a substrate by ITPK to produce inositol 1,3,4,5-tetrakisphosphate (IP<sub>4</sub>), which has multiple functions in the cell (reviewed by Irvine, 2001<sup>302</sup>), including activating calcium channels in epithelial cells<sup>303</sup>. The down regulation of *ITPKA* in the IBDi samples implies the inhibition of cellular signalling, in IBD involved mucosa.

#### *Sulphate transporter (DTDST) mRNA (U14528\_at)*

This gene is called absent in 100% of the UCi samples and is down regulated in CDi compared to the CDu and NI samples (figure 4.47), i.e. there is a general pattern of decreased expression in the IBDi samples.

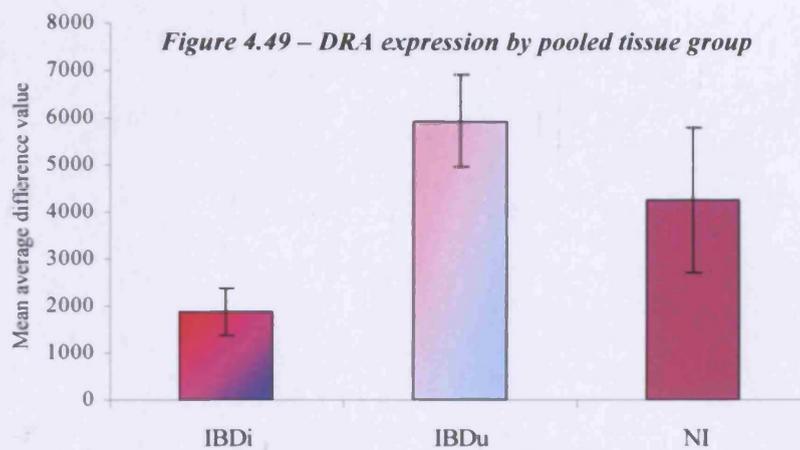
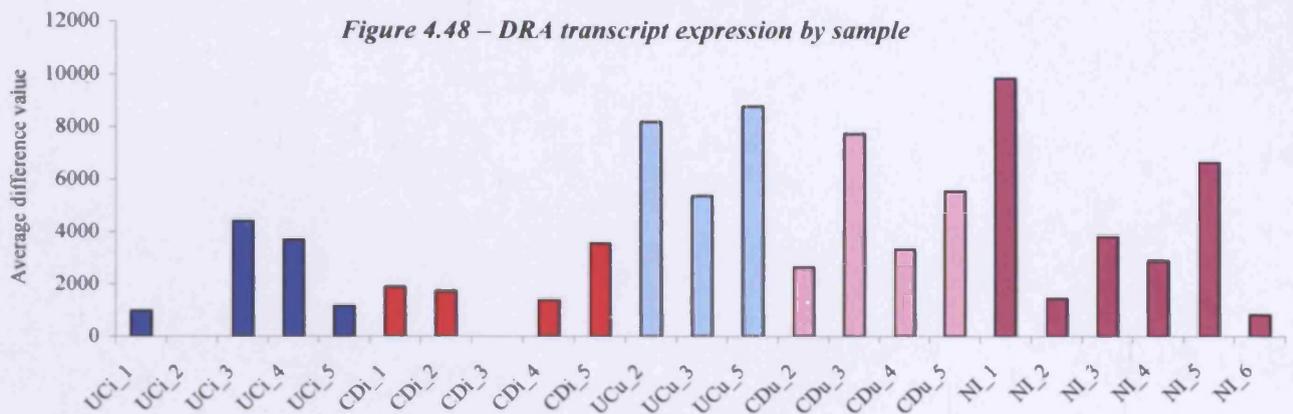
DTDST is a ubiquitously expressed sulphate transporter. Mutations in the *DTDST* gene cause three related bone diseases of varying severity (diastrophic dysplasia, achondrogenesis type IB and atelosteogenesis type II), with the mutation determining the disease. DTDST is part of the 'solute carrier 26' family of proteins, named SLC26A1 to SLC26A11. DTDST is therefore also known as SLC26A2<sup>304</sup>.



The under expression of *DTDST* in IBDi has only previously been seen from a microarray experiment<sup>181</sup>. However, the under expression of another member of the SLC26 family, SLC26A3 (or 'down-regulated in adenoma' (DRA)) has previously been seen in UCi tissues at the mRNA level by microarray<sup>181</sup> and *in situ* hybridisation<sup>170</sup> and at the protein level by immunohistochemistry<sup>170</sup>. DRA is also a sulphate transporter and is represented on the HuFL array (L02785\_at). In agreement with the previous studies, *DRA* was identified as being under expressed in UCi in preliminary mining efforts (figure 4.48). By the stringent method used in the final analysis *DRA* was not selected in any of the queries described here. The previous microarray study<sup>181</sup> does not report any differential expression between CDi and control tissues, or CDi and UCi tissues. The current

\* This graph includes data from ulcerative colitis patient UC\_5.

data shows a similar level of expression in both the CDi and UCi tissues (figure 4.48), which is reflected in the decreased *DRA* expression in the IBDi samples (figure 4.49).



None of the other members of the SLC26 family were represented on the arrays used. The under expression of two sulphate transporters in the IBDi samples, implies that sulphate transport in general is decreased in IBD affected tissues and this is discussed in detail in section 4.8.

#### 4.6.2.2 Genes with increased expression in IBDi

The 29 genes that show increased expression in IBDi compared to NI are listed in appendix A4.5. Nine of the probe sets had no associated information in the Incyte functional hierarchy column. One was an unknown KIA protein. The remaining genes are summarised in figure 4.50. In some cases, the genes were called absent in all of the NI samples, so a fold change could not be calculated (table 4.10.)

Figure 4.50 – Genes that are over expressed in IBDi by fold change from NI samples

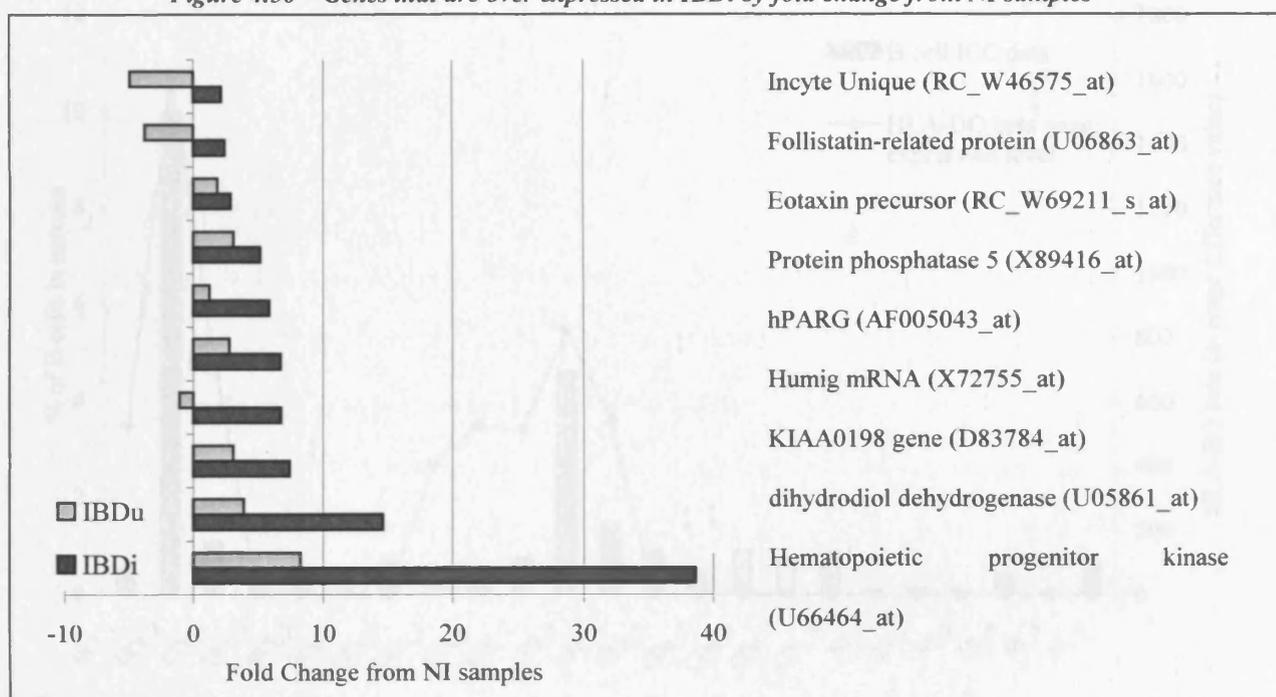


Table 4.10 – Genes that are called present in IBDi samples and called absent in NI samples

Accession Number	Description from array
RC_AA158396_s_at	Human mRNA for HLA-D class II antigen DO beta chain.
RC_AA191454_at	Human FGF-1 intracellular binding protein (FIBP) mRNA, complete cds.
M16336_s_at	Human T-cell surface antigen CD2 (T11) mRNA, complete cds, clone PB1.
Y08374_ma1_at	Human glycoprotein mRNA, complete cds.
RC_AA621096_at	Human okadaic acid-inducible and cAMP-regulated phosphoprotein 19 (ARPP-19) mRNA
RC_H80901_s_at	Human mRNA for Hakata antigen, complete cds.
RC_N24732_at	Human nuclear respiratory factor-2 subunit alpha mRNA, complete cds.
RC_N33920_at	Human mRNA for diubiquitin.
RC_N53534_s_at	Human beta-arrestin 2 mRNA, complete cds.

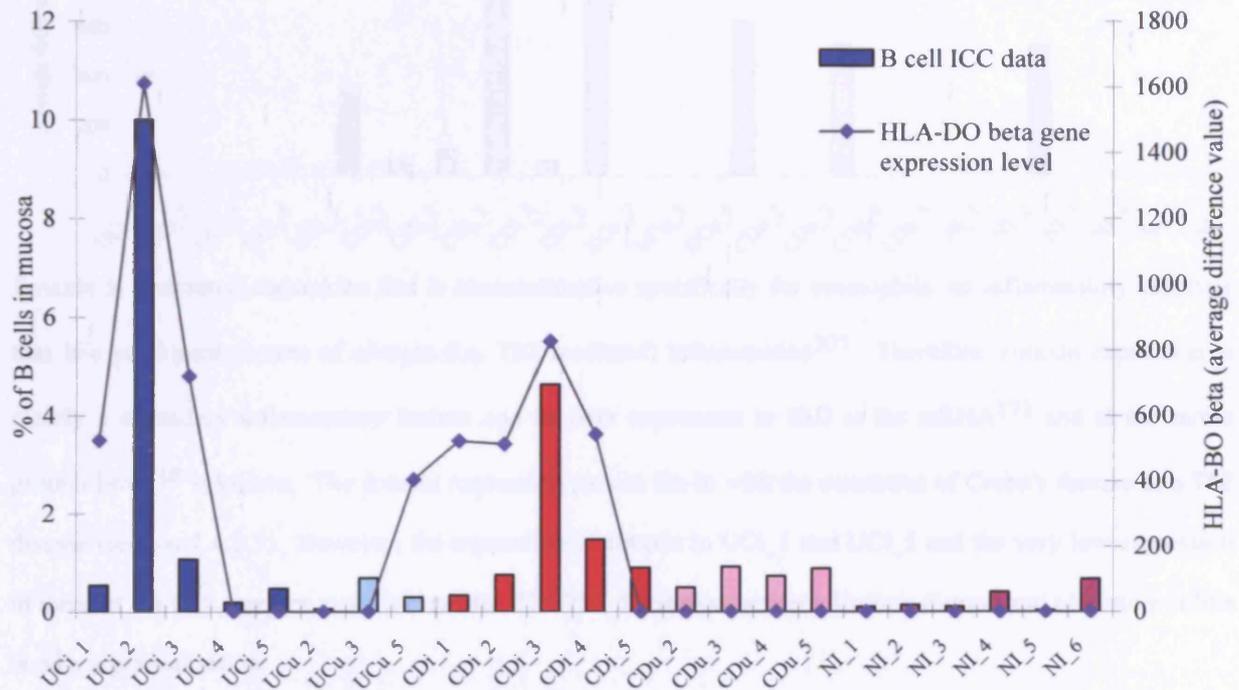
As this query does not represent the main focus of this thesis, the annotations of all the genes from figure 4.50 and table 4.10 are not detailed below. As expected, the majority of genes identified in this query were inflammation related. Some of these were known in IBD and some represent a novel expression pattern, but all are believed to be over expressed due to the inflammatory process in the IBDi tissues.

*HLA-D class II antigen DO  $\beta$  (RC\_AA158396\_s\_at)*

In humans, the DO $\beta$  chain of the HLA class II antigen is expressed exclusively by B cells<sup>305</sup>. Therefore the array data was normalised to the ICC B cells data and the over expression of this gene was no longer of interest.

The expression of the HLA-D class II DO $\beta$  antigen was found to be proportional to the percentage volume of mucosa corresponding to B lymphocytes (figure 4.51).

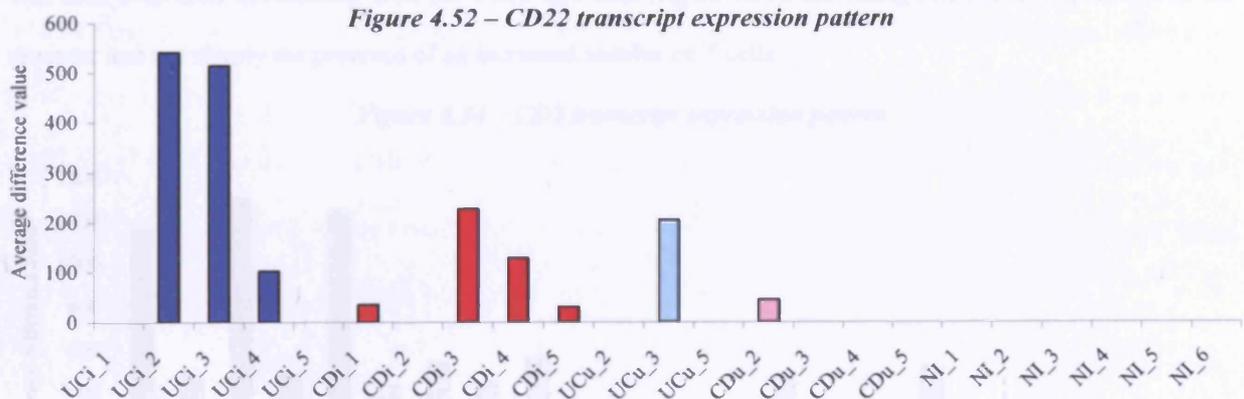
**Figure 4.51 - Comparing B lymphocyte ICC data to HLA-DO $\beta$  antigen transcript expression**



*$\beta$ -arrestin 2 (RC\_N53534\_s\_at)*

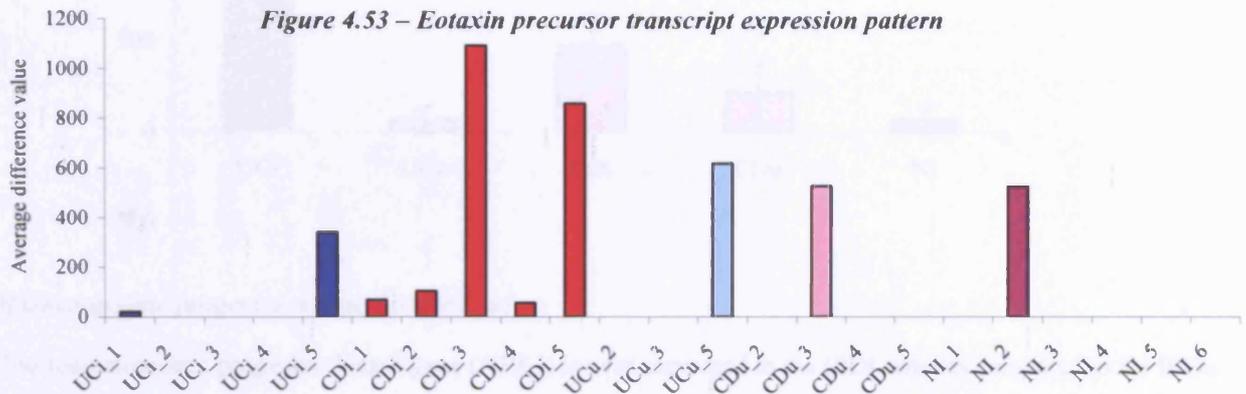
This gene is also known as CD22. CD22 is the precursor of the B cell receptor and is specifically expressed by B lymphocytes<sup>306</sup>. Thus, its expression in 3 of 5 of the UCI samples and its absence in 100% of the NI samples (figure 4.52) is unremarkable.

**Figure 4.52 - CD22 transcript expression pattern**



*Eotaxin precursor (R\_W69211\_s\_at)*

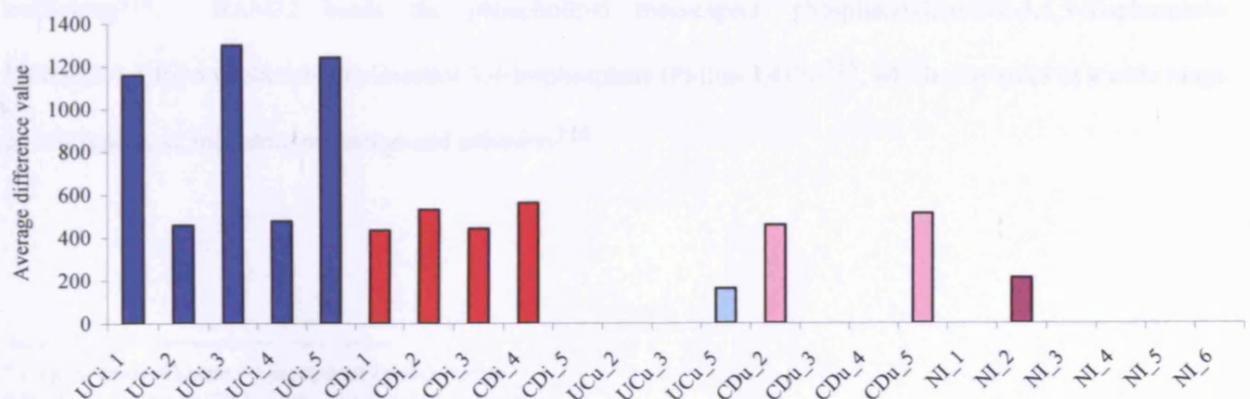
The eotaxin precursor tends to be expressed by the CDi samples; however, the expression level across the CDi samples is inconsistent (figure 4.53). It was detected at a very low level in UCi\_1 and at a medium level in UCi\_5.

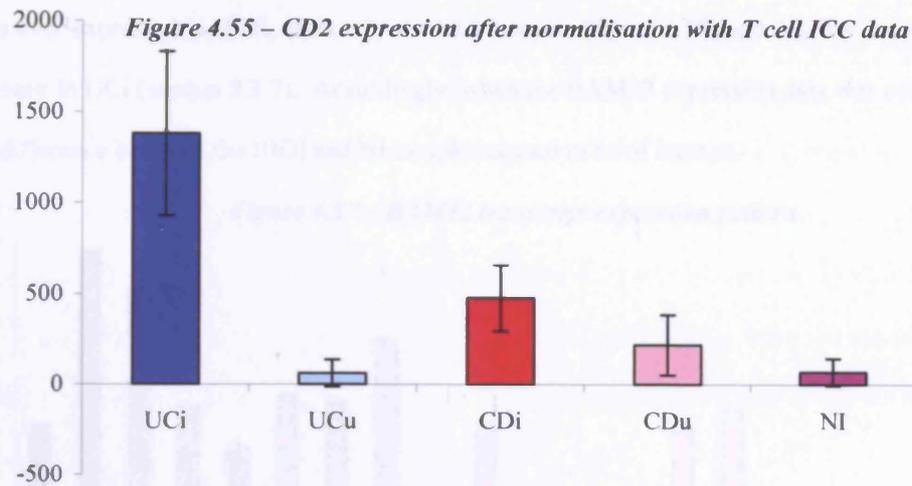


Eotaxin is a secreted chemokine that is chemoattractive specifically for eosinophils, an inflammatory cell type that is a prominent feature of allergic (i.e. Th2 mediated) inflammation<sup>307</sup>. Therefore, eotaxin expression is clearly a secondary inflammatory feature and its over expression in IBD at the mRNA<sup>172</sup> and at the serum protein level<sup>308</sup> is known. The eotaxin expression pattern fits in with the consensus of Crohn's disease as a Th2 disease (section 1.4.2.3). However, the expression of eotaxin in UCi\_1 and UCi\_5 and the very low expression in three of the CDi samples make it clear that Th1/Th2 distinction between Crohn's disease and ulcerative colitis is only a generalisation.

*T-cell surface antigen CD2 (M16336\_s\_at)*

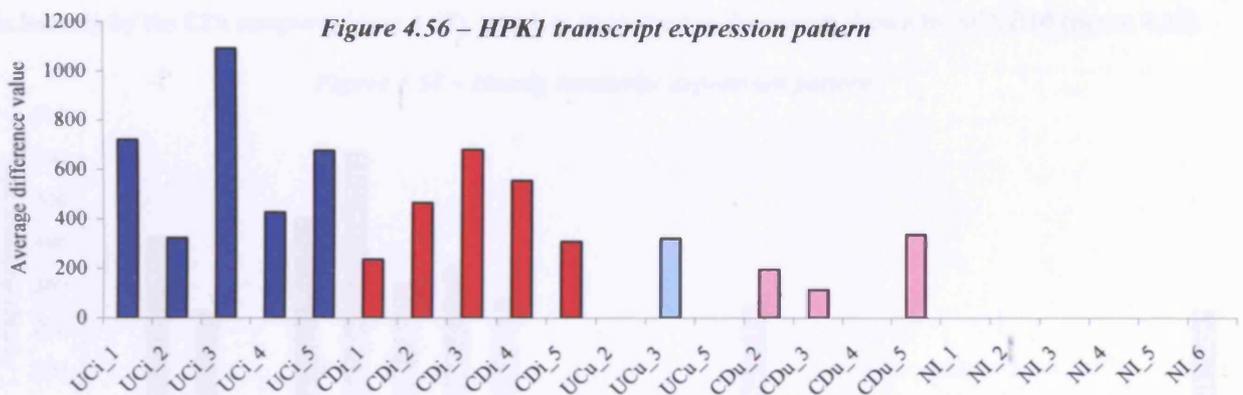
The CD2 receptor is expressed by T cells and binding of the receptor to its ligand (or anti-CD2 antibody) is partly responsible for T cell activation<sup>309</sup>. The current data shows expression of CD2 by 100% of IBDi samples; with a higher expression level in 3 of 5 of the UCi sample (figure 4.54). The over expression of CD2 was seen even after normalising with the T cell ICC data (figure 4.55), indicating actual over expression of the receptor and not simply the presence of an increased number of T cells.

**Figure 4.54 – CD2 transcript expression pattern**



#### *Haematopoietic progenitor kinase (U66464\_at)*

The haematopoietic progenitor kinase gene (*HPK1*) is over expressed in the IBDi samples compared to the IBDu and NI samples (figure 4.56). *HPK1* is part of the SAPK/JNK\* pathway<sup>310</sup>, which is vital to the activation of a number of haematopoietic cells, including macrophages<sup>311</sup> and B cells<sup>312</sup>. Therefore the over expression of *HPK1* in the IBDi tissues is highly likely due to the inflammatory process. It is possible however, that the cellular origin of this transcript is different in Crohn's disease and ulcerative colitis.



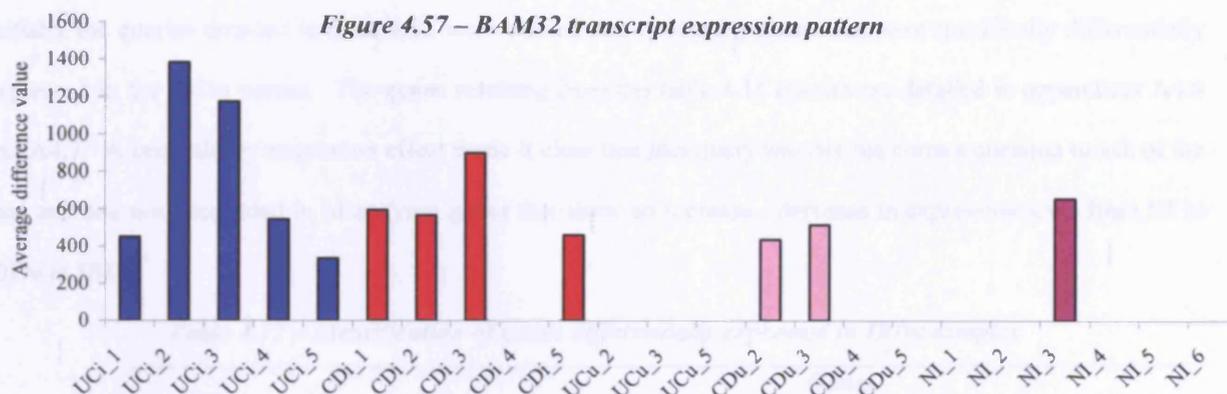
#### *B lymphocyte adapter protein BAM32 (RC\_AA459123\_at)*

*BAM32* (also called *DAPP1†*) was observed in B cell lines, but not in T cell, epithelial or fibroblast cell lines<sup>313</sup>. Additionally, B cell activation increased *BAM32* expression<sup>313</sup> and the protein is implicated in endosomal trafficking<sup>314</sup>. *BAM32* binds the phospholipid messengers, phosphatidylinositol-3,4,5-trisphosphate (PtdIns(3,4,5)P<sub>3</sub>) and phosphatidylinositol-3,4-bisphosphate (PtdIns(3,4)P<sub>2</sub>)<sup>315</sup>, which play roles in a wide range of cell functions, including migration and adhesion<sup>316</sup>.

\* c-Jun N-terminal kinase / stress-activated protein kinase

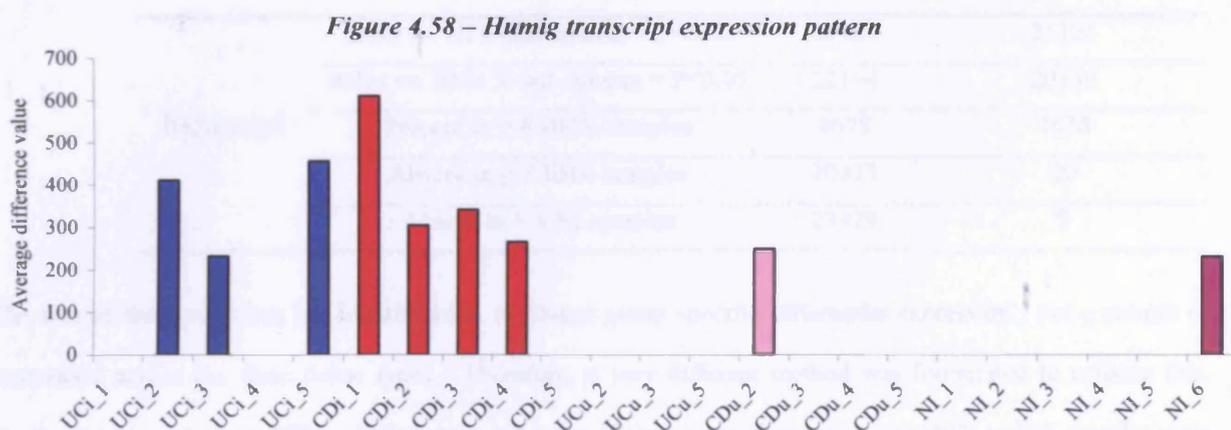
† Dual Adaptor for Phosphotyrosine and 3-Phosphoinositides

*BAM32* is over expressed in IBDi, especially the UCi samples (figure 4.57) and this fits in with the higher B cell numbers seen in UCi (section 2.3.7). Accordingly, when the *BAM32* expression data was normalised to the ICC data, the difference between the IBDi and NI samples ceased to be of interest.



#### *HuMig* mRNA (*X72755\_at*)

This probe set represents the human MIG (Monokine Induced by Gamma interferon); a CXC chemokine closely related to SCYB10 (discussed in section 4.5.4.3). MIG and SCYB10 are both ligands of the chemokine receptor CXCR3 and they have a number of common functions (reviewed in Faber, 1997<sup>317</sup>). MIG is not expressed exclusively by the CDi samples (figure 4.58), which is in contrast to the pattern shown by SCYB10 (figure 4.32).



A difference in expression pattern between MIG and SCYB10 is also shown in experimental viral infections in mice<sup>317</sup>, which suggests that although the effects mediated by the two proteins are similar, they are likely to play non-redundant roles *in vivo*. MIG and SCYB10 are expressed by intestinal epithelial cells<sup>277, 318</sup>, are chemoattractive specifically for activated T lymphocytes<sup>317</sup> and are expressed in the late stages of the inflammatory process<sup>319</sup>. Thus, the expression of *MIG* in the IBDi samples simply indicates late phase inflammation.

#### 4.7 Which genes are specifically differentially regulated between uninvolved IBD and controls?

##### 4.7.1 Rationale and Method

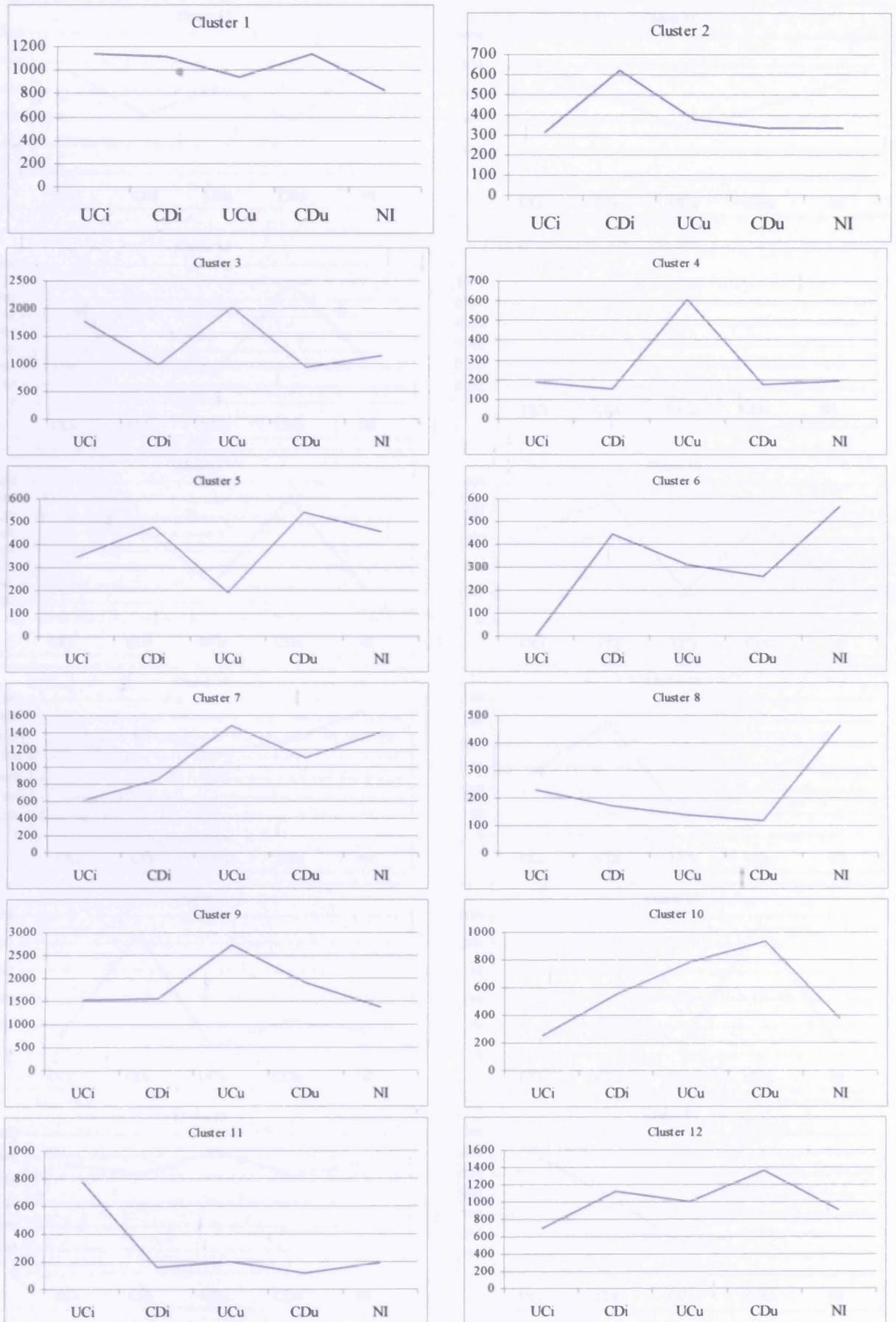
To identify genes that show a 'pre-involved' gene expression, the last query focused on the IBDu samples. Initially the queries detailed in table 4.11 were carried out, to identify genes that were specifically differentially expressed in the IBDu tissues. The genes resulting from the table 4.11 queries are detailed in appendices A4.6 and A4.7. A preliminary annotation effort made it clear that this query was not the correct question to ask of the data and had not succeeded in identifying genes that show an increase / decrease in expression level from NI to IBDu to IBDi.

*Table 4.11 – Identification of genes differentially expressed in IBDu samples*

<i>IBDu expression pattern</i>	<i>Criteria</i>	<i>Genes returned by individual query</i>	<i>Genes returned in cumulative query</i>
<b>Decreased</b>	IBDu vs. NI T-test column = $P < 0.05$	21105	21105
	IBDu vs. IBDi T-test column = $P < 0.05$	22164	20136
	Absent in $\geq 6$ IBDu samples	18702	5215
	Present in $\geq 7$ IBDi samples	7708	54
	Present in $\geq 4$ NI samples	7199	16
<b>Increased</b>	IBDu vs. NI T-test column = $P < 0.05$	21105	21105
	IBDu vs. IBDi T-test column = $P < 0.05$	22164	20136
	Present in $\geq 6$ IBDu samples	4678	4678
	Absent in $\geq 7$ IBDi samples	20373	20
	Absent in $\geq 4$ NI samples	23429	9

The aim of this query was not identification of 'tissue group specific differential expression', but a pattern of expression across the three tissue types. Therefore, a very different method was formulated to achieve this. Firstly, the genes that are differentially expressed in the IBDu tissues compared to the IBDi and NI samples were identified with the T-test columns and, as shown in the first two rows of table 4.11, this returned 20,136 genes. K-means hierarchy (Stanford's version) was applied to the tissue group 'mean average difference values', i.e. five values per gene (UCi, CDi, UCu, CDu, NI) were entered. The genes were organised into 24 nodes containing varying numbers of genes (table 4.12), in a maximum of 1000 cycles. Graphs were plotted of the mean average difference value for each node (figure 4.59) and these were used to select nodes showing patterns of interest. Annotations for the genes of particular interest within the selected nodes are detailed in the sections 4.7.2 and 4.7.3.

Figure 4.59 – K-means clustering of IBDu genes into 24 nodes\*



\* The y-axes in these graphs represent the mean 'average difference values' of the genes in that cluster.

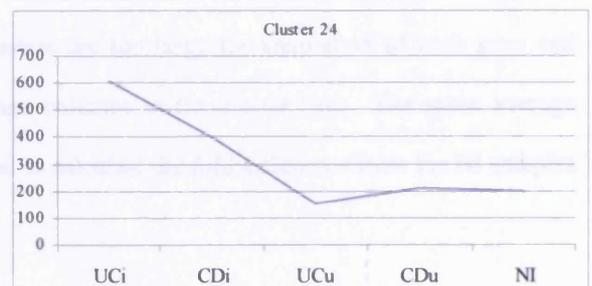
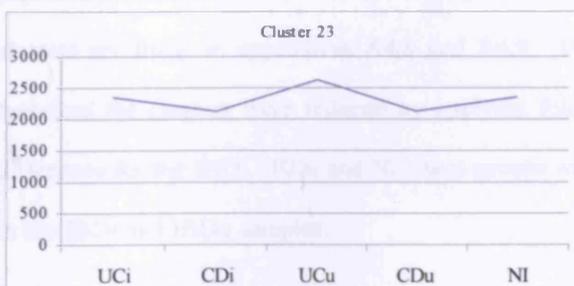
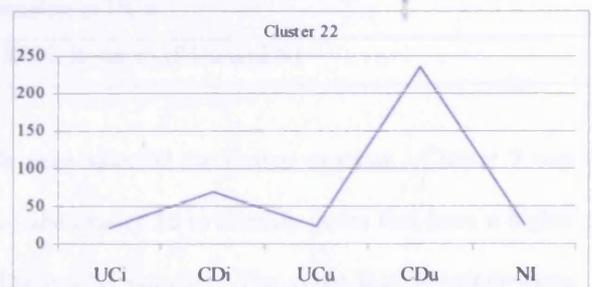
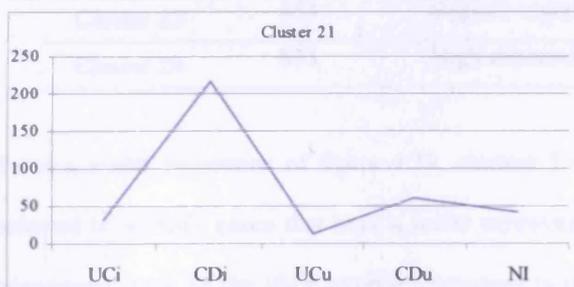
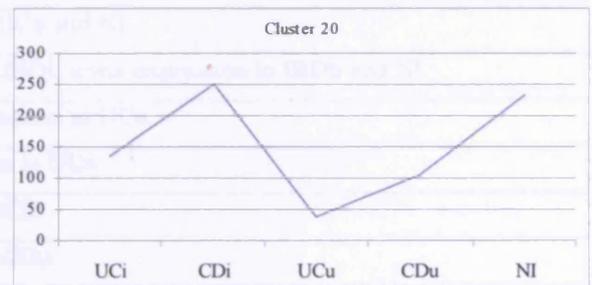
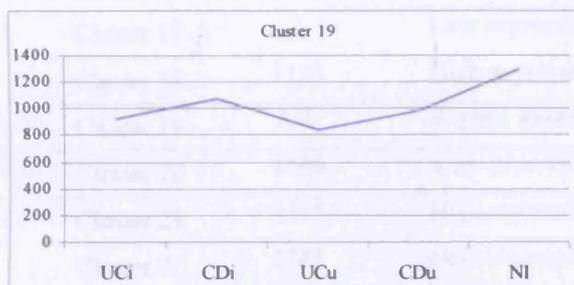
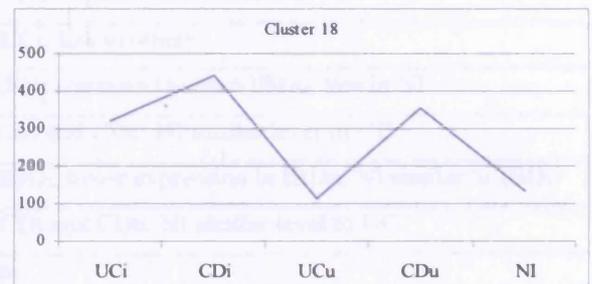
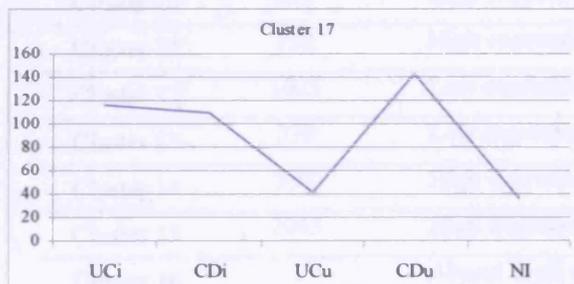
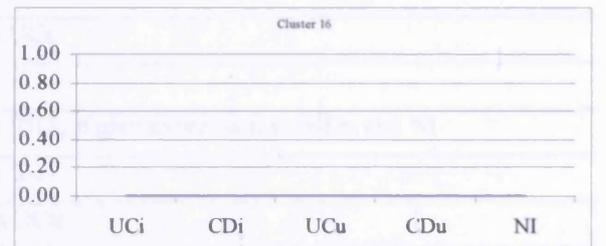
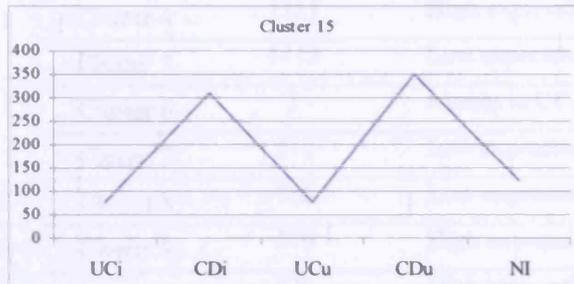
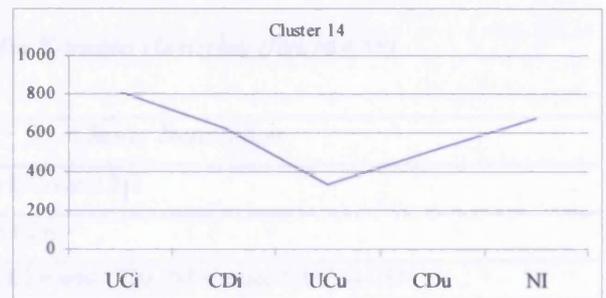
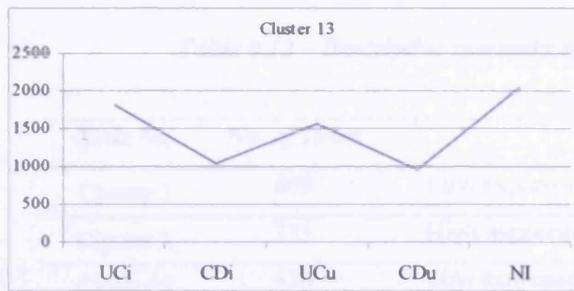


Table 4.12 – Descriptive summary of IBDu K-means clustering (figure 4.59)

<i>Node No.</i>	<i>No. of Items</i>	<i>Cluster Description</i>
Cluster 1	869	Low expression in UCu and NI
Cluster 2	735	High expression in CDi
Cluster 3	424	Low expression in CDi and CDu, NI similar level to CD
Cluster 4	1111	High expression in UCu
Cluster 5	1413	Low expression in UCu
Cluster 6	1	Absent in UCi
Cluster 7	511	Low expression in IBDi, higher expression in IBDu and NI
Cluster 8	986	Low expression in IBD
Cluster 9	579	High expression in UCu
Cluster 10	920	High expression in CDu, medium in UCu, low in NI
Cluster 11	732	High expression in UCi, low in others
Cluster 12	1025	Low expression in IBDi compared to each IBDu, low in NI
Cluster 13	339	Low expression in CDi and CDu, NI similar level to UC
Cluster 14	998	High expression in IBDi, lower expression in IBDu, NI similar to IBDi
Cluster 15	2043	High expression in CDi and CDu, NI similar level to UC
Cluster 16	2	Absent in all samples
Cluster 17	1	Low expression in UCu and NI
Cluster 18	1116	High expression in IBDi, lower expression in IBDu and NI
Cluster 19	838	Slightly lower expression in UCu
Cluster 20	1106	Very low expression in UCu
Cluster 21	1115	High expression in CDi
Cluster 22	1785	High expression in CDu
Cluster 23	654	Slightly higher expression in UCu
Cluster 24	833	High expression in IBDi, lower in IBDu and NI

From a visual inspection of figure 4.59, clusters 7 and 18 were selected for further analysis. Cluster 7 was selected to identify genes that have a lower expression level and cluster 18 to identify genes that have a higher expression level, in the IBDi samples compared to the IBDu and NI samples. The genes that constitute these clusters are listed in appendices A4.8 and A4.9. The clusters are too large for annotation of each gene and therefore the clusters were reduced by applying fold change columns to the cluster lists. The mean average difference for the IBDi, IBDu and NI tissue groups was used to calculate the fold difference from the NI samples in the IBDi and IBDu samples.

#### 4.7.2 Genes with a lower expression level in IBDi tissues (Cluster 7)

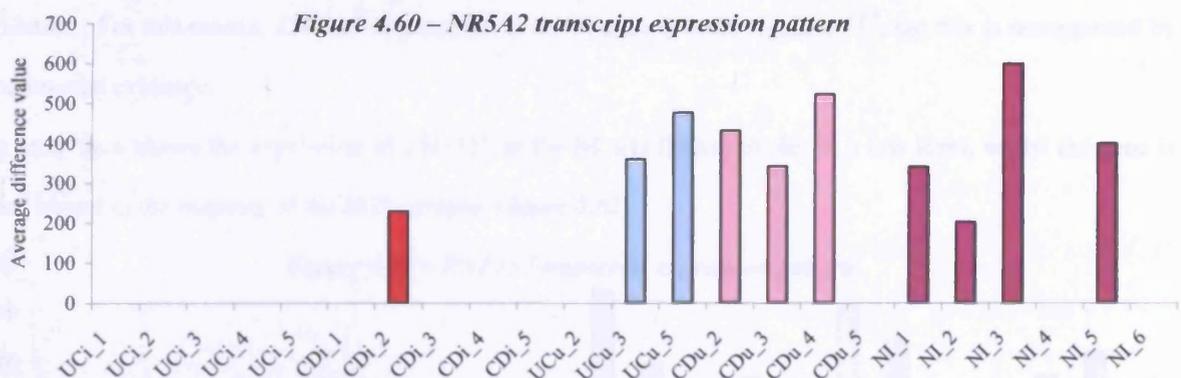
Cluster 7 contains genes that are expressed at higher levels in the IBDu and NI samples compared to the IBDi samples. All the genes clustered to produce the 24 nodes in figure 4.59 showed an interesting difference between IBDi & IBDu and between IBDu & NI tissues. To decrease the number of genes in cluster 7 to a manageable number, the fold change criteria was applied. Genes that showed at least a 5 fold decrease in IBDi compared to NI and a fold change between  $-1.5$  and  $1.5$  in IBDu compared to NI, (i.e. no difference from the NI samples) were selected from the 511 genes in cluster 7. 21 probe sets were returned.

As this criteria does not include those probe sets that may have been called absent in 100% of the IBDi samples, a second query was carried out on the cluster 7 data set. The filters were set to select genes that had a mean average difference of 0 in the IBDi samples (i.e. called absent in 100% of IBDi) and had a fold change between  $-1.5$  and  $1.5$  in the IBDu compared to NI samples. This returned 14 genes.

Therefore, in total 35 genes were selected from cluster 7. Of these, 20 represented EST sequences with no annotation in the public domain, beyond the chromosomal location. From the remaining 15 genes, only 3 showed interesting 'sample' expression patterns. The remaining genes were inconsistently expressed across the IBDu samples and therefore, were not investigated further.

##### *Incyte Unique (RC\_R10138\_at)*

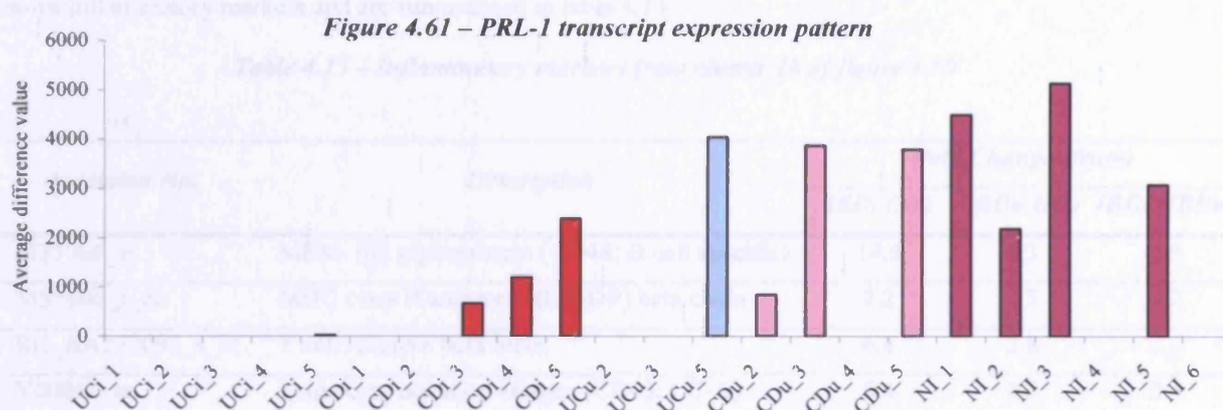
This EST sequence links to nuclear receptor subfamily 5, group A, member 2 (NR5A2). NR5A2 is an orphan nuclear receptor, which activates the  $\alpha$ 1-fetoprotein gene in early differentiating hepatocytes<sup>320</sup>. NR5A2 was called present in 5/7 IBDu and 4/6 NI samples, but in only 1/10 IBDi samples (figure 4.60). The level of expression in the IBDu and the NI samples is similar, with mean average differences of 304.5 and 256.8 respectively. NR5A2 has not been studied in IBD previously.



##### *Protein tyrosine phosphatase PTPCAAX1 (RC\_T40895\_at)*

Protein tyrosine phosphatases (PTPs) hydrolyse the phosphate monoesters of tyrosine residues, share a common active site motif and are classified into 3 groups (receptor-like PTPs, intracellular PTPs and dual-specificity

PTPs). PTPCAAX1 (PRL-1) is a member of a fourth class, along with PRL-2 and PRL-3, two closely related genes<sup>321</sup>. PRL-1 has not been investigated in IBD previously, but is known to be expressed by intestinal<sup>322</sup> and gastric<sup>323</sup> epithelium and in T cells<sup>324</sup>. Specifically PRL-1 protein is expressed in terminally differentiated cells in the colon<sup>322, 323</sup>. A recent study has shown that PRL-1 binds a novel bZIP protein ATF-7<sup>325</sup>, which is an activating transcription factor<sup>326</sup>, whose target gene/s are yet to be determined.



The current study shows the expression of PRL-1 in the NI and IBDu tissues and in 3 of the CDi and none of the UCI samples (figure 4.61). The levels of expression range from medium (665 in CDi\_3) to high (5146 in NI\_3). The gene is called absent in 100% of the UCI samples. The decreased expression of *PRL-1* in IBDi tissues (especially UCI) is clear, but the significance of this will remain unclear until the downstream effects of *PRL-1* under expression are deduced.

#### Zinc finger protein ZNF137 (*RC\_AA043458\_s\_at*)

The zinc finger domain is a DNA binding domain and therefore zinc finger proteins tend to be transcriptional regulators. For this reason, ZNF137 is proposed to be a transcriptional regulator<sup>327</sup>, but this is unsupported by experimental evidence.

The array data shows the expression of ZNF137 in the NI and IBDu samples at a low level, whilst the gene is called absent in the majority of the IBDi samples (figure 4.62).



#### 4.7.3 Genes with a higher expression level in IBDi tissues (Cluster 18)

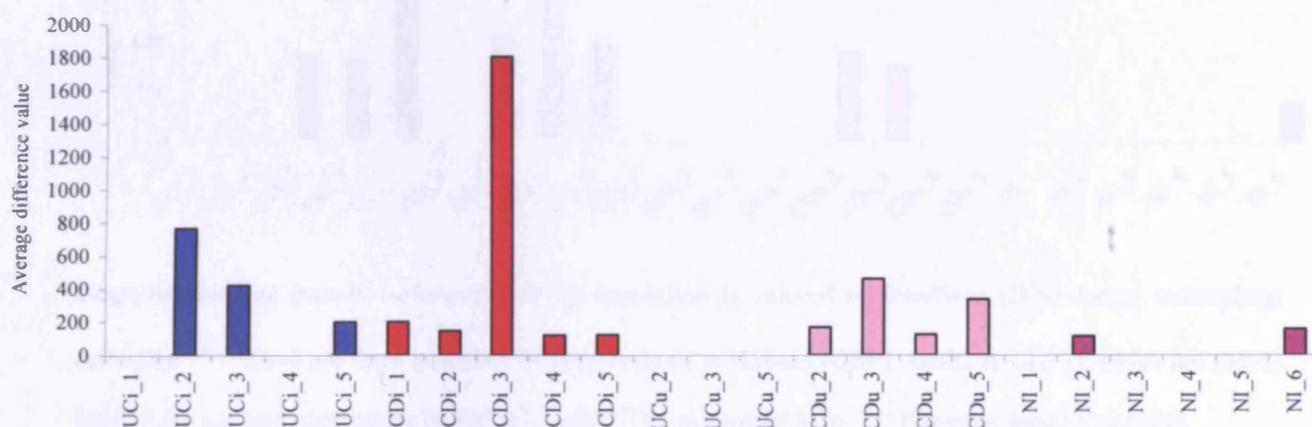
To identify genes that are up regulated gradually from NI to IBDu to IBDi, the fold change columns were applied to cluster 18. 23 genes were selected from the cluster 18 data set, that had an increase of, at least 6 fold in IBDi compared to NI), at least 2 fold in IBDi compared to IBDu) and at least 2 fold in IBDu compared to NI. Of these 23, 10 showed consistent expression patterns in the IBDi and IBDu samples. Five of these are well known inflammatory markers and are summarised in table 4.13.

Table 4.13 – Inflammatory markers from cluster 18 of figure 4.59

Accession No.	Description	Fold Change (from)		
		IBDi (NI)	IBDu (NI)	IBDi (IBDu)
M37766_at	MEM-102 glycoprotein (CD48; B cell specific)	14.5	6.3	2.3
M57466_s_at*	MHC class II antigen (HLA-DP) beta chain	7.2	2.3	3.2
RC_AA257093_s_at	T cell receptor beta locus	6.4	2.8	2.3
Y00062_at	Leukocyte common antigen (CD45, LC-A)	8.4	3.5	2.4
X58529_at	IgM heavy chain	9.1	2.7	3.3

Regenerating protein (*J05412\_at* & *L08010\_at*)

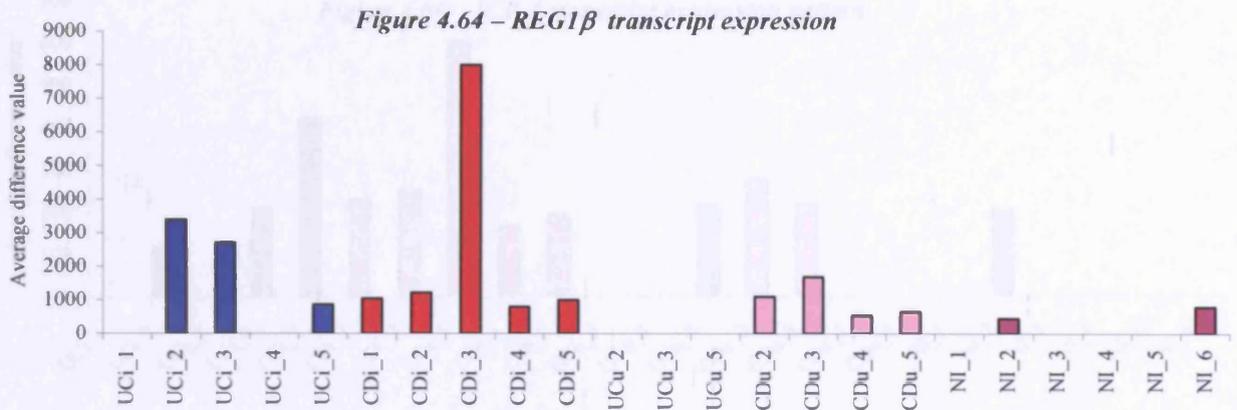
Figure 4.63 – *REG1α* transcript expression



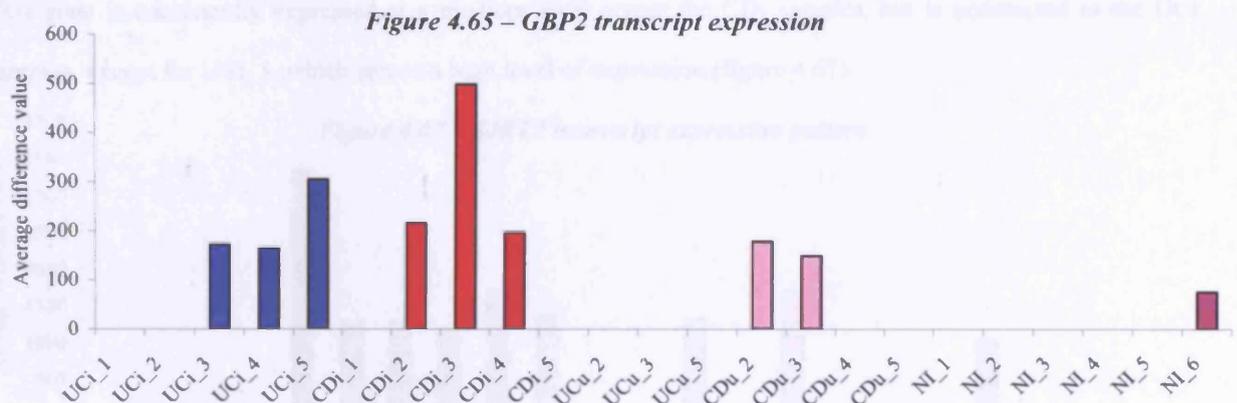
The regenerating protein (REG) family consists of four members to date<sup>328</sup>, two of which are represented on the arrays and have been identified in this query. *REG1α* and *REG1β* are over expressed by the involved IBD tissues. The expression pattern for both genes is very similar across the samples, although the level of expression of *REG1β* is much higher (figures 4.63 & 4.64).

*REG1α* is over expressed in regenerating, compared to normal, pancreatic islet  $\beta$  cells and is widely believed to inhibit the formation of calcium carbonate stones in the renal and pancreatic ducts. The expression of REG transcripts has also been associated with early death from colorectal carcinoma<sup>329</sup>.

The over expression of both REG1 $\alpha$  and REG1 $\beta$  in involved IBD has been noted in two previous microarray studies<sup>180, 181</sup>. The over expression of another REG mRNA, *REGIV*, in IBDi tissues has also been noted previously<sup>328</sup>. It is plausible that these genes are being induced in response to the mucosal degeneration that occurs during IBD pathogenesis. The difference in expression of the REG genes between the UCu and CDu samples is interesting, but is inexplicable from data in the public domain.



Guanylate binding protein isoform II (*RC\_AA487252\_at*)



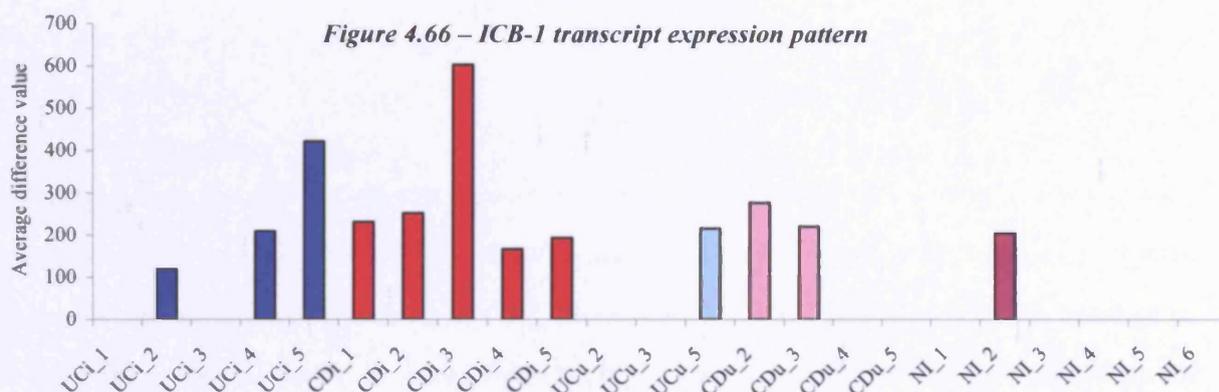
Guanylate binding protein isoform II (*GBP2*) expression is induced by interferon (IFN) during macrophage activation<sup>330</sup>. There are three guanylate binding proteins in humans (GBP1, GBP2 & GBP3), which are able to bind all the guanine nucleotides (GMP, GDP and GTP), in contrast to the GTP specific binding proteins.

The array data shows that GBP2 is expressed in 60% of IBDi compared to 29% of IBDu and 17% of NI samples (figure 4.65). The expression of GBP2 probably simply reflects the inflamed nature of the samples, but the precise function of GBP2 remains unknown. GBP2 is only expressed in 2 IBDu samples and thus does not represent a true 'gradually increasing' gene.

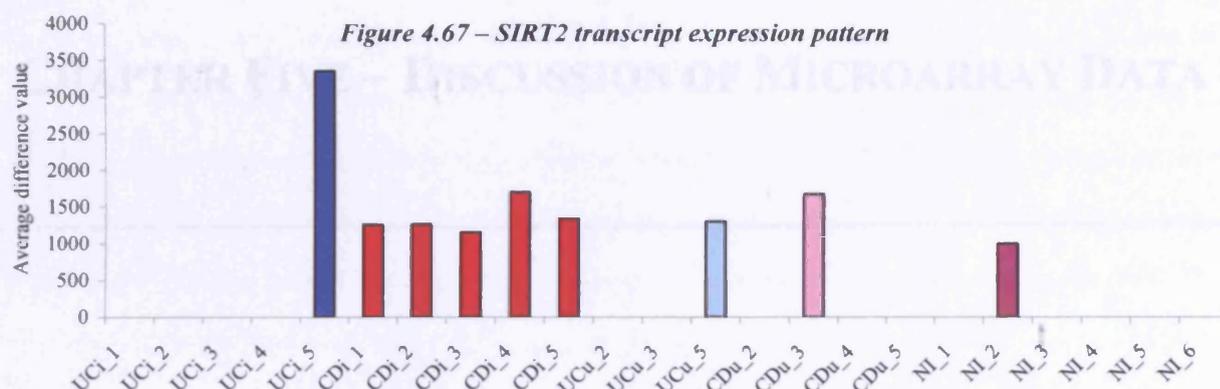
\* This gene was expressed at a 9-fold greater level in UCI tissues compared to controls in a previous microarray study<sup>181</sup>.

*ICB-1 mRNA (T59929\_s\_at)*

The observation that *ICB-1* is over expressed in involved IBD tissues (figure 4.66) is novel. *ICB-1* was originally identified in human endometrial adenocarcinoma<sup>331</sup>, but further studies investigating the function of this gene have not been reported. The gene is present in only a few of the CDu and UCu samples, so this gene expression pattern is not of interest in terms of the original query.

*Silencing information regulator 2-like protein (W32305\_f\_at)*

This gene is consistently expressed at a medium level across the CDi samples, but is undetected in the UCI samples, except for UCI\_5, which shows a high level of expression (figure 4.67).



*SIRT2* is part of the sirtuin family of proteins (*SIRT1* to *SIRT5*)<sup>332</sup>, which are conserved from prokaryotes to eukaryotes. Studies on the human sirtuins suggest that they may function as intracellular regulatory proteins with mono-ADP-ribosyltransferase activity. Recombinant human *SIRT2* is able to transfer <sup>32</sup>P radioactivity from NAD to bovine serum albumin, but this is blocked by converting a conserved histidine residue to tyrosine<sup>332</sup>. It is not clear why this gene is expressed almost exclusively in the CDi samples.

**4.7.4 Section summary**

Apart from the inflammation related genes, the remainder cannot be demonstrated to have any relevance to IBD pathogenesis. This maybe because the IBD samples represent end stage disease and although the IBDu samples appear normal at the macroscopic and microscopic levels, there maybe changes at the transcript level.

**CHAPTER FIVE – DISCUSSION OF MICROARRAY DATA**

## **5.1 General Discussion - Study Limitations**

---

This thesis has described the analysis of the gene expression patterns of two related but distinct diseases. It also describes the evaluation of commonly available gene expression analysis tools and the results of applying them to the data generated. At the start of this project the technology and the analysis methods used were not in routine use and therefore a brief discussion of the limitations of such a study is warranted.

### **5.1.1 Sample Issues**

#### **5.1.1.1 Collection**

With mucosal sample collection, important considerations include the inadvertent inclusion of submucosa and the unknown extent of ulceration. For the control samples this was not a problem as the injection of saline underneath the mucosa was very efficient at lifting the mucosa away from the underlying layers. However for the IBD tissues, especially where the mucosa was very fibrotic, the mucosa did not always lift up cleanly and these samples could not be used. In severely ulcerated colon specimens, sections of submucosa were exposed to the lumen as the mucosa had been shed and these areas were avoided for sample collection. All samples were checked by routine histopathology to ensure the absence of submucosa before being processed for array hybridisation.

For future projects using colonic mucosa microdissection<sup>333</sup> or the use of a biopsy claw would be considered. Using the biopsy claw would ensure that only a thin layer of mucosa was sampled from the resected colon. However, separate biopsies would have to be taken for histopathology and ICC procedures. Microdissection would ensure that only the cells of interest were included in the starting material for array hybridisation studies. Whilst the biopsy claw would be relatively easy to access, specialist tools are required for microdissection. The amount of tissue generated by either method would have been insufficient for use with the GeneChip system as described in this study. Recent publications have described the linear amplification of mRNA for use in microarray expression studies of very small amounts of tissues. For example, the gene expression profiling of individual *Drosophila* imaginal discs has been achieved using an RNA amplification method<sup>334</sup>. This study also utilised RNA amplification in the reduction of the starting amount of total RNA from 700 µg to 40 µg (section 2.2.4); however, the amount of amplification achieved was insufficient for the use of biopsy material. A future study could investigate the adaptation of the RNA amplification protocols described recently for the use of colonic biopsies in microarray studies.

#### **5.1.1.2 Variability**

The use of tissues rather than pure cell cultures in this study meant that the cellular make-up of each tissue had a bearing on the results. Each cell type would have a distinct gene expression pattern and if there were more of one cell type in a particular sample this could heavily bias the final gene expression pattern of that sample. IBD

encompasses a large group of pathologically distinct conditions and even within the subgroups of Crohn's disease and ulcerative colitis there is a large variation of types. As this study was mainly focused on differences in epithelial gene expression the influence of other cell types had to be accounted for. The ICC data (section 2.3) was used to normalise the array data wherever it was clear that a particular cell was responsible for the expression of a particular gene. This was a rather crude method of trying to isolate only the gene expression patterns of the epithelial cells and with hindsight was not very successful due to the cellular origin of most genes being unknown. If subsequent studies were to be carried out *in situ* hybridisation for the gene (or ICC for the protein) of interest would be performed to determine cellular origin more precisely. This would require a thorough knowledge of colonic mucosal histopathology to accurately identify which cells were producing the gene or protein of interest.

### 5.1.2 Microarray Issues

Issues of sensitivity, specificity and reproducibility are important in any scientific study and apply equally well to microarray experiments. As the array technology used was a commercial system some of these issues are not directly influenced by the experimenter, but by the manufacturer. The GeneChip system included array generation, hardware for array hybridisation and software for image analysis.

#### 5.1.2.1 Specificity

At the time the arrays used here were designed, the entire human genome had not been sequenced and thus, much of the genome may have been omitted in this study. However, until the final draft of the genome has been published and all ESTs represented on the arrays assigned to a full-length gene, it is not possible to estimate how much of the genome was analysed in this study.

The specificity of the probes to the genes that were represented was entirely determined by the manufacturer. Lists of the probe sequences were available, but with the large number of genes profiled it was not practical to check the specificity of each one. However, for the 70 genes that showed differential expression between the CDi and UCi samples the probe sequences were checked. In all cases the perfect match probe sequence retrieved the correct gene sequence from the GenBank database.

#### 5.1.2.2 Sensitivity

The sensitivity of the GeneChip system as measured with a pure cell culture has been calculated between 1 and 10 mRNA copies per cell<sup>192, 335</sup>. Bertucci *et al* calculated that the GeneChip system is capable of detecting mRNA at 1 in 300,000 transcripts<sup>335</sup> and Kane *et al* found no difference between the sensitivities of PCR probes on glass slides and the oligonucleotide probes on GeneChip arrays<sup>192</sup>. Investigations into the sensitivity of the GeneChip using whole tissues have not been performed. It is clear that the amount of starting material would greatly influence the sensitivity of the detection system. In this study care was taken to ensure that an equal

amount of total RNA was applied to each array. Each sample was hybridised to each of the five arrays in a specific order to further ensure that comparisons between samples for each probe set would be as unbiased as possible.

Rare transcripts at 1-2 copies per cell are at the limit of detection and the apparent expression level may not be accurate. In this study average difference levels of less than 20 were negated. It is therefore possible that some very rare transcripts have been overlooked, but it was felt that false negatives were preferable to false positives.

#### ***5.1.2.3 Reproducibility***

For cDNA spotted arrays the issue of consistent array generation is very important. If spot sizes are too varied or if the exact location of each cDNA is not consistently known the results become meaningless. With the GeneChip arrays quality is less in the control of the experimenter. Each array was used within its 6-month expiry date and the batch number of each array used was noted in case the company became aware of any irregularities with a particular batch in the future.

The array technology used in this study was very costly and funding was not available for inter-chip reliability experiments. However, the accepted strategy is to only consider fold-changes of greater than 2 for cell cultures and 3 for tissues as significant. In this study fold change was generally not used as a measure of differential expression, but where it was used only expression changes greater than 3 fold were considered significant.

#### ***5.1.2.4 Variability***

The number of samples and arrays used to produce the data presented in this thesis meant that it would have been impossible to run all the hybridisations simultaneously. Variability can be introduced at any of the stages and therefore it was important that the quality of each array could be relied upon and that each step of the procedure introduced as little variation as possible.

### **5.1.3 Mining Issues**

There are two main aims of gene expression analysis: the identification of groups of genes with similar expression patterns and the identification of genes that have different levels of expression in pathologically distinct diseases. These two aims are related in that the same tool can be used to identify both patterns, but generally clustering tools are used to identify groups of genes, whilst matrix-based methods tend to be better for identifying individual genes. The fact that there are gradients of gene expression with cells and tissues expressing varying levels of any particular mRNA means that the gene expression profile of a tissue or cell is highly complex. Although the tools used in this study are entirely typical of gene expression techniques available at the time of writing they are still very basic considering the complexity of the cell. Only very rudimentary statistics and mathematics were applied to the data in this thesis and with a more intricate analysis subtler gene expression patterns may emerge.

#### 5.1.4 Issues specific to this Study

##### 5.1.4.1 Sample size

One of the most important confounding factors was the small sample set. This problem is exemplified by the similarity of the pathological features of UC and CD. The two diseases may have a single aetiology, but it seems increasingly the case that a diagnosis of ulcerative colitis at least may actually cover a range of similar diseases. In the small sample set used in this study it is virtually impossible to find any general gene expression patterns that describe ulcerative colitis or Crohn's disease. The identification of one ulcerative colitis sample as being unequivocally 'different' to the others is testimony to this difficulty. The only way to overcome the small sample set would have been to increase the number of samples arrayed. As this was not possible due to the cost of the arrays the small sample set remains a major limitation of this study.

##### 5.1.4.2 Type II errors

As discussed above the microarray data analysis was weighted towards generating false negatives rather than false positives. However, with a 95% threshold of significance in the T-test analyses, the increasing number of T-tests increases the probability that some genes identified as having a significant expression pattern would be selected by chance. For example, in cases where two samples (or two tissue groups) were being compared, a single T-test was used to determine whether the expression of the gene in the first sample was significant compared to its expression in the second samples. 5% of the returned genes in this case would be due to chance, i.e. a type II error would be responsible for the selection of these genes as being significantly expressed. Given the nature of the data there is no complete way to resolve this problem. Grouping the genes into those that were significantly differentially expressed in sample 1 vs. sample 2 AND significantly differentially expressed in sample 1 vs. controls, aimed to find truly significant differentially expressed genes and reduce the number of type II errors. Therefore, the 70 genes identified as being significantly differentially expressed in CDi and UCi when gene expression analysis was carried out between UCi vs. CDi vs. controls are more likely to be truly differentially expressed than genes selected on the basis of a single comparison, for example CDi vs. controls. Considering groups of related genes that are all differentially expressed is another way to reduce the chances of being misled by type II errors. However, it must be remembered that the use of the T-test in this way does not indicate statistical significance, and only serves as a way of identifying potentially interesting gene expression patterns.

##### 5.1.4.3 Use of steroid refractory samples

Another limitation of this study is the fact that all the samples represented steroid-refractory disease. This meant that none of the gene changes detected would be those seen in the early stages and are likely to represent late or secondary changes. Using biopsies from patients with early-stage IBD as the starting material would be more likely to yield any primary gene changes. However, biopsies extra to those taken for clinical purposes would be required and due to the amount of starting material needed several would have to be taken especially for this

purpose. Additionally, colonoscopy is usually carried out in order to determine the diagnosis and for this study it was necessary that a diagnosis of either Crohn's disease or ulcerative colitis be known in advance so that arrays were not used unnecessarily.

#### ***5.1.4.4 Cellular make-up of samples***

If a gene is called present in a sample it has a high likelihood of being expressed, whereas for a gene that has been called absent the possibility of expression, albeit at a low level, remains. The generation of false negatives is unavoidable as there is always a limit to detection sensitivity with any system. Therefore it can be assumed that a gene that is called present in a particular sample was being expressed at the time of collection. However, when it comes to correlating gene expression patterns with each tissue group the same assumption cannot be made. Confounding factors such as the different proportion of cell types in each tissue and the small sample size could lead to correlations wrongly being observed as significant. As discussed above an attempt was made to correct for the different proportion of cell types in each sample.

#### ***5.1.4.5 Gene expression to protein function***

Determining the protein expression is important in determining the actual biological relevance of the gene expression patterns seen in the samples. Levels of mRNA do not always correlate to levels of protein expression<sup>336</sup>. Post-translational modifications are an important aspect of final protein function, but are notoriously difficult to investigate and identifying the human proteome is the next major challenge in biochemistry. This means that the gene expression patterns described here for Crohn's disease and ulcerative colitis are not necessarily reflective of the protein expression in such tissues. Thus, whilst microarray analysis of gene expression does not provide any answers as to the cause and effect of gene regulation and protein function it does generate multiple novel hypotheses for future studies. Protein expression studies such as ICC or western blotting for the genes of interest would be the next step in a project aiming to further characterise the proteins that contribute to IBD pathology, whereas this project focused on the gene expression experiment itself with particular emphasis on the bioinformatic tools available to analyse such data.

## 5.2 General Discussion – Microarray Data

---

The genes discussed in chapter four represent those with distinctive expression patterns in two or more tissue groups. Thus the assumption is that expression of those genes is dependant on the pathological state of the tissue. This assumption is not necessarily accurate for each gene identified. The microarray technique is useful for the elucidation of correlations, but does provide any clues as to the cause of altered expression of a gene in one sample compared to another or the effect of that altered expression. The cause and effect of a gene's expression pattern can only be shown with studies focused on the elucidation of the function and regulation of the gene of interest.

Microarrays have an obvious potential as diagnostic tools. If enough samples could be profiled in each sample group the sheer number of genes included in each profile would mean that the expression patterns of a few key genes in each sample group would become evident fairly quickly. This gene expression 'fingerprint' would be unique to each sample group and therefore the classification of an unknown sample into one of the groups based only on its gene expression profile would be possible - as has been done in B-cell lymphoma<sup>208</sup>. The use of the microarray as a diagnostic tool for the classification of IBD samples was one of the potential outcomes of this study. However, the gene expression profiles of CD and UC did not prove to be as different as expected and with the limited number of samples, the experiment to try to classify a sample based only its gene expression profile was not done. It would be interesting to see in a future study if an unclassified IBD sample could be classified as CDi or UCi based only on the expression of the 70 genes found to be differentially expressed in CDi and UCi.

Literally thousands of novel hypotheses can be generated in each microarray experiment. Some of the gene-specific hypotheses generated by this study were briefly discussed in chapter four. Two of the areas emanating from the microarray data were investigated further (and are discussed in the following sections) as the microarray data seemed to fit with previously published data and these areas are of great interest in the IBD field. The first hypothesis was originally published by William Roediger in 1980 and concerns the theory that ulcerative colitis is the result of an energy deficient mucosa<sup>337</sup>. The literature in this area is reviewed with respect to the microarray data, which shows the decreased expression of two sulphate transporter genes in the involved IBD samples. The second area concerns the over-expression of cancer related genes in ulcerative colitis and apoptosis related genes in Crohn's disease. These are examples of where the microarray data show a correlation between the expression of certain genes and the tissues they were over-expressed in; however, upon further investigation the association is not as significant as initially appeared. The discussion is only based on previously published data and supporting experiments have not been carried out for this study. However, it was felt that to analyse the data set with a global view would be a worthwhile exercise and would provide more of an overall insight into the gene expression pattern of IBD.

### 5.2.1 Ulcerative Colitis Energy Deficiency Theory

The theory that ulcerative colitis may be caused by an energy deficient mucosa was first proposed by Roediger in 1980. Based on his observations that (i) colonocytes (colonic epithelial cells) prefer to utilise butyrate as an energy source and (ii) butyrate oxidation is diminished in ulcerative colitis affected mucosa, he theorised that diminished butyrate oxidation results in an energy deficient mucosa<sup>337</sup>. The original observations have since been augmented; it has been shown that the distal colon prefers to utilise a limited supply of butyrate over an unlimited supply of glucose<sup>338</sup> and the diminished ability of inflamed mucosa to oxidise butyrate has been shown by many<sup>337, 339-341</sup>, although this latter finding is not universal<sup>342-344</sup>. The original theory proposed that the failure of butyrate oxidation was a primary cause of ulcerative colitis. However, as diminished butyrate oxidation is not a universal finding it is now thought unlikely that it represents a primary feature of ulcerative colitis.

The suppression of the enzymes involved in butyrate metabolism is believed to be mediated by sulphides, especially hydrogen sulphide (H<sub>2</sub>S)<sup>345-347</sup>. Sulphides are present in the colon as the result of sulphate fermentation by the gut flora. Levels of luminal sulphate would be regulated by the sulphate transporters expressed by the colonocyte; however, the microarray data (figures 4.47 to 4.49) and previous studies<sup>170, 181</sup> indicate that the gene expression of two sulphate transporters, DRA and *DTDST*, may be decreased in IBD affected mucosa. Mutations in both transporter genes have been linked to other diseases<sup>170, 348-353</sup> and they both transport a number of different anions<sup>304, 354-358</sup>. However, this discussion is based on their common function of sulphate transport and the following section aims outline a mechanism by which an energy deficient state may potentially be linked to the observed decrease in the gene expression of colonocyte sulphate transporters.

### 5.2.2 Colonic Butyrate

Colonic flora produce short chain fatty acids (SCFA) such as butyrate by fermenting dietary and endogenous carbohydrates that have escaped digestion in the small intestine. As discussed above butyrate is a major energy source for colonocytes<sup>337</sup> and is metabolised via the  $\beta$ -oxidation pathway to acetyl-CoA, which can then be used to generate ATP via the citric acid cycle and oxidative phosphorylation. Recent studies have uncovered several other vital roles played by butyrate in the colonic mucosa including the induction of mucin secretion in the rat colon<sup>359</sup>, maintaining a balance between growth, differentiation and apoptosis<sup>347, 360</sup> and preventing excess inflammation by inhibiting NF- $\kappa$ B<sup>361, 362</sup>. Thus butyrate is a vital nutrient for colonocytes with functions far beyond that of a simple metabolic substrate.

### 5.2.2.1 Butyrate Oxidation in Ulcerative Colitis

One of the initial findings of the current study was the general decrease in the expression of genes encoding enzymes that are part of various metabolic pathways in the UCI samples (detailed in appendix A5.1). A broad repression of metabolic pathways in UCI tissues has also been noted in a previous microarray study<sup>181</sup>. The general decrease in metabolic enzymes could be due to the drugs indicated for ulcerative colitis. However, previous studies in colonocytes from IBD patients<sup>337</sup> and in animal models of colitis<sup>338, 340</sup> have shown a decrease in the activity of one particular metabolic pathway, the fatty acid  $\beta$ -oxidation pathway, which is necessary for the utilisation of butyrate and other SCFAs. The theory is that the decreased activity of the  $\beta$ -oxidation pathway may result in an energy deficient state in the distal colon of ulcerative colitis patients<sup>337</sup>.

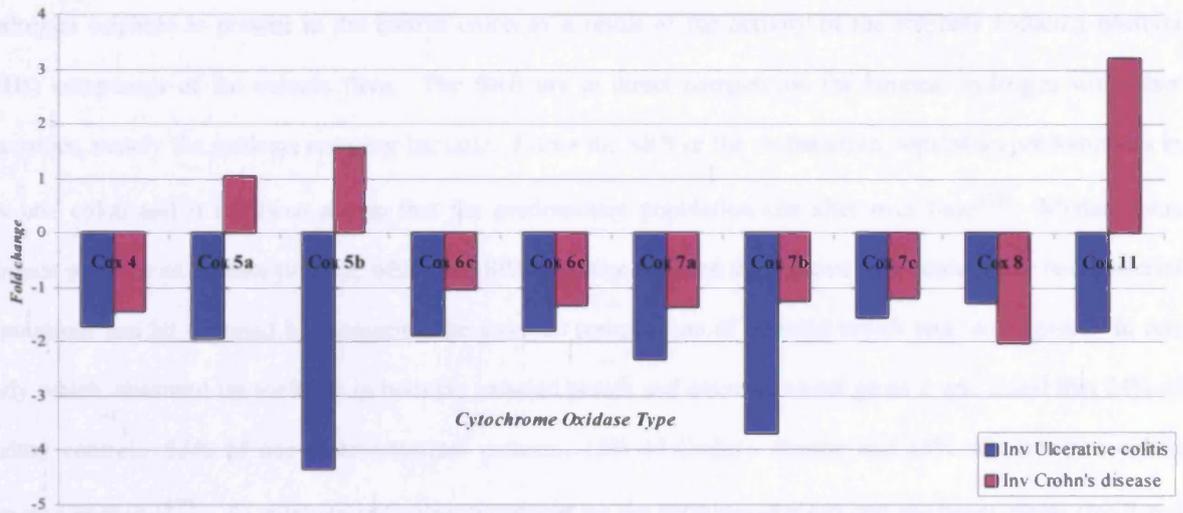
### 5.2.3 Hydrogen Sulphide as an Inhibitor of Butyrate Oxidation

It has previously been proposed that hydrogen sulphide may have an etiologic role in ulcerative colitis via its inhibition of the  $\beta$ -oxidation pathway<sup>363</sup>. It is known that hydrogen sulphide inhibits a vital component of the  $\beta$ -oxidation pathway, short-chain acyl-CoA dehydrogenase<sup>364-366</sup> and that it is also able to inhibit cytochrome c oxidase in the oxidative phosphorylation pathway<sup>367</sup>.

Studies suggest that the amount of H<sub>2</sub>S produced by UC patients is greater than that seen in healthy subjects as measured by faecal concentrations<sup>346, 368, 369</sup>. It certainly seems possible that the healthy colon absorbs (and presumably detoxifies) the majority of sulphides produced by the colonic flora. In a study that measured the amount of sulphide in rat colon, less than 1% was found to be in the form of free sulphide<sup>370</sup>. The study also reported the absorption of potentially detrimental levels of sulphide by healthy rat colon, which would have to be detoxified to avoid local and systemic injury. If sulphide detoxification systems were diminished for any reason, then presumably mucosal injury would occur with 'normal' amounts of H<sub>2</sub>S. Many detoxification systems have been studied in ulcerative colitis<sup>371-374</sup>, but which of these have a role to play in UC aetiology remains to be determined.

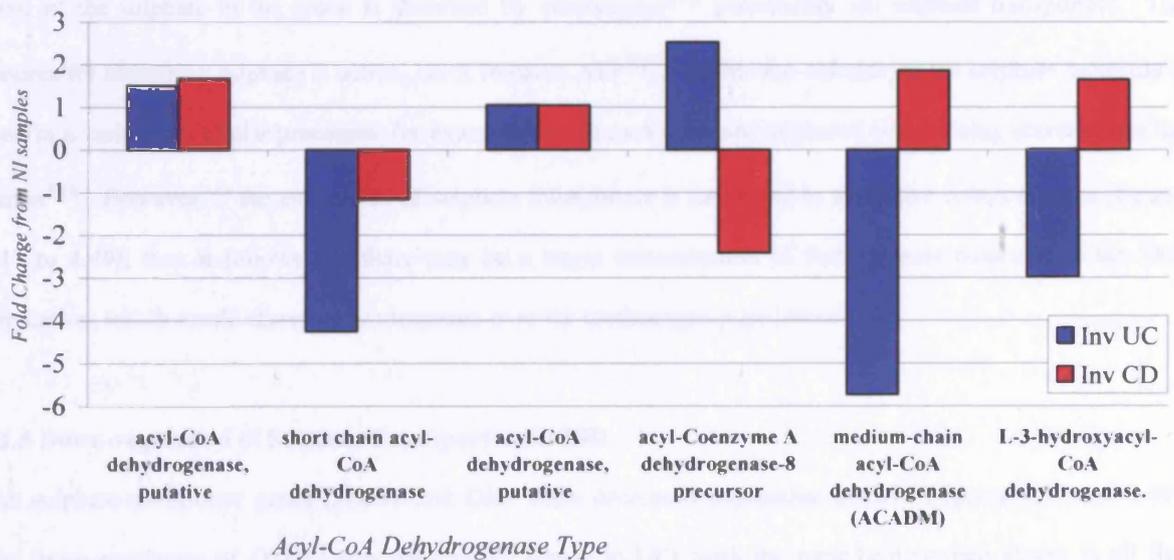
If hydrogen sulphide were to cause decreased expression of genes coding for  $\beta$ -oxidation pathway enzymes in ulcerative colitis mucosa *in vivo*, then the expression of short-chain acyl-CoA dehydrogenase and cytochrome c oxidase would show as being decreased compared to the controls and CDi samples on the array data. A number of cytochrome c oxidase (COx) and acyl-CoA dehydrogenase genes were represented on the arrays (figure 5.1 and 5.2 respectively). The COx genes were generally decreased in all IBDi tissues however, whilst the genes were consistently decreased in UCI, this was not seen in the CDi tissues.

Figure 5.1 - Fold Change in COx gene expression in IBDi compared to NI samples



There was only one short chain acyl-CoA dehydrogenase gene on the arrays; the rest represent different forms of the enzyme. Expression of the short chain acyl-CoA dehydrogenase was decreased in both CDi and UCI compared to the NI samples, with a larger fold decrease in the UCI samples. This data is consistent with the hypothesis, but does not provide direct evidence for H<sub>2</sub>S causing the decrease in the expression of cytochrome c oxidase and acyl-CoA dehydrogenase genes in ulcerative colitis.

Figure 5.2 - Fold Change in Acyl-CoA dehydrogenase gene expression in IBDi compared to NI samples



If the progression of ulcerative colitis is dependent on the inhibition of these metabolic pathways, then these data also imply that the disease mechanism of Crohn's disease may differ from ulcerative colitis in this regard. In support of the data shown here previous studies report decreased butyrate oxidation in ulcerative colitis, but not in Crohn's disease<sup>337, 339, 375</sup>.

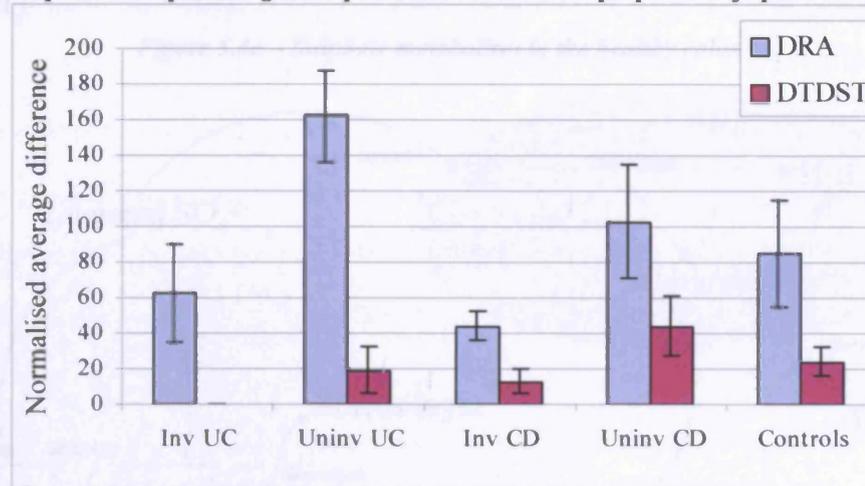
#### 5.2.4 Sulphate Reducing Bacteria

Hydrogen sulphide is present in the human colon as a result of the activity of the sulphate reducing bacteria (SRB) component of the colonic flora. The SRB are in direct competition for luminal hydrogen with other anaerobes, mainly the methane reducing bacteria. Either the SRB or the methanogen population predominates in any one colon and it has been shown that the predominant population can alter over time<sup>376</sup>. Methanogens produce methane as a waste product, whilst the SRB produce H<sub>2</sub>S and the relative magnitude of the two bacterial populations can be assessed by measuring the gaseous composition of exhaled breath and rectal gases. In one study which measured the methane in both the exhaled breath and excreted rectal gases it was found that 54% of healthy controls, 53% of non-gastrointestinal patients, 13% of Crohn's disease and 15% of ulcerative colitis excreted methane<sup>377</sup>. As sulphate reduction (mediated by the SRB population) and methanogenesis (mediated by the methanogen population) are mutually exclusive this implies that the SRB population tends to predominate in IBD patients and in support of this, another study found that the concentration of SRB in the faeces of ulcerative colitis patients with active disease was almost 3 log greater than that of ulcerative colitis patients with inactive disease<sup>346</sup>. Which population predominates in the colon has been shown to be dependent on factors such as colonic pH and sulphate availability<sup>378</sup>. When sulphate availability is high SRB have a higher affinity for hydrogen than methanogens<sup>378</sup> therefore the SRB population will predominate.

Most of the sulphate in the colon is absorbed by colonocytes<sup>379</sup> presumably *via* sulphate transporters. The process for absorbing sulphate is active, i.e. it requires ATP<sup>380</sup>. Within the colonocyte the sulphate molecule is used in a variety of cellular processes; for example, mucin molecules are sulphated before being secreted into the lumen<sup>146</sup>. However, if the expression of sulphate transporters is decreased in ulcerative colitis mucosa (figures 4.47 to 4.49), then it follows that there may be a larger concentration of free sulphate available to the SRB population, which would therefore predominate over the methanogen population.

#### 5.2.5 Down-regulation of Sulphate Transporters in IBD

The sulphate transporter genes *DTDST* and *DRA* show decreased expression in IBDi (figures 4.47 and 4.49). The down-regulation of *DTDST* was particularly severe in UCi, with the gene being called absent in all five samples. The *DTDST* and *DRA* proteins are most likely to be expressed by epithelial cells<sup>304, 356</sup> and epithelial degeneration is a feature of active ulcerative colitis. Therefore the genes' expression was normalised to the epithelial cell ICC data (figure 5.3) and this showed that the decreased expression of both *DRA* and *DTDST* in the IBDi samples was still significant. The cause of the decreased expression of the sulphate transporter genes is not clear, but it is unlikely to be a primary cause of IBD, as the uninvolved samples do not show similar gene expression.

**Figure 5.3 – Sulphate transporters' gene expression normalised to proportion of epithelial cells in sample**

### 5.2.6 The Proposed Mechanism

The proposal is that a decrease in sulphate transport gene expression leads to an energy deficient mucosa. This is illustrated in figures 5.4a and 5.4b.

The decreased uptake of sulphate by the colonocytes would produce a higher sulphate concentration in the lumen. Excess luminal sulphate would give an advantage to the SRB population over the methanogens, explaining the prevalence of SRB in the colons of ulcerative colitis patients. This enhanced SRB population would produce greater than normal amounts of hydrogen sulphide and butyrate, which would be secreted into the colonic lumen.

A greater net production of  $H_2S$  has been observed in ulcerative colitis<sup>346, 369</sup> and the levels of butyrate in the colonic lumen of ulcerative colitis patients have also been reported as greater than in controls<sup>381</sup>. Some of the butyrate would be absorbed by the colonocytes, but the colonocytes would be unable to oxidise it, as the  $\beta$ -oxidation pathway may be impaired by the presence of excess  $H_2S$ <sup>364-366</sup>. Whether the excess  $H_2S$  is able to inhibit the  $H_2S$  detoxification mechanism is not known. If as suggested by Roediger's initial hypothesis a lack of butyrate availability contributed to ulcerative colitis aetiology, butyrate enemas would be expected to be therapeutic for ulcerative colitis. However, it has been shown that additional butyrate in the form of butyrate enemas are not universally therapeutic for ulcerative colitis<sup>382</sup>. As butyrate is the preferred (and therefore main) colonocyte energy source, impaired butyrate oxidation would result in an energy deficient state.

#### 5.2.6.1 5-ASA

There is additional support for the mechanism from the effects of 5-ASA on sulphate metabolism. As discussed in section 1.2.2, 5-ASA is a component of the many of the drugs indicated in IBD. 5-ASA has been shown to significantly inhibit the ability of SRB to produce sulphides<sup>346, 372</sup>. Additionally, it has been shown that decreasing the intake of sulphur containing amino acids (i.e. methionine and cysteine) may benefit some ulcerative colitis patients<sup>383</sup>.

Figure 5.4a – Sulphate metabolism in the healthy colon\*

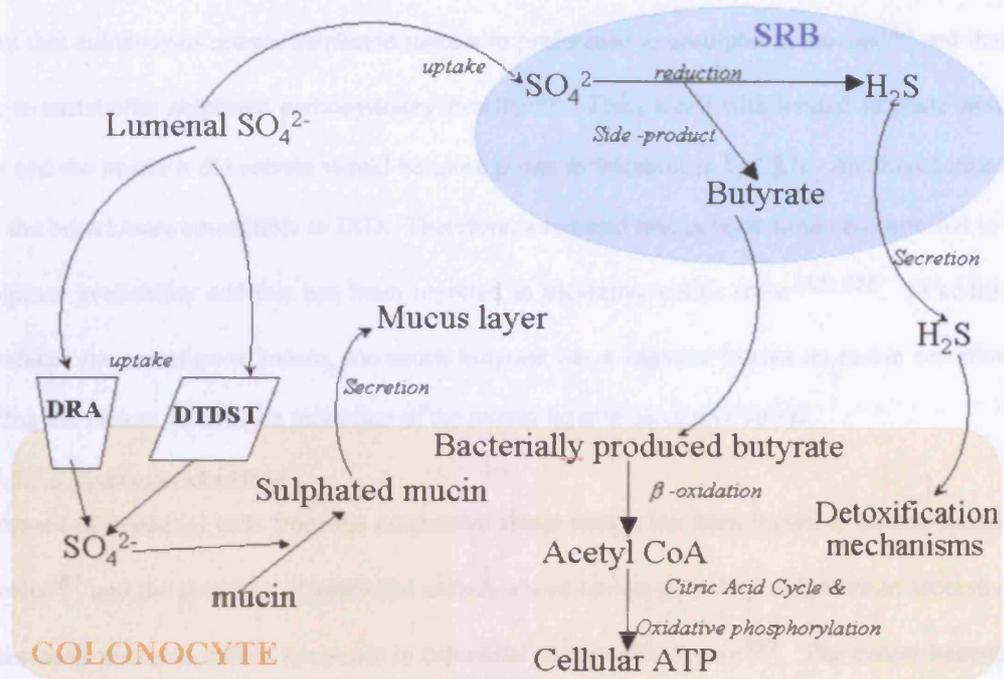
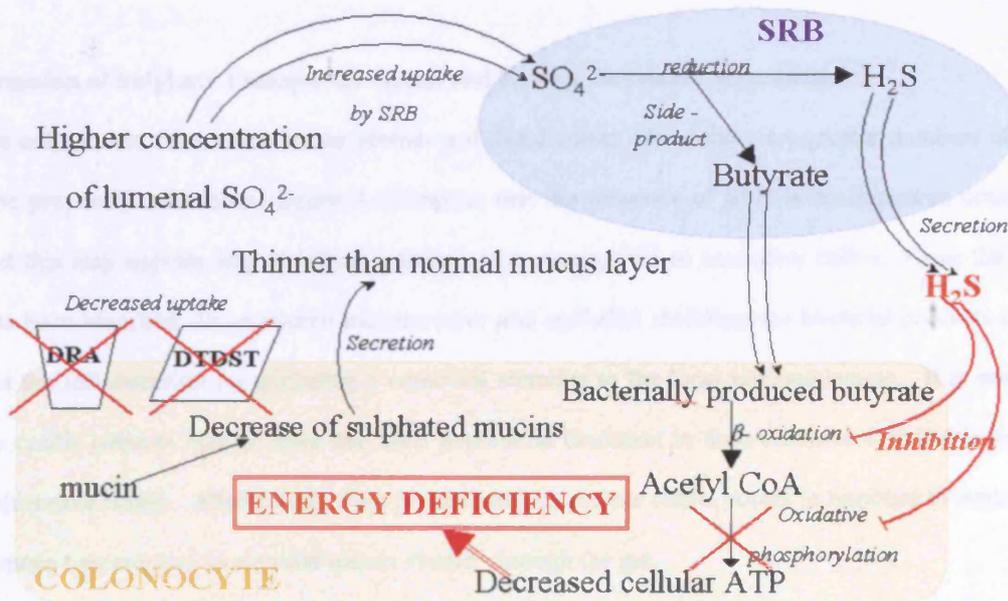


Figure 5.4b - Reduced Uptake of Sulphate Results in Energy Deficiency



5.2.6.2 The mucus layer

The mucus layer is a vital part of the epithelial barrier function, as discussed in chapter 1. It consists of high molecular weight, complex glycoconjugates called mucins, which consist of a polypeptide core (the apomucin) covered by O-linked carbohydrate chains, sulphates and other non-peptide molecules<sup>384</sup>.

\* SO<sub>4</sub><sup>2-</sup> represents the sulphate molecule in figure 5.4.

It has been shown that UC mucosa has a reduced ability to sulphate mucins<sup>145-147</sup>. This is illustrated in figure 5.4b; less sulphate available to the colonocyte implies less sulphate available for mucin sulphation. It has also been shown that colonocytes secrete sulphated mucins in preference to unsulphated mucins<sup>146</sup> and that the SRB are unable to metabolise sulphated carbohydrates directly<sup>346</sup>. Thus, a cell with limited sulphate would secrete less mucin and the mucin it did secrete would be more prone to metabolism by SRB. An altered mucus layer in may leave the bowel more susceptible to IBD. Therefore, a reduced mucus layer would be expected in cells with limited sulphate availability and this has been reported in ulcerative colitis colon<sup>385, 386</sup>. In addition, whilst butyrate induces the secretion of mucin, too much butyrate has a negative impact on mucin secretion<sup>359</sup>, thus compounding the factors resulting in reduction of the mucus layer in ulcerative colitis.

### 5.2.6.3 Epithelial shedding

The detachment of epithelial cells from the connective tissue matrix has been shown to induce apoptosis in the detached cells<sup>387</sup> and the shedding of epithelial cells is a well-known pathological feature of ulcerative colitis<sup>8</sup>. Butyrate has been shown to induce apoptosis in colorectal cancer cells *in vivo*<sup>360</sup>. The excess luminal butyrate (present according to the mechanism illustrated in figure 5.4) may therefore be partly responsible for epithelial shedding in ulcerative colitis.

### 5.2.7 Discussion of Sulphate Transporter Genes and Energy Deficiency Hypothesis

Ulcerative colitis most often starts in the rectum and distal colon, where there are greater numbers of bacterial cells. The proposed mechanism (figure 5.4) implies that the presence of SRB is an important component of colitis and this may explain why the distal colon is more susceptible to ulcerative colitis. Once the epithelial barrier has been breached via an altered mucous layer and epithelial shedding, the bacterial products could then compound the inflammation by providing a continual stimulus to the local immune system. It is possible that ulcerative colitis patients tend to have the SRB population dominant in their colons and that this predisposes them to ulcerative colitis. Alternatively it is possible that ulcerative colitis occurs in response to something that becomes more concentrated as material moves distally through the gut.

The initial theory considered the failure of butyrate oxidation as a primary causative factor of UC, whereas the proposed mechanism views it as a secondary feature; a concept which has some experimental support. Histological abnormalities precede measurable defects in butyrate oxidation in experimental rodent colitis<sup>338</sup>, abnormalities in butyrate oxidation could not be found in the mucosa of patients in histological remission<sup>344</sup>. If the failure of butyrate oxidation were a primary cause of UC, any abnormalities in butyrate oxidation would precede histological defects and would be found in quiescent, as well active colitis.

Several studies in rat colonocytes suggest that colitis may be caused by excessive H<sub>2</sub>S in the colon<sup>374, 388, 389</sup>. However, although UC patients generate larger amounts of colonic H<sub>2</sub>S than controls<sup>346, 368, 369</sup>, it does not seem likely that a high level of colonic H<sub>2</sub>S is the primary cause of colitis. In one study, rats were fed dextran sulphate, with or without bismuth subsalicylate. The dextran sulphate fed rat is a recognised animal model of colitis. Bismuth subsalicylate is a compound that avidly binds H<sub>2</sub>S and the faecal release of H<sub>2</sub>S in the bismuth subsalicylate fed rats was reduced well below that of controls. However, the bismuth subsalicylate fed rats still developed colitis, as did the control animals, implying that the level of colonic H<sub>2</sub>S is not a primary aetiological factor in colitis<sup>390</sup>.

It also seems unlikely that the reduced expression of sulphate transporters is a primary cause of ulcerative colitis. The uninvolved samples did not show gene expression levels of *DRA* and *DTDST* similar to the involved samples and in a previous study it was found that stimulating the colonic epithelial cell line Caco-2 with the pro-inflammatory cytokine IL-1 $\beta$ , decreased the transcript level expression of *DRA*<sup>170</sup>, implying that the inflammatory process occurs prior to the down-regulation of *DRA* expression.

The proposed mechanism aims to provide a potential hypothesis as to how the observed decrease in gene expression of two sulphate transporter genes may result in an energy deficient mucosa. The energy deficiency hypothesis is more relevant to ulcerative colitis than Crohn's disease, although a decrease in *DRA* and *DTDST* gene expression was seen in both forms of IBD. This may mean that diminished sulphate transport has a different consequence in the two diseases. The microarray samples were from refractory IBD and it must be remembered that the untreated or unoperated IBD mucosa may yield different results. It seems highly likely that the down-regulation of *DRA* and *DTDST* is a secondary feature of inflammation and the initial cause of this remains to be established.

## 5.2.8 Cancer and Regeneration in Inflammatory Bowel Disease

### 5.2.8.1 The microarray data

Abnormal cell proliferation is an essential part of cancer pathogenesis both in UC-CRC and sporadic CRC. The decreased expression of proliferation-regulating genes and the increased expression of proliferation-promoting genes would be expected in all cancer types. Table 5.1 summarises the genes that were identified as being differentially expressed in the IBDi tissues and through annotation were identified as potentially being involved in the progression of cancer. However, an increase in cell proliferation would also be expected in an inflamed environment, due to the regeneration of cells that have been destroyed by the inflammatory process and it is likely that this is the true cause of the increased expression of these genes in the ulcerative colitis tissues.

**Table 5.1 – Cancer-related genes that are over expressed in IBDi tissues**

<i>Gene Name</i>	<i>Gene Function</i>	<i>Summary of microarray data (figure)</i>	<i>Previous studies supporting gene expression data</i>
<i>REG</i> gene family	Mucosal regeneration <sup>328</sup>	Over expressed in IBDi tissues ( <i>REG1</i> α – figure 4.63; <i>REG1</i> β – figure 4.64)	All family members over expressed in IBDi samples <sup>180, 181, 328</sup>
<i>TRRAP</i>	Blastocyte proliferation <sup>266</sup> Cell cycle progression <sup>267</sup>	Over expressed in UCi samples (figure 4.28)	None
<i>MDG1</i>	Co-chaperone to HSP70 protein chaperone family <sup>284</sup>	Over expressed in UCi samples (figure 4.34)	Rat homologue up regulated during proliferation <sup>285</sup> Rat homologue down regulated during terminal differentiation <sup>286</sup>
<i>CD9</i> antigen	Inhibits the motility of colorectal tumour cells <i>in vitro</i> <sup>250</sup>	Higher expression in UCu compared to UCi tissues (figure 4.22)	None

The *REG* gene family represents a set of genes that are likely to be expressed in response to inflammation and the subsequent need for cellular regeneration. It is interesting however, to note that the expression of *REG* gene family transcripts in CRC patients with non-invasive disease had a highly significant adverse effect on survival. The *REG* transcripts were found to be present in the tumour cells themselves and the theory was that in addition to inhibition of apoptosis, the over expression of this gene family may contribute to carcinogenesis<sup>329</sup>.

The *TRRAP* protein is a blastocyte pro-proliferation factor<sup>266</sup>. This gene was called present in all samples except one (NI\_5) and it is concluded that although it has a role to play in the blastocyte, it may also have a role in the adult cell. In support of this adult role *TRRAP* has previously been noted as an essential co-factor in the c-

myc and E1A/E2F transcription factor pathways, both of which regulate cell cycle progression in higher eukaryotes<sup>267</sup>. It is plausible that over expression of this gene may lead to rapid cell cycle progression, contributing to abnormal cell proliferation.

The human MDG1 protein is up regulated *in vitro* during angiogenesis<sup>284</sup>. If this up regulation during angiogenesis *in vitro* is mirrored *in vivo*, the over expression of this gene in the UCi samples may indicate a higher propensity for carcinogenesis, as angiogenesis is an important part of carcinogenesis.

Another essential part of the progression of any cancer is the ability to metastasise. This is the ability of tumour cells to leave the area where they divided, travel undetected through the body, attach themselves to another tissue or organ and grow as a secondary tumour at that site. The CD9 antigen inhibits the motility of colorectal tumour cells *in vitro*<sup>250</sup> and it is therefore possible that the protein acts to inhibit metastasis *in vivo*. Its under expression in the UCi samples is comparable to the decreased expression of *MRP-1* in a number of different cancers, including colon cancer<sup>249</sup>. In many cases, a low *CD9* expression level is associated with a poor prognosis for the cancer patient. The higher expression of *MRP-1* in the UCu tissues compared to the UCi tissues implies that the local environment causes this reduction in expression. The increase in cellular motility of UCi affected mucosa has obvious consequences in terms of CRC progression. However, as *CD9* expression is only decreased in the involved ulcerative colitis samples, it is more likely that cell motility is increased in order to promote re-epithelialisation in response to the epithelial shedding that is characteristic of ulcerative colitis involved mucosa.

#### 5.2.8.2 Discussion of cancer related genes

The abnormal expression of each of these genes in the UC tissues implies at first glance a role in the development of CRC. However, the protein functions coded by these genes means that they are also likely to show the same expression patterns due to inflammation of the colonic mucosa. The expression of the *REG* genes in both CDi and UCi samples supports this view. It would have been useful to compare the IBD gene expression data to that of an inflammatory disease from another part of the body, e.g. psoriasis, as a further control to check for gene expression changes potentially due to inflammation only. It is not possible to distinguish between genes that are over expressed due to inflammation and those that are over expressed due to a UC-CRC pathway from the microarray. However, it is also possible that the expression pattern of some genes is in response to inflammation, but contributes to carcinogenesis as a side effect.

It is important to remember that the 'normal' control samples used in this study were from colorectal cancer patients. Although the tissues used were macroscopically normal, there may well be inherent differences between these controls and normal mucosa from non-cancer patients. The average difference values used as the 'normal' baseline may actually represent some intermediary expression level between truly normal and truly cancerous. The tumour gene expression levels from the NI patients were not measured and so the level of expression of the 'cancer-related' genes in cancerous tissue was unavailable as an additional control.

### 5.2.9 Apoptosis and Inflammatory Bowel Disease

Cells can die in two ways, by necrosis or by apoptosis. Necrosis is a passive process, initiated by factors other than the cell undergoing death. It results in the lysis of cell membranes and the spillage of cellular contents into the extracellular environment. Apoptosis (or programmed cell death) is an active process, requiring the synthesis of specific RNA and proteins. Crucially, every cell has the ability to initiate apoptosis in itself. A cell's survival is dependent on survival signals from neighbouring cells and the extracellular matrix; if these signals are not received, or the cell becomes detached from the extracellular matrix, it will undergo apoptosis. Other cells can also 'actively' induce apoptosis. Additionally, apoptotic cells die 'tidily'; they do not leak any of their cellular contents but shrink into small 'apoptotic' bodies, which are engulfed by phagocytic cells leaving no trace of the dead cell. In order to understand how the 'apoptosis related' genes identified in chapter 4 and discussed below are linked to apoptosis, a brief overview of the process is discussed first.

#### 5.2.9.1 Apoptotic pathways

There are many molecular mechanisms controlling apoptosis and whilst they are not all fully understood, three protein families that play vital roles in apoptosis have been described.

##### *The Caspases*

The caspases are a family of at least 10 cysteine proteases, which act in a cascade to promote apoptosis. They are synthesized as proenzymes and are activated by cleavage at specific aspartate cleavage sites. The caspases can be divided into two functional groups: those that initiate the cascade (e.g. caspases-8 & -10) by cleaving and thus activating all other caspases (including their own precursors) and those that cleave other proteins to induce the morphological changes associated with apoptosis (e.g. caspases -2 & -6). It is thought highly likely that the caspase family includes other as yet undiscovered members and that the functions of the currently known caspases may yet be expanded.

##### *The Bcl-2 family*

The Bcl-2 family currently consists of 15 proteins, two of which are anti-apoptotic (Bcl-2 & Bcl-x<sub>l</sub>) and the rest of which are pro-apoptotic (e.g. Bax, Bad, Bid). The proteins form a heterodimer consisting of an anti-apoptotic Bcl-2 family member and a pro-apoptotic Bcl-2 family member<sup>391</sup>, thus blocking the function of each. In the normal state the proportion of anti-apoptotic proteins is greater than that of the pro-apoptotic proteins. In apoptotic cells, this ratio is reversed.

Anti-apoptotic Bcl-2 is one of the targets of caspase-mediated cleavage, as is the pro-apoptotic Bid. The Bcl-2 fragments are inactive, whereas the Bid fragment triggers the release of cytochrome c from mitochondria; thus the cleavage of both proteins results in an acceleration of the apoptotic process<sup>392</sup>. Neither of the anti-apoptotic

Bcl-2 proteins is believed to bind any of the caspases directly, but their inhibitory actions are thought to affect the cascade prior to the processing of caspases-3 and -7<sup>393</sup>.

*The IAP (inhibitors of apoptosis) family*

All members of the IAP family inhibit apoptosis and are characterised by Cys-His repeats and a COOH terminal ring finger<sup>394</sup>. The first member to be identified was discovered in the baculovirus and is referred to as baculoviral IAP repeat-containing protein 1, or BIRC1. The IAP family is believed to inhibit caspases / pro-caspases directly by binding to them<sup>395, 396</sup>, for example BIRC1 is thought to be a CASP-3 antagonist.

**5.2.9.2 Apoptosis Induction**

Apoptosis can be induced in a number of ways by both internal (i.e. damage to intracellular components) and external factors. The mechanism of only a few apoptosis triggers is known and these are described below. Once induced, apoptosis proceeds along the pathways discussed above.

*Fas pathway*<sup>392, 393</sup>

All cells display the Fas receptor (CD95) on their extracellular membranes. The ligand for this receptor, Apo-1 or Fas ligand, is expressed by activated killer lymphocytes. Binding of CD95 to Fas ligand, causes the aggregation of other activated CD95 molecules. The aggregated receptors recruit adaptor proteins, which in turn recruit pro-caspase-8 molecules. The clustered pro-caspase-8 molecules cleave and activate each other, thus initiating the caspase cascade and apoptosis.

*Perforin-granzyme B pathway*<sup>392, 393</sup>

A second method employed by killer lymphocytes to induce apoptosis in virally infected cells is the perforin-granzyme B pathway. The lymphocytes secrete proteins including perforin onto the infected cell surface. The perforin molecules assemble into trans-membrane channels that allow the entry of other proteins such as granzyme B into the cell. Granzyme B cleaves and activates pro-caspase-7, thus initiating the apoptotic pathway.

*DNA damage*<sup>391</sup>

Damage to a cell's DNA initiates the transcription of the cell cycle repressor, p53. If the DNA is unreparable, p53 accumulates and inhibits the Bcl-2 protein. This increases the ratio of pro-apoptotic proteins to anti-apoptotic proteins, thereby triggering apoptosis.

*Cytoplasmic release of cytochrome c*<sup>392</sup>

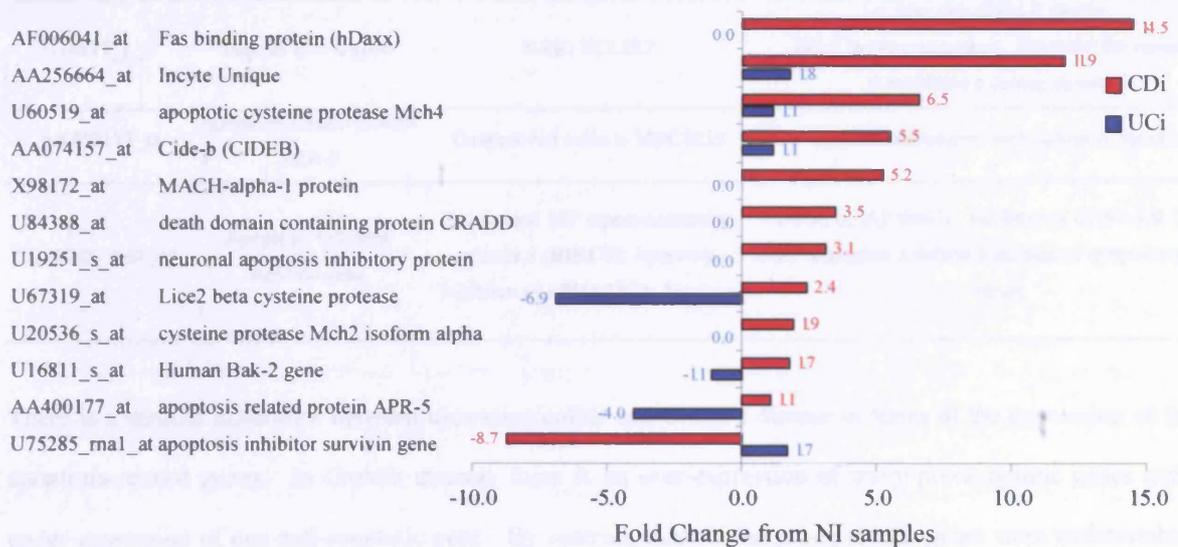
Damage to the mitochondria results in the release of cytochrome c into the cytoplasm. In the mitochondria, cytochrome c is a vital component of the electron transport chain and thus the respiratory process, but in the cytoplasm it can cause the death of the cell. Cytochrome c binds to (and thus activates) an adaptor protein, CED-4-like. This complex cleaves pro-caspase-9, resulting in the active caspase-9 which triggers the caspase

cascade and results in apoptosis. As discussed above, the caspase mediated cleavage of Bid also results in release of mitochondrial cytochrome c, thus providing a mechanism by which extracellular pro-apoptotic triggers also induce the release of cytochrome c.

**5.2.9.3 The microarray data**

Apart from the EST cDNA DKFZp434N103 (RC\_AA446944\_at), which represents the human homologue of the CED-6 gene, the other genes discussed here have not been detailed in chapter four. These genes were identified in the initial mining effort (detailed in section 4.4) and were not discussed due to the large number of genes returned. The ‘functional pathway’ and ‘fold change’ columns were used to select apoptosis related genes from 1148 genes selected in query 4.4 (tables 4.4 & 4.5). The fold change value for each gene is illustrated in figure 5.5. All the genes show an over expression in the CDi samples, except for BIRC5, which is under expressed in CDi. Two genes show under expression in the UCi samples; APR-5 and CASP-7. The protein functions of the genes returned are detailed in table 5.2.

**Figure 5.5 – Fold Change of Apoptosis-related genes from NI samples\***



All the apoptosis associated genes that are over expressed in the CDi samples including CED-6 (figure 4.23) promote apoptosis. These include several members of the caspase family and one pro-apoptotic member of the Bcl-2 family. Interestingly, BIRC5 has been investigated as a prognostic marker in colorectal cancer. It was found that patients with BIRC5 negative tumours (less apoptotic inhibition) had a five-year survival rate of 94.4% compared with 44.8% for patients with BIRC5 positive tumours (more apoptotic inhibition)<sup>397</sup>. BIRC5 is the only apoptotic gene to be under-expressed in the CDi samples and slightly over-expressed in the UCi samples when compared to the controls.

Table 5.2 – Protein function of differentially regulated apoptosis related genes†

Accession number	Array description	Other names‡	Protein Function
AF006041_at	Fas binding protein ( <b>Daxx</b> )	Death associated protein 6	Enhances Fas-mediated apoptosis; mediates Casp-independent activation of JNK pathway.
AA256664_at	Incyte Unique	LOC51283	Apoptosis regulator
U60519_at	Apoptotic cysteine protease Mch4	<b>CASP-10</b> ; FADD-like ICE2 (FLICE2)	Member of caspase family. Cleaves & activates CASP-3, -4, -6, -7, -8, -9.
AA074157_at	Cide-b ( <b>CIDEB</b> )	Cell death-inducing DFFA-like effector B	Apoptotic activator.
X98172_at	MACH-alpha-1 protein	<b>CASP-8</b>	Member of caspase family.
U84388_at	Death domain containing protein CRADD	<b>CASP-2</b> ; RAIDD	Member of caspase family
U19251_s_at	Neuronal apoptosis inhibitory protein (NAIP)	Baculoviral IAP repeat-containing protein 1 ( <b>BIRC1</b> )	Member of IAP family. May be a CASP-3 antagonist.
U67319_at	Lice2 beta cysteine protease	<b>CASP-7</b> ; MCH3, LICE2	Member of caspase family. Cleaves poly(ADP-ribose) polymerase (PARP).
U20536_s_at	cysteine protease Mch2 isoform alpha	<b>CASP-6</b>	Member of caspase family. Cleaves PARP & lamins <i>in vitro</i> . Over-expression promotes apoptosis.
U16811_s_at	Human Bak-2 gene	<b>BAK</b> ; BCL2L7	Member of Bcl-2 family. Bcl-2 protein antagonist. Essential for release of cytochrome c during apoptosis.
AA400177_at	Apoptosis related protein <b>APR-5</b>	Unapproved name is MGC2835	Hypothetical protein with unknown function.
U75285_mal_at	Apoptosis inhibitor survivin gene	Baculoviral IAP repeat-containing protein 5 ( <b>BIRC5</b> ); Apoptosis Inhibitor 4 (API4); IAP4; Survivin	Member of IAP family. Inhibitor of CASP-3 & CASP-7. May counteract a default induction of apoptosis in G2/M phase.

There is a distinct difference between ulcerative colitis and Crohn's disease in terms of the expression of these apoptosis-related genes. In Crohn's disease, there is an over-expression of many pro-apoptotic genes and an under-expression of one anti-apoptotic gene. By contrast, most of the pro-apoptotic genes were undetectable in the UCi samples and two of the pro-apoptotic genes (caspase-7 and APR-1) showed significant under-expression in the UCi samples compared to both NI and CDi.

Other caspase genes represented on the arrays included caspases-1, -2, -4, -5 and -9. As the average difference values for these were relatively low, they were not identified in the initial mining effort. However, the general expression pattern of these caspase genes does correspond to the expression patterns discussed above. Caspases-1, -4 and -9 had the highest expression levels in the Crohn's disease samples (figure 5.6). Caspase-2 was only called present in one sample, NI\_4. Although caspase-5 showed the highest mean average difference value in

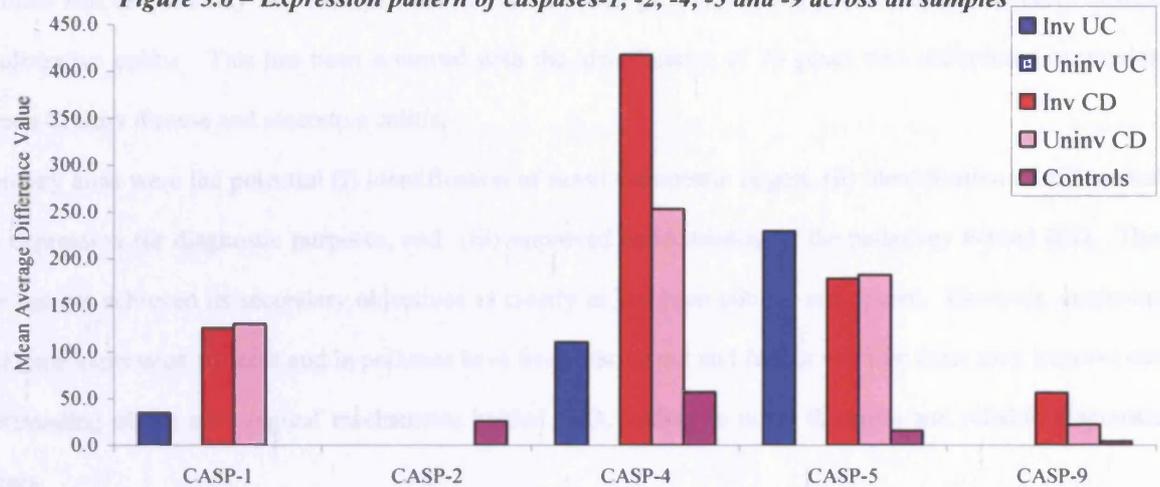
\* Genes that were absent in 100% of the UCi samples are identified with a fold change of '0'.

† The protein functions of the genes was determined by searching on the 'GeneCards' website (<http://bioinfo.weizmann.ac.il/cards/>)

‡ Where a gene has more than one name, the name used in this thesis is highlighted by the use of a bold typeface.

the UCi samples it was only called present in three of these, whilst being called present in eight of the nine Crohn's disease samples.

**Figure 5.6 – Expression pattern of caspases-1, -2, -4, -5 and -9 across all samples**



#### 5.2.9.4 Discussion of apoptosis related genes

The increased expression of pro-apoptotic genes could be attributed to the propensity of neutrophils to undergo apoptosis and the increased proportion of neutrophils in IBD tissues compared to controls. However, as discussed in section 2.1.4 and shown in section 2.3.7, the UCi samples had greater proportions of neutrophils than the CDi samples. If the increased expression of pro-apoptotic genes was due entirely to the neutrophils in the samples, the UCi samples should show a larger fold change than the CDi samples in figure 5.5. However, many of the pro-apoptotic genes were called absent in 100% of the UCi samples, which implies that the expression pattern of the apoptosis-related genes seen in the CDi tissues is not due to the presence of neutrophils. One theory of IBD pathology links the over-active immune system in IBD to apoptotic resistance of activated T lymphocytes. The theory maintains that once activated by an unknown trigger factor, the IBD immune system remains activated due to the resistance of activated T lymphocytes to apoptosis. In a normal immune response, activated killer T lymphocytes undergo apoptosis once the activating antigen has been cleared from the body. If these cells became resistant to apoptosis, then the immune system would remain active, in spite of the activating antigen having been cleared from the body. Activation of the T cell specific receptor CD2 induces the apoptotic cascade. In a previous study, lamina propria lymphocytes from Crohn's disease patients showed defective CD2-pathway induced apoptosis, associated with elevated levels of Bcl-2 protein<sup>398</sup>. Increased Bcl-2 protein would support the observation of the Bcl-2 protein antagonist BAK being slightly under-expressed in the CDi samples compared to the controls (figure 5.5). However, the data points to increased pro-apoptotic gene expression; for example many of the caspase genes show over-expression in the CDi samples compared to the NI samples. It is possible that this reduced apoptotic potential would only become clear in a microarray study if RNA from a purified sample of activated Crohn's disease T lymphocytes were used for target preparation.

## 5.3 Achievements

---

### 5.3.1 Objectives reached

The main aim of this study was to discover differences in the gene expression pattern between Crohn's disease and ulcerative colitis. This has been achieved with the identification of 70 genes with differential expression between Crohn's disease and ulcerative colitis.

Secondary aims were the potential (i) identification of novel therapeutic targets, (ii) identification of differential gene expression for diagnostic purposes, and (iii) improved understanding of the pathology behind IBD. This study has not achieved its secondary objectives as clearly as had been initially anticipated. However, numerous novel gene expression patterns and hypotheses have been discovered and further work on these may improve our understanding of the pathological mechanisms behind IBD, leading to novel therapies and reliable diagnostic markers.

### 5.3.2 Study design improvements

A major drawback of this study was the lack of an appropriate statistical method to test the data. This was partly due to the low number of samples, but was also due to the author's lack of experience in statistical methods. Statisticians are becoming involved in microarray studies and as expertise in this interdisciplinary field grows, statistical approaches appropriate for microarray analysis are becoming available. What is clear however, is that incorporating the relevant statistical method into the design of a microarray study is vital for the success of the experiment. This was not the case here, as the study had not been designed with any particular analysis or mining method in mind. The mining and analysis issue was only considered after the initial microarray data had been gathered.

In an idealised repeat of this study the number of samples would be increased. Allowing more time for sample collection could allow the grouping of samples based on histology rather than diagnosis. This would allow the generation of gene expression patterns for specific histological features such as fibrosis, and this may result in some interesting results regarding the similarity of UC and CD. Another difference would be the amount of expertise involved. Statisticians would be consulted at the start of the study, and the mining method determined before any microarray data were generated. This would hopefully allow the establishment of data mining methods specifically for microarray studies with IBD tissues.

### 5.3.3 Future studies

This study has generated numerous hypotheses, the study of which may yield important clues as to the pathogenesis of IBD. In particular, the expression of the sulphate transporters, *DRA* and *DTDST* would be investigated further. *In situ* hybridisation studies for the gene or immunohistochemistry for the protein

(depending on antibody availability) could be utilised to confirm expression patterns across a larger sample set. Having confirmed DRA and DTDST protein expression, the expression of other members of this family, which were not represented on the microarrays, could also be investigated. Ascertaining the expression levels of these additional sulphate transporters in CD and UC could give some indication as to whether the amount of sulphate being absorbed in IBD affected colonic mucosa is actually decreased, as it is possible that there is redundancy between these family members.

Other studies could also focus on ESTs that may now be associated with a full-length gene. This would involve the BLASTing of each gene individually against the human genome data and annotation as described in chapter 4, to find further gene expression patterns of interest.

## CONCLUDING REMARKS

The genes and pathways discussed in this thesis may prove to be important in understanding the pathogenesis of IBD. In particular, further research into the down regulation of the sulphate transporter genes *DRA* and *DTDST* may uncover some of the molecular mechanisms behind the pathology of ulcerative colitis and could lead to novel therapies targeting the pathway regulating their expression. The apparent failure of butyrate oxidation in ulcerative colitis is another interesting area and worthy of further investigation. If this is the case in ulcerative colitis (or a sub-set of patients), how does this influence disease progression and pathology? The apparent over expression of pro-apoptotic genes in Crohn's disease is another interesting observation and also deserves further investigation.

The techniques applied to the data generated in this project are typical of the types of tools available for massively parallel gene expression analysis. Very rudimentary statistics and mathematics have been applied to the data in this thesis, but as gene expression studies become routine mathematicians and statisticians are also becoming involved in the challenge of analysing gene expression data, resulting in more powerful techniques becoming available. Combining this with the increasingly detailed annotation of the human genome and proteome this data set will continue to generate novel gene expression patterns of interest to researchers in the IBD field and remains a valuable resource for further mining studies.

## REFERENCES

1. Baron, J.H., *Inflammatory bowel disease up to 1932*. Mount Sinai Journal of Medicine, 2000. 67(3): 174-189.
2. Crohn, B.B., L. Ginzburg, and G.D. Oppenheimer. *Regional ileitis: A pathological and clinical entity*. Journal of American Medical Association, 1932. 99: 1323-1329.
3. Dalziel, T.K., *Chronic Intestinal Enteritis*. British Medical Journal, 1913. 2: 1068-70.
4. Goligher, J.C., *Crohn's disease (granulomatous enteritis)*, in *Surgery of the anus, rectum and colon*, 1980, Bailliere Tindall: London. 827-857.
5. Brooke, B.N., *et al.*, *Crohn's disease*. 1977, London: The Macmillan Press Ltd. 113.
6. Kelly, M.P., *Colitis*. The experience of illness, ed. R. Fitzpatrick and S. Newman. 1992, London: Routledge. 127.
7. Farthing, M.J.G., *Crohn's disease in childhood and adolescence*, in *Inflammatory bowel disease*, A.A. Anagnostides, H.J.F. Hodgson, and J.B. Kirsner, Editors. 1991, Chapman & Hall: London. 12-25.
8. Goligher, J.C., *Ulcerative colitis*, in *Surgery of the anus, rectum and colon*. 1980, Balliere Tindall: London. 689-826.
9. Lockhart-Mummery, H.E. and B.C. Morson, *Crohn's disease (regional enteritis) of the large intestine and its distinction from ulcerative colitis*. Gut, 1960. 1: 87-105.
10. Anagnostides, A.A., H.J.F. Hodgson, and J.B. Kirsner, eds. *Inflammatory bowel disease*. 1 ed. . 1991, Chapman & Hall: London. 390.
11. Thomas, G.A., *et al.*, *Incidence of Crohn's disease in Cardiff over 60 years: 1986-1990 an update*. European Journal of Gastroenterology & Hepatology, 1995. 7(5): 401-5.
12. Burkitt, H.G., B. Young, and J.W. Heath, *Gastrointestinal Tract*, in *Wheater's Functional Histology*. 1997, Churchill Livingstone: Newcastle. Australia. 247-270.
13. Jass, J.R., N.A. Shepherd, and J.D. Maybee, *Atlas of surgical pathology of the colon, rectum and anus*. 1989, London: Churchill Livingstone. 227.
14. Rhodes, J.M., *Clinical patterns in inflammatory bowel disease*, in *Inflammatory bowel disease*, A.A. Anagnostides, H.J.F. Hodgson, and J.B. Kirsner, Editors. 1991, Chapman & Hall: London. 3-11.
15. Morson, B.C. and I.M.P. Dawson, *Gastrointestinal Pathology*. 2 ed. 1979, Oxford: Blackwell Scientific Publications.
16. Cotran, R.S., V. Kumar, and T. Collins, *Robbins pathologic basis of disease*. 1999, Philadelphia, USA: WB Saunders Company. 815-821.

17. Hamilton. S.R., J.L. Farber, and E. Rubin, *The Gastrointestinal Tract*, in *Pathology*, E. Rubin and J.L. Farber, Editors. 1999, Lippincott-Raven: Philadelphia. 668-755.
18. Ross, M.H., L.J. Romrell, and G.I. Kaye, *Digestive system II: Esophagus and gastrointestinal tract*, in *Histology: A text and atlas*, P.A. Coryell, Editor. 1995, Williams & Wilkins: Baltimore, Maryland. 440-495.
19. Geboes, K., *Crohn's disease, ulcerative colitis or indeterminate colitis--how important is it to differentiate?* Acta Gastroenterol Belg, 2001. **64**(2): 197-200.
20. Moum, B., *et al.*, *Incidence of ulcerative colitis and indeterminate colitis in four countries in southeastern Norway, 1990-93*. Scandinavian Journal of Gastroenterology, 1996. **31**(4): 362-6.
21. Swan, N.C., *et al.*, *Fulminant colitis in inflammatory bowel disease*. Diseases of the Colon and Rectum, 1998. **41**: 1511-1515.
22. Riegler, G., *et al.*, *Clinical evolution in an outpatient series with indeterminate colitis*. Diseases of the Colon and Rectum, 1997. **40**: 437-439.
23. Misiewicz, J.J., *et al.*, *Gastroenterology: Diseases of the colon and rectum*, ed. D. Bennett. Vol. 3. 1988, London: Gower Medical Publishing. 140.
24. Dobbins, W.O., M. Stock, and A.L. Ginsberg, *Early detection and prevention of carcinoma of the colon in patients with ulcerative colitis*. Cancer, 1977. **40**(5): 2542-8.
25. Fiocchi, C., *Inflammatory bowel disease: etiology and pathogenesis*. Gastroenterology, 1998. **115**: 182-205.
26. Lapidus, A., *The changing epidemiology of inflammatory bowel diseases*. Acta Gastroenterol Belg, 2001. **64**(2): 155-59.
27. Rubio, C.A. and R. Befrits, *Colorectal adenocarcinoma in Crohn's disease: a retrospective histologic study*. Diseases of the Colon and Rectum, 1997. **40**: 1072-1978.
28. Lapidus, A., *et al.*, *Incidence of Crohn's disease in Stockholm County 1955-1989*. Gut, 1997. **41**(4): 480-6.
29. Levine, J.B. and D. Lukawski-Trubish, *Extraintestinal considerations in inflammatory bowel disease*. Gastroenterology Clinics of North America, 1995. **24**(3): 633-646.
30. Snook, J.A., H.H.d. Silva, and D.P. Jewell, *The association of autoimmune disorders with inflammatory bowel disease*. Quarterly Journal of Medicine, 1989. **72**(269): 835-840.
31. Spira, A., R. Grossman, and M. Balter, *Large airway disease associated with inflammatory bowel disease*. Chest, 1998. **113**(6): 1723-1726.
32. Korelitz, B.I. and N. Sohn, *Management of inflammatory bowel disease*. 1992, St. Louis: Mosby.
33. Rowe, W.A., *Inflammtory bowel disease*. eMedicine Journal, 2001. **2**(8).

34. Barnes, P.J., *Anti-inflammatory actions of glucocorticoids: molecular mechanisms*. Clin Sci (Colch), 1998. **94**(6): 557-72.
35. Ghosh, S., A. Shand, and A. Ferguson, *Ulcerative colitis*. British Medical Journal, 2000. **320**(7242): 1119-23.
36. Coltery, C., *et al.*, eds. *Therapeutic Drugs*. . Vol. 1 and 2. 1991, Churchill Livingstone: London.
37. Kronke, M., *et al.*, *Cyclosporin A inhibits T-cell growth factor gene expression at the level of mRNA transcription*. Proceedings of the National Academy of Sciences U.S.A., 1984. **81**(16): 5214-8.
38. Bell, S.J. and M.A. Kamm, *Review article: the clinical role of anti-TNF $\alpha$  antibody treatment in Crohn's disease*. Alimentary Pharmacology & Therapeutics, 2000. **14**(5): 501-514.
39. Deventer, S.v., *Immunomodulation of Crohn's disease using TNF-  $\alpha$  neutralizing monoclonal antibodies*. Clinical Nutrition, 1997. **16**: 271-275.
40. Rutgeerts, P., *et al.*, *Efficacy and safety of retreatment with anti-tumor necrosis factor antibody (Infliximab) to maintain remission in Crohn's disease*. Gastroenterology, 1999. **117**(4): 761-9.
41. Sandborn, W.J., *et al.*, *An engineered human antibody to TNF (CDP571) for active Crohn's disease: a randomized double-blind placebo-controlled trial*. Gastroenterology, 2001. **120**(6): 1330-8.
42. Sands, B.E., *Biological therapies for ulcerative colitis*. Acta Gastroenterol Belg, 2001. **64**(2): 205-209.
43. Ramakrishna, J., *et al.*, *Combined use of cyclosporine and azathioprine or 6-mercaptopurine in pediatric inflammatory bowel disease*. Journal of Pediatric Gastroenterology and Nutrition, 1996. **22**(3): 296-302.
44. Fazio, V., *Current status of surgery for inflammatory bowel disease*. Digestion, 1998. **59**: 470-480.
45. Hofer, B., *et al.*, *The impact of clinical types of disease manifestation on the risk of early postoperative recurrence in Crohn's disease*. Hepatogastroenterology, 2001. **48**(37): 152-5.
46. Hogezaand, R.A.v. and W.A. Bemelman, *Management of recurrent Crohn's disease*. Neth J Med, 1998. **53**(6): S32-8.
47. Orholm, M., *et al.*, *Concordance of inflammatory bowel disease among Danish twins. Results of a nationwide study*. Scandinavian Journal of Gastroenterology, 2000. **35**(10): 1075-81.
48. Tysk, C., *Genetic susceptibility in Crohn's disease - review of clinical studies*. European Journal of Surgery, 1998. **164**(12): 893-896.
49. Binder, V., *Genetic epidemiology in inflammatory bowel disease*. Digestive Diseases, 1998. **16**: 351-355.
50. Breslin, N.P., *et al.*, *Monozygotic twins with Crohn's disease and ulcerative colitis: a unique case report*. Gut, 1997. **41**(4): 557-60.
51. Karlinger, K., *et al.*, *The epidemiology and the pathogenesis of inflammatory bowel disease*. European Journal of Radiology, 2000. **35**(3): 154-67.
52. Logan, R.F.A., *Inflammatory bowel disease incidence: up, down or unchanged?* Gut, 1998. **42**(3): 309-11.

53. Bjornsson, S., J.H. Johannsson, and E. Oddsson. *Inflammatory bowel disease in Iceland, 1980-89. A retrospective nationwide epidemiologic study*. Scandinavian Journal of Gastroenterology, 1998. **33**(1): 71-7.
54. Fonager, K., H.T. Sorensen, and J. Olsen. *Change in incidence of Crohn's disease and ulcerative colitis in Denmark. A study based on the National Registry of Patients, 1981-1992*. International Journal of Epidemiology, 1997. **26**(5): 1003-8.
55. Pajares, J.M. and J.P. Gisbert. *Epidemiology of inflammatory bowel disease in Spain - A systematic review*. Rev Esp Enferm Dig, 2001. **93**(1): 9-20.
56. Primatesa, P. and M.J. Goldacre. *Crohn's disease and ulcerative colitis in England and the Oxford record linkage study area: a profile of hospitalized morbidity*. International Journal of Epidemiology, 1995. **24**(5): 922-8.
57. Blanchard, J.F., et al., *Small-area variations and sociodemographic correlates for the incidence of Crohn's disease and ulcerative colitis*. American Journal of Epidemiology, 2001. **154**(4): 328-335.
58. Farrokhyar, R., E.T. Swarbrick, and E.J. Irvine. *A critical review of epidemiological studies in inflammatory bowel disease*. Scandinavian Journal of Gastroenterology, 2001. **36**(1): 2-15.
59. Saro-Gismera, C., et al., *Epidemiology of chronic inflammatory bowel disease in Gijon, Asturias*. Gastroenterologia y Hepatologia, 2001. **24**(5): 228-235.
60. Jayanthi, V., et al., *Low incidence of ulcerative colitis and proctitis in Bangladeshi migrants in Britain*. Digestion, 1992. **52**(1): 34-42.
61. Shivananda, S., et al., *Epidemiology of Crohn's disease in Regio Leiden, The Netherlands. A population study from 1979 to 1983*. Gastroenterology, 1987. **93**(5): 966-74.
62. Carr, I. and J.F. Mayberry, *The effects of migration on ulcerative colitis: a three-year prospective study among Europeans and first- and second- generation South Asians in Leicester (1991-1994)*. American Journal of Gastroenterology, 1999. **94**(10): 2918-22.
63. Feehally, J., et al., *Disease variations in Asians in Leicester*. Q J Med, 1993. **86**(4): 263-9.
64. Yang, S.K., E.V.L. Jr, and W.J. Sandborn, *Epidemiology of inflammatory bowel disease in Asia*. Inflammatory Bowel Disease, 2001. **7**(3): 260-270.
65. Cosnes, J., et al., *Smoking cessation and the course of Crohn's disease: an intervention study*. Gastroenterology, 2001. **120**(5): 1093-9.
66. Timmer, A., L.R. Sutherland, and F. Martin. *Oral contraceptive use and smoking are risk factors for relapse in Crohn's disease*. Gastroenterology, 1998. **14**(6): 1143-50.
67. Beaugerie, L., et al., *Impact of cessation of smoking on the course of ulcerative colitis*. American Journal of Gastroenterology, 2001. **96**(7): 2113-6.

68. Tysk, C. and G. Jarnerot, *Has smoking changed the epidemiology of ulcerative colitis?* Scandinavian Journal of Gastroenterology, 1992. **27**(6): 508-512.
69. Sandborn, W.J., *et al.*, *Transdermal Nicotine for Mildly to Moderately Active Ulcerative Colitis.* Annals of Internal Medicine, 1997. **126**(5): 364-71.
70. Rubin, D.T. and S.B. Hanauer, *Smoking and inflammatory bowel disease.* European Journal of Gastroenterology & Hepatology, 2000. **12**(8): 855-62.
71. Thomas, G.A.O., *et al.*, *Role of smoking in inflammatory bowel disease: implications for therapy.* Postgraduate Medical Journal, 2000. **76**(895): 273-9.
72. Vohra, P., *Inflammatory bowel disease.* Indian J Pediatr, 2000. **67**(10): 747-756.
73. Zijlstra, F.J., *Smoking and nicotine in inflammatory bowel disease: good or bad for cytokines?* Mediators of Inflammation, 1998. **7**(3): 153-155.
74. Corrao, G., *et al.*, *Risk of inflammatory bowel disease attributable to smoking, oral contraception and breastfeeding in Italy: a nationwide case-control study.* International Journal of Epidemiology, 1998. **27**(3): 397-404.
75. Green, B.T. and J.A. DiPalma, *Crohn's disease activity not affected by oral contraceptives.* American Journal of Gastroenterology, 2000. **95**(6): 1585-6.
76. Qiu, B.S., *et al.*, *The role of CD4+ lymphocytes in the susceptibility of mice to stress-induced reactivation of experimental colitis.* Nature Medicine, 1999. **5**(10): 1178-82.
77. Jacobsohn, W.Z. and Y. Levine, *Incidence and prevalence of ulcerative colitis in the Jewish population of Jerusalem.* Israel Journal of Medical Sciences, 1986. **22**(7-8): 559-63.
78. Peeters M, *et al.*, *Familial and sporadic inflammatory bowel disease: different entities?* Inflammatory Bowel Disease, 2000. **6**(4): 314-20.
79. van Heel D.A., McGovern D.P.B., and J. D.P., *Crohn's disease: genetic susceptibility, bacteria and innate immunity.* The Lancet, 2001. **357**: 1902-1904.
80. Lee, J.C.W., *et al.*, *Why children with inflammatory bowel disease are diagnosed at a younger age than their affected parent.* Gut, 1999. **44**(6): 808-11.
81. Polito J.M., *et al.*, *Preliminary evidence for genetic anticipation in Crohn's disease.* The Lancet, 1996. **347**(9004): 798-800.
82. Grandbastien, B., *et al.*, *Anticipation in familial Crohn's disease.* Gut, 1998. **42**(2): 170-4.
83. Rosenberg, R.N., *DNA-triplet repeats and neurologic disease.* New England Journal of Medicine, 1996. **335**(16): 1222-4.
84. Hampe J, *et al.*, *Anticipation in inflammatory bowel disease: a phenomenon caused by an accumulation of confounders.* American Journal of Medical Genetics, 2000. **92**(3): 178-83.

85. Picco MF, *et al.*, *Methodologic pitfalls in the determination of genetic anticipation: the case of Crohn disease*. *Ann Intern Med*, 2001. **134**(12): 1124-9.
86. Hugot, J.-P., *et al.*, *Mapping of a susceptibility locus for Crohn's disease on chromosome 16*. *Nature*, 1996. **379**: 821-823.
87. Ohmen, J.D., *et al.*, *Susceptibility locus for inflammatory bowel disease on chromosome 16 has a role in Crohn's disease, but not in ulcerative colitis*. *Human Molecular Genetics*, 1996. **5**(10): 1679-83.
88. Colombel, J.F., *Genetics of IBD: Where do we stand?* *Digestive Diseases*, 1998. **16**(6): 349-350.
89. Hugot, J.P. and G. Thomas, *Genome-wide scanning in inflammatory bowel diseases*. *Digestive Diseases*, 1998. **16**(6): 364-9.
90. Hampe J, *et al.*, *A genomewide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort*. *American Journal of Human Genetics*, 1999. **64**(3): 808-16.
91. Mirza, M.M., *et al.*, *Evidence of linkage of the inflammatory bowel disease susceptibility locus on chromosome 16 (IBD1) to ulcerative colitis*. *Journal of Medical Genetics*, 1998. **35**(3): 218-21.
92. van Heel D.A., *et al.*, *Inflammatory bowel disease: progress toward a gene*. *Canadian Journal of Gastroenterology*, 2000. **14**(3): 207-18.
93. Parkes M, *et al.*, *The IBD2 locus shows linkage heterogeneity between ulcerative colitis and Crohn disease*. *American Journal of Human Genetics*, 2000. **67**(6): 1605-10.
94. Satsangi, J., *Genetics of inflammatory bowel disease: from bench to bedside?* *Acta Odontol Scand*, 2001. **59**(3): 187-92.
95. Dechairo B, *et al.*, *Replication and extension studies of inflammatory bowel disease susceptibility regions confirm linkage to chromosome 6p (IBD3)*. *European Journal of Human Genetics*, 2001. **9**(8): 627-33.
96. Cho J.H., *et al.*, *Linkage and linkage disequilibrium in chromosome band 1p36 in American Chaldeans with inflammatory bowel disease*. *Human Molecular Genetics*, 2000. **9**(9): 1425-32.
97. Paavola P, *et al.*, *Genetic analysis in Finnish families with inflammatory bowel disease supports linkage to chromosome 3p21*. *European Journal of Human Genetics*, 2001. **9**(5): 328-34.
98. Ma Y, *et al.*, *A genome-wide search identifies potential new susceptibility loci for Crohn's disease*. *Inflammatory Bowel Disease*, 1999. **5**(4): 271-8.
99. Vermeire S, *et al.*, *Evidence for inflammatory bowel disease of a susceptibility locus on the X chromosome*. *Gastroenterology*, 2000. **120**(4): 834-40.
100. Hugot, J.-P., *et al.*, *Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease*. *Nature*, 2001. **411**(6837): 599-603.
101. Hampe, J., *et al.*, *Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations*. *Lancet*, 2001. **357**(9272): 1925-1928.

102. Ogura, Y., *et al.*, *A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease*. *Nature*, 2001. **411**(6837): 603-606.
103. Pabst, R., *The anatomical basis for the immune function of the gut*. *Anatomy and Embryology*, 1987. **176**(2): 135-144.
104. Podolsky D.K., Fournier D.A., and L. K.E., *Human colonic goblet cells. Demonstration of distinct subpopulations defined by mucin-specific monoclonal antibodies*. *Journal of Clinical Investigation*, 1986. **77**(4): 1263-71.
105. Podolsky, D.K. and K.J. Isselbacher, *Composition of human colonic mucin. Selective alteration in inflammatory bowel disease*. *Journal of Clinical Investigation*, 1983. **72**(1): 42-53.
106. Smith, A.C. and D.K. Podolsky, *Biosynthesis and secretion of human colonic mucin glycoproteins*. *Journal of Clinical Investigations*, 1987. **80**(2): 300-7.
107. Podolsky, D.K. and D.A. Fournier, *Alterations in mucosal content of colonic glycoconjugates in inflammatory bowel disease defined by monoclonal antibodies*. *Gastroenterology*, 1988. **95**(2): 379-87.
108. Smithson J.E., *et al.*, *Altered expression of mucins throughout the colon in ulcerative colitis*. *Gut*, 1997. **40**(2): 234-40.
109. Evans, C.M., *et al.*, *Activation of lamina propria T cells induces crypt epithelial proliferation and goblet cell depletion in cultured human fetal colon*. *Gut*, 1992. **33**(2): 230-5.
110. Leiper, K. and J.M. Campbell BJ, Milton J, Yu LG, Democratis J, Rhodes JM., *Interaction between bacterial peptides, neutrophils and goblet cells: a possible mechanism for neutrophil recruitment and goblet cell depletion in colitis*. *Clin Sci (Lond)*, 2001. **101**(4): 395-402.
111. Kagnoff, M.F., *Current concepts in mucosal immunity III. Ontogeny and function of  $\gamma\delta$  T cells in the intestine*. *American Journal of Physiology*, 1998. **274**(Gastrointestinal Liver Physiology 37): G455-G458.
112. Cerf-Bensussan, N., E.E. Schneeberger, and A.K. Bhan, *Immunohistologic and immunoelectron microscopic characterization of the mucosal lymphocytes of human small intestine by the use of monoclonal antibodies*. *Journal of Immunology*, 1983. **130**(6): 2615-22.
113. Selby, W.S., *et al.*, *Lymphocyte subpopulations in the human small intestine. The findings in normal mucosa and in the mucosa of patients with adult coeliac disease*. *Clin Exp Immunol*, 1983. **52**(1): 219-28.
114. Hirata, I., *et al.*, *Immunohistological characterization of intraepithelial and lamina propria lymphocytes in control ileum and colon in inflammatory bowel disease*. *Digestive Diseases and Sciences*, 1986. **31**(6): 593-603.
115. Ferguson, A., *Intraepithelial lymphocytes of the small intestine*. *Gut*, 1977. **18**: 921-937.

116. James, S.P., *Treatment implications of immunological abnormalities*, in *Inflammatory bowel diseases: pathophysiology as a basis of treatment*, J. Scholmerich, et al., Editors. 1993, Kluwer Academic Publishers BV: Lancaster, UK. 92-100.
117. Kohne, G., T. Schneider, and M. Zeitz, *Special features of the intestinal lymphocytic system*. Bailliere's Clinical Gastroenterology, 1996. **10**(3): 427-42.
118. MacDermott, R.P., *Alterations in the mucosal immune system in ulcerative colitis and Crohn's disease*. Inflammatory Bowel Disease, 1994. **78**(6): 1207-1231.
119. Roitt, I., J. Brostoff, and D. Male, *Immunology*. 5 ed. 1998, London: Mosby International. 423.
120. Bland, P.W. and L.G. Warren, *Antigen presentation by epithelial cells of the rat small intestine*. Immunology, 1986. **58**: 1-7.
121. Hommes, D.W., et al., *Production and cellular source of interleukin-8 in ulcerative colitis*. Inflammatory Bowel Disease, 1995. **1**: 108-116.
122. Salerno, A. and F. Dieli, *Role of  $\gamma\delta$  T lymphocytes in immune response in humans and mice*. Critical Reviews in Immunology, 1998. **18**(4): 327-257.
123. Havran, W.L., Y. Chen, and R. Boismenu, *Innate functions of epithelial  $\gamma\delta$  T cells*. Advances in Experimental Medicine & Biology, 1998. **452**: 29-35.
124. Boismenu, R. and W.L. Havran,  *$\gamma\delta$  T cells in host defense and epithelial cell biology*. Clinical Immunology & Immunopathology, 1998. **86**(2): 121-133.
125. Mosmann, T.R. and S. Sad, *The expanding universe of T-cell subsets: Th1, Th2 and more*. Immunology Today, 1996. **17**(3): 138-146.
126. Romagnani, S., *T-cell subsets (Th1 versus Th2)*. Ann Allergy Asthma Immunol, 2000. **85**(1): 9-18.
127. Brandtzaeg, P., *The human intestinal immune system: basic cellular and humoral mechanisms*. Baillieres Clin Rheumatol, 1996. **10**(1): 1-24.
128. Radford-Smith, G., *Ulcerative colitis: an immunological disease?* Bailliere's Clinical Gastroenterology, 1997. **11**(1): 35-52.
129. Chamberlin W, et al., *Review article: Mycobacterium avium subsp. paratuberculosis as one cause of Crohn's disease*. Alimentary Pharmacology and Therapy, 2001. **15**(3): 337-46.
130. Chiodini, R.J., *Crohn's disease and the mycobacterioses: a review and comparison of two disease entities*. Clinical Microbiological Reviews, 1989. **2**(1): 90-117.
131. El-Zaatari F.A.K., Osato M.S., and G. D.Y., *Etiology of Crohn's disease: the role of Mycobacterium avium paratuberculosis*. Trends in Molecular Medicine, 2001. **7**(6): 247-52.

132. Olsen I, *et al.*, *Elevated antibody responses in patients with Crohn's disease against a 14-kDa secreted protein purified from Mycobacterium avium subsp. paratuberculosis*. Scandinavian Journal of Immunology, 2001. **53**(2): 198-203.
133. Naser S.A, *et al.*, *Specific seroreactivity of Crohn's disease patients against p35 and p36 antigens of M. avium subsp. paratuberculosis*. Vet Microbiol, 2000. **77**(3-4): 497-504.
134. Schwartz D, *et al.*, *Use of short-term culture for identification of Mycobacterium avium subsp. paratuberculosis in tissue from Crohn's disease patients*. Clin Microbiol Infect, 2000. **6**(6): 303-7.
135. Wakefield, A.J., *et al.*, *Evidence of persistent measles virus infection in Crohn's disease*. J Med Virol, 1993. **39**(4): 345-53.
136. Lewin, J., *et al.*, *Persistent measles virus infection of the intestine: confirmation by immunogold electron microscopy*. Gut, 1995. **36**(4): 564-9.
137. Miyamoto, H., *et al.*, *Detection of immunoreactive antigen, with a monoclonal antibody to measles virus, in tissue from a patient with Crohn's disease*. J Gastroenterol, 1995. **30**(1): 28-33.
138. Feeney, M., *et al.*, *A case-control study of measles vaccination and inflammatory bowel disease*. Lancet, 1997. **350**(9080): 764-6.
139. Chadwick, N., *et al.*, *Measles virus RNA is not detected in inflammatory bowel disease using hybrid capture and reverse transcription followed by the polymerase chain reaction*. J Med Virol, 1998. **55**(4): 305-11.
140. Haga Y, *et al.*, *Absence of measles viral genomic sequence in intestinal tissues from Crohn's disease by nested polymerase chain reaction*. 1996.
141. Davis, R.L., *et al.*, *Measles-mumps-rubella and other measles-containing vaccines do not increase the risk for inflammatory bowel disease: a case-control study from the Vaccine Safety Datalink project*. Arch Pediatr Adolesc Med, 2001. **155**(3): 354-9.
142. Meulen, V.T., *Measles virus and Crohn's disease: view of a medical virologist*. Gut, 1998. **43**: 733-734.
143. Morris D.L., *et al.*, *Measles vaccination and inflammatory bowel disease: a national British Cohort Study*. American Journal of Gastroenterology, 2000. **95**(12): 3507-12.
144. Robertson D.J. and S. R.S., *Measles virus and Crohn's disease: a critical appraisal of the current literature*. Inflammatory Bowel Disease, 2001. **7**(1): 51-7.
145. Klinken, B.J.-W.v., *et al.*, *Sulphation and secretion of the predominant secretory human colonic mucin MUC2 in ulcerative colitis*. Gut, 1999. **44**: 387-93.
146. Raouf, A.H., *et al.*, *Sulphation of colonic and rectal mucin in inflammatory bowel disease: reduced sulphation of rectal mucus in ulcerative colitis*. Clinical Science, 1992. **83**: 623-6.

147. Corfield, A.P., *et al.*, *Loss of sulphate in human colonic mucosa during ulcerative colitis*. Biochemical Society Transactions, 1992. **20**: 95S.
148. Corfield, A.P., *et al.*, *Mucins and mucosal protection in the gastrointestinal tract: new prospects for mucins in the pathology of gastrointestinal disease*. Gut, 2000. **47**: 589-94.
149. Schiffrin, E.J. and D. Brassart, *Intestinal microflora and the mucosal mechanisms of protection*, in *Colonic Microbiota, Nutrition and Health*, G.R. Gibson and M.B. Roberfroid, Editors. 1999, Kluwer Academic Publishers: Dordrecht. 201-211.
150. Sartor, R.B., *Pathogenesis and immune mechanisms of chronic inflammatory bowel disease*. American Journal of Gastroenterology, 1997. **92**(12): 5S-11S.
151. Hollander, D., *et al.*, *Increased intestinal permeability in patients with Crohn's disease and their relatives*. Annals of Internal Medicine, 1986. **105**: 883-885.
152. May, G.R., L.R. Sutherland, and J.B. Meddings, *Is small intestinal permeability really increased in relatives of patients with Crohn's disease?* Gastroenterology, 1993. **104**: 1627-1632.
153. Ogura Y, *et al.*, *Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF-kappaB*. Journal of Biological Chemistry, 2001. **276**(7): 4812-8.
154. Inohara N, *et al.*, *Human Nod1 confers responsiveness to bacterial lipopolysaccharides*. Journal of Biological Chemistry, 2001. **276**(4): 2551-4.
155. Neurath, M.F., C. Becker, and K. Barbulescu, *Role of NF-kappaB in immune and inflammatory responses in the gut*. Gut, 1998. **43**(6): 856-860.
156. Beutler, B., *Autoimmunity and apoptosis. the crohn's connection*. Immunity, 2001. **15**(5): 5-14.
157. Anderson, M., *Genetics research might help to answer questions about the etiology of Crohn's disease*, in *HMS Beagle*. 2001.
158. Lee, J.C., J.E. Lennard-Jones, and G. Cambridge, *Antineutrophil antibodies in familial inflammatory bowel disease*. Gastroenterology, 1995. **108**: 428-433.
159. Shanahan, F., *et al.*, *Neutrophil autoantibodies in ulcerative colitis: familial aggregation and genetic heterogeneity*. Gastroenterology, 1992. **103**: 456-461.
160. Podolsky, D.K., *Lessons from genetic models of inflammatory bowel disease*. Acta Gastroenterol Belg, 1997. **60**(2): 163-5.
161. Mayer, L. and R. Shlien, *Evidence for function of Ia molecules on gut epithelial cells in man*. Journal of Experimental Medicine, 1987. **166**: 1471-1483.
162. Cerf-Bensussan N, *et al.*, *Intraepithelial lymphocytes modulate Ia expression by intestinal epithelial cells*. Journal of Immunology, 1984. **132**(5): 2244-52.

163. Blumberg, R.S., *et al.*, *Antigen presentation by intestinal epithelial cells*. Immunology Letters, 1999. **69**(1): 7-11.
164. Hershberg, R.M., *et al.*, *Intestinal epithelial cells use two distinct pathways for HLA class II antigen processing*. Journal of Clinical Investigation, 1997. **100**(1): 204-215.
165. Mayer, L. and D. Eisenhardt, *Lack of induction of suppressor T cells by intestinal epithelial cells from patients with inflammatory bowel disease*. Journal of Clinical Investigation, 1990. **86**: 1255-01260.
166. Kelleher, D., *et al.*, *Defective suppression in the autologous mixed lymphocyte reaction in patients with Crohn's disease*. Gut, 1989. **30**(6): 839-44.
167. Bradford-Hill, A., *The Environment and Disease: Association or Causation?* Proc. Royal Soc. Med, 1965. **58**: 295-300.
168. Dammeier J, B.M., Falk W, Grotendorst GR, Werner S., *Connective tissue growth factor: a novel regulator of mucosal repair and fibrosis in inflammatory bowel disease?* Int J Biochem Cell Biol, 1998. **30**(8): 909-22.
169. Hendel J, N.O., *Expression of cyclooxygenase-2 mRNA in active inflammatory bowel disease*. Am J Gastroenterol, 1997. **92**(7): 1170-3.
170. Yang, H., *et al.*, *Intestinal inflammation reduces expression of DRA, a transporter responsible for congenital chloride diarrhea*. American Journal of Physiology, 1998. **275**(6 (Pt 1)): G1445-53.
171. Keates S, K.A., Mizoguchi E, Bhan A, Kelly CP., *Enterocytes are the primary source of the chemokine ENA-78 in normal colon and ulcerative colitis*. Am J Physiol, 1997. **273**(1 Pt1): G75-82.
172. Garcia-Zepeda E.A., *et al.*, *Human eotaxin is a specific chemoattractant for eosinophil cells and provides a new mechanism to explain tissue eosinophilia*. Nature Medicine, 1996. **2**(4): 449-56.
173. Funakoshi K, S.K., Anezaki K, Bannai H, Ishizuka K, Asakura H., *Spectrum of cytokine gene expression in intestinal mucosal lesions of Crohn's disease and ulcerative colitis*. Digestion, 1998. **59**(1): 73-8.
174. Arai F, T.T., Furukawa K, Matsushima K, Asakura H., *Mucosal expression of interleukin-6 and interleukin-8 messenger RNA in ulcerative colitis and in Crohn's disease*. Dig Dis Sci, 1998. **43**(9): 2071-9.
175. Holub MC, M.E., Devay T, Dank M, Szalai C, Fenyvesi A, Falus A., *Increased interleukin-6 levels, interleukin-6 receptor and gp130 expression in peripheral lymphocytes of patients with inflammatory bowel disease*. Scand J Gastroenterol Suppl, 1998. **228**: 47-50.
176. McLaughlan, J.M., *et al.*, *Interleukin-8 and inducible nitric oxide synthase mRNA levels in inflammatory bowel disease at first presentation*. Journal of Pathology, 1997. **181**: 87-92.
177. Zhang XJ, T.J., Mannick EE, Correa P, Miller MJ., *Localization of inducible nitric oxide synthase mRNA in inflamed gastrointestinal mucosa by in situ reverse transcriptase-polymerase chain reaction*. Nitric Oxide, 1998. **2**(3): 187-92.

178. Berrebi D, B.A., Paris R, Potet F, Aigrain Y, Emilie D, Cezard JP, Hugot JP, Navarro J, Peuchmaur M., *In situ* Rantes and interferon-gamma gene expression in pediatric small bowel Crohn's disease. *J Pediatr Gastroenterol Nutr.* 1997. **25**(4): 371-6.
179. Waldegger, S., *et al.*, *h-sgk serine-threonine protein kinase gene as transcriptional target of transforming growth factor beta in human intestine.* *Gastroenterology.* 1999. **116**(5): 1081-8.
180. Dieckgraefe, B.K., *et al.*, *Analysis of mucosal gene expression in inflammatory bowel disease by parallel oligonucleotide arrays.* *Physiological Genomics.* 2000. **4**(1): 1-11.
181. Lawrance, I.C., C. Fiocchi, and S. Chakravarti, *Ulcerative colitis and Crohn's disease: distinctive gene expression profiles and novel susceptibility candidate genes.* *Human Molecular Genetics.* 2001. **10**(5): 445-456.
182. Alwine, J.C., D.J. Kemp, and G.R. Stark, *Method for detection of specific RNAs in agarose gels by transfer to diazobenzoyloxymethyl-paper and hybridization with DNA probes.* *Proceedings of the National Academy of Sciences U.S.A.*, 1977. **74**(12): 5350-4.
183. Berk A.J. and S. P.A., *Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids.* *Cell.* 1977. **12**(3): 721-32.
184. Hedrick S.M., *et al.*, *Isolation of cDNA clones encoding T cell-specific membrane-associated proteins.* *Nature.* 1984. **308**(5955): 149-53.
185. Adams, M.D., *et al.*, *Complementary DNA sequencing: Expressed sequence tags and human genome project.* *Science.* 1991. **252**: 1651-1656.
186. Liang, P. and A.B. Pardee, *Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction.* *Science.* 1992. **257**: 967-971.
187. Velculescu, V.E., *et al.*, *Serial analysis of gene expression.* *Science.* 1995. **270**: 484-487.
188. Maskos, U. and E.M. Southern. *A novel method for the parallel analysis of multiple mutations in multiple samples.* *Nucleic Acids Research.* 1993. **21**(9): 2269-2270.
189. Maskos, U. and E.M. Southern. *A novel method for the analysis of multiple sequence variants by hybridisation to oligonucleotides.* *Nucleic Acids Research.* 1993. **21**(9): 2267-2268.
190. Schena, M., *et al.*, *Quantitative monitoring of gene expression patterns with a complementary DNA microarray.* *Science.* 1995. **270**(5235): 467-470.
191. Schena, M., *et al.*, *Parallel human genome analysis: microarray-based expression monitoring of 1000 genes.* *Proceedings of the National Academy of Sciences U.S.A.*, 1996. **93**(20): 10614-9.
192. Kane, M.D., *et al.*, *Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays.* *Nucleic Acids Research.* 2000. **28**(22): 4552-7.

193. Sirover, M.A., *New insights into an old protein: the functional diversity of mammalian glyceraldehyde-3-phosphate dehydrogenase*. *Biochim Biophys Acta*, 1999. **1432**(2): 159-84.
194. Savonet, V., *et al.*, *Pitfalls in the use of several "housekeeping" genes as standards for quantitation of mRNA: the example of thyroid cells*. *Analytical Biochemistry*, 1997. **247**(1): 165-7.
195. Eickhoff, B., *et al.*, *Normalization of array hybridization experiments in differential gene expression analysis*. *Nucleic Acids Research*, 1999. **27**(22): e33.
196. Bhatia, P., *et al.*, *Comparison of glyceraldehyde-3-phosphate dehydrogenase and 28S-ribosomal RNA gene expression as RNA loading controls for northern blot analysis of cell lines of varying malignant potential*. *Analytical Biochemistry*, 1994. **216**(1): 223-6.
197. Heid, C.A., *et al.*, *Real time quantitative PCR*. *Genome Research*, 1996. **6**(10): 986-94.
198. Gibson, U.E.M., C.A. Heid, and P.M. Williams, *A novel method for real time quantitative RT-PCR*. *Genome Research*, 1996. **6**(10): 995-1001.
199. Lennox, B., *Observations on the accuracy of point counting including a description of a new graticule*. *Journal of Clinical Pathology*, 1975. **28**: 99-103.
200. Rogge, L., *et al.*, *Transcript imaging of the development of human T helper cells using oligonucleotide arrays*. *Nature Genetics*, 2000. **25**: 96-101.
201. Dozmorov, I., A. Bartke, and R.A. Miller, *Array-based expression analysis of mouse liver genes: effect of age and of the longevity mutant Prop1df*. *Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 2001. **56**(2): B72-80.
202. Baldi, P. and A.D. Long, *A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes*. *Bioinformatics*, 2001. **17**(6): 509-19.
203. Long, A.D., *et al.*, *Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework*. *Journal of Biological Chemistry*, 2001. **276**(23): 19937-44.
204. Glynn, R., *et al.*, *How self-tolerance and the immunosuppressive drug FK506 prevent B-cell mitogenesis*. *Nature*, 2000. **403**(6770): 672-6.
205. Heller, R.A., *et al.*, *Discovery and analysis of inflammatory disease-related genes using cDNA microarrays*. *Proceedings of the National Academy Sciences USA*, 1997. **94**(6): 2150-2155.
206. Tamayo, P., *et al.*, *Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation*. *Proceedings of the National Academy of Sciences*, 1999. **96**(6): 2907-2912.
207. Lu, J., *et al.*, *Gene expression profile changes in initiation and progression of squamous cell carcinoma of esophagus*. *International Journal of Cancer*, 2001. **91**(3): 288-294.

208. Alizadeh, A.A., *et al.*, *Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling*. *Nature*, 2000. **404**: 503-511.
209. Bassett(Jr), D.E., M.B. Eisen, and M.S. Bouguski, *Gene expression informatics - it's all in your mine*. *Nature Genetics Supplement*, 1999. **21**: 51-55.
210. Herrero, J., A. Valencia, and J. Dopazo, *A hierarchical unsupervised growing neural network for clustering gene expression patterns*. *Bioinformatics*, 2001. **17**(2): 126-136.
211. Campos, M.M. and G.A. Carpenter, *S-TREE: self-organizing trees for data clustering and online vector quantization*. *Neural Networks*, 2001. **14**(4-5): 505-25.
212. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for CSNK1G3* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997  
<http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?CSNK1G3>
213. Zhai, L., *et al.*, *Casein kinase I gamma subfamily*. *Journal of Biological Chemistry*, 1995. **270**(21): 12717-24.
214. PubMed (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>) search under *IBD AND casein kinase on*
215. Kimura, H., *et al.*, *Increased nitric oxide production and inducible nitric oxide synthase activity in colonic mucosa of patients with active ulcerative colitis and Crohn's disease*. *Digestive Disease and Sciences*, 1997. **42**(5): 1047-54.
216. Armour, K.E., *et al.*, *Evidence for a pathogenic role of nitric oxide in inflammation-induced osteoporosis*. *Journal of Bone Mineral Research*, 1999. **14**(12): 2137-42.
217. Kubes, P. and D.-M. McCafferty, *Nitric oxide and intestinal inflammation*. *American Journal of Medicine*, 2000. **109**(2): 150-8.
218. Jankowski S.A., *et al.*, *SAS, a gene amplified in human sarcomas, encodes a new member of the transmembrane 4 superfamily of proteins*. *Oncogene*, 1994. **9**(4): 1205-11.
219. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for SAS* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997  
<http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?SAS>
220. Clapper, M.L. and C.E. Szarka, *Glutathione S-transferases--biomarkers of cancer risk and chemopreventive response*. *Chemico-biological interactions*, 1998. **111-112**: 377-88.
221. Lieshout, E.M.M.v., *et al.*, *Nonsteroidal anti-inflammatory drugs enhance glutathione S-transferase theta levels in rat colon*. *Biochim Biophys Acta*, 1998. **1381**(3): 305-11.
222. Landi, S., *Mammalian class theta GST and differential susceptibility to carcinogens: a review*. *Mutation Research*, 2000. **463**(3): 247-83.

223. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for GSTT2* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997  
<http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?GSTT2>
224. Coggan, M., *et al.*, *Structure and organization of the human theta-class glutathione S-transferase and D-dopachrome tautomerase gene complex*. *Biochemical Journal*, 1998. **334**(3): 617-23.
225. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for CRSP2* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997  
<http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?CRSP2>
226. Gray, G.E., *et al.*, *Human ligands of the Notch receptor*. *American Journal of Pathology*, 1999. **154**(3): 785-794.
227. Deng Y, *et al.*, *Characterization, chromosomal localization, and the complete 30-kb DNA sequence of the human Jagged2 (JAG2) gene*. *Genomics*, 2000. **63**(1): 133-8.
228. Jiang, R., *et al.*, *Defects in limb, craniofacial, and thymic development in Jagged2 mutant mice*. *Genes Dev*, 1998. **12**(7): 1046-4057.
229. Lindner, V., *et al.*, *Members of the Jagged/Notch gene families are expressed in injured arteries and regulate cell phenotype via alterations in cell matrix and cell-cell interaction*. *American Journal of Pathology*, 2001. **159**(3): 875-83.
230. Tsai, S., J. Fero, and S. Bartelmez, *Mouse Jagged2 is differentially expressed in hematopoietic progenitors and endothelial cells and promotes the survival and proliferation of hematopoietic progenitors by direct cell-to-cell contact*. *Blood*, 2000. **96**(3): 950-7.
231. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for ERF* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997  
<http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?ERF>
232. Sacchi N, *et al.*, *Hu-ets-1 and Hu-ets-2 genes are transposed in acute leukemias with (4;11) and (8;21) translocations*. *Science*, 1986. **231**(4736): 379-82.
233. Sumarsono S.H., *et al.*, *Down's syndrome-like skeletal abnormalities in Ets2 transgenic mice*. *Nature*, 1996. **379**(6565): 534-540.
234. Dooley, S., *et al.*, *Constitutive expression of c-fos and c-jun, overexpression of ets-2, and reduced expression of metastasis suppressor gene nm23-H1 in rheumatoid arthritis*. *Annals of the Rheumatic Diseases*, 1996. **55**(5): 298-304,.
235. Nantais D.E., *et al.*, *Prenylation of an interferon-gamma-induced GTP-binding protein: the human guanylate binding protein, huGBP1*. *Journal of Leukocyte Biology*, 1996. **60**(3): 423-31.

236. Toggas S.M., Krady J.K., and B. M.L., *Molecular neurotoxicology of trimethyltin: identification of stannin, a novel protein expressed in trimethyltin-sensitive cells*. *Molecular Pharmacology*, 1992. **42**(1): 44-56.
237. Thompson T.A., *et al.*, *Induction of apoptosis by organotin compounds in vitro: neuronal protection with antisense oligonucleotides directed against stannin*. *Journal of Pharmacology and Experimental Therapy*, 1996. **276**(3): 1201-16.
238. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for Snn* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997  
<http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?Snn>
239. Millenium Pharmaceuticals, I., *22105, a human thioredoxin family member and uses thereof*. . 2001: USA.
240. Hofer, D., *et al.*, *From cytoskeleton to polarity and chemoreception in the gut epithelium*. *Annals of New York Academy of Sciences*, 1998. **859**: 75-84.
241. Karki, S. and E.L. Holzbaur, *Cytoplasmic dynein and dynactin in cell division and intracellular transport*. *Current Opinion in Cell Biology*, 1999. **11**(1): 45-53.
242. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for CD9* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997  
<http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?CD9>
243. Okochi, H., *et al.*, *Expression of tetraspans transmembrane family in the epithelium of the gastrointestinal tract*. *Journal of Clinical Gastroenterology*, 1999. **29**(1): 63-7.
244. Miyamoto S, *et al.*, *Loss of motility-related protein 1 (MRP1/CD9) and integrin alpha3 expression in endometrial cancers*. *Cancer*, 2001. **92**(3): 542-8.
245. Kusakawa J, *et al.*, *Reduced expression of CD9 in oral squamous cell carcinoma: CD9 expression inversely related to high prevalence of lymph node metastasis*. *J Oral Pathol Med*, 2001. **30**(2): 73-9.
246. Yau J.C., *et al.*, *Expression of transmembrane 4 superfamily member, CD9, is related to improved progression-free survival in patients with diffuse non-Hodgkin's lymphoma*. *Oncol Rep*, 1998. **5**(6): 1507-11.
247. Sho M, *et al.*, *Transmembrane 4 superfamily as a prognostic factor in pancreatic cancer*. *International Journal of Cancer*, 1998. **79**(5): 509-16.
248. Huang C.I., *et al.*, *Correlation of reduction in MRP-1/CD9 and KAI1/CD82 expression with recurrences in breast cancer patients*. *American Journal of Pathology*, 1998. **153**(3): 973-83.
249. Mori M, *et al.*, *Motility related protein 1 (MRP1/CD9) expression in colon cancer*. *Clinical Cancer Research*, 1998. **4**(6): 1507-10.

250. Ono M, *et al.*, *GM3 ganglioside inhibits CD9-facilitated haptotactic cell motility: coexpression of GM3 and CD9 is essential in the downregulation of tumor cell motility and malignancy*. *Biochemistry*, 2001. **40**(21): 6414-21.
251. Miyake M, *et al.*, *Suppression of pulmonary metastasis using adenovirally motility related protein-1 (MRP-1/CD9) gene delivery*. *Oncogene*, 2000. **19**(46): 5221-6.
252. Liu Q.A. and H. M.O., *Candidate adaptor protein CED-6 promotes the engulfment of apoptotic cells in C. elegans*. *Cell*, 1998. **93**(6): 961-72.
253. Smits, E., *et al.*, *The human homologue of Caenorhabditis elegans CED-6 specifically promotes phagocytosis of apoptotic cells*. *Current Biology*, 1999. **9**(22): 1351-4.
254. Brennan, F.E. and P.J. Fuller, *Rapid upregulation of serum and glucocorticoid-regulated kinase (sgk) gene expression by corticosteroids in vivo*. *Molecular and Cellular Endocrinology*, 2000. **166**(2): 129-36.
255. Shigaev, A., *et al.*, *Regulation of sgk by aldosterone and its effects on the epithelial Na(+) channel*. *American Journal of Physiology - Renal Physiology*, 2000. **278**(4): F613-9.
256. Bhargava, A., *et al.*, *The serum- and glucocorticoid-induced kinase is a physiological mediator of aldosterone action*. *Endocrinology*, 2001. **142**(4): 1587-94.
257. Rotin, D., *Regulation of the epithelial sodium channel (ENaC) by accessory proteins*. *Current Opinion in Nephrology and Hypertension*, 2000. **9**(5): 529-34.
258. Greig, E. and G.I. Sandle, *Diarrhea in ulcerative colitis. The role of altered colonic sodium transport*. *Ann N Y Acad Sci*, 2000. **915**: 327-32.
259. Greener, T., *et al.*, *Role of cyclin G-associated kinase in uncoating clathrin-coated vesicles from non-neuronal cells*. *Journal of Biological Chemistry*, 2000. **275**(2): 1365-70.
260. Umeda, A., A. Meyerholz, and E. Ungewickell, *Identification of the universal cofactor (auxilin 2) in clathrin coat dissociation*. *European Journal of Cell Biology*, 2000. **79**(5): 336-42.
261. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for PCTK1* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997  
<http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?PCTK1>
262. Besset V, Rhee K, and W. D.J., *The cellular distribution and kinase activity of the Cdk family member Pctaire1 in the adult mouse brain and testis suggest functions in differentiation*. *Cell Growth Differ*, 1999. **10**(3): 173-81.
263. Charrasse S, *et al.*, *PCTAIRE-1: characterization, subcellular distribution, and cell cycle-dependent kinase activity*. *Cell Growth Differ*, 1999. **10**(9): 611-20.
264. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for RNASE6* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997

<http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?RNASE6>

265. Rosenberg, H.F. and K.D. Dyer, *Molecular cloning and characterization of a novel human ribonuclease (RNase k6): increasing diversity in the enlarging ribonuclease gene family*. Nucleic Acids Research, 1996. **24**(18): 3507-13.
266. Herceg Z. *et al.*, *Disruption of Trrap causes early embryonic lethality and defects in cell cycle progression*. Nature Genetics, 2001. **29**(2): 206-11.
267. McMahon, S.B., *et al.*, *The novel ATM-related protein TRRAP is an essential cofactor for the c-Myc and E2F oncoproteins*. Cell, 1998. **94**(3): 363-74.
268. Salicioni, A.M., *et al.*, *Identification and structural analysis of human RBM8A and RBM8B: two highly conserved RNA-binding motif proteins that interact with OVCA1, a candidate tumor suppressor*. Genomics, 2000. **69**(1): 54-62.
269. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for RBM8A* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997
- <http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?RBM8A>
270. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for RBM8B* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997
- <http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?RBM8B>
271. Kataoka, N., *et al.*, *Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm*. Molecular Cell, 2000. **6**(3): 673-82.
272. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for B2M* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997
- <http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?B2M>
273. Sartor, R.B., *Colitis in HLA-B27/beta 2 microglobulin transgenic rats*. International Reviews in Immunology, 2000. **19**(1): 39-50.
274. Nielsen, O.H., *et al.*, *Established and emerging biological activity markers of inflammatory bowel disease*. American Journal of Gastroenterology, 2000. **95**(2): 359-67.
275. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for SCYB10* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997
- <http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?SCYB10>
276. Uguccioni, M., *et al.*, *Increased expression of IP-10, IL-8, MCP-1, and MCP-3 in ulcerative colitis*. American Journal of Pathology, 1999. **155**(2): 331-6.
277. Dwinell, M.B., *et al.*, *Regulated production of interferon-inducible T-cell chemoattractants by human intestinal epithelial cells*. Gastroenterology, 2001. **120**(1): 49-59.

278. Liu, M.T., *et al.*, *Expression of Mig (monokine induced by interferon-gamma) is important in T lymphocyte recruitment and host defense following viral infection of the central nervous system.* Journal of Immunology, 2001. **166**(3): 1790-5.
279. Odh, G., *et al.*, *Isolation of a new tautomerase monitored by the conversion of D-dopachrome to 5,6-dihydroxyindole.* Biochemical and Biophysical Research Communications, 1993. **197**(2): 619-24.
280. Sugimoto, H., *et al.*, *Crystal structure of human D-dopachrome tautomerase, a homologue of macrophage migration inhibitory factor, at 1.54 Å resolution.* Biochemistry, 1999. **38**(11): 3268-79.
281. Nishihira, J., *et al.*, *Molecular Cloning of Human D-Dopachrome Tautomerase cDNA: N-terminal Proline Is Essential for Enzyme Activation.* Biochemical and Biophysical Research Communications, 1998. **243**(2): 538-44.
282. Esumi, N., *et al.*, *Conserved gene structure and genomic linkage for D-dopachrome tautomerase (DDT) and MIF.* Mammalian Genome, 1998. **9**(9): 753-7.
283. Fitzgerald, L.M. and H. Li, *Macrophage Migration Inhibitory Factor-3*, in *Delphion Intellectual Property Network*. 1995: WO World Intellectual Property Organization (WIPO). 41.
284. Pröls, F., *et al.*, *Upregulation of the Cochaperone Mdgl in Endothelial Cells Is Induced by Stress and during in Vitro Angiogenesis.* Experimental Cell Research, 2001. **269**(1): 42-53.
285. Prols F, *et al.*, *Assignment of the microvascular endothelial differentiation gene 1 (MDG1) to human chromosome band 14q24.2-->q24.3 by fluorescence in situ hybridization.* Cytogenet Cell Genet, 1997. **79**(1-2): 149-50.
286. Pröls, F., B. Loser, and M. Marx, *Differential expression of osteopontin, PC4, and CEC5, a novel mRNA species, during in vitro angiogenesis.* Experimental Cell Research, 1998. **239**(1): 1-10.
287. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for SEC14L1* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997  
<http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?SEC14L1>
288. Heike, M., *et al.*, *Expression of stress protein gp96, a tumor rejection antigen, in human colorectal cancer.* International Journal of Cancer, 2000. **86**(4): 489-93.
289. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for TRAI* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997  
<http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?TRAI>
290. Schild, H. and H.-G. Rammensee, *gp96—The immune system's Swiss army knife.* Nature Immunology, 2000. **1**(2): 100-101.
291. Binder R.J., Han D.K., and S. P.K., *CD91: a receptor for heat shock protein gp96.* Nature Immunology, 2000. **1**(2): 151-5.

292. Joberty, G., R.R. Perlungher, and I.G. Macara, *The Borgs, a new family of Cdc42 and TC10 GTPase-interacting proteins*. *Molecular and Cellular Biology*, 1999. **19**(10): 6585-97.
293. Hirsch, D.S., D.M. Pirone, and P.D. Burbelo, *A New Family of Cdc42 Effector Proteins, CEPs, Function in Fibroblast and Epithelial Cell Shape Changes*. *Journal of Biological Chemistry*, 2001. **276**(2): 875-883.
294. Burbelo, P.D., *et al.*, *MSE55, a Cdc42 effector protein, induces long cellular extensions in fibroblasts*. *Proceedings of the National Academy of Sciences U.S.A.*, 1999. **96**(16): 9083-8.
295. Fukuda M.N., *et al.*, *Trophinin and tascin, a novel cell adhesion molecule complex with potential involvement in embryo implantation*. *Genes Dev*, 1995. **9**(10): 1199-210.
296. Ogryzko V.V., *et al.*, *Histone-like TAFs within the PCAF histone acetylase complex*. *Cell*, 1998. **94**(1): 35-44.
297. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for HS2ST1* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997  
<http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?HS2ST1>
298. Rong, J., *et al.*, *Substrate specificity of the heparan sulfate hexuronic acid 2-O-sulfotransferase*. *Biochemistry*, 2001. **40**(18): 5548-5555.
299. Merry, C.L.R., *et al.*, *The Molecular Phenotype of Heparan Sulfate in the Hs2st/ Mutant Mouse*. *Journal of Biological Chemistry*, 2001. **276**(38): 35429-35434.
300. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for PEX14* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997  
<http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?PEX14>
301. Communi, D., V. Vanweyenberg, and C. Erneux, *Molecular study and regulation of D-myo-inositol 1,4,5-trisphosphate 3-kinase*. *Cell Signalling*, 1995. **7**(7): 643-50.
302. Irvine, R., *Inositol phosphates: Does IP(4) run a protection racket?* *Current Biology*, 2001. **11**(5): R172-4.
303. Luckhoff, A. and D.E. Clapham, *Inositol 1,3,4,5-tetrakisphosphate activates an endothelial Ca(2+)-permeable channel*. *Nature*, 1992. **355**(6358): 356-8.
304. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for SLC26A2* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997  
<http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?SLC26A2>
305. Liljedahl M, *et al.*, *HLA-DO is a lysosomal resident which requires association with HLA-DM for efficient intracellular transport*. *EMBO Journal*, 1996. **15**(18): 4817-24.
306. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for CD22* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997  
<http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?CD22>

307. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for SCYA11* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997  
<http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?SCYA11>
308. Chen W, *et al.*, *Increased serum levels of eotaxin in patients with inflammatory bowel disease.* Scandinavian Journal of Gastroenterology, 2001. **36**(5): 515-20.
309. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for CD2* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997  
<http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?CD2>
310. Hu M.C., *et al.*, *Human HPK1, a novel human hematopoietic progenitor kinase that activates the JNK/SAPK kinase cascade.* Genes Dev, 1996. **10**(18): 2251-64.
311. Rao, K.M.K., *MAP kinase activation in macrophages.* Journal of Leukocyte Biology, 2001. **69**(1): 3-10.
312. Tsuji, S., *et al.*, *B cell adaptor containing src homology 2 domain (BASH) links B cell receptor signaling to the activation of hematopoietic progenitor kinase 1.* Journal of Experimental Medicine, 2001. **194**(4): 529-39.
313. Marshall A.J., *et al.*, *A novel B lymphocyte-associated adaptor protein, Bam32, regulates antigen receptor signaling downstream of phosphatidylinositol 3-kinase.* Journal of Experimental Medicine, 2000. **191**(8): 1319-32.
314. Anderson, K.E., *et al.*, *DAPP1 undergoes a PI 3-kinase-dependent cycle of plasma-membrane recruitment and endocytosis upon cell stimulation.* Current Biology, 2000. **10**(22): 1403-12.
315. Rao V.R., *et al.*, *Expression cloning of protein targets for 3-phosphorylated phosphoinositides.* Journal of Biological Chemistry, 1999. **274**(53): 37893-900.
316. Brooksbank, C., *Pocket the difference.* Nature Reviews Molecular Cell Biology, 2000. **1**(1): 9.
317. Farber, J.M., *Mig and IP-10: CXC chemokines that target lymphocytes.* Journal of Leukocyte Biology, 1997. **61**(3): 246-57.
318. Shibahara, T., *et al.*, *Characterization of epithelial chemoattractants for human intestinal intraepithelial lymphocytes.* Gastroenterology, 2001. **120**(1): 60-70.
319. Goebeler, M., *et al.*, *Differential and sequential expression of multiple chemokines during elicitation of allergic contact hypersensitivity.* American Journal of Pathology, 2001. **158**(2): 431-40.
320. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for NR5A2* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997  
<http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?NR5A2>
321. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for PTP4A1* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997

<http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?PTP4A1>

322. Diamond RH, P.C., Jung SP, Greenbaum LE, Haber BA, Silberg DG, Traber PG, Taub R., *Expression of PRL-1 nuclear PTPase is associated with proliferation in liver but with differentiation in intestine.* American Journal of Physiology, 1996. **271**(1): G121-9.
323. Kong W, *et al.*, *PRL-1 PTPase expression is developmentally regulated with tissue-specific patterns in epithelial tissues.* American Journal of Physiology - Gastrointestinal & Liver Physiology, 2000. **279**(3): G613-21.
324. Gyorloff-Wingren A, *et al.*, *Subcellular localization of intracellular protein tyrosine phosphatases in T cells.* European Journal of Immunology, 2000. **30**(8): 2412-21.
325. Charles S. Peters, *et al.*, *ATF-7, a Novel bZIP Protein, Interacts with the PRL-1 Protein-tyrosine Phosphatase.* Journal of Biological Chemistry, 2001. **276**(17): 13718-13726.
326. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for ATF7* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997

<http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?ATF7>

327. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for ZNF137* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997

<http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?ZNF137>

328. Hartupée J.C., *et al.*, *Isolation and characterization of a cDNA encoding a novel member of the human regenerating protein family: Reg IV.* Biochimica et Biophysica Acta, 2001. **1518**(3): 287-93.
329. Macadam R.C., *et al.*, *Death from early colorectal cancer is predicted by the presence of transcripts of the REG gene family.* British Journal of Cancer, 2000. **83**(2): 188-95.
330. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for GBP2* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997

<http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?GBP2>

331. Treeck, O., E. Strunck, and G. Vollmer, *A novel basement membrane-induced gene identified in the human endometrial adenocarcinoma cell line HEC1B.* FEBS Letters, 1998. **425**(3): 426-30.
332. Frye, R.A., *Characterization of five human cDNAs with homology to the yeast SIR2 gene: Sir2-like proteins (sirtuins) metabolize NAD and may have protein ADP-ribosyltransferase activity.* Biochemical and Biophysical Research Communications, 1999. **260**(1): 273-9.
333. Bowtell, D.D.L., *Options available - from start to finish - for obtaining expression data by microarray.* Nature Genetics Supplement, 1999. **21**: 25-32.
334. Klebes, A., *et al.*, *Expression profiling of Drosophila imaginal discs.* Genome Biology, 2002. **3**(8): 1-16.

335. Bertucci, F., *et al.*, *Sensitivity issues in DNA array-based expression measurements and performance of nylon microarrays for small samples*. Human Molecular Genetics, 1999. **8**(9): 1715-22.
336. Lee, P.S. and K.H. Lee, *Genomic analysis*. Current Opinion in Biotechnology, 2000. **11**: 171-5.
337. Roediger, W.E.W., *The colonic epithelium in ulcerative colitis: an energy deficiency disease?* The Lancet, 1980. **2**: 712-5.
338. Ahmed, M.S., *et al.*, *Butyrate and glucose metabolism by colonocytes in experimental colitis in mice*. Gut, 2000. **46**: 493-9.
339. Chapman, M.A., *et al.*, *Butyrate oxidation is impaired in the colonic mucosa of sufferers of quiescent ulcerative colitis*. Gut, 1994. **35**(1): 73-6.
340. Roediger, W.E.W. and S. Nance, *Selective reduction of fatty acid oxidation in colonocytes: correlation with ulcerative colitis*. Lipids, 1990. **25**(10): 646-52.
341. Moore, J.W., *et al.*, *Hydrogen sulphide produces diminished fatty acid oxidation in the rat colon in vivo: implications for ulcerative colitis*. Aust N Z J Surg, 1997. **67**(5): 245-9.
342. Jorgensen, J. and P.B. Mortensen, *Substrate utilization by intestinal mucosal tissue strips from patients with inflammatory bowel disease*. American Journal of Physiology - Gastrointestinal & Liver Physiology, 2001. **281**(2): G405-11.
343. Simpson, E.J., *et al.*, *In vivo measurement of colonic butyrate metabolism in patients with quiescent ulcerative colitis*. Gut, 2000. **46**: 73-7.
344. Allan, E.S., *et al.*, *Mucosal enzyme activity for butyrate oxidation; no defect in patients with ulcerative colitis*. Gut, 1996. **38**: 886-893.
345. Hulin, S.J., *et al.*, *Sulphide-induced energy deficiency in colonic cells is prevented by glucose but not by butyrate*. Alimentary Pharmacology Therapeutics, 2002. **16**(2): 325-31.
346. Pitcher, M.C.L., E.R. Beatty, and J.H. Cummings, *The contribution of sulphate reducing bacteria and 5-aminosalicylic acid to faecal sulphide in patients with ulcerative colitis*. Gut, 2000. **46**: 64-72.
347. Christl, S.U., *et al.*, *Antagonistic effects of sulphide and butyrate on proliferation of colonic mucosa*. Digestive Diseases and Sciences, 1996. **41**(12): 2477-81.
348. Francomano, C.A., I. McIntosh, and D.J. Wilkin, *Bone dysplasias in man: molecular insights*. Current Opinion in Genetics and Development, 1996. **6**: 301-308.
349. Hastbacka, J., *et al.*, *The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping*. Cell, 1994. **78**(1073-87).
350. Haila, S., *et al.*, *The congenital chloride diarrhea gene is expressed in seminal vesicle, sweat gland, inflammatory colon epithelium, and in some dysplastic colon cells*. Histochemistry and Cell Biology, 2000. **113**(4): 279-86.

351. Markovich, D., *Physiological Roles and Regulation of Mammalian Sulfate Transporters*. *Physiological Reviews*, 2001. **81**(4): 1499-1533.
352. Everett, L.A. and E.D. Green, *A family of mammalian anion transporters and their involvement in human genetic diseases*. *Human Molecular Genetics*, 1999. **8**(10): 1883-91.
353. Satob, H., *et al.*, *Functional analysis of diastrophic dysplasia sulfate transporter*. *The Journal of Biological Chemistry*, 1998. **273**(20): 12307-15.
354. Byeon, M.K., *et al.*, *Human DRA functions as a sulfate transporter in Sf9 insect cells*. *Protein Expr Purif*, 1998. **12**(1): 67-74.
355. Melvin, J.E., *et al.*, *Mouse down-regulated in adenoma (DRA) is an intestinal Cl(-)/HCO(3)(-) exchanger and is up-regulated in colon of mice lacking the NHE3 Na(+)/H(+) exchanger*. *Journal of Biological Chemistry*, 1999. **274**(32): 22855-61.
356. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for SLC26A3* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997  
<http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?SLC26A3>
357. Lohi, H., *et al.*, *Mapping of Five New Putative Anion Transporter Genes in Human and Characterization of SLC26A6, A Candidate Gene for Pancreatic Anion Exchanger*. *Genomics*, 2001. **70**(1): 102-112.
358. Alrefai WA, T.S., Mansour F, Saksena S, Syed I, Ramaswamy K, Dudeja PK., *Sulfate and chloride transport in Caco-2 cells: differential regulation by thyroxine and the possible role of DRA gene*. *Am J Physiol Gastrointest Liver Physiol*, 2001. **280**(4): G603-13.
359. Barcelo, A., *et al.*, *Mucin secretion is modulated by luminal factors in the isolated vascularly perfused rat colon*. *Gut*, 2000. **46**(2): 218-224.
360. Mandal, M., *et al.*, *Butyric Acid Induces Apoptosis by Up-regulating Bax Expression via Stimulation of the c-Jun N-Terminal Kinase/Activation Protein-1 Pathway in Human Colon Cancer Cells*. *Gastroenterology*, 2001. **120**(1): 71-78.
361. Inan, M.S., *et al.*, *The luminal short-chain fatty acid butyrate modulates NF- $\kappa$ B activity in a human colonic epithelial cell line*. *Gastroenterology*, 2000. **118**(4): 724-734.
362. Lührs, H., *et al.*, *Cytokine-activated degradation of inhibitory kappaB protein alpha is inhibited by the short-chain fatty acid butyrate*. *International Journal of Colorectal Disease*, 2001. **16**(4): 195-201.
363. Roediger, W.E. and S. Nance, *Metabolic induction of experimental ulcerative colitis by inhibition of fatty acid oxidation*. *British Journal of Experimental Pathology*, 1986. **67**(6): 773-82.
364. Babidge, W., S. Millard, and W. Roediger, *Sulfides impair short chain fatty acid beta-oxidation at acyl-CoA dehydrogenase level in colonocytes: implications for ulcerative colitis*. *Mol Cell Biochem*, 1998. **181**(1-2): 117-24.

365. Moore, J.W., *et al.*, *Effect of sulphide on short chain acyl-CoA metabolism in rat colonocytes*. Gut, 1997. **41**(1): 77-81.
366. Roediger, W.E.W., *et al.*, *Sulphide impairment of substrate oxidation in rat colonocytes: a biochemical basis for ulcerative colitis?* Clinical Science, 1993. **85**: 623-7.
367. Volkel, S. and M.K. Grieshaber, *Sulphide oxidation and oxidative phosphorylation in the mitochondria of the lugworm Arenicola marina*. The Journal of Experimental Biology, 1997. **200**: 83-92.
368. Jorgensen, J. and P.B. Mortensen, *Hydrogen sulfide and colonic epithelial metabolism: implications for ulcerative colitis*. Digestive Diseases and Sciences, 2001. **46**(8): 1722-32.
369. Levine, J., *et al.*, *Fecal hydrogen sulfide production in ulcerative colitis*. American Journal of Gastroenterology, 1998. **93**(1): 83-7.
370. Levitt, M.D., *et al.*, *Physiology of sulfide in the rat colon: use of bismuth to assess colonic sulfide production*. Journal of Applied Physiology, 2002. **92**(4): 1655-60.
371. Roediger, W.E. and W.J. Babidge, *Thiol methyltransferase activity in inflammatory bowel disease*. Gut, 2000. **47**(2): 206-10.
372. Roediger, W.E., W. Babidge, and S. Millard, *Methionine derivatives diminish sulphide damage to colonocytes--implications for ulcerative colitis*. Gut, 1996. **39**(1): 77-81.
373. Moore, J.W.E., *et al.*, *Thiolmethyltransferase activity in the human colonic mucosa: Implications for ulcerative colitis*. Journal of Gastroenterology and Hepatology, 1997. **12**: 678-84.
374. Levitt, M.D., *et al.*, *Detoxification of hydrogen sulfide and methanethiol in the cecal mucosa*. The Journal of Clinical Investigation, 1999. **104**(8): 1107-14.
375. Latella, G., *et al.*, *Carbonic anhydrase 1 expression in colonic mucosa of patients with ulcerative colitis*. Gastroenterology, 1995. **108**(4): A858.
376. Stocchi, A., *et al.*, *Study of constancy of hydrogen-consuming flora of human colon*. Digestive Diseases and Sciences, 1994. **39**(3): 494-7.
377. McKay, L.F., M.A. Eastwood, and W.G. Brydon, *Methane excretion in man - a study of breath, flatus, and faeces*. Gut, 1985. **26**: 69-74.
378. Gibson, G.R., *et al.*, *Alternative pathways for hydrogen disposal during fermentation in the human colon*. Gut, 1990. **31**: 679-83.
379. Florin, T., *et al.*, *Metabolism of dietary sulphate: absorption and excretion in humans*. Gut, 1991. **32**: 766-73.
380. Chou, H.-F., M. Passage, and A.J. Jonas, *ATP stimulates lysosomal sulphate transport at neutral pH: evidence for phosphorylation of the lysosomal sulphate carrier*. Biochemical Journal, 1997. **327**: 781-6.

381. Roediger, W.E., *The starved colon--diminished mucosal nutrition, diminished absorption, and colitis*. Diseases of the Colon and Rectum, 1990. **33**(10): 858-62.
382. Breuer, R.I., *et al.*, *Short chain fatty acid rectal irrigation for left-sided ulcerative colitis: a randomised, placebo controlled trial*. Gut, 1997. **40**: 485-91.
383. Roediger, W.E.W., *Decreased sulphur aminoacid intake in ulcerative colitis*. The Lancet, 1998. **351**: 1555.
384. Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: *Gene Card for MUC2* GeneCards: encyclopedia for genes, proteins and diseases. (Rehovot, Israel), 1997  
<http://bioinformatics.weizmann.ac.il/cards-bin/carddisp?MUC2>
385. Pullan RD, T.G., Rhodes M, Newcombe RG, Williams GT, Allen A, Rhodes J., *Thickness of adherent mucus gel on colonic mucosa in humans and its relevance to colitis*. Gut, 1994. **35**(3): 353-9.
386. Pullan, R.D., *Colonic mucus, smoking and ulcerative colitis*. Ann R Coll Surg Engl, 1996. **78**(2): 85-91.
387. Tarnawski, A.S. and I. Szabo, *Apoptosis - Programmed cell death and its relevance to gastrointestinal epithelium: survival signal from the matrix*. Gastroenterology, 2001. **120**(1): 294-298.
388. Roediger, W.E. and W.J. Babidge. *Nitric oxide effect on colonocyte metabolism: co-action of sulfides and peroxide*. Mol Cell Biochem, 2000. **206**(1-2): 159-167.
389. Suarez, F.L., *et al.*, *Bismuth subsalicylate markedly decreases hydrogen sulfide release in the human colon*. Gastroenterology, 1998. **114**: 923-9.
390. Furne, J.K., *et al.*, *Binding of hydrogen sulfide by bismuth does not prevent dextran sulfate-induced colitis in rats*. Dig Dis Sci, 2000. **45**(7): 1439-43.
391. King, R.J.B., *Cancer Biology*. 1996, Edinburgh: Longman. 227.
392. Raff, M., *Cell suicide for beginners*. Nature, 1998. **396**: 119-122.
393. Cohen, G.M., *Caspases: the executioners of apoptosis*. Biochemical Journal, 1997. **326**(1): 1-16.
394. Fogt, F., *et al.*, *Expression of survivin, YB-1, and KI-67 in sporadic adenomas and dysplasia-associated lesions or masses in ulcerative colitis*. Appl Immunohistochem Molecul Morphol, 2001. **9**(2): 143-149.
395. Altieri, D.C. and P.C. Marchisio. *Survivin apoptosis: an interloper between cell death and cell proliferation in cancer*. Lab Invest, 1999. **79**(11): 1327.
396. Goyal, L., *Cell death inhibition: keeping caspases in check*. Cell, 2001. **104**(6): 805-808.
397. Sarela, A.I., *et al.*, *Expression of the antiapoptosis gene, survivin, predicts death from recurrent colorectal carcinoma*. Gut, 2000. **46**(5): 645-50.
398. Boirivant, M., *et al.*, *Lamina propria T cells in Crohn's disease and other gastrointestinal inflammation show defective CD2 pathway-induced apoptosis*. Gastroenterology, 1999. **116**(3): 557-65.

**APPENDIX**

**A2 Appendix for Chapter 2 (Materials & Methods)**

---

**A2.1 1% Ethidium Bromide Gel**

5 g agarose (GibcoBRL #15510-027) and 10 µl ethidium bromide (Sigma #E-1510) was made up to 500 ml with 0.5% TBE buffer (10X TBE buffer from GibcoBRL #15581-036). Heating for 2-3 min in a microwave dissolved the agarose, which was then left to cool slightly before being poured into gel racks.

**A2.2 5X Fragmentation Buffer (20 ml)**

*(200 mM Tris-acetate, pH8.1, 500 mM KOAc, 150 mM MgOAc)*

4.0 ml 1 M Tris acetate pH8.1 {Trizma base (Sigma #T-1503), pH adjusted with glacial acetic acid (Sigma #A-6283)}

0.64 g MgOAc (Sigma #M-2545)

0.98 g KOAc (Sigma #P-5708)

DEPC-treated H<sub>2</sub>O to 20 ml

Mixed and filtered through a 0.2 µm vacuum filter unit (Nalgene #), and stored at room temperature.

**A2.3 12X MES Stock (1 L) (1.22 M MES, 0.89 M [Na<sup>+</sup>])**

70.4 g MES free acid monohydrate (Sigma #M-5287)

193.3 g MES Sodium salt (Sigma #M-3058)

800 ml sterile ultra-pure H<sub>2</sub>O.

Mix and adjust volume to 1000 ml. pH should be between 6.5 and 6.7. Filter through a 0.2 µm filter (Nalgene #). Stored at 2-8°C, discarded if discoloured and fresh solution made up.

**A2.4 2X Hybridisation Buffer (50 ml)**

*(final 1X concentration is 100 mM MES, 1 M [Na<sup>+</sup>], 20 mM EDTA, 0.01% Tween 20)*

8.3 ml 12X MES stock

17.7 ml 5 M NaCl (RNase & DNase free) (Sigma #S-5150)

4.0 ml 0.5 M EDTA

0.1 ml 10% Tween 20 (CalBiochem #655206)

19.9 ml DEPC-treated H<sub>2</sub>O

Store at 2-8°C.

**A2.5 Antifoam 0-30 (Sigma #A8082)**

Prepared a 5% solution in DEPC-treated H<sub>2</sub>O and stored at room temperature.

**A2.6 Wash buffer A - Non-Stringent Wash Buffer (1 L)**

*(6X SSPE, 0.01% Tween 20, 0.005% Anti-foam)*

300 ml 20X SSPE

1.0 ml 10% Tween 20

698 ml Ultra-pure water

Filtered through 0.2 µm filter, and 1.0 ml 5% Anti-foam 0-30 added. Stored at room temperature.

**A2.7 Wash buffer B - Stringent Wash Buffer (1 L)***(100 mM MES, 0.1 M [Na<sup>+</sup>], 0.01% Tween 20)*

83.3 ml 12X MES stock

5.2 ml 5 M NaCl

1.0 ml 10% Tween 20

910.5 ml Ultra-pure water

Filtered through 0.2 µm filter and stored at room temperature.

**A2.8 2X Stain Buffer (250 ml)***(1X concentration: 100 mM MES, 1 M [Na<sup>+</sup>], 0.05% Tween 20, 0.005% Anti-foam)*

41.7 ml 12X MES stock

92.5 ml 5M NaCl

2.5 ml 10% Tween 20

112.8 ml Ultra-pure water

Filtered through 0.2 µm filter and 0.5 ml 5% Anti-foam 0-30 added. Stored at room temperature.

**A2.9 SAPE stain solution (600 µl)**

300 µl 2X Stain buffer

270 µl DEPC-treated H<sub>2</sub>O

24 µl of 50 mg/ml acetylated BSA (final concentration of 10 µg/ml)

6 µl of 1 mg/ml streptavidin phycoerythrin (SAPE) (final concentration of 10 µg/ml)

3 µl of 5% antifoam solution

For 1200 µl of this solution, the amounts were doubled, and 600 µl was aliquoted into two 1.5 ml Eppendorf tubes.

**A2.10 Antibody solution (600 µl)**

300 µl 2X Stain buffer

266.4 µl DEPC-treated H<sub>2</sub>O

24 µl of 50 mg/ml acetylated BSA (final concentration of 10 µg/ml)

6 µl of 10 mg/ml normal goat IgG (final concentration of 0.1 mg/ml)

3.6 µl of 0.5 mg/ml biotinylated anti-streptavidin, goat antibody (final concentration of 3 µg/ml)

3 µl of 5% antifoam solution

**A2.11 Absolute Analysis Parameters**

The GeneChip software calculates the following analysis parameters automatically after the array has been scanned.

***A2.11.1 Probe Cell Average***

The digital image of each feature (or probe cell) is further divided into 64 pixels, and the array is scanned at a resolution of 3 µm per pixel. The bordering pixels of the probe cell are excluded and the remaining pixel intensity distribution is calculated. The intensity value associated with 75% of the distribution is used as the 'Average Intensity' of the probe cell.

### A2.11.2 Background Calculation

This is a measure of any fluorescence caused by the array surface and any non-specific binding of the target or SAPE stain molecules to the array surface. The array is divided into 16 sectors, and the software ranks each probe cell by its fluorescence intensity. The lowest 2% for each sector is identified and averaged to give the sector's 'background'. The sector's background is subtracted from the average intensities of all the probe cells within that sector.

### A2.11.3 Noise Calculation

Noise (Q) is a distinct phenomenon to the background, resulting from small variations in the digitised signal observed by the scanner as it scans the array surface. It is calculated using the standard deviations of the pixel intensities of the probe cells used to calculate the background. The calculation carried out is shown below.

$$Q = \frac{1}{N} \sum_{i \in \text{all bg cells}} \left( \frac{\text{stdev}_i}{\sqrt{\text{pixel}_i}} \right) * \text{SF} * \text{NF}$$

↓ Noise for a given probe array    ↓ Total # of background cells - lowest 2% for each sector    ↓ Standard deviation of the intensities of the pixels making up background cell i    ↓ Total # of pixels in cell i    ↓ Scaling factor    ↓ Normalisation factor

### A2.11.4 Positive & Negative Probe Pairs

In order to determine if a gene transcript is detectable in a sample, the data from the corresponding probe set is analysed to produce a set of 'Absolute Metrics'. The positive and negative probe pair calculation is the first of these. This calculation essentially determines the difference in intensity between the PM and MM probe cells. If the PM probe cell has a higher intensity than the MM probe cell, the probe pair is termed positive. If the MM probe cell has a higher intensity than the PM probe cell, the probe pair is termed negative. The significance is determined by calculating both the ratio (PM/MM) and the difference (PM-MM) for each probe pair. These values are then compared against two threshold values, the 'Statistical Difference Threshold' (SDT) and the 'Statistical Ratio Threshold' (SRT), as shown below.

$$\text{SRT} = 1.5$$

$$\text{SDT} = (Q) * (\text{SDT}_{\text{mult}})$$

A probe pair is **positive** if:

(1)  $\text{PM-MM} \geq \text{SDT}$

And

(2)  $\text{PM/MM} \geq \text{SRT}$

A probe pair is **negative** if:

(1)  $\text{MM-PM} \geq \text{SDT}$

And

(2)  $\text{MM/PM} \geq \text{SRT}$

\*  $\text{SDT}_{\text{mult}} = 2.0$  with standard wash method (section 2.2.10.3) and 4.0 with antibody amplification wash method (section 2.2.10.4)

### A2.11.5 Absolute Call Metrics

The numbers of positive and negative probe pairs, and the PM and MM intensities are used to derive three additional 'Absolute Call Metrics' for each probe set.

(1) Positive Fraction = # positive probe pairs / # probe pairs used

(2) Pos/Neg Ratio = # positive pairs / # negative pairs

(3) Log Average Ratio =  $10 * [\sum \log (PM / MM)] / (\text{Pairs in Avg})$

N.B. Pairs in Avg = probe pairs used (if probe pairs used  $\leq 8$ ). If probe pairs used  $>8$ , superscoring\* is performed.

A log avg ratio of zero indicates random cross hybridisation; the higher the log avg ratio, the more likely the transcript is present.

### A2.11.6 The Absolute Call

The three matrices determined above 'Positive Fraction', 'Pos/Neg Ratio' and 'Log Average Ratio' are entered into a decision matrix to determine whether a transcript is scored 'Present', 'Marginal' or 'Absent'. The default values are shown in the table below, and were used for the array experiments described in this thesis.

	Absent	Marginal	Present
	<u>min</u>		<u>max</u>
<b>Positive/Negative Ratio</b>	3.0		4.0
<b>Positive Fraction</b>	0.33		0.43
<b>Log Average Ratio</b>	0.9		1.3

If the calculated values were greater than the maximum thresholds, the transcript was scored present, if they were less than the minimum, the transcript was scored absent, and if they were in between, the transcript was scored marginal.

### A2.11.7 The Average Difference

The 'Average Difference' (Avg Diff) is an indicator of the relative level of expression of a transcript. It is the raw expression level value used in subsequent analyses.

$$\text{Avg Diff} = \sum (PM - MM) / (\text{Pairs in Avg})$$

## A2.12 Tris buffered saline (T.B.S.)

71.0 g Tris (BDH #27119), and 85.0 g sodium chloride (BDH #27800) were sequentially dissolved in 500 ml distilled water. The pH was adjusted to 7.6 with concentrated hydrochloric acid. 4.0 g magnesium chloride (BDH #26123), and 10 g bovine serum albumin was dissolved into the solution sequentially. The solution was made up to 10 L with distilled water, and stored at room temperature.

\* Superscoring filters probe pairs that are out of a given range when calculating 'average difference' and 'log average ratio'. The mean and standard deviation are calculated for intensity differences (PM - MM) across the entire probe set (excluding highest and lowest values). Values within 3 standard deviations are not included in the calculation.

**A2.13 Phosphate buffered saline (P.B.S.)**

2.0 g potassium dihydrogen orthophosphate (BDH #10203), 13.7 g disodium hydrogen orthophosphate and 80.0 g sodium chloride were sequentially dissolved in 600 ml distilled water. The solution was made up to 10 L with distilled water, and stored at room temperature.

**A2.14 Diaminobenzidine (DAB)**

5 mg of solid DAB (BDH #36237) was dissolved in 0.5 ml P.B.S. and frozen until needed. Approximately 30 min before use, one vial of frozen DAB for every 20 slides, was removed from the freezer. 70 µl 3% hydrogen peroxide was added to each thawed vial, and the DAB was made up to 10 ml with P.B.S.

**A2.15 Haematoxylin**

100 g aluminium potassium sulphate (BDH #27085) was dissolved in 2 L distilled water. 2 g haematoxylin (BDH #34037) was dissolved into 99% IMS, and added to the dissolved aluminium potassium sulphate, mixing thoroughly. 0.4 g sodium iodate (BDH #30172), citric acid (BDH #44445) and chloral hydrate (BDH #27668) were added in the order stated. The haematoxylin was filtered before being stored at room temperature.

**A2.16 1% Eosin**

20 mg eosin (BDH # 34026) was dissolved in 2 L tap water, and 2 ml formaldehyde (BDH 28421) added.

**A2.17 10 mM Citrate Buffer (pH 6.0)**

Dissolve 21 g citric acid crystals (BDH #277804L) in 500 mL distilled water. Adjust pH to 6.0 with 2 M sodium hydroxide. Make solution up to 10 L with distilled water.

### A3 Appendix for Chapter 3 (Data Mining)

#### A3.1 Comparative Analysis Parameters

The GeneChip Data Mining Tool calculates the following analysis matrices when two arrays are compared, to determine whether the expression level of each transcript has changed.

##### A3.1.1 Increase and Decrease Probe Pairs

This parameter compares two arrays at the probe pair level to determine changes in expression level between the two. If the difference between the PM and the MM probe cells in the experimental array (exp) is significantly greater than the difference in the comparative sample (comp), then the probe pair is considered **increased**. Significance is tested with two criteria:

- (1)  $(PM - MM)_{exp} - (PM - MM)_{comp} \geq \text{Change Threshold (CT)}$
- (2)  $[(PM - MM)_{exp} - (PM - MM)_{comp}] / (PM - MM)_{comp} \geq \text{Percent Change Threshold} / 100$

If the difference between the PM and MM probe cells on the experimental array is significantly less than the difference in the comparative sample, then the probe pair is considered **decreased**. Again, this significance is tested with two criteria:

- (1)  $(PM - MM)_{comp} - (PM - MM)_{exp} \geq \text{Change Threshold (CT)}$
- (2)  $[(PM - MM)_{comp} - (PM - MM)_{exp}] / (PM - MM)_{comp} \geq \text{Percent Change Threshold} / 100$

*Change Threshold (CT)* – calculated by the software using SDT of exp and comp data.

*Percent Change Threshold (PCT)* – the default value of 80 was used throughout this thesis.

##### A3.1.2 Difference Call Metrics

Four comparison metrics are calculated to determine any changes in the expression level of a transcript.

###### A3.1.2.1 Max (Increase / PP used, Decrease / PP used)

This metric determines the number of probe pairs that have changed in a certain direction:

$$\text{Increase / PP used} = \# \text{ increased probes} / \# \text{ probe pairs used}$$

$$\text{Decrease / PP used} = \# \text{ decreased probes} / \# \text{ probe pairs used}$$

The greater of these values is used in the decision matrix, and is reported in the *Max Inc & Dec* column.

###### A3.1.2.2 Ratio of Increased probe pairs / Decreased probe pairs

$$\text{Increase / Decrease Ratio} = \# \text{ Increased probe pairs} / \# \text{ Decreased probe pairs}$$

###### A3.1.2.3 Log Average Ratio Change

The log average ratios are recomputed for each probe set based on probe pairs used in both the baseline and the experimental probe arrays (see section A2.11.5 for formula).

$$\text{Log Average Ratio Change} = \text{Log Avg}_{exp} / \text{Log Avg}_{comp}$$

#### A3.1.2.4 Difference positive – Difference negative Ratio (Dpos – Dneg Ratio)

This metric combines the changes in the number of positive probe pairs and the change in the number of negative probe pairs between the comparative and the experimental arrays, into one metric for each probe set.

$$\text{Dpos – Dneg Ratio} = (\text{Positive Change}) - (\text{Negative Change}) / \# \text{ probe pairs used}$$

Where: Positive Change = # Positive Probe Pairs<sub>exp</sub> - # Positive Probe Pairs<sub>comp</sub>  
 Negative Change = # Negative Probe Pairs<sub>exp</sub> - # Negative Probe Pairs<sub>comp</sub>

The *Log Avg Ratio Change* and the *Dpos – Dneg Ratio* are typically positive when a transcript has a very low expression level in one sample and a relatively high expression level in the other. If a transcript has a very high expression level in one sample and a low or undetected expression level in the other, two metrics are typically negative. If a transcript is present in both samples, the metrics may be close to zero, despite a change in the expression level of the transcript.

#### A3.1.3 The Difference Call

The Difference Call is based on the metrics described in A3.1.2, which are entered into a decision matrix (see below) for every transcript. This results in the one of five outcomes for each transcript, Increase (I), Marginally Increase (MI), Decrease (D), Marginal Decrease (MD) and No Change (NC).

The default values are shown, and these were used to generate for this thesis. Each metric has a different weighting in its ability to influence the outcome of the matrix.

	Decrease	No Change	Increase
	<u>min</u>		<u>max</u>
<b>Inc / PP used</b>	0.33		0.43
<b>Inc / Dec Ratio</b>	3.0		4.0
<b>Log Avg Ratio Change</b>	0.9		1.3
<b>Dpos – Dneg Ratio</b>	0.2		0.3

#### A3.1.4 Fold Change Calculation

As the average difference of a transcript is directly related to its expression level, an estimate of the Fold Change of the transcript between the comp and exp samples can be calculated. The scaled average difference values are recomputed in both arrays to include only those probe pairs that used in both arrays. The Average Difference Change is determined by:

$$\text{Average Difference Change} = \text{Average Difference}_{\text{exp}} - \text{Average Difference}_{\text{comp}}$$

The Fold Change (FC) is calculated using:

$$\text{FC} = \left( \frac{\text{Avg Diff Change}}{\text{Max} [\text{min} (\text{Avg Diff}_{\text{exp}}, \text{Avg Diff}_{\text{comp}}), Q_M * Q_C]} \right) + \left\{ \begin{array}{l} +1 \text{ if Avg Diff}_{\text{exp}} \geq \text{Avg Diff}_{\text{comp}} \\ -1 \text{ if Avg Diff}_{\text{exp}} < \text{Avg Diff}_{\text{comp}} \end{array} \right\}$$

Where:  $Q_C = \max (Q_{\text{exp}}, Q_{\text{comp}})$

And the Q multiplier  $Q_M = 2.1$  in 50  $\mu\text{m}$  feature arrays; 2.8 in 24  $\mu\text{m}$  feature arrays\*

\* Test1 (50  $\mu\text{m}$ ), Test 2 (50  $\mu\text{m}$ ), Hu6800 set (50  $\mu\text{m}$ ), HuFL (24  $\mu\text{m}$ ), Hu35K set (24  $\mu\text{m}$ ).

Therefore, when the transcript expression has increased in the exp sample compared to the comp sample, the fold change is expressed as a positive number. When the transcript expression has decreased, the fold change is expressed as a negative number. If the noise ( $Q * QM$ ) is greater than the average difference of the transcript in either array, an approximate fold change is calculated using the noise. This is indicated by a tilde '~' in front of the fold change value.

#### A4 Appendix for Chapter 4 (Data Mining Results)

##### A4.1 25 kinase / phosphatase genes over expressed in CDi compared to UCi (including UC\_5)

<i>Gene Description</i>	<i>UCi_1</i>	<i>UCi_2</i>	<i>UCi_3</i>	<i>UCi_4</i>	<i>UCi_5</i>	<i>CDi_1</i>	<i>CDi_2</i>
phosphatase 2A B56-alpha	0	0	0	0	315.7	487.8	487.8
protein kinase C iota isoform	0	0	0	0	131.1	219.4	219.4
casein kinase II alpha' subunit	0	0	0	0	309.4	362.9	362.9
brk mRNA for tyrosine kinase.	0	0	0	0	274	573.8	573.8
kinase A anchor protein.	0	0	0	0	161.5	304.8	304.8
casein kinase I gamma 2	0	0	0	0	0	148.4	148.4
PtdIns 4-kinase (PI4Kb)	0	0	0	0	200.7	228.5	228.5
protein tyrosine phosphatase	275.9	0	0	0	0	585.3	585.3
protein kinase Dyrk2	0	0	0	0	0	157.2	157.2
tyrosine kinase-type receptor (HER2)	0	0	0	0	326.3	186.6	186.6
serine/threonine protein kinase sgk	0	0	0	0	0	379.7	379.7
protein serine/threonine kinase	0	0	0	0	0	209.9	209.9
diacylglycerol kinase (DAGK)	0	0	1081.8	0	0	905.3	905.3
testis-specific cAMP-dependent protein kinase catalytic subunit	0	0	0	0	116.9	244.4	244.4
A-kinase anchor protein (AKAP100)	0	0	0	84.8	0	0	0
inducible 6-phosphofructo-2-kinase/fructose 2,6-bisphosphatase	0	0	0	0	0	583.8	583.8
acylphosphatase, erythrocyte (CT) isoenzyme	0	0	0	0	39.8	74.3	74.3
mitogen activated protein kinase activated protein kinase	0	0	0	0	0	110.6	110.6
protein tyrosine phosphatase precursor	0	0	0	0	353.2	406	406
5'-AMP-activated protein kinase beta-1	0	0	0	190.6	0	225.9	225.9
AMP-activated protein kinase alpha-1	0	0	0	85.8	0	235.9	235.9
Rho-associated, coiled-coil containing protein kinase p160ROCK	0	0	0	272	0	301.6	301.6
putative phosphatidylinositol 3-kinase	0	0	0	0	0	203.7	203.7
inhibitor 2 of protein phosphatase 1	0	0	0	114.3	0	127.1	127.1
6-phosphofructo-2-kinase	0	0	0	0	0	57.8	57.8

*N.B. The figures are the average difference values for the gene in each sample.*

## A4.2 Genes specifically differentially expressed in CDI

Acc #	Incyte Description	Fold Change from NI	
		UCi	CDi
RC_AA056557_at	Incyte Unique	A in UCi	9.9
RC_AA291429_at	Human mRNA; cDNA DKFZp586M151 (from clone DKFZp586M151).	A in UCi	8.9
RC_AA236254_at	Incyte Unique	1.4	8.6
AA134488_at	Human mRNA; cDNA DKFZp566O134 (from clone DKFZp566O134).	1.4	8.2
L38503_at	Homo sapiens glutathione S-transferase theta 2 (GSTT2) mRNA, complete cds	A in UCi	7.3
RC_AA129268_s_at	Human mRNA for KIAA0650 protein, partial cds.	A in UCi	7.1
RC_AA460683_at	Human EXLM1 mRNA, complete cds.	1.2	6.8
W04732_at	Human 14q32 Jagged2 gene, complete cds; and unknown gene	A in UCi	6.6
RC_AA424515_at	Incyte Unique	A in UCi	6.5
U15655_at	Human ets domain protein ERF mRNA, complete cds	A in UCi	5.9
RC_AA446943_at	Incyte Unique	A in UCi	5.6
R39374_at	Human guanylate binding protein isoform I (GBP-2) mRNA, complete cds	A in UCi	5.5
RC_W58611_at	Human mRNA for 2-hydroxyphytanoyl-CoA lyase	-1.0	5.5
RC_AA446944_at	Human mRNA; cDNA DKFZp434N103 (from clone DKFZp434N103).	-1.5	5.4
RC_AA350771_s_at	Incyte Unique	-1.3	5.3
RC_AA434225_at	Human serum constituent protein (MSE55) mRNA, complete cds.	A in UCi	4.6
RC_AA425378_i_at	Human mRNA for KIAA1191 protein, partial cds.	-3.1	4.6
RC_D51169_s_at	Human replication protein A 32-kDa subunit mRNA, complete cds.	-1.2	4.2
L10378_at	Human (clone CTG-B43a) mRNA sequence	A in UCi	4.0
U04811_at	Human mRNA for KIAA1114 protein, complete cds.	1.0	3.4
RC_AA348001_at	Human NY-REN-24 antigen mRNA, partial cds.	-2.8	3.4
RC_AA461174_at	Human rec mRNA, complete cds.	-1.5	3.3
RC_AA477743_at	Human CpG island DNA genomic MseI fragment, clone 54b5, reverse read cpg54b5.rt1a.	-3.4	3.2
RC_AA253339_at	Incyte Unique	-2.6	3.1
RC_AA598636_at	Human mRNA; cDNA DKFZp564F2072 (from clone DKFZp564F2072).	-2.3	3.1
U01160_at	Human transmembrane 4 superfamily protein (SAS) mRNA, complete cds	-2.5	3.0
RC_AA450078_at	Human mRNA; cDNA DKFZp564J1616 (from clone DKFZp564J1616).	-1.9	3.0
RC_AA075048_at	Incyte Unique	-2.1	2.9
D29675_s_at	Human inducible nitric oxide synthase gene, promoter and exon 1 /gb=D29675 /ntype=DNA /annot=exon - Also Represents: X73029	-5.4	2.5
RC_AA063431_f_at	Incyte Unique	-3.1	2.4
RC_AA599986_at	Incyte Unique	-1.8	2.3
RC_AA488978_at	Human PRO0461 mRNA, complete cds.	-2.9	2.2
RC_AA232208_at	Human CGI-149 protein mRNA, complete cds.	-3.0	2.2
RC_AA287111_at	Human mRNA for HsGAK, complete cds.	-2.1	1.3
RC_AA056557_at	Incyte Unique	A in UCi	9.9
RC_AA291429_at	Human mRNA; cDNA DKFZp586M151 (from clone DKFZp586M151).	A in UCi	8.9

## A4.3 Genes specifically differentially expressed in UCI

<i>Acc #</i>	<i>Incyte Description</i>	<i>Fold Change from NI</i>	
		<i>UCi</i>	<i>CDi</i>
RC_H57166_at	Incyte Unique	-9.9	-1.2
T34752_s_at	Human dynactin p62 subunit mRNA, complete cds.	-9.6	1.3
M38690_at	Human CD9 antigen mRNA, complete cds	-6.1	-1.4
AA401894_at	Incyte Unique	-5.6	1.3
RC_AA452245_s_at	Human mRNA; cDNA DKFZp434M162 (from clone DKFZp434M162).	-5.1	-1.1
AA215970_at	Incyte Unique	-4.9	-1.1
U49785_at	Human D-dopachrome tautomerase mRNA, complete cds.	-3.0	1.5
H46831_at	Incyte Unique	-2.4	-1.0
RC_AA479362_at	Human mRNA; cDNA DKFZp586N0819 (from clone DKFZp586N0819).	-1.9	1.2
RC_N58463_at	Human mRNA PCTAIRE-1 for serine/threonine protein kinase.	-1.4	-1.0
RC_R20732_at	Human PCAF-associated factor 400 (PAF400) mRNA, complete cds.	1.6	1.3
RC_R20669_f_at	Human tra1 mRNA for Human homologue of murine tumor rejection antigen gp96.	1.8	-1.1
RC_AA132689_at	Human mRNA for KIAA0746 protein, partial cds.	1.8	1.1
RC_AA598702_at	Human transforming growth factor-beta (tgf-beta) mRNA, complete cds.	1.9	1.0
RC_AA054515_at	Incyte Unique	1.9	1.2
U79266_at	Human clone 23627 mRNA, complete cds.	1.9	-1.8
RC_AA257993_at	Human protein-tyrosine kinase (JAK1) mRNA, complete cds.	1.9	1.4
S82297_at	beta 2-microglobulin	2.0	-1.5
RC_AA045793_s_at	Human mRNA for microvascular endothelial differentiation gene 1 product, complete cds.	2.5	1.3
RC_N49002_at	Human ribonuclease k6 precursor gene, complete cds.	2.6	1.4
RC_T10088_at	Human genomic DNA, chromosome 22q11.2, clone KB1572G7.	2.6	-1.5
RC_AA122317_s_at	Human SEC14L mRNA, complete cds.	2.8	1.1
M63438_s_at	Human Ig rearranged gamma chain mRNA, V-J-C region and complete cds - Also Represents: X96754	2.9	-2.0
AA429539_f_at	Incyte Unique	3.1	1.3
RC_W95841_at	Human mRNA for KIAA0392 gene, partial cds.	3.1	-1.0
RC_R00037_at	Incyte Unique	3.2	1.1
RC_AA211388_at	Novel Human gene mapping to chromosome 1.	4.0	2.6
S71043_ma1_s_at	Ig alpha 2=immunoglobulin A heavy chain allotype 2 (constant region, germ line) [human, peripheral blood neutrophils, Genomic, 1799 nt] - Also Represents: S55735, J00220_cds5	4.2	-1.2
RC_AA489045_at	Human stannin mRNA, complete cds.	4.9	2.5
RC_AA425937_at	Incyte Unique	11.2	A in CDi
M87789_s_at	Human (hybridoma H210) anti-hepatitis A IgG variable region, constant region, complementarity-determining regions mRNA, complete cds - Also Represents: J00221_cds2, J00231	14.8	2.5
RC_N52440_at	Incyte Unique	21.0	7.3
RC_N69207_at	Incyte Unique	27.7	10.7
Y10032_at	H. sapiens mRNA for putative serine/threonine protein kinase	A in UCI	1.5

**A4.4 Genes with decreased expression in IBDi compared to NI samples**

<i>Acc #</i>	<i>Incyte Description</i>	<i>Fold Change from NI</i>	
		<i>IBDi</i>	<i>IBDu</i>
RC_R63388_at	Incyte Unique	-6.7	-1.1
RC_AA283046_at	Human mRNA for KIAA0448 protein, complete cds.	-4.6	-1.6
U14528_at	Human sulfate transporter (DTD) mRNA, complete cds.	-4.5	1.3
RC_AA449355_at	Human mRNA; cDNA DKFZp586O0223 (from clone DKFZp586O0223).	-4.1	1.1
RC_Z41366_at	Human mRNA for KIAA0872 protein, complete cds.	-3.9	-1.6
RC_AA282523_at	Human genomic DNA, chromosome 22q11.2, clone KB1269D1.	-3.5	1.2
RC_AA055932_at	Human mRNA for putative transcription factor, partial.	-3.3	-1.5
RC_T16315_s_at	Human mRNA for inositol 1,4,5-triphosphate 3-kinase.	-3.2	2.3
RC_T79586_at	Human genomic DNA of 8p21.3-p22 anti-oncogene of hepatocellular colorectal and non-small cell lung cancer , segment 8/11.	-2.4	1.6
RC_T64896_at	Incyte Unique	-2.3	1.2
RC_AA403305_at	Incyte Unique	-2.1	1.7
AA459536_at	Human lysyl hydroxylase isoform 3 (PLOD3) mRNA, complete cds.	-2.1	-1.3
RC_AA450302_s_at	Human Pex14 mRNA for peroxisomal membrane anchor protein, complete cds.	-1.3	-1.2
RC_R93029_at	Public Unique	-1.2	3.1

**A4.5 Genes with increased expression in IBDi compared to NI samples**

<i>Acc #</i>	<i>Incyte Description</i>	<i>Fold Change from NI</i>	
		<i>IBDi</i>	<i>IBDu</i>
RC_AA158396_s_at	Human mRNA for HLA-D class II antigen DO beta chain.	A in NI	A in NI
RC_AA191454_at	Human FGF-1 intracellular binding protein (FIBP) mRNA, complete cds.	A in NI	A in NI
RC_AA281006_at	Incyte Unique	A in NI	A in NI
RC_AA621096_at	Human okadaic acid-inducible and cAMP-regulated phosphoprotein 19 (ARPP-19) mRNA, complete cds.	A in NI	A in NI
RC_H80901_s_at	Human mRNA for Hakata antigen, complete cds.	A in NI	A in NI
RC_N24732_at	Human nuclear respiratory factor-2 subunit alpha mRNA, complete cds.	A in NI	A in NI
RC_N33920_at	Human mRNA for diubiquitin.	A in NI	A in NI
RC_N53534_s_at	Human beta-arrestin 2 mRNA, complete cds.	A in NI	A in NI
M16336_s_at	Human T-cell surface antigen CD2 (T11) mRNA, complete cds, clone PB1.	A in NI	A in NI
Y08374_ma1_at	Human glycoprotein mRNA, complete cds.	A in NI	A in NI
U66464_at	Human hematopoietic progenitor kinase (HPK1) mRNA, complete cds.	38.6	8.2
RC_R43162_s_at	Incyte Unique	28.8	A in IBDu
U05861_at	Human dihydrodiol dehydrogenase mRNA, complete cds.	14.7	3.9
RC_N52440_at	Incyte Unique	13.8	3.1
RC_N74958_at	Incyte Unique	11.4	2.4
RC_AA218544_at	Incyte Unique	8.8	1.3
D83784_at	Human mRNA for KIAA0198 gene, partial cds.	7.4	3.1
RC_AA478112_at	Human genomic DNA, chromosome 22q11.2, clone KB1125A3.	7.4	-1.7
X72755_at	Human Humig mRNA.	6.8	-1.1
AF005043_at	Human poly(ADP-ribose) glycohydrolase (hPARG) mRNA, complete cds.	6.8	2.8
RC_AA115933_r_at	Human mRNA for KIAA1098 protein, partial cds.	6.5	1.3
N76780_at	Incyte Unique	5.9	2.5
RC_AA459123_at	Human B lymphocyte adapter protein BAM32 (BAM32) mRNA, complete cds.	5.9	1.2
X89416_at	Human mRNA for protein phosphatase 5.	5.1	3.1
RC_AA053401_at	Incyte Unique	4.4	2.0
RC_AA424346_at	Incyte Unique	3.1	-1.5
RC_W69211_s_at	Human eotaxin precursor mRNA, complete cds.	2.9	1.9
U06863_at	Human follistatin-related protein precursor mRNA, complete cds.	2.4	-3.8
RC_W46575_at	Incyte Unique	2.1	-5.0

**A4.6 Genes with decreased expression in IBDu compared to NI samples**

<i>Acc #</i>	<i>Incyte Description</i>	<i>Fold Change from NI</i>	
		<i>IBDi</i>	<i>IBDu</i>
AA007282_at	Incyte Unique	-1.6	A in IBDu
AA082171_at	Human clone 24921 mRNA sequence.	-1.3	A in IBDu
D54358_at	Human sperm membrane protein BS-63 mRNA, complete cds.	-2.3	A in IBDu
RC_AA243654_at	Incyte Unique	2.2	-67.9
RC_AA441792_at	Human CHORD containing protein-1 (CHP1) mRNA, complete cds.	-1.2	-19.8
RC_AA521097_at	Incyte Unique	-1.5	-19.4
RC_H59722_at	Human X-linked anhidrotic ectodermal dysplasia protein gene (EDA), exon 2 and flanking repeat regions.	1.1	-17.7
RC_H65650_at	Incyte Unique	6.0	-8.1
RC_H97521_at	Incyte Unique	1.1	-6.9
RC_N20027_at	Incyte Unique	-1.5	-6.0
RC_R08929_at	Human ubiquitin conjugating enzyme G2 (UBE2G2) mRNA, complete cds.	1.2	-5.7
RC_N46429_at	Human full length insert cDNA clone YY75G10.	1.0	-5.5
U69141_at	Human glutaryl-CoA dehydrogenase mRNA, complete cds.	1.1	-5.1
RC_AA598468_at	Human mRNA for hSNF2H, complete cds.	1.1	-4.3
X77548_at	Human cDNA for RFG.	1.3	-2.4
RC_N33144_at	Incyte Unique	2.5	-2.2

**A4.7 Genes with increased expression in IBDu compared to NI samples**

<i>Acc #</i>	<i>Incyte Description</i>	<i>Fold Change from NI</i>	
		<i>IBDi</i>	<i>IBDu</i>
RC_AA521097_at	Incyte Unique	A in NI	A in NI
RC_AA243654_at	Incyte Unique	1.8	9.6
D54358_at	Human sperm membrane protein BS-63 mRNA, complete cds.	1.8	8.7
RC_H65650_at	Incyte Unique	2.6	6.2
RC_AA441792_at	Human CHORD containing protein-1 (CHP1) mRNA, complete cds.	1.8	5.8
AA007282_at	Incyte Unique	-1.7	4.7
RC_H59722_at	Human X-linked anhidrotic ectodermal dysplasia protein gene (EDA), exon 2 and flanking repeat regions.	1.2	4.3
AA082171_at	Human clone 24921 mRNA sequence.	-1.1	3.8
RC_H97521_at	Incyte Unique	-3.0	1.9

## A4.8 Genes in cluster 7 of figure 4.59

Acc No	Gene Description	Mean Average Difference value				
		UCi	CDi	UCu	CDu	NI
RC_AA258352_at	Human Ig lambda gene locus DNA, clone:288A10.	516.5	842.8	1099.6	1213.5	1109.7
L25119_at	Human opioid receptor mRNA, complete cds.	0	15.9	39.4	24.9	21.7
RC_AA496986_at	Human mRNA; cDNA DKFZp434L108 (from clone DKFZp434L108).	83.7	147	181.7	175.2	186.1
RC_W73774_at	Incyte Unique	0	20.1	30.5	10.6	26.9
RC_AA449355_at	Human mRNA; cDNA DKFZp586O0223 (from clone DKFZp586O0223).	222.7	85.1	894.5	548	634.4
U04840_at	Human onconeural ventral antigen-1 (Nova-1) mRNA, complete cds.	11.6	43.5	46	33.9	58.5
RC_W70079_at	Human full length insert cDNA clone ZD42A08.	72.8	52.9	152.1	205.7	179.6
RC_AA446609_at	Human mRNA for transcription elongation factor TFIIS.h.	501.4	1051.2	1072.2	821.9	1211.5
J04469_at	Human mitochondrial creatine kinase (CKMT) gene, complete cds.	1275	1832	2587	2240.4	2142.4
U00943_at	Human glioma pathogenesis-related protein (GliPR) mRNA, complete cds.	0	21.5	61.6	34.9	75.9
RC_AA491295_at	Human mRNA for KIAA0787 protein, partial cds.	486.2	459.6	995.9	672.7	743
R78444_at	Incyte Unique	0	5.3	24.8	27.3	45.8
L02785_at	Human colon mucosa-associated (DRA) mRNA, complete cds.	2051.6	1712.7	7436.6	4807.7	4248.9
J03910_ma1_at	Human (clone 14VS) metallothionein-IG (MT1G) gene, complete cds.	411	3343.6	5465	3695.5	3882.7
L05188_f_at	Human SPRR2A gene encoding small proline rich protein.	0	19.1	55.1	32.3	60.1
RC_W72079_at	Human full length insert cDNA clone ZD69D05.	644	886.5	1526.8	748.1	1176.7
RC_W72382_at	Human oxidative 3 alpha hydroxysteroid dehydrogenase mRNA, complete cds.	91.7	172	248	131.1	171.9
U00951_at	Human M-phase phosphoprotein homolog mRNA, complete cds.	0	0	0	0	0
RC_AA496366_at	Incyte Unique	913	745	2860.4	665.9	2409.8
AA481201_at	Human chromosome 17 HSPC009 mRNA, complete cds.	276	886.6	834	708.2	981.4
R67763_at	Human mRNA for myosin phosphatase target subunit 1 (MYPT1).	0	7.5	24.2	27.4	29.2
L08096_s_at	Human CD27 ligand mRNA, complete cds.	0	0	0	0	0
J05243_at	Human alpha II spectrin mRNA, complete cds.	314.3	606.3	706.6	278.7	590.8
RC_AA449475_at	Human mRNA for brk kinase substrate (BKS gene).	1116.7	976.7	1660.5	1721	2054.5
R66772_at	Human mRNA; cDNA DKFZp434M082 (from clone DKFZp434M082).	0	0	7.4	10	13.1
H44269_at	Incyte Unique	332.4	451.9	540.9	425	566.4
RC_W81079_at	Human beta1-syntrophin (SNT B1) gene, complete cds.	0	0	25.3	20.9	35.8
AA320369_s_at	Human RGS-GAIP interacting protein GIPC mRNA, complete cds.	2753.2	1645.3	5138.6	2160.4	5182.3
RC_Z41366_at	Human mRNA for KIAA0872 protein, complete cds.	0	238.1	287.3	286.3	461
RC_AA478487_at	fibroin-3	0	0	78.9	0	93.8
M76378_at	Human cysteine-rich peptide mRNA, complete cds.	1002	1883.3	3953.2	1668.6	2569.2
RC_AA459905_at	Incyte Unique	132.1	232.4	357.3	326.7	305.7
D82284_at	Human mRNA for KIAA0733 protein, partial cds.	16.7	148.8	328.9	464.3	681.1
X69908_ma1_at	Human gene for mitochondrial ATP synthase c subunit (P2 form).	1821.1	2009.3	2946.1	1476.3	2865.2
AA410480_at	Incyte Unique	441.1	699.4	923.6	528.1	789.8
RC_AA459690_s_at	Human HSPC040 protein mRNA, complete cds.	771.6	927.1	1233.7	1463.8	1372.3
RC_AA251906_at	Incyte Unique	47.5	28.4	273.4	153.9	230
W96311_at	Human mRNA; cDNA DKFZp434N024 (from clone DKFZp434N024); complete cds.	0	24.9	52.6	56.9	70.8
RC_AA477330_at	Incyte Unique	0	87.9	99.7	87.7	93.9
W92836_at	Incyte Unique	101	588.6	508.4	469.5	649.1
RC_AA253471_at	Incyte Unique	208.1	512	751.2	458.2	672.7
RC_AA436841_at	Incyte Unique	0	0	96.4	56	69.3
RC_Z40805_at	Human c-ERBB-2 gene, exons 1', 2', 3', 4'.	411.9	441.1	1457.3	1194.4	841.5
M55513_s_at	Human potassium channel (HPCN1) mRNA, complete cds.	0	411.2	523.3	381.8	476.3
W69964_at	Human mRNA, chromosome 1 specific transcript KIAA0495.	139.1	60.6	302	350.9	315.6
Z83821_cds1_at	Human APE2 protein (APE2) mRNA, complete cds.	1717.7	2024.3	3389.2	2772.5	3223.7
AA448409_at	Human mRNA; cDNA DKFZp586G011 (from clone DKFZp586G011); partial cds.	278.5	388.6	513.4	367.7	437.2
M12125_at	Human fibroblast muscle-type tropomyosin mRNA, complete cds.	865.5	979.7	3616.9	966.8	3311.9
RC_AA437258_at	Incyte Unique	549.6	531.4	846.1	585.2	1031.7
RC_AA437235_s_at	Human full length insert cDNA clone ZD69D05.	236.1	589.2	667.4	472.5	545.8
M13452_s_at	Human lamin A mRNA, 3'end.	972.5	967.5	1546.5	1110.8	1451.9
RC_Z40379_at	Public Unique	0	46.5	63.5	87.4	75.1

X86032_at	Human mRNA for thioesterase II.	70.9	401.2	341.8	437.5	548.6
RC_Z40489_at	Human mRNA for KIAA0957 protein, complete cds.	0	0	25.4	29.7	56
X99802_at	Human mRNA for ZYG homologue.	107	378.1	451.6	282.2	414
RC_AA262943_at	Human ribosomal protein s4 Y isoform gene, complete cds.	717.5	881	1298.4	891.8	1421.4
RC_Z40573_at	Human mRNA for KIAA0672 protein, complete cds.	329.5	763.8	845.4	988.2	953
M16364_s_at	Human creatine kinase B mRNA, complete cds.	308.9	393.2	3797	798.9	5009.6
W69967_at	Incyte Unique	116.1	0	469.5	0	540.8
L40904_at	Rhesus monkey peroxisome proliferator-activated receptor gamma 1-a (PPARgamma) mRNA, complete cds.	572.6	720.5	919.4	796.4	919.3
AF007111_at	Human MDM2-like p53-binding protein (MDMX) mRNA, complete cds.	0	0	0	0	0
U83908_at	Human nuclear antigen H731 mRNA, complete cds.	543.8	634.3	832.9	1258.6	1183.2
W27237_at	Human Cdc42 effector protein 4 mRNA, complete cds.	159	181.9	284.4	115.1	247.6
W26996_at	Human mRNA; cDNA DKFZp434B2411 (from clone DKFZp434B2411).	0	303.1	1250.5	1126.9	1228.1
M38690_at	Human mRNA for MRP-1.	704.6	1806.3	2519.8	2569	2472.1
U70732_ma1_at	Human mRNA for alanine aminotransferase.	833.6	856.6	2174.7	1939.8	1391.2
RC_AA252598_at	Incyte Unique	79.9	113.1	130.1	211.7	225
RC_AA435738_at	Incyte Unique	199.6	427.4	566	413.1	407.9
W27330_at	Incyte Unique	594.1	542.9	1129.3	450	1242.7
W20094_at	Human mRNA; cDNA DKFZp586A0522 (from clone DKFZp586A0522); partial cds.	850	800.2	1563.8	1359.8	1316.3
AA427468_s_at	Human hCPE-R mRNA for CPE-receptor, complete cds.	8159.9	5224.1	11807.9	7884.5	11990.1
W01587_s_at	Human full length insert cDNA clone YN88C07.	81.2	116.7	165.5	135.6	147.3
AA426304_s_at	Human mRNA; cDNA DKFZp434P1721 (from clone DKFZp434P1721); partial cds.	959.2	728.5	1585	1080.7	1341.7
H43790_at	Human mRNA for plakophilin 2a and b.	0	42.3	73	35.3	47.7
H19580_at	Human clone 25076 mRNA sequence.	0	28.8	21.8	14.3	37
RC_AA465692_at	Human mRNA for KIAA0648 protein, partial cds.	66.3	100.6	133.5	107.3	102.2
M29273_at	Human myelin-associated glycoprotein (MAG) mRNA, complete cds.	0	37.8	69.5	0	91.3
RC_AA436624_s_at	Human mRNA for hB-FABP.	0	25.9	33.6	33.3	39.1
W44681_at	Human JAW1-related protein MRV1A long isoform (MRV1) mRNA, complete cds.	735.6	919.5	1286.2	782.6	1109.6
D88153_at	Human mRNA for HYA22, complete cds.	120	400.2	375.3	360.2	400
RC_AA252147_at	Incyte Unique	14.8	21.6	31.2	24.9	41.8
RC_AA476448_at	Human ribosomal S6 protein kinase mRNA, complete cds.	0	0	156.8	127.2	138.8
W48808_s_at	Incyte Unique	284.7	420.6	566.7	264.4	521.5
L25880_s_at	Human mRNA for microsomal epoxide hydrolase (EC 3.3.2.3).	237.3	364	409.2	244.1	460.2
T83154_at	Human origin recognition complex subunit 6 (ORC6) mRNA, complete cds.	0	0	22.8	0	27.9
U12255_at	Human IgG Fc receptor hFcRn mRNA, complete cds.	5508.5	5712.4	10242.3	8234.7	10450.9
M24485_s_at	Human (clone pHGST-pi) glutathione S-transferase pi (GSTP1) gene, complete cds.	2382.2	2761.8	3977.6	2798.6	3555.1
H89896_s_at	Human full length insert cDNA clone ZD76G03.	0	26	50.8	59.2	110
AA419186_at	Human NF2 gene.	0	423.1	464.8	456	478.4
W38597_s_at	Human peroxisomal membrane protein 20 mRNA, complete cds.	3122.2	4948.2	6506.6	4352.8	4835.1
U00943_at	Human glioma pathogenesis-related protein (GliPR) mRNA, complete cds.	0	21.5	61.6	34.9	75.9
RC_AA255903_at	Human CD39L4 (CD39L4) mRNA, complete cds.	238.2	980.5	2075.5	1975.6	1613.3
RC_AA456955_at	similar to ankyrin of Chromatium vinosum.	0	52.3	50.6	56.7	91.8
AF000237_s_at	Human lysophosphatidic acid acyltransferase mRNA, complete cds.	2432.2	1718.8	4737.3	2540.4	3391
RC_AA243285_at	Incyte Unique	0	0	27	0	29.4
RC_AA452036_at	Human mRNA for KIAA0879 protein, complete cds.	0	244	239.5	157	352.3
N99617_at	Human mRNA; cDNA DKFZp434E1515 (from clone DKFZp434E1515).	153.8	150.4	265.5	202	235.5
N94832_at	Human m6b1 mRNA, complete cds.	0	9.7	12.3	0	16.2
AA352656_at	Human lysyl hydroxylase 1 gene, promoter region and partial cds.	469.5	463.1	848.7	756.1	850.3
N89302_s_at	Human mRNA for KIAA1248 protein, partial cds.	930.2	1240	1846.2	1090.1	2206.5
L40904_at	Rhesus monkey peroxisome proliferator-activated receptor gamma 1-a (PPARgamma) mRNA, complete cds.	572.6	720.5	919.4	796.4	919.3
RC_W92449_at	Human AP-mu chain family member mu1B (HSMU1B) mRNA, complete cds.	1349.2	1830.4	2387.5	1907.6	2193.6
D78014_at	Human mRNA for dihydropyrimidinase related protein-3, complete cds.	119.1	332.2	375.5	168.7	304.4
RC_W94281_s_at	Incyte Unique	11451.6	8805.2	17401.8	12447.2	17709
RC_W94672_at	Incyte Unique	44	82.5	151.3	122.5	162.4
M95787_at	Human 22kDa smooth muscle protein (SM22) mRNA, complete cds.	2601.1	3603.8	8545.9	3108.1	5552.1

C02548_s_at	Human full length insert cDNA clone ZA70C11.	0	35.7	25.6	27.7	44.7
RC_AA451896_at	Human HSPC154 mRNA, complete cds.	0	150.2	164.9	196	377.6
C00808_s_at	Incyte Unique	489.8	779.4	1718.3	1718.1	1258.5
RC_AA489009_at	Incyte Unique	25	33	62.5	48.2	89.8
L42611_f_at	Human keratin 6 isoform K6e (KRT6E) mRNA, complete cds.	0	356.6	660.9	359	337.5
RC_AA489383_at	Human bone morphogenetic protein 2A (BMP-2A) mRNA.	43.6	145.5	179.6	152.5	129.4
RC_AA237017_at	Human mRNA for KIAA1068 protein, partial cds	347.2	189.7	488.3	384.9	448.4
RC_W86185_at	Incyte Unique	77.7	58.5	208.6	84.8	240.3
RC_W86375_s_at	Human CGI-10 protein mRNA, complete cds.	747.3	1806.9	2371.5	3132.2	2998
R27140_at	Human full length insert cDNA clone YI72E07.	0	95.4	108.8	184.4	195.1
S69265_s_at	Human mRNA for PLE21 protein, complete cds.	0	20.9	139.1	135.8	227.9
RC_AA451836_at	Incyte Unique	30.5	67.8	119	68	68.1
RC_W87533_at	Incyte Unique	3225	4805.9	6739.3	5475.1	5024.4
RC_W88425_at	Human APCL gene, exon 9.	900.2	607	2475.9	1619.4	1628.9
L38820_at	Human (lambda-gt11ht-5) MHC class I antigen-like glycoprotein (CD1D) mRNA, complete cds	7.6	63.8	73.4	31.1	60.3
R14606_at	Incyte Unique	0	43.2	78.9	20.9	53.4
L42601_f_at	Human keratin 6 isoform K6e (KRT6E) mRNA, complete cds	242.3	237.1	580.1	393	533.6
RC_AA454566_at	Incyte Unique	434.2	728.8	1061	597.8	782.5
RC_AA479900_at	Incyte Unique	194.4	131.2	326.1	171.1	268.3
RC_Z39652_at	Human oligodendrocyte myelin glycoprotein (OMG) exons 1-2; neurofibromatosis 1 (NF1) exons 28-49; ecotropic viral integration site 2B (EVI2B) exons 1-2; ecotropic viral integration site 2A (EVI2A) exons 1-2; adenylate kinase (AK3) exons 1-2.	139.7	756.4	1753	1121.9	1075.9
RC_Z39283_at	Incyte Unique	92.3	0	221	62.6	195.9
RC_Z39338_at	Human mRNA, cDNA DKFZp566B0846 (from clone DKFZp566B0846); partial cds.	477.6	683.4	768.3	800.8	870.3
D14662_at	Human mRNA for KIAA0106 gene, complete cds.	510.1	1436	1680.6	1010	1095.6
RC_AA481437_at	Incyte Unique	1022.1	1340	1762	1327	1694.3
RC_Z39569_at	Human mRNA, cDNA DKFZp434A132 (from clone DKFZp434A132).	0	229.5	420.4	203	557.8
M10050_at	Human liver fatty acid binding protein (FABP) mRNA, complete cds.	3319.4	7420.2	15942	11924.5	8928.1
RC_AA481883_at	Incyte Unique	0	160.7	113.7	51	174.1
D54949_at	Human calmodulin-1 (CALM1) mRNA, 3'UTR, partial sequence.	1219.2	2072.5	3201.1	2263.2	2245
RC_Z39818_at	Incyte Unique	166.7	173.6	307.4	192.8	405.7
D57341_at	Incyte Unique	142.7	384.2	374.8	341.7	390.8
RC_AA456821_at	Incyte Unique	41.6	74.5	120.3	137.3	138.2
RC_Z39929_at	Incyte Unique	163.6	299.2	568	421.9	403.5
D17408_s_at	Human mRNA for calponin, complete cds.	338.7	506.6	2168.3	728.3	2235.9
N35781_at	Human lambda-crystallin mRNA, complete cds.	963.5	1088.2	1392.1	1418.4	1672.1
RC_AA431468_s_at	Public Unique	1972.8	2536.3	3636.8	2909.6	3377
RC_AA454599_s_at	Incyte Unique	702.9	1566.7	2142.1	1931.4	2094.1
RC_AA485140_at	Human HepG2 3' region Mbol cDNA, clone hmd6a08m3.	10.2	7.8	32	36.1	67.7
RC_AA486868_s_at	Human SLIT2 (SLIL2) mRNA, complete cds.	158.1	282.8	530.2	313.1	339.7
RC_AA485443_at	Human hypothetical SBBI03 protein mRNA, complete cds.	72.1	175.2	232.6	150.4	219
RC_AA485308_at	Human mRNA, cDNA DKFZp434C2016 (from clone DKFZp434C2016).	70.6	322.2	374.5	231.3	234.2
D31840_s_at	Human atrophin-1 mRNA, complete cds	28.3	70.1	115.7	130.1	184.8
RC_AA256365_s_at	Human mRNA expressed in thyroid gland.	6441.8	5519.1	14066.5	7532	14499.1
RC_AA485060_at	Human transcription factor (MEF2) mRNA, complete cds.	459.9	826.2	942.7	974.1	1367.5
RC_AA256273_at	Human mRNA, cDNA DKFZp434P174 (from clone DKFZp434P174).	766.2	1673.7	2277.1	1749.1	1667.1
RC_AA485039_at	Incyte Unique	144.5	426.3	648.2	529.8	679.7
D30949_at	Human TRPM-2 mRNA, complete cds.	0	0	190.6	270.6	365.4
M91083_at	Human DNA-binding protein (HRC1) mRNA, complete cds.	270	610	1222.6	787.1	727.1
AA389673_at	Human CGI-135 protein mRNA, complete cds.	188.4	207.3	384.8	159.7	334.1
RC_AA279418_at	Incyte Unique	0	233.3	190.2	238.6	265.1
RC_AA497018_at	Human full length insert cDNA clone ZD51B03.	0	4	104.6	108.5	103.4
RC_N26184_at	Human MYLE mRNA, complete cds.	340.2	494.8	592	224	662.8
RC_AA055697_f_at	Incyte Unique	0	133	286.8	368.5	365.7
RC_R25375_at	Human expressed pseudo TCTA mRNA at t(1;3) translocation site, complete cds.	0	0	187.3	0	210.7

AA307896_at	Human nuclear localization signal containing protein deleted in Velo-Cardio-Facial syndrome (Nlvcf) mRNA, complete cds.	0	0	37.7	23.7	26.7
RC_AA115533_at	Incyte Unique	2519.8	2899.7	4806.6	4048.3	3647.1
RC_AA055932_at	Human mRNA for putative transcription factor, partial.	95.5	96.9	295.6	156.2	320.6
RC_AA055811_s_at	Human A33 antigen precursor mRNA, complete cds.	3450.3	6557.8	7247	8207.1	8838.4
RC_F13789_at	Human mRNA; cDNA DKFZp586D2223 (from clone DKFZp586D2223); partial cds.	0	29	51.9	45.1	38.7
RC_R21762_at	Incyte Unique	479.5	0	764.5	1136.7	825.6
RC_N57896_at	Incyte Unique	25.1	17.2	44.3	48.2	77.7
RC_AA180163_s_at	Human mRNA export protein (RAE1) mRNA, complete cds.	0	180.2	182.8	98.7	254.3
RC_AA406197_at	Incyte Unique	0	41.9	45.7	0	60
RC_R19183_at	Incyte Unique	25.7	58.6	128.4	87	77.6
AA149507_at	Human mRNA; cDNA DKFZp586I0923 (from clone DKFZp586I0923).	0	25.7	41.2	15.8	20.7
RC_AA406484_at	Incyte Unique	95	34.6	204.9	127.4	171.8
RC_R33498_s_at	Human gene for hippocalcin, exon 2, 3 and complete cds.	6863.4	9536.5	13677.6	12375.9	11523.5
X16135_at	Human mRNA for novel heterogeneous nuclear RNP protein, L protein.	750.5	821.9	1412.9	961.2	1268.9
RC_AA149652_at	Incyte Unique	0	0	30.3	0	36.2
RC_R28370_at	Human carboxypeptidase M precursor, mRNA, partial cds.	203.2	168.4	455.7	476.5	338.6
RC_H86351_at	Incyte Unique	0	39.3	80	0	102.8
RC_AA181600_at	Incyte Unique	39.3	33.9	134.4	102.9	180
AA133029_at	Human TACC2 protein (TACC2) mRNA, partial cds.	0	164.7	258.6	287.8	242
RC_N57905_at	Incyte Unique	43.3	4	67.2	46.8	58.2
RC_R17067_s_at	Human mRNA for KIAA0038 gene, partial cds.	2135.5	2546.5	3782.1	3040.5	3226
RC_AA187144_s_at	Human endothelin-1 (EDN1) gene, complete cds.	128.6	194.6	291.5	125.6	240.5
RC_AA063280_at	Incyte Unique	106.2	258	320.6	272	514
RC_H79148_at	Incyte Unique	243.2	429.9	708.7	507.7	467.8
RC_H08371_at	Human cDNA similar to C. elegans RNA binding protein UI4946, Q10572, complete cds.	117.7	360.8	458.4	164	426.8
RC_R01712_at	Public Unique	99.9	0	156.5	96.6	137.9
RC_AA063394_at	Human WAVE3 mRNA for WASP-family protein, complete cds.	47.6	76.1	116.7	78.4	84
RC_N58193_at	Incyte Unique	164.5	283.2	318.8	203.1	273.2
RC_AA069425_at	Incyte Unique	168.4	216.7	401.8	233.9	299.2
RC_AA405379_at	Human mRNA full length insert cDNA clone EUROIMAGE 42138.	669.5	562.5	1006.9	644.5	1031.2
RC_H06233_at	Incyte Unique	40.6	135.4	180.4	85.7	172.8
RC_H05890_at	Incyte Unique	0	219.4	246.8	192	284.2
RC_H04649_at	Incyte Unique	40.7	290.2	293	122.8	312.9
RC_R10138_at	Incyte Unique	0	45.8	278.8	323.9	256.8
RC_R10720_at	Public Unique	25.8	17.6	120.4	59.7	150.2
RC_R10657_s_at	Human clone 24951 mRNA sequence.	114	251.9	281.8	303.1	367.6
RC_N58065_at	Public Unique	65.4	61.3	152.6	137.7	102.1
X54162_at	Human mRNA for a 64 Kd autoantigen expressed in thyroid and extra-ocular muscle.	67.9	103.6	433.3	0	451.2
RC_H05599_at	Incyte Unique	67.2	61.1	167.4	74.6	118.7
RC_AA058911_at	Human mRNA; cDNA DKFZp586O1624 (from clone DKFZp586O1624); partial cds.	403.4	528.1	715	390.7	723.7
RC_AA173573_at	Incyte Unique	38.1	54.5	92.5	35.5	96.4
RC_AA053962_at	Incyte Unique	30.6	84.7	117.7	5	130
AA149783_at	Human SYBL1 gene, exons 6-8.	0	83	130	69.7	62.4
RC_AA283935_at	Human antigen NY-CO-43 mRNA, complete cds.	123	355.5	329	256	308.8
AA099391_s_at	Incyte Unique	396.6	1238	2210.6	1187.6	2373.6
RC_R42698_at	Incyte Unique	96.1	230.3	256.5	133.1	194.7
RC_AA411114_at	Human lysyl hydroxylase 1 gene, promoter region and partial cds.	1235.2	1446.9	1799.7	2158.8	2353.5
RC_R43317_at	Human glioma amplified on chromosome 1 protein (GAC1) mRNA, complete cds.	29.9	235.6	861.7	526.6	496.3
RC_AA193593_at	Incyte Unique	0	111.5	169.2	0	140.1
RC_R42365_at	Human mRNA for KIAA0972 protein, complete cds.	405.3	1042	1511.2	682.3	935.7
AA171913_at	Human carbonic anhydrase precursor (CA 12) mRNA, complete cds.	1181.2	1349.5	2548.2	2584.9	3477.3
RC_F09118_at	Human insulinoma rig-analog mRNA encoding DNA-binding protein, complete cds.	119.7	139.9	318.1	293.9	315
RC_H89132_s_at	Human P2Y1 gene.	0	16.5	17.6	16.9	17.2
RC_AA149007_at	Human transcription factor ESE-3A (ESE-3) mRNA, complete cds.	116.3	266.9	303.3	250.9	305.5

RC_AA148983_at	Human full length insert cDNA YO73E04.	182.3	587.1	722.9	704.2	593.1
RC_F09255_at	Incyte Unique	0	167.1	254.9	115.7	395.5
RC_F04052_at	Human clone 23763 unknown mRNA, partial cds.	338.7	615	739	607.9	623.9
RC_R48589_at	Incyte Unique	442.5	777.7	876.2	755.3	1240.6
RC_N51868_at	Human ribosomal protein S20 (RPS20) mRNA, complete cds.	30.5	68	146.4	73.6	186.7
RC_R45562_at	Incyte Unique	81.7	0	144.9	132.8	229.7
RC_AA155749_at	Human mRNA expressed in thyroid gland.	3296.5	3720.6	8375.7	4995.5	6456.9
RC_AA194830_at	Incyte Unique	87.6	298.5	330.9	206.3	311.5
RC_N52151_at	Incyte Unique	87	111.9	397	218.9	591.2
RC_F04446_at	Incyte Unique	209.3	310.3	371.5	328.9	472.2
RC_AA193384_at	Incyte Unique	0	25.2	63.1	63.8	55.6
RC_AA047291_at	Human Tigger1 transposable element, complete consensus sequence.	21	177.6	213.9	48	131.1
AA057555_at	Human T-cell receptor alpha delta locus from bases 250472 to 501670 (section 2 of 5) of the Complete Nucleotide Sequence.	0	91.1	66.7	70.1	107.1
RC_R36939_f_at	Human mRNA; cDNA DKFZp434H1215 (from clone DKFZp434H1215); partial cds.	72	127.6	237.2	154	157.3
RC_AA188785_at	Incyte Unique	0	136.4	187.5	0	217
RC_AA053022_i_at	Human mRNA for serine/threonine protein kinase EMK.	0	23.9	29	21.8	36.2
RC_AA149635_at	Incyte Unique	57.6	23.6	111.1	39.7	93.2
X12876_s_at	Human mRNA for XPAC protein.	3400.1	3800	6584	4966.2	5938.7
RC_F10199_f_at	Human mRNA; cDNA DKFZp434N1427 (from clone DKFZp434N1427); partial cds.	463.7	410.5	806.9	1100.6	1192.2
RC_N55274_at	Human gene for transforming growth factor-beta 3 (TGF-beta 3) exon 6.	0	227.4	189.3	0	242.7
RC_AA053400_at	Incyte Unique	0	0	86	34.3	88
X93036_at	Human mRNA for MAT8 protein.	3167.2	2637.8	12384.9	7896.3	8717
RC_R40093_at	Human mRNA for KIAA0969 protein, complete cds.	559	786.8	1138.3	647.8	1001.2
RC_F09902_s_at	Incyte Unique	219.2	212.5	305.4	379.4	408.7
RC_N53560_at	Incyte Unique	179.2	166.9	308.1	170.3	417
RC_AA121166_s_at	Human zinc finger protein (ZNF139) mRNA, partial cds.	161.4	657.1	682.7	536.7	780.8
RC_R41811_at	Incyte Unique	37.1	97.8	378.3	287.7	262.7
X96969_at	Human mRNA for urea transporter.	39.6	68.7	104.9	43.5	103.3
RC_AA191744_at	Human mRNA for KIAA0661 protein, complete cds.	275	61.2	419.9	310.2	364.1
RC_R41682_at	Incyte Unique	0	4	19.3	0	20.8
RC_N55072_at	Incyte Unique	342.2	291.2	464	640.8	700.5
RC_AA191708_at	Incyte Unique	1120.8	1496.4	1979.7	1608.4	1919
RC_R40978_at	Incyte Unique	0	186.4	179.3	194.3	273.7
RC_AA169154_at	Incyte Unique	14.7	47.4	264	211.4	221.6
RC_AA114970_i_at	Human mRNA for AMP-activated protein kinase gamma2 subunit (AMPK gamma2 gene).	0	63	72.1	53.4	101.4
RC_F03321_at	Incyte Unique	0	88	55.8	69.8	95.2
RC_H30270_at	Incyte Unique	356.9	1055	1188.4	580.9	1216.7
RC_H29870_s_at	Incyte Unique	93.9	241.8	365.5	144	285.3
RC_AA102718_at	Incyte Unique	0	29.3	169	140	127
RC_AA401302_at	Human P2Y1 gene.	47.6	143.3	154.1	114.5	264.8
RC_H39669_at	Human mRNA; cDNA DKFZp434I116 (from clone DKFZp434I116); partial cds.	0	0	112.4	79.7	216.3
RC_H28985_at	Human clone HQ0319.	2098.7	2466.8	3672.2	2996.4	3076.4
RC_H66858_at	Human mRNA for KIAA1019 protein, partial cds.	282.4	396.9	504.5	508.6	496.8
X69878_at	Human receptor tyrosine kinase Flt4 (short form) mRNA, complete cds.	70.4	215.8	369.2	137.9	203.2
RC_N70577_at	Human chromosome 17 HSPC009 mRNA, complete cds.	1718.9	3119.8	4215.3	3019	3386
RC_AA084286_at	Human mRNA for KIAA0287 gene, partial cds.	22	12	71.7	60.5	43.2
X69908_ma1_at	Human gene for mitochondrial ATP synthase c subunit (P2 form).	1821.1	2009.3	2946.1	1476.3	2865.2
RC_N73080_at	Human clone 23688 mRNA sequence.	58.9	75.4	112.3	83.1	114.6
RC_H24460_s_at	Human immunophilin (FKBP52) mRNA, complete cds.	1376.5	1621.8	2200.6	1495.6	2540.8
RC_N73702_at	Human mRNA; cDNA DKFZp434H1235 (from clone DKFZp434H1235); partial cds.	19.4	60.3	617.1	671	508.5
RC_N71299_at	Incyte Unique	21.4	10.3	43.1	55.7	57.7
RC_N72708_at	Incyte Unique	50.7	157.6	310.2	158.3	281.3
RC_AA084162_at	Incyte Unique	0	0	32.6	0	42.1
X69111_at	Human HLH 1R21 mRNA for helix-loop-helix protein.	859.9	1364.5	1763.7	1307.1	1325.9

RC_N71072_at	Incyte Unique	4075.7	6814.7	9233.8	6296.7	11437.8
RC_AA157857_s_at	Human 40-kDa keratin intermediate filament precursor gene.	11120.2	12692.8	24341.6	15321.5	18697.4
X73358_s_at	Human gp130 associated protein GAM mRNA, complete cds.	1224.9	573.7	1689.1	1282.4	1646.1
RC_AA287131_at	Human NAD <sup>+</sup> -specific isocitrate dehydrogenase beta subunit precursor, mRNA, nuclear gene encoding mitochondrial protein, complete cds.	384.8	394.5	676.4	673.8	1176.2
X65724_at	Human mRNA NDP.	0	0	38.6	35.9	35.7
RC_H57130_at	Incyte Unique	51.7	0	105	77.2	117.8
RC_H56679_at	Incyte Unique	0	116.2	92.2	79.7	174.3
RC_N67239_at	Human (clone NG6-4) processed pseudogene of nucleophosmin/B23, complete insertion site.	48.3	171.1	303.8	172.7	265.3
RC_AA152323_at	Incyte Unique	0	0	31.6	0	33.5
RC_N63646_at	Incyte Unique	795.4	731.3	1228.6	824.2	1320.8
RC_H54565_at	Incyte Unique	0	14.2	16.8	0	22.5
X80763_s_at	Human serotonin 5-HT1C receptor mRNA, complete cds.	174	224.2	624.2	250.2	399.2
RC_N66357_s_at	Human mRNA for delta-subunit of mitochondrial F1F0 ATP-synthase (clone #5).	3162.4	5219.8	7242.3	5721.5	7724.4
RC_H56424_at	Incyte Unique	51.9	69.8	145.2	93.6	98.2
RC_H66642_f_at	Incyte Unique	175.7	158.4	550.8	634.9	703.1
RC_N68390_i_at	Incyte Unique	0	0	15.6	0	16.3
RC_N67906_at	Human mRNA; cDNA DKFZp564O043 (from clone DKFZp564O043).	19.7	65.1	160.7	111.3	103.4
RC_N63597_at	Incyte Unique	0	10.7	24.8	19.2	15.3
RC_N63611_at	Incyte Unique	23.1	22.3	67.1	66.3	69.1
RC_H50876_at	Human hepatocyte nuclear factor-3 gamma (HNF3G) gene, exon 2; and complete cds.	384.1	1508.8	1668.2	664.7	2392.8
RC_H66644_at	Incyte Unique	189.2	256.1	309.1	219.3	427.2
X77777_s_at	Human vasoactive intestinal peptide receptor mRNA, complete cds.	34.4	123.3	185.8	246.8	204.4
RC_AA156821_at	Incyte Unique	200.1	285.8	552.1	354	448.3
RC_H53572_at	Human mRNA full length insert cDNA clone EUROIMAGE 202740.	75.6	280.4	582.4	423	376.7
RC_AA079758_f_at	Human citrate transporter protein mRNA, nuclear gene encoding mitochondrial protein, complete cds.	3074.6	3414.3	4809.3	5142.1	5564.3
RC_N74699_at	Human malonyl-CoA decarboxylase precursor (MLYCD) mRNA, complete cds.	0	83.6	105.3	80.7	91.5
RC_N74624_at	Human mRNA for collectin 34, complete cds.	236.8	44.1	306.9	255.1	257
AA075599_at	Human NADH:ubiquinone oxidoreductase B22 subunit mRNA, nuclear gene encoding mitochondrial protein, complete cds.	1886	2343	3015.4	1822.9	3657
RC_H12470_at	Incyte Unique	58.4	105.6	130.5	80.2	123.7
RC_H12325_s_at	Incyte Unique	34.2	51.2	91.9	114.5	132.7
RC_N92934_s_at	Human cysteine-rich heart protein (hCRHP) mRNA, complete cds.	5926	4909	9723.3	8367.6	13408.3
RC_N91347_at	Incyte Unique	0	75.1	73.9	0	105.5
AA095473_at	Incyte Unique	93.8	246	363.1	231.7	340.7
RC_N93138_at	Human PHEX gene.	0	143.4	93.3	101.7	150.9
RC_AA285019_at	Human mRNA for dicarboxylate carrier protein.	109.6	245.6	478.7	298.5	381.3
RC_N91278_at	Incyte Unique	20	82.9	65	66.4	88.3
RC_N93105_f_at	Human mRNA; cDNA DKFZp566E0224 (from clone DKFZp566E0224).	880.6	972.4	1735.2	1367.9	1683.8
RC_H11463_at	Human HSPC301 mRNA, partial cds.	76.3	134.3	187.1	129.5	260.8
RC_AA151402_at	Human phosphoglucomutase-related protein (PGMRP) gene, complete cds.	562.1	611.3	2034.4	295.9	2499.6
RC_H09813_at	Human mRNA for KIAA1034 protein, partial cds.	837.2	1231.8	1426	1837	2243.3
RC_H77617_at	Human hook1 protein (HOOK1) mRNA, complete cds.	39.8	75.4	101.7	46.9	129.2
RC_N95585_at	Incyte Unique	0	0	147.5	0	168.8
RC_AA291137_at	Incyte Unique	25.2	70.2	94.6	106.3	184
RC_H77597_f_at	Human mRNA for metallothionein.	2477.8	8120.2	12227.1	8504.1	8237.1
RC_AA113303_at	Human tetraspan NET-7 mRNA, complete cds.	2229.4	2012.9	3509.4	3218.6	3436.6
RC_N95230_at	Incyte Unique	0	40.3	68.7	24.2	41
RC_H10661_at	Incyte Unique	519.5	624.3	1088.5	1334.3	1279.2
RC_H14089_at	Human synaptobrevin 2 (VAMP2) gene, complete cds.	118.5	123.2	499.6	71.3	436.9
RC_AA074350_at	Incyte Unique	0	0	77.6	0	83.3
AA095059_at	Human mRNA for KIAA0826 protein, partial cds.	0	0	36.9	105.8	95.8
RC_N76012_r_at	Incyte Unique	0	73.9	103	48.4	114.5
RC_AA402035_at	Incyte Unique	98.6	250.2	241.2	246	400.5
RC_AA076249_at	Human alpha-methylacyl-CoA racemase (RM) mRNA, complete cds.	44	170.1	191.7	82.5	256.2
RC_AA11895_at	Public Unique	43.6	18.7	72.7	75.5	57

RC_AA160320_s_at	Human ARSE gene, complete cds.	190.9	643.8	725.5	50.3	720.5
RC_AA401721_s_at	Incyte Unique	229.2	477.1	494.6	263.9	780.1
RC_AA160775_s_at	Human Bcl-2 binding component 6 (bbc6) mRNA, complete cds.	504.5	830.4	1138.7	1122.4	1171.6
RC_AA151778_at	Human mRNA for Claudin-7.	9172.6	10307.2	18723.4	17991.4	15759.2
RC_AA159980_s_at	Human mRNA for serine/threonine protein kinase EMK.	462.4	925.3	2297.4	1762.3	1390.8
X64177_f_at	Human mRNA for metallothionein.	1113.1	2439	5228.4	2991.5	4041
RC_AA075299_at	Human DSC2 mRNA for desmocollins type 2a and 2b.	1450.5	1473.7	2097.5	2303.2	2843
RC_AA290679_at	Human selenium-binding protein (hSBP) mRNA, complete cds.	4785.6	5511.3	15009.8	7295.3	12526.4
RC_N63192_s_at	Human phenylethanolamine-N-methyltransferase mRNA, complete cds.	308.6	131	481.9	370.7	523
RC_AA151676_at	Human mRNA for KIAA0994 protein, partial cds.	451	1112.4	1796.5	1422	2215.9
RC_AA074759_f_at	Human platelet-derived growth factor receptor alpha (PDGFRA) mRNA, complete cds.	269.6	654.5	863	439.1	630.3
X85137_s_at	Human kinesin-like spindle protein HKSP (HKSP) mRNA, complete cds.	15.3	73.1	57.9	90.1	112.1
RC_AA290674_s_at	Human 4E-binding protein 1 mRNA, complete cds.	30.7	322.8	412.1	43	454.3
RC_H18836_at	Incyte Unique	2181.8	2999.2	4434.5	3796.4	3821.6
RC_N80129_i_at	Human mRNA for Ig kappa light chain, anti-RhD, therad 7.	0	687.9	1905.6	361.6	1595.1
RC_N80129_f_at	Human mRNA for Ig kappa light chain, anti-RhD, therad 7.	3335.6	5457.8	8511	6087.9	7011.8
RC_AA195626_at	Incyte Unique	130.6	315.8	386	255.9	630.1
RC_R46074_at	Human TACC2 protein (TACC2) mRNA, partial cds.	0	216.2	417.8	80.1	364.6
RC_AA394258_s_at	Human mRNA for RD protein, RNA-binding.	1835.2	1478.3	3463.2	3118.6	3152.6
RC_T79842_at	Human clone 23781 mRNA sequence.	0	0	396.6	149.8	317.5
RC_AA017402_f_at	Human DNA for Ig heavy-chain variable region, complete sequence, 1 of 5.	0	0	86.2	74.1	53.6
RC_T80627_at	Incyte Unique	192.4	494	523.7	534.4	577.5
RC_T68873_f_at	Human mRNA for metallothionein isoform 1R.	1708.8	5257.7	7546.2	4563.4	6846.5
RC_N21548_at	Incyte Unique	208.9	198.1	290.7	230.4	322.2
RC_AA424155_s_at	Incyte Unique	52.2	109.2	172.6	93.9	123.3
RC_AA131919_at	Human mRNA for type II membrane protein, complete cds, clone:HP10328.	2363.9	2474.1	4029.5	3513.1	3533.6
AA253330_s_at	Human carbonic anhydrase precursor (CA 12) mRNA, complete cds.	3008.1	2189	5628.9	4643.3	5344.9
RC_AA609831_at	Incyte Unique	0	0	14.2	0	17
RC_AA609645_at	17-kDa PKC-potentiated inhibitory protein of PPI	97.2	153.6	268.3	121.6	349.3
RC_AA609650_s_at	Incyte Unique	0	7.1	29	21.2	27.5
RC_T63141_at	Incyte Unique	176.3	543.7	613.8	555.7	528.1
Z47038_s_at	Human microtubule-associated protein 1a (MAP1A) mRNA, complete cds.	404.8	415.6	776.9	569.5	623
RC_T94862_at	Incyte Unique	231.6	356.6	391	307.6	405.4
RC_AA609346_at	Human CGI-40 protein mRNA, complete cds.	257.5	129.6	712.4	567.4	578.3
RC_AA609460_at	Incyte Unique	370.3	666.1	1580.5	778.4	1037.7
RC_T93263_s_at	Human T-cell receptor alpha delta locus from bases 1 to 250529 (section 1 of 5) of the Complete Nucleotide Sequence.	485.9	758.9	1013.5	653.8	806.5
RC_AA609491_at	Human clone lambda MEN1 region unknown protein mRNA, complete cds.	228.7	102	367.1	552.4	613
RC_N21673_at	Incyte Unique	185.5	322	544.4	195.2	451.7
RC_AA342049_at	Incyte Unique	0	0	118.1	107.2	76.7
Z24727_at	Human mRNA (exon 6-9 part.) for smooth muscle tropomyosin.	1435.3	2455.1	2678.5	3017.8	3135.2
RC_T62857_at	Incyte Unique	0	11.4	54	13.9	46.9
RC_N32071_at	Incyte Unique	0	212.2	315.6	159.5	384
RC_N37048_at	Incyte Unique	0	41.3	73.9	110.1	117.1
RC_AA130235_at	Human mRNA for KIAA0567 protein, partial cds.	173.6	96.1	214.6	283.1	335
RC_T40895_at	Human protein tyrosine phosphatase PTPCAAX1 (hPTPCAAX1) mRNA, complete cds.	4	853.8	1347.9	2128.5	2485.9
RC_AA394121_at	Incyte Unique	2165.1	3247.3	5626.3	2407	3857.5
RC_AA129777_s_at	Human monocarboxylate transporter (MCT3) mRNA, complete cds.	0	38.7	50.8	0	48.3
RC_AA232678_at	Incyte Unique	0	65.2	98.6	0	88.3
RC_C20749_at	Human mRNA for KIAA1201 protein, partial cds.	0	0	66.2	87.1	123.7
RC_T24106_at	Incyte Unique	639.3	1524.2	3158.2	1886.6	2341.6
RC_N34918_at	Incyte Unique	20.8	46.3	87.7	47.5	69.7
RC_T49655_at	Incyte Unique	306.1	180.4	440	296.5	425.8
RC_AA024661_at	Human NPD002 mRNA, complete cds.	276.9	448.8	624.5	445.7	562.9
RC_AA621399_at	Incyte Unique	241.9	1389	6241.9	5975.1	4995.6
RC_AA349792_s_at	Human mutY homolog (hMYH) gene, complete cds.	197.8	526.3	935.2	622.1	1002.8

RC_T58137_at	Human mRNA for repressor protein, partial cds.	4	149.1	114.8	105.8	173.8
RC_AA348592_at	Incyte Unique	0	161.9	153.7	105.4	160.6
RC_AA131328_at	cytochrome oxidase subunit II	2158	3977.2	5082.4	7183.4	7101.1
U40282_at	Human integrin-linked kinase (ILK) mRNA, complete cds.	560.9	852.3	1149.5	948.1	929.2
RC_T50397_at	Human glycophorin B gene, exon 4.	0	0	75.7	39.6	62.2
RC_T51572_at	Public Unique	0	15.9	58.8	41.1	30.3
RC_T51539_s_at	Human hepatocyte growth factor-like protein homolog mRNA, partial cds.	0	0	246.4	133.2	287.1
U31875_at	Human Hep27 protein mRNA, complete cds.	0	0	0	0	0
RC_AA609113_at	Human mRNA, chromosome 1 specific transcript KIAA0503.	376.4	591.8	749.5	410.4	1066.7
RC_AA025887_at	Incyte Unique	395.1	963.6	1182.4	651.5	1303.2
RC_N24761_at	Incyte Unique	0	108.7	417.6	0	548.8
RC_W44798_at	Incyte Unique	0	0	973.6	426.6	958.7
RC_W44796_at	Incyte Unique	736.2	1039.5	1240.9	1675	1891.7
RC_AA004648_at	Incyte Unique	130.7	199.3	268.7	184.2	330.5
U09716_s_at	Human ERGIC-53 mRNA.	14.8	136.2	175.2	89	104.4
RC_W47175_at	Human 3'phosphoadenosine 5'-phosphosulfate synthase 2b isoform mRNA, complete cds.	3304	3645	5141.3	5415.7	5345.2
AA293334_s_at	Incyte Unique	3428.8	2987.2	5408.8	4070.3	4474
RC_AA133590_at	Incyte Unique	488.5	659.1	827.2	815.6	873.1
Z68280_cds2_s_at	Human (clone: SS20B/E6.0) alpha-adducin gene, exon 12.	43.4	359.3	326.6	373.3	501.5
RC_N26801_at	Human clone 25116 mRNA sequence.	2564.1	4443	5900.4	4817	7021.6
RC_N24899_at	Incyte Unique	392	746.6	1061.1	1113.5	1414.9
RC_AA133395_at	Human T-cell receptor alpha delta locus from bases 501613 to 752736 (section 3 of 5) of the Complete Nucleotide Sequence.	53.7	106.6	220.2	61.9	193.9
RC_AA428244_at	Human echinoderm microtubule-associated protein-like EMAP2 mRNA, complete cds.	217.9	385.9	489.1	559.7	584
RC_W63728_at	Incyte Unique	95.8	124.4	223.5	174	263.8
RC_W60007_s_at	Human mRNA for KIAA0203 gene, complete cds.	0	0	31.5	30.6	29.2
RC_N25718_at	Human serine/threonine protein phosphatase catalytic subunit mRNA, complete cds.	64.2	45	203.9	101.4	126.2
RC_AA236037_at	Human CGI-68 protein mRNA, complete cds.	153.8	342	682	565.9	456.7
RC_AA134576_at	Human GalNAc-T4 gene.	425.6	754.4	793.2	1099.3	1420.7
RC_AA001386_at	Incyte Unique	178.5	317.9	462.5	260.9	380.6
RC_AA427924_at	Human mRNA for KIAA0762 protein, partial cds.	302.5	280.9	587.7	294	660.6
U12255_at	Human IgG Fc receptor hFcRn mRNA, complete cds.	5508.5	5712.4	10242.3	8234.7	10450.9
RC_AA235236_f_at	Human CpG island DNA genomic MseI fragment, clone 47c5, reverse read cpg47c5.rt1a.	112.6	194.1	229.4	208.8	419.6
RC_T96374_at	Human myo-inositol monophosphatase 2 mRNA, complete cds.	1244.6	1592.5	2040.5	1619.3	3106.6
RC_W04569_at	Incyte Unique	88.7	0	237.3	159.5	217.2
RC_W15263_at	Incyte Unique	1096.2	4921.5	11458.8	12404.2	11916.3
RC_W04687_at	Human Ig lambda gene locus DNA, clone:288A10.	244.2	353.7	424.8	431.7	443.3
RC_N22268_at	Human mRNA for KIAA1135 protein, partial cds.	51.8	130.3	235.7	89.3	197.6
RC_T99639_s_at	Human KH type splicing regulatory protein KSRP mRNA, complete cds.	4580.8	4316.5	8403.2	4785.7	7024.9
AA017469_at	Human mRNA; cDNA DKFZp434B187 (from clone DKFZp434B187); partial cds.	76.1	91.6	124	113.8	148.5
RC_AA234442_at	Human NTF2-related export protein NXT1 (NXT1) mRNA, complete cds.	19	46.5	47.9	23.5	45
RC_T97982_at	Incyte Unique	0	104.5	180.8	118.5	322.1
RC_AA600328_at	Human Wolf-Hirschhorn syndrome candidate 2 protein (WHSC2) mRNA, complete cds.	4.3	26.1	44.7	9.7	42.7
RC_W16424_at	Human mRNA; cDNA DKFZp434H1235 (from clone DKFZp434H1235); partial cds.	45.2	350.8	807.1	1184.8	1082.7
RC_AA009593_at	Incyte Unique	25.9	94.7	119.6	171.6	217
Z49989_at	Human mRNA for smoothelin-B.	227.3	189.9	1210.6	421.6	1112
RC_N30034_at	Incyte Unique	69.2	169.1	298	142.2	183.3
RC_AA426011_at	Human mRNA; cDNA DKFZp434N161 (from clone DKFZp434N161).	484.9	810.3	1363.6	690	1194.5
RC_W32405_at	Human full length insert cDNA clone ZE12H05.	0	884.6	883.9	837.1	778.6
RC_AA007158_f_at	Incyte Unique	0	175.5	184.8	232.3	349.6
RC_AA297783_at	Human mRNA; cDNA DKFZp434F0217 (from clone DKFZp434F0217).	508.2	729.6	1000	1017.6	986.9
RC_AA135406_at	Incyte Unique	0	140.5	244.9	192.7	305.7
RC_AA599954_at	Human mRNA for KIAA1254 protein, partial cds.	264.8	174.6	304.8	370.6	382.5
RC_W23441_at	Incyte Unique	0	60	162.7	277	256.3

RC_AA025883_f_at	Human zinc finger protein (MAZ) mRNA.	8057.3	7251.2	12352.9	7793.2	13111.8
RC_AA232114_s_at	Human cytosolic epoxide hydrolase mRNA, complete cds.	566.7	1604.5	1815.1	1455.4	1547.3
U47105_at	Human H105e3 (H105e3) mRNA, complete cds.	40.6	158.3	224.2	43.8	261.3
RC_AA211613_at	Human genomic DNA, chromosome 22q11.2, clone N27C7.	9052.1	11612.4	21114.8	12958.5	16872.1
AA044781_at	Human mRNA for KIAA1187 protein, partial cds.	0	173.3	606.2	67.1	616.9
RC_AA037410_s_at	Incyte Unique	103.9	205.5	268.6	188.4	277.2
RC_R66436_at	Incyte Unique	0	22.4	24.8	44.4	39.9
AA215970_at	Incyte Unique	327.8	756.1	795.3	782.4	859.3
U72209_at	Human YY1-associated factor 2 (YAF2) mRNA, complete cds.	16.3	0	21.5	60.2	44.8
RC_F01534_at	Incyte Unique	50.6	148.4	210.4	262	295.9
RC_AA039854_at	Human mRNA full length insert cDNA clone EUROIMAGE 26539.	0	91.8	133.6	103.3	122.2
RC_R69589_at	Public Unique	26.5	51.2	95	36.1	121.7
RC_R71225_at	Human T-cell gamma receptor locus, complete sequence	128.4	476.8	810.4	353.6	568.2
RC_R63388_at	Incyte Unique	0	282.3	1148.8	596.6	949.5
RC_AA034910_at	Human HSPC326 mRNA, partial cds.	245.2	373.5	413	287.5	504.6
RC_R85880_at	Human mRNA for KIAA0994 protein, partial cds.	954.7	2586.7	8465.2	5501.7	6090.6
U67171_at	Human selenoprotein W (seW) mRNA, complete cds.	1392	2118.6	2669.6	2397.4	2734
RC_R79617_at	Human mRNA; cDNA DKFZp564E1962 (from clone DKFZp564E1962); partial cds.	0	108.5	188.5	120.7	120.5
RC_AA142849_at	Incyte Unique	938.1	577.7	1479.3	986.2	1880.9
RC_R71689_s_at	Human mRNA for KIAA0338 gene, partial cds.	164.8	94.2	475.5	481.6	430.8
U70732_ma1_at	Human mRNA for alanine aminotransferase.	833.6	856.6	2174.7	1939.8	1391.2
Y00503_at	Human 40-kDa keratin intermediate filament precursor gene.	3018	3238.7	5762.8	4481.6	4780
U78722_at	Human zinc finger protein 165 (Zpf165) mRNA, complete cds.	11.3	88.1	171.4	80.4	109.3
RC_R62371_at	Incyte Unique	29	128	150.3	23.3	109.4
RC_R88711_at	Incyte Unique	207.7	286.9	456.7	423.7	364.1
RC_R53467_at	Incyte Unique	53.1	98.1	113.4	52	135.7
RC_R54494_at	Human mRNA; cDNA DKFZp564C1416 (from clone DKFZp564C1416).	55.8	59.6	78.6	80.1	97.5
U83908_at	Human nuclear antigen H731 mRNA, complete cds.	543.8	634.3	832.9	1258.6	1183.2
RC_N51406_at	Incyte Unique	109.8	85.4	402.2	336.2	306.6
RC_AA043458_s_at	Human zinc finger protein ZNF137 mRNA, complete cds.	30	5.6	157.1	110.4	129
RC_AA393812_at	Human full length insert cDNA clone YB21E09.	228.5	203	489.7	230.3	360.9
RC_AA283046_at	Human mRNA for KIAA0448 protein, complete cds.	0	106.4	174.4	139.5	242.2
RC_AA195718_at	Human TPI (triosephosphate isomerase) pseudogene psi-13C.	150.8	214.8	448.7	416.5	635
RC_N51053_s_at	Incyte Unique	20.7	80.1	196.2	40.1	234.4
RC_R56121_s_at	Incyte Unique	0	0	548.2	466.9	357.4
RC_R62173_f_at	Human UDP-glucose dehydrogenase (UGDH) mRNA, complete cds.	1496.1	2304.7	3122.2	3709.6	3459
RC_N49405_s_at	Human mRNA for ubiquitin-like protein, complete cds.	0	91.3	256.2	95.1	223.9
RC_F01930_at	Incyte Unique	5	96.7	77.8	70.3	113.8
AA044842_at	Human Autosomal Highly Conserved Protein (AHCP) mRNA, complete cds.	0	58.2	46.1	37.9	54.2
U81599_at	Human homeodomain protein HOXB13 mRNA, complete cds.	145.7	137.5	291.5	202	390.7
RC_F02245_at	Human DNA for monoamine oxidase type A (14) (partial).	2517.2	6410.1	7589.2	8340.4	8487.8
RC_R56557_f_at	Human mRNA; cDNA DKFZp434K2235 (from clone DKFZp434K2235); partial cds.	0	5.2	12.8	0	11.4
RC_R56570_at	Incyte Unique	0	172	286.6	182.3	338.7
AA046674_at	Incyte Unique	113.5	419.7	462	415.8	529.6
RC_H99370_at	Human mRNA for KIAA0706 protein, complete cds.	772.1	1324.2	1868.7	1625.3	2039.6
RC_D59328_at	Human PPAR gamma coactivator-1 (PPARGC1) gene, exon 13 and complete cds.	11.5	62.8	62.1	78.7	83.4
U65579_at	Human mitochondrial NADH dehydrogenase-ubiquinone Fe-S protein 8, 23 kDa subunit precursor (NDUFS8) nuclear mRNA encoding mitochondrial protein, complete cds.	135.9	229.9	327	147.7	412
AA234665_at	Human supervillin mRNA, complete cds.	185.7	375.3	632.3	231.3	806
AA237089_at	Human ribosomal protein S20 (RPS20) mRNA, complete cds.	233.3	222.3	372.2	285	420
RC_T16232_at	Human mRNA for synaptogyrin 3.	17.6	0	31.7	19.9	34.4
RC_AA372018_at	Incyte Unique	654.2	1057.7	1523.6	1337.4	1385.8
RC_AA137034_at	Human glutathione synthetase mRNA, complete cds.	40.4	104.4	219.4	77.8	243
RC_D20113_at	Human unknown mRNA.	1069.5	1116.1	2041.7	1193.4	1472.6
RC_AA029651_f_at	Human COX VIa-L mRNA for cytochrome c oxidase liver-specific subunit	10126.5	15894.6	19323.5	20314.3	24674.8

Vla (EC 1.9.3.1).						
RC_AA029697_at	Incyte Unique	4	59.9	75.5	68.6	93.6
RC_N45300_at	Human leukemogenic homolog protein (MEIS1) mRNA, complete cds.	73.6	105.1	251.1	69.4	237.8
RC_AA027253_s_at	Human mRNA for cytochrome c oxidase subunit VIb (EC 1.9.3.1).	4490.7	6072.6	7677.4	7306	8494.9
RC_R98105_at	Incyte Unique	1073.9	1730.7	1843.8	1483.4	1943.2
RC_T23640_s_at	myotonin protein kinase=thymopoietin homolog [Human, muscle, mRNA, 2503 nt].	297	446.2	568.7	476.3	954.2
RC_AA026349_r_at	Human mRNA for myopodin.	86.3	183	338.9	99.5	242.9
RC_N40917_at	Human mRNA: cDNA DKFZp434N014 (from clone DKFZp434N014); partial cds.	0	16.4	29.2	0	33.4
RC_D11628_i_at	Human znf-xp protein mRNA, complete cds.	0	0	18.2	13.4	35.2
U51095_at	Human homeobox protein Cdx1 mRNA, complete cds.	630.6	759.9	1169.1	961.9	1094.5
RC_AA227913_at	Human Pig11 (PIG11) mRNA, complete cds.	89	546.1	543.5	207.9	569.7
RC_D11925_r_at	Novel Human gene mapping to chromosome 1.	46.5	70.6	95.3	81.8	78.5
RC_D45492_at	Human mRNA for KIAA0974 protein, partial cds.	0	24.2	30.2	0	27.8
RC_D25755_s_at	Human mRNA: cDNA DKFZp586G1219 (from clone DKFZp586G1219); partial cds.	0	0	39.6	28	21.8
RC_N64355_s_at	Human mRNA for electron transfer flavoprotein beta subunit.	2290.6	3275.2	5968.6	3797.8	5238.3
RC_AA031379_at	Incyte Unique	0	0	33.2	23.4	32.3
RC_AA033798_s_at	Human mRNA for KIAA0141 gene, complete cds.	87.9	196.6	206.8	168.8	300.7
RC_R97804_at	Incyte Unique	808.5	1964.2	3737.8	1588.8	2047.6
RC_R93068_at	Incyte Unique	146.9	305.8	462.6	227.5	372.5
RC_N47972_at	Human mRNA for small GTP-binding protein Rab36, complete cds.	92.5	46.8	119.8	208.6	160.8
RC_R93176_f_at	Human carbonic anhydrase I (CAI) mRNA, complete cds.	311.5	905.6	1690.4	1897	1931.6
RC_R91819_at	Incyte Unique	284.6	212.8	1193.7	805	715.4
RC_R97176_at	Incyte Unique	158.4	395.2	591.9	582	483.8
RC_AA292305_s_at	Human cytokine-inducible SH2-containing protein (G18) mRNA, complete cds.	0	61.2	93	48.2	52.4
RC_AA398090_at	Human CG1 mRNA, complete cds.	0	19	0	0	41.7

## A4.9 Genes in cluster 18 of figure 4.59

Acc No	Gene Description	Mean Average Difference value				
		UCi	CDi	UCu	CDu	NI
M33600_f_at	Human MHC class II HLA-DR-beta-1 (HLA-DRB1) mRNA, complete cds.	4973.4	6677.9	1700.5	5517.6	1491.1
N42272_s_at	Human mRNA for hDj9, complete cds.	183.2	223.3	59	141	100.1
U90552_s_at	Human butyrophilin (BTF5) mRNA, complete cds.	642.4	1872	797.7	1307.4	394.7
RC_N49846_at	Human mRNA; cDNA DKFZp434B1021 (from clone DKFZp434B1021).	918.7	889.7	264	745.4	116.1
RC_N51617_at	Public Unique	38.2	20.4	0	51.9	12.8
RC_Z40325_at	Incyte Unique	92.9	119.3	39.6	136.6	27.2
RC_R99084_at	Incyte Unique	113	119.9	0	112	13.1
AA194091_at	Incyte Unique	152.6	403	157.7	275.7	57
U88726_at	Human mRNA for symplekin.	41.3	45.4	0	36.8	8.4
N44756_at	Incyte Unique	65.5	43.4	0	65.8	0
M37766_at	Human MEM-102 glycoprotein mRNA, complete cds.	577.4	535.8	0	421.4	38.5
M57466_s_at	Human MHC class II lymphocyte antigen (HLA-DP) beta chain mRNA, complete cds.	1223.6	1479.4	0	738.3	186.5
RC_AA029317_at	Incyte Unique	192.2	166.7	74.7	223.8	86.6
U96781_cds1_at	Human Ca <sup>2+</sup> ATPase of fast-twitch skeletal muscle sarcoplasmic reticulum ad and neonatal isoforms (ATP2A1) gene, exon 15.	74.4	82.2	0	67.9	41
RC_AA279667_s_	Human cyclophilin-related protein mRNA, complete cds.	926	729.1	175.8	760.5	355.2
RC_AA461463_at	Human mRNA for KIAA0836 protein, partial cds.	86.5	150.9	45.6	130.9	70.8
AA402095_s_at	Incyte Unique	185.3	303.2	66.9	289.5	154
AA167268_at	Human ras inhibitor mRNA, 3' end.	50	84.5	27.3	55.4	18.8
RC_R98774_at	Incyte Unique	118.4	144.6	65	80.6	0
M74719_at	Human SEF2-1A protein (SEF2-1A) mRNA, 5' end.	573.9	596.1	49	379.4	259.3
M62486_at	Human proline-rich protein (PRP) mRNA, complete cds.	93.5	106.4	0	95.9	35.4
RC_H37820_at	Incyte Unique	454.7	665	122.1	647.6	199.3
T30341_s_at	Incyte Unique	98.7	127.9	59.9	88.4	32.5
RC_T15665_at	Incyte Unique	736	1003.6	223.5	1215.8	481
RC_AA034014_at	Incyte Unique	143	192.6	0	178.2	91.2
RC_T02889_at	Public Unique	91	34.2	0	102.7	0
M62424_at	Human thrombin receptor mRNA, complete cds.	45	46.9	0	69.4	0
AA403202_at	Human Ikb kinase-a (IKK-alpha) mRNA, complete cds.	121.5	190.3	61.8	229.5	76
RC_H12297_at	Human intestine N-acetylglucosamine 6-O-sulfotransferase (I-GlcNAc-6-ST) mRNA, complete cds.	564	683.7	510.4	795.1	100.7
RC_AA160890_s_	Human mRNA for KIAA0389 gene, complete cds.	92.4	88.8	0	94	29.3
RC_AA461458_at	Incyte Unique	204.8	280.3	125	226.7	128.1
RC_AA126855_at	Incyte Unique	181.6	346.7	119.1	302.1	104.2
RC_H39853_at	Human olfactory receptor (OR7-86) pseudogene, partial sequence.	67.4	131.2	0	127.1	39.7
RC_AA428240_at	Incyte Unique	75.3	192.2	45.8	95.1	39.1
AA421370_at	Human hypoxanthine phosphoribosyltransferase (HPRT) gene, complete cds.	5	4	0	5	0
RC_AA476621_at	Human mRNA; cDNA DKFZp586N2119 (from clone DKFZp586N2119).	1188.8	931	0	1003.2	297.9
RC_AA430675_at	Human mRNA for Fanconi anemia group G.	206.9	275.1	70.8	79.7	0
RC_T03229_f_at	Incyte Unique	1774.1	1645.6	426	1763.8	972.6
RC_R93914_s_at	Human genomic DNA, chromosome 6p21.3, HLA Class I region, section 5/20.	215.1	220.6	0	195.6	30.5
X00274_at	Human gene for HLA-DR alpha heavy chain a class II antigen (immune respon gene) of the MHC (MHC).	7955.7	9252.1	3331.6	9847.5	2571.3
RC_AA425836_at	Incyte Unique	146.1	245.3	114.1	228.4	52.6
RC_AA252395_at	Human NY-REN-45 antigen mRNA, complete cds.	57.8	81.8	27.7	57.1	12.6
RC_AA126803_at	Human 17-beta-hydroxysteroid dehydrogenase IV (HSD17B4) gene, exon 23.	34.4	50.2	0	9.6	9.7
RC_AA476609_at	Incyte Unique	455.2	828.2	153.7	704.3	451.8
RC_AA428204_at	Human vitamin D3 receptor interacting protein (DRIP80) mRNA, complete cd	176.5	207.1	84.3	110.8	91.3
RC_R87160_at	Incyte Unique	233.1	301.5	50.9	398.1	129
RC_Z38779_at	Human clone 24420 mRNA sequence.	64	72.1	0	49.9	0
U78521_at	Chlorocebus aethiops HBV-X associated protein 2 (XAP2) mRNA, complete cds.	1188.3	1268.3	246.6	1170.1	690.7
AA203630_at	Incyte Unique	172.2	389.3	155.2	280.5	66.8
RC_N47905_at	Human pancreas-enriched phospholipase C mRNA, complete cds.	35.1	30.4	0	53.8	11.9
M55621_at	Human alpha-1,3-mannosyl-glycoprotein beta-1, 2-N-acetylglucosaminyltransferase (MGAT) gene, complete cds.	340.9	335.7	123.5	304.7	100.3

M34455_at	Human interferon-gamma-inducible indoleamine 2,3-dioxygenase (IDO) mRNA complete cds.	255.2	319.9	0	207.3	127.6
RC_AA128462_at	Incyte Unique	74.2	115.6	0	94.3	32
M62424_at	Human thrombin receptor mRNA, complete cds.	45	46.9	0	69.4	0
U77846_ma1_s_at	Human elastin mRNA, complete cds.	111.4	191.4	0	195.3	77.1
RC_AA252454_at	Incyte Unique	61.3	79.1	0	58.3	25.4
RC_T23490_s_at	Incyte Unique	598.9	528.4	216.6	590.5	308.6
RC_AA428139_s_	Human mRNA for KIAA0936 protein, complete cds.	59.5	85.1	30.7	63.8	14.6
RC_H16790_at	Incyte Unique	121.5	174.3	0	212.9	73.4
RC_R94756_at	Incyte Unique	286.9	420	104	311.2	148.9
RC_AA029328_s_	Human mRNA for KIAA0073 gene, partial cds.	177.6	150.7	48.3	142.9	69.1
RC_Z38777_f_at	Incyte Unique	161.2	218.4	125.8	190.5	44.6
RC_AA034040_at	Incyte Unique	82.1	71	0	50.6	0
RC_AA251902_at	Human clone 23753 mRNA sequence.	239.7	516.4	288	448.7	150.8
M69066_at	Human cysteine-rich repeat-containing protein S52 precursor, mRNA, complete cds.	1452.3	2161.6	695.7	1561.2	1006.4
M74558_at	Human SIL mRNA, complete cds.	43.7	56	0	34.1	19.4
RC_Z39762_s_at	Human mRNA for KIAA0882 protein, partial cds.	493	414.3	111.1	492	93.7
M32304_s_at	Human mRNA; cDNA DKFZp586J021 (from clone DKFZp586J021).	329.8	447.1	244.8	580.8	183.4
RC_H17861_at	Incyte Unique	530.7	680.4	181	483.1	111.3
RC_AA429809_at	Incyte Unique	59	78.7	34.3	53.2	21.4
M32053_at	Human H19 gene, complete sequence.	402.2	599.4	0	139	150.8
RC_T10100_f_at	Human genomic DNA, chromosome 22q11.2, Cat Eye Syndrome region, clone:c60H5.	99.9	135.8	0	105.3	62.8
RC_T16645_s_at	Human partial C1 mRNA.	711.7	1305	350.1	1155.4	516.4
U94585_at	Human requiem (HREQ) mRNA, complete cds.	369.4	631.4	100.6	538.4	213
M34996_s_at	Human mRNA for DC classII histocompatibility antigen alpha-chain.	805.1	1645.9	403.6	988.9	381.8
M64929_at	Human protein phosphatase 2A alpha subunit mRNA, complete cds.	219.4	334.8	79.4	292.7	118.6
U83461_at	Human putative copper uptake protein (hCTR2) mRNA, complete cds.	75.9	149.5	0	145.6	8.4
RC_AA280627_at	Human CpG island DNA genomic MseI fragment, clone 15a10, reverse read cpg15a10.rt1b.	200.8	123.6	0	244.2	64.5
L44338_at	Human aminopeptidase PILS (APPILS) mRNA, complete cds.	13.2	9.7	0	15.7	0
U79267_at	Human protein serine/threonine phosphatase 4 regulatory subunit 1 (PP4R1) mRNA, complete cds.	85.7	208.3	0	177.5	35.6
M64174_at	Human protein-tyrosine kinase (JAK1) mRNA, complete cds.	158.4	275.7	0	287.3	104.3
N28889_at	Human mRNA for NRD1 convertase.	284.3	256.7	0	302.1	92.7
RC_AA252293_at	Incyte Unique	136.2	124.8	0	85.3	52.6
AA187579_at	Human MCT-1 mRNA, complete cds.	146.7	237.6	123.7	251	33.7
M59829_at	Human genomic DNA, chromosome 6p21.3, HLA Class I region, section 2/20.	44	68.8	0	52.9	25.6
M60091_at	Human galactose-1-phosphate uridyl transferase (GALT) mRNA, complete cds.	246.1	522.7	155.5	336.6	43.6
U89606_at	Human mRNA; cDNA DKFZp434P051 (from clone DKFZp434P051).	148.5	114.8	0	160.9	61.3
AA426156_at	Human vesicle trafficking protein sec22b mRNA, complete cds.	89.8	152.9	35.5	126.8	22.6
RC_AA429998_at	Incyte Unique	85.1	136	46.3	149.6	58
AA174183_s_at	Incyte Unique	299.8	463.5	145.9	359.4	226.4
RC_Z41680_at	Human mRNA; cDNA DKFZp566P013 (from clone DKFZp566P013).	73.5	87.9	0	75.2	18.1
M91488_at	Incyte Unique	26.6	17	0	22.3	7
M64595_at	Human HSPC022 mRNA, complete cds.	94.3	122.4	0	38.2	0
RC_AA428364_s_	Human genomic DNA, chromosome 22q11.2, Cat Eye Syndrome region, clone c17H1.	219.9	424	91.1	276.9	219.4
RC_AA127263_at	Human mRNA; cDNA DKFZp434L1028 (from clone DKFZp434L1028).	36.9	40.3	0	31.2	6.5
RC_AA164633_at	Incyte Unique	110.2	83.6	34.2	108.3	43.7
RC_AA252603_at	Human BACH1 mRNA, complete cds.	95.8	126.3	55.9	116.7	18.7
RC_AA279799_at	Incyte Unique	895.6	1073.3	678.2	1016.7	204
RC_T16346_at	Incyte Unique	120.8	207.8	97	156.3	76.5
L41067_at	Human NF-AT4c mRNA, complete cds.	349.7	549.9	123.4	565	251.2
RC_H27560_at	Public Unique	123.8	146	47.9	210.4	52
RC_T16429_at	Human mRNA; cDNA DKFZp434C136 (from clone DKFZp434C136).	94.5	183.7	0	91.1	53.9
RC_AA429861_at	Incyte Unique	262.6	319.2	98.2	331.8	37.9
RC_R91828_at	Incyte Unique	147.1	244.4	80.6	177.4	106.9
RC_AA428288_at	Human mRNA; cDNA DKFZp564L0864 (from clone DKFZp564L0864); partial cds.	127.1	243.3	144.3	253.2	48.7

RC_AA280588_at	Human mRNA for KIAA0784 protein, partial cds.	97.5	195.2	57	201.9	28.9
AA410529_s_at	Human mRNA for KIAA0191 gene, partial cds.	124.3	181.5	44	164.4	48.2
RC_AA253216_at	Incyte Unique	137.6	127	34.6	225.1	23.5
RC_AA252355_at	Human mRNA for putative membrane protein, complete cds.	266.4	283.5	95.6	253	180.9
RC_R98192_at	Incyte Unique	58.6	107	0	60.1	45.1
RC_T15977_at	Human mRNA for KIAA0392 gene, partial cds.	84.3	90.6	0	72.2	45
RC_AA026969_at	Incyte Unique	23.9	37.3	0	30.3	0
RC_T03927_at	Incyte Unique	343.4	311.4	93.5	366.7	205.3
RC_AA253168_s_	Human mRNA for KIAA0904 protein, partial cds.	246.9	324.3	79	209	187.2
RC_Z40234_at	Incyte Unique	277.8	355.3	54.7	141.9	108.7
M30269_at	Human nidogen mRNA, complete cds.	166.3	220.5	0	112.1	106.2
RC_AA122350_at	Human full length insert cDNA clone ZD40H07.	234.4	256	61.5	217.6	3.3
RC_AA164851_at	Human genomic DNA, chromosome 6p21.3, HLA Class I region, section 8/20.	36.4	39.5	9.3	58.7	5.4
RC_T10682_at	Incyte Unique	115.5	108.7	0	124.2	9.3
RC_N50675_at	Incyte Unique	66.4	86.6	0	60.1	29.2
RC_AA029823_f_	Human mRNA for HLA-D class II antigen DQw1.1 beta chain.	1059.1	2537.8	634.3	1569.3	365.4
RC_AA427608_at	Human mRNA for transcription factor, E4TF1-47, complete cds.	102	198.5	70.8	160.2	30.6
RC_H29285_at	Incyte Unique	79.4	48.9	0	77.4	35
N41849_at	Human cytokine receptor related protein 4 (CYTOR4) mRNA, complete cds.	45.5	67.1	36	66.8	0
AA422160_at	Human NAP (nucleosome assembly protein) mRNA, complete cds.	138	124	17.4	125.6	66
RC_Z41208_at	Incyte Unique	192.9	174.1	68.8	204.4	102.5
S53911_at	CD34=glycoprotein expressed in lymphohematopoietic progenitor cells {alternatively spliced, truncated form} [Human, UT7, mRNA, 2657 nt].	217.1	453.8	0	201	83.6
RC_AA463697_s_	Human carboxypeptidase D mRNA, complete cds.	75.6	110.6	0	115.7	47.9
M63838_s_at	Human interferon-gamma induced protein (IFI 16) gene, complete cds.	183.8	238.9	18.6	127.3	57.3
RC_AA465194_at	Incyte Unique	117	253.3	55.7	159.5	35.6
RC_AA032250_at	Human mRNA; cDNA DKFZp434H2019 (from clone DKFZp434H2019).	218.3	276.1	128.5	264.7	86.7
RC_Z41517_at	Incyte Unique	37.5	89.6	0	66.6	0
RC_AA279777_at	Human mRNA; cDNA DKFZp566M043 (from clone DKFZp566M043).	163.3	322.7	88.2	264.1	151.6
RC_Z39983_s_at	Human mRNA for KIAA0561 protein, partial cds.	434.2	854.2	0	729.9	116.9
RC_Z39976_at	Human mRNA; cDNA DKFZp564D0462 (from clone DKFZp564D0462).	138.3	151	75.7	136.9	13.1
RC_AA159962_at	Incyte Unique	35.3	28	0	36.7	10.9
RC_AA027806_i_	Human NY-REN-25 antigen mRNA, partial cds.	35.8	45.9	0	16.5	12.1
RC_H22956_at	Human clone 23736 mRNA sequence.	45.3	67.5	0	22.3	0
M68864_at	Human ORF mRNA, complete cds.	307	598.9	129.1	588.7	150.7
M37238_s_at	Human phospholipase C mRNA, complete cds.	165.9	279.3	0	115	34.4
RC_Z98502_at	Incyte Unique	245.5	315.5	103.8	183.6	140
RC_R91401_at	Public Unique	101.8	45.3	0	116.3	3.3
L38961_at	Human putative transmembrane protein precursor (B5) mRNA, complete cds.	243.7	312.7	163.2	151.5	61.7
RC_H20568_at	Human mRNA; cDNA DKFZp434J1217 (from clone DKFZp434J1217).	129.8	201.9	92.6	178.7	30
M60750_f_at	Human histone H2A.1 (H2A) gene, complete cds.	45.8	70.6	0	76.3	28.1
RC_N48290_at	Human S164 gene, partial cds; PS1 and hypothetical protein genes, complete cds and S171 gene, partial cds.	412.3	428.1	190.5	320	227
RC_AA136569_at	Human mRNA for KIAA0187 gene, complete cds.	134.2	145.7	51.3	109.8	62.7
U77665_at	Human RNaseP protein p30 (RPP30) mRNA, complete cds.	69.7	106.9	0	84.5	39.3
U16282_at	Human ELL mRNA, complete cds.	197.7	180.4	56.2	140.5	0
RC_T90369_at	Human clone IMAGE:110987 mRNA sequence.	34.5	52.3	0	40.3	21.2
U45878_s_at	Human TNFR2-TRAF signalling complex protein mRNA, complete cds.	718.2	1126.5	276.4	780.8	51.7
RC_AA134138_at	Human leucine aminopeptidase mRNA, complete cds.	303.5	912.6	247.6	568	120.5
RC_T90476_at	Incyte Unique	347.1	550.3	271.8	502.5	227.3
U16031_at	Human transcription factor IL-4 Stat mRNA, complete cds.	133.3	263.3	0	174.6	120.3
RC_T90549_s_at	Human granule membrane protein-140 mRNA, complete cds.	253.7	420.6	0	123.6	0
RC_AA257090_at	Incyte Unique	433.1	789.7	272	573.3	208.6
AA248582_at	Human mRNA for KIAA0737 protein, complete cds.	142.1	320.7	47.8	191.4	71.9
RC_T90217_f_at	Incyte Unique	180.2	121.9	0	141.9	3.3
RC_AA257093_r_	Human mRNA for T-cell specific protein.	1456.3	1851.5	510.2	1724.9	357.5
U46571_at	Human tetratricopeptide repeat protein (tpr2) mRNA, complete cds.	120.2	320.5	39.4	221.5	151.8
AA287840_at	Human full length insert cDNA clone YR30C05.	17.7	36.6	0	29.2	0

U15085_at	Human mRNA for KIAA0295 gene, partial cds.	781.7	1801.3	294.5	1589	335
U15128_at	Human beta-1,2-N-acetylglucosaminyltransferase II (MGAT2) gene, complete cds.	105.2	93.1	39.9	103.4	55.6
RC_AA443993_at	Incyte Unique	258.7	293.4	0	318.4	6
U15782_at	Human cleavage stimulation factor 77kDa subunit mRNA, complete cds.	217.6	355.5	85.1	261	180.6
RC_AA017130_at	Incyte Unique	32.6	40.2	0	18.8	0
RC_T90074_at	Human clone IMAGE:110582 mRNA sequence.	105.7	54	42.4	95.3	0
U45285_at	Human specific 116-kDa vacuolar proton pump subunit (OC-116kDa) mRNA complete cds.	407.7	961.8	242	490.7	344.6
U44772_at	Human palmitoyl-protein thioesterase gene, complete cds.	144.5	329	32.6	146.4	136.1
U14417_at	Human Ral guanine nucleotide dissociation stimulator mRNA, partial cds.	224.1	310.3	0	169.9	79.6
U18321_at	Human DAP-3 mRNA.	48.4	60.1	0	59.7	24.2
RC_AA147680_at	Human signal transducer and activator of transcription Stat5B mRNA, complete cds.	913.7	934.1	321	842.8	587.8
RC_N25798_at	Incyte Unique	67.7	66.2	0	54.3	40.1
RC_AA450294_s.	Human CGI-60 protein mRNA, complete cds.	49.2	116.1	43.6	95.5	12
RC_N25555_at	Incyte Unique	67.1	108.8	0	107.6	0
RC_AA446570_at	Human cyclin-D binding Myb-like protein mRNA, complete cds.	83.4	132.1	43.7	147.3	61.9
RC_AA450281_at	Human CGI-111 protein mRNA, complete cds.	76	124.8	45.9	119.2	30.5
RC_T92735_at	Human cig41 mRNA, partial sequence.	409.5	587.6	52.4	382.6	223.4
U18300_at	Human damage-specific DNA binding protein p48 subunit (DDB2) mRNA, complete cds.	178	268.3	46	288.1	71.7
RC_AA001046_at	Incyte Unique	264.7	394.6	0	210.4	182.3
RC_AA257093_s.	Human mRNA for T-cell specific protein.	958.7	1205.3	45.2	798.6	169.8
RC_N26398_at	Incyte Unique	62.5	25.7	0	71	15
RC_N26482_at	Human mRNA for KIAA0448 protein, complete cds.	62.7	46.1	0	44	0
RC_AA451703_at	Human U5 snRNP 100 kD protein mRNA, complete cds.	379	444	183.6	284.7	79.9
RC_T91057_at	Incyte Unique	184	168.5	0	143.1	86
U43899_at	Human signal transducing adaptor molecule STAM mRNA, complete cds.	59	69.3	0	46.8	17.5
RC_AA446461_at	Human mRNA; cDNA DKFZp434I1714 (from clone DKFZp434I1714); partial cds.	138	287.1	84.7	225.2	147.5
AA285293_at	Human gene for JKTBP2, JKTBP1, complete cds.	93	103.9	22.3	92.1	67.6
U43522_at	Human cell adhesion kinase beta (CAKbeta) mRNA, complete cds.	322.4	256.2	123.8	381.5	140.8
RC_AA452161_at	Human mRNA for ATP-dependent metalloprotease YME1L.	412.9	493.8	174.8	548.9	241.7
R82411_at	Human dek mRNA.	135.5	304.8	73.3	149.1	74.6
RC_AA261907_at	Human CGI-114 protein mRNA, complete cds.	306.7	341.3	115.3	185.6	146.9
RC_T92946_at	Human mRNA for five-lipoxygenase activating protein (FLAP).	467.1	644.4	152.4	412.1	58.7
RC_N30161_at	Human proline-rich Gla protein 1 (PRGP1) mRNA, complete cds.	244.6	313.2	0	152.1	166.4
RC_W69660_at	Human 5'-AMP-activated protein kinase beta-1 mRNA, complete cds.	426.6	661.2	385.2	478.2	124.6
AA247643_at	Incyte Unique	136.1	175.3	69.8	126	60.9
U11875_s_at	Human interleukin-8 receptor type B (IL8RB) mRNA, splice variant IL8RB4, partial cds.	17.8	19.9	0	22.1	8
U11791_at	Human cyclin H mRNA, complete cds.	115.5	194.4	0	148.5	98.5
U11090_at	Human hydroxyindole-O-methyltransferase promoter B-derived (HIOMT) mRNA, complete cds.	373.8	680.8	0	547.9	86.9
RC_T85248_at	Incyte Unique	52.2	96.4	0	58.8	7.1
RC_W69134_at	Human full length insert cDNA clone ZD44G02.	264.5	338.1	0	216.2	181.8
RC_AA443330_at	Human mRNA for T200 leukocyte common antigen (CD45, LC-A).	312.3	336.4	35.6	164.5	71
RC_AA443321_at	Incyte Unique	283	487.7	168.8	324.5	159.2
RC_T85314_at	Incyte Unique	455	565.1	243.2	582.4	247.9
RC_AA443271_at	Human mRNA for KIAA0546 protein, partial cds.	256.8	339.8	98.8	275.7	210.1
U51432_at	Human nuclear receptor coactivator NCoA-62 mRNA, complete cds.	190.4	211	0	181.3	145.2
U51240_at	Human lysosomal-associated multitransmembrane protein (LAPTm5) mRNA, complete cds.	494.4	634.7	0	304.6	132.6
RC_AA132969_s.	Human mRNA for KIAA1104 protein, complete cds.	196.8	279.2	119.5	230.4	110.2
U10439_at	Human IFI-4 mRNA for type II protein.	795.2	1014.1	259.6	858.3	438.9
RC_AA256996_at	Human mRNA; cDNA DKFZp434D222 (from clone DKFZp434D222); partial cds.	73.6	109.3	31.2	124.3	30
R78119_at	Incyte Unique	71.6	106.6	28.5	74.6	23.3
RC_AA018346_at	Human CGI-15 protein mRNA, complete cds.	234.2	357.5	126.5	368.1	3.3
RC_H87608_f_at	Human partial 5-HT4 receptor gene, exons 2 to 5.	25.7	49.1	0	37	17.5
RC_AA262499_at	Human mRNA for KIAA0748 protein, complete cds.	21.9	32.7	0	28.9	9.1

RC_W65344_at	Human mRNA; cDNA DKFZp586A181 (from clone DKFZp586A181); partial cds.	77.2	136.1	0	61.6	0
R82229_at	Human mRNA for KIAA0542 protein, partial cds.	150.6	123.6	0	153.2	59.4
RC_AA452411_at	Human vitamin D3 receptor interacting protein (DRIP130) mRNA, complete cds.	106.9	132.4	43.4	71.6	40.1
U13737_at	Human cysteine protease CPP32 isoform alpha mRNA, complete cds.	90.9	111.2	0	74.6	18.8
RC_AA443958_at	Human MUC18 gene exon 16.	447.7	660.4	182	270.6	0
RC_AA452265_at	Human mRNA; cDNA DKFZp434D0935 (from clone DKFZp434D0935).	102.2	181.4	102.4	153.5	44.6
U12779_at	Human mRNA for MAP kinase activated protein kinase.	249.1	459.4	0	240.8	0
RC_N29484_at	Incyte Unique	291.7	577	0	296.8	0
RC_H89561_at	Human ZNF202 beta (ZNF202) mRNA, complete cds.	9.6	14.6	0	5	4.8
RC_AA262179_at	Human mRNA; cDNA DKFZp434M2023 (from clone DKFZp434M2023); partial cds.	712.7	819.8	320.5	563.3	456
RC_AA149051_at	Human mRNA; cDNA DKFZp586G0321 (from clone DKFZp586G0321).	57.9	119.8	23.8	74.5	60.5
RC_AA443671_at	Human transcriptional coactivator PC4 mRNA, complete cds.	70.7	107.2	40	103.5	26.3
RC_AA262351_f_1	Human mRNA expressed in placenta, 3'UTR.	255.8	175.7	128.4	193.3	0
U49835_s_at	Human chitinase precursor (HUMTCHIT) mRNA, exon 1a form, complete cds	354.4	456.7	0	325.6	129.9
RC_AA262470_at	Human FSHD-associated repeat DNA, proximal region.	84.9	182.8	43.8	123	74.1
RC_AA017410_at	Incyte Unique	369.6	425.1	128.1	234.2	220.3
R80048_at	Incyte Unique	167.4	256.7	59.9	203	122.6
R80332_at	Incyte Unique	177.1	255.1	130.9	283.8	91.4
U49395_at	Human ATP receptor subunit (P2X5) mRNA, complete cds.	435.6	463.4	53	321.4	96.9
RC_AA262276_at	Incyte Unique	279.4	250.6	142.8	184.9	41.2
U42031_at	Human 54 kDa progesterone receptor-associated immunophilin FKBP54 mRNA partial cds.	232.7	149.3	0	260.1	64.9
RC_AA001507_at	Incyte Unique	143.5	133.7	86.8	163.8	16.3
RC_AA453022_at	Human NADH-cytochrome b5 reductase isoform mRNA, complete cds.	209.9	410	227.2	269.3	41.2
RC_W38203_at	#N/A	121.9	147.9	0	110.9	51.2
RC_AA258572_at	Human mRNA for seven transmembrane domain orphan receptor, complete cds	721	649.9	195.8	572.3	77
U28015_at	Human cysteine protease (ICERel-III) mRNA, complete cds.	229.6	178.7	0	182.8	15.3
RC_AA136884_at	Incyte Unique	46	59.2	28.1	58.7	23
RC_AA135748_at	Incyte Unique	19.3	13.1	0	26.2	0
RC_AA449361_at	Human RING zinc finger protein (RZF) mRNA, complete cds.	120.8	156.7	47.7	172.8	36.5
U28811_at	Human Golgi membrane sialoglycoprotein MG160 (GLG1) mRNA, complete cds.	883.5	639.5	159.8	753.6	367.9
RC_AA447791_at	Human putative RNA-binding protein Q99 mRNA, complete cds.	740.8	843.5	309.1	674.7	402.7
RC_W15219_at	Human FcERI gamma-chain gene sequence.	59.4	67.7	0	58.1	8.7
RC_AA449419_at	Incyte Unique	30.8	83.6	27.2	59.6	29.7
RC_AA449238_s_1	Incyte Unique	91.8	190.6	82.9	160.9	18.2
RC_AA449458_at	Human integrin B-6 mRNA, complete cds.	365.4	360.8	79.8	396.9	172.1
RC_AA449469_at	Incyte Unique	76.1	102.8	0	77.7	20.2
RC_AA258638_at	Public Unique	88.3	151.6	0	113.2	86.4
RC_AA447726_at	Human androgen receptor gene, partial exon.	100.3	104.4	55.7	61.6	18.8
RC_H98835_at	Incyte Unique	154.8	237.8	129.8	315.5	0
U26710_at	Human cbl-b truncated form 1 lacking leucine zipper mRNA, complete cds.	225.9	318.6	0	224.1	62.7
U26648_at	Human syntaxin 5 mRNA, complete cds.	253.4	531.6	136.2	395.4	168.9
RC_W42674_at	FK506-binding protein FKB23 isoform	901.2	1367.2	0	717.9	643.5
U33936_s_at	Human adenosine kinase mRNA, complete cds.	107.9	130.6	37	104.6	44.5
RC_AA449197_at	Incyte Unique	89	110.4	33.9	87.2	14.4
U24704_at	Human 26S protease subunit S5a mRNA, complete cds.	95.3	160.2	46	146.1	50.9
U31556_at	Human transcription factor E2F-5 mRNA, complete cds.	122.3	249.8	31.4	213.2	47.9
RC_AA007160_at	Human mRNA; cDNA DKFZp564D016 (from clone DKFZp564D016).	360.1	752.2	100.4	516.3	208.5
RC_N20099_i_1	Human mRNA; cDNA DKFZp586A191 (from clone DKFZp586A191).	94.5	126.8	25.6	46.1	59.4
RC_AA448850_at	Human mRNA; cDNA DKFZp564B2463 (from clone DKFZp564B2463).	151.4	319.8	112.8	199.6	84.6
RC_AA448679_at	Incyte Unique	19.1	15.3	0	31	0
RC_W31352_at	Incyte Unique	349.4	653.6	220.2	648.5	273.3
RC_W32470_at	Incyte Unique	87.3	88.2	0	88.5	0
U31384_at	Human G-protein gamma-11 subunit mRNA, complete cds.	190.7	322.6	84.2	284.5	162.4
RC_W31356_at	Human HSPC160 mRNA, complete cds.	965.1	1530.3	477.1	1237.8	869
RC_AA004900_at	Human mRNA; cDNA DKFZp586A0422 (from clone DKFZp586A0422).	150.9	337.1	70.5	257.3	92.4

AA280200_at	Incyte Unique	103	88.5	0	175.5	49
AA262132_at	Human mRNA; cDNA DKFZp586F1318 (from clone DKFZp586F1318); parti cds.	59.1	151.9	41.7	108.8	20.9
RC_AA009489_at	Human CGI-74 protein mRNA, complete cds.	145.6	210.7	134	192.9	34
AA259102_at	Human Ste-20 related kinase SPAK mRNA, complete cds.	118.4	276.4	125.5	226	93.2
U29953_ma1_at	Human pigment epithelium-derived factor gene, complete cds.	723.5	1008.6	100.3	361.4	247.7
U33818_at	Human polyadenylate binding protein mRNA, complete cds.	293.2	433.3	110.4	281.1	235.7
RC_AA136079_at	Incyte Unique	947.3	1162.9	323.1	963.2	324.8
U30246_at	Human bumetanide-sensitive Na-K-Cl cotransporter (NKCC1) mRNA, comple cds.	195.9	332.5	52.5	166.2	100.6
RC_W20487_s_at	Human mRNA full length insert cDNA clone EUROIMAGE 327506.	936.9	1013.3	241.2	734.6	488.5
RC_H98527_at	Incyte Unique	32.7	35.1	0	16.1	0
RC_W02102_at	Incyte Unique	132.7	274	67.6	149	102.1
RC_H97940_at	Human alpha 1,2-mannosidase IB mRNA, complete cds.	245.9	184	0	165.2	0
RC_H94648_at	Incyte Unique	413.6	527.6	148.3	241.3	278.4
RC_W49487_at	Incyte Unique	81.4	108.4	0	79.9	64.7
RC_AA011414_s_	Human fibrinogen alpha subunit and fibrinogen alpha subunit precursor, genes complete cds.	16.9	46.3	0	33.8	3.3
RC_AA147439_s_	Incyte Unique	110.5	130.7	0	151.7	66.6
RC_N23167_r_at	Human MAFB/Kreisler basic region/leucine zipper transcription factor (MAFE mRNA, complete cds.	356.3	309.3	98.1	319	173.6
RC_N23174_at	Human mRNA for SLC7A8 protein.	776.2	792.9	338.1	1337.3	192
RC_N24614_at	Incyte Unique	72.6	178.7	62.3	152.3	36.9
U20158_at	Human 76 kDa tyrosine phosphoprotein SLP-76 mRNA, complete cds.	63.5	164.2	20.6	93.7	24.7
RC_AA450047_at	Human mRNA for breast cancer associated protein BRAP1, complete cds.	114.1	94	56.5	105.6	0
AA252436_at	Human clone 23753 mRNA sequence.	163	438.7	80.8	288	175.1
RC_AA011654_at	Incyte Unique	154.6	267.8	164.2	243.5	86.6
RC_AA258138_at	Incyte Unique	71.7	73.1	33.7	79.6	34.3
RC_W55858_at	Human ZNF202 beta (ZNF202) mRNA, complete cds.	117	139.6	52.7	174.9	60.3
U18932_at	Human heparan sulfate-N-deacetylase/N-sulfotransferase mRNA, clone HSST: 3'UTR.	20.1	44.2	0	31.3	22.2
RC_AA001607_at	Incyte Unique	65.1	70.7	0	110.8	3.3
U40990_at	Human kidney and cardiac voltage dependent K+ channel (KvLQT1) mRNA, complete cds.	158.7	346.1	0	154	138.5
RC_AA258032_at	Human endosome-associated protein (EEA1) mRNA, complete cds.	37.6	63.4	38.5	73.2	12.5
U19713_s_at	Human full length insert cDNA clone ZD18G05.	113.2	273.3	0	172.5	74.1
AA250870_s_at	Human mRNA; cDNA DKFZp434P1650 (from clone DKFZp434P1650); parti cds.	269.6	502.8	96.4	314.3	76
RC_N24772_at	Human putative transmembrane protein (CLN5) mRNA, complete cds.	291.3	280.5	92.9	360.5	127.2
RC_T96965_at	Incyte Unique	18.4	26.5	0	10.5	0
RC_T97307_at	Incyte Unique	265.9	442.5	73.5	403.8	204.9
RC_AA447617_at	Incyte Unique	419.1	573.6	194.8	298.1	222.6
RC_AA449756_at	Human hereditary haemochromatosis region, histone 2A-like protein gene, hereditary haemochromatosis (HLA-H) gene, RoRet gene, and sodium phospho: transporter (NPT3) gene, complete cds.	84.7	152.4	31.7	118.5	56.3
RC_AA146979_at	Incyte Unique	172.1	107.8	0	156.6	0
U37546_s_at	Human TNFR2-TRAF signalling complex protein mRNA, complete cds.	664.9	829.7	485.9	838.5	96.3
RC_H97809_at	Incyte Unique	362.3	474.5	118.3	536	246.3
RC_AA002007_at	Human genomic DNA, chromosome 6p21.3, HLA Class I region, section 15/20	67.6	105.6	32.3	110	53
RC_N22332_at	Incyte Unique	68.2	15.9	0	73.2	0
RC_T99385_f_at	Human genomic DNA, chromosome 6p21.3, HLA Class I region, section 19/20	29	21.9	0	23.5	0
U23736_s_at	Human mRNA for zinc-finger DNA-binding protein, complete cds.	0	0	0	0	0
AA255638_at	Human thioredoxin-like protein mRNA, complete cds.	621.9	565.6	148.7	426.9	226.5
U22055_at	Human 100 kDa coactivator mRNA, complete cds.	1242.5	1583.5	378.5	1260.7	836.2
RC_AA447110_at	Human CHD2 mRNA, complete cds.	72.6	152.9	37.6	150	9.4
U21128_at	Human lumican mRNA, complete cds.	753.4	1287.6	298	567.4	423.3
RC_AA449951_at	Human topoisomerase-related function protein (TRF4-1) mRNA, partial cds.	241.2	341.9	107.6	315.7	152
RC_AA447142_at	Incyte Unique	68.7	84.9	0	81.9	17
U38964_s_at	Human PMS2 related (hPMSR2) gene, complete cds.	103.9	179.2	61.7	139.4	54.7
RC_AA447155_at	Human mRNA; cDNA DKFZp586F1918 (from clone DKFZp586F1918); parti cds.	368.1	482.1	156.1	545.3	189.2
RC_H96908_at	Incyte Unique	131.2	145.9	53.9	144.3	61.4

RC_H96712_at	Incyte Unique	433.6	406.2	115	395	152.5
RC_AA449818_s_	Human modulator recognition factor I (MRF-1) mRNA, 3' end.	110.1	103.4	62.2	145.3	24.6
RC_W72231_at	Incyte Unique	128.4	215	112.5	167.9	65
RC_N30728_at	Incyte Unique	155.5	221.1	0	112.2	69.5
RC_AA026417_at	Human PIL protein mRNA, complete cds.	132.2	251.9	83.9	217.8	95.7
RC_AA458919_at	Human signalosome subunit 2 (SGN2) mRNA, complete cds.	224.1	283.5	110.7	236.8	38.5
RC_AA025086_at	Incyte Unique	333.8	701.1	367.6	667	202.4
AA355201_at	Human mRNA for SOX-4 protein.	154.5	209.7	59.6	77.3	88
M97935_s_at	Human transcription factor ISGF-3 mRNA, complete cds.	461.3	1186.1	251.2	544.1	281.6
RC_H55915_at	Incyte Unique	326.5	605.8	432.7	554.4	14.6
RC_AA279024_at	Human mRNA for triple LIM domain protein.	110.6	142.5	0	188.9	52.8
AA349630_at	Incyte Unique	474.7	692.7	0	703.3	0
R09286_at	Human LIM protein mRNA, complete cds.	77.4	100.3	0	38.3	42.9
M96954_s_at	Human nucleolysin TIAR mRNA, complete cds.	155.1	314.6	64.8	266.7	134.8
RC_N38930_at	Incyte Unique	530.4	574.5	208.6	683.3	308
RC_AA255601_at	Incyte Unique	9.6	15.3	0	18.5	0
AA373890_at	Human splicing factor SRp30c gene, exons 3 and 4, and complete cds.	212.6	275.6	154.2	294.4	93.1
RC_T34611_at	Human HPTP epsilon mRNA for protein tyrosine phosphatase epsilon.	149.7	180.1	96.1	224.5	92.7
AA227366_at	Human mRNA; cDNA DKFZp564C246 (from clone DKFZp564C246); complete cds.	147.2	206.3	23.4	151.1	69
N95507_at	Incyte Unique	202.2	171.1	34.3	126.9	0
AA372630_s_at	Human GW112 protein (GW112) mRNA, complete cds.	4504.2	7807.2	2962.2	5058.2	1260.8
U70321_at	Human herpesvirus entry mediator mRNA, complete cds.	715.1	886.3	282.5	825.3	277.1
U70660_at	Human copper transport protein HAH1 (HAH1) mRNA, complete cds.	572	987.8	159.2	954.3	517.3
RC_AA279060_at	Human CARD-containing apoptotic signaling protein (BCL10) mRNA, complete cds.	334.8	393.8	65.7	220.5	171.6
RC_AA433947_at	Human gamma-glutamyl hydrolase gene, exons 1 and 2.	276.6	383.8	225	423.6	114.2
S49592_s_at	Human (E2F-1) pRB-binding protein mRNA, complete cds.	222.4	322.7	0	390.3	109.5
U71203_s_at	Human mRNA for RIT protein.	20.1	14.3	0	25.7	0
RC_AA456598_at	Human CGI-149 protein mRNA, complete cds.	122.4	158.9	48.5	188	73.4
RC_AA435633_at	Human clone 23965 mRNA sequence.	324.6	308.5	68.7	451.2	165.5
U66052_at	Human clone Z'3-1 placenta expressed mRNA from chromosome X.	96.7	136.2	0	111.2	83
RC_AA156230_at	Human mRNA; cDNA DKFZp434I1817 (from clone DKFZp434I1817).	279.2	306.6	124.1	239	140
RC_AA456589_at	Human mRNA; cDNA DKFZp761N2124 (from clone DKFZp761N2124).	177	335.3	153.3	318.7	54.2
AA329211_s_at	Human NSAP1 protein (NSAP1) mRNA, complete cds.	83.6	193.9	47	150.3	59.2
AA234791_at	Incyte Unique	255.8	329.7	172	165.3	64.2
RC_T51620_at	Incyte Unique	21.6	44.3	0	22.4	0
RC_AA255886_at	Human clone YDD19 mRNA sequence.	95.9	112.8	27.9	58.2	31.6
RC_T49030_s_at	Novel Human gene mapping to chromosome X.	158.5	234	159.7	241.1	60.5
RC_AA129968_at	Similar to phosphoprotein phosphatase 2A regulatory subunit	107	127.3	25.6	122.4	17.9
RC_AA278845_at	Human mRNA for KIAA0871 protein, complete cds.	91.9	89.2	0	164.3	40.8
S53911_at	CD34=glycoprotein expressed in lymphohematopoietic progenitor cells {alternatively spliced, truncated form} [Human, UT7, mRNA, 2657 nt].	217.1	453.8	0	201	83.6
R15268_at	Human mRNA for KIAA1255 protein, partial cds.	107	207.5	32.4	118.9	67.1
RC_AA130349_at	Incyte Unique	734.7	645	163.2	650.6	204.1
U67369_at	Human growth factor independence-1 (Gfi-1) mRNA, complete cds.	78.4	125.3	0	133	18.5
RC_AA156247_at	Human mRNA; cDNA DKFZp564C203 (from clone DKFZp564C203).	200.1	262.7	0	232	0
S57132_s_at	Human alpha-1 type XVI collagen (COL16A1) mRNA, complete cds.	239.4	551.7	172.2	454.5	186.5
RC_N36001_at	Human bystin mRNA, complete cds.	1109.4	1202.3	291.6	1079.2	671.5
RC_W93382_at	Human 959 kb contig between AML1 and CBR1 on chromosome 21q22, segment 2/3.	421.6	440.9	95.7	352.4	275.2
RC_AA279162_s_	Human heterogeneous nuclear ribonucleoprotein R mRNA, complete cds.	375.4	649.7	274.5	548.9	277.8
RC_AA025534_at	Human mRNA; cDNA DKFZp434H0717 (from clone DKFZp434H0717); partial cds.	128.3	206.6	61.9	151.5	113.8
U64197_at	Human chemokine exodus-1 mRNA, complete cds.	117.1	175.2	17.8	127.5	19.4
U75362_at	Human isopeptidase T-3 (ISOT-3) mRNA, complete cds.	89.7	133.4	0	85.1	75.9
RC_AA431454_at	Incyte Unique	82.8	181	38	172.7	13.8
M85276_at	Human NKG5 gene, complete cds.	72.9	167.1	0	93.1	0
RC_AA157568_at	Human CpG island DNA genomic MseI fragment, clone 87h7, forward read cpg87h7.ft1a.	41.5	87.7	32.7	53.6	0

M86667_at	Human NAP (nucleosome assembly protein) mRNA, complete cds.	228	607	63.4	328.8	189.3
N75274_at	Incyte Unique	34	45.7	0	10	0
RC_AA460350_at	Human mRNA; cDNA DKFZp564O0122 (from clone DKFZp564O0122).	47.8	119.5	38.6	67.4	23.5
RC_AA025905_f_	Human TNF-inducible protein CG12-1 mRNA, complete cds.	1062.3	1267	633.3	850.7	519.5
RC_AA431448_at	Incyte Unique	285.3	367.3	126	385.8	198.3
RC_AA460532_at	Human mRNA for ankyrin repeat protein, complete cds.	60.9	86.1	39.6	90	20.7
RC_AA431482_s_	Human mRNA for actin binding protein ABP620, complete cds.	112.3	269.5	112.7	199.7	63.5
RC_H41244_at	Human chromosome 16 open reading frame 5 (C16orf5) mRNA, complete cds	1527.1	2692.3	0	1727.2	299.3
RC_H40509_at	Incyte Unique	68.2	53.6	0	71.9	0
AA401047_at	Human serine protease ovasin mRNA, complete cds.	24.1	28.5	0	46.8	0
RC_Z38299_at	Human DNA cytosine methyltransferase 3 alpha (DNMT3A) mRNA, complete cds.	66.2	143.6	39.4	117.1	48
RC_AA431426_at	Human chromosome 4q35 subtelomeric sequence.	91.3	178.6	44.8	176.4	56.8
RC_AA253473_at	Human DD96 mRNA, complete cds.	2135.1	2464	934.9	1981.2	281.3
RC_AA128654_at	Incyte Unique	106.7	172.8	0	193.5	47.7
RC_H46018_at	Human clone CCA11 locus D20S101 mRNA containing CCA trinucleotide repeat.	1210.9	2006.3	0	1596.2	824.5
RC_AA431479_at	Human mRNA; cDNA DKFZp586I0521 (from clone DKFZp586I0521).	46	95.5	0	54.4	19.2
AA216017_at	Incyte Unique	322.7	350.5	63.7	320	27.1
RC_T33260_at	Human DiGeorge syndrome critical region, centromeric end.	57.6	90.8	12.5	33.8	14.5
M92642_at	Human alpha-1 type XVI collagen (COL16A1) mRNA, complete cds.	314.8	993.8	0	692.4	0
RC_H51050_at	Incyte Unique	937.4	589.5	0	1128.5	424.9
RC_AA459673_at	Human mRNA; cDNA DKFZp434F205 (from clone DKFZp434F205); complete cds.	59.2	66.6	0	54.2	46.1
RC_T32438_at	Incyte Unique	99.5	153.9	0	116.9	0
RC_AA255546_at	Incyte Unique	226.9	456.6	121	395	163.6
AA216562_at	Human mRNA for KIAA0885 protein, complete cds.	61.8	145.4	0	80.2	30.1
RC_AA459293_at	Human mRNA for Hmob33 protein, 3' untranslated region.	91.1	163.7	39.4	147.6	24
N90820_at	Human mRNA; cDNA DKFZp434M1115 (from clone DKFZp434M1115); partial cds.	424.8	382.5	185.8	468.7	83.7
RC_W94774_s_at	Human mRNA for KIAA0226 protein, partial cds.	409.3	734.6	250.7	475.6	159.6
AA216256_at	D10S102=FBRNP [Human, fetal brain, mRNA, 3043 nt].	32.7	66.3	18.6	61.1	9.4
RC_W96222_at	Human FYN binding protein mRNA, complete cds.	205.6	144.8	43.9	195.1	20.5
RC_T25732_f_at	Human mRNA for KIAA0252 gene, partial cds.	319.2	390.5	0	444.7	129.7
RC_AA460343_at	Incyte Unique	133.3	200.1	44.6	189.7	97.9
RC_H49252_at	Incyte Unique	100.1	75.8	38.9	124.7	0
RC_AA460243_at	Human CpG island DNA genomic MseI fragment, clone 21g11, forward read cpg21g11.ft1b.	78.1	84.8	0	102.8	52.4
RC_N41059_at	Incyte Unique	501.8	676.9	178.9	514	384.1
RC_AA431571_at	Incyte Unique	819.4	808.4	534	1063.5	49.8
U73377_at	Human p66shc (SHC) mRNA, complete cds.	590.9	917.3	120.4	925.1	482.6
RC_W95041_at	Incyte Unique	43.5	85.4	24	52.7	17.2
AA234817_at	Incyte Unique	158.6	170.1	92.5	177.4	0
RC_AA456584_at	clone 4-3 {Alu sequences, splice acceptor sites} [Human, Pre-mRNA, 545 nt].	337.9	446.5	142.9	510	88.3
RC_T79615_at	Incyte Unique	346.4	176.1	0	411	111.3
RC_T67105_s_at	Incyte Unique	334.7	395	77.3	293.7	148.8
U57316_at	Human GCN5 (hGCN5) gene, complete cds.	119.1	187.2	78	102.9	26.7
U06088_at	Human N-acetylgalactosamine 6-sulphatase (GALNS) gene, exon 14.	19.1	55.5	0	41.1	5.7
U07158_at	Human syntaxin mRNA, complete cds.	273.7	661.2	59.9	411.9	274.4
RC_T67161_at	Incyte Unique	97.8	105.8	0	138.7	27.6
AA236843_s_at	Human (D8S135) DNA segment containing GT repeat.	185.5	290.3	117.6	231.1	90
R64459_at	Human MRC OX-2 gene, V-like region.	124.5	226.2	45	63.6	39
RC_AA256672_at	Incyte Unique	18.6	27.3	16.8	35.1	0
U05875_at	Human clone pSK1 interferon gamma receptor accessory factor-1 (AF-1) mRNA, complete cds.	385.4	451.5	117.1	533.9	194.7
R60117_at	Human mRNA; cDNA DKFZp586F1019 (from clone DKFZp586F1019); partial cds.	139.4	106	89.5	155.6	0
RC_AA256680_at	Human mRNA; cDNA DKFZp564H1916 (from clone DKFZp564H1916).	67.5	150.3	19.8	108.6	38.9
RC_AA436705_at	Human mRNA for KIAA0766 protein, complete cds.	126.1	165.3	67.6	206	35.4
RC_W80467_at	Incyte Unique	777.5	733.8	213.1	693.2	395.3
RC_AA454840_s_	Incyte Unique	82.1	232.2	77.5	166.8	82.5

RC_T65802_at	Human lymphocyte antigen mRNA, 3' flank.	666.9	1035.3	370.2	468.6	344.6
RC_H69021_f_at	Human GT mitochondrial solute carrier protein homologue mRNA, complete cds.	38.9	50.3	22.4	75.7	0
RC_H70047_at	Human regulator of G-protein signaling (RGS13) mRNA, complete cds.	55.9	65.8	0	67.4	15.6
RC_AA437106_at	Human ORF (CEI5) mRNA, 3' flank.	728.1	919	265.6	674.9	217.7
RC_AA151570_at	Incyte Unique	205.3	284.9	76.6	176.5	118.5
RC_AA150776_at	Human clone 24405 mRNA sequence.	313.1	422.8	0	270.5	181.6
U08006_s_at	Human complement protein C8 alpha subunit mRNA, complete cds.	4	6.6	0	5	0
RC_W80666_at	Incyte Unique	124.5	181.1	79.3	148	90.6
R71427_at	Human putative phenylalanyl-tRNA synthetase beta-subunit mRNA, complete cds.	64.5	109.6	26.6	66.3	50.3
RC_T79020_at	Incyte Unique	117.9	40.5	0	119.6	0
RC_W72455_at	Incyte Unique	675.4	795.8	166.1	359.4	405.1
RC_N30868_at	Human mRNA for integrin alpha-4 subunit.	131.4	198.4	0	145.7	102.7
RC_H82527_at	Human antigen NY-CO-33 (NY-CO-33) mRNA, complete cds.	510.8	473.5	0	469.9	235.9
RC_H80860_at	Incyte Unique	150.1	192.4	0	95.5	84.6
R73170_at	Human mRNA for synaptosome associated protein of 23 kilodaltons, isoform /	58.5	93.8	0	51	20.8
RC_N30824_at	Human heparanase (HPA) mRNA, complete cds.	177.2	142	48.3	165	42.5
RC_AA442142_at	Incyte Unique	61.2	103	32.8	101.5	25.9
R69673_at	Incyte Unique	102.7	113.2	0	56.2	0
U08021_at	Human nicotinamide N-methyltransferase gene, exon 1 and 5' flanking region.	627.1	720.8	132.5	291.6	117.9
RC_AA441802_at	Human prenylcysteine carboxyl methyltransferase (PCCMT) mRNA, complete cds.	112.1	245.5	0	148.9	72.9
U53786_at	Human envoplakin (EVPL) mRNA, complete cds.	318	368.5	0	311.9	242.6
RC_AA453619_at	Incyte Unique	139.8	171.4	138.8	242.6	0
RC_H78370_at	Incyte Unique	62.9	33.6	0	86.5	0
R67751_at	Human UMP-CMP kinase mRNA, complete cds.	151.1	239.6	33.2	237.1	63.9
RC_W72748_at	Human guanylate binding protein isoform II (GBP-2) mRNA, complete cds.	158.9	313.7	0	200.5	98.4
RC_N32019_at	Human choline/ethanolaminophosphotransferase (CEPT1) mRNA, complete c	96.7	119.5	51.7	136.5	46.5
RC_AA441939_at	Human Ig lambda gene locus DNA, clone:80A10.	54.6	124.2	51.2	96	21
RC_T64465_at	Public Unique	11.7	15.8	0	10.6	0
RC_AA262999_at	Incyte Unique	82	72	0	50.6	14
RC_AA256380_at	Human transformer-2 alpha (htra-2 alpha) mRNA, complete cds.	14.6	4	0	14.5	0
RC_W87710_s_at	Human mRNA, cDNA DKFZp564F093 (from clone DKFZp564F093).	98.8	123.1	31.7	79.5	74.1
U62317_ma3_at	Human platelet-derived endothelial cell growth factor gene, exons 1 through 1(	692.3	1191.2	727.6	1215.8	344.9
R25326_at	Human mRNA, cDNA DKFZp564O0582 (from clone DKFZp564O0582); partial cds.	119.9	195.9	47.2	148.8	80.2
RC_AA435896_at	Human CGI-78 protein mRNA, complete cds.	23.4	34.7	0	42.8	9.4
RC_W86221_at	Human mRNA for KIAA0737 protein, complete cds.	218.1	327.7	70.2	401	129.9
R26213_at	Human TWIK-related acid-sensitive K <sup>+</sup> channel (TASK) mRNA, complete cds	131.9	325.8	158.8	250.3	0
RC_N34804_at	Human mRNA, cDNA DKFZp434J214 (from clone DKFZp434J214); partial cds.	362.7	454.2	67.7	418.4	238
RC_T56048_at	Incyte Unique	67.7	172.3	0	135.5	47.5
RC_AA256208_at	Human mRNA for KIAA0891 protein, partial cds.	226.3	436.4	166	395.8	96.8
RC_AA021590_at	Incyte Unique	138.7	123.6	0	94.6	76.7
RC_N34871_at	Human mRNA, cDNA DKFZp566G0224 (from clone DKFZp566G0224); partial cds.	189	584.7	74.2	378.6	146.2
AA315935_at	Human unknown mRNA.	182.8	312.8	50.2	265.9	113
R20168_at	Human plasma membrane calcium ATPase isoform 1 (ATP2B1) gene, alternative splice products, partial cds.	22.3	30.5	0	22.8	0
RC_T51933_at	Incyte Unique	361.6	432.5	124.9	257.7	239.8
RC_AA435748_at	Incyte Unique	359.9	451.2	90.7	327.2	146.8
U62962_at	Human Int-6 mRNA, complete cds.	655.7	1546.9	251.3	1088.5	713.1
RC_AA456325_at	Human TDE homolog mRNA, complete cds.	264.9	383.4	178.1	372.1	97.6
RC_AA435840_at	Human mRNA for high mobility group protein HMG2a.	67.2	84.7	46.4	116.6	13.6
RC_W87415_at	Human genomic DNA, chromosome 6p21.3, HLA Class I region, section 5/20.	53.8	95	0	31.3	16.4
R20358_at	Human mRNA for proteasome subunit p58, complete cds.	206.7	458	217.5	389.9	159.5
RC_T57317_at	Incyte Unique	32.2	74.9	0	47.6	0
S79639_at	EXT1=putative tumour suppressor/hereditary multiple exostoses candidate gen [Human, placenta, mRNA, 3183 nt].	352.5	412.4	111.5	309.5	212.6
RC_AA436613_at	Human mRNA, cDNA DKFZp564I1922 (from clone DKFZp564I1922); partia cds.	76.9	107.1	0	28.6	38.4

RC_AA151712_at	Human HEB helix-loop-helix protein (HEB) mRNA, complete cds.	442	323.6	113.8	382.3	191
RC_AA256323_at	Human mRNA, cDNA DKFZp434N126 (from clone DKFZp434N126).	214.7	286.7	122.8	288.2	106.9
RC_AA256326_at	Incyte Unique	139.4	154.3	83.9	144.1	61.5
RC_AA436570_at	Human mRNA for pre-mRNA cleavage factor I subunit.	217.5	276.1	47.9	196.8	103.9
RC_AA256376_s_	Human metaxin 2 (MTX2) mRNA, nuclear gene encoding mitochondrial prote: complete cds.	62.7	103.3	51.5	105.4	25.7
RC_N33920_at	Human mRNA for diubiquitin.	263.2	445.9	0	312.6	0
RC_T63857_at	Human Ras-like GTP-binding protein (RAB27A) gene, exons 1b and 2.	248.3	202.6	74.3	227.1	126.9
RC_W80750_at	Incyte Unique	227.1	466.8	0	391.6	94.9
RC_H65790_s_at	Incyte Unique	324.5	618.4	0	410.2	242.9
RC_AA455864_at	Incyte Unique	280.6	226.3	0	331.5	0
RC_AA152408_at	Human KIAA0433 mRNA, partial cds.	70	117.8	45.5	105.8	19.1
RC_AA436149_at	Incyte Unique	55.9	76.5	0	19.8	11.7
RC_AA435920_s_	orf1 5' to PD-ECGF/TP...orf2 5' to PD-ECGF/TP [Human, epidermoid carcinoma cell line A431, mRNA, 3 genes, 1718 nt].	86.1	94.1	0	49.2	0
RC_AA435999_at	Human CGI-83 protein mRNA, complete cds.	110.7	209.2	54.4	110	54.6
RC_AA455970_at	Human multiple membrane spanning receptor TRC8 (TRC8) mRNA, complete cds.	151.7	179.3	82	259.5	35.7
RC_AA436420_at	Incyte Unique	160.7	60.6	0	197.4	0
RC_AA455938_at	Incyte Unique	95.4	138.6	0	100	87.7
RC_AA278329_f_	Incyte Unique	402.8	681.5	111.7	665	337.1
RC_T58885_at	Human KpnI fragment upstream of apolipoprotein(a) gene.	41.8	97	18.5	59.2	0
RC_AA478596_at	Human mRNA for KIAA0854 protein, complete cds.	147.6	342.8	95	260.7	40.1
U46116_at	Human receptor-type protein tyrosine phosphatase gamma (PTPRG) mRNA, complete cds.	49.3	26.4	0	58.2	0
L11015_s_at	Human lymphotoxin-beta mRNA, complete cds.	253.5	433.8	0	137.9	0
RC_AA055197_at	Human gamma-glutamyl hydrolase gene, exons 8 and 9 and complete cds.	84.9	158.4	35.6	119.8	78.4
AC002486_at	Simian mRNA for ral protein.	76.4	78.6	0	63.7	15
AC002477_s_at	Human ZNF183 gene.	93.8	139.9	0	124.3	0
RC_AA598710_at	Incyte Unique	76.3	135.5	69.7	119.3	21.9
RC_N67324_at	Incyte Unique	95.4	110.9	0	64.8	16.1
RC_N67394_at	Incyte Unique	275.2	291.7	51.5	199.4	178.5
RC_AA088851_s_	Human S-adenosylmethionine decarboxylase (AMD1) gene, exons 5-9.	71.6	99	30.9	100.5	32.9
RC_AA405004_at	Human mRNA, cDNA DKFZp564H092 (from clone DKFZp564H092).	203.8	268.6	79.2	370	38
RC_AA598831_f_	Incyte Unique	118.7	205.2	77.3	142.3	93.4
D59253_at	Human mRNA, cDNA DKFZp586F0221 (from clone DKFZp586F0221).	60.6	73.2	24	39.2	36.1
RC_AA213410_at	Human mRNA for ADP ribosylation factor-like protein, complete cds.	106.4	61.1	14.3	101.6	15.6
J05070_at	Human type IV collagenase mRNA, complete cds.	559.3	676.3	0	232.2	87.6
RC_AA214369_at	Human mRNA for KIAA0037 gene, complete cds.	41.1	40.9	15.8	33.2	13.1
AA090257_at	Human SOD-2 gene for manganese superoxide dismutase.	625.3	959.3	569.9	869.3	268.5
RC_AA235295_at	Human mRNA, cDNA DKFZp761H171 (from clone DKFZp761H171); partial cds.	24.3	24.7	0	39.7	0
RC_AA404427_at	Human mRNA for KIAA1249 protein, partial cds.	134.2	233.1	47.3	183.5	27.4
J04080_at	Human inducible 6-phosphofructo-2-kinase/fructose 2,6-bisphosphatase (PFK 2) mRNA, complete cds.	2093	2837	1192.9	1873.2	1332.1
RC_AA404564_at	Incyte Unique	162.5	220.5	0	204.5	116.1
AF005043_at	Human poly(ADP-ribose) glycohydrolase (hPARG) mRNA, complete cds.	51.6	97	0	53.6	11
D58604_s_at	Human mRNA for KIAA0573 protein, partial cds.	334.3	558.7	212.6	550.3	182.2
RC_AA235553_at	Incyte Unique	381.2	632.7	400.4	744.2	56.4
RC_R32361_at	Human chromosome 1 Ig V(K)I gene, part. with 5' breakpoint between orphon and neighbouring non-amplified region.	96.2	66.8	0	94.9	0
RC_AA404342_at	Incyte Unique	159.9	106.6	47.2	148.7	40.2
RC_R36109_at	Human CpG island DNA genomic MseI fragment, clone 165g2, forward read cpg165g2.ft1a.	338.8	382.6	94.3	271.3	237.2
RC_N67098_s_at	Human mRNA for KIAA0640 protein, partial cds.	168.7	227.1	40.5	210.4	133.6
RC_AA235799_s_	Human candidate tumor suppressor p33 ING1 homolog mRNA, complete cds.	269.4	402.7	112.9	342.7	0
L00972_at	Human cystathionine-beta-synthase (CBS) mRNA.	69	129.4	0	126.4	18.2
AB000460_at	Human mRNA, complete cds, clone:RES4-22A.	1073.3	1332.9	310.4	1247.9	648.2
K02777_s_at	Human T-cell receptor active alpha-chain mRNA from JM cell line, complete cds.	22.1	42	0	40.3	13.7
RC_AA235964_at	Incyte Unique	447	585.6	396.4	555.7	124.5
D63079_s_at	Incyte Unique	358.7	862.3	196	616.5	184.9

RC_AA284694_at	Human CG1 mRNA, complete cds.	98.3	184.9	53.2	175.7	82.2
RC_AA099360_at	Incyte Unique	234.4	276	0	161.8	161.7
AB001325_at	Human AQP3 gene for aquaporine 3 (water channel), partial cds.	655	905.2	198.1	605.4	490.6
RC_AA284844_at	Human mRNA; cDNA DKFZp434D0428 (from clone DKFZp434D0428); partial cds.	66.2	112.9	35	55.8	21.5
RC_N67227_at	KIAA0790 protein	44.1	128.6	37.4	75.7	26
RC_AA405488_at	Incyte Unique	1200.4	2441.8	1243	2339.4	518.9
J05412_at	Human regenerating protein (reg) gene, complete cds.	9132	12637.7	1078.6	6826.7	908.5
AB002356_s_at	Human mRNA for KIAA0358 gene, complete cds.	369.3	738	46.2	547.3	250.4
RC_AA405663_at	Human calmodulin-dependent protein phosphatase catalytic subunit (PPP3CA) mRNA, complete cds and alternative exon.	102.3	183.9	63.5	151.6	70.4
RC_R34249_at	Incyte Unique	4	4	0	5	0
RC_N67190_s_at	Human mRNA for serum response factor-related protein, RSRFC9.	31.2	56.3	0	34.6	25.2
RC_AA405669_at	Incyte Unique	41.5	107.6	0	74	47
J03915_s_at	Human chromogranin A mRNA, complete cds.	250.3	652.5	0	476.2	166.3
AA491188_at	Human clone HQ0310 PRO0310p1 mRNA, complete cds.	62.5	44.3	0	54	20.1
RC_AA406125_s	Incyte Unique	362.5	455.3	245.2	299.9	131.2
RC_AA056727_at	Human (clone 1NIB-138) normalized cDNA library sequence.	38.2	56.9	29.6	39.2	8.2
RC_R16833_at	Human mRNA for triple LIM domain protein.	24.8	29.9	0	26.9	7.2
RC_D60062_f_at	Public Unique	182.6	267.7	0	144.3	88.8
RC_AA402494_at	Human mRNA; cDNA DKFZp566E034 (from clone DKFZp566E034); complete cds.	102	250	53.2	170	55.6
RC_N68905_f_at	Human mRNA for KIAA0392 gene, partial cds.	1821.5	2968.7	0	2651.1	0
RC_N68990_s_at	Human MHC HLA-Dw12 DQ-beta chain mRNA, 3' end.	257.8	594.8	0	352.3	38.8
RC_AA599751_at	Human candidate tumor suppressor gene 21 protein isoform I mRNA, complete cds.	155	219.4	88.4	193.5	110.6
X87212_at	Human mRNA for cathepsin C.	469.5	525.2	166.7	415.7	293.2
RC_N68937_at	Incyte Unique	44.8	94.1	0	71.1	32.9
RC_AA401809_at	Human RNA-binding protein (RBMS3) mRNA, complete cds.	183.9	213.6	0	196.7	64.9
RC_AA234957_at	Human myotubularin-related protein 1a mRNA, partial cds.	214	283.7	107.5	250.3	128.5
RC_D59787_f_at	Incyte Unique	239.5	344.2	140.5	382.1	145.5
RC_AA085613_at	Incyte Unique	201.7	229	65	165.4	109.7
RC_AA287042_at	Incyte Unique	118.6	249	111.6	240.6	40.9
RC_AA600012_at	Human mRNA for KIAA1249 protein, partial cds.	179.4	265.9	68.5	256.7	131.3
RC_AA600125_at	Incyte Unique	177	256.7	141.9	286.2	82.2
AB002306_at	Human mRNA for KIAA0308 gene, partial cds.	91.9	146.2	55.1	141	29.5
RC_AA057193_at	Incyte Unique	20.8	35.1	0	20	0
X87160_at	Human mRNA for gamma subunit of epithelial amiloride-sensitive sodium channel.	77.6	29.5	0	74.7	12.9
RC_D60296_at	Incyte Unique	64.4	90.1	0	23.9	0
RC_AA404277_at	Incyte Unique	111.4	123.1	45	108	49.9
RC_AA599214_at	Incyte Unique	137.6	349.3	103.6	260	54.3
RC_AA403121_at	Human mRNA; cDNA DKFZp566I174 (from clone DKFZp566I174); partial cds.	58.4	121.2	38.5	88.4	18.8
J03756_at	Human placental lactogen hormone (PL-4) mRNA, complete cds.	263.4	385.1	0	222.6	145.9
RC_AA285153_at	Human mRNA for M-phase phosphoprotein, mpp4, 2043bp.	27.7	71.4	0	56.9	9.3
RC_AA235112_at	Human mRNA for Musashi, complete cds.	429.7	647.9	136.5	352.1	120.1
J03473_at	Human poly(ADP-ribose) synthetase mRNA, complete cds.	305.7	496.7	74	366.5	282.1
RC_R25114_at	Incyte Unique	269.8	404.4	224.3	298.6	0
RC_AA404263_at	Public Unique	192.9	272.3	163.4	324	96.6
RC_AA055829_at	Incyte Unique	71.3	109.3	77.4	123.5	13.1
RC_AA285167_at	Incyte Unique	149.9	271.9	131.3	242.9	96.8
X85372_at	Human mRNA for Sm protein F.	215.5	363.4	48.1	225.8	216.2
RC_R20655_f_at	Human ribosomal protein L9 pseudogene.	2600.3	3161	774.5	3704.3	725.3
D11327_s_at	Human mRNA for protein-tyrosine phosphatase, complete cds.	51.7	39.6	0	64.3	0
J02923_at	Human 65-kilodalton phosphoprotein (p65) mRNA, complete cds.	349.7	469.9	0	216.6	42.4
RC_AA286710_at	Human adaptor protein Lnk mRNA, complete cds.	180.9	319.1	103	310.4	112.4
RC_AA599501_at	Human mRNA for p0071 protein.	129.2	329.9	84.1	248.2	57.8
RC_AA599526_at	Human mRNA for cartilage-associated protein (CASP).	1075.8	1038.5	0	1327.9	323.3
RC_R22206_at	Incyte Unique	9.6	4	0	8.8	0

J03171_at	Human interferon-alpha receptor (HuIFN-alpha-Rec) mRNA, complete cds.	21.9	17.3	0	17.7	0
D10537_s_at	Human mRNA for major structural protein of myelin, complete cds.	132.4	136.4	55.7	91.7	68.9
RC_N66900_at	Public Unique	43.9	64.6	28.7	73.6	15.4
RC_AA520989_at	Human mRNA for C3G protein, complete cds.	833.2	677.6	271.9	934.1	113.8
RC_AA284356_at	Human CpG island DNA genomic MseI fragment, clone 21e5, forward read cpg21e5.f1b.	260.2	356.1	186.9	341.2	142.4
RC_N69068_at	Human full length insert cDNA clone YO64F11.	140.5	180.9	0	130.2	0
W80658_at	Human mRNA; cDNA DKFZp434J0713 (from clone DKFZp434J0713).	294.3	397.9	243.3	361.3	99.7
RC_AA283035_f_1	Incyte Unique	262.6	383.2	73.5	348.6	213.4
RC_AA283007_s_1	Human mRNA for KIAA0660 protein, complete cds.	527.4	815.9	340.6	848.3	166.4
RC_F04597_at	Incyte Unique	166.2	400.6	0	295.7	149.8
RC_AA412420_at	Incyte Unique	261.5	386.9	0	433.6	149.9
RC_R42338_at	Incyte Unique	156.8	64.8	0	169	36
RC_N64489_at	Incyte Unique	38.8	56.5	0	70.6	0
RC_AA236493_at	Human BH3 interacting domain death agonist (BID) mRNA, complete cds.	41.3	21	0	34.1	0
RC_AA496374_at	Incyte Unique	120.6	146.6	62.2	129.9	33.6
RC_F04564_at	Human aryl hydrocarbon receptor nuclear translocator (ARNT) mRNA, complete cds.	994.7	947	218.3	843.9	475.3
X69819_at	Human ICAM-3 mRNA.	453.4	461	126.2	396.4	156.9
L13852_at	Human ubiquitin-activating enzyme E1 related protein mRNA, complete cds.	993.7	1150.3	268.4	974	368.5
L12392_at	Human Huntington's Disease (HD) mRNA, complete cds.	257.1	292.8	0	295	0
X04476_s_at	Human lck mRNA for membrane associated protein tyrosine kinase.	229.3	304.8	51.4	157.8	0
X04011_at	Human mRNA of X-CGD gene involved in chronic granulomatous disease located on chromosome X.	117.1	272.4	59.6	161.5	22.7
RC_AA411818_at	Human mRNA; cDNA DKFZp434D1335 (from clone DKFZp434D1335); partial cds.	124.6	233.7	62.4	228.9	83
RC_N66093_at	Human BRCA1, Rho7 and vat1 genes, complete cds, and ipf35 gene, partial cds.	83.2	132.3	0	111.2	0
RC_AA194833_at	Human senescence-associated epithelial membrane protein (SEMP1) mRNA, complete cds.	291	333.7	63.3	242.7	218.2
RC_AA496573_at	Human genomic DNA of 8p21.3-p22 anti-oncogene of hepatocellular colorectal and non-small cell lung cancer, segment 3/11.	891.3	1736.5	533.8	1686.6	304.3
RC_AA236482_at	Human mRNA for KIAA0680 protein, complete cds.	24.6	12.3	0	31.4	0
L13848_at	Human RNA helicase A mRNA, complete cds.	45	88.4	0	77.1	44.5
W77734_at	Incyte Unique	113.4	192.4	13.4	60.8	65.2
RC_AA496963_at	Human full length insert cDNA clone YZ83H06.	51.6	60.1	24	46.2	20.7
RC_AA046405_at	Human mRNA for KIAA0022 gene, complete cds.	326.4	541.7	261.6	570.2	161.3
X66401_cds1_at	Human DMA, DMB, HLA-Z1, IPP2, LMP2, TAP1, LMP7, TAP2, DOB, DQB and RING8, 9, 13 and 14 genes.	761.8	1451.5	348.2	1155.7	274.5
RC_F09173_at	Human mRNA for large subunit of ribosomal protein L21.	56.8	108.8	0	80.3	60.9
RC_AA416686_at	Incyte Unique	23.5	32.1	0	36.4	7
L15702_at	complement factor B [Human, liver, mRNA, 2447 nt].	1764.5	1708.8	336.3	1492.4	391.5
L15309_at	Human zinc finger protein (ZNF141) mRNA, complete cds.	71.7	60.9	0	118.1	15.7
RC_AA416697_at	Incyte Unique	695.8	765	472.4	841	248.6
RC_F09458_at	Human clone 23674 mRNA sequence.	161	213.3	80.6	75.3	19.6
RC_F09051_at	Incyte Unique	288.5	424	64.2	153.7	182.2
RC_F08904_at	Incyte Unique	268	327.7	0	238.4	192
AA477046_at	Incyte Unique	110.9	184.5	126.7	188.5	13.5
RC_AA282914_at	Incyte Unique	27.1	45	0	32.7	0
RC_AA046457_at	Incyte Unique	837.4	811.8	364	931.8	435.7
RC_AA101878_at	Human apelin gene, complete cds.	102.3	206.6	62.3	154.2	0
W75976_at	Human dolichyl-phosphate beta-glucosyltransferase (ALG5) mRNA, complete cds.	395.4	575.5	140	327.7	264.4
RC_AA412548_at	Incyte Unique	148.1	299.5	37.6	185.6	71.6
X67698_at	Chimpanzee (Pan troglodytes) epididymal secretory protein precursor (EPI-1) mRNA, complete cds.	1497	1902.9	897.5	1832.4	1049.1
RC_AA194045_at	Human jumonji putative protein (jumonji) mRNA, complete cds.	73.2	115.7	36.9	83.6	22.8
RC_R43365_at	Human Cdc14B2 phosphatase mRNA, complete cds.	128	216.9	78.8	221.2	104.2
W74158_at	Incyte Unique	285	457.3	82.1	409.8	189.9
RC_N66132_s_at	Human gene for thymidylate synthase, exons 1, 2, 3, 4, 5, 6, 7, complete cds.	198.2	324.7	75.6	272.6	130.8
X15949_at	Human mRNA for interferon regulatory factor-2 (IRF-2).	128.4	254.4	75.3	180.1	0
RC_F02094_at	Human EVI5 homolog mRNA, complete cds.	281.8	444.4	109.7	389.5	164.2
X76105_at	Human DAP-1 mRNA.	150.1	218.9	0	218.1	46.4

RC_AA504631_at	Incyte Unique	85.4	179	41.2	164.6	52.8
X76302_at	Human RY-1 mRNA for putative nucleic acid binding protein.	33.9	68.3	0	37.2	28
RC_AA236223_at	Human T-cell activation protein (PGR1) gene, complete cds.	84.9	138.1	72	117.3	32.4
RC_R38436_at	Human mRNA for G9a.	252.7	531.5	0	327	194.2
Y11651_at	Human mRNA for phosphate cyclase.	109.3	127.1	45.2	116.3	28.9
RC_AA406629_at	Incyte Unique	87.7	155.9	69.7	139.3	58.7
RC_R38511_s_at	Human PP35 mRNA, complete cds.	57.4	120.3	82.1	104.7	0
Y10659_at	Human mRNA for IL13 receptor alpha-1 chain.	108.4	154.3	47.9	111.6	69.8
RC_AA504832_at	Human nuclear phosphoprotein mRNA, complete cds.	317.7	423.1	73.1	346.4	104.9
RC_AA406383_at	Human caveolin 1 (CAV1) gene, exon 3 and partial cds.	137.9	160.4	0	97.2	62.4
AA165564_at	Incyte Unique	96.5	91.6	33.8	105	35.7
RC_N66769_at	Incyte Unique	59.4	131.5	0	98.9	68.3
AA480828_at	Human mRNA for reticulocalbin, complete cds.	11.6	25.5	0	16.9	0
Z21420_at	Human clone C40 unknown mRNA.	147.5	266.9	80.5	222.5	139.4
Z19813_at	Human alpha gene sequence.	90.1	103.2	33.9	127	40.2
Z21081_at	Incyte Unique	332.3	257.3	168.4	333.1	71.5
RC_AA406373_at	Human DMA, DMB, HLA-Z1, IPP2, LMP2, TAP1, LMP7, TAP2, DOB, DQB and RING8, 9, 13 and 14 genes.	426.8	273.5	49.3	463.5	82.1
RC_AA406363_at	Human partial mRNA; ID YG81-2B.	52.2	153.5	37.6	98.4	10.8
RC_F02252_at	Incyte Unique	263.7	435.8	81.1	470.1	224.2
AA091932_at	Human CGI-04 protein mRNA, complete cds.	28.1	33.9	0	26	6.1
RC_AA199649_at	Human reclin mRNA, complete cds.	285.1	263.5	0	378.5	140.4
X16901_at	Human mRNA for RAP30 subunit of transcription initiation factor RAP30/74.	58.8	112.2	24.2	41.4	26.3
D82422_at	Human full length insert cDNA clone YB21H04.	779	942.3	597.1	1208.1	241
RC_R40431_at	Human mRNA; cDNA DKFZp564D016 (from clone DKFZp564D016).	951.4	985.2	293	1205	509.9
RC_AA504110_at	Human Ets-1 gene, partial sequence.	701.2	1303.2	651.3	1015.8	250.9
RC_F04015_at	Incyte Unique	160.7	261.2	32.2	237.9	46.7
RC_AA410986_at	Human mRNA full length insert cDNA clone EUROIMAGE 248114.	104.5	177.8	0	70.9	69.9
RC_AA410505_s_cds.	Human mRNA; cDNA DKFZp434L0117 (from clone DKFZp434L0117); parti	1308.7	2293.3	878.8	1815.2	1093.2
RC_AA411207_at	Human clone 24906 hJTB protein mRNA sequence, complete cds.	312.7	570	160.3	553.1	191
RC_AA496993_at	Incyte Unique	217.7	374.7	108.9	241.1	145.1
RC_AA195399_at	Human partial mRNA; ID YG31-2, YG36-2.	32	25.4	0	25.8	0
L09708_at	Human mRNA for histone deacetylase-like protein (JM21).	566.4	1045.5	134.1	634.1	143.7
RC_AA410336_at	Incyte Unique	136	270.7	64.7	256.6	113.2
RC_N66607_at	Incyte Unique	192.8	259	0	234.7	154.6
RC_N66577_at	Incyte Unique	131.4	220.6	109.2	230	88.9
RC_F02855_at	Human double stranded RNA activated protein kinase (PKR) gene, 3' flanking region.	868.8	686.1	0	827.9	357.8
RC_AA236280_at	Human mRNA for KIAA0965 protein, partial cds.	72.3	160.3	39.7	105.1	57.5
RC_AA196977_at	Human IGF-II mRNA-binding protein 1 (IMP-1) mRNA, complete cds.	87.2	80.7	0	128.8	0
X75346_s_at	Human mRNA for MAP kinase activated protein kinase.	23.5	50.1	0	32.5	0
RC_R39869_at	Incyte Unique	447.9	452.1	0	642.2	181.6
D82307_at	Human HSPC130 mRNA, complete cds.	176.2	279	85.1	129.9	90.6
X75208_at	Human HEK2 mRNA for protein tyrosine kinase receptor.	206.8	133	0	284.9	61.1
RC_R11673_at	Incyte Unique	55.2	32.2	0	42	0
D49676_s_at	Human U2AF1-RS2 mRNA, complete cds.	332.9	401.1	205.2	359	68.1
X65867_at	Human adenylosuccinate lyase (ADSL) mRNA, alternatively spliced, complete cds.	214	226.8	0	151.7	74
RC_AA393808_at	Human CGI-101 protein mRNA, complete cds.	236.4	195.2	63.4	197.6	102.4
RC_AA070801_at	Human X (inactive)-specific transcript (XIST) complete exon.	1478.2	1829.1	577.5	1527.6	899.5
RC_AA620629_at	Human mRNA for DCRA, complete cds.	244.2	277.5	0	359.6	125.3
RC_N73865_at	Human F-box protein FBX9 mRNA, complete cds.	318.7	356.3	118.9	355.8	224
RC_AA620598_at	Incyte Unique	125.1	163.2	41.7	193.8	80.2
D59253_at	Human mRNA; cDNA DKFZp586F0221 (from clone DKFZp586F0221).	60.6	73.2	24	39.2	36.1
RC_N73861_at	Incyte Unique	123.8	94.7	36.6	106.5	13.7
RC_D19708_at	Human Gu protein mRNA, partial cds.	39.1	64.2	0	28	14.8
D63475_at	Human mRNA for KIAA0109 gene, complete cds.	591.4	1178.6	243.1	830.1	629
RC_AA379085_at	Human HSPC313 mRNA, partial cds.	63.8	172.3	28.5	94.9	50.6

RC_D20045_at	Human clone KDB2.12 (CAC) <sub>m</sub> /(GTG) <sub>n</sub> repeat-containing mRNA.	12	12.4	0	9.5	4.6
D63485_at	Human mRNA for KIAA0151 gene, complete cds.	373.9	688	189.2	424.9	262.6
RC_AA357999_at	Human mRNA for INADL, C-term variant2.	252	355.5	85.7	424.5	159.9
RC_AA358109_at	Human mRNA for AMP-activated protein kinase alpha-1, complete cds.	42.1	76.3	31.1	75.5	11.4
RC_N74353_at	Incyte Unique	156.6	224.4	0	140.1	136.3
Z29678_at	Human mitF mRNA.	143.9	233.5	0	173.2	80.9
C00695_s_at	Human survival of motor neuron protein interacting protein 1 (SIP1) mRNA, complete cds.	21.7	22.8	0	15.3	0
Z29574_at	Human mRNA for BCMA peptide.	226.9	285.4	0	195.5	86.4
RC_AA365691_at	Incyte Unique	58.4	85.2	21.3	22.7	10.3
RC_AA070649_s_	Human hereditary haemochromatosis region, histone 2A-like protein gene, hereditary haemochromatosis (HLA-H) gene, RoRet gene, and sodium phospho transporter (NPT3) gene, complete cds.	2043.3	1817.6	517.2	1920	985.8
RC_AA227780_at	Incyte Unique	100.9	161.2	66	152.8	31.2
RC_AA227934_at	Human Ntera2D1 cell line mRNA containing L1 retroposon, clone R8.	34	54.1	12.9	39.4	19.1
D43949_at	Human mRNA for KIAA0082 gene, partial cds.	203.4	249.6	0	190.8	93
RC_N72295_at	Human full length insert cDNA clone ZD64D01.	458	376.5	128.8	388.6	88
RC_AA083478_s_	Human Staf50 mRNA.	453.1	596.4	60.5	418.8	287.1
RC_AA227483_at	Human messenger RNA fragment for the beta-2 microglobulin.	54.2	85.4	47.1	103.4	14.5
RC_AA291553_at	Incyte Unique	191.5	242.1	80.2	223.5	123.1
Y11651_at	Human mRNA for phosphate cyclase.	109.3	127.1	45.2	116.3	28.9
D45248_at	Human mRNA for proteasome activator hPA28 subunit beta, complete cds.	1268.5	3352.1	778.7	1824.3	762.8
RC_AA398109_at	Human PROS-27 mRNA.	115.5	173.6	68.3	156	70.3
RC_D20313_at	Human full length insert cDNA clone ZD69A07.	4	8.2	0	5	0
RC_AA291710_at	Human goodpasture antigen-binding protein (COL4A3BP) mRNA, complete cds.	83	204.3	64.5	155.1	65.7
RC_AA620520_at	Incyte Unique	58.1	92.7	0	28.7	0
D50525_at	Human mRNA for TI-227H.	158.8	259.5	0	177.2	54.2
RC_D20158_at	Human mRNA, cDNA DKFZp586A0422 (from clone DKFZp586A0422).	37.3	100	15.1	54.1	16.3
RC_N93853_at	Incyte Unique	575.9	728.7	222.7	846.6	266.9
RC_AA227537_at	Incyte Unique	254.9	208.3	0	263	117.2
D49490_at	Human mRNA for protein disulfide isomerase-related protein (PDIR), complete cds.	111	188.6	0	162.6	82.7
D49489_at	Human mRNA for protein disulfide isomerase-related protein P5, complete cds.	643.2	1131.1	104.3	818.4	640
RC_N91968_at	Incyte Unique	47.4	32.1	6.7	61.9	16.7
RC_N91773_at	Incyte Unique	82.5	139.5	0	64.1	33.1
RC_D20483_f_at	Human tapasinas (tapasin) mRNA, alternatively spliced, complete cds.	68.2	190.1	34.8	142.8	21.2
RC_AA347359_s_	Human monocyte/macrophage Ig-related receptor MIR-7 (MIR cl-7) mRNA, complete cds.	7795.6	13854.4	2550.3	6089.9	4658.9
RC_C13977_at	Incyte Unique	704	786.0	260.8	842.8	400.7
D87073_at	Human mRNA for KIAA0236 gene, complete cds.	155.3	122.4	0	144.3	31.9
RC_AA074729_at	Incyte Unique	130.8	79.6	38.3	132.5	22.7
RC_AA074880_at	Incyte Unique	219.9	248.2	152.8	228.6	47.4
RC_AA232644_s_	Human protein-tyrosine phosphatase mRNA, complete cds.	357.2	279.4	102.6	331.3	159.9
RC_N89820_at	Incyte Unique	117	155.2	0	105.3	66.6
RC_N89829_at	Incyte Unique	41.1	60.4	0	68.4	30.2
D86979_at	Human mRNA for KIAA0226 protein, partial cds.	171.7	470.4	0	297.5	148.4
RC_AA342842_at	Human PHEX gene.	25.6	64.9	0	36.6	0
D86978_at	Human mRNA for KIAA0225 gene, partial cds.	97.4	239.1	0	172.9	89.6
D79992_at	Human mRNA for KIAA0170 gene, complete cds.	158.8	236.9	60.1	89.3	85.9
RC_C14845_f_at	Human ubiquitin-protein ligase E3-alpha (UBR1) gene, exon 9.	90.4	223.1	52.1	124.7	76.9
RC_AA076133_at	Incyte Unique	159.5	220.5	0	254.7	25.7
Z84497_s_at	Human mRNA for KIAA9001 gene, complete cds.	403.8	480.3	201.9	446.5	269.5
D86967_at	Human mRNA for KIAA0212 gene, complete cds.	631.3	631	127.1	549.8	289.3
RC_AA232904_at	Human mRNA, cDNA DKFZp570I0164 (from clone DKFZp570I0164); partial cds.	225.4	267.9	104.6	270.7	32.1
RC_AA340539_at	Human DNA sequence.	153.6	185.1	67.4	103.7	79.6
RC_AA075642_at	Human mRNA for DMBT1 6 kb transcript variant 1 (DMBT1/6kb.1).	16132.5	17618.8	10900.8	12212.4	1665.8
C01747_at	Human mRNA, cDNA DKFZp586B1922 (from clone DKFZp586B1922).	224.6	243.6	124.2	346	55.1
RC_N89836_at	Human constitutive fragile region FRA3B sequence.	308.2	354.5	125	258.5	172.6

D79991_at	Human mRNA for KIAA0169 protein, partial cds.	327.4	461.5	95.3	305.7	251.4
D88422_at	Human radiated keratinocyte mRNA for cysteine protease inhibitor.	79.8	191.4	0	164.5	0
RC_AA621159_at	Incyte Unique	182.9	262.1	127.6	193.3	82.6
RC_AA228002_i :	Incyte Unique	70.8	140.4	63.3	143.3	20.4
RC_AA233261_at	Incyte Unique	36	81.8	0	52.6	0
RC_AA071294_at	Public Unique	40.8	16.8	0	37.6	0
D63880_at	Human mRNA for KIAA0159 gene, complete cds.	106.3	209.7	38.6	136.2	111.5
RC_AA357189_at	Human beta-arrestin 2 mRNA, complete cds.	238.4	500.1	190.2	382.2	200.4
D87989_at	Human mRNA for UDP-galactose transporter related isozyme 1, complete cds.	407.7	779	0	465.9	274.3
RC_AA071089_at	Incyte Unique	158.2	313.3	186.7	266.9	76.6
D67029_at	Human SEC14L mRNA, complete cds.	215.7	342.9	65.1	88.3	89.4
RC_AA350817_at	Human clone 25088 mRNA sequence.	21.6	15.6	0	16.7	0
RC_N90882_s_at	Human transglutaminase E3 (TGASE3) mRNA, complete cds.	286.9	391.1	0	403.8	0
RC_AA292701_at	Human mRNA; cDNA DKFZp564I052 (from clone DKFZp564I052).	116.8	98.8	51.9	99	19.3
RC_AA350690_at	Human protein associated with Myc mRNA, complete cds.	441	697.1	443.8	570.3	174.6
C02352_s_at	Human CGI-121 protein mRNA, complete cds.	47.3	64.1	33.7	55.1	12.2
RC_AA621413_at	Human orphan neurotransmitter transporter NTT5 (NTT5) mRNA, complete c	56.5	83.2	0	46.3	0
RC_AA621340_at	Incyte Unique	155	131.2	62.3	151.1	19.4
RC_AA350729_at	Human clone KDB2.12 (CAC) <sub>n</sub> /(GTG) <sub>n</sub> repeat-containing mRNA.	181.5	361.4	81.7	278.1	40.1
RC_AA350796_at	Human mRNA for KIAA0939 protein, partial cds.	326.4	220.3	135.8	272.5	0
D87465_at	Human mRNA for KIAA0275 gene, complete cds.	158.5	407.1	0	270.7	32.7
RC_AA074038_s_.	Human mRNA; cDNA DKFZp564B167 (from clone DKFZp564B167); complete cds.	33.5	80.7	0	50.2	16.4
Y10936_at	Human mRNA for hypothetical protein downstream of DMPK and DMAHP.	121.4	170.7	0	216.6	67.9
RC_N72094_at	Incyte Unique	76.4	111.4	38.9	100.1	54.7
Y10659_at	Human mRNA for IL13 receptor alpha-1 chain.	108.4	154.3	47.9	111.6	69.8
RC_N69101_at	Incyte Unique	770.8	813.5	0	748.9	528.2
RC_AA400410_at	Human NNX3 (C19orf2) mRNA, complete cds.	183.3	304.4	95.7	292.9	123.3
RC_AA400277_at	Incyte Unique	97.8	259.4	0	152.1	117.4
RC_AA608777_s_.	Human mRNA for cystinosis.	497.5	725	240.2	736.1	249
RC_AA400482_at	Human mRNA for KIAA0511 protein, partial cds.	165.1	220	70.7	203.7	104.5
RC_AA608908_at	Incyte Unique	444.8	758.5	0	596.9	297.2
RC_AA400512_at	Human RU1 (RU1) mRNA, complete cds.	43.6	131.3	39.7	69.2	14.5
RC_R08850_at	Incyte Unique	727.9	727.3	316.7	742.9	319.1
RC_N69656_at	Incyte Unique	91	80.4	0	105.4	27.1
D25216_at	Human mRNA for KIAA0014 gene, complete cds.	0	0	0	0	0
D31313_s_at	Human unknown mRNA.	399.4	394	161	363.4	108.6
RC_AA400517_at	Human mRNA; cDNA DKFZp434G1115 (from clone DKFZp434G1115); partial cds.	77.1	60.8	12.7	57.9	0
RC_AA084769_at	Incyte Unique	142.3	142.9	0	131.8	0
RC_R06852_at	Human HSPC076 mRNA, partial cds.	424.9	610.7	0	430.7	87.7
RC_D54289_s_at	Human CLP mRNA, partial cds.	2122.7	2731.1	1070.4	2037.5	1547.1
RC_AA400074_at	Similar to Rat trg gene product; coded for by C. elegans cDNA yk31e7.5; code for by C. elegans cDNA yk40d6.5; coded for by C. elegans cDNA yk31e7.3; coded for by C. elegans cDNA yk40d6.3; coded for by C. elegans cDNA yk149g5.3; coded for by C>	58.7	82.8	15.3	57.1	0
RC_AA223209_at	Human high-glucose-regulated protein 8 (HGRG8) mRNA, complete cds.	404.9	388.6	196.3	367.5	151.1
RC_AA400229_at	Incyte Unique	155.4	274.5	59.4	214.1	45.6
RC_D59276_at	Incyte Unique	143.7	230.3	120.8	225.1	30.7
X95808_s_at	Human mRNA for protein encoded by a candidate gene, DXS6673E, for mental retardation.	99.5	271	89.7	189.8	59.8
RC_AA400528_at	Incyte Unique	53.4	91.2	0	48	14
RC_AA062731_at	Human thyroid hormone receptor-associated protein complex component TRAP150 mRNA, complete cds.	169.4	236.6	114.3	173.5	91.3
RC_AA600363_at	Human mRNA for KIAA0709 protein, complete cds.	732.5	674	125.6	571.6	253.2
D14889_at	Human mRNA for small GTP-binding protein, S10, complete cds.	93.8	201	0	118.6	74.5
D14874_at	Human mRNA for adrenomedullin precursor, complete cds.	288.5	418.2	88.3	190.2	157.8
D14812_at	Human mRNA for KIAA0026 gene, complete cds.	1664.7	2258.6	1011.3	1954	1257.8
RC_AA401258_at	Human mRNA; cDNA DKFZp566J153 (from clone DKFZp566J153); complete cds.	62.7	166.3	0	97.6	56.3
AB002328_at	Human calcineurin binding protein cabin 1 mRNA, complete cds.	103.5	179.6	0	172.6	0

X91809_at	Human mRNA for GAIP protein.	125.3	278.4	0	226.9	46.1
D14695_at	Human clone 24560 unknown mRNA, complete cds.	824.2	975.7	140.2	897.5	435
RC_AA287109_at	Incyte Unique	360.9	317.1	104.2	330.9	165.5
D16105_at	Human ltk mRNA for tyrosine kinase.	426.3	505.3	0	640.5	273.2
RC_AA401144_s_	Incyte Unique	250	511.7	315.5	510.8	55.1
RC_D59553_f_at	Human mRNA: cDNA DKFZp434O159 (from clone DKFZp434O159).	699.2	1025.6	190.6	746.8	369.9
RC_AA400760_at	Human metabotropic glutamate receptor 6 (mGluR6) gene, complete cds.	222.3	447.4	207.8	414.8	81.4
RC_R09547_at	Incyte Unique	225.4	110.8	0	190.9	0
RC_AA608546_at	Human clone 24670 mRNA sequence.	313	358.7	209.3	303.3	56.9
RC_AA608699_at	Incyte Unique	105.5	103.4	23.5	107.3	31.5
RC_AA400948_at	Human CpG island DNA genomic MseI fragment, clone 21g11, forward read cpg21g11.ft1b.	1004.1	952.4	145.1	798.7	476.4
RC_AA058686_at	Incyte Unique	453	453.1	171.9	364.4	96.6
AA039789_at	Incyte Unique	99.6	76	6.7	111.3	3.3
RC_AA290630_at	Human mRNA for KIAA1268 protein, partial cds.	443.3	633.9	350.9	650.9	50.8
AF004292_at	Human HSPC284 mRNA, partial cds.	323.6	256.7	89	290.8	49.9
D10537_s_at	Human mRNA for major structural protein of myelin, complete cds.	132.4	136.4	55.7	91.7	68.9
RC_AA065114_at	Incyte Unique	134.3	150.2	64	93.8	0
RC_AA290603_at	Incyte Unique	52.2	107.9	0	88.3	58.3
D38548_at	Human mRNA for KIAA0076 gene, complete cds.	91.5	37.9	0	83.8	0
Y00796_at	Human mRNA for leukocyte-associated molecule-1 alpha subunit (LFA-1 alpha subunit).	211.3	408	0	201.9	98.7
RC_N70964_s_at	Incyte Unique	37.7	33.7	0	21.2	0
RC_AA398606_at	Incyte Unique	64	58.2	16	47.8	12.7
Y08976_at	Human mRNA for FEV protein.	393.3	599.6	0	590.4	76.7
AA040839_at	Human putative 13 S Golgi transport complex 90kD subunit brain-specific isoform mRNA, complete cds.	56.6	81.9	0	37.1	0
RC_AA398745_at	Human HRIHFB2216 mRNA, partial cds.	487.6	409.3	13.3	435.5	280.5
RC_AA398368_at	Human mRNA: cDNA DKFZp586K2123 (from clone DKFZp586K2123).	96.2	211.7	69.7	102.2	55.1
RC_AA398346_at	Incyte Unique	45	55.1	10.8	56	13.7
RC_AA620307_at	Incyte Unique	132	153.7	0	126.7	0
RC_AA610070_at	Human CASK mRNA, complete cds.	110.1	215.9	40.3	200.4	72.5
RC_AA610073_at	Human mRNA for GTPase activating protein ID-GAP, complete cds.	68.1	136.8	50	91.9	30.7
RC_AA233620_at	Incyte Unique	60.7	78.1	44.4	48.8	0
D38555_at	Human mRNA for KIAA0079 gene, complete cds.	212.8	449.3	0	328.3	94
RC_AA233609_s_	Human spindle pole body protein spc98 homolog GCP3 mRNA, complete cds.	63.7	45.8	38.6	70.2	0
RC_D25917_at	Incyte Unique	13.5	42.3	24.4	32.2	0
Y00705_at	Human pancreatic secretory trypsin inhibitor (PSTI) mRNA, complete cds.	507	981.1	79.7	509.6	272
D26362_at	Human mRNA for KIAA0043 gene, complete cds.	507.7	657.5	144.8	761.5	335
Y00062_at	Human mRNA for T200 leukocyte common antigen (CD45, LC-A).	278.1	479.1	0	272.7	45
RC_N70546_at	Human breast cancer putative transcription factor (ZABC1) mRNA, complete cds.	410.4	409.6	0	289.1	0
RC_AA084640_at	Human mRNA for Musashi, complete cds.	249.3	444.9	118.9	382.9	152.6
X99325_at	Human mRNA for YSK1, complete cds.	214.9	490.9	158.9	284.3	89.2
RC_AA224324_at	Human Ikapab kinase complex associated protein (IKAP) mRNA, complete cds.	88.6	152.5	83.8	129.6	40.9
RC_AA399418_at	Human mRNA for JM23 protein, complete coding sequence (clone IMAGE 34581 and IMAGE 45355 and LLNLc1101133Q7 (RZPD Berlin)).	194.1	204.7	61.2	127.7	37.2
AB002353_at	Human mRNA for KIAA0355 gene, complete cds.	121.4	255.8	48.7	200.1	55.6
RC_R02354_at	Incyte Unique	249.7	235.7	0	161.9	94.3
D30851_at	Human genomic DNA (chromosome 3; clone NRL063D).	108.5	244.8	0	147.6	112.8
RC_AA609717_at	Incyte Unique	285	333.4	74.2	168.8	100.6
RC_AA609773_at	Human PRO0530 mRNA, complete cds.	1520	1370.1	467.7	1358.4	839.8
RC_D45719_at	Public Unique	375	263.8	173.3	466.4	0
RC_AA084349_at	Human mRNA: cDNA DKFZp586I031 (from clone DKFZp586I031).	687.3	701.3	135.3	583.4	412.8
RC_AA084408_at	unidentified reading frame 4	2441.3	1537.7	0	2435.4	389.9
RC_AA063618_at	Incyte Unique	211	341.9	84.6	264.3	51.1
RC_N70711_at	Human CGI-116 protein mRNA, complete cds.	579.1	737.6	278	631.8	321.7
AA045870_at	Human mRNA: cDNA DKFZp564A072 (from clone DKFZp564A072).	108	183.5	55.5	178.4	80.4
RC_R01901_f_at	Human T-cell receptor alpha delta locus from bases 501613 to 752736 (section of 5) of the Complete Nucleotide Sequence.	73.3	91.2	0	112	0

RC_AA046085_at	Human genomic DNA of 8p21.3-p22 anti-oncogene of hepatocellular colorectal and non-small cell lung cancer, segment 7/11.	584.4	673.2	0	520.1	399.8
L08010_at	Human mRNA for regenerating protein 1 beta, complete cds.	4966.5	7023.3	231.4	3028.7	559.9
RC_W33134_s_at	Incyte Unique	181.6	183.9	92.9	155	17.3
RC_AA035741_at	Human CpG island DNA genomic MseI fragment, clone 171a11, reverse read cpg171a11.rt1a.	154.2	274.5	133.3	222.6	55
L38961_at	Human putative transmembrane protein precursor (B5) mRNA, complete cds.	243.7	312.7	163.2	151.5	61.7
RC_AA114091_at	Incyte Unique	42.6	101.3	0	66.5	42.2
U37546_s_at	Human TNFR2-TRAF signalling complex protein mRNA, complete cds.	664.9	829.7	485.9	838.5	96.3
RC_AA035649_at	Incyte Unique	17.2	29.3	0	20.7	17.1
RC_AA418900_f_1	Human male-specific lethal-3 homolog 1 (MSL3L1) mRNA, complete cds.	260.3	299.2	67.4	259.3	72.6
RC_H05960_at	Incyte Unique	66.4	93.3	0	89.7	16.9
RC_AA043360_at	Incyte Unique	662.8	1027.1	213.2	815.8	427.1
RC_R53765_at	Human mRNA for KIAA0981 protein, partial cds.	63	104.5	19.8	57	38.6
AA454462_at	Human full length insert cDNA clone ZD87B10.	129.5	269.1	20	234.4	13.3
AA454908_s_at	Human genomic DNA, chromosome 22q11.2, BCRL2 region, clone:KB1440D	314.8	73.7	0	341.1	91.9
RC_AA488432_at	Human mRNA for L-3-phosphoserine phosphatase.	194.2	340	54.4	226	148.8
X56741_at	MEL=RAS-related [Human, Genomic/mRNA, 1980 nt].	78.5	111.8	0	102.2	16.2
RC_AA113356_at	Human mRNA for Drg1 protein.	298.9	230.5	0	352.1	147.2
RC_AA035630_at	Human U4/U6 snRNP 60 kDa protein gene, complete cds.	249.1	349.3	125.2	265	141.3
L38951_at	Human importin beta subunit mRNA, complete cds.	345.9	665.5	137.8	640.1	301.1
W25781_at	Human MOP2 mRNA, complete cds.	51	37.8	19.2	37	0
M19650_s_at	Human tetratricopeptide repeat protein (tpr2) mRNA, complete cds.	91.7	160.3	0	111.1	94.5
RC_R71324_at	Human full length insert cDNA clone YI54B09.	42.9	53.6	0	20	17.1
RC_AA251089_i_1	Human mRNA; cDNA DKFZp43411226 (from clone DKFZp43411226).	51.1	50.1	0	40.9	15.5
W19201_at	Human vascular endothelial cell growth factor 165 receptor 2 (VEGF165R2) mRNA, complete cds.	51.1	43.8	25.5	90.2	20.8
L40586_at	Human iduronate-2-sulphatase (IDS) mRNA, complete cds.	45.9	58.7	0	35.8	8.1
RC_AA481407_at	Incyte Unique	157.2	218.2	0	187.7	51.7
W21426_at	Human mRNA for KIAA0806 protein, complete cds.	119.6	173.3	0	135.7	39.9
U54999_at	Human LGN protein mRNA, complete cds.	20	15.3	0	17	0
RC_AA419263_at	Incyte Unique	48.4	78	19.8	46.8	16.1
RC_R55270_at	Human mRNA for KIAA1043 protein, partial cds.	399.7	622.5	119	517.8	189.7
RC_AA481256_at	Incyte Unique	147.9	130.7	0	110.3	58.5
RC_AA488199_at	Human mRNA for KIAA0693 protein, partial cds.	32.8	44.6	26.5	33.7	13
RC_AA243281_at	Incyte Unique	50.8	57.6	28.4	89.2	0
M19645_at	Human mRNA for BiP protein.	816.6	1756.6	134.3	1018.2	806.6
RC_AA424517_at	Incyte Unique	111.4	147.6	86.4	186.9	31
RC_AA281214_s_1	Human neuroblastoma-amplified protein mRNA, complete cds.	224.8	438.5	222.5	344	139.2
RC_AA243443_at	Human mRNA capping enzyme (HCE) mRNA, complete cds.	234.1	257.2	107.7	212.3	92.1
X55733_at	Human initiation factor 4B cDNA.	758.1	1036.5	188	982.6	573.8
RC_AA418824_at	Human Nedd-4-like ubiquitin-protein ligase WWP1 mRNA, partial cds.	190.9	315.8	121.6	328.9	129.9
W26496_at	Human WSB-1 mRNA, complete cds.	215.1	300.3	58.2	236.2	106.8
U57316_at	Human GCN5 (hGCN5) gene, complete cds.	119.1	187.2	78	102.9	26.7
RC_AA480855_at	Incyte Unique	64.5	147.7	66.6	92.2	22.6
L34587_at	Human RNA polymerase II elongation factor SIII, p15 subunit mRNA, complete cds.	291.7	656.5	108.3	315.5	288.6
L35545_at	Human endothelial cell protein C/APC receptor (EPCR) mRNA, complete cds.	32.2	48.1	0	20.8	15.7
L34657_at	Human platelet/endothelial cell adhesion molecule-1 (PECAM-1) gene, exon 1 and complete cds.	240.5	434	0	203.4	22.8
RC_N53471_at	Incyte Unique	93.5	125.3	58.8	101.4	61.2
U15128_at	Human beta-1,2-N-acetylglucosaminyltransferase II (MGAT2) gene, complete cds.	105.2	93.1	39.9	103.4	55.6
RC_AA417956_at	Incyte Unique	145.7	130.8	0	144.8	85.2
RC_AA424813_at	Incyte Unique	105.4	299.1	103.9	202.9	0
RC_AA242753_at	Incyte Unique	22.5	42.1	0	19.1	9.4
H19597_at	Human ZIS1 mRNA, complete cds.	32.2	71.2	23.9	58.9	21.9
H85185_at	Human M-phase phosphoprotein homolog mRNA, complete cds.	41.4	107.7	34.7	71.8	33.6
RC_N53388_at	Incyte Unique	368	383.3	92.7	415	236.3
X06700_s_at	Human mRNA for pro-alpha-1 type 3 collagen.	917.2	1412	331.1	854.6	494.5

W27873_at	N-CAM=145 kda neural cell adhesion molecule [Human, small cell lung cancer cell line OS2-R, mRNA, 2960 nt].	66.8	101.5	35.7	55	0
L33799_at	Human mRNA for type 1 procollagen C-protease enhancer protein, complete cds.	235.1	183.7	0	223.8	110.7
RC_R73565_at	Human mRNA; cDNA DKFZp564M113 (from clone DKFZp564M113).	225.8	171.8	6.7	186.7	82.6
X58529_at	Human mRNA for IgM heavy chain complete sequence.	1391.2	2406.9	0	997	207.7
RC_AA489041_at	Incyte Unique	52.4	89.8	32.4	68.5	6.2
L38616_at	Human brain and reproductive organ-expressed protein (BRE) gene, complete cds.	269.5	277.2	66	318.2	169.5
RC_AA488885_at	Human partial cDNA sequence, complete sequence of clone 98D12.	180.1	302.8	114.9	311.6	86.4
RC_AA035621_at	Incyte Unique	102.2	232.9	92.9	156.2	28.3
RC_AA418726_at	Human clone 24640 mRNA sequence.	142.5	265	80.5	235.3	61.3
U30246_at	Human bumetanide-sensitive Na-K-Cl cotransporter (NKCC1) mRNA, complete cds.	195.9	332.5	52.5	166.2	100.6
RC_AA243007_at	Human mRNA for Npw38-binding protein NpwBP, complete cds.	125.8	172.9	70	119.2	28
AA437171_at	Human transcriptional regulatory protein p54 mRNA, complete cds.	78.6	202.5	46.6	128.6	63.2
RC_AA043790_at	Human mRNA for KIAA0937 protein, partial cds.	83.7	104.9	15.3	96.3	5.6
RC_N53943_at	Human serum deprivation response (SDPR) mRNA, complete cds.	207.9	323.1	86.7	307.5	159.2
RC_R51872_at	Public Unique	235.7	346	82.5	327.5	57.9
RC_AA044016_at	Human P-selectin glycoprotein ligand (SELPLG) gene, exon 2, and complete cds.	212.3	240.4	0	221.4	108.6
AA100219_at	Incyte Unique	119.2	187.9	38.3	104.3	65.1
L36922_at	Human Met-ase gene, exon 1.	251	483.7	0	243.3	145.1
RC_N62122_at	Human mRNA; cDNA DKFZp564E1962 (from clone DKFZp564E1962); partial cds.	78.7	114.7	34.6	156.1	18.9
RC_AA242868_at	Incyte Unique	90.2	136.5	0	65	62.3
AA456471_s_at	Human mRNA for KIAA0963 protein, complete cds.	74	243.7	0	169.8	0
RC_AA251237_at	Incyte Unique	157.6	253.9	115.3	109.3	37.1
L41067_at	Human NF-AT4c mRNA, complete cds.	349.7	549.9	123.4	565	251.2
X13444_at	Human mRNA for CD8 beta-chain glycoprotein (CD8 beta 1).	331.4	440.4	0	102.8	0
RC_AA042813_at	Human mRNA for TPCR26 protein.	363	448.7	92.6	308.3	197.5
RC_AA489461_at	Human mRNA for KIAA0540 protein, partial cds.	103.8	111.3	0	62.2	0
RC_AA243765_at	Human lyn mRNA encoding a tyrosine kinase.	573	549.1	120.3	476.8	173.5
X15187_at	Human tra1 mRNA for Human homologue of murine tumor rejection antigen gp96.	1333.1	1444.6	0	878.1	788
RC_AA120785_at	Incyte Unique	150.5	236.9	90.2	176.7	123.5
RC_AA281545_at	Human mRNA; cDNA DKFZp434I0812 (from clone DKFZp434I0812); partial cds.	24.7	43.6	0	14.8	16.8
RC_AA482112_at	Human choline/ethanolaminophosphotransferase (CEPT1) mRNA, complete cds.	242.8	414.3	77.3	334.5	93.1
RC_AA424147_at	Human mRNA; cDNA DKFZp566G184 (from clone DKFZp566G184).	88	131.9	53.9	157.6	64.7
H60661_at	Human Staf50 mRNA.	168.3	222.9	0	124	94.6
RC_AA120944_s_at	Incyte Unique	539.4	649.7	204.3	553.9	374.5
RC_AA039461_at	Human arginine methyltransferase mRNA, complete cds.	220.8	309.8	0	345.3	151.1
RC_AA281508_at	Human genomic DNA, chromosome 6p21.3, HLA Class I region, section 15/21	203.1	165.1	0	210.4	32.9
M14219_at	Human chondroitin/dermatan sulfate proteoglycan (PG40) core protein mRNA complete cds.	82.9	153.6	0	128.7	77.3
RC_AA485212_at	Human mRNA; cDNA DKFZp434K1412 (from clone DKFZp434K1412).	284.6	306.4	97.9	376.1	147.1
L78833_cds4_at	Human breast and ovarian cancer susceptibility (BRCA1) mRNA, complete cds.	13.4	10.8	0	16.7	0
U87281_s_at	Human hVps41p (hVPS41) mRNA, alternative splice variant, partial cds.	160	257.6	92.3	183.6	109.4
RC_AA179510_at	Human clone ICRFy900g0523 from Xqter pseudoautosomal region.	85.9	157.7	0	100.1	52.9
RC_AA424175_at	Human CGI-65 protein mRNA, complete cds.	34.2	55.8	20.8	46	0
RC_R62469_s_at	Human mRNA for KIAA0849 protein, partial cds.	1157.9	1365.3	379.8	950.4	710.8
RC_AA040107_at	Incyte Unique	140.4	243.1	0	174.4	80
RC_R67283_s_at	Incyte Unique	166.6	92.9	0	147.2	32.4
H59417_s_at	Incyte Unique	161	290.7	71.1	299	30.4
AA147425_s_at	Human S164 gene, partial cds; PSI and hypothetical protein genes, complete cds and S171 gene, partial cds.	117.6	118	44.7	90	39.9
X15949_at	Human mRNA for interferon regulatory factor-2 (IRF-2).	128.4	254.4	75.3	180.1	0
M12959_s_at	Human T-cell receptor active alpha-chain mRNA from JM cell line, complete cds.	594.7	858	479.8	675.5	182.7
RC_H02848_s_at	Human receptor protein-tyrosine kinase (TEK) mRNA, complete cds.	222.2	324.2	43.3	154.7	76.9
RC_AA250736_at	Human mRNA for KIAA1268 protein, partial cds.	1971	1753.9	595.4	1562.6	801.4
RC_R63879_at	Human intergenic DNA between SURF-2 and SURF-4.	117.1	268.2	0	210.9	106.6

M12529_at	Human apolipoprotein E mRNA, complete cds.	870.6	948.1	365.8	515.3	314.1
RC_R63802_at	Human ring finger protein BAP-1 mRNA, complete cds.	99.4	201.9	51.8	117.4	94.7
RC_R65803_at	Incyte Unique	266.3	544.2	0	265.8	183.8
X16832_at	Human mRNA for cathepsin H (EC 3.4.22.16).	1416.2	1898.9	199.5	1210.2	495.3
X17042_at	Human mRNA for hematopoietic proteoglycan core protein.	612.8	1611.9	249.3	918.3	439.2
X16663_at	Human HS1 gene for hematopoietic lineage cell specific protein.	508.1	634.6	0	460.8	115.2
X16665_at	Human HOX2H mRNA from the Hox2 locus.	73.5	166.4	44.8	101.4	72.5
RC_AA039887_at	Incyte Unique	336.7	462.5	305	411.3	60.4
X15414_at	Human mRNA for aldose reductase (EC 1.1.1.2).	470	609.8	192.8	390.6	283.9
U66672_at	Human putative transcription factor CA150 mRNA, complete cds.	381.9	743.1	305.1	656.6	146.9
RC_AA421473_s	Incyte Unique	30.7	41.8	0	36.4	9.6
RC_R67996_at	Incyte Unique	177	154.9	0	122.5	68.6
RC_AA421011_at	Human vesicle trafficking protein (SEC22C) mRNA, complete cds.	203.6	226.4	56.1	219.9	58
RC_AA281796_at	Human SL15 protein mRNA, complete cds.	268.7	486.1	290.8	401.2	47.3
RC_AA281760_at	Incyte Unique	92.7	110.8	0	51.7	40.6
L42324_at	Human (clone GPCR W) G-protein-linked receptor gene (GPCR) gene, 5' end + cds.	64.7	56.4	0	49	12.3
RC_AA487297_at	Incyte Unique	455.8	475.3	245.5	290.6	181.5
W02027_s_at	Human PTD012 mRNA, complete cds.	31.4	45.4	0	24.4	13.1
RC_AA487252_at	Human guanylate binding protein isoform II (GBP-2) mRNA, complete cds.	128.3	182	0	81.8	12.6
RC_N59137_at	Incyte Unique	36	61	0	43	21.6
RC_N58561_s_at	Human cyclophilin-related protein (NKTR) gene, complete cds.	872.5	922.8	379.9	1073.7	446.1
RC_AA251003_at	Human mRNA for KIAA0697 protein, partial cds.	60.1	110.4	54	74.1	23.9
RC_AA176475_at	Human mRNA for KIAA0892 protein, partial cds.	465.3	755.1	403.4	645.5	276
W16684_at	Human mRNA for KIAA0276 gene, partial cds.	55.4	90.5	11.5	56.8	15.8
M16505_at	Human steroid sulfatase (STS) mRNA, complete cds.	27.6	47.4	0	23.8	9.3
RC_AA041535_at	Incyte Unique	435.1	438.1	114.4	466.5	151.6
W07097_at	Human tetraspan NET-4 mRNA, complete cds.	294.7	544.9	94.6	370.1	133.5
RC_F13642_at	Human HSPC055 mRNA, complete cds.	592.7	1037.4	307.8	828.2	538
RC_N58295_f_at	Human mRNA for KIAA0065 gene, partial cds.	45.5	93.5	30.8	80.6	45
RC_R60047_at	Human S164 gene, partial cds; PS1 and hypothetical protein genes, complete cds and S171 gene, partial cds.	73.7	65.4	0	101.5	26.1
RC_AA115535_at	Incyte Unique	110.2	149.5	62	129.8	66.6
X52151_at	Human arylsulphatase A mRNA, complete cds.	491.5	526	213.2	851.4	217.2
RC_AA486072_i	Human CC chemokine gene cluster, complete sequence.	911.1	1550.7	525.3	1117.9	552.8
M15059_at	Human Fc-epsilon receptor (IgE receptor) mRNA, complete cds (H107 epitope)	100.3	124.5	0	123.5	22.2
RC_R61293_at	Human 959 kb contig between AML1 and CBR1 on chromosome 21q22, segment 3/3.	121	130.4	25.7	145.7	43.3
RC_AA115461_at	Human mRNA; cDNA DKFZp564B102 (from clone DKFZp564B102); partial cds.	97.3	153.1	74	132.6	73
RC_AA179602_at	Human CpG island DNA genomic MseI fragment, clone 15e10, reverse read cpg15e10.rt1b.	239.5	271.7	183.9	184.7	42
U95822_at	Human clone 23777 putative transmembrane GTPase mRNA, partial cds.	253.3	335.2	109.6	381	179.7
RC_R60512_s_at	Human mRNA for KIAA0191 gene, partial cds.	97.3	154.8	0	43.1	47.9
RC_AA040699_at	Human mRNA for Net transcription factor.	53.8	145.5	27.6	100.5	30.2
RC_AA481526_at	Human F-box protein FBL5 mRNA, partial cds.	174.4	355	120.2	305.4	129.7
RC_N54991_at	Incyte Unique	140.2	132.9	77	152.9	19.2
RC_R60192_s_at	Human peroxisome targeting signal 2 receptor (Pex7) mRNA, complete cds.	85.4	167.4	54.9	104.2	59.2
RC_AA180208_at	Human CGI-32 protein mRNA, complete cds.	200.1	294.7	79.7	284.3	101.2
W01094_at	Human melanoma-associated antigen MG50 mRNA, partial cds.	163.8	209.2	100.5	95.9	49.8
RC_N57964_at	Human CCR6 chemokine receptor (CMKBR6) gene, complete cds.	90.2	131.9	0	89.3	20.2
X13956_at	Human 12S RNA induced by poly(rI), poly(rC) and Newcastle disease virus.	164.9	234.4	0	269.5	123.7
X60003_s_at	Human transactivator protein (CREB) mRNA, complete cds.	158.4	263	0	264.4	62.5
RC_F10779_at	Human mRNA for KIAA0615 protein, complete cds.	332.5	412	107.1	338.9	131.7
RC_AA121543_at	Human mRNA for KIAA0758 protein, partial cds.	227.4	394.7	0	180.1	144.2
M13792_at	Human protein kinase inhibitor gamma (PKIG) mRNA, complete cds.	291.7	268.7	88.3	298.8	88
X62744_at	Human RING6 mRNA for HLA class II alpha chain-like product.	1180.7	1246.8	236.6	1103.2	27.4
X03100_cds2_at	Human HLA-SB(DP) alpha gene.	1710	2788.4	1031.2	1710.3	969.1
T59929_s_at	Human ICB-1 mRNA, complete cds.	149.9	288.9	71.8	124	34.2

L23808_at	Human metalloprotease (HME) mRNA, complete cds.	412.4	838	0	358.6	95.3
X03068_f_at	Human mRNA for HLA-D class II antigen DQw1.1 beta chain.	1541	2191.2	0	1589.1	459.4
X63337_at	Human HB2A gene for high sulfur keratin.	70.4	185.3	0	124.1	41.4
AA094999_at	Human zinc finger protein 216 splice variant 2 (ZNF216) mRNA, complete cd	302.4	353.2	181.1	392.2	158.6
RC_AA417067_at	Human F-box protein Fbx22 (FBX22) gene, partial cds.	115.8	150.8	41.6	170.3	24.9
RC_AA282473_s_	Human mRNA for KIAA0083 gene, partial cds.	205.3	181.5	31.3	216.5	99.4
RC_R46036_at	Human clone 24665 mRNA sequence.	52.8	44.5	0	39.7	21.1
RC_R45656_i_at	Incyte Unique	40.4	48.3	0	64.7	18.6
RC_AA188981_at	Human retinoblastoma-associated protein HEC mRNA, complete cds.	26.9	27.6	0	38.3	16.1
X62534_s_at	Human HMG-2 mRNA.	140.2	381.9	38.8	260.4	192.6
RC_R80664_at	Incyte Unique	209.7	154.6	0	158.9	81.4
RC_AA490814_at	Incyte Unique	57	92.4	16	22.9	0
RC_AA121879_s_	Human DMA, DMB, HLA-Z1, IPP2, LMP2, TAP1, LMP7, TAP2, DOB, DQB and RING8, 9, 13 and 14 genes.	2438.2	3249.6	1433.4	2640.9	1078.7
L25270_at	Human XE169 mRNA, complete cds.	338.6	247	151.4	319.1	6.9
X04011_at	Human mRNA of X-CGD gene involved in chronic granulomatous disease located on chromosome X.	117.1	272.4	59.6	161.5	22.7
RC_AA417102_at	Human MUC18 gene exon 16.	193.5	240.3	54.8	213.3	93.3
L25085_at	Human Sec61-complex beta-subunit mRNA, complete cds.	616.1	1185	93.5	857	672.3
RC_AA490261_s_	Human DNA, CpG island, clone R27-2.	177.4	162.7	0	285.4	78.2
X03934_at	Human gene for 20K T3 glycoprotein (T3-delta-chain) of T-cell receptor complex.	392.8	455.1	0	383.2	119.7
RC_AA251428_at	Human HOM-TES-103 tumor antigen mRNA, complete sequence.	222.4	339.2	99.1	195.8	75.8
X03663_at	Human mRNA for c-fms proto-oncogene.	42.8	120.5	0	73.5	6
X62055_at	Human PTP1C mRNA for protein-tyrosine phosphatase 1C.	521.7	559.5	171.9	524	340.6
RC_AA251352_r_	Human HeLa mRNA isolated as a false positive in a two-hybrid-screen.	182.6	191.8	43.6	148.1	100.6
RC_AA479082_at	Human HMT-1 mRNA for beta-1,4 mannosyltransferase, complete cds.	92	149.2	0	72.9	0
RC_AA034499_s_	Human mRNA for ZNF198 protein.	180.5	311.8	0	256.2	82
RC_AA282521_at	Human HRIHFB2157-like protein mRNA, partial cds.	126.9	230.9	115.3	217.5	94
W36290_s_at	Human MAFB/Kreisler basic region/leucine zipper transcription factor (MAFE mRNA, complete cds.	135.7	190.1	139	131.1	0
RC_AA165520_at	Public Unique	58.7	104.6	27.5	65.2	9.6
X02317_at	Human mRNA for Cu/Zn superoxide dismutase (SOD).	1848.8	3095.7	1340.4	2937.9	1469.4
W52821_at	Human leucine aminopeptidase mRNA, complete cds.	203.1	394.3	72.5	225.5	71.3
RC_AA251561_at	Incyte Unique	191.1	233.6	85.4	119	31.7
RC_AA425753_at	Human CpG island DNA genomic MseI fragment, clone 58h11, reverse read cpg58h11.rt1a.	364.7	425.2	224.3	377	171.8
RC_F09845_at	Incyte Unique	993	1404	0	381.3	230.7
RC_F09926_at	Human clone 23773 mRNA sequence.	34.7	42.7	0	33.7	8.5
RC_AA490894_at	Human mRNA for calpastatin, complete cds.	648.6	1020	502.1	947.6	535.9
AA165144_f_at	Human mRNA for KIAA0997 protein, complete cds.	35.1	41.2	0	44.4	14
M28130_ma1_s_at	Human gene for LUCT/interleukin-8, complete cds.	45.1	56.1	0	41.9	5.5
RC_AA102652_at	Human mRNA; cDNA DKFZp434K0926 (from clone DKFZp434K0926).	616.7	587.1	190.5	483.5	337.6
RC_AA478727_at	Incyte Unique	918.5	488.8	0	744.3	0
L10125_s_at	Human ALK-4 mRNA, complete cds.	21.4	29.4	0	41.1	0
RC_AA416886_at	Human mRNA; cDNA DKFZp564C1563 (from clone DKFZp564C1563).	115.2	217.7	64.5	237.5	24.2
RC_F10024_at	Incyte Unique	106.7	148.2	46.6	174.5	43.5
RC_AA251555_at	Human full length insert cDNA clone ZD86E06.	98.6	101.2	47.6	118	54.6
W47082_at	Novel Human mRNA from chromosome 1, which has similarities to BAT2 genes.	62.1	84.9	17	34	21.3
H03686_f_at	Human GAP SH3 binding protein mRNA, complete cds.	233.8	418.5	163.9	331	92.9
RC_R82055_at	Human clone HQ0149 PRO0149 mRNA, complete cds.	179.5	134.2	43.1	286.1	79.4
AA094989_at	Human voltage dependent anion channel protein mRNA, complete cds.	153.3	208.1	0	149.3	71
RC_AA282571_at	Human FRG1 (FRG1) gene, complete cds; 5S ribosomal RNA gene, complete sequence; TUB4q and TIG2 pseudogenes, complete sequence.	114.2	162.9	74.6	109.5	67.6
T54762_s_at	Incyte Unique	23.6	20.9	6.7	24.1	14.5
RC_AA121974_at	Public Unique	45.4	72.6	0	78.1	0
M27436_s_at	Human placental tissue factor (two forms) mRNA, complete cds.	321.3	669.4	107.6	402.4	188.6
RC_AA236950_at	Human HSPC302 mRNA, partial cds.	156	174.9	85.5	139.6	35.4
AA094744_at	Human RRM RNA binding protein Gry-rbp (GRY-RBP) mRNA, complete cds	158	105.9	37.7	157.8	35.7
L20298_at	Human transcription factor (CBFB) mRNA, 3' end.	204.5	502	0	314.5	184.1

RC_R79413_at	Incyte Unique	188.1	143.7	0	204.1	73.1
M25280_at	Human lymph node homing receptor mRNA, complete cds.	184.7	251.5	0	131.4	0
RC_AA169226_at	Incyte Unique	21.2	59.3	28.8	46.2	9.8
RC_AA187634_at	Human eukaryotic translation initiation factor eIF3, p35 subunit mRNA, complete cds.	87.4	93.2	29.8	52.1	22.7
AA436291_at	Human mRNA for KIAA1073 protein, complete cds.	28.9	38.4	0	40.2	8.2
RC_AA237018_at	Human constitutive fragile region FRA3B sequence.	426.6	720	188.7	481.8	410.9
RC_H09533_at	Incyte Unique	1575.7	1670.8	1011.8	2684.8	525.8
RC_AA187490_at	Incyte Unique	89.8	212.3	94.4	129.9	39.3
X61123_at	Human BTG1 mRNA.	627.5	1142.5	358.6	937.3	475.7
U03644_at	Human recepim mRNA, complete cds.	126.5	142.3	47.3	115.2	27.8
AA156897_s_at	Human mRNA; cDNA DKFZp564I1922 (from clone DKFZp564I1922); partial cds.	770.3	766.1	146	484.9	161.5
RC_AA489668_at	Human HZF6 mRNA for zinc finger protein.	61.9	115.5	0	50.4	41.2
T89072_at	Human mRNA for KIAA1268 protein, partial cds.	90.3	103.3	0	145.6	0
RC_AA237014_at	Human mRNA; cDNA DKFZp564I112 (from clone DKFZp564I112).	145.5	220	75.6	148	92.8
RC_AA417568_s_.	Human mRNA; cDNA DKFZp566D244 (from clone DKFZp566D244); partial cds.	439.7	529.5	133.7	580.9	74.2
RC_R49046_s_at	Human clone 24410 ABC transporter mRNA, partial cds.	497.5	545.1	129.6	516.8	125.4
W32305_f_at	Human silencing information regulator 2-like protein (SIR2L) mRNA, complete cds.	670.2	1342.8	432.6	417.3	166.7
AA436202_at	Human brain secretory protein hSec10p (HSEC10) mRNA, complete cds.	117.7	221.6	55.3	219.1	90.1
W28747_at	Incyte Unique	51.7	63.8	0	48.2	10.4
X06272_at	Human mRNA for docking protein (signal recognition particle receptor).	161.4	297.3	77.3	215.6	135.4
RC_AA490069_at	Incyte Unique	16.2	33.7	17.7	24.2	0
RC_AA425128_at	Incyte Unique	60.9	111.5	0	40.9	14.2
W28362_at	Human mRNA for KIAA0974 protein, partial cds.	136.5	188.1	43.6	126.8	79.4
RC_AA479727_s_.	Incyte Unique	690.8	646	328.2	508.8	133.5
RC_R49295_at	Incyte Unique	24.4	43.1	0	19.9	9.9
AA156925_at	Incyte Unique	30.7	108.2	18.5	66	28.2
W28106_at	Incyte Unique	262.4	222.9	44.2	289.7	139.6
RC_N52908_at	Incyte Unique	54.8	112.7	0	97.2	60.1
RC_AA237034_at	Human full length insert cDNA clone ZB96E04.	199.5	229.3	118.9	322.8	63.4
RC_AA282405_at	Human inducible 6-phosphofructo-2-kinase/fructose 2,6-bisphosphatase (IPFK 2) mRNA, complete cds.	227	536.1	79.9	245.3	178.9
D50840_at	Human mRNA for ceramide glucosyltransferase, complete cds.	1429.3	1144.8	1190.2	1585.5	1667

## A5 Appendix for Chapter 5 (Discussion of Microarray Data)

## A5.1 The general decrease of metabolic pathway genes in UCi tissues

From the initial array experiments (before the CD samples were available), it was noticed that many of the genes showing reduced expression in the UCi samples compared to the NI samples were part of a metabolic pathway. This appendix summarises the results of the investigation to amalgamate this gene expression data. Table A5.1 lists the genes that showed decreased expression in 3 or more UCi samples. The SRS portal was used to access pathway databases to enable identification of the main pathways each enzyme has a role in. Many protein complexes are made up of more than one peptide and reduction of more than one chain per enzyme was seen in many cases.

Table A5.1 – Genes decreased in 60% or more of the UCi samples compared to the NI samples

Decreased in n / 5 UCi samples	Gene Acc No*	Gene Name	Enzyme ID	Metabolic Pathways
5	L05144	Phosphoenolpyruvate carboxykinase	4.1.1.32	Citric acid cycle Pyruvate metabolism
4	L12760			Glycine/Threonine/Serine metabolism Arginine/Proline metabolism Histidine metabolism Tyrosine metabolism Phenylalanine metabolism Tryptophan metabolism β-Alanine metabolism Alkaloid biosynthesis II
4	U11862	Diamine Oxidase	1.4.3.6	Arginine/Proline metabolism β-Alanine metabolism
4		γ-aminobutyraldehydedehydrogenase	1.2.1.19	Oxidative phosphorylation
4	X15822	COxVIIa – L Liver specific cytochrome c oxidase	1.9.3.1 (complex 4)	Synthesis & degradation of ketone bodies Valine/isoleucine/leucine degradation Butanoate metabolism
4	X83618	3-hydroxy-3-metabolismhylglutarylcoenzyme A synthetase	4.1.3.5	Oxidative phosphorylation
4	Z14244	Cytochrome C oxidase subVIIb COxVIIb	1.9.3.1	Citric acid cycle Lysine degradation
3	D26535	Dihydrolipoamide succinyltransferase	2.3.1.61	Specific electron acceptor for several dehydrogenases
3	J04058	Electron transfer flavoprotein α-subunit		Oxidative phosphorylation (complex III)
3	X71129			Oxidative phosphorylation (complex III)
3	J04973	Cytochrome bc-1 complex core protein II	1.10.2.2	Glycolysis Fatty Acid metabolism Bile acid biosynthesis Tyrosine metabolism Glycerolipid metabolism
3	L32977	Rieske iron-sulphur protein (of bc1 complex)	1.10.2.2	Urea cycle Arginine/Proline metabolism
3	M12963	Class I alcohol dehydrogenase (ADH1)	1.1.1.1	Oxidative phosphorylation (complex III)
3	M16364	Creatine kinase	2.7.3.2	Urea cycle Arginine/Proline metabolism
3	M19961	Cytochrome c oxidase subunit Vb	1.9.3.1	Oxidative phosphorylation (complex III)
3	M22348	Ubiquinonebinding complex (part of cytochrome c reductase complex) (ie complex III)	1.10.2.2	Urea cycle Arginine/Proline metabolism
3	M29927	Ornithine aminotransferase gene	2.6.1.13	Oxidative phosphorylation
3	M37104	mitoATPase coupling factor 6 subunit (ATPSA)	3.6.1.34	Glutamate Alanine metabolism Cysteine metabolism Arginine/Proline metabolism Tyrosine metabolism
3	M37400	Cytosolic aspartate aminotransferase mRNA	2.6.1.1	

\* When a gene was represented by more than one accession number and both were selected, they are shown in the same row in this table.

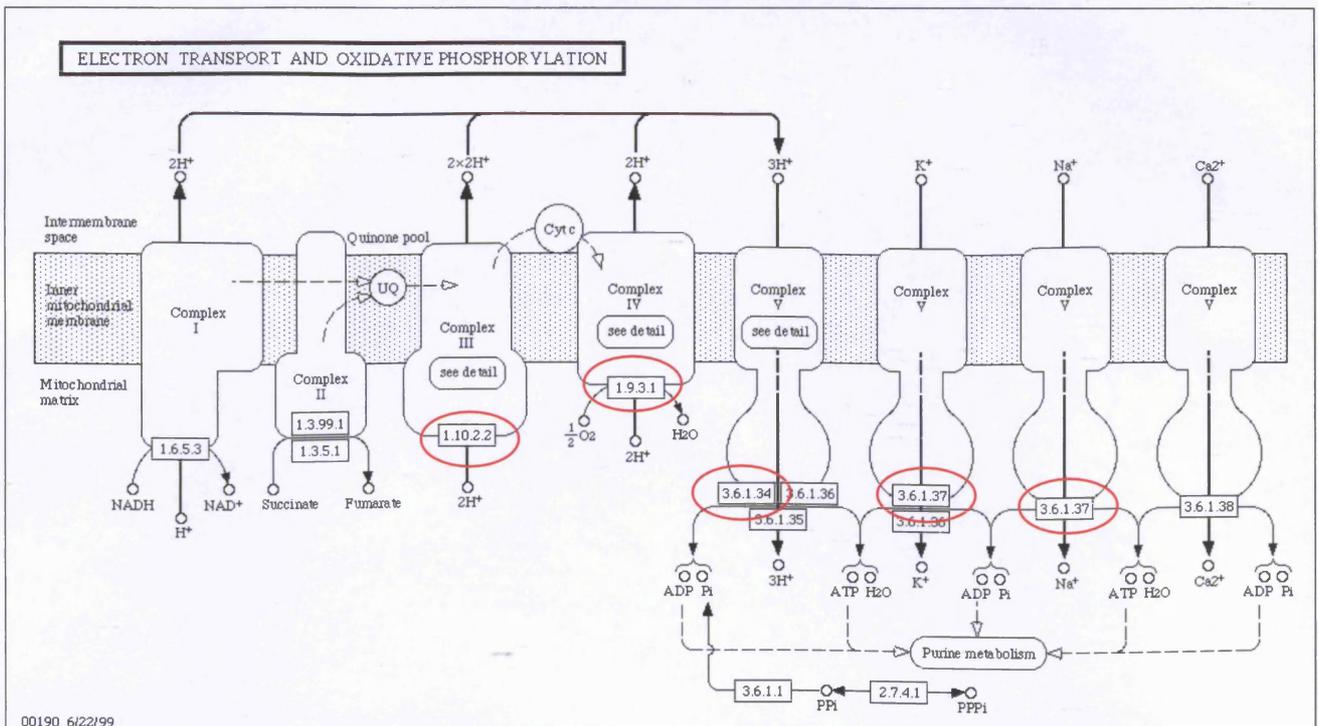
				Phenylalanine/Tyrosine/Tryptophan biosynthesis (carbon fixation) Alkaloid biosynthesis I
3	M37457	Na <sup>+</sup> ,K <sup>+</sup> - ATPase catalytic subunit $\alpha$ -III isoform	3.6.1.37	Oxidative phosphorylation
3	M61832	s-adenosylhomocysteine hydrotase (AHCY)	3.3.1.1	Methionine metabolism Selenoamino acid metabolism
3	M68840	Monoamine oxidase A (MAOA)	1.4.3.4	Glycine/Serine/Threonine metabolism Arginine/Proline metabolism Histidine metabolism Tyrosine metabolism Phenylalanine metabolism Tryptophan metabolism
3	M90516	Glutamine:fruc-tase-6-phosphate amidotransferase (GFAT)	2.6.1.16	Glutamate metabolism Aminosugers metabolism
3	M91432	Mediumchain acyl-CoAdetrydogenase (MCAD)	1.3.99.3	Fatty acid oxidation Valine/Leucine/Isoleucine degradation $\beta$ - alanine metabolism Propanoate metabolism
3	U09813	Mitochondrial ATP synthase subunit 9	3.6.1.34	Oxidative phosphorylation
3	U27460	Uridine disphosphoglucose pyrophosphatylase	2.7.7.9	Pentose/Glucuronate Galadose metabolism Starch/Sucrose metabolism Nucleotide sugar metabolism
3	U90915	Cytochrome c oxidase subunit IV	1.9.3.1	Oxidative phosphorylation
3	X13238	Cytochrome c oxidase subunit VIc	1.9.3.1	Oxidative phosphorylation
3	X14813	3-oxoacyl-coa thiolase	2.3.1.16	Fatty acid synthesis (path2) Fatty acid metabolism Bile acid biosynthesis Valine/Leucine/Isoleucine degradation Phenylalanine metabolism
3	X69141	Squalene synthase	2.5.1.21	Sterol biosynthesis Terpenoid biosynthesis
3	X76228	H <sup>+</sup> ATPase E subunit	3.6.1.34	Oxidative phosphorylation
3	Y00339	Carbonic anhydrase II	4.2.1.1	Nitrogen metabolism
3	Z68204	Succinyl CoAsynthase	6.2.1.4	Citric acid cycle Propanoate metabolism

Figures A5.1 and A5.2 illustrate the mapping of genes that are decreased onto metabolic maps of the citric acid cycle and oxidative phosphorylation respectively. Pathways were only highlighted if at least 3 genes in the pathway appear in table A5.1. Here the example of the citric acid cycle and oxidative phosphorylation have been used, but many other pathways were also represented but do not appear in figure A5.1.

Many of these pathways would be decreased in defunctioned colon too, so may not be specific to ulcerative colitis. However, there is some evidence to support the decreased activity of some pathways, for example the production of ketone bodies has been reported as inhibited in UC mucosa[Jorgensen, 2001 #524], and the diminished activity of the fatty acid oxidation pathway as discussed in chapter 5.



Figure A5.2 – The oxidative phosphorylation pathway



Enzymes that were decreased in 3 or more UCi samples are circled in red