Gene expression profiling of colorectal tissues in the early adenoma-carcinoma sequence

Thesis submitted for the Degree of

Doctor of Philosophy

at the University of Leicester

By

Katie Mervyn Thomas BSc (Hons) (Leicester)

Department of Cancer Studies and Molecular Medicine

University of Leicester

December 2011

Gene expression profiling of colorectal tissues in the early adenoma-carcinoma sequence

Katie Mervyn Thomas

ABSTRACT

Colorectal cancer (CRC) is the third most common cancer worldwide. Colorectal polyps are pre-cursors of CRC; however hyperplastic polyps (HPs) lack malignant potential. The aim of this thesis was to describe differences in gene expression and pathway activation between colorectal tissues (normal mucosa and polyp) from early stages of CRC development and to validate novel candidate genes identified by qRT-PCR.

Differential gene expression was investigated in 48 colorectal tissues from the early stages of the adenoma-carcinoma sequence using DASL Whole-Genome expression microarrays and appropriate bioinformatics. Particular emphasis was placed on the comparison between Adenomatous polyps (APs) and HPs as lesions with and without malignant potential. In the comparison between HP and AP tissues 1633 significantly differentially expressed genes (DEGs) ($p<0.05$) and 33 pathways were identified, which confirms the fundamental differences between these polyps. Moreover, DEGs associated with Wnt-Signalling, MAPK Signalling, p53 Signalling, cell cycle and apoptosis were noted between HPs and APs. In addition, a novel network was created using COXPRESSdb, which found connections between genes comparing HP and AP tissues.

Six candidate genes were selected based on their differential expression across the range of colorectal tissues; *ASCL2, ANXA2, AXIN2, ETS2, G3BP1* and *TFF2*. qRT-PCR was employed to investigate expression of these candidate genes in a larger Validation Set (n=143) of colorectal tissues. With the exception of *TFF2,* significant differential expression was identified for all genes, which supported the results of the DASL microarray. Again the most significant differences identified were between HPs and APs.

In conclusion, HPs and APs have different malignant potential, with associated differential gene expression profiles, which could be exploited for screening and to provide potential therapeutic candidates.

ACKNOWLEDGEMENTS

Firstly, I would like to say a huge thank you to my supervisors, Dr. Jacqui Shaw and Dr. Howard Pringle. Thank you for giving me the opportunity to undertake this PhD under your supervision. Without your input and support I would not be where I am today. I would also like to thank Dr. Kevin West for his help, support and advice throughout this project.

I am very grateful to all of the laboratory staff (past and present) on the 3rd floor of the RKCSB, including Angie Gillies, Lindsay Primrose, Janine Moreton, James Brown and Dr. Karen Page. Thank you for your support and input over these last few years. In particular I would like to thank Linda Potter for all of her hard work, guidance, support and dedication in helping me succeed. Throughout the problems and successes your support was invaluable.

My sincerest thanks go to Medisearch for providing me with financial support during my degree. I would also like to thank Mrs. Anne Peppitt for her help in finding this financial support. In addition I would like to thank the Leicester Medical School for allowing me to take time out to undertake this Intercalated PhD.

I would like to say a warm thank you to my family; Mum, Dad, Amma, Appa, Sarah, Louise, Acha, Uncle Ian, Auntie Marlene and Uncle Frank. Your support has never wavered and your encouragement is what has helped me to get to where I am today. Thank you for all of the support and love that you have given me.

Finally I would like to say a very special thank you to my husband, Mervyn. You were the person who persuaded me to undertake the BSc all those years ago. Who knew that I would end up doing a PhD! You have always supported me every step of the way. Thank you for your friendship, love and support. You really are my better half.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Chapter 3

Chapter 4

Chapter 5

LIST OF TABLES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

LIST OF ABBREVIATIONS

ACF	Aberrant Crypt Foci
AMV	Avian Myeloblastosis Virus
AN	Normal colonic mucosa from patients with adenomatous polyps
ANOVA	Analysis of variance
ANXA2	Annexin A2
APC	Adenomatous Polyposis Coli
AP	Adenomatous Polyp
ASCL2	Achaete-scute complex homolog 2
AXIN2	Axin2
BLAST	Basic Local Alignment Search Tool
BRAF	V-raf murine sarcoma viral oncogene homolog B1
cDNA	Complimentary DNA
CIMP	CpG Island Methylator Phenotype
CIN	Chromosomal Instability
COSMIC	Catalogue of Somatic Mutations in Cancer
CRC	Colorectal Cancer
Ct	Cycle Threshold
CTNNB1	Beta-catenin
DASL	cDNA-mediated Annealing, Selection, Extension and Ligation
DAVID	Database for Annotation Visualization and Integrated Discovery
dCt	Delta Cycle Threshold
DEG	Differentially expressed gene
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphates
EC	Endogenous control
EGF	Epidermal growth factor
EGFR	EGF Receptor
ERK	Extracellular signal-related kinase
ETS2	V-ETS erythroblastosis virus E26 oncogene homolog 2
FAP	Familial Adenomatous Polyposis
FDR	False discovery rate
FF	Fresh frozen
FFPE	Formalin fixed paraffin embedded
G3BP1	GTPase activating protein (SH3 domain) binding protein 1
GAPDH	Glyceraldehydes-3-phosphate dehydorgenase
GEO	Gene expression Omnibus
H and E	Haematoxylin and Eosin
HCL	Hierarchical Cluster
HP	Hyperplastic Polyp
IAA	Isoamyl alcohol
KN	Normal colonic mucosa from patients with colorectal polyps and CRC
KP	Cancer-associated colorectal polyp
KRAS	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
LOH	Loss of heterozygosity
MAPK	Mitogen-activated protein kinase
MEV	MultiExperimental Viewer
miRNA	Micro Ribonucleic acid
MMR	Mismatch Repair
mRNA	Messenger ribonucleic acid

MSI	Microsatellite Instability
NC	Normal control
NHS-BCSP	National Health Service Bowel Cancer Screening Programme
NTC	No template control
OG	Oncogene
PCA	Principal component analysis
PCR	Polymerase chain reaction
PK	Proteinase K
PPAR	Peroxisome proliferator-activated receptors
QC	Quality control
qPCR	Quantitative PCR
qRT-PCR	Quantitative reverse transcriptase PCR
RNA	Ribonucleic acid
RPL13a	Ribosomal protein L13a
rpm	Revolutions per minute
RQ	Relative quantification
RT	Reverse Transcription
RTP	Reverse transcription primer
SAM	Significance Analysis of Microarrays
SSA/P	Sessile Serrated Adenoma/Polyp
TA	Tubular Adenoma
TFF2	Trefoil Factor 2
TME	Total meso-rectal excision
TPT1	Tumour protein, translationally-controlled 1
TNM	Tumour Node Metastasis
TSA	Traditional Serrated Adenoma
TSG	Tumour Suppressor Gene
TVA	Tubulovillous Adenoma
UBC	Ubiquitin C
UMFIX	Universal Molecular Fixative
UP	Ultrapure
VA	Villous adenoma
VST	variance stabilizing transformation
WT	Wild-type
5-FU	fluoropyrimidine 5-FU

CHAPTER 1: INTRODUCTION

1.1. CLINICAL FEATURES OF COLORECTAL CANCER (CRC)

1.1.1. EPIDEMIOLOGY AND RISK FACTORS ASSOCIATED WITH CRC

Colorectal Cancer (CRC) is the third most common cancer in men and the second most common cancer in women in the United Kingdom (UK) (CRUK, 2011). In 2008 there were approximately 40,000 new cases and 16,000 CRC related deaths (CRUK, 2011).

The risk factors associated with CRC are largely associated with "Western" diet and lifestyle; with higher risk associated with diets low in fibre, high in fat and red meat, and physical inactivity (Stein and Colditz, 2004, Markowitz and Winawer, 1997, Boyle and Langman, 2000). It has been suggested that 50% of all sporadic cancers can be prevented by modifying diet and lifestyle choices (Stein and Colditz, 2004).

Epidemiological studies also support the link with Western diet and lifestyle. Studies in Japan have shown an increasing incidence of CRC, which correlates with Japan's recent westernization and provides evidence to reinforce the causal link between CRC risk and a "Western" diet and lifestyle. The incidence of CRC in Japan now mimics that of the West (Benson, 2007, Goh, 2007). Additional studies have been undertaken in Israel, where the incidence of CRC is directly related to differences in diet and lifestyles between different ethnic populations (Center et al., 2009). Furthermore, studies of migratory information have shown that individuals who migrate from low to high risk CRC countries/regions adopt an increased risk of CRC within one generation of living in the new region due to changes in diet and lifestyle (Center et al., 2009, Boyle and Langman, 2000).

The incidence of CRC is also increasing in countries undergoing economic development, which tend to adopt a more "Western" lifestyle. Conversely,

economically stable countries are showing stabilized and/or reduced incidences of CRC (Center et al., 2009). This stabilization could be due to the development of greater health awareness and the introduction of screening programmes. However, overall, the incidence of CRC remains higher in the developed world.

Additional risk factors for CRC include Inflammatory Bowel Disease (IBD), family history/hereditary predisposition, advancing age and the presence of adenomatous polyps (APs). These APs represent a pre-malignant precursor to CRCs, providing an opportunity to identify and remove potential CRCs before malignant transformation. Lynch Syndrome and Familial Adenomatous Polyposis (FAP) are hereditary conditions that have an increased risk of CRC development and account for approximately 5% of all CRC cases (Jasperson et al., 2010, Weitz et al., 2005). Both conditions are associated with AP formation prior to CRC development, providing evidence that APs are pre-malignant lesions.

1.1.2. CRC SCREENING

In recent years CRC screening programmes have been introduced to reduce the incidence of CRC by removing pre-cancerous lesions and to reduce mortality by detecting earlier stage cancers. In the United States, the incidence of CRC has declined over recent years, which is thought to be due to the introduction of the CRC screening programme in the late 1990s (Center et al., 2009).

The NHS Bowel Cancer Screening Programme (NHS-BCSP) started in July 2006 with a successful pilot study and by 2010 the NHS-BCSP was nationwide. The NHS-BCSP screening programme targets individuals aged 60-75, who are at greater risk of developing CRC. The late age of presentation of sporadic CRC is due to an

accumulation of genetic events over time (Ballinger and Anggiansah, 2007, Benson, 2007). Therefore, the identification of APs offers a window of opportunity to detect and remove the risk of CRC, as APs tend to form 5-10 years prior to the development of CRC (Leslie et al., 2002, Scholefield, 2000).

1.1.3. SYMPTOMS AND DIAGNOSIS OF CRC

Symptoms of CRC include changes in bowel habit, rectal bleeding, abdominal pain, weight loss and fatigue; these are often non-specific (Schofield and Jones, 1992a) and associated with conditions other than CRC. Unfortunately in the UK the majority of CRC patients present at later stages of the disease, with associated poorer survival (Ballinger and Anggiansah, 2007) hence the UK has a relatively high mortality rate of CRC. With steps to improve public awareness of the risk factors of CRC, and with the implementation of the NHS-BCSP, the incidence and mortality are expected to decrease with time.

Size, location, type, grade and stage are all used to help determine CRC treatment and prognosis. Tumour grade and stage are of particular importance as they reflect how advanced the cancer is. Tumour grade is a measure of tumour differentiation, however inter-observer variation exists when classifying tumours as being poor, moderately or well differentiated, making this grading system unreliable and of little use in predicting prognosis (Chandler and Houlston, 2008).

Tumour stage is a measure of tumour spread and histopathologists use a combination of the traditional Dukes' system (Dukes, 1949), and the more modern Tumour Node Metastasis (TNM) staging (Greene et al., 2002). Originally Dukes' staging had three categories; A, B and C. However in practice Dukes' Stage D is often used to describe

tumours that have metastasized. Table 1.1 shows the TNM and Dukes' guidelines for CRC staging. Overall CRCs are divided into four stages.

Table 1.1: Overview of Dukes' and TNM staging systems

TNM stage	T	N	M	Dukes' Stage	Description
1	1	0	0	A	Tumour invades submucosa. No nodal involvement. No distant metastasis.
	2	0	0		Tumour invades muscularis propria. No nodal involvement. No distant metastasis.
2	3	0	0	B	Tumour invades subserosa. No nodal involvement. No distant metastasis.
	4	0	0		Tumour invades into other organs or structures. No nodal involvement. No distant metastasis.
3	1,2,	1,	0	C	Tumour invades muscularis propria. 1 to 3 lymph nodes involved. No distant metastasis.
	3,4,	2,	0		Tumour invades subserosa. 4 or more lymph nodes involved. No distant metastasis.
4	Any	Any	1	D	Distant metastasis identified.

T is for tumour and is a measure of tumour invasion. T1 is the lowest level of invasion with the tumour being restricted to the submucosa. T4 is the highest level of invasion, past the serosa and into other organs and structures. N0, N1 and N2 refer to 0, less than 4 and 4 or more lymph nodes involved respectively. M stands for metastasis with M0 tumours showing no signs of metastasis while in M1 cancers there are distant metastases identified. (Weitz et al., 2005, Ballinger and Anggiansah, 2007).

1.1.4. TREATMENT AND SURVIVAL OF CRC PATIENTS

Tumour staging is important in establishing the best course of treatment for each patient who requires an individual treatment plan (Young and Rea, 2000). The type of treatment depends on a variety of factors such as grade and stage of the tumour, tumour size and location, patient age and the general health of the patient. Surgical resection of the tumour, chemotherapy and radiotherapy are the most widely used therapies.

Surgery will most likely be performed on the majority of all CRC patients. Unfortunately this form of treatment can result in a local recurrence of the cancer despite the removal of all visible tumour cells (McArdle, 2000a). The type of surgery undertaken and the surgeon that performs it can influence the risk of local recurrence

(McArdle, 2000b, Schofield and Jones, 1992b, Dorrance et al., 2000). A new surgical technique known as a total meso-rectal excision (TME) (Heald and Ryall, 1986) has recently been introduced for middle and lower third rectal cancers to help reduce recurrence rates to less than 5% (Ridgway and Darzi, 2003).

The majority of CRC chemotherapy regimens include the use of fluoropyrimidine 5- FU (5-FU) both in adjuvant and palliative settings (Winder and Lenz, 2010). 5-FU is often coupled with other chemotherapy agents to increase patient response. As a single agent in the treatment of advanced CRC, 5-FU has a response rate of 20-25% (Winder and Lenz, 2010). A combination of chemotherapy and radiotherapy is widely used as a pre-surgical treatment for rectal cancer as studies have suggested that this helps to reduce local recurrence of the cancer (Lee et al., 2002a). This neo-adjuvant therapy helps to reduce the size of the tumour making it easier for surgeons to excise completely.

Additional therapies are now being evaluated such as anti-epidermal growth factor receptor (anti-EGFR) treatment, for example cetuximab. Recent studies have shown that the presence of *KRAS* (V-Ki-Ras2 Kirsten Rat Sarcoma Viral Oncogene Homolog) gene mutations is a negative marker for patient response to cetuximab, predicting resistance to anti-EGFR treatment and a poorer patient prognosis (Lievre et al., 2006, Winder and Lenz, 2010, Karapetis et al., 2008). Interestingly, anti-EGFR therapy is also ineffective in 50% of *KRAS* wild-type tumours. Other studies have shown that CRCs with *BRAF* (V-Raf Murine Sarcoma Viral Oncogene Homolog B1) mutations also show a poorer response to anti-EGFR treatment. (Winder and Lenz, 2010). These findings support the theory that tumours that do not respond to anti-EGFR treatment could

have activating mutations in *BRAF*/*PIK3CA* or other unknown mutations (Baldus et al., 2010). These findings highlight the importance of knowing the genotype of individual CRCs to ensure an appropriate treatment plan.

Table 1.2 shows the survival statistics for CRC according to Dukes' stage. An increased stage at presentation is associated with a much poorer survival rate. However, following the introduction of the NHS-BCSP, CRCs should present at earlier stages and hence increase survival. The identification and removal of precancerous polyps by the NHS-BCSP coupled with novel surgical techniques and chemotherapy advances, should further improve the survival of all stages of CRC.

Table 1.2: CRC survival rates according to stage

Dukes' Stage	5 year survival rates (%)
A	80-95
B	65-75
C	25-60
D	0-7

This table shows the 5 year survival rates as a percentage according to the stage of the disease. Patients presenting later are more likely to have more advanced disease and therefore poorer survival rates than those presenting at an earlier stage. Figures adapted from Weitz et al., 2005

1.2. TURNOVER AND MAINTENANCE OF THE COLORECTAL MUCOSA

The colonic mucosa is replaced frequently, every few days (Salama and Platell, 2009). This constant shedding of the colonic mucosa necessitates a large turnover and differentiation of new healthy colonic cells.

Stem cells, unlike normal cells can accumulate mutations over a long period of time before any phenotypic changes are recognized (Kim et al., 2004, Kim and Shibata, 2002). It is postulated that CRC could be a stem cell disease due to the presence of cells within cancers, which have the capability to generate new primary tumours.

These subpopulations of cancer cells are termed cancer stem cells (Pinto and Clevers, 2005).

The crypts of the large intestine contain a small subset of stem cells within a niche environment, which provides optimum surroundings for stem cell homeostasis (Spradling et al., 2001). These adult stem cells have the capability to produce differentiated cell types that are required for the normal functioning of the large bowel. The adult stem cells reside at the base of the colonic crypts and divide slowly, thus reducing the rate of DNA mutations (Potten et al., 1997). Their division produces an undifferentiated stem cell, which will replace the parent stem cell, and a transit amplifying cell that will undergo mitosis at a much higher rate as it travels to the surface of the lumen and becomes terminally differentiated (Figure 1.1) (McDonald et al., 2006).

If CRC is a disease that initiates in the stem cells, it is logical to conclude that the mutations required for adenoma formation and subsequent carcinogenesis occur in the stem cells at the base of the crypt. The “bottom-up” theory (Preston et al., 2003) suggests that the mutations necessary for transformation are present in the stem cells and that they undergo expansion in an upwards direction until the whole crypt is occupied.

Aberrant Crypt Foci (ACF) or microadenomas can be identified microscopically from as few as one abnormal crypt (Cheng and Lai, 2003) and subsequently develop into macroscopically identifiable adenomas. ACF represent an interim phase between normal colonic mucosa and colorectal adenomas. The theory of a single crypt being

affected and then undergoing expansion into macroscopic adenomas is a theory that fits with clinical findings (Preston et al., 2003).

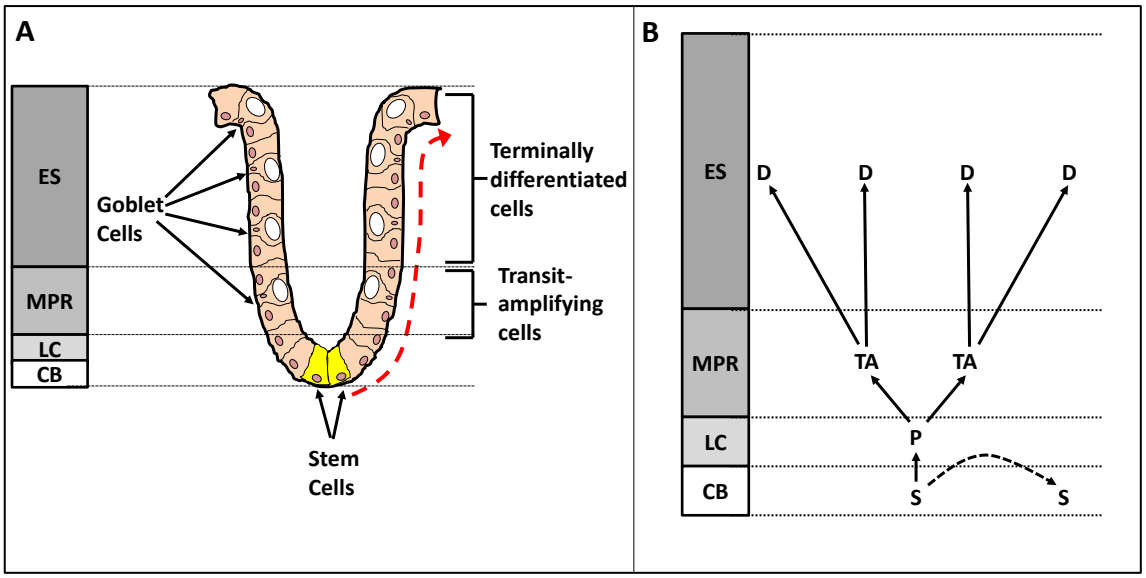

Figure 1.1: Stem cell division and crypt structure in the large intestine.
A: The structure of a normal colonic crypt where the stem cells are located at the base of the crypt. The red line indicates the direction of cell movement from the undifferentiated stem cells to the differentiated cells on the epithelial surface. B: The type of cell division that occurs in the crypt. Stem cells divide to produce one daughter stem cell to maintain the stem cell population, and one progenitor cell. The progenitor cell divides to produce two daughter transit-amplifying cells, which undergo regular mitosis and produce differentiated cells that end up at the surface of the crypt. ES, Epithelial Surface; MPR, Midpoint Region; LC, Lower Crypt; CB, Crypt Base; S, Stem Cell; P, Progenitor Cell; TA, Transit-Amplifying Cell; D, Differentiated Cell. Based on McDonald *et al.*, 2006 and Salama and Platell, 2009.

1.3. COLORECTAL POLYPS (ADENOMAS)

Colorectal adenomas are growths occurring on the surface of the colonic epithelium that are associated with an increased risk of CRC. They are considered to be potential pre-cursors of CRC (Risio, 2010, Markowitz and Winawer, 1997). The majority of colorectal adenomas remain as benign lesions, with only a small proportion (5%) progressing to malignancy (Eide, 1986, Boyle and Leon, 2002, Winder and Lenz, 2010). Worldwide prevalence rates of adenomas vary geographically and mimic the incidence rates of CRC (Markowitz and Winawer, 1997, Clark et al., 1985, Correa et al., 1977). As with CRC, migrants who move from low risk regions to westernized societies develop

increased rates of adenoma formation (Correa, 1978). Additionally, as with CRC, diet and lifestyle factors appear to be associated with an increased risk of adenoma development (Giovannucci et al., 1992).

Colorectal adenomas are common in the general population but older individuals have an increased risk (Rickert et al., 1979). The majority of adenomas and CRCs appear to have a left-sided distribution (O'Brien et al., 1990, Shinya and Wolff, 1979, Matek et al., 1986).

Histologically adenomas can be broadly classified into two categories: neoplastic and non-neoplastic (Markowitz and Winawer, 1997). Adenomatous polyps (APs) are neoplastic in nature and exhibit a malignant potential. Hyperplastic polyps (HP's) are examples of non-neoplastic polyps, which are considered to be benign lesions.

1.3.1. ADENOMATOUS POLYPS

APs were initially linked to increased CRC risk in patients with the hereditary condition FAP, which is due to a germline mutation of the *APC* (Adenomatous Polyposis Coli) gene, and results in the development of multiple polyps throughout the colon (Fearnhead et al., 2001, Wu et al., 1998, Nugent et al., 1994). Individuals with FAP inevitably develop CRC without treatment (Boland and Goel, 2005). The removal of APs in patients with FAP removed the risk of CRC and hence APs were identified as potential pre-cursors to CRC (Boland and Goel, 2005). Additional studies supported this finding as removal of APs in the general population also reduces the risk of CRC (Winawer et al., 1993, Lau and Sung, 2010).

Histologically APs (neoplastic) are classified into three groups; Tubular (TA), Tubulovillous (TVA) and Villous (VA). TA's are the most common type accounting for 75-87% of APs, whereas VA's are the least common accounting for 5-10% of cases (Muto et al., 1975, O'Brien et al., 1990). These adenomas require greater than 75% of the polyp to show a villous architecture. TVA's show a mix of both tubular and villous architecture but need 25-75% of the polyp to have a villous appearance for diagnosis as a TVA (Markowitz and Winawer, 1997).

The size of the adenoma correlates with its histology. The majority of TAs (75%) are smaller than 10mm whereas VAs tend to be larger, with 60% being greater than 20mm (Markowitz and Winawer, 1997, Muto et al., 1975). As increased diameter, especially over 20mm, is associated with an increased risk of tumorigenesis (Risio, 2010) it is logical to conclude that VA's are also associated with an increased risk of CRC development since the majority of these adenomas are larger (>10mm) in size (O'Brien et al., 1990).

Each AP has neoplastic potential. However, the vast majority of APs remain as benign lesions, with only a minority progressing to malignancy. This property means that each adenoma has a degree of dysplasia that is classified histologically as low, moderate or high grade. High grade dysplasia is associated with a higher chance of malignant transformation (Winawer et al., 1993). High grade dysplasia is associated with larger adenomas, increased patient age and villous architecture. The combination of these factors therefore provides an increased risk of CRC (Muto et al., 1975).

1.3.2. SERRATED POLYPS

Serrated polyps represent a family of polyps that differ histologically from APs, due to their 'saw-tooth' morphology, which could result from decreased apoptosis in these polyps (Parfitt and Driman, 2007, Tateyama et al., 2002) The serrated polyp family encompasses HPs, sessile serrated adenomas/polyps (SSA/P), traditional serrated adenomas (TSA) and mixed hyperplastic/adenomatous polyps (MPs). HPs represent the most common type of serrated polyp, comprising 80-90% of all serrated polyps while TSA and MPs are the least common (Ensari et al., 2010).

HPs are histologically distinct from adenomatous type polyps. They represent non-neoplastic lesions that are believed to have no metastatic potential (Bond, 2000, Lau and Sung, 2010, Risio, 2010, Winawer et al., 1990, Bensen et al., 1999, Bauer and Papaconstantinou, 2008). The majority of HP's are smaller than 5mm with only a small proportion exceeding 15mm (Markowitz and Winawer, 1997).

Larger HPs were previously associated with an increased risk of CRC. However, it is now recognized that these larger lesions represent a histologically distinct type of serrated polyp; SSA/Ps. SSA/Ps were identified during investigation of HPs removed from patients with hyperplastic polyposis. This condition is characterized by large numbers of HPs within the colon. The histology of these HPs showed subtle differences to conventional HPs, being more similar to TSAs (Harvey and Ruszkiewicz, 2007). However, further analysis revealed a more sessile configuration resulting in a new classification of serrated polyp, the SSA/P. The identification and definition of SSA/Ps is relatively new and it is estimated that 8.3% of SSA/Ps have previously been

misdiagnosed as HPs (Risio, 2010). SSA/Ps have an increased risk of malignancy via the serrated pathway (Snover, 2011) described in 1.4.3.

SSA/Ps are thought to represent 8-20% of all serrated polyps (Ensari et al., 2010). They are larger in size than HPs, tending to be greater than 5mm in diameter. Strict histological criteria have recently been published to differentiate between HPs and SSA/Ps (Ensari et al., 2010, Higuchi and Jass, 2004) due to the increased risk of CRC associated with SSA/Ps. A review of the literature revealed limited research into conventional HPs, with most research concentrating on SSA/Ps and the serrated pathway. However, the lack of malignant transformation of HPs makes them attractive targets for research. The differences existing between HPs and APs could provide insight into new therapeutic targets.

1.4. SPORADIC VS FAMILIAL CRC

Sporadic and familial CRC arise from distinct carcinogenic pathways (Figure 1.2), but all are thought to arise from pre-cancerous colorectal polyps. FAP and Lynch Syndrome are the most well described familial forms of CRC, arising through germline mutations in *APC* (Suppressor Pathway) (Fearnhead et al., 2001) and DNA mismatch repair (MMR) genes (Mutator Pathway) (Lynch and Smyrk, 1996, Lynch et al., 1997) respectively. Sporadic cancers can arise through either the Suppressor Pathway or Serrated Pathway.

The Suppressor pathway gives rise to two-thirds of sporadic CRCs (Jass, 2007, Fearon and Vogelstein, 1990). These sporadic CRCs develop due to acquired mutations/loss of *APC*.

The Suppressor pathway accounts for 60% and the Mutator Phenotype accounts for 5% of all CRCs (Snover, 2011). These CRCs develop from APs. The Serrated Pathway accounts for one third of sporadic cancers (Snover, 2011), arising from sessile serrated adenomas/polyps (SSA/Ps) instead of APs.

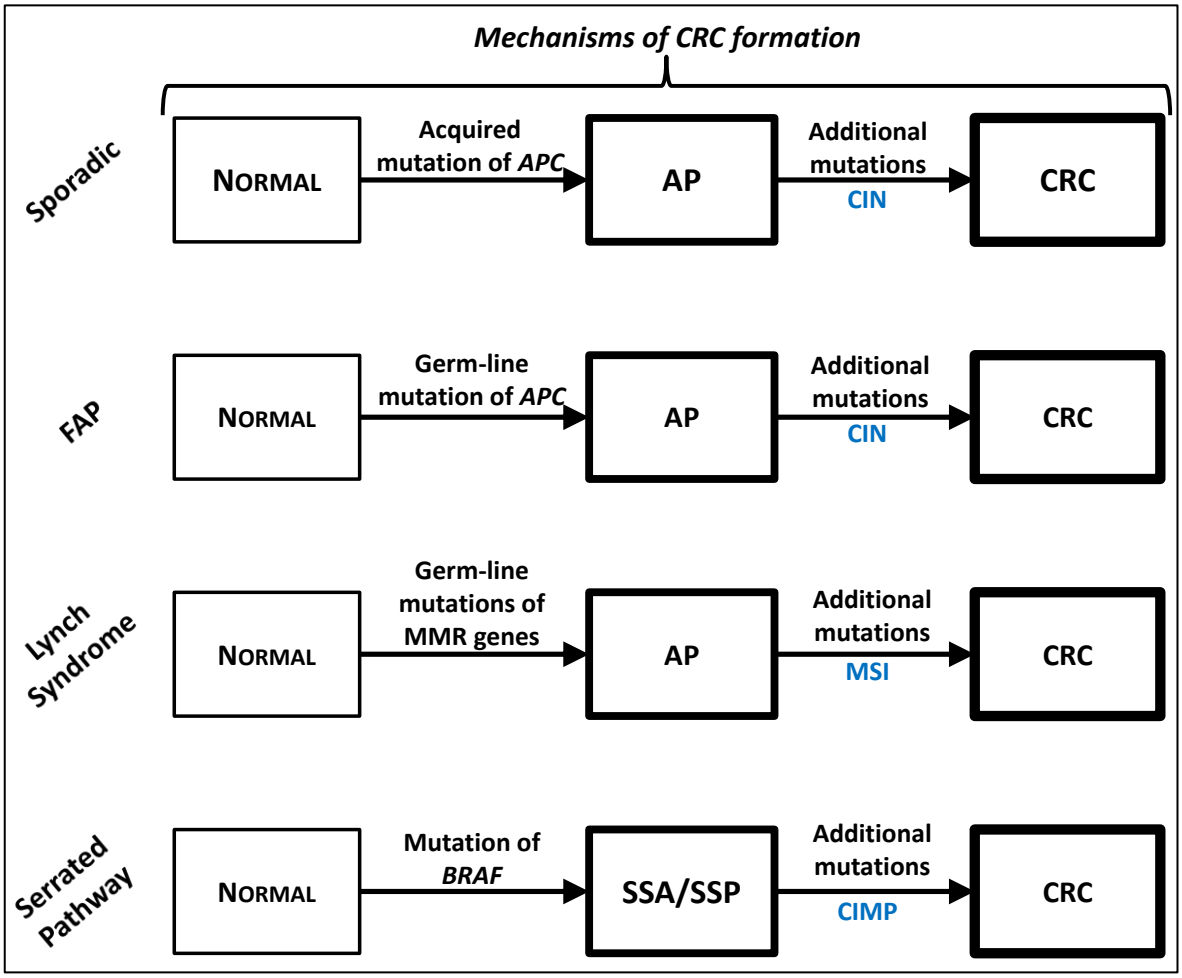

Figure 1.2: Mechanisms of CRC formation in sporadic and familial forms of CRC
Sporadic CRC and FAP arising from APs develop according to the Suppressor pathway, with initiating mutations occurring in *APC.* Although developing through the same pathway, FAP is more aggressive and occurs at a younger age due to the germ-line mutation in *APC* and earlier development of multiple APs. Both of these cancers are associated with CIN. Lynch Syndrome develops via the Mutator Pathway, with germ-line mutations of the MMR genes. These cancers are associated with MSI. The Serrated Pathway arises through initiating mutations in *BRAF,* leading to the development of SSA/Ps instead of APs. These cancers are associated with CIMP. Abbreviations: CIN, Chromosomal Instability; MSI, Microsatellite Instability; CIMP, CpG Island Methylator Phenotype; MMR, mismatch repair genes.

1.4.1. THE SUPPRESSOR PATHWAY

The adenoma-carcinoma sequence outlined in Figure 1.3 shows the stepwise accumulation of genetic events over time in the suppressor pathway, for both sporadic and FAP-associated CRCs.

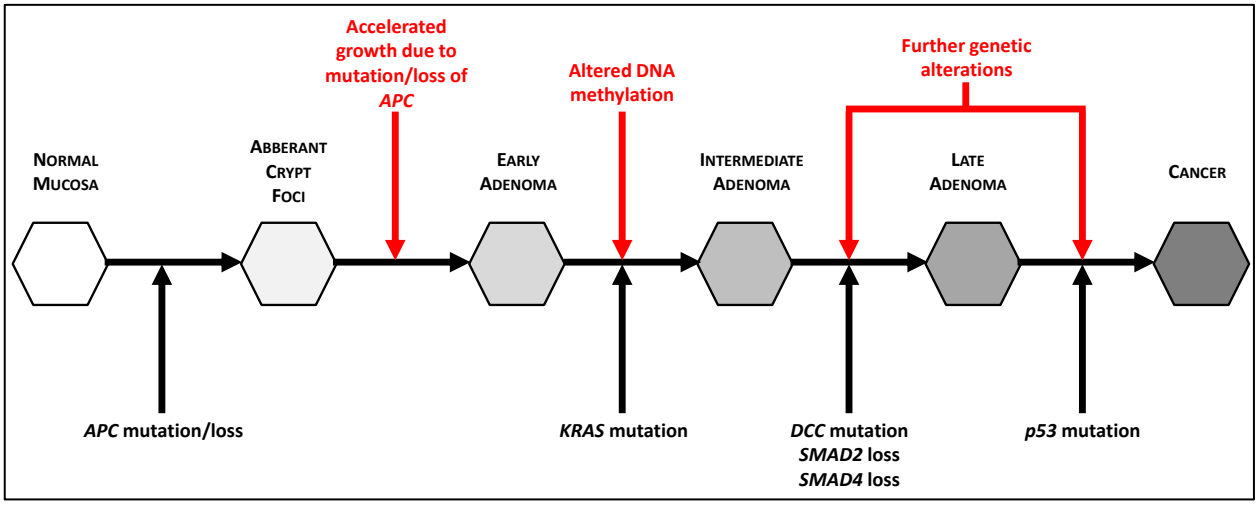

Figure 1.3: The Adenoma-Carcinoma Sequence
This pathway shows the phenotypical steps that occur during the development of CRC. In addition, mutations and genetic/epigenetic changes are also shown, along with the developmental stages associated with them. This sequence of events relates only to APs, since SSA/Ps develop through the distinct serrated pathway.

Mutation or loss of the tumour suppressor *APC* (5q) is the initiating mutation (Fearon and Vogelstein, 1990) in FAP, and *APC* is also mutated in approximately 80% of sporadic suppressor pathway CRCs (Fearnhead et al., 2001). In the absence of Wnt Signalling, *APC* is a member of the "β-catenin (*CTNNB1*) destruction complex" along with Axin and glycogen synthase kinase 3β (GSK-3β). Binding of *CTNNB1* to this complex initiates GSK-3β-mediated phosphorylation, which marks *CTNNB1* for degradation (Munemitsu et al., 1995). Mutations in *APC* prevent its incorporation into this complex, thus inhibiting the degradation of *CTNNB1* and allowing its accumulation in the cytoplasm. *CTNNB1* subsequently enters the nucleus and interacts with TCF, resulting in changes in gene expression of target genes (Mann et al., 1999). The

oncogene *C-MYC* is one of the target genes activated and overexpressed via this pathway, which leads to uncontrolled cell growth (Leslie et al., 2002). In addition to the regulation of *CTNNB1*, *APC* also has a role in cell-cell adhesion, cell cycle regulation and apoptosis (Senda et al., 2007). These findings provide additional evidence that the loss of *APC* plays a big role in the development of CRC.

Other mutations along this sequence include mutations or losses of *KRAS*, *DCC*, *SMAD2*, *SMAD4*, and *p53*. These losses and mutations are often accompanied by widespread DNA hypomethylation and focal promoter hypermethylation (Leslie et al., 2002). Although this pathway exhibits some promoter hypermethylation, CRCs arising through this pathway are CpG Island Methylator Phenotype (CIMP) negative (Snover, 2011). APs and CRCs exhibit a multitude of mutations and genetic abnormalities. The loss of tumour suppressor genes (TSGs) and the presence of activating mutations of oncogenes (OGs) is commonly seen (Winder and Lenz, 2010). Most of these events are surplus/coincidental and are unnecessary for carcinogenesis (Boland and Goel, 2005). It is the accumulation of genetic events over time and not the order in which they occur that appears to be the most important factor in colorectal tumorigenesis (Fearon and Vogelstein, 1990), with each mutation providing its own growth advantage and clonal expansion.

Chromosomal Instability (CIN) occurs in 65% of sporadic CRCs and is associated with global hypomethylation (Pino and Chung, 2010, Gopalakrishnan et al., 2008). During normal cell division, the daughter cells should receive the same chromosomal content as that of the parent cell. However, in the presence of CIN the rate of chromosomal loss, gain and rearrangement is increased. Cancer cells are often aneuploid (having an

abnormal number of chromosomes), and show a loss of heterozygosity (LOH). The presence of CIN could be due in part to the lack of regulation of cell check points due to defects in regulatory pathways or genes. Whether CIN occurs as an initiating mechanism of tumorigenesis, or in response to other abnormalities remains a debate.

1.4.2. THE MUTATOR PHENOTYPE/PATHWAY

The mutator phenotype (seen in Lynch Syndrome) is initiated by germline mutations in DNA MMR genes, which are responsible for repairing mutations of microsatellites. Individuals with Lynch Syndrome are not only at an increased risk of developing CRC, but due to the nature of their mutation, they are also at risk of developing other malignancies (Wheeler et al., 2000, Mecklin et al., 1986).

Microsatellites are short simple sequences of DNA, which are repeated throughout the genome. These repeats vary in length between individuals but are usually less than 10 DNA base pairs in length. Under normal conditions, mutations of microsatellites would be recognized and repaired by MMR genes.

In Lynch syndrome the lack of functioning MMR genes results in Microsatellite Instability (MSI) (Boland and Goel, 2005, Strand et al., 1993, Papadopoulos et al., 1994, Prolla et al., 1994), as accumulating mutations in microsatellites go unrepaired and escalate in severity by interfering with adjacent genes. Germline mutations in *hMLH1* and *hMSH2* (mutS homolog 2, colon cancer, nonpolyposis type 1) are the most common initiating mutations (90%) (Kinzler and Vogelstein, 1996) in Lynch syndrome. However other mutations are also linked to this disease such as *hMSH6* (mutS homolog 6) and *hPMS2* (PMS2 postmeiotic segregation increased 2) (Kinzler and Vogelstein, 1996, Jasperson et al., 2010). As with the suppressor pathway, additional

mutations occur due to the absence of adequate DNA MMR resulting in the formation of colorectal polyps.

1.4.3. THE SERRATED PATHWAY

The serrated pathway is thought to be initiated with a *BRAF* mutation, which results in the formation of a serrated lesion (Harvey and Ruszkiewicz, 2007, Jass et al., 2006). When mutated, *BRAF* causes a lack of normal apoptosis and subsequent increase in cell growth and division. Serrated lesions are susceptible to the methylation of CIMP regions and the majority of cancers derived from this pathway are considered to be CIMP+ (Snover, 2011). Localised areas of hypermethylation are present at CpG islands of promoter genes, leading to transcriptional silencing (Strathdee and Brown, 2002, Herman et al., 1994, Boland and Goel, 2005, Esteller, 2005, Boland et al., 2009). Promoter hypermethylation is a candidate mechanism for the down-regulation/epigenetic silencing of well known TSG's in CRC, such as *p53*.

Although random gene silencing exists, a common gene to be silenced in this group of CRCs is *hMLH1.* Silencing of this gene leads to MSI. These cancers subsequently develop numerous additional mutations at an increased rate due to the lack of DNA repair. These cancers are often described as CIMP+MSI+.

1.5. GENETIC EVENTS AND PATHWAYS ASSOCIATED WITH NEOPLASTIC PROGRESSION

1.5.1. THE EPIDERMAL GROWTH FACTOR (EGF) PATHWAY

Activation of the Epidermal Growth Factor (EGF) pathway results in increased cell survival by inhibiting apoptosis and increasing cell proliferation via a wide spectrum of downstream interactions. The EGF receptor (EGFR) is currently a hot topic of research

and the target for recent drug therapies; anti-EGFR monoclonal antibodies. Mutations in EGFR and its downstream targets are commonly seen in cancer (Winder and Lenz, 2010). *KRAS*, *BRAF* and *PIK3CA* are downstream targets of EGFR and are commonly mutated in cancer.

When EGFR is activated by ligand binding, there is phosphorylation of important tyrosine kinases. Subsequent protein interactions ensue followed by the activation of downstream signalling pathways including the RAS/extracellular signal regulated kinase (ERK) pathway and PI3K/AKT pathway. Interaction between the different signalling networks promotes cell growth and survival. Prolonged activation of EGFR via activating mutations or mutation of one or more of downstream effector molecules results in increased cell proliferation and reduced apoptosis (Figure 1.4) (Fakih and Vincent, 2010). By targeting EGFR in CRC, these properties for tumour survival can be reduced.

KRAS mutations are found in all stages of adenomas and cancers (Velho et al., 2008). Mutations of this gene typically occur in codons 12, 13 or 61, with mutations in 12 and 13 being the most common (Vakiani and Solit, 2011). Normal KRAS protein acts as a molecular on-off switch, however when mutated KRAS is locked in an active state due to a loss of its intrinsic GTPase activity. Activated KRAS signals to downstream effectors such as Erk/MAPkinase and PI3K/Akt pathways resulting in changes to apoptosis and proliferation (Vakiani and Solit, 2011). Mutations in *KRAS* have been associated with large and aggressive APs and CRCs (Monticone et al., 2008).

The *BRAF* gene encodes for the BRAF protein, which is involved with processes related to cell growth and regulation of the Erk/MAPkinase pathway. Mutations in the *BRAF*

gene prevent normal regulation by the BRAF protein, which proves to be advantageous for APs and CRCs. *BRAF* mutations are more commonly seen in serrated adenomas and those tumours exhibit MSI, however they have been identified in all stages of AP and CRC development. Mutations usually occur at residue 600 where there is a change of amino acid from Valine to Glutamate, commonly referred to as V600E. Mutations at this location account for the majority of *BRAF* mutations (Carragher et al., 2010, Deng et al., 2004).

The presence of *KRAS*, *BRAF* or *PIK3CA* mutations in polyps suggests that they are at higher risk of tumorigenesis (Velho et al., 2008). Identifying these mutations earlier would be helpful in predicting those patients that are at increased risk of CRC and thus enable proactive screening in these individuals. Associations between mutation status and gene expression could also provide insight into other potential predictive/therapeutic targets, which could be used in individuals who cannot receive anti-EGFR therapy.

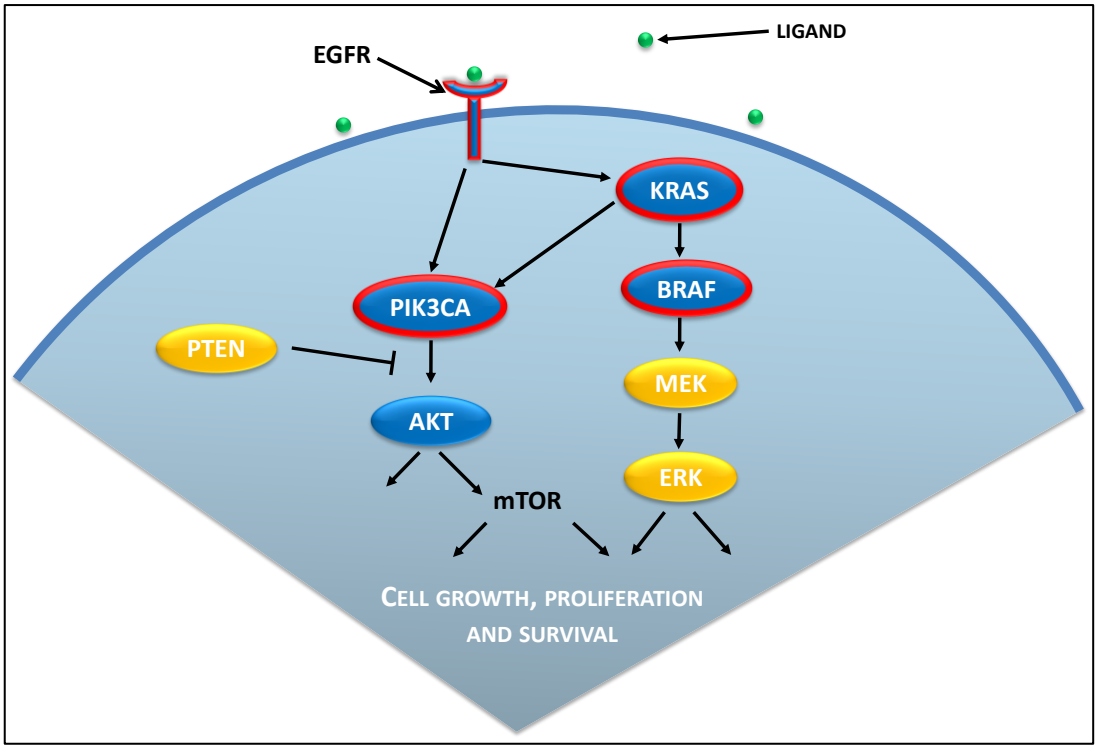

Figure 1.4: EGFR pathway and interacting signalling pathways
The red outlines in this figure indicate common sites of gene mutation. By targeting the EGFR pathway, it is possible to prevent the effects of activating mutations in *KRAS, BRAF* and *PIK3CA*.

1.5.2. MAPK PATHWAY

Mitogen-activated protein kinases (MAPK) belong to a large family of serine-threonine kinases, and form cell-proliferating pathways. There are three major MAPK subfamilies: Extracellular-signal-regulated kinases (ERK MAPK, Ras/Raf1/MEK/ERK), c-jun N-terminal/stress-activated protein kinases (JNK or SAPK) and MAPK14 (p38). The most widely described MAPK subfamily is ERK MAPK (Sebolt-Leopold, 2000). This pathway is important for cellular proliferation (Troppmair et al., 1994) differentiation (deFazio et al., 2000, Taupin and Podolsky, 1999) and protecting against apoptosis (Lewis et al., 1998). Key growth factors and proto-oncogenes promote growth and differentiation through this pathway (Taupin and Podolsky, 1999). Activation of this pathway is involved with pathogenesis, progression and oncogenic behaviour of CRC (Wang et al., 2004). Multiple targets associated with the MAPK pathway are currently being investigated as potential anti-cancer targets (Sebolt-Leopold, 2000). The JNK and

p38 pathways are also associated with cellular proliferation and differentiation, however JNK pathway has pro-apoptotic functions (Soh et al., 2001) while p38 has an anti-apoptotic role (Liao and Hung, 2003).

The ERK MAPK pathway is activated by signals from protein kinase C or Ras, both of which bring about their effects via EGF. Either of these initiating signals leads to activation of Raf1, a proto-oncogene, which results in a cascade leading to the phosphorylation of MEK1/2 and ERK1/2. On activation, ERK enzymes then migrate to the nucleus to activate transcription factors such as c-Fos and Elk1. Some of these transcription factors are important regulators of genes associated with cell proliferation and anti-apoptosis. Hence activation of this pathway can cause increased expression of these genes and thus reduced apoptosis and increased proliferation. Short activation of ERK MAPK is associated with increased proliferation, whereas prolonged activation is associated with increased cellular differentiation (Yen et al., 1998). Dis-regulation of the MAPK pathway can therefore provide growth advantages during the development of CRC.

1.5.3. LOSS/MUTATION OF *TP53*

TP53 is a vital transcription regulator of genes associated with cell cycle regulation, promotion of apoptosis and restricting angiogenesis (Vousden and Prives, 2009). In addition, the p53 protein is involved with microRNA (miRNA) processing via interactions with Drosha (Suzuki et al., 2009).

Approximately 70% of CRCs have loss of heterozygosity (LOH) at 17p, the location of *TP53* (Vogelstein et al., 1988, Fearon and Vogelstein, 1990), with a somatic mutation occurring in the remaining allele (Baker et al., 1990). Mutation/loss of *TP53* appears to

be a later event in the development of CRC (Figure 1.3), since most adenomas do not possess *TP53* mutations (Baker et al., 1990).

1.6. COLORECTAL CANCER: EVIDENCE FROM ANIMAL MODELS

Mouse (*mus musculus)* models have provided insight into the development and treatment of both sporadic and familial CRC. Since most CRCs develop via an initiating mutation in *APC* most animal models have been generated to harbor different mutations of *Apc*. However, the development of the serrated pathway and its initiating mutation in *Braf* and not *Apc* has led to increasing studies into this pathway.

The first *Apc* mutant mouse was the Apc^{Min}*,* where *Min* (Multiple Intestinal Neoplasia) was identified in a group of mice undergoing random mutagenesis (Moser et al., 1990). Additional knock-out mice have been created with a variety of *Apc* mutations. The histology of the adenomas created in the different *Apc* mutant mouse models remain similar however the number of adenomas that develop vary significantly (Taketo and Edelmann, 2009). Apc^{Min} mice are associated with ~30 adenomas whereas $Apc^{\Delta 716}$ and Apc^{1638N} develop 300 and 3 adenomas respectively. These three mutations are all truncating mutations occurring at positions 850 (Apc^{Min}), 716 ($Apc^{\Delta 716}$) and 1638 (Apc^{1638N}). Apc^{1638N} mice are thought to be representative of the human condition FAP (Smits et al., 1998). Although in humans FAP is associated with large numbers of adenomas, in Apc^{1638N} mice the low adenoma burden allows the mice to live longer and develop more advanced adenomas and tumours. The presence of a mutation in *Apc* is sufficient to generate adenomas yet *Apc* mutations alone do not lead to the development of invasive cancer (Taketo and Edelmann, 2009).

It is thought that the lack of tumour progression could be due to diet and lifestyle differences and also due to a lack of additional mutations such as *KRAS, SMAD4* and *p53* that occur in human CRCs (Taketo and Edelmann, 2009). This hypothesis was partly answered when the addition of *Smad4* mutations to $Apc^{\Delta 716}$ mutant mice resulted in the development of locally invasive adenocarcinomas (Takaku et al., 1998). Due to the short lifespan of these mice, no metastases were seen. In addition, mouse models have also shown increased size, number and invasiveness of adenomas with both *APC* loss and *KRAS* mutation (Janssen et al., 2006, Sansom et al., 2006). A recent study in zebrafish suggests that following loss of *APC*, disregulation of C-terminal binding protein 1(Ct-BP1) is associated with initiation of adenoma formation, while *KRAS* mutations and nuclear localization of β-catenin are associated with adenoma progression and carcinoma formation (Phelps et al., 2009). This finding is corroborated by other studies finding increased levels of nuclear β-catenin in advanced adenomas, rather than in early or microadenomas (Anderson et al., 2002, Amos-Landgraf et al., 2007, Blaker et al., 2003).

Additional mutations have been introduced into mouse models including *β-catenin*, *Braf* (Pritchard et al., 2007) *,* and MMR genes (Taketo and Edelmann, 2009). CRC has not developed in the presence of single mutations in these mouse models, most likely because of their short lifespan (Taketo and Edelmann, 2009). However, as with APC, the ability to induce these mutations has provided information on the phenotype of adenomas that exhibit specific mutations. It has also enabled the development of candidate treatment options of CRC.

Recent *in vivo* mouse models have demonstrated that *Braf* mutations (^{V600E}*Braf)* result in the development of CRCs, which are reminiscent of serrated CRCs (Carragher et al., 2010). ^{V600E}*Braf* activates the Wnt and Erk pathways, which results in the development of hyperplastic crypts. These crypts remain in a senescent state until acquired epigenetic inactivation of $p16^{Ink4a}$, allowing subsequent development of serrated CRCs (Carragher et al., 2010).

It is reported that western style diets (high fat, low fibre) result in an increased number of polyp formation in mice (Yang et al., 1998, Mai et al., 2003). Additionally, exercise can reduce the number of polyps that develop (Mehl et al., 2005). This evidence from the mouse model reinforces the association between diet and lifestyle risk factors and development of adenomas.

Mouse models provide information on the natural progression of CRC in the presence of known mutations. However, the results obtained from mouse models need to be used appropriately due to the differences that exist between mice and humans. Adenomas in mouse models develop in the small intestine rather than the large intestine as in humans, and most of the adenomas do not develop into CRC (Taketo and Edelmann, 2009, Heijstek et al., 2005). This lack of cancer progression could be due to the short lifespan of mice, especially in the presence of multiple adenomas, and the inability of the CRC to develop in such a short amount of time (Edelmann and Edelmann, 2004, Taketo and Edelmann, 2009). It could also be due to the lack of additional mutations that are known to exist in humans, which prevents the adenomas from advancing further.

Overall mouse models provide useful information about the development of the disease and enable treatment options to be investigated in a controlled environment before patients are involved.

1.7. MOLECULAR CLASSIFICATION OF CRC

The three pathways described previously (Adenoma-Carcinoma Sequence, Serrated Pathway and Mutator Phenotype) provide researchers and clinicians with information on CRC development, progression and treatment. However, not all CRCs develop entirely through one pathway; a mixture of several pathways or features (CIN, MSI, CIMP) can exist, making the pathogenesis of CRC complex. In 2006 Jass *et al* suggested key molecular features of colorectal polyps (Table 1.3) (Jass et al., 2006). Subsequently in 2007 they went on to suggest five molecular subtypes of CRC, based on the type of genomic instability and the presence or absence of CIMP (Table 1.4) (Jass, 2007).

Table 1.3: Genetic features of colorectal polyps based on Jass *et al* (2006)

Feature	HP	SSA	AP
KRAS	negative	negative	M
BRAF	M	H	L/negative
CIMP	L	H	L/negative
MSI	NA	positive	S
CIN	NA	S	positive

This table was created based on a paper by Jass *et al* (2006). The information presented in this table highlights the molecular differences that exist between different types of colorectal polyps. L, Low; M, medium; H, high; NA, not applicable; S, stable.

Table 1.4: Genetic and pathological features of CRC subsets based on Jass *et al* (2007)

Feature	Group 1	Group 2	Group 3	Group 4	Group 5
MSI	H	S/L	S/L	S/L	H
CIN	S	S	positive	positive	S
CIMP	H	H	L	negative	negative
KRAS	negative	L	H	M	M
BRAF	H	M	negative	negative	Negative
APC	L/negative	l/negative	L	H	M
TP53	negative	L	M	H	L
origin	SSA	SSA	SSA/AP	AP	AP

This table was adapted from a table presented by (Jass, 2007). The information presented in this table identifies the diversity of CRCs and the combination of genetic and epigenetic events that can coexist. MSI, microsatellite instability; CIN, chromosomal instability; CIMP, CpG Island Methylation Phenotype; SSA, sessile serrated adenoma; AP, adenomatous polyp; H, high; M, medium; L, low; S, stable

From Table 1.3 and Table 1.4 it is clear that *BRAF* and *KRAS* are mutually exclusive mutations and SSAs have a higher frequency of *BRAF* mutations and CIMP than HPs and APs. In addition *APC* mutations occur more frequently in APs. However, data in Table 1.4 show that although the suppressor, mutator and serrated pathways to CRC are well documented, in reality the pathology of CRC is more heterogeneous.

The investigation of colorectal tissues belonging to the earlier stages of CRC development may provide useful insights into the underlying sequence of events involved in colorectal carcinogenesis. In particular, the comparison between HP and AP tissues could be particularly useful in understanding the genetic differences between polyps with and without malignant potential. This comparison is currently lacking in the literature.

1.8. GENE EXPRESSION ANALYSIS OF COLORECTAL TISSUES

Microarray technology has enabled researchers to investigate gene expression changes between different tissue types and diseases. The comparison between diseased and matched disease-free normal tissues has been utilized for CRC. Additionally, investigation into the adenoma-carcinoma sequence has also been

undertaken. A recent study in 2010 identified 463 probe sets that mark the progression from “normal” colorectal mucosa (n=60) to APs (n=72) and finally CRC (n=99) (Tang et al., 2010), supporting the hypothesis that there are gene expression changes occurring during the development of CRC. A limitation of this study is the use of cancer adjacent normal tissues instead of true normal colorectal tissue from healthy controls. Cancer adjacent normal tissues could undergo field effect changes (Jothy et al., 1996, Nonn et al., 2009), which would distort the results obtained when comparing with APs and CRC.

Investigations into specific cancer types such as serrated CRCs (Laiho et al., 2007) and mucinous adenocarcinomas (Kim et al., 2011b) have also identified lists of differentially expressed genes (DEGs), which help to differentiate between the different types of CRC. Although studies have been undertaken to compare the different types of CRC, there are currently no studies that have investigated gene expression changes between different types of colorectal polyp. This gap needs to be addressed as it is the changes in gene expression during the early phases of adenoma development that could reveal novel candidate therapeutic targets.

The comparison between benign (HPs) polyps and those with malignant potential (APs) is of interest. Only two microarray studies have been published that include analysis of HPs (Galamb et al., 2008b, Galamb et al., 2010). Both included HPs as a minor focus of their study but the results suggest that HPs and APs are distinguishable from one another using DEGs.

1.9. SUMMARY

CRC is an important disease with a complex development. The ability to understand the development of CRC would enable novel therapeutic interventions to be developed and implemented. Advances in research now allow whole genomes to be analysed and compared. The use of microarray technology for the comparison of gene expression profiles in different cohorts of patients and samples is particularly useful. In addition, the use of Pathway analysis software can help to identify pathways that are depressed/activated in different tissue types and can be used to suggest potential therapeutic targets.

Gene expression microarrays will be utilized in this thesis to differentiate between colorectal tissues representing the early stages of the adenoma-carcinoma sequence. In particular this thesis will concentrate on the differences between neoplastic (APs) and non-neoplastic tissues (HPs) as this comparison is likely to provide new insight into the genes associated with carcinogenic potential. Gene expression and pathway analysis will also be performed on the data in the hope of finding new insight into the development of adenomas and carcinomas of the large bowel.

1.10. AIMS AND OBJECTIVES

The aims of this thesis were to investigate differences in gene expression and pathway activation between neoplastic and non-neoplastic colorectal tissues from early stages of CRC development.

In Chapter Three the aim was to establish the best method of tissue fixation to preserve the quality and integrity of nucleic acids, for subsequent expression

microarray studies, while maintaining tissue architecture. Three methods of tissue fixation were compared; UMFIX, FFPE and FF.

The specific objectives were as follows:

- To establish the effect of different methods of tissue fixation on DNA (*GAPDH*) and RNA (RPL13a) quality using samples from the Fixation Set.
- To determine the effect of different methods of tissue fixation on tissue histomorphology.

In Chapter Four the aims were to investigate patterns of differential gene expression between different colorectal tissues, perform pathway analysis and explore the comparison between HPs and APs in detail. Prior to this, validation of the expression microarray is required.

Objectives:

- QC steps available in lumi will be utilized to ensure that the results of the microarray experiment are reliable, and identify outlier samples that require removal from the analysis.

- Replicate analysis of the repeated cases A190N and A190P will be performed to ascertain if the microarray data are reproducible.

- *In-silico* analysis will be used to confirm the gene expression changes seen on the microarray using a review of previously published data.

- Limma will be used to create top-tables of differential expression between different types of colorectal tissues.

- To investigate differential gene expression of genes between specific tissues and perform pathway analysis to identify pathways that show significant differential expression using limma, MEV and DAVID.
- To investigate specifically the differential gene expression and pathway associations between HP and AP tissues.

In Chapter Five, the aim was to perform gene expression profiling in different colorectal tissues by qRT-PCR to identify potential genetic markers of CRC development and progression.

Objectives:

- Identify a panel of genes showing differential expression across a range of colorectal tissues that will be selected for qRT-PCR analysis using a range of colorectal tissues.
- Identify associations between candidate gene expression and clinico-pathological parameters such as adenoma type, mutation status and polyp location.

CHAPTER 2: MATERIALS AND METHODS

2.1. ETHICAL APPROVAL AND TISSUE COLLECTION

2.1.1. ETHICAL APPROVAL

Tissues collected prospectively were obtained with informed consent and local ethics approval. Retrospective tissues were obtained from the Leicester Royal Infirmary (LRI) tissue bank under approval of the local ethics committee.

Patients scheduled for CRC surgery were sent patient information leaflets, outlining the details of this study, one week prior to their surgery. Informed consent was then obtained from patients who wished to be involved with the study on their arrival at the hospital one day prior to the surgery. All patients who consented to the study had the opportunity to withdraw their consent at any point.

2.1.2. EUKARYOTIC CELL LINES

All cell lines were purchased from the American Type Tissue Culture Collection (ATCC, Rockville, MD., USA) with the exception of SW626, SW837 and GP2d. DNA from the SW626, SW837 and GP2d cell lines was purchased from the Health Protection Agency Cultures Collection (HPA Cultures, Salisbury, UK). Details of the cell lines utilized can be found in Table 2.1.

Table 2.1: Details of cell lines used in this study

Cell Line	Gender	Age	Type	Grade	Clinical description
SW626	F	46	CRC	Stage 3	Human ovarian metastasis of a primary colorectal adenocarcinoma (Furlong et al., 1999) first isolated in 1974.
SW837	M	53	CRC	Stage 4	Human rectal adenocarcinoma first isolated in 1976 (Leibovitz et al., 1976).
GP2d	F	71	CRC	Dukes' B	Human colorectal adenocarcinoma derived from the same adenocarcinoma as GP5d.
LOVO	M	56	CRC	Stage 4	Human supraclavicular metastasis from a primary colorectal adenocarcinoma which was first isolated in 1971.
SW480	M	50	CRC	Dukes' B	Human primary adenocarcinoma of the colon described in 1976.
HCT116	M	adult	CRC	n/k	Human colorectal cancer of unknown stage occurring in an adult male of unknown age.

2.1.3. IDENTIFICATION AND COLLECTION OF NORMAL AND COLORECTAL TISSUES

All colorectal tissues were reviewed by a consultant histopathologist (Dr. KP West, LRI) prior to inclusion in the respective studies. Four cohorts of colorectal tissue cases were utilized during this thesis as detailed below with the inclusion and exclusion criteria for cases present on the microarray. Sample location was broadly separated into left and right sided tissues based on embryological development. Right sided colonic tissues represent those obtained from the caecum, ascending colon and proximal two-thirds of the transverse colon. Left sided tissues represent those obtained from the distal third of the transverse colon, sigmoid and rectum.

2.1.3.1. FIXATION STUDY

CRC and normal colorectal tissues were isolated from the surgical specimens of 8 CRC patients to investigate the effect of the tissue fixation method on nucleic acid quality and tissue histomorphology. Under the supervision of a leading pathologist (Dr. KP West, LRI), normal and CRC tissues were obtained, ensuring that enough tissue remained for diagnostics.

2.1.3.2. INCLUSION AND EXCLUSION CRITERIA FOR THE SELECTION OF MICROARRAY SAMPLES

Patient inclusion criteria were strict and followed the guidelines outlined in Table 2.3.

Common guidelines applied to each category and were as follows:

- Patients must have undergone a total colonoscopy and not sigmoidoscopy.
- No additional disease or extra-colonic cancer could be present.
- Sufficient tissue was available for RNA extraction and clinical purposes.

The abbreviations used to describe the different types of colorectal tissue are described in Table 2.2.

Table 2.2: Explanation of sample group abbreviations

Abbreviation	Definition
NC	Normal Controls
HP	Hyperplastic Polyp
AN	Normal colonic mucosa from patients with Adenomatous Polyps (APs)
AP	Adenomatous Polyp
KN	Normal colonic mucosa from patients with colorectal polyps and CRC
KP	Colorectal Polyp from patients with CRC

This table explains the meaning of the abbreviations used throughout this thesis to describe the sample groups. This table should be used in conjunction with Table 2.3 to see the specific inclusion and exclusion criteria for sample collection.

Table 2.3: Inclusion and exclusion criteria for samples on the DASL microarray

Tissue Type	Inclusion Criteria	Exclusion Criteria
NC	No abnormal findings on colonoscopic investigation	Colonoscopy identified colorectal disease or presence of APs
	No abnormal findings on histological examination	Histological review revealed that the tissue was not normal
	No colorectal disease including; DD, IBD, CRC or PP	Presence of CRC, IBD, DD, PP or other colorectal disorder
HP	No abnormal findings on colonoscopic investigation	Colonoscopy identified colorectal disease such as CRC or IBD
	Diagnosis of HP on histological review	Histological review identified presence of inflammation, necrosis, or AP/CRC
	No additional colorectal disease including; DD, IBD or CRC (PP were allowed)	Patient had additional colorectal disease
AN	Presence of polyps on colonoscopic investigation but no other abnormalities identified	Colonoscopy revealed additional colorectal disease such as CRC or IBD
	Diagnosis of normal mucosa on histological review	Histological review identified non-normal tissue
	Normal tissue must be collected at the same time as an AP is removed and is not retrieved from previous or future normal blocks from the same patient	No normal tissue was removed at the same time as the AP
	Presence of AP identified histologically but separately to the associated normal tissue	No AP's identified on histological review
	No additional colorectal disease including; DD, IBD or CRC	Additional colorectal disease was present
AP	Presence of polyps on colonoscopic investigation but no other abnormalities identified	Colonoscopy revealed additional colorectal disease such as CRC or IBD
	Diagnosis of AP on histological review	On histological review APs showed signs of invasion or foci of adenocarcinoma. Presence of CRC or HP instead of AP
	No additional colorectal disease including; DD, IBD or CRC	Additional colorectal disease was present
KN	Presence of AP and CRC in resected tissue	AP in the presence of CRC was not identified
	Diagnosis of normal tissue on histological review	Histological review identified non-normal tissue
	Normal tissue must be collected at the same time as the AP and CRC and is not retrieved from previous or future normal blocks from the same patient	No normal tissue was retained at the time of AP and CRC removal
	No additional colorectal disease such as IBD or DD	Additional colorectal disease was observed
KP	Presence of AP and CRC in resected tissue	AP/CRC not identified
	Diagnosis of separate AP and CRC on histological review	Histological review did not identify AP or CRC
	No additional colorectal disease such as IBD or DD	Additional colorectal disease was identified

This table outlines the specific inclusion and exclusion criteria for each tissue type. These criteria were strictly adhered to. PP, Previous Adenomatous Polyps; DD, Diverticular Disease

2.1.3.3. TRAINING SET CASES

46 formalin fixed, paraffin embedded (FFPE) tissue samples referred to as the Training Set were identified from the LRI tissue bank (archived during 2008) for use in a microarray experiment. These samples are described in Table 2.4.

Table 2.4: Sample details of the Training Set

Sample ID	Gender	Age	Sample Group	Tissue Type	Location	Matched Sample (Y/N)
NC2	M	63	NC		n/k	N
NC3	M	57				N
NC5	M	41				N
NC37	F	50				N
NC49	M	41				N
NC54	F	61				N
NC68	F	59				N
NC292	F	48				N
HP3	F	61	HP		n/k	N
HP10	M	67				N
HP13	F	52				N
HP25	F	54				N
HP27	M	63				N
HP39	F	80				N
HP75	M	47				N
A161HP	M	76				N
A4N	F	61	AN		L	Y
A53N	M	64			L	Y
A161N	M	76			R	Y
A190N_1	M	73			R	Y
A190N_2	F	73			R	Y
A257N	M	61			R	Y
A329N	F	70			R	Y
A4P	F	61	AP	TA	L	Y
A10P	M	54		TV	L	N
A53P	M	64		TA	L	Y
A151P	M	72		TA	L	N
A161P	M	76		TA	R	Y
A190P_1	F	73		TV	R	Y
A190P_2	F	73		TV	R	Y
A257P	M	61		TV	R	Y
A329P	F	70		TA	R	Y
KT01N	F	67	KN		R	Y
KT05N	F	79			R	Y
KT26N	M	75			R	Y
KT27N	M	50			L	Y

KT28N	M	66			L	Y
KT50N	M	66			L	Y
KT65N	F	68			R	Y
KT71N	F	49			L	Y
KT01P	F	67	KP	TVA	R	Y
KT05P	F	79		TA	R	Y
KT26P	M	75		TV	R	Y
KT27P	M	50		TA	L	Y
KT28P	M	66		TA	L	Y
KT50P	M	66		TA	L	Y
KT65P	F	68		TA	R	Y
KT71P	F	49		TA	L	Y

This table presents the patient details for the samples in the Training Set. These samples were used for the microarray. The abbreviations are as follows: F, female; M, male; TA, Tubular Adenoma; TV, Tubulovillous Adenoma; VA, Villous Adenoma; R, Right; L, Left; Y, yes; N, No. Samples that are identified as being "matched" are paired with a normal or polyp tissue with the same prefix. For example, A4N is paired to A4P and KT01N is paired with KT01P.

2.1.3.4. VALIDATION SET

The results of the microarray experiment were validated using the Validation Set, which consists of 143 FFPE colorectal tissues. Some of the cases in the Validation Set were also present in the Training Set. The majority of samples were new and selected from the LRI tissue bank. Samples used were archived during 2008. A Summary of the Clinico-pathological features of the Validation Set is presented in Table 2.5.

Table 2.5: Summary of Validation Set cases

Sample Type	n	Gender	Age Range	Average Age	Location	Adenoma Type
NC	16	M=5, F=11	15-74	45.6	n/k	NA
HP	40	M=28, F=12	29-86	62.1	L=11, R=6, n/k=23	HP
AP	60	M=31, F=33	29-88	66.3	L=39, R=12, n/k=9	TA=36, TV=15, VA=9,
KN	12	M=9, F=3	49-85	69.9	L=5, R=7,	NA
KP	15	M=11, F=4	49-85	71.6	L=6, R=9,	TA=8, TV=3, VA=0, HP=3, flat=1

This table outlines the clinico-pathological features of the Validation Set cases.

2.1.3.5. MUTATION COHORT

APs and KPs from 61 patients were screened for *BRAF* (V600E) and *KRAS* (Codons 12 and 13) mutations (Table 2.6), which includes samples from the Training and Validation Set.

Table 2.6: Summary of clinico-pathological features of cases screened for mutations of *BRAF* and *KRAS*

Adenoma Type	n	Gender (M:F)	Age Range	Average Age	Location (L/R)
TA	37	F =18 M = 19	23-83	61.5	L=25, R=7, n/k=5
TVA	14	F = 8 M = 6	43-85	70.7	L=6, R=7, n/k=1
VA	8	F = 2 M = 6	69-88	76.8	L=4, R=1, n/k=3
Serrated	2	F=1. M=1	62-77	69.5	L=0, R=2

This table outlines the details of the Mutation Set. 61 polyps (APs and KPs) were investigated for their mutation status of *KRAS* and *BRAF*.

2.2. PROCESSING OF COLORECTAL TISSUES

2.2.1. FFPE TISSUES

FFPE samples were fixed in 10% formal saline for a minimum of 24 hours. FFPE tissues were processed by a senior laboratory technician in the department following standard fixation protocols on the Leica ASP3000 automated vacuum tissue processor. The Leica processing schedule is outlined in Table 2.7. Processing of FFPE tissues took place either overnight or over the weekend depending on when the tissues were collected and how long they were kept in 10% formalin saline.

Table 2.7: Processing schedule of FFPE tissues

Step	Solution/Temperature	Time (minutes)
1	70% IMS/10% Formalin	60
2	99% IMS	60
3	99% IMS	60
4	99% IMS	60
5	99% IMS	60
6	99% IMS	60
7	99% IMS	60
8	99% IMS	60
9	Xylene	90
10	Xylene	90
Wax bath 1	62 °C	60
Wax bath 2	62 °C	90
Wax bath 3	62 °C	90

This table represents the processing schedule for FFPE tissues on the Leica processor. This program can be used overnight or over the weekend. If the processing stage occurs over a weekend the first step is delayed until Sunday evening. The remaining steps continue as outlined above.

2.2.2. UMFIX FIXATION OF TISSUES

Tissues were fixed in UMFIX (90% methanol, 10% Polyethylene glycol 300 (Sigma-Aldrich, UK)) for 24 hours at room temperature. Following initial fixation, samples were processed using the Leica Processor using a 4 hour processing programme. The processing details are outlined in Table 2.8. This process was performed by a senior laboratory technician in the department.

Table 2.8: UMFIX four hour LEICA processor schedule

Step	Solution/Temperature	Time (minutes)
1	70% IMS	2
2	99% IMS	2
3	99% IMS	2
4	99% IMS	2
5	99% IMS	5
6	99% IMS	5
7	99% IMS	10
8	99% IMS	10
9	Xylene	30
10	Xylene	30
Wax Bath 1	62°C	30
Wax Bath 2	62°C	45
Wax Bath 3	62°C	60

This table shows the processing schedule for UMFIX tissues following their 24 hour fixation in UMFIX solution.

2.2.3. FROZEN TISSUES

Frozen tissues were fixed by immersion of colorectal tissues in isopentane, which was chilled using liquid nitrogen. Samples were then transferred to containers and stored at -20 °C in liquid nitrogen.

2.3. PREPARATION AND QUANTIFICATION OF NUCLEIC ACIDS

2.3.1. EXTRACTION OF TOTAL RNA FROM FFPE AND UMFIX TISSUES

Total RNA was extracted from UMFIX and FFPE tissues using a combination of tri-reagent (Sigma-Aldrich, UK) and the Qiagen RNeasy mini kit (Qiagen, UK) according to the manufacturer's instructions.

Tissues were dewaxed in xylene and subsequently rehydrated by immersion in increasingly dilute IMS down to 95%. The slides were then air dried. Tissues were removed from the slides using a 1 µl pipette tip using a Haematoxylin and Eosin (H and E) slide as a reference. All tissues, except those from the Fixation Set were microdissected to obtain foci of target tissue (normal, hyperplastic or adenomatous). Tissues were re-suspended in 500 µl TRIS pH8/0.1% SDS with 5 µl of Proteinase K (PK) (10mg/ml) (Roche Diagnostics Ltd, UK) and incubated at 56˚C overnight. Samples were chilled on ice and 500 µl of Tri-reagent was added, vortexed, left at room temperature and subsequently centrifuged at 13000rpm for 15 minutes at 4˚C. The aqueous phase of each sample was removed and transferred to a clean eppendorf. An appropriate volume of absolute ethanol was added to each sample (x1.2 volume of sample). The sample was vortexed before completing the extraction using the Qiagen RNeasy mini kit according to the manufacturer's instructions. Samples were re-suspended in 30 µl of RNase free H_2O.

2.3.2. EXTRACTION OF TOTAL RNA FROM FF TISSUES

Frozen tissue samples were cut using a cryostat and stored in an eppendorf with 1ml of tri-reagent. Samples were thawed on ice for five minutes followed by the addition of 200 µl of chloroform, then vortexed and left at room temperature for three minutes. Samples were centrifuged for 15 minutes at 13000rpm at 4˚C. The aqueous layer was transferred to a clean eppendorf and the process was then repeated using half volumes of both tri-reagent and chloroform. Following the addition of second chloroform and subsequent spin, the aqueous layer was transferred to a clean eppendorf with 500 µl of isopropanol. The sample was then vortexed and incubated at room temperature for 10 minutes. Samples were centrifuged for 15 minutes at 13000rpm at 4˚C. The supernatant was carefully removed and 500 µl 70% ethanol was added and centrifugation was repeated. The supernatant was carefully removed and the pellet was allowed to air-dry. The samples were re-suspended in 25 µl of sterile ultrapure (UP) H_2O. Samples were stored at -20˚C.

2.3.3. EXTRACTION OF DNA FROM UMFIX AND FFPE TISSUES

UMFIX and FFPE tissue sections were dewaxed and rehydrated using the same method as described in 2.3. Tissues were scraped off the slide and re-suspended in 300 µl TRIS pH8/0.1% SDS then 30 µl of PK (10mg/ml) was added and digested at 56˚C for three days with an additional 30 µl of PK added each day.

Following digestion with PK, 300 µl of phenol-chloroform isoamyl alcohol (IAA) (Sigma-Aldrich, UK) was added and samples were vortexed and centrifuged at 13000rpm for three minutes at room temperature. The aqueous phase was transferred to a clean eppendorf with an additional 300 µl of phenol-chloroform IAA and extraction and

centrifugation was repeated. The aqueous layer was again transferred to a clean eppendorf and 1/10th volume of 1M NaCl and 1.25x volume of cold absolute ethanol was added. Following overnight incubation the samples were centrifuged at 13000rpm for 15 minutes at 4˚C and pellets were rinsed with 500 µl 70% ethanol and re-centrifuged. Pellets were then air-dried, re-suspended in 30 µl of UP H_2O and stored at 4˚C.

2.3.4. EXTRACTION OF DNA FROM FF TISSUES

Frozen tissues (10 sections) were cut using a cryostat and digested with 1 ml TRIS pH8/0.1% SDS and 50 µl PK (10mg/ml) overnight at 56˚C. Following overnight incubation DNA was isolated following standard phenol/chloroform extraction and ethanol precipitation as described previously.

2.3.5. NUCLEIC ACID QUANTIFICATION

The concentration of RNA and DNA was determined using the NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, USA). 1 µl of sample was loaded onto the NanoDrop pedestal and the absorbance measured at 230, 260 and 280 nm. The purity of DNA and RNA was established by comparing the ratios with pure DNA having a desired $A_{260/280}$ ratio of 1.8, and RNA having a ratio of 2.1. Additionally, for RNA, the $A_{260/230}$ was also compared with an ideal value being greater than 1.8 to help ensure that samples were not contaminated with protein or reagents from the RNA extraction process. The ratios were calculated using the NanoDrop interface.

2.4. ENZYMATIC MANIPULATION OF NUCLEIC ACIDS

2.4.1. PRIMER DESIGN

Oligonucleotide primers were designed in house and ordered from Sigma-Genosys, UK. All forward and reverse primer sequences were generated using a combination of the Primer_3 software program (Rozen and Skaletsky, 2000) followed by sequence confirmation using NCBIs Nucleotide Basic Local Alignment Search Tool (BLAST) (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

The details of FAM-MGB probe design will be discussed in 2.4.3. All primer sequences can be found in Table 2.9.

For each primer, lyophilized pellets were re-suspended in sterile Ultra Pure (UP) H_2O to a concentration of 200 pm/µl (200 µM) and stored at -20 °C. Working aliquots (10pm/ µl (10 µM) were taken from the stock primers as needed.

The Fixation Set required the adaptation of RPL13a primer sequences provided by Illumina to allow the design of a reverse transcription primer (RTP) and FAM-MGB probe. Using the Primer Express Software Program (Applied Biosystems, UK), the sequences were modified to allow the inclusion of the additional two sequences. The sequence was then checked, using BLAST, prior to synthesis. GAPDH primers were also utilized by the Fixation Set. These primers were designed and validated by a Post-Doctoral Fellow prior to the commencement of this project.

2.4.2. GENERATION OF MULTIPLEX REVERSE TRANSCRIPTION PRIMERS

The Validation set required the synthesis of eight primers that could be multiplexed together in the RT reaction due to a limited sample quantity. These primers were used to generate multiplexed cDNA for use in qRT-PCR reactions.

Forward and reverse primers were first synthesized using the methods outlined above. Care was taken to ensure that the sequences selected were gene specific and did not contain single nucleotide polymorphisms (SNPs) using SNP: GeneView (NCBI). Following this the sequences were input into MuPlex (Rachlin et al., 2005), a software used to generate multiplexed PCR assays, and the relevant parameters selected (e.g. amplicon size, melting temperature (Tm), etc). The oligonucleotide sequences identified were further checked for specificity to the target gene using BLAST prior to ordering, and some were adjusted using Primer Express where necessary.

2.4.3. DESIGN AND SYNTHESIS OF TAQMAN LABELLED PROBES

All FAM™-labeled MGB probes were designed using the Primer Express Software. Design guidelines were adhered to as set out by the Primer Express Software as follows:

- Long sequences of identical nucleotides were avoided
- The 5' end of the probes did not contain a guanosine residue to prevent quenching of the reporter fluorescence
- The Tm of the probes was designed to be ~10 °C higher than that of the primers
- When performing allelic discrimination (e.g. *KRAS* probes) VIC™ and FAM™ reporter dyes were used to label the probes

KRAS probes were designed for codons 12 (bases 1 and 2) and 13 (base 2). These probes are referred to as *KRAS* 121, 122 and 132 where the first two digits refer to the codon, and the last digit refers to the base. In addition, a wild-type probe (*KRAS* WT) was also synthesized. *KRAS* WT probe was used as a template for the generation of both mutant probes in codon 12 by inputting a degenerate base at the appropriate position in the WT probe. The use of degenerate bases allowed probes to be made that contain all of the alternative bases that are seen in the mutant forms. Codon 13 is associated only with one major base change (G>A) and therefore degenerate bases were not used. The Catalogue of Somatic Mutations in Cancer (COSMIC) database available at the Sanger Institute was used to identify the mutations in *KRAS* and the degenerate bases required for codon 12. The WT probe was labeled with a VIC™ reporter dye, whereas mutant KRAS probes were labeled with FAM™ reporter dyes.

The mutation dataset also investigates *BRAF* mutations. *BRAF* primer and probe sequences were designed and validated by previous members of the department based on published methods (Benlloch et al., 2006). All primer and probe sequences can be found in Table 2.9.

Table 2.9: Primer and probe sequences used during this project

GENE NAME	Primer/ Probe	Sequence	Primer/ Probe Tm	Primer/ Probe GC content	Primer/ Probe length	Amplicon length
ASCL2	FP	CGCAACCGCGTGAAGC	59.6	69	16	78-86
	RP	CTTCTTGCTGGCGCCG	58.7	69	16	
	RTP	TTGCTCAGCTTCTTGCTG	59.4	50.0	18	
	Probe	CAGGCGCTGCGGCA	71.0	79	14	
ANXA2	FP	CTCAGCTTGGAGGGTGATCACT	59.3	55	22	94-109
	RP	TGTTCAAAGCATCCCGCTC	58.4	53	19	
	RTP	TGATGGCTGTTTCAATGTT	58.4	36.8	19	
	Probe	AAGTGCATATGGGTCTGTC	70.0	47	19	
AXIN2	FP	GAAGGAGACAGGTCGCAGGAT	59.3	57	21	79-89
	RP	GGGCACTATGGGGCTTGG	59.9	67	18	
	RTP	TTTGTGCTTTGGGCACTA	59.8	44.4	18	
	Probe	CTGGCAGTGGATGCT	69.0	60	15	
ETS2	FP	CAGTTTCTCCTGGAGCTGCTATC	58.3	52	23	79-87
	RP	GCTTAAACTCCCATCCGTCTCC	59.9	55	22	
	RTP	GTCGGCGAGCTTAAACTC	55.6	60.2	18	
	Probe	ACAAATCCTGCCAGTCAT	70.0	44	18	
G3BP1	FP	CTGAAGAAGAAGTAGAGGAACCTGAAG	58.5	44	27	73-100
	RP	AAAGTTCCAGAATCATCAGGTACCA	59.0	40	25	
	RTP	TCATTACTGACAACTGCCTG	58.4	45.0	20	
	Probe	AAGACAGCAAACACC	68.0	47	15	
TFF2	FP	GAATCACCAGTGACCAGTGTTTTG	59.4	46	24	87-110
	RP	TTGGGAGGGGGTGGAAAC	59.1	61	18	
	RTP	GACGCACTGATCCGACTC	61.1	60.2	18	
	Probe	AATGGATGCTGTTTCGACTC	70.0	45	20	
UBC	FP	AGGTGGGATGCAGATCTTCGT	59.5	52	21	71-90
	RP	TGTCACTGGGCTCCACCTC	58.2	63	19	
	RTP	GCCTTGACATTCTCGATG	57.9	50.0	18	
	Probe	ACCCTGACTGGTAAGAC	69.0	53	17	
TPT1	FP	CCATCACCTGCAGGAAACAAGT	60.0	50	22	84-99
	RP	AGTTTCCCTTTGATTGATTTCATGTA	58.1	31	26	
	RTP	GGTCTCTGTTCTTCAAGTTTC	57.7	42.9	21	
	Probe	ACAAAAGAAGCCTACAAGAA	70.0	35	20	
RPL13a	FP	GAAGGCATCAACATTTCTGGC	60.0	48	21	84-137
	RP	AGGGTTGGTGTTCATCCGC	59.5	58	19	
	RTP	ACGGTCCGCCAGAAGATG	59.7	61	18	
	Probe	TACCTGGCTTTCCTCC	69.2	56	16	
KRAS	FP	AGGCCTGCTGAAAATGACTGA	59.4	48	21	78
	RP	TGTATCGTCAAGGCACTCTTGC	58.9	50	22	
	WT Probe	CTACGCCACCAGCTC	70	67	15	
	121 Probe	TACGCCACDAGCTC	66	57-64	14	
	122 Probe	TACGCCADCAGCTC	66	57-64	14	
	132 Probe	CTACGTCACCAGCTC	66	60	15	

The amplicon range given represents the range from the FP to RP, followed by the FP to RTP amplicon length. The range of Tm's given for KRAS 121 and 122 probes represent the difference in Tm according to the base inserted at the degenerate base. The degenerate base used in these probes is D, allowing A, T and G bases to be included in the sequence.

2.4.4. GENERATION OF COMPLEMENTARY DNA USING REVERSE TRANSCRIPTION

Synthesis of complementary DNA (cDNA) from total RNA was achieved with the use of AMV-Reverse Transcription (AMV-RT) reagents. Unless otherwise stated, all reagents described in this section were supplied from Promega, USA.

Due to limited RNA quantities, 200ng of total RNA was used in each AMV-RT reaction. The required volume of RNA (containing 200ng) was added to 1.5µl of 10 µM RTP. The volume was made up to 15µl with the appropriate volume of UP H_2O or RNase free H_2O. Samples were vortexed and then incubated at 70°C for 5 minutes and subsequently allowed to return to room temperature. 5 µl of 5x AMV-RT buffer (250 mM Tris-HCl, pH 8.3; 250 mM KCl; 50 mM MgCl2; 2.5 mM spermidine & 50 mM DTT); 2.5 µl 10 mM dNTPs; 0.62 µl RNasin; and 0.5 µl AMV-RT were added to the samples. The samples were vortexed and subsequently incubated at 42 °C for 60 minutes in a GeneAmp 9700 96-well thermal cycler. Samples were stored at 4 °C until required for qRT-PCR.

Cell line cDNA was generated using the same method as outlined above with the exception that the amount of input RNA was 1 µg. The validation set was also constructed using the same method as outlined above with the exception that a master mix of the eight RTPs was created. The master mix consisted of a 1:8 dilution of each working aliquot (1.25 pmol/RTP). 4 µl of this 8 gene master mix was added to the RT reaction instead of a single primer.

A control for each sample was also included where AMV-RT was omitted from the reaction and replaced with UP H_2O. This step was included to test for genomic DNA contamination during the extraction of RNA.

2.4.5. QUANTITATIVE PCR FOR ALLELIC DISCRIMINATION

BRAF and *KRAS* primers and probes were used to test for mutations in these genes using DNA templates from cell lines and colorectal adenomas. Each reaction was set-up on ice and consists of: 3.6 µl DNA (10ng/reaction), 5 µl TaqMan Genotyping Master Mix (Applied Biosystems, UK), 0.6 µl of forward and reverse primers; 0.2 µl WT (VIC) and 0.2 µl (FAM) probes. Each sample was performed in duplicate. In addition to the samples a positive cell line control (LOVO, HCT116, GP2d, SW626 and SW837) was included for each probe set on every plate, along with a no template control (NTC) to test for genomic DNA contamination during the PCR set-up. Prior to the commencement of the PCR, plates were spun at 3000rpm for 30 seconds. All reactions were performed on the Step-One thermal cycler (Table 2.10).

Table 2.10: qRT-PCR cycling conditions for allelic discrimination

Stage	Temperature (°C)	Time	Number or Cycles
Initial denaturation	95	10 minutes	1
Denaturation	95	15 seconds	40
Annealing and Extension	63/64	1 minute	

KRAS 121 and 122 were annealed and extended at 63°C whereas KRAS 132 worked better at 64 °C

2.4.6. QUANTITATIVE PCR FOR GENE EXPRESSION

Each gene expression reaction (with the exception of *GAPDH*) was set-up on ice and consists of: 3.6 µl cDNA (diluted 1:10); 5 µl 2x TaqMan Fast Universal Master Mix (Applied Biosystems, UK); 0.6 µl forward and reverse primers; and 0.2 µl FAM-MGB probe. The cycling conditions of the Step-One thermal cycler can be seen in Table 2.11. These conditions apply for both the Fixation, Training and Validation sets. Where DNA

was used as the input template, the same method was used as described above with the substitution of cDNA for 10ng DNA.

The *GAPDH* primers were the exception to the above method as they require the use of Fast SYBR Green (Applied Biosystems, UK) instead of 2x TaqMan Fast Universal Master Mix. This was due to the absence of a designed probe for this primer set. The reaction was set up in the same way as described above using Fast SYBR Green instead of 2x TaqMan Fast Universal Master Mix and substituting the probe volume for an equivalent volume of UP H_2O.

Table 2.11: qPCR cycling conditions for gene expression

Stage	Temperature (°C)	Time (seconds)	Number of cycles
Initial denaturation	95	20	1
Denaturation	95	3	40/50
Annealing and Extension	60	20	

This table shows the cycling conditions on the Step-One thermal cycler. The number of cycles varies according to the gene investigated. When investigating *GAPDH* expression, 40 cycles were used, however when looking for gene expression of candidate genes in the Validation Set 50 cycles were used. More cycles were used due to the lower expression levels of the genes in some tissues.

2.4.7. STANDARD CURVES

To test the efficiencies of the individual primer/probe sets, standard curves were constructed for each gene using cell line cDNA templates. Template cDNA was serially diluted for each probe set and the efficiency of the probe set was calculated from the gradient of the slope.

The hallmarks of an optimized qRT-PCR reaction rely on three things; linear standard curve (R^2 >0.98); high amplification efficiency (90-110%) and; consistency across the replicate reactions.

2.4.8. RELATIVE EXPRESSION OF TARGET GENES

The relative expression of target genes was determined by comparing the average Cts of target genes with those of endogenous controls to generate a ΔCt using Equation 1.

Equation 1: Calculation of ΔCt

ΔCt = Average Ct of sample – Average Ct of endogenous control

2.4.9. STATISTICAL ANALYSIS OF QRT-PCR DATA

The results of the qRT-PCR experiments were analysed using GraphPad Prism (GraphPad Software, inc. USA). The data sets were assessed for normality using the KS, D'Agostino and Pearson, and Shapiro-Wilk Normality tests. If sample groups showed a normal distribution, parametric tests were performed. Parametric tests were performed as long as one of the sample groups exhibited a normal distribution. If a test was performed whereby all sample groups did not exhibit a normal distribution, non-parametric tests were performed.

When parametric tests cannot be used, a non-parametric alternative was utilized. Table 2.12 outlines the different statistical tests used and the post-hoc tests associated with them.

Table 2.12: Selection of statistical test according to normality status of the sample groups

Parametric Test	Non-Parametric Alternative	Number of groups
One way Analysis of Variance (one way ANOVA) with Bonferroni Multiple Comparison Test	Kruskal-Wallis test with Dunns Post-hoc test	≥3 groups
Unpaired T-test	Mann-Whitney Test	2 groups
Paired T-test	Wilcoxon matched-pairs signed rank test	2 groups

The tests described in this table were utilised for all analysis involving qRT-PCR data. The chosen test was determined by the normality status of the groups investigated and the number of sample groups investigated. If one or more of the groups exhibited a normal distribution, parametric tests were utilised. It was determined that since some of the sample groups are not "normal" tissues, there would be some alteration in the normal distribution as a result of the disease process. If these groups were being compared to a normally distributed group, parametric tests were used.

2.5. MICROARRAY

2.5.1. QUALITY CONTROL OF MICROARRAY SAMPLES

The samples belonging to the Training Set were used to perform a Whole-Genome DASL Gene Expression Assay. The DASL (cDNA-mediated Annealing, Selection, Extension and Ligation) microarray was performed by Cambridge Genomics Services (CGS). The initial quality control of the samples was performed prior to shipment of the samples to CGS. Each sample was screened using RPL13a following the RNA extraction, reverse transcription and qRT-PCR methods described in sections 2.3, 2.4.4 and 2.4.6. Samples were deemed of sufficient quality for microarray analysis if the cycle threshold (Ct) was less than 29. In addition, CGS required samples to meet the following criteria: concentration > 40ng/µl; $A_{260/230} > 1.8$; and $A_{260/280} > 1.8$. The criteria provided by CGS were prioritised if the Ct was slightly above 29.

2.5.2. ARRAY HYBRIDISATION

Briefly, RNA was converted to cDNA in a reverse transcription reaction using biotinylated primers. The cDNA was annealed to assay oligonucleotides and bound to streptadivin conjugated paramagnetic particles (SA-PMPs). Following oligo hybridisation, any mis-hybridised and non-hybridised oligonucleotides were washed away. The hybridised oligonucleotides were extended and ligated forming the synthetic template for the PCR reaction. The template was transferred to a PCR reaction containing a fluorescently labelled primer, which labelled the PCR template. The labelled PCR product was then isolated and hybridised to a whole-genome expression BeadChip: Human Ref-8 WG-DASL. The BeadChip was then washed and imaged on the BeadArray Reader.

2.5.3. ANALYSIS OF ARRAY DATA

2.5.3.1. NORMALISATION AND TOP-TABLE GENERATION

Analysis of the DASL microarray data was primarily done using R bioconductor (www.bioconductor.org/) using lumi and limma packages. Quantile normalisation, transformation and production of top tables of gene expression were produced using R bioconductor using scripts written by the author.

The top-tables contain a log-fold change, t-statistic, raw p-value, adjusted p-value and B statistic for each gene. Significant genes were identified as having a p-value of 0.05 or less. The adjusted p-value takes into account the large number of tests performed on the data and reduces the false discovery rate (FDR). The Benjamini and Hochberg's (BH) method was employed to adjust the data and control the FDR, producing an adjusted p-value and limiting the number of significant genes identified. The adjusted

p-value was given priority in the interpretation of the results. However in cases where the adjusted p-values identified very few/no differentially expressed genes (DEGs), the raw p-values were also considered. Raw p-values do not take into account multiple testing and identify more DEGs but with a higher number of false positive results.

The B statistic reflects the likelihood of a gene being differentially expressed. Equation 2 and Equation 3 show the formulas used to interpret the B-statistic and arrive at a probability of a gene being differentially expressed.

Equation 2: Calculating the odds of differential expression using the B statistic

$$\text{Exp}(B) = n$$

B, B-statistic provided in the top tables; n, the odds of differential expression relative to 1

Equation 3: Calculating the probability of a gene being differentially expressed

$$= (n/1+n)*100$$

2.5.3.2. GENE EXPRESSION ANALYSIS

The top tables were used to identify DEGs. In addition, the normalised data set was also investigated using MultiExperiment Viewer (MEV) (www.tm4.org/mev/) (Saeed et al., 2006, Saeed et al., 2003). Hierarchical cluster (HCL), Significance Analysis of Microarrays (SAM) (Tusher et al., 2001) and Principle Component Analysis (PCA) tests were performed in MEV to identify significant genes and sample clustering. The default statistical settings of the tests were utilized as they were most appropriate for these data.

2.5.3.3. PATHWAY ANALYSIS

Pathway analysis was performed using the Database for Annotation Visualization and Integrated Discovery (DAVID) (Huang da et al., 2009b, Huang da et al., 2009a, Huang da et al., 2007) and COXPRESdb (Obayashi et al., 2008, Obayashi and Kinoshita, 2011).

These two programmes when used in conjunction enable relationships to be identified between the DEGs.

2.5.3.4. ADDITIONAL RESOURCES

In addition to the DASL microarray data, the GSE4183 (Galamb et al., 2008a) data set obtained from NCBI's GEO datasets (http://www.ncbi.nlm.nih.gov/gds) was also investigated. This array was analysed in a similar way to the DASL microarray, with the exception that the affy package was utilized in R Bioconductor instead of lumi.

The use of PubMed and NCBI's Gene was also used to identify the function and significance of genes identified in the analysis. The use of GSE4183 and NCBI provided additional ways to validate the results of the DASL microarray.

CHAPTER 3: COMPARISON OF EFFECTS OF FIXATION METHODS ON RECOVERY OF NUCLEIC ACIDS FROM COLORECTAL TISSUES

3.1. INTRODUCTION

3.1.1. PRESERVATION AND ASSESSMENT OF HUMAN TISSUE SAMPLES

The preservation of tissue architecture is essential for accurate histopathological diagnosis. Recent technologies such as expression arrays and more routine techniques such as PCR require extraction of high quality nucleic acids to obtain reliable results.

The preservation of tissue requires appropriate tissue fixation. The first and most important variable is the time between tissue excision and exposure to fixative, as delays in this step can cause a substantial reduction in the quality of the macromolecules obtained (Dash et al., 2002). Secondly, tissue should be immersed in an adequate amount of fixative to ensure complete fixation. Thirdly, the tissue should not be extensively exposed to the fixative prior to processing (Masuda et al., 1999) and care should be taken to prevent autolysis and to inhibit microorganism growth.

To date, the preservation of tissue architecture has been achieved predominantly using formalin fixation. Formalin Fixed Paraffin Embedded (FFPE) tissues represent the largest available source of human tissues for retrospective analysis, but DNA and RNA extracted from FFPE tissues is often degraded and of poor quality. Fresh Frozen tissue provides better quality DNA and RNA; however the histomorphology is inadequate for clinical diagnosis. Therefore FF tissues are less abundant, and are an unrealistic alternative to FFPE tissues for routine analysis due to the extensive storage requirements.

The introduction of alternative methods of fixation, such as Universal Molecular Fixative (UMFIX), into the clinical and research setting has been suggested to maintain good tissue histomorphology while preserving DNA, RNA and protein (Vincek et al.,

2003, Cox et al., 2006). In contrast to FFPE tissues the length of exposure of tissues to UMFIX does not cause a reduction in nucleic acid quality. UMFIX also has the advantage of being less toxic than formalin, which is a known carcinogen, irritant and poison (Cleary et al., 2005). If the claims about UMFIX are validated, this fixative could replace formalin as the fixative of choice for both clinicians and researchers alike.

3.1.2. ENDOGENOUS CONTROLS FOR THE INTERPRETATION OF QRT-PCR DATA

Quantitative RT-PCR (qRT-PCR) generates cycle threshold (Ct) values that can be used for the accurate quantification of mRNA expression. The Ct is the cycle number at which signal is detected above the background fluorescence. Endogenous control (EC) genes are used to normalise gene expression between samples by relative quantification. The ideal EC gene should show stable expression in the tissue independent of disease or treatment. However, in practice many commonly used EC genes show unequal gene expression across different tissues (Warrington et al., 2000); therefore a panel of EC genes is sometimes used for normalisation.

GAPDH and *UBC* show stable expression over a number of tissues (Andersen et al., 2004) and are therefore widely used as EC genes. Illumina (the platform used in this thesis) recommend the use of *RPL13a* for prediction of RNA quality (Reinholz et al., 2010) with an inverse relationship seen between the Ct value for *RPL13a* RNA quality (Waddell et al., 2010). Other variables that can alter the qRT-PCR output include amount of starting material, enzymatic efficiency and cellular differences between the tissues of interest. Normalisation using validated EC genes helps to minimise the

effects of these variables on the qRT-PCR results and enables a more accurate prediction of nucleic acid quality.

3.2. AIMS AND OBJECTIVES

The aim of this chapter was to establish the best method of tissue fixation to preserve the quality and integrity of nucleic acids for subsequent expression microarray studies, while maintaining tissue architecture. Three methods of tissue fixation were compared; UMFIX, FFPE and FF.

The specific objectives were as follows:

- To establish the effect of different methods of tissue fixation on DNA (*GAPDH*) and RNA (RPL13a) quality using samples from the Fixation Set.
- To determine the effect of different methods of tissue fixation on tissue histomorphology.

3.3. RESULTS

Eight patients were recruited to obtain samples for the Fixative Study. Normal and tumour tissue was collected from each patient as described in 2.1.3. A sample of each tissue was fixed using the three fixation methods: FFPE, UMFIX and FF. The preservation of histomorphology was assessed by a consultant histopathologist. *RPL13a* and *GAPDH* were used to assess RNA and DNA quality respectively using qRT-PCR.

3.3.1. VALIDATION OF RPL13A AND GAPDH PRIMER SETS

RPL13a primer sequences were adapted from sequences provided by Illumina to allow the design of an RTP and TaqMan probe. Standard curves were constructed using cDNA derived from the HCT116 cell line.

A 1:5 serial dilution was carried out on HCT116 cDNA (Figure 3.1) to generate a standard curve. The efficiency seen in Figure 3.1 is 93.52%, which is within the 90-110% range desired for maximal accuracy and efficiency. The R^2 value for the *RPL13a* standard curve was 0.9934.

GAPDH primers (for DNA expression) were available in house and had been designed and validated previously by a Post-doctoral researcher in the group. No additional validation step was performed on this primer set.

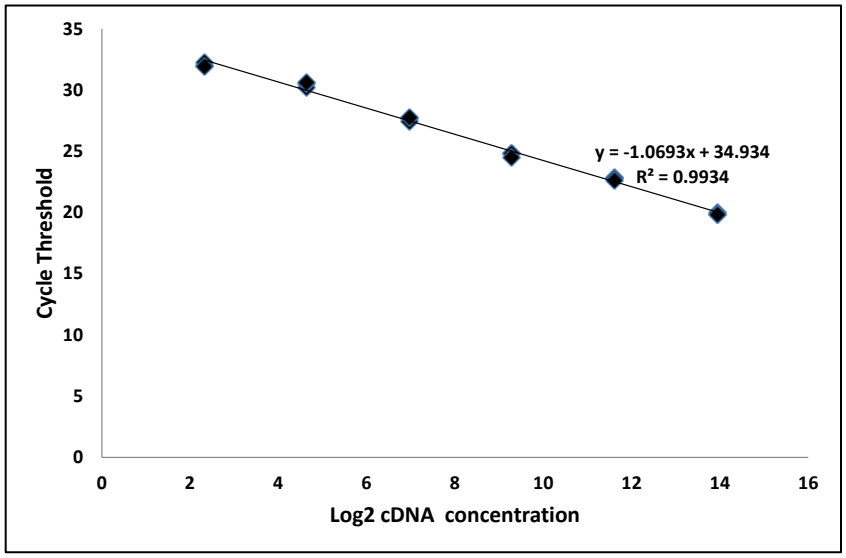

Figure 3.1: Standard curve for RPL13a
This standard curve shows that the efficiency of the *RPL13a* primers lies within the optimal range for efficiency. In addition, the three replicate points overlap, showing minimal variation.

3.3.2. THE EFFECT OF TISSUE FIXATION ON TISSUE HISTOLOGY

Tissue histomorphology was assessed using H and E stained slides of UMFIX, FFPE and FF for each normal and cancer tissue. No significant difference was seen in tissue architecture or morphology when comparing UMFIX and FFPE fixation. Frozen tissues (FF) were deemed less desirable due to a poorer haematoxylin stain and the presence of artefacts created during the fixation process, supporting previous studies (Vincek et

al., 2003). Figure 3.2 shows examples of H and E sections from the same tissue for each fixation method.

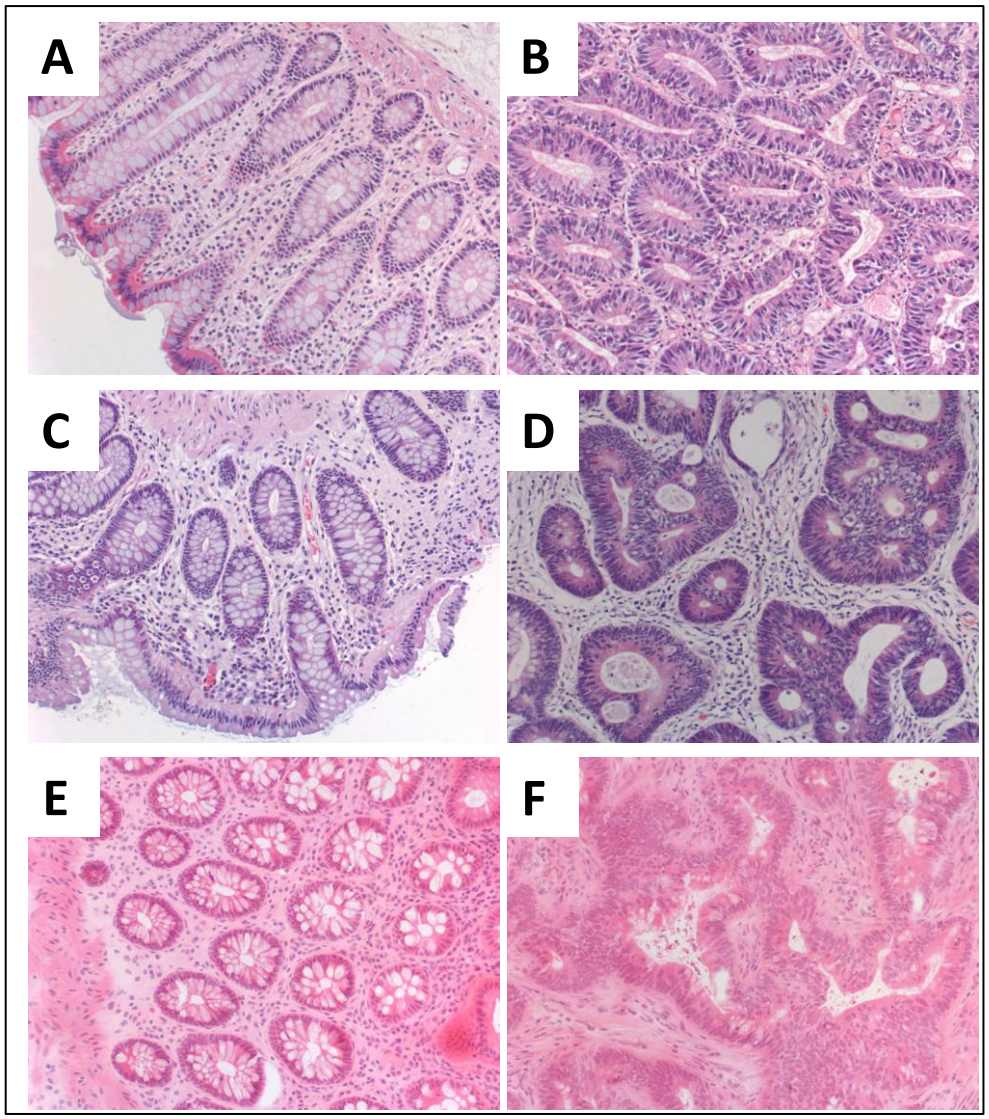

Figure 3.2: Histology of UMFIX, FFPE and FF tissues from CRC patients
All photographs are taken at x20 magnification and interpreted by a Consultant Histopathologist. Matched normal and tumour tissues were fixed using UMFIX, FFPE and FF fixation methods. Normal tissue sections; fixed with UMFIX (A), FFPE (C) and FF (E). Tumour tissue sections; fixed with UMFIX (B), FFPE (D) and FF (F).

3.3.3. QUANTITATIVE PCR ANALYSIS OF DNA FROM MATCHED UMFIX, FFPE AND FF TREATED COLORECTAL TISSUES

GAPDH was used to investigate the quantity and quality of DNA preservation using FFPE, UMFIX and FF fixative methods using a fixed amount of starting DNA (10ng). Figure 3.3 shows average Ct values for each fixation method. There was a clear difference between UMFIX and FFPE tissues, with FFPE showing higher average Cts, suggesting poorer quality DNA.

UMFIX tissues showed a greater homogeneity and a lower average Ct than FFPE or FF tissues. The difference between UMFIX and FF tissues was minimal; suggesting that the quality of DNA obtained from tissues fixed using these methods was comparable.

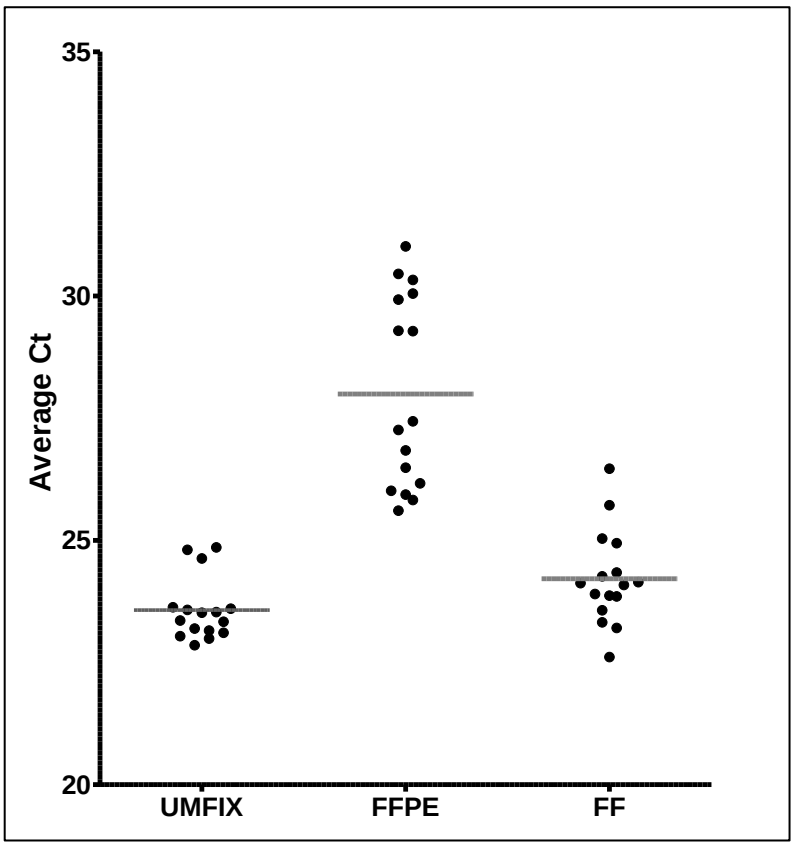

Figure 3.3: Scatter plots of GAPDH average Cts for all normal and tumour samples
A scatter plot showing the Average Ct values and mean of *GAPDH* in DNA from UMFIX, FFPE and FF treated samples.

The average Ct values (Table 3.1) obtained from qPCR analysis were compared using non-parametric tests as the dataset was not normally distributed (Table 3.2).

Table 3.1: Average Ct value of GAPDH in UMFIX, FFPE and FF treated tissues

SAMPLE	UMFIX	FFPE	FF
Case 1N	23.6	26.4	22.6
Case 1T	23.4	25.8	23.9
Case 2N	23.6	27.4	24.1
Case 2T	23.5	26.0	23.2
Case 3N	23.2	29.9	23.9
Case 3T	23.0	29.3	25.7
Case 4N	24.9	31.0	25.0
Case 4T	24.8	30.5	24.1
Case 5N	23.1	26.2	24.9
Case 5T	22.9	26.8	24.3
Case 6N	23.3	30.3	23.3
Case 6T	23.2	29.3	23.9
Case 7N	23.6	30.1	24.1
Case 7T	23.0	26.0	26.4
Case 8N	24.6	25.6	23.6
Case 8T	23.5	27.2	24.3

This table presents the average Ct values of *GAPDH* in normal and tumour tissue treated with UMFIX, FFPE and FF. Average Cts are given according to one decimal place.

Table 3.2: Comparison of DNA quantity by Ct isolated from UMFIX, FFPE and FF treated tissues

Statistical Test	Post-test	Sample Comparisons	p-value	Significance
Kruskal-Wallis	Dunn's Multiple Comparison Test	UMFIX-FFPE-FF	P<0.0001	Yes
		UMFIX-FFPE	***	Yes
		FFPE-FF	***	Yes
		UMFIX-FF	NS	No

Kruskal-Wallis test compared the average Cts of *GAPDH* in UMFIX, FFPE and FF treated tissues. A Dunn's Multiple comparison test was performed to compare the individual groups. ***, p<0.0001.

Table 3.2 shows that there was a statistically significant differences in mean Ct values across the three methods of tissue fixation (p<0.0001, Kruskal-Wallis). In the individual comparisons the biggest differences were between UMFIX and FFPE (***, p<0.0001, Dunn' Multiple Comparison Test) and FFPE and FF (***, p<0.0001, Dunn's Multiple Comparison Test), whereas no significant difference was seen between UMFIX and FF tissues.

Results were also compared for normal and tumour tissues. Scatter plots are presented in Figure 3.4. These charts show similar results to those found in Figure 3.3. UMFIX and FF tissues appear comparable, with less variation and spread of the data. FFPE tissues show a much greater variation in average Ct in both normal and tumour tissues.

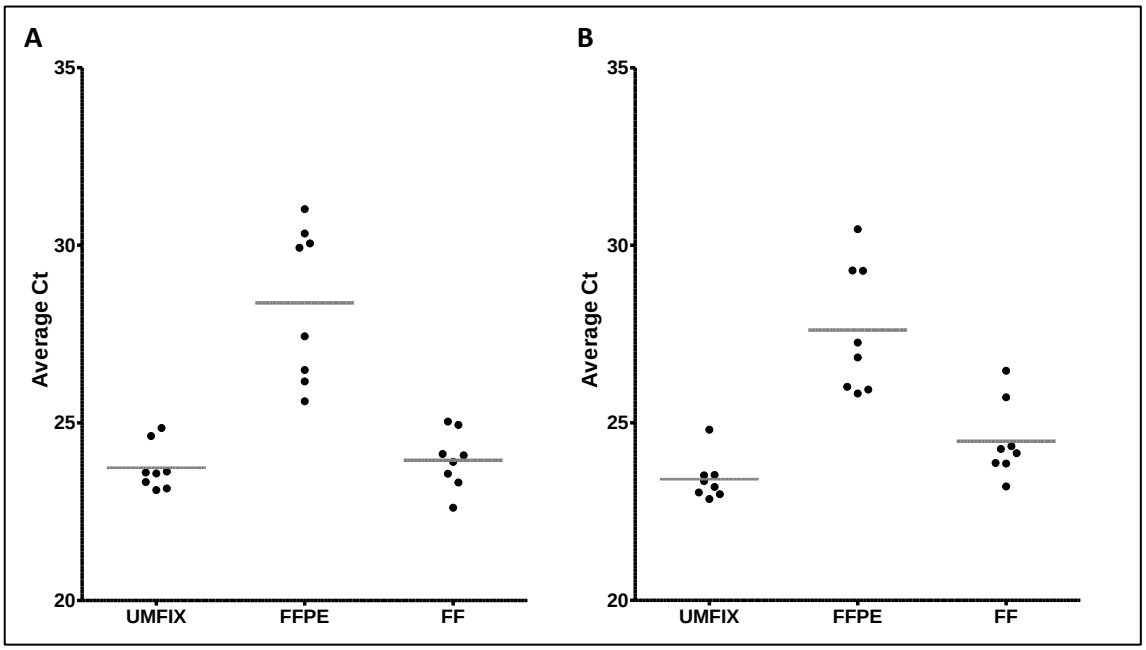

Figure 3.4: Scatter plots of GAPDH average Cts for normal and tumour samples
A) Scatter plot with mean, showing average Ct values of *GAPDH* from Normal colonic tissue treated with UMFIX, FFPE and FF protocols. B) As for (A) but showing average Ct values of Tumour tissues.

The results confirm the differences in DNA quality identified between UMFIX and FFPE tissues (Table 3.3). Normal and tumour tissues did not show a significant difference when comparing UMFIX and FF tissues. These findings suggest that these two methods of fixation are comparable.

In contrast, the difference between UMFIX and FFPE tissues was highlighted in both normal and tumour tissues by highly significant p-values.

Table 3.3: Comparing GAPDH average Ct s from normal and tumour DNA samples treated with UMFIX, FFPE and FF

Sample Type	Statistical Test	Post-test	Comparison	p-value	Significance
Normal Colonic Mucosa	Kruskal-Wallis	Dunn's	UMFIX-FFPE-FF	p=0.0004	Yes
			UMFIX-FFPE	***	Yes
			FFPE-FF	**	Yes
			UMFIX-FF	p>0.05	No
Colorectal tumour tissue	Kruskal-Wallis	Dunn's	UMFIX-FFPE-FF	p=0.0002	Yes
			UMFIX-FFPE	***	Yes
			FFPE-FF	p>0.05	No
			UMFIX-FF	p>0.05	No

Normal mucosa and colorectal tumour tissues were investigated separately using non-parametric Kruskal-Wallis test with Dunn's Multiple Comparison post-hoc test. **, p<0.01; ***p<0.0001.

A 2 way ANOVA of the grouped data set was performed where normal and tumour samples from the same cases were paired, and then compared with the three methods of fixation. The result of this test ($p<0.0001$) indicates that the method of tissue fixation chosen can significantly affect the results obtained.

3.3.4. QUANTITATIVE PCR ANALYSIS OF RNA FROM MATCHED UMFIX, FFPE AND FF TREATED COLORECTAL TISSUES

RPL13a was used to evaluate the quantity of RNA obtained from UMFIX, FFPE and FF tissues based on average Ct values using equivalent concentrations of starting RNA. One FF sample (Case 6N) was omitted from the data set as an outlier, being greater than two standard deviations away from the mean.

An example of an amplification plot showing the average Ct values for each method of fixation for one colorectal tumour tissue samples using *RPL13a* is presented in Figure 3.5. The FF samples have lower Ct than UMFIX or FFPE samples. The FFPE sample has the highest average Ct, suggesting that FFPE samples are less effective at preserving good quality RNA than FF or UMFIX samples.

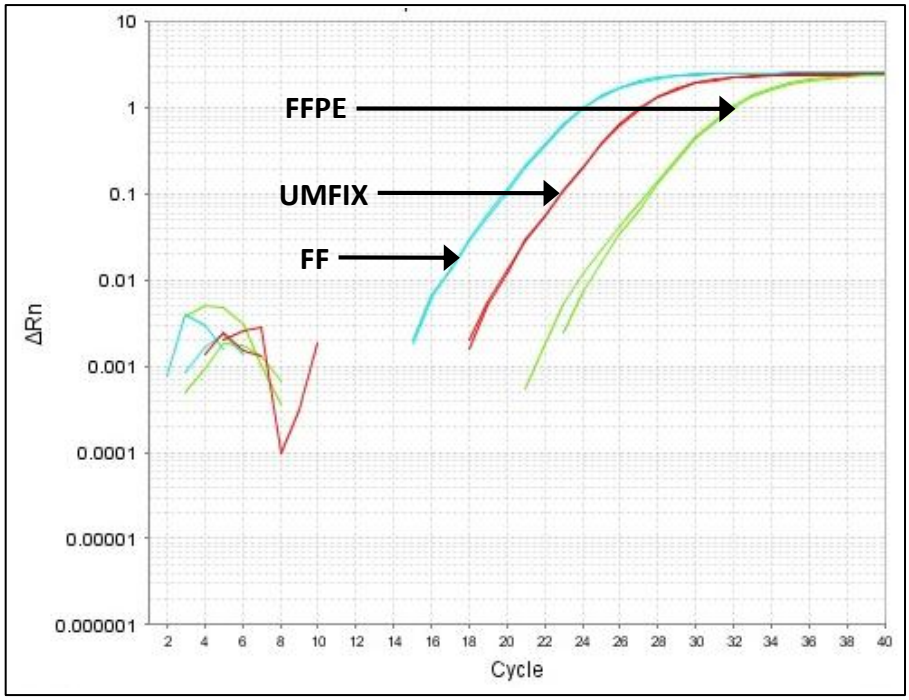

Figure 3.5: Amplification plot of colorectal tumour tissue using *RPL13a*
FF tissues show the lowest Ct, suggesting that this method of tissue fixation is best at preserving RNA. FFPE tissues represent the poorest quality RNA, with a higher Ct than either UMFIX or FF tissues. UMFIX tissues have an intermediate Ct, lying halfway between FFPE and FF tissues.

Figure 3.6 shows scatter plots of average Ct values for the RNA samples. The difference seen between UMFIX and FFPE was smaller than for DNA analysis (Figure 3.3), although this was still statistically significant (Table 3.5).

FF samples have lower average Ct values than UMFIX and FFPE samples (Table 3.4). However, there was a greater variation in the average Ct values of FF samples compared to UMFIX or FFPE fixation.

Table 3.4: *RPL13a* average Ct values for UMFIX, FFPE and FF treated samples

SAMPLE	UMFIX	FFPE	FF
Case 1N	22.6	21.7	19.3
Case 1T	23.0	22.4	22.3
Case 2N	24.8	25.6	21.9
Case 2T	21.2	25.1	21.4
Case 3N	25.0	26.7	24.5
Case 3T	23.4	26.5	18.0
Case 4N	23.5	25.9	21.9
Case 4T	21.1	26.5	21.0
Case 5N	25.9	22.6	23.9
Case 5T	26.1	22.8	24.4
Case 6N	23.9	26.8	28.9
Case 6T	21.9	26.8	22.9
Case 7N	22.5	24.7	20.5
Case 7T	21.4	22.6	16.3
Case 8N	24.1	25.6	19.0
Case 8T	23.7	25.2	17.4

Average Ct values of *RPL13a* show some variation compared to the *GAPDH* Ct values from DNA samples. FF treated samples had lower average Cts in 11/16 samples.

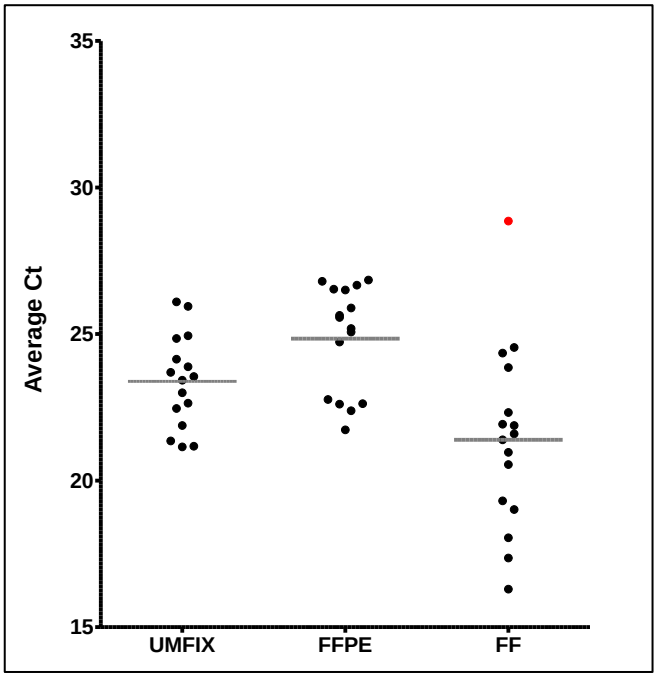

Figure 3.6: *RPL13a* average Ct values for combined normal and tumour samples
A scatter plot showing the Average Ct of *RPL13a* in RNA from UMFIX, FFPE and FF treated samples. The red point indicates the outlier sample which has been omitted from the statistics.

A one way ANOVA test was performed on all samples to compare the three methods of tissue fixation, with Bonferonni's Multiple Comparison test used to compare all pairs of columns (Table 3.5).

Table 3.5: Effect of UMFIX, FFPE and FF fixation on RNA yield

Statistical Test	Post-test	Sample Comparisons	p-value	Significance
One way ANOVA	Bonferroni's Multiple Comparison Test	UMFIX:FFPE:FF	P=0.0004	Yes
		UMFIX:FFPE	NS	No
		UMFIX:FF	NS	No
		FFPE:FF	***	Yes

One way ANOVA with Bonferroni's multiple comparison test was performed to compare the average Cts of samples fixed using the three methods. No significant difference was seen between UMFIX and FF or FFPE tissues. The difference between FFPE and FF treated tissues is highly significant (*, p<0.0001).**

There was a statistically significant difference between the three methods of fixation (p=0.0004, one way ANOVA). In the two group analysis no statistically significant difference occurs when comparing UMFIX with FFPE or FF tissues. Overall, UMFIX appeared to be better than FFPE fixation due to the lower average Ct values, but the difference was not as pronounced as for DNA samples (Table 3.2).

As for DNA, normal and tumour RNA values were analysed separately with the outlier omitted from the analysis (Table 3.6 and Figure 3.7). RNA from normal samples did not show any statistically significant differences between the three fixatives when looking at the one-way ANOVA statistics or Bonferroni's multiple comparison tests.

In the tumour tissues, a statistically significant difference was identified when comparing all three fixatives (p=0.0021, ANOVA). There was no significant difference when comparing UMFIX with FFPE or FF tissues. A statistically significant difference was seen when comparing FFPE and FF tissues (**, p<0.01) (Table 3.6).

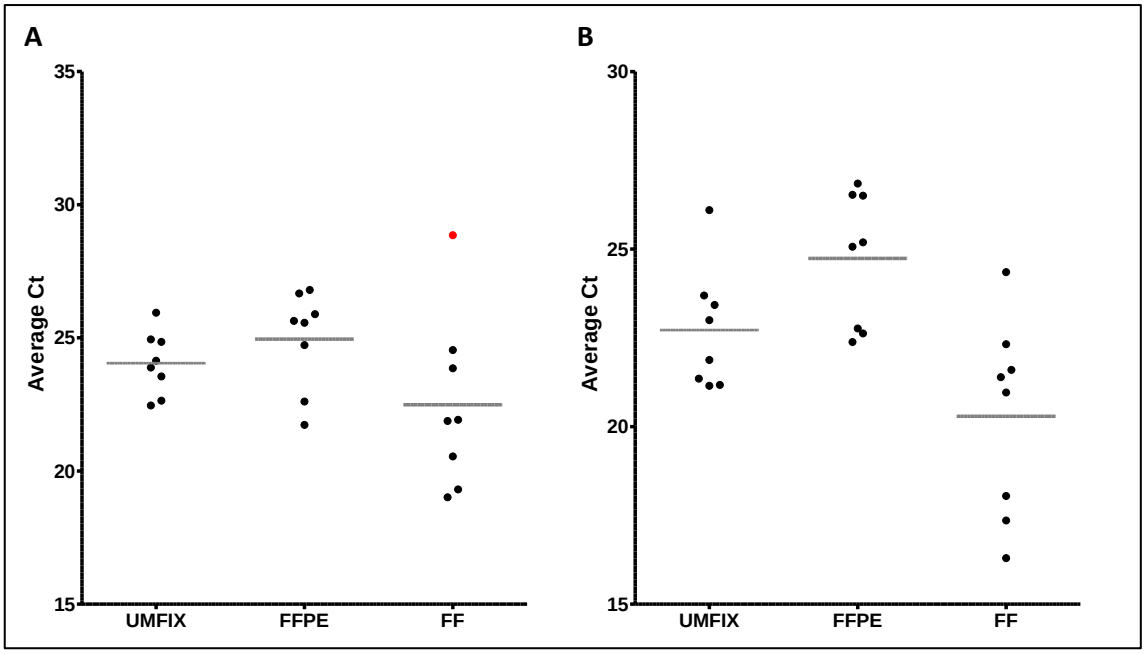

Figure 3.7: RPL13a average Ct values for normal and tumour samples
A) Scatter-plot with mean showing the average Ct values of *RPL13a* in RNA from UMFIX, FFPE and FF treated NORMAL colonic tissue. All normal have been included, the red point is the outlier result, which was omitted from the statistics. B) Scatter plot showing the same information as (A) but for TUMOUR tissue.

Table 3.6: *RPL13a* average Ct values from normal and tumour RNA isolated tissue treated with UMFIX, FFPE and FF

Sample Type	Statistical Test	Post-test	Comparison	p-value	Significance
Normal Colonic Mucosa	One way ANOVA	Bonferroni's Multiple Comparison Test	UMFIX-FFPE-FF	p=0.1118	No
			UMFIX-FFPE	p>0.05	No
			FFPE-FF	p>0.05	No
			UMFIX-FF	p>0.05	No
Colorectal tumour tissue	One way ANOVA	Bonferroni's Multiple Comparison Test	UMFIX-FFPE-FF	P=0.0021	Yes
			UMFIX-FFPE	p>0.05	No
			FFPE-FF	**	Yes
			UMFIX-FF	p>0.05	No

The data were normally distributed and therefore parametric tests were undertaken. The results of the normal tissues did not identify any significant difference between any of the three methods of tissue fixation. Tumour samples identified a statistically significant difference across the three groups (p=0.0021, ANOVA) and between FFPE and FF tissues (, p<0.01). The results suggest that FFPE and UMFIX tissues are comparable in their ability to preserve RNA, since no significant difference was identified.**

A 2-way ANOVA grouped paired analysis revealed that for *RPL13a* the type of fixative (p=0.0060) and also type of tissue (p=0.0127) influence the results (i.e. RNA yield and quality).

3.4. DISCUSSION

3.4.1. THE EFFECT OF TISSUE FIXATION ON NUCLEIC ACID QUALITY AND TISSUE HISTOMORPHOLOGY

This study investigated the tissue histomorphology and nucleic acid quality obtained from matched colorectal tissue fixed using three protocols: UMFIX, FFPE and FF. The tissue histomorphology of the tissues revealed that FFPE and UMFIX tissues were comparable, supporting other studies (Gugic, 2007, Vincek et al., 2003). FF tissues showed poorer staining and were therefore less useful for histological interpretation.

DNA analysis by qPCR revealed that there were significant differences in the average Ct values for each fixative. Lower Ct values are associated with a higher quality starting template since each sample had the same initial quantity of DNA. UMFIX and FF tissues were comparable and gave the lowest average Ct values. This finding was mirrored in both the combined (normal and tumour) and individual (normal vs tumour) analyses. FFPE tissues had higher average Ct values and showed more variation between samples. These findings indicate that UMFIX and FF tissues best preserved the quality of DNA.

The findings for RNA were not as marked as for DNA. UMFIX showed a lower average Ct than FFPE in both combined and individual data sets. However, FF tissues show a lower average Ct and therefore better quality RNA than either UMFIX or FFPE samples. Interestingly, the type of tissue and the fixative were both seen to have an effect on the results obtained. The effect of tissue type may be an artefact but could reflect a difference in expression of *RPL13α* in normal and tumour samples from CRC patients. More cases need to be investigated to confirm this result. Overall however, UMFIX

samples appear to improve the quality of RNA in comparison to FFPE tissues but the difference in this study was not as striking as described previously (Vincek et al., 2003).

Results found in this study confirm those found by previous researchers (Vincek et al., 2003, Cox et al., 2006, Cleary et al., 2005), UMFIX could be introduced as a routine fixative in the future. According to the results presented in this study, the main benefit of replacing formalin fixation with UMFIX lies in UMFIX's ability to better preserve DNA. However, given that both research and clinical fellows are familiar with formalin fixation, it is unlikely that a widespread conversion to UMFIX will take place. Perhaps if further evidence can demonstrate a marked improvement in both RNA and DNA quality, along with adequate histomorphology, UMFIX may become the fixative of choice.

3.4.2. WEAKNESSES AND LIMITATIONS

Unfortunately, only eight prospective CRC patients were consented for this study and other prospective tissues were not collected for use on the DASL microarray. From the results, it appears that the preservation of RNA by UMFIX and FFPE methods is comparable. This finding could be due to the short duration of fixation and exposure to formalin in the Fixation Set samples compared to the Training Set samples, which were fixed for 2-3 years. Overall, more samples are required for full analysis of the different fixatives to fully investigate the advantage of using UMFIX (if any) over FFPE tissues in clinical practice. At present, the main benefit appears to lie in the ability of UMFIX to better preserve DNA.

A further limitation of this study lies in the lack of follow-up of the UMFIX samples. The long term storage of tissues in formalin leads to a reduction in the quality of nucleic

acids. It is thought that the long term storage of tissues in UMFIX does not have this effect. However, since long term follow-up was not performed, for example repeating the experiment after the samples had been stored in UMFIX/FFPE/FF for greater than 12 months, this finding cannot be confirmed or denied. Future follow-up experiments using the samples collected in this study may provide the answer to this question.

Finally the ability of each method of tissue fixative to preserve protein was not investigated. Perhaps in the future, the samples collected in this study can be investigated further to see the difference, if any, in protein quality according to the method of tissue fixation used.

3.5. CONCLUSIONS

In conclusion, UMFIX and FFPE methods were shown to be comparable in their ability to preserve tissue architecture and morphology, but there were differences in DNA and RNA yield. UMFIX treated samples had lower average Cts than FFPE tissues for both RNA and DNA, indicating that the quality of these nucleic acids in UMFIX tissues was better, but FF tissues were better at preserving the quality of RNA than either UMFIX or FFPE.

.

CHAPTER 4: EVALUATION AND VALIDATION OF THE DASL WHOLE-GENOME EXPRESSION ASSAY IN COLORECTAL TISSUES

4.1. INTRODUCTION

The 'adenoma-carcinoma sequence' described by Vogelstein *et al* (1988) represents a model for the development of the majority of CRCs. The sequence of events leading to CRC occurs over a period of 10-15 years, providing a therapeutic window for CRC prevention (Heijink et al., 2011). With the advent of microarray technology, the ability to perform whole genome expression analysis has become possible. Microarray technology has the ability to reveal potential diagnostic and molecular targets of human diseases. The gene expression profiles of the tissues in the adenoma-carcinoma sequence could provide insights into CRC tumorigenesis.

Although colorectal adenomas (APs) represent the precursor to CRC, relatively few studies have profiled APs and other tissues from early stages of the 'adenoma-carcinoma sequence'(Heijink et al., 2011, Lin et al., 2002, Kita et al., 2006, Sabates-Bellver et al., 2007, Galamb et al., 2008b, Galamb et al., 2008a). Conversely, hyperplastic polyps (HPs) represent largely non-neoplastic lesions with little neoplastic potential (Bond, 2000). Only limited literature is available comparing gene expression profiles of HPs and APs (Chen et al., 2008, Galamb et al., 2008b), however, the results demonstrate clear gene expression differences between HPs and APs, warranting further investigation.

With the explosion of microarray applications, data analysis and interpretation has become a potential bottleneck. Before in-depth analysis of gene expression profiles and pathway analysis can be undertaken, the results must first be validated through the use of appropriate quality control (QC) steps and independent validation tools such as qRT-PCR and literature review (Chuaqui et al., 2002). Since the technology

behind the DASL microarray is novel, the need for an effective pre-processing and QC analysis package is paramount. Lumi, a package available on R Bioconductor, enables Illumina specific pre-processing and QC methods to be performed. These include several methods of data normalisation as well as VST, which takes into account the technical replicates present on the Illumina microarray (Du et al., 2008).

This chapter describes the design of the microarray experiment and presents the findings of the QC methods performed using lumi. In addition this chapter describes the differential gene and pathway expression between different colorectal tissues. Candidate EC genes were selected based on the results of the microarray to allow normalisation of candidate genes chosen for validation by qRT-PCR (Chapter 5). *TPT1* and *UBC* were selected as endogenous control (EC) genes based on previous data in colorectal tissues (Andersen et al., 2004).

Hypothesis: The hypothesis to be tested is that Whole-Genome DASL microarrays will identify candidate genes showing differential expression across different colorectal tissues.

4.2. AIMS AND OBJECTIVES

The aims of this chapter were to investigate patterns of differential gene expression between different colorectal tissues, perform pathway analysis and explore the comparison between HPs and APs in detail. Prior to this, validation of the expression microarray is required.

Objectives:

- QC steps available in lumi will be utilized to ensure that the results of the microarray experiment are reliable, and identify outlier samples that require removal from the analysis.
- Replicate analysis of the repeated cases A190N and A190P will be performed to ascertain if the microarray data are reproducible.
- *In-silico* analysis will be used to confirm the gene expression changes seen on the microarray using a review of previously published data.
- Limma will be used to create top-tables of differential expression between different types of colorectal tissues.
- To investigate differential gene expression of genes between specific tissues and perform pathway analysis to identify pathways that show significant differential expression using limma, MEV and DAVID.
- To investigate specifically the differential gene expression and pathway associations between HP and AP tissues.

4.3. VALIDATION OF RNA TEMPLATES IN TRAINING AND VALIDATION SETS

Training Set samples were first evaluated for RNA quality with *RPL13a* qRT-PCR, as recommended by Illumina (samples with a Ct<29 are deemed to be of significant quality for use with the DASL array).

The average Ct values of *RPL13a* in Training Set samples were compared with FFPE samples from the Fixative Set (Chapter 3). The Training Set samples have a higher

average Ct than the FFPE tissues from the Fixative Set ($p<0.0001$, unpaired t-test) – (Figure 4.1).

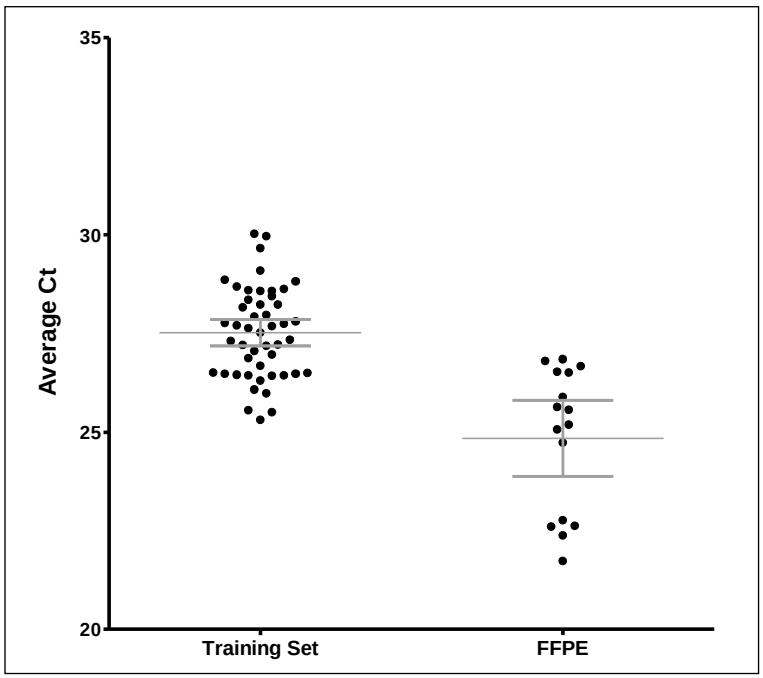

Figure 4.1: Comparison between average Ct values of *RPL13α* for the Training and Fixative Set

The majority of samples in the Training Set achieved the criteria to ensure adequate RNA quality (2.5.1) with the exception of two samples with mean Ct values between 29 and 30. These samples met the other criteria and were therefore included in the Training Set.

Figure 4.2 shows a breakdown of the Training Set samples by tissue type. The data were normally distributed and one way ANOVA revealed no significant differences between tissue type ($p=0.2381$).

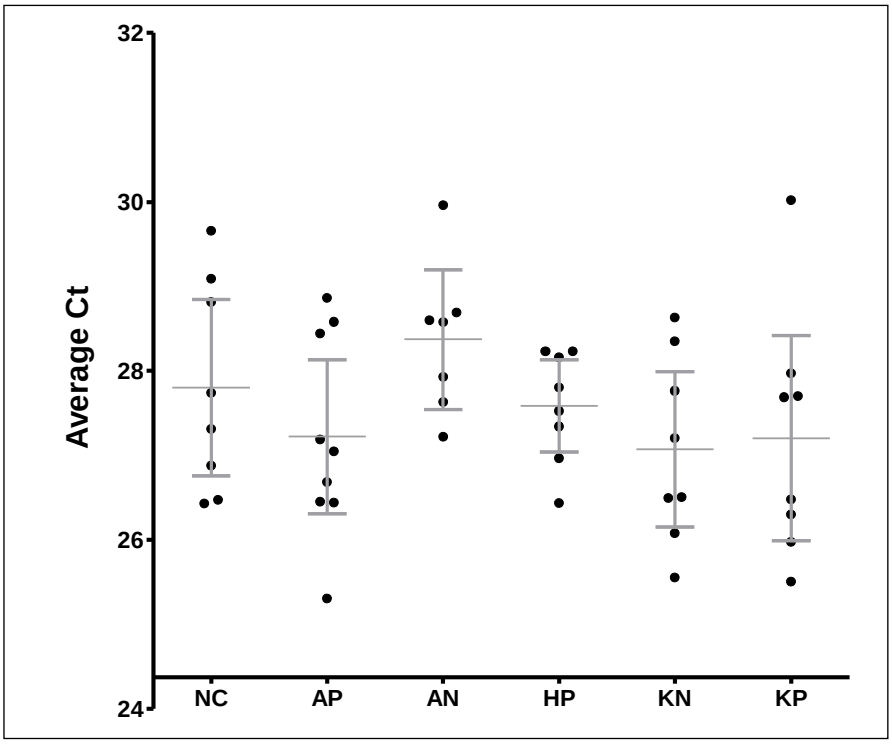

Figure 4.2: *RPL13a* average Ct values in the different tissues present in the Training Set

The Validation Set consists of other FFPE samples collected to validate the results of the DASL microarray experiment by qRT-PCR. *UBC* was used as an EC gene to help identify samples that were not suitable for further analysis. *UBC* was used instead of *RPL13a* as samples were precious and *UBC* was the EC gene that would be used to normalise the qRT-PCR results from the Validation Set. The average Ct of *UBC* in the Validation Set was compared to the average Ct of *UBC* in 27/48 samples from the Training Set. There was insufficient RNA available after the DASL microarray for the remainder of samples to enable analysis with *UBC*.

The *UBC* results showed no significant difference between the two groups (p=0.7399, unpaired t-test). However, some outliers greater than 2 standard deviations away from the mean were identified (Figure 4.3). In addition these samples did not fulfill the desired $A_{260/280}$ and $A_{280/230}$ ratio >1.8. Therefore these samples were not included in any other analyses.

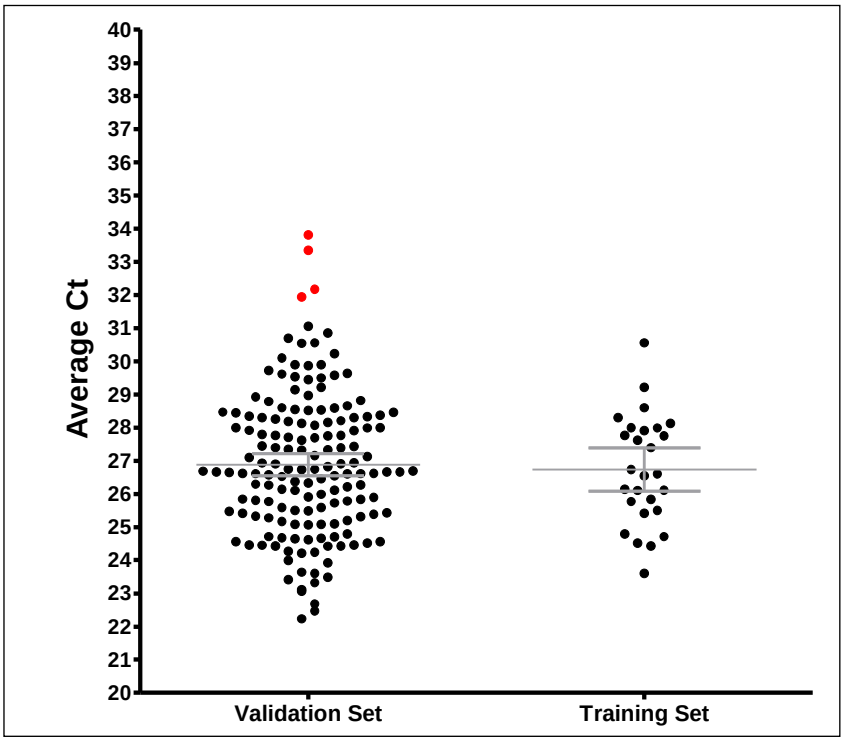

Figure 4.3: Average Ct of *UBC* in the Validation and Training Sets

Samples highlighted in red represent those samples that lie greater than two standard deviations from the mean. These samples were subsequently removed from future experiments/analyses.

4.4. MICROARRAY DESIGN

The design of the microarray experiment allowed multiple group comparisons to be made (Figure 4.4).

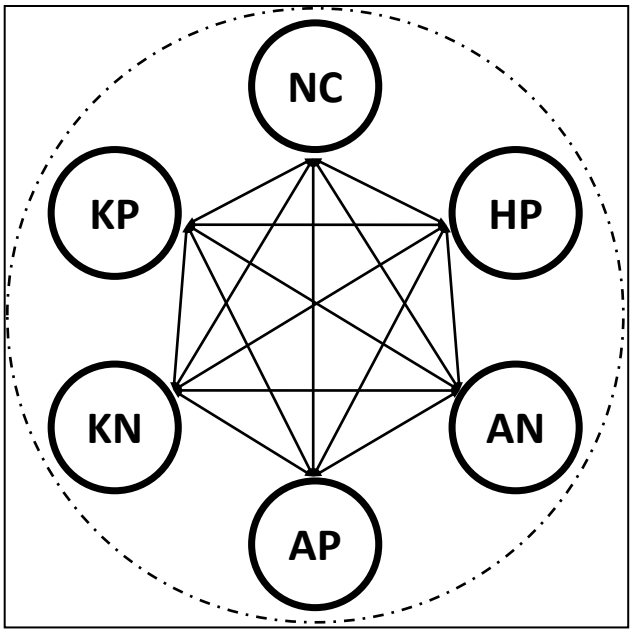

Figure 4.4: Microarray Design Matrix
This figure represents the six different tissue types included on the microarray. The matrix in the centre of the figure shows the potential comparisons between the different groups. NC, normal controls; HP, Hyperplastic Polyp; AN, Normal colonic mucosa from patients with adenomatous polyps but no CRC; AP, Adenomatous polyps from patients without CRC; KN, normal colonic mucosa from patients with adenomatous polyps (APs) and CRC; KP, Adenomatous polyps from patients with CRC.

4.5. QUALITY CONTROL ANALYSIS OF DASL MICROARRAY

4.5.1. DATA TRANSFORMATION AND BACKGROUND CORRECTION

Following the creation of the raw data charts presented in 4.5.2, the data was background corrected and transformed using the variance stabilizing transformation (VST) method available in lumi. VST is crucial for downstream analysis to identify statistically significant changes in gene expression.

4.5.2. NORMALISATION OF MICROARRAY DATA

The raw data from the microarray was assessed using the lumi package in R Bioconductor. Box-plots, MA plots and distance matrix heat maps revealed three

outlier samples (NC292, A161N and A161HP) that were subsequently removed from the analysis.

Normalisation of the data was undertaken using the quantile normalisation method available in the lumi package. Box-plots of the Log_2 Intensities of the samples before and after normalisation are presented in Figure 4.5. This chart clearly identified a lack of intensity in A161N, warranting its removal from subsequent analysis. In addition, NC292 showed a wider range of intensities and a lower mean suggesting that this sample should be removed from the analysis. Following normalisation, the intensity of this sample was corrected and became comparable to the remaining samples. Since the raw intensities of these samples clearly identified a problem they were omitted from further analysis to prevent the incorporation of sample bias into the results.

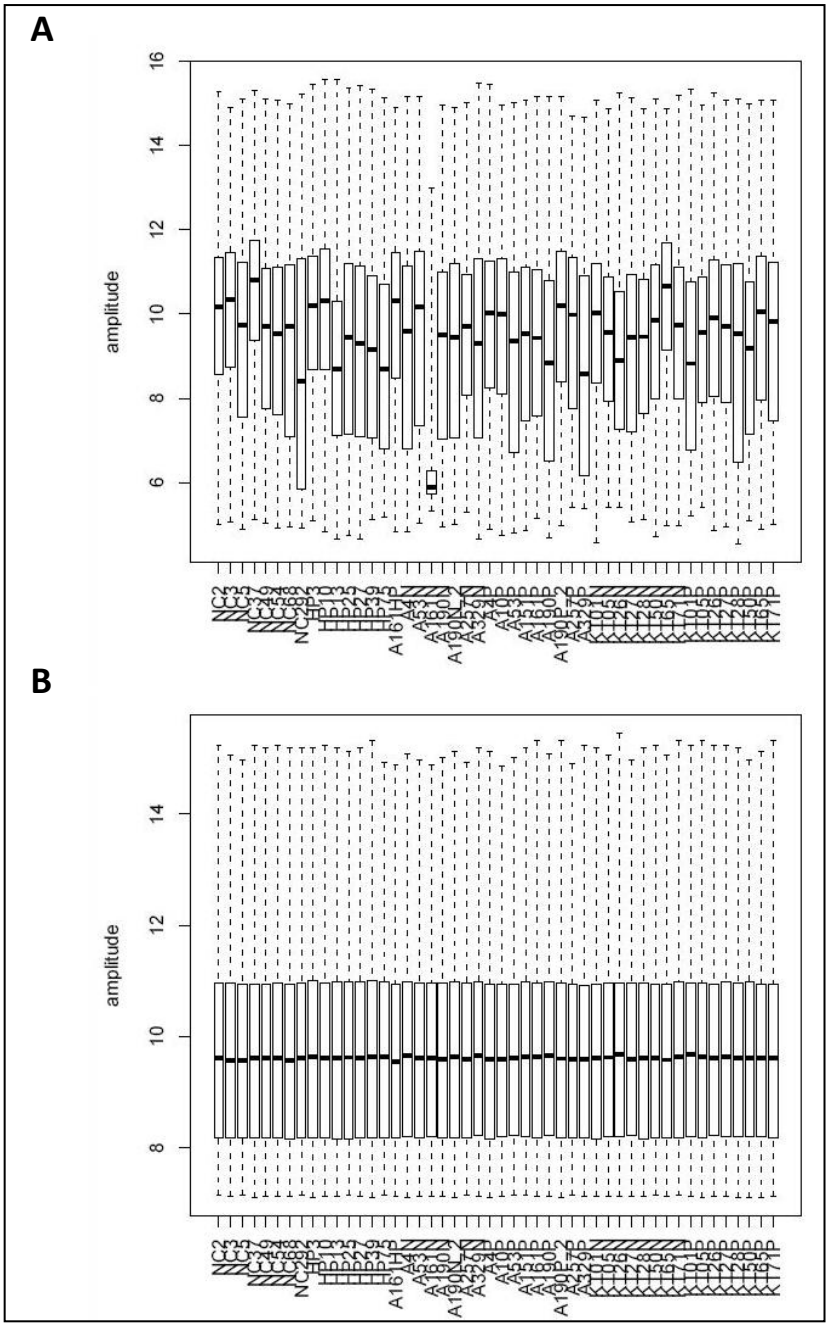

Figure 4.5: Box-plots of arrays before and after normalisation
A) Raw Log_2 intensities of the samples. A161N is identified as an outlier as depicted by the low-level Log_2 intensities as compared to the other samples. B) Log_2 intensities of the samples after quantile normalisation. All samples are now standardised.

MA plots were constructed to allow the pair-wise comparison of Log-intensities of each array to a reference array. This allowed the identification of intensity-dependent biases. A sample of MA plots is presented in Figure 4.6. A161N is included to show additional evidence that this array needs to be removed from the analysis.

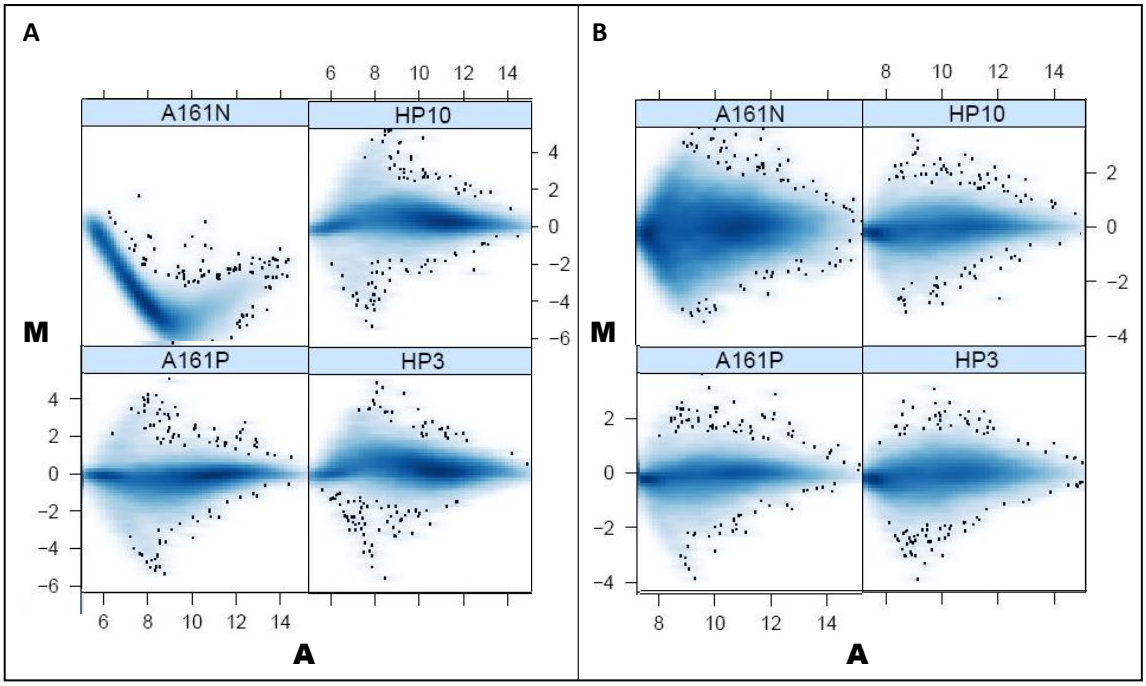

Figure 4.6: MA plots before and after normalisation

M (y-axis) and A (x-axis) are defined as M=log2(I1) – log2(I2) and A=1/2(log2(I1)+log2(I2)). I1 is the intensity of the array of interest and I2 is the intensity of a pseudo array, which has the median values of all arrays. The distribution should be concentrated along the M=0 axis. A) MA plots of four samples before normalisation. A161N shows an abnormal plot indicating that this array has not worked. The remaining three plots show the MA plot centered on M=0 indicating little intensity-dependent biases. B) MA plot of the same four samples shown in (A) after normalisation has occurred. The normalisation process has corrected the A161N intensity profile however this array will still be omitted from future analysis.

A sample distance matrix, which represents the homogeneity of the samples, is shown in Figure 4.7. The generation of these matrices helped to identify A161HP as an outlier in the data.

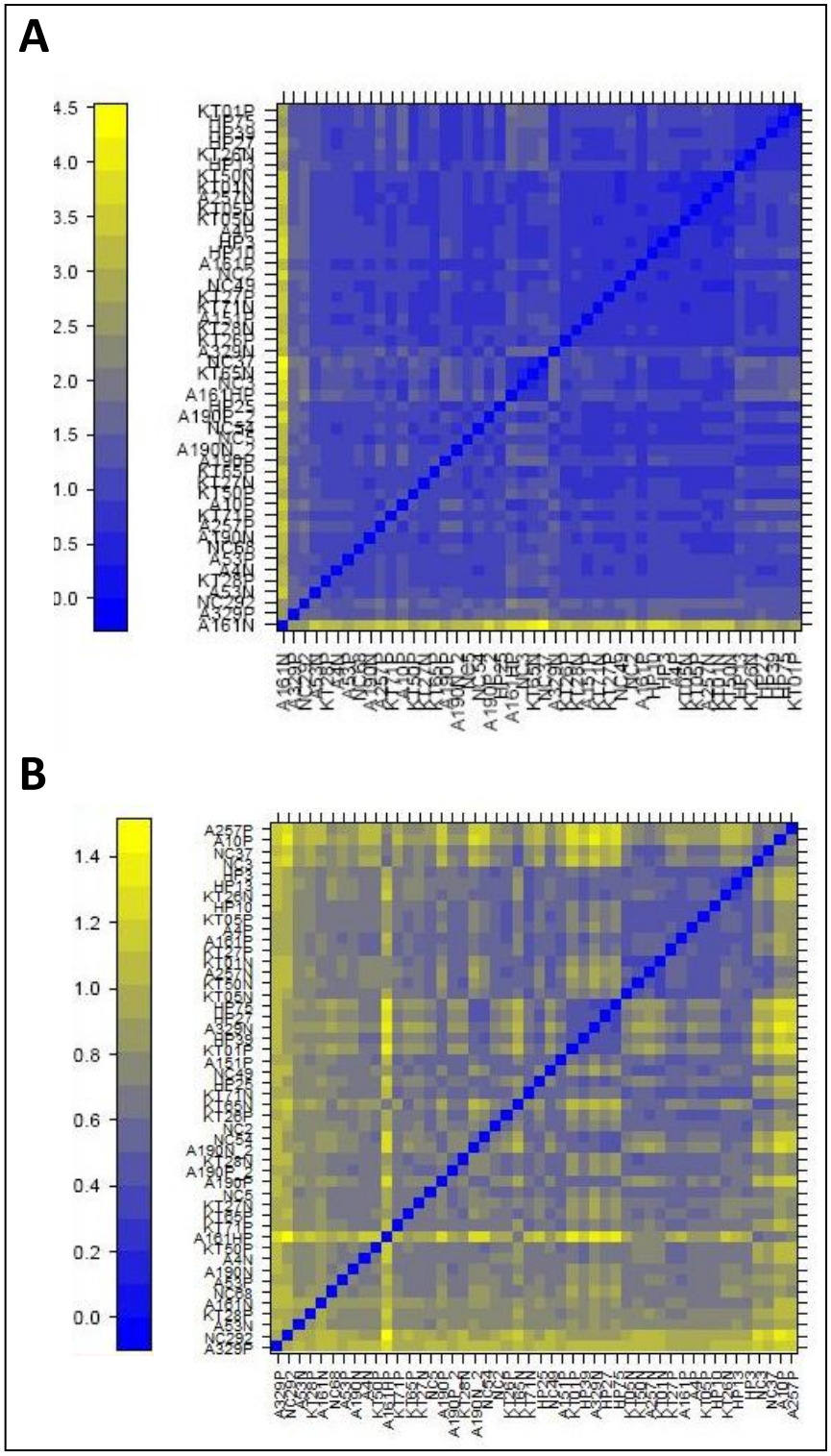

Figure 4.7: Sample distance matrices of all arrays before and after normalisation
These matrices reveal the similarity between arrays by comparing the median absolute distance of the M-value for each pair of arrays. The blue diagonal line shows 100% sample similarity as each sample intersects with itself. A) Before normalisation matrix indicating that A161N is dissimilar to the other arrays. B) Following normalisation, more variation is seen across the arrays with A161HP being identified as an outlier due to its highly dissimilar profile to other arrays and "channel-effect" produced.

To further assess the homogeneity between the arrays, Density plots of Log-intensity distributions and Empirical Cumulative Distribution Functions (ECDFs) were created. The ECDFs plot the cumulative distribution of intensity values for each of the arrays. Altered distributions can represent outlier or failed arrays, which can be removed from subsequent analysis. A161N showed an abnormal profile on both of these charts, however all other samples followed the expected profile. Figure 4.8 shows a representative for each chart before and after normalisation for a subset of arrays.

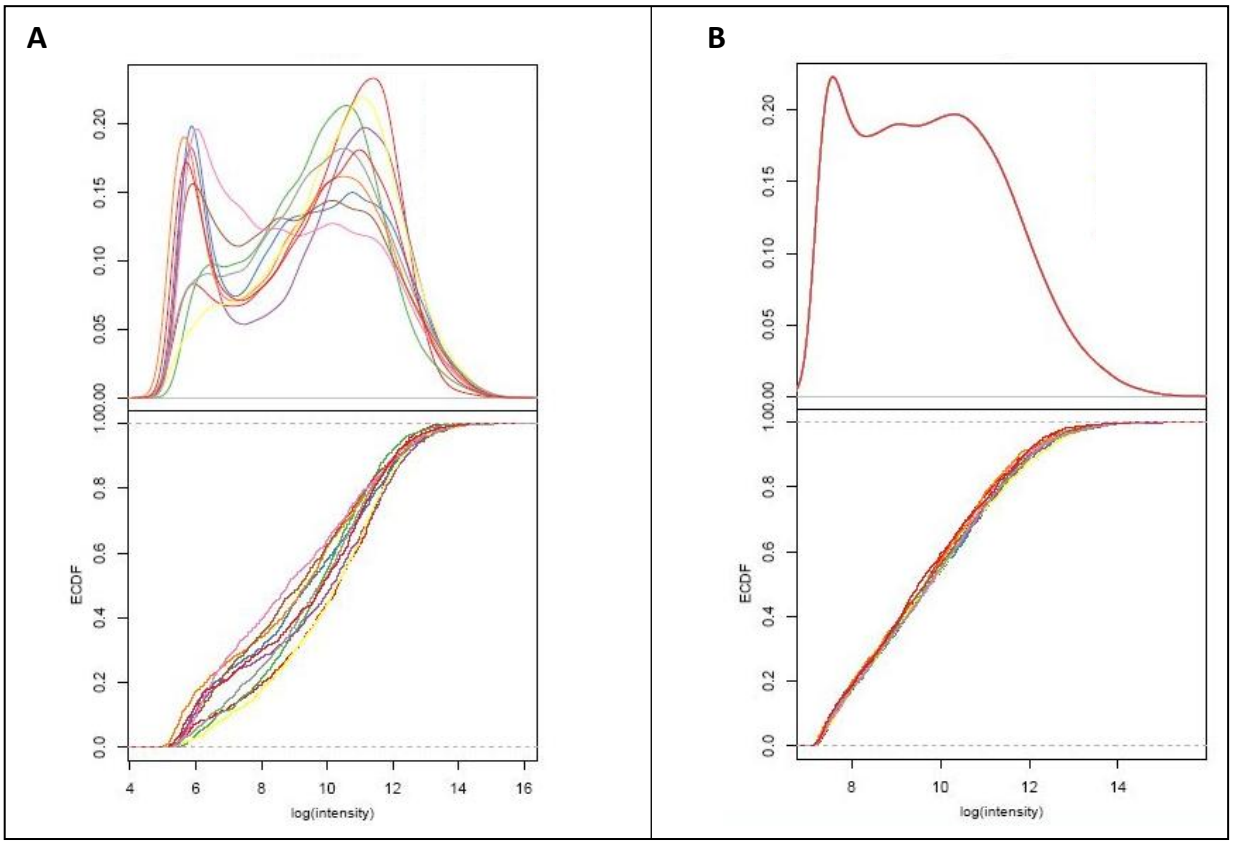

Figure 4.8: Density plots of Log-intensity distributions and ECDFs
A) Density and ECDF plots before normalisation for a sample of arrays. In the log-intensity density plot there is no significant shift on the x-axis, and all arrays follow a similar distribution, with slight variation seen between arrays. The ECDF reveals some differences in the intensity values for the arrays investigated but the profile remains similar for all arrays studied. B) Density and ECDF plots for the same arrays as shown in (A) following normalisation. The distribution of all samples is now uniform with little variation.

Variance mean dependency plots were also generated and are presented in Figure 4.9. Normally, higher intensities show a greater variation. Following normalisation, the red-curve should be approximately horizontal showing no substantial trend.

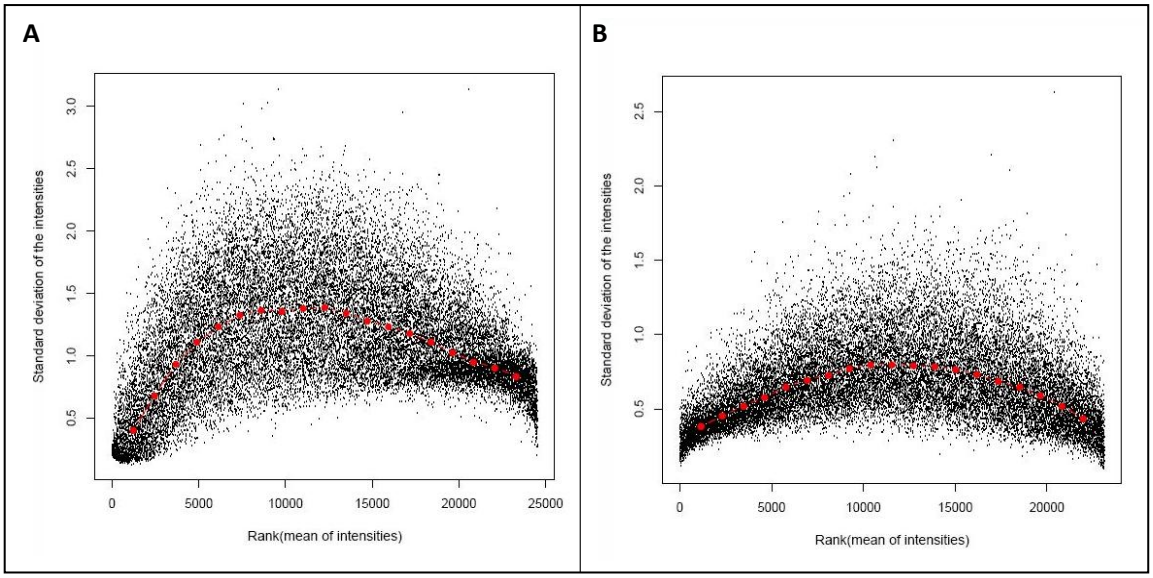

Figure 4.9: Variance mean dependency plots

These plots chart the standard deviation of the intensities against the rank of the mean intensities. The red-line depicts the median of the standard deviation. A) Before normalisation, the red-curve appears to be skewed, however following normalisation (B) the curve appears to follow a normal distribution with no trend identified.

4.5.3. TECHNICAL REPLICATE ANALYSIS

Duplicate RNA extractions were obtained from A190N and A190P in order to assess the reliability of the RNA extraction protocol. The Bland-Altman method and Intraclass correlation coefficients were calculated for both A190N and A190P repeats. Scatter plots and Bland-Altman plots with 95% limits of agreement are shown in Figure 4.10. The bias was not significantly different from zero and both plots show narrow 95% limits of agreement as shown in Figure 4.10 and Table 4.1. This suggests that the results were reproducible. The intraclass correlation coefficient for A190N and A190P were 0.90 (95% Confidence Interval 0.89-0.90) and 0.92 (95% Confidence Interval 0.91-0.92) respectively.

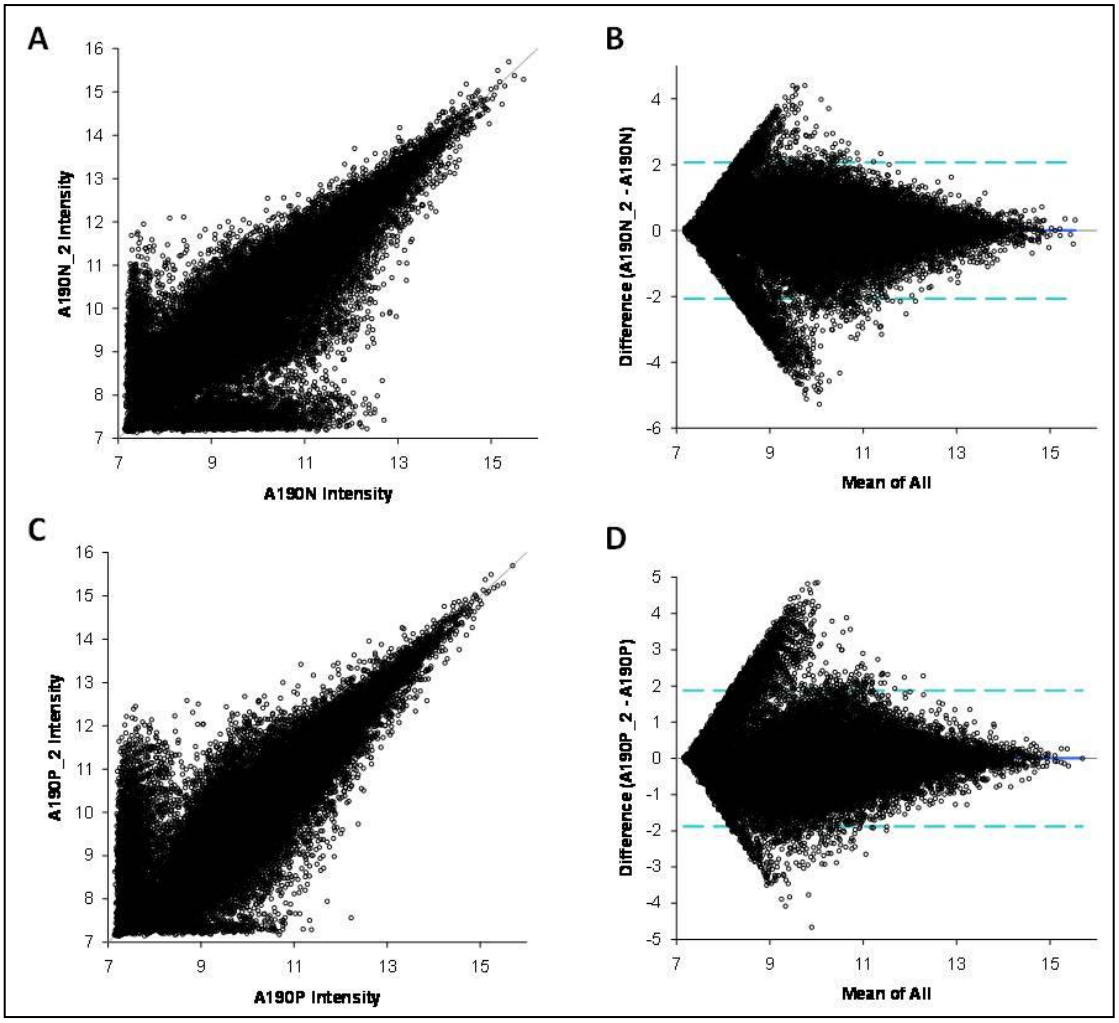

Figure 4.10: Scatter and Bland-Altman plots for technical replicate analysis of A190N and A190P

A) Scatter plot comparing the intensities of A190N and A190N_2 for all genes present on the microarray. B) Bland-Altman plot showing the 95% limits of agreement for A190N and A190N_2. C) Scatter plot comparing the intensities of A190P and A190P_2 for all genes present on the microarray. D) Bland-Altman plot showing the 95% limits of agreement for A190P and A190P_2.

Table 4.1: Bland-Altman plot summary

Replicate Pair	Bias	Limits of Agreement	P-value
A190N and A190N_2	0.00	-2.07 to 2.07	1.00
A190P and A190P_2	0.00	-1.88 to 1.88	1.00

This table shows the bias (average difference between variables), the 95% limits of agreement and the p-value, which is the probability of rejecting the null hypothesis (that the bias is equal to zero). Here we find that the p-value is not significant, implying that the bias is not significantly different from zero. This suggests that there is good reproducibility between the replicates.

4.5.4. SELECTION OF ENDOGENOUS CONTROLS

To verify that *UBC* and *TPT1* were appropriate EC genes, NormFinder (Andersen et al., 2004) was employed to investigate a panel of potential EC genes from the array data. The EC genes were selected from two publications (Andersen et al., 2004, Kheirelseid et al., 2010). These potential EC genes were identified in the normalised Training Set microarray data and analysed using NormFinder. The samples were divided into the six histologically distinct colorectal tissue groups previously mentioned; NCs, HPs, ANs, APs, KNs and KPs, and expression of the EC genes was compared across the six groups.

Figure 4.11 shows the stability value for each of the candidate EC genes. *TPT1* was identified as the most stably expressed gene across the different groups of colorectal tissues, followed by *UBC*. These genes have been highlighted in red.

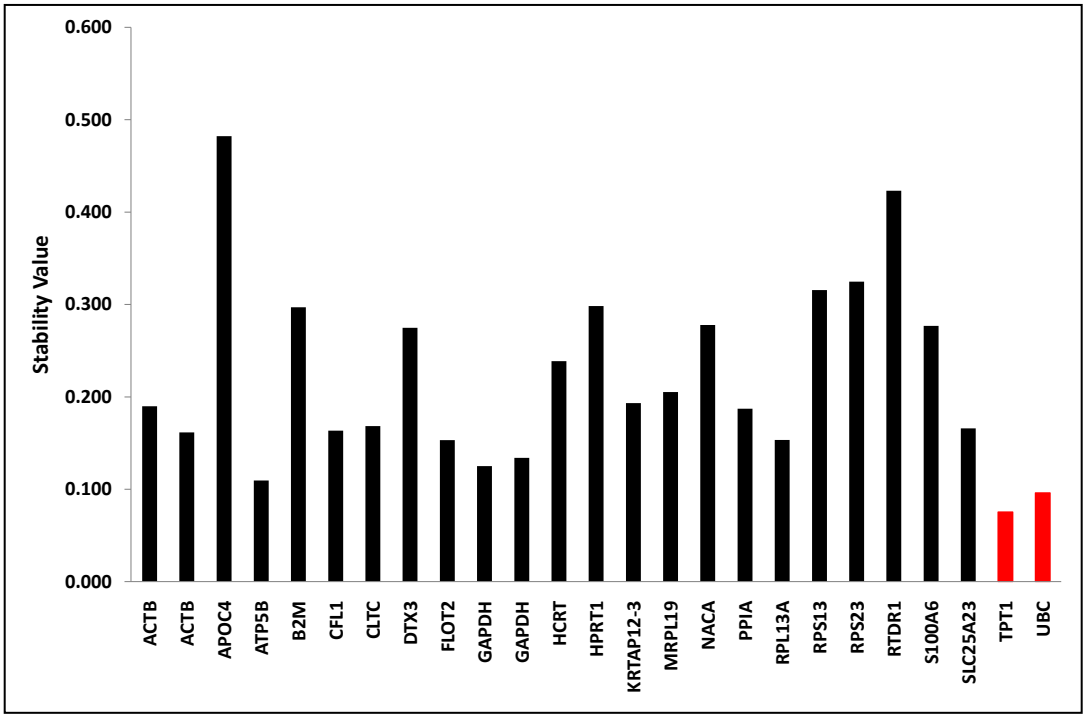

Figure 4.11: Stability values of candidate EC genes
This chart represents the variation in gene expression across the six sample groups from the normalised Training Set microarray data. The most stably expressed genes have a stability value close to zero since a stability value of zero represents no variation in gene expression across the different groups. The best two EC genes were *TPT1* and *UBC*, which have been highlighted in red.

4.6. INTERROGATION OF MICROARRAY DATA

4.6.1. GENERATION OF TOP-TABLES

Following normalisation, comparisons were made between the different groups to identify changes in gene expression. The limma (Linear Models for Microarray Analysis) package within R Bioconductor was utilized to create top-tables of gene expression using a moderated t-statistic based on comparisons between two groups of samples e.g. Group A vs Group B.

Log-fold changes were noted for all genes with a significant adjusted p-value ($p<0.05$). The log-fold changes enabled the directionality of a gene's expression to be established. Genes showing similar log-fold changes across several groups were removed from the candidate gene list as they are unlikely to enable differentiation between tissue types. For example, *KLK11* showed a 2.54 fold change in APs relative to NCs, and a 2.6 fold change in HPs relative to NCs. However, only a 0.3 fold change was noted in APs relative to HPs. This gene would therefore not be useful in differentiating between HPs and APs.

Table 4.2 shows a summary of the sample comparisons performed using limma, along with the number of genes identified as having significantly different expression, according to raw and adjusted p-values. Several of the comparisons did not identify any significant differences in gene expression according to the adjusted p-value. However, for all comparisons, the raw p-value yielded many DEGs. The vast number of genes deemed significant according to raw p-value makes it difficult to recognise biologically meaningful genes, which have altered gene expression in tissues of interest.

Table 4.2: The number of differentially expressed genes identified in the different comparisons

Comparisons	Raw P-value (p<0.05)			Adjusted P-value (p<0.05)		
	<0.05	<0.03	<0.01	<0.05	<0.03	<0.01
Q1: Is there differential gene expression between NC, HP and AP tissues?						
HP vs AP	6685	5212	2949	1633	854	196
NC-HP	5573	4205	2243	504	164	8
NC vs AP	2901	2099	1096	197	128	31
Q2: Does gene expression alter across different types of normal-appearing colorectal mucosa?						
NC-AN	2748	1781	656	0	0	0
NC vs KN	1155	723	302	0	0	0
AN vs KN	2807	1852	669	0	0	0
Q3: Does gene expression alter across different types of colorectal polyp?						
NC vs KP	3131	2150	935	8	4	0
HP vs KP	2357	1582	675	2	0	0
AP vs KP	2755	1824	736	0	0	0
HP vs All Polyps (APs + KPs)	5008	3712	1850	302	110	11
Q4: Is differential gene expression observed in matched cancer-associated normal and polyp tissues (KN and KP)?						
KN vs KP (MATCHED)	845	376	94	0	0	0
Q5: Are changes in gene expression associated with clinicopathological parameters?						
Left vs Right ANs	1688	1141	536	25	20	9
Left vs Right KNs	500	274	92	0	0	0
NC vs Left Normals (ANs + KNs)	1423	889	337	0	0	0
NC vs Right Normals (ANs + KNs)	1334	797	276	0	0	0
NC vs Left Polyps (APs + KPs)	2402	1713	756	14	4	0
NC vs Right Polyps (APs + KPs)	2601	1828	825	31	20	7
HP vs Left Polyps (APs + KPs)	4175	3020	1447	73	12	0
HP vs Right Polyps (APs + KPs)	3882	2728	1292	49	14	8
Left vs Right APs	1214	713	238	0	0	0
Left vs Right KPs	1093	664	258	0	0	0
Left Normals vs Left Polyps	738	450	149	0	0	0
Right Normals vs Right Polyps	1499	939	383	3	3	0
Additional comparisons						
AN vs AP (MATCHED)	934	548	197	15	11	6
NC vs All Normals (ANs + KNs)	1479	934	346	1	1	1
NC vs All Polyps (APs + KPs)	2892	2116	1060	119	59	13

This table shows the different sample comparisons that were made during the generation of the top tables according to questions asked. The number of DEGs identified using raw (no FDR adjustment) and adjusted p-values (FDR correction) are identified at the 0.05, 0.03 and 0.01 significance level.

Differential gene expression is of interest in normal colonic tissue (NC vs AN vs KN), colorectal polyps (HP vs AP vs KP), and a mixture of all three (NC vs HP vs AP). In addition to the top-tables (Table 4.2), three Venn diagrams are presented in Figure 4.12 to represent the number of common significant genes between the different colorectal tissues.

Comparing NCs, ANs and KNs no genes were identified as showing differential expression across all three groups according to the adjusted p-value ($p<0.05$).

Comparing NC, HP and AP tissues there were 358 significant genes, according to raw p-value, which were common to all three tissue types. However, using the adjusted p-value only one gene was significant across all three groups; *CARS,* which showed significant up-regulation in both HPs and APs relative to NCs, but did not show differential expression between HPs and APs (as seen with *KLK11* above).

The final Venn diagram identified DEGs between different colorectal polyps (HP, AP and KP). Again, no genes were identified as being differentially expressed across all three tissues using the adjusted p-value. Additionally, when comparisons are made with KP samples, only two genes are identified as having differential expression according to the adjusted p-values; *HLA-G* and *ANXA2*. Both of these genes were down-regulated in KPs relative to HPs.

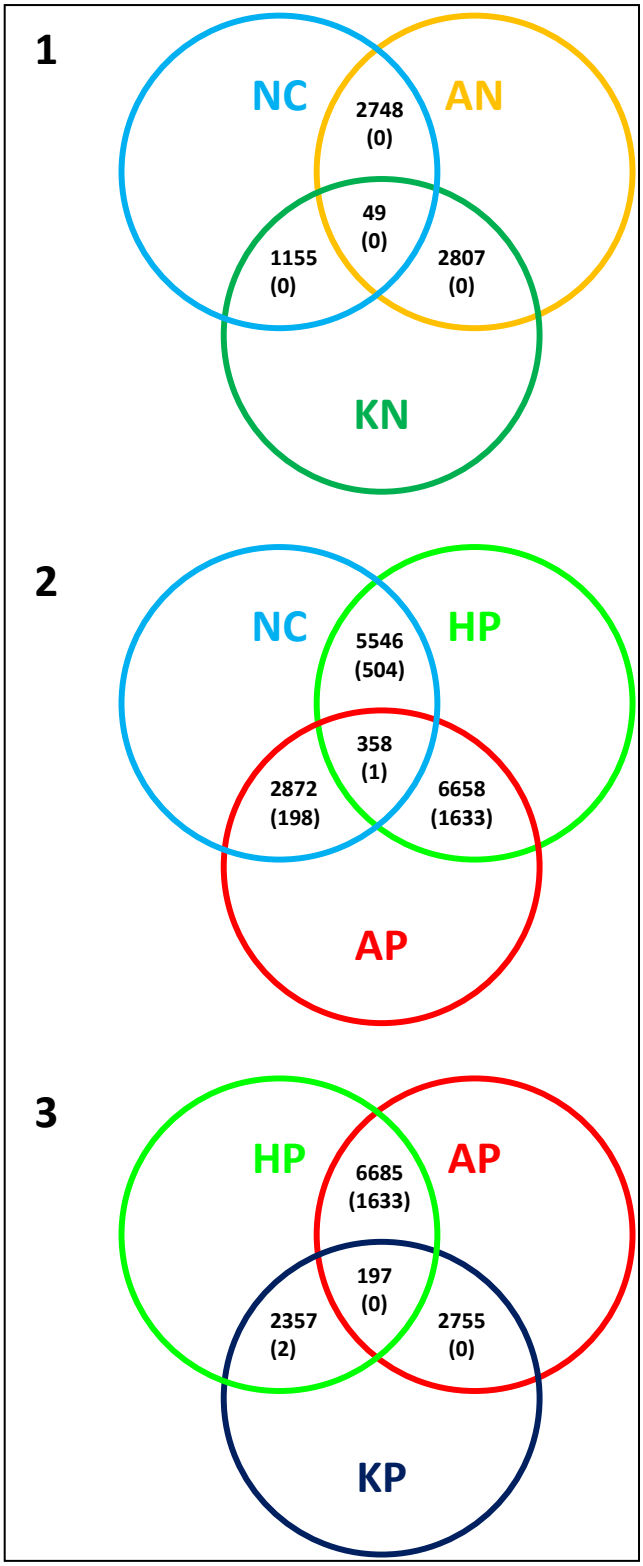

Figure 4.12: Venn Diagrams showing the number of DEGs in different colorectal tissues
This figure represents the number of significant DEGs between different colorectal tissues according to the raw and adjusted p-values ($p<0.05$). The adjusted p-value is presented in brackets. 1 – NC, AN, KN comparison reveals no DEGs using the adjusted p-value. 2 – NC, HP and AP samples have one commonly differentially expressed gene across the three tissues (*CARS*). This Venn diagram is the only one to show significantly DEGs according to the adjusted p-value. 3 – HP-AP, KP comparison does not identify any commonly differentially expressed gene across the three groups. The coloured circles represent NCs (light blue), HPs (light green), ANs (yellow), APs (red), KNs (dark green) and KPs (dark blue).

Table 4.2 and Figure 4.12 show that the majority of DEGs were found when comparing NC, HP and AP samples. Based on these data, candidate genes for qRT-PCR validation were selected from the top-tables that compared these three tissues groups: NC vs HP, NC vs AP and HP vs AP. Specific criteria used to select the candidate genes were:

- Limma top-table findings including:
    - adjusted p-value (<0.05),
    - Actual FC (-2<, >2)
    - B statistic (>2)
- Differential gene expression across NCs, HPs and APs

The top-tables for the three comparisons comparing NCs, HPs and APs have been compiled together to provide the top 20 up- and down- regulated genes presented in Table 4.3 and Table 4.4.

Table 4.3: Top 20 up-regulated DEGs in NCs, HPs and APs

Gene Symbol	Fold Change (Log/Adjusted)	Actual p-value	B statistic	Comparison
GALE	1.53/2.90	0.000098	10.23	NC vs HP
AXIN2	2.02/4.05	0.00057	7.729	NC-AP and HP vs AP
ASCL2	1.59/3.01	0.00113	6.23	NC vs AP and HP vs AP
SDC3	3.46/11.03	0.00121	6.81	NC vs AP
C21ORF49	1.13/2.19	0.00124	2.19	HP vs AP
ANXA2	1.09/2.13	0.0013	7.10	NC vs HP
CARS	1.15/2.23	0.0013	6.82	NC vs HP and NC vs AP
REG4	2.27/4.83	0.0013	6.72	NC vs HP
CAPS	1.56/2.94	0.00155	5.67	HP vs AP
SIL1	0.88/1.84	0.00177	5.85	NC vs AP
CAPN12	1.47/2.77	0.00177	6.26	NC vs AP
LY6G6D	2.72/6.60	0.00177	5.91	NC vs AP
ETS2	1.35/2.55	0.00188	5.69	NC vs AP and HP vs AP
MID1IP1	1.94/3.84	0.00197	6.15	NC vs HP
G3BP1	1.13/2.19	0.00266	4.60	NC vs AP and NC vs HP
HPCAL1	1.68/3.21	0.00269	4.59	HP vs AP
PCDH17	1.02/2.04	0.00306	4.22	HP vs AP
FGFBP1	2.14/4.40	0.00315	5.43	NC vs HP
SPINK4	2.56/5.88	0.00315	5.45	NC vs HP
NPDC1	1.84/3.59	0.00342	4.44	NC vs AP

The top 20 up-regulated genes were compiled from the following comparisons: NC vs HP, NC vs AP, and HP vs AP. The most significant genes were selected according to adjusted p-value. The genes are up-regulated in the second group relative to the first group. For example GALE is up-regulated in HPs relative to NCs and AXIN2 is up-regulated in APs relative to HPs and NCs.

Table 4.4: Top 20 down-regulated DEGs in NCs, HPs and APs

Gene Symbol	Fold Change (Log/Adjusted)	Actual p-value	B statistic	Comparison
KRT20	-2.36/0.20	0.00025	8.98	HP vs AP
CHGA	-2.83/0.14	0.00025	9.32	HP vs AP
MALL	-2.52/0.17	0.00057	7.60	HP vs AP
SLC26A3	-1.87/0.27	0.00082	7.77	NC vs AP
GUCA2A	-1.47/0.36	0.00082	8.00	NC vs AP
CLDN23	-2.10/0.23	0.001	6.70	HP vs AP and NC vs AP
ENTPD5	-1.28/0.41	0.001	6.70	HP vs AP
PRDX6	-1.26/0.42	0.001	6.53	HP vs AP
CIDEC	-2.70/0.15	0.001	6.42	HP vs AP
ABCC13	-1.86/0.28	0.001	6.99	NC vs AP
HIGD1A	-1.47/0.36	0.00124	6.05	HP vs AP
SLC26A3	-1.65/0.32	0.00155	5.56	HP vs AP
AAK1	-1.52/0.35	0.00155	5.54	HP vs AP
FABP1	-2.43/0.19	0.00175	5.34	HP vs AP
TICAM1	-0.76/0.59	0.00175	5.31	HP vs AP
SLC25A34	-1.35/0.39	0.00177	5.86	NC vs AP
AHCYL2	-2.19/0.22	0.00192	5.17	HP vs AP
PCDH24	-2.07/0.24	0.00254	4.81	HP vs AP
GCNT3	-1.15/0.45	0.00254	4.84	HP vs AP
AQP8	-2.44/0.18	0.00266	4.97	NC vs AP

The top 20 down-regulated genes were compiled the same methods for Table 4.3. The genes are down-regulated in the second group relative to the first group e.g. *KRT20* is down-regulated in APs relative to HPs and *SLC26A3* is down-regulated in APs relative NCs.

4.6.2. IN-SILICO ANALYSIS AS VALIDATION OF MICROARRAY DATA

Results obtained from the DASL microarray were compared with the literature and previously published and publicly available gene expression data from Gene Expression Omnibus (GEO).

Gene expression changes of the 15 genes identified in the NC vs AP comparisons from Table 4.3 and Table 4.4 were reviewed in a GSE4183 (Gyorffy et al., 2009, Galamb et al., 2010), a gene expression dataset. Genes identified in the other comparisons (NC vs HP and HP vs AP) were not reviewed in the GEO dataset as HPs were not included in its study design. GSE4183 looks at the differences between colorectal adenomas and normal colonic mucosa from healthy controls (equivalent to NC vs AP) and is therefore comparable to the design of this DASL microarray study.

14 of the 15 genes identified in the GSE4183 dataset showed agreement with the gene expression profile seen in this thesis, with the remaining sample (*LY6G6D*) showing no significant change in expression. 12/15 genes were significant ($p<0.05$) according to the adjusted p-values with two additional genes being significantly different using the raw p-values ($p<0.05$). In total 14/15 genes showed similar gene expression profiles between the two microarrays. The gene expression changes for GSE4183 can be seen in Table 4.5.

Table 4.5: Gene expression changes in GSE4183

Gene	Fold change (log/adjusted)	Raw p-value	Adjusted p-value
AXIN2	1.45/2.70	0.000057	0.0064
SCD3	0.66/1.58	0.000179	0.01067
CARS	0.58/1.50	0.001233	0.025679
SIL1	0.31/1.24	0.001978	0.032788
CAPN12	0.47/1.39	0.00034	0.013552
LY6G6D	0.08/1.06	0.850696	0.944268
ETS2	0.85/1.80	0.000419	0.014926
G3BP1	0.36/1.29	0.001315	0.026628
NPDC1	0.66/1.58	0.012012	0.092281
SLC26A3	-2.32/0.20	0.000126	0.009164
GUCA2A	-2.04/0.24	0.003037	0.041595
CLDN23	-1.65/0.32	0.001514	0.028439
ABCC13	-0.18/0.88	0.01858	0.119292
SLC25A34	-0.70/0.61	0.000308	0.01291
AQP8	-3.49/0.089	0.00015	0.009902

This table presents the 15 genes selected from the top-tables comparing NC and AP samples, which were validated using microarray data from the GSE4183 data set.

A review of the literature provided supporting evidence for some of the genes identified (Table 4.6). 25 of the genes identified are of unknown biological relevance. Further validation of these genes is required to investigate their potential role in carcinogenesis.

The combined results of the literature review and GEO datasets supported results for 24 out of 39 genes present in the top-tables (Table 4.3 and Table 4.4). The changes in gene expression of the remaining 15 genes could not be confirmed using this method of analysis.

Table 4.6: Evidence from the literature to support the results of the DASL microarray

Gene of Interest	Findings	References
AXIN2	Low levels of expression in colorectal tissues	(Dong et al., 2001)
	Up-regulation in colorectal adenomas	Obrador-Hevia et al., 2010
ASCL2	Up-regulation at all stages of colorectal tumorigenesis secondary to Wnt Signalling pathways	Jubb et al., 2010
	Up-regulation in intestinal neoplasias	Jubb et al., 2006
REG4	Up-regulation in hyperplastic polyps and adenomas in the colon	Rafa et al., 2010
	Up-regulation occurs early in CRC development	Li et al., 2010 Zhang et al., 2003a
	Up-regulation in colorectal adenomas	Lu et al., 2007 Lu et al., 2006 Zhang et al., 2003b
ETS2	Expression is not seen in NC or HP tissues	Ito et al., 2002
FGFBP1	Up-regulation in colorectal adenomas and cancers	Tassi et al., 2006
	Up-regulation occurs in early lesions in the progression to CRC	Ray et al., 2003
KRT20	Altered expression in HPs relative to APs	Tatsumi et al., 2005
CHGA	Overexpression is associated with more aggressive cancer	Indinnimeo et al., 2002
MALL	Differences in gene expression in NCs and HPs (serrated polyps)	Kim et al., 2008
SLC26A3	Down-regulated in adenomas compared to normal tissue. SLC26A3 may have a tumour suppressor role	Mlakar et al., 2009
GUCA2A	Up-regulated in normal colorectal tissues	Chen et al., 2009
ENTPD5	Down-regulated continuously along the adenoma-carcinoma sequence	Mikula et al., 2010
FABP1	Down-regulated in colorectal adenomas	Lee et al., 2006
AHCYL2	Down-regulated in CRC	ME et al., 2006
AQP8	Down-regulated in colorectal adenomas and adenocarcinomas	Fischer et al., 2001

This table presents the supporting evidence for the results of the top-tables (Table 4.3 and Table 4.4) obtained from a literature review.

4.7. GENE EXPRESSION PROFILES OF "NORMAL" COLORECTAL TISSUES

SAM two-way (NC vs AN, NC vs KN and AN vs KN) and multiclass (all three groups) statistics were performed in MEV (Figure 4.13). No significant DEGs were identified in this analysis.

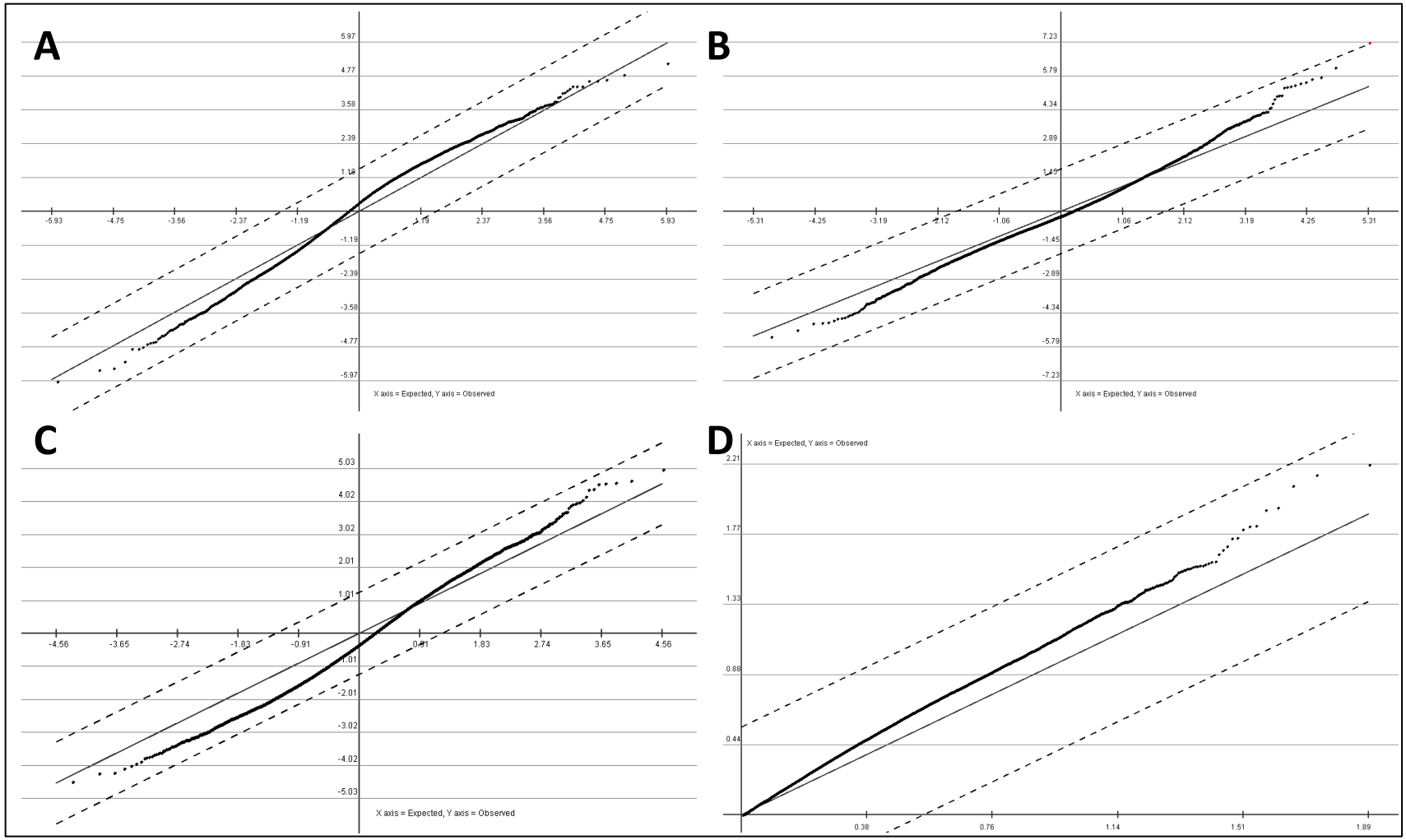

Figure 4.13: SAM graphs produced during the analysis of normal colorectal tissues (NC, AN and KN)

This figure represents the SAM graphs produced when analyzing NC vs AN (A), NC vs KN (B), AN vs KN (C) and three-way analysis (NC vs AN vs KN)(D). No significant genes were identified in any of the analyses. The FDR for each analysis was set at zero.

4.7.1. PATHWAYS SHOWING DIFFERENTIAL GENE EXPRESSION BETWEEN NORMAL COLORECTAL TISSUES BY MICROARRAY

The top 300 genes from each comparison (Table 4.2) were uploaded to DAVID for pathway analysis.

Seven genes showing differential expression between NC and AN samples (*ATP2B2, CACNA1G, CAMK2B, ITPR2, MYLK3, PPP3CB* and *TACR3*) were found to belong to the calcium signalling pathway (p=7.6E-2). All of these genes were down-regulated in ANs relative to NCs with the exception of *ITPR2* and *PPP3CB*.

When comparing NC and KN samples, genes showing differential expression were associated with six pathways (Table 4.7). *CACNB1, CACNB4, ITGA5, ITGA8, LMNA* and *TPM2* were common to several of these pathways. These genes were largely up-regulated in KNs relative to NCs, with the exception of *CACNB1* and *CACNB4*.

Table 4.7: Pathway analysis in the comparison of NC and KN tissues

Pathway	Number of genes	p-value	Genes identified
Hypertrophic Cardiomyopathy	6	2.1E-2	*CACNB1, CACNB4, ITGA5, ITGA8, LMNA, TPM2*
Dilated cardiomyopathy	6	2.8E-2	*CACNB1, CACNB4, ITGA5, ITGA8, LMNA, TPM2*
Nucleotide excision repair	4	4.8E-2	*ERCC2, ERCC8, RFC3, XPC*
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	5	5.3E-2	*CACNB1, CACNB4, ITGA5, ITGA8, LMNA*
ECM-receptor interaction	5	7.1E-2	*CD36, ITGA5, ITGA8, SDC3, TNXA*
Haematopoietic cell lineage	5	7.6E-2	*CD36, CD38, ITGA5, IL7R, HLA-DRB4*

Six pathways were identified by DAVID using the top 300 genes from the NC vs KN comparison. These pathways are largely cardiology related; however several genes identified in these pathways have previously been linked with various cancers.

Only one pathway was identified during the comparison of AN and KN samples; the PPAR Signalling pathway (p=9.7E-2). Four genes were differentially expressed in this pathway (*HMGCS2, ACSL4, PPARG* and *RXRA*), all showing down-regulation in KNs relative to ANs.

4.7.2. DIFFERENTIAL EXPRESSION IN NORMAL COLORECTAL TISSUES ACCORDING TO ANATOMICAL LOCATION

AN and KN samples were separated based on their anatomical location into left and right sided tissues. This distinction was made both in limma, which identified 25 DEGs (see Table 4.2), and MEV.

Analysis within MEV identified four genes that were all up-regulated in right sided ANs using SAM: *OSBPL5*, *CLEC4F*, *LRRC17*, and *PSAT1*. However pathway analysis performed in DAVID did not identify any pathway associations with the genes identified by limma or MEV.

Further analysis was performed in DAVID using the top 300 significant genes identified in the limma top-tables. Two pathways were identified when using this group of genes (Table 4.8). All genes in these pathways show down-regulation in right-sided ANs with the exception of *IL11* and *MAD2L2*.

Table 4.8: Pathways showing differential expression between left and right ANs

Pathway	Number of genes	p-value	Genes identified
Jak-STAT Signalling pathway	6	6.1E-2	*CISH, IL11, IL12RB1, PIK3CD, STAT5A, SOCS5*
Cell Cycle	5	9.3E-2	*CDC14A*, *MAD2L2*, *ANAPC11, CCNE2, GADD45G*

This table shows the results of the pathway analysis performed in DAVID on the top 300 differentially expressed genes between left and right ANs. The p-value represents a modified Fisher exact p-value. This test determines whether the number of genes identified in the gene list, which belong to a particular pathway, are identified by random chance or whether they are significant associations.

Similar analysis was performed to compare left and right KN samples. The results of the top-tables did not reveal any significant DEGs according to the adjusted p-value ($p<0.05$). SAM analysis performed in MEV identified two DEGs; *ENTPD2, HTR4*. These genes showed marginal up-regulation in right sided KNs. Further analysis was performed on the top 300 genes showing differential expression between left and

right sided KNs according to raw p-values. No pathways were found to be associated with the 300 genes investigated.

When combining all left and all right sided normal samples (all AN and KN samples), no significant genes were identified by any approach.

4.8. DIFFERENCES IN GENE EXPRESSION BETWEEN HYPERPLASTIC AND CANCER-ASSOCIATED POLYPS (HP VS KP)

The differential gene expression between HPs and KPs was investigated using the same approach (limma top-tables and MEV). Two DEGs were identified in the top-tables according to the adjusted p-value ($p<0.05$): *ANXA2* and *HLA-G.* Both of these genes were down-regulated in KPs relative to HPs.

Uploading the data into MEV identified 24 additional DEGs using SAM analysis (Figure 4.14). *FZD9* was the only gene showing up-regulation in KPs relative to HPs, the remaining genes were down-regulated in KPs relative to HPs. Similarly no pathways were identified by DAVID analysis.

The top 300 genes were also up-loaded to DAVID. This analysis identified three genes that are associated with Steroid Biosynthesis (p=3.1E-2); *CYP51A1, SC4MOL* and *SC5DL*. All three of these genes show down-regulation in KPs relative to HPs.

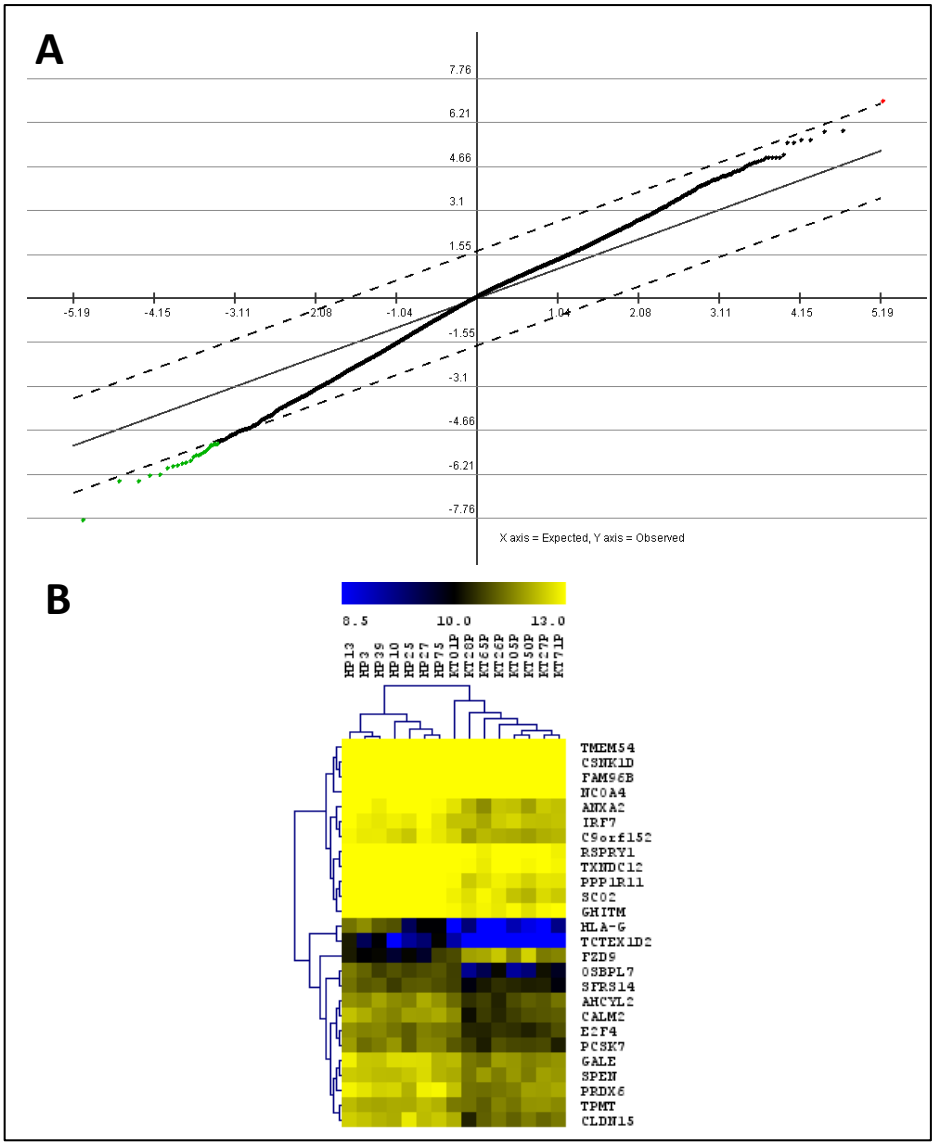

Figure 4.14: SAM graph and HCL showing the DEGs between HP and KP samples
The SAM graph showing the 20 DEGs identified in the SAM analysis. One gene is up-regulated in KPs (*FZD9*) and is shown as a red point; the remaining genes are down-regulated in KPs and are represented as green points. An FDR of zero was utilized in the analysis. B) HCL produced during the SAM analysis. *HLA-G* and *ANXA2* are identified within the analysis, both showing reduced gene expression in KPs as seen in the limma top-tables. Blue represents low expression, Yellow represents high expression.

4.8.1. HYPERPLASTIC (HP) AND CANCER-ASSOCIATED POLYPS (KP) SHOW DIFFERENTIAL EXPRESSION BASED ON ANATOMICAL LOCATION

The difference in gene expression between HPs and KPs was investigated based on anatomical location. KP tissues were separated into left and right and compared with all HP samples.

4 DEGs were identified in limma according to the adjusted p-value when comparing HPs with right-sided KPs (*GCG*, *CHGA*, *CCL27* and *HLA-G*). With the exception of *CCL27*, all of these genes were up-regulated in right-sided KPs relative to HPs.

The data were analysed in MEV using SAM. 12 genes were identified during the analysis. Two of these genes were also identified during the limma top-table analysis; *GCG* and *CHGA* (Figure 4.15). Five pathways were identified in DAVID (Table 4.9).

Table 4.9: Pathways showing differential gene expression between HPs and right-sided KPs

Pathway	Number of genes	p-value	Genes identified
Hedgehog Signalling Pathway	5	1.3E-2	*ZIC2, BMP8A, CSNK1D, PTCH2, PRKACB,*
SNARE interactions in vesicular transport	4	2.4E-2	*GOSR1, BET1, STX12, STX2,*
Steroid Biosynthesis	3	3.2E-2	*CYP51A1, SC4MOL, SC5DL*
Phenylalanine metabolism	3	5.0E-2	*ALDH3B2, AOC2, PRDX6*
Calcium Signalling pathway	7	7.0E-2	*ATP2A2, DRD5, CACNA1C, NOS3, PRKACB, SLC25A5, SLC8A1*

This table shows the results of the pathway analysis performed in DAVID on the top 300 differentially expressed genes between HPs and right-sided KPs. The p-value represents a modified Fisher exact p-value as described in the legend of Table 4.8.

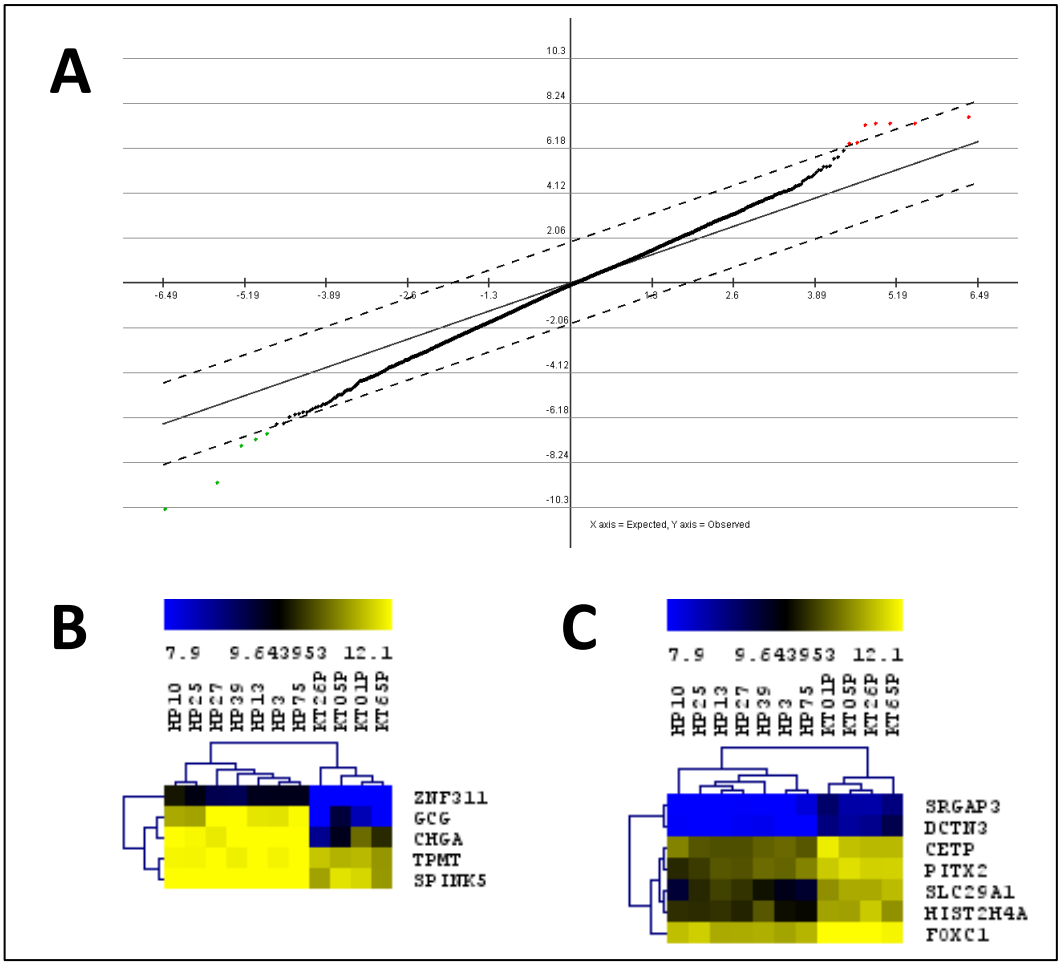

Figure 4.15: SAM analysis comparing HPs with right-sided KPs

This figure presents the genes which were identified to be differentially expressed between HPs and right-sided KPs during SAM analysis. The FDR was zero. A) SAM graph showing the up- and down-regulated DEGs between HPs and right-sided KPs. B) HCL produced during the SAM analysis showing the genes that are up-regulated in HPs relative to right-sided KPs. C) HCL showing genes that are down-regulated in HPs relative to right sided KPs as identified during the SAM analysis.

Concerning HPs with left-sided KPs, *GUSBL2* was identified, which was up-regulated in left-sided KPs relative to HPs (Limma). SAM analysis identified 19 DEGs between HPs and left-sided KPs but *GUSBL2* was not identified (Figure 4.16), and no pathways were identified in DAVID for the 19 genes identified during the SAM analysis.

Up-loading the top 300 genes into DAVID identified four genes that are associated with gluconeogenesis/glycolysis (p=8.3E-2) (*ALDOA*, *ALDOC*, *ENO1* and *PKM2*). With the exception of *ALDOA,* all genes were up-regulated in left-sided KPs relative to HPs.

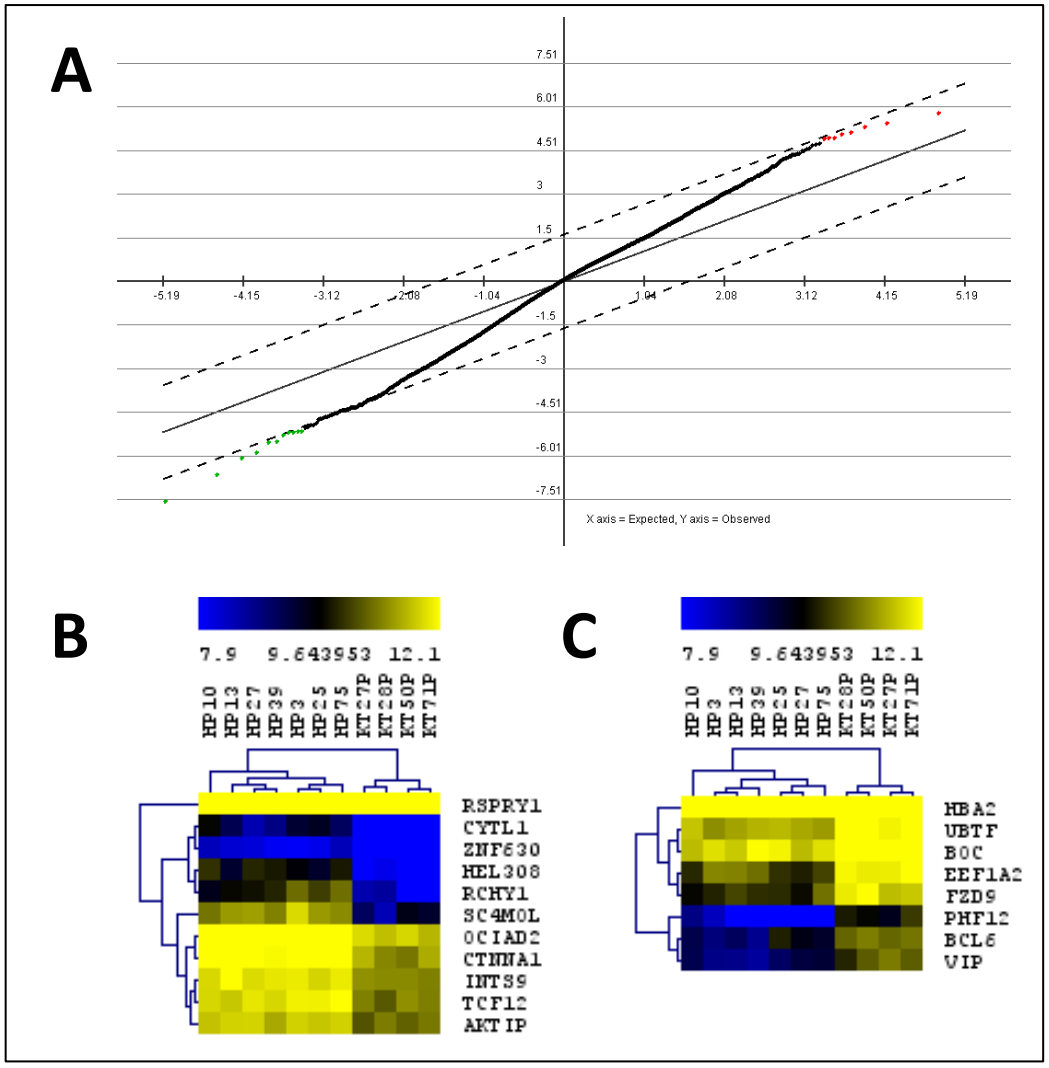

Figure 4.16: SAM analysis comparing HPs with left-sided KPs

This figure presents the results of the SAM analysis. The FDR was zero. A) SAM graph showing the up- and down-regulated differentially expressed genes between HPs and left KPs. B) HCL produced during the SAM analysis showing the genes that are up-regulated in HPs relative to KPs. C) HCL produced during the SAM analysis showing the genes that are down-regulated in HPs relative to KPs.

4.9. DIFFERENTIAL GENE EXPRESSION BETWEEN NON-CRC AND CRC ASSOCIATED POLYPS (AP VS KP)

The results of the top-tables did not identify any DEGs between AP and KP samples. The dataset was analysed in MEV using a two class unpaired SAM test, however no significant DEGs were identified. However, 2755 genes were identified as showing significant differential expression using the raw p-values in the top-tables and the top 300 genes were selected for pathway analysis in DAVID. 5 genes were identified, showing involvement with the PPAR Signalling pathway (p=4.3E-2) (*ACSL3, EHHADH,*

FABP1, PPARD and *RXRG*). The genes were up-regulated in KPs relative to APs with the exception of *PPARD*.

4.9.1. DIFFERENTIAL EXPRESSION IN NON-CRC (AP) AND CRC ASSOCIATED (KP) POLYPS ACCORDING TO ANATOMICAL LOCATION

AP and KP tissues were separated based on their anatomical location. In addition, all left and right sided polyps were pooled together, regardless of CRC association.

The top-table comparing right sided APs and KPs did not identify any significant genes according to the adjusted p-values but there were 1413 genes according to the raw p-value. The top 300 genes were up-loaded into DAVID but no pathway involvement and no DEGs were identified by SAM analysis.

Top-table analysis comparing left sided APs with left-sided KPs identified *EDARADD,* according to the adjusted p-value (p=0.0066), as a DEG that is up-regulated in left-sided KPs relative to left-sided APs. 2287 genes were identified as showing differential expression when using the raw p-values. Analyzing the top 300 genes in DAVID identified three pathways (Table 4.10), including genes associated with the MAPK Signalling pathway. The dataset was also analysed using SAM in MEV, however no significant DEGs were identified.

Table 4.10: Pathways showing differential expression between left sided AP and KP tissues.

Pathway	Number of genes	p-value	Genes Identified
PPAR Signalling pathway	6	7.5E-3	*CD36, CPT1A, FABP1, PPARA, PPARD, RXRG*
MAPK Signalling pathway	11	2.0E-2	*CD14, CACNA1E, CACNA1H, CACNA2D4, DUSP7, FGFR1, GADD45A, GNG12, MAP2K3, MAP2K4, RAC1*
Haematopoietic cell lineage	5	6.7E-2	*CD14, CD36, CD9, IL4, MS4A1*

Three pathways showed differential expression between right sided AP and KP tissues using the top 300 genes identified in the top-table. Of interest, 11 genes were identified in the MAPK Signalling pathway which is known to be associated with colorectal carcinogenesis. The statistics performed in DAVID are explained in the legend of Table 4.8.

To further increase the numbers present in the left and right sided sample groups, all AP and KP samples were then pooled according to anatomical location. No DEGs were identified in the top-table according to the adjusted p-value, but there were 1076 significant genes according to the raw p-values. For the top 300 genes no pathways were identified in DAVID and SAM analysis did not identify any significant DEGs. This finding emphasizes the importance of the adjusted p-value, since the raw p-value can identify false positives.

4.10. DIFFERENTIAL GENE EXPRESSION BETWEEN PAIRED NORMAL AND POLYP SAMPLES

Matched normal and polyp samples from non-CRC (AN-AP) and CRC (KN-KP) associated patients were investigated for differential gene expression.

In matched AN-AP samples, 15 genes were identified in the top-table according to the adjusted p-value, but there were no pathways identified by DAVID. Using SAM analysis 2 genes (*CA1* and *FRMD6*) were identified as being down-regulated in APs relative to the paired AN samples.

The limma top-tables did not identify any DEGs when comparing matched KN and KP tissues using the adjusted p-values. In addition, no DEGs were identified using SAM in MEV. The top 300 genes from the top-tables were up-loaded to DAVID for pathway analysis. Two pathways were identified (Table 4.11).

.

Table 4.11: Pathways showing differential expression between matched KN and KP tissues

Pathway	Number of genes	p-value	Genes Identified
Cytokine-cytokine receptor interaction	9	4.9E-2	*CCL25, CCL27, CCL28, CCL4L2, IL1R2, IL20RA, IL28A, MET, PDGFB*
Purine metabolism	6	8.8E-2	*ADA, ADK, ADCY5, NME1-NME2, PDE4B, PKLR*

The two pathways outlined in the table show altered gene expression among matched KN-KP tissues. The statistics are the same as those described in Table 4.8.

Matched samples were pooled together to include all non-CRC (AN and AP) and CRC associated (KN and KP) matched tumour and normal samples. Six DEGs were identified using a SAM two-class paired analysis. 3 genes were up-regulated in normal tissues (*GCG, PAMR1 (DKFZP586H2123),* and *C5orf29*) and 3 were down-regulated in normal tissues (*AXIN2, NPDC1* and *IGFBP2*).No pathways were identified by this analysis.

4.10.1. DIFFERENTIAL GENE EXPRESSION IN MATCHED TISSUES BASED ON ANATOMICAL LOCATION

Matched AN-AP and KN-KP samples were analysed based on their anatomical location. This analysis did not identify any significant DEGs for either comparison.

All left-sided matched samples (AN-AP and KN-KP) were pooled together to see the effect of increased numbers on DEG identification. However, no significant genes were identified using SAM analysis in MEV (Figure 4.17). When doing the same for right sided samples (matched AN-AP and KN-KP samples) one significant gene (*TRIM29*) was identified, showing increased expression in right-sided polyps compared to the matched normal tissues.

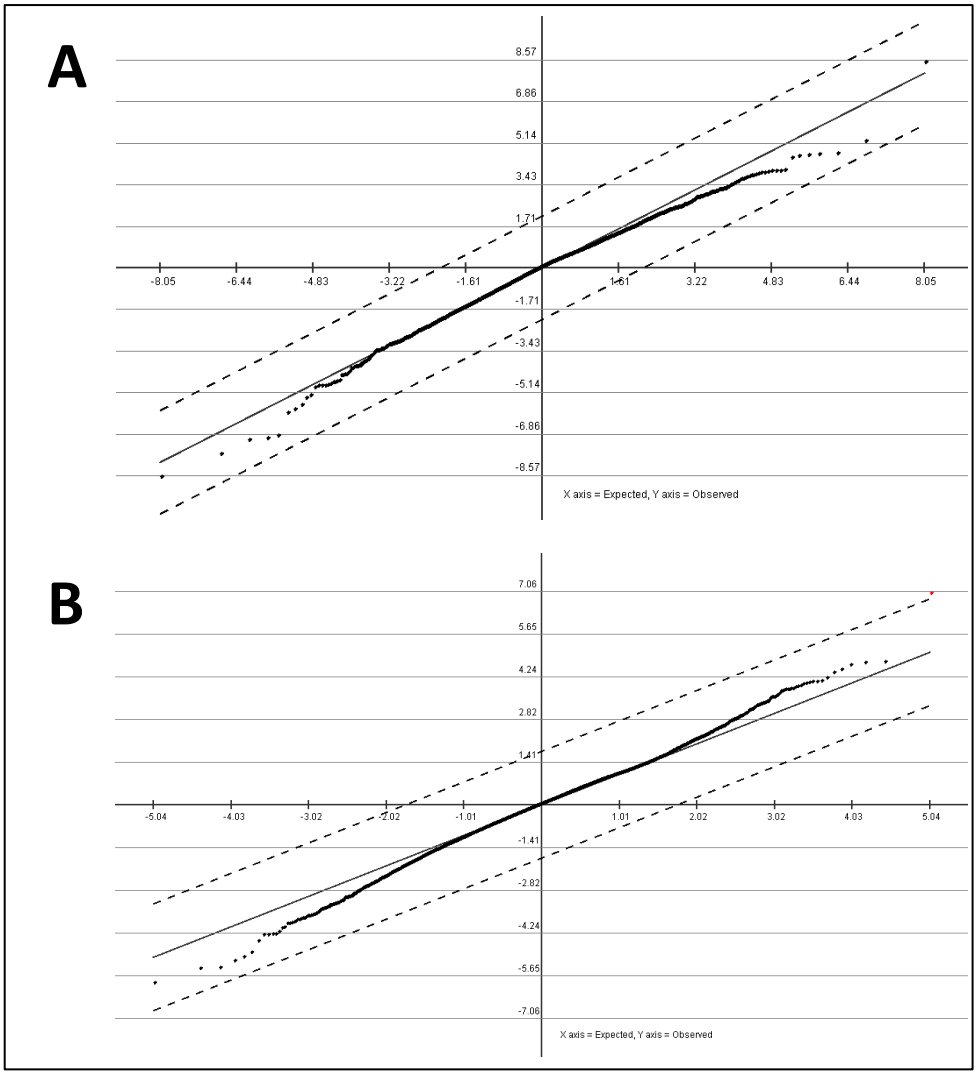

Figure 4.17: SAM graphs produced during the analysis of pooled left and right matched samples

SAM two-classed paired analysis was performed on left and right sided paired samples. An FDR of zero was achieved for all SAM analysis. A) SAM graph of left-sided matched normal and polyp tissues. No significant DEGs were identified. B) SAM graph of right-sided matched normal and polyp tissues identifying one significant DEG which is up-regulated in polyps relative to normal tissues; *TRIM29.*

4.11. DIFFERENTIAL GENE EXPRESSION BETWEEN NORMAL CONTROLS, HYPERPLASTIC AND ADENOMATOUS POLYPS

Differential gene expression across NC, HP and AP samples was investigated using SAM statistics in MEV. 102 genes were identified as showing differential gene expression across the three groups. These genes were up-loaded into DAVID for pathway analysis, identifying one significant pathway; Nitrogen Metabolism (p=1.1E-2). Three genes were identified in this pathway; *CA1*, *CA2* and *CA4*. All three genes

showed down-regulation in APs relative to NCs and HPs. The HCL and SAM graph produced during the analysis, along with a PCA plot, are presented in Figure 4.18. The HCL identifies patterns of differential gene expression across the three groups.

PCA was also performed on the entire gene data set and on the genes identified during the SAM analysis. Using the entire dataset, the three sample groups could not be distinguished from one another. However, when performing PCA on the 102 genes identified during the SAM analysis, the three groups were clearly separated from one another (Figure 4.18).

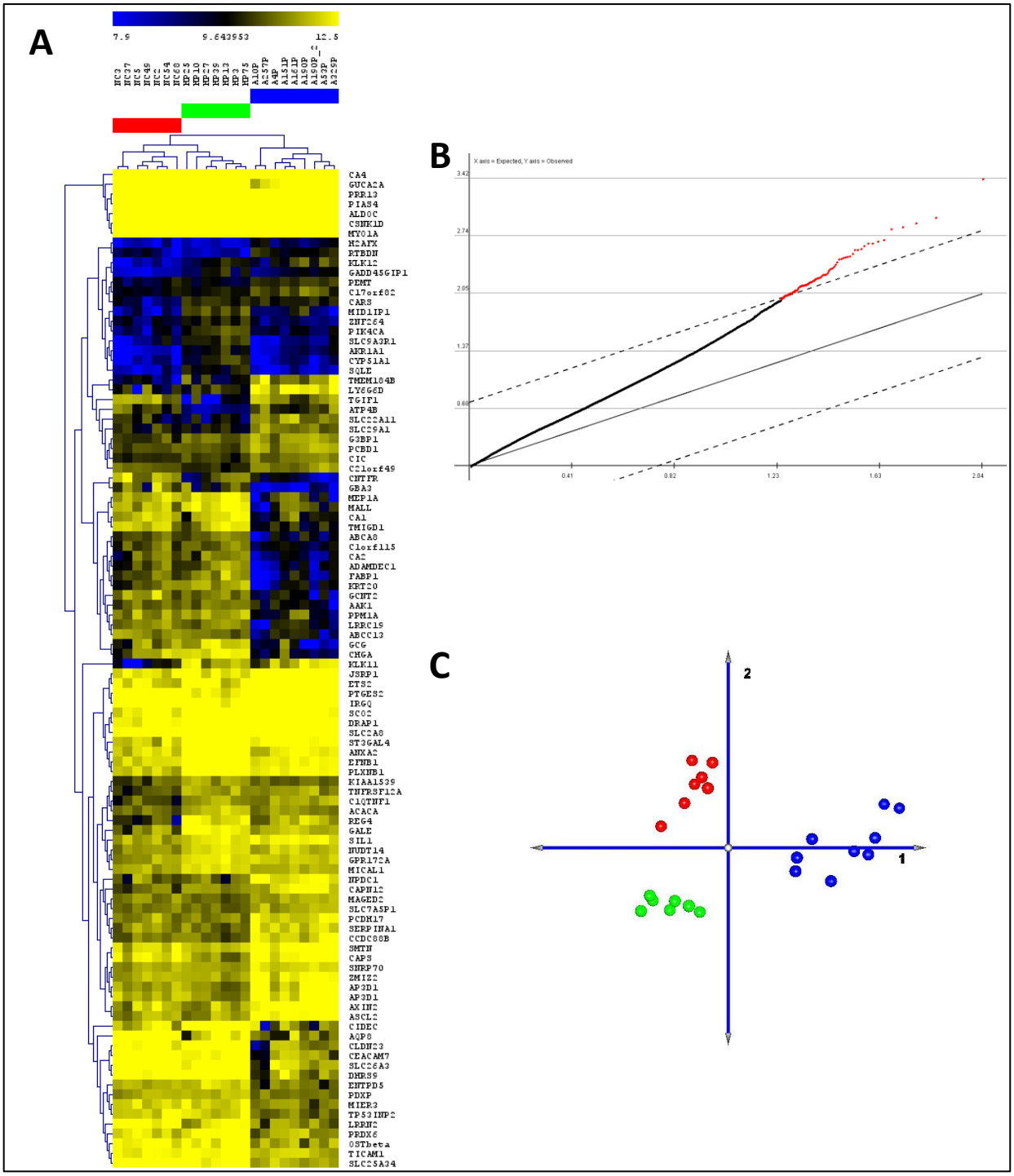

Figure 4.18: Results of SAM and PCA analysis of NC, HP and AP tissues

102 genes were identified during the SAM analysis. A) HCL created to show the gene expression profile of the 102 genes identified during the multiclass SAM analysis. B) SAM multiclass graph showing the DEGs in red. An FDR of zero was maintained during this analysis. C) PCA analysis showing how the 102 genes identified during the SAM analysis help to separate the three sample groups into distinct histological groups based on gene expression. Red, NCs; Green, HPs; Blue, APs.

4.11.1. PATHWAYS SHOWING DIFFERENTIAL GENE EXPRESSION BETWEEN HEALTHY CONTROLS AND HYPERPLASTIC TISSUES (NC VS HP)

The top-table analysis identified 504 DEGs between NCs and HPs according to the adjusted p-values. This comparison represents the second highest list of DEGs in the top-tables. Using the 504 DEGs identified in the top-table between NCs and HPs, five pathways were identified during analysis in DAVID (Table 4.12). Of interest, six genes were identified showing differential gene expression in the p53 Signalling pathway, up-regulated in HPs relative to NCs, with the exception of *STEAP3*.

Table 4.12: Pathways showing significant differential expression between NCs and HPs

Pathway	Number of genes	p-value	Genes Identified
Proteasome	8	6.8E-4	*PSMC5, PSMD4, PSMA1, PSMA3, PSMA4, PSMB3, PSMB8, PSMC4,*
Aminoacyl-tRNA biosynthesis	6	9.6E-3	*CARS, SARS2, SARS, TARS, WARS, VARS*
Axon guidance	10	2.3E-2	*EPHA2, EFNB1, MAPK3, MAPK1, NFAT5, PLXNB1, PPP3CB, SEMA3B, SEMA7A, SLIT2*
Pyruvate metabolism	5	3.9E-2	*ACAT2, ACACA, GRHPR, ME1, PKM2*
p53 Signalling pathway	6	6.7E-2	*STEAP3, CCNB1, IGFBP3, SERPINB5, SHISA5*

The top 504 differential expressed genes between NCs and HPs identified changes in the five pathways presented in the table. The statistics presented in the table are derived from DAVID and were explained previously in Table 4.11.

NC and HP samples were compared in MEV using SAM. 111 genes were identified as showing significant differential expression. The SAM graph, HCL and PCA analysis are presented in Figure 4.19. The PCA analysis shows how the NC and HP samples cluster together before and after the SAM analysis. By using the list of DEGs identified during the SAM analysis, the NC and HP samples cluster into two distinct groups as shown in Figure 4.19.

The genes identified during the SAM analysis were uploaded to DAVID for pathway analysis, which identified five pathways (Table 4.13).

Table 4.13: Pathways showing differential gene expression between NC and HP samples according to genes identified during SAM analysis

Pathway	Number of genes	p-value	Genes Identified
Axon guidance	6	1.9E-3	*EPHA2, ABLIM1, EFNB1, NGEF, PLXNB1, SEMA3B*
Steroid Biosynthesis	3	6.2E-3	*CYP51A1,SQLE, SC4MOL*
p53 Signalling pathway	4	1.2E-2	*STEAP3, CCND1, IGFBP3*
Aminoacyl-tRNA biosynthesis	3	3.4E-2	*CARS, SARS2, VARS*
PPAR Signalling pathway	3	8.5E-2	*ACOX2, AQP7, CHKB,*

These pathways were identified following analysis of the DEG list identified during SAM analysis of NC and HP samples. Three of the pathways were previously described in Table 4.12 (p53 signalling pathway, Axon guidance and Aminoacyl-tRNA biosynthesis). The other two pathways were not identified in the analysis of the top table genes.

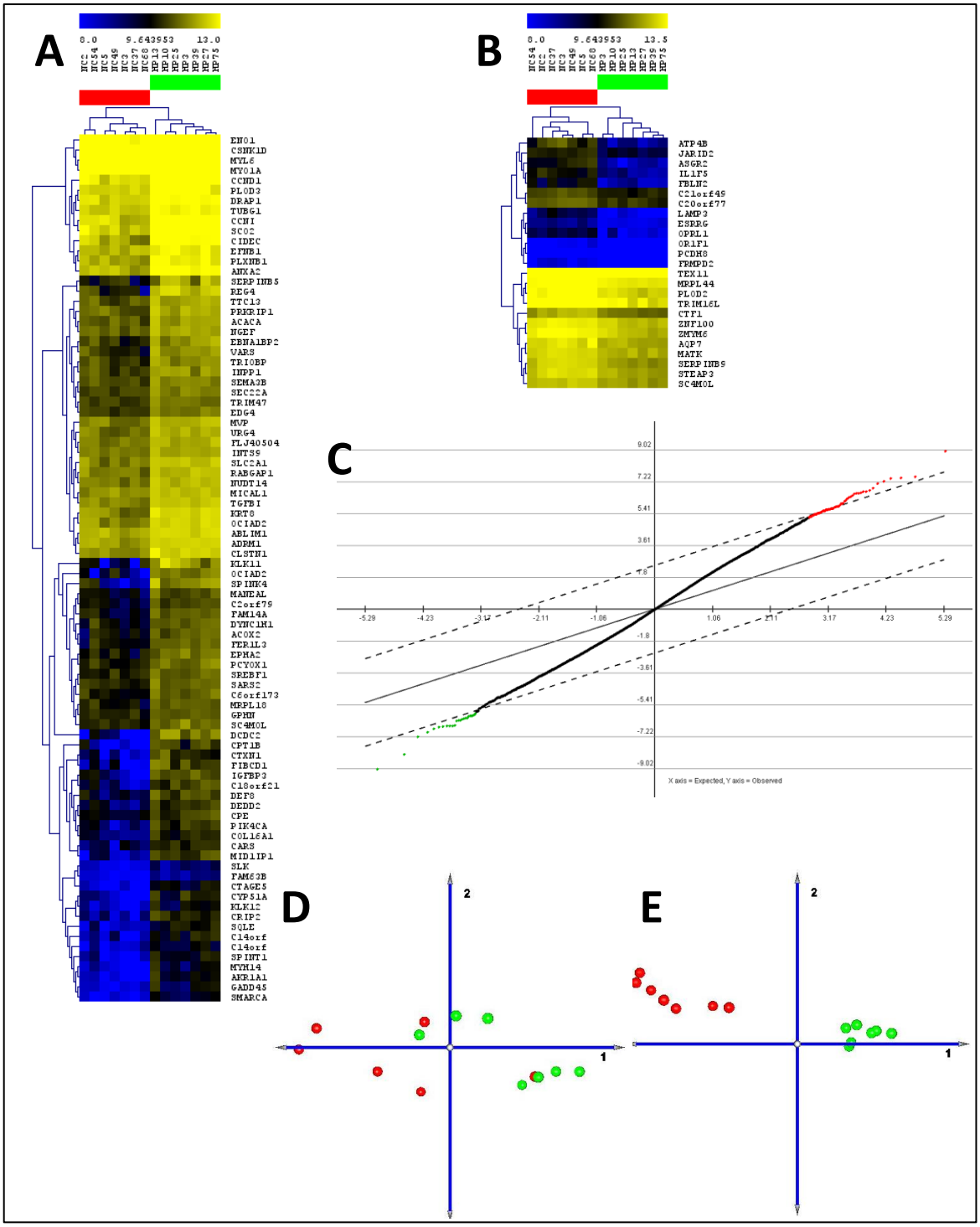

Figure 4.19: SAM and PCA results when comparing NCs with HPs

These figures were created during the analysis of NC and HP samples in MEV. A) HCL showing DEGs that are up-regulated in HPs relative to NC. B) HCL showing DEGs that are down-regulated in HPs relative to NCs. C) SAM graph representing the 111 DEGs identified during the analysis. An FDR of zero was used in this analysis. D) PCA using the entire gene set shows that the NC and HP samples do not cluster into tight little groups. Following the SAM analysis, a PCA was performed on the 111 genes identified during the SAM analysis. Using the 111 genes from the SAM analysis enables the NC and HP samples to separate into two diverse groups (E).

4.11.2. PATHWAYS SHOWING DIFFERENTIAL EXPRESSION BETWEEN NC AND AP SAMPLES

Top-table analysis identified 197 DEGs between NCs and APs according to adjusted p-values but no pathways were revealed in DAVID using this gene list. Subsequently the top 300 DEGs were uploaded to DAVID, which revealed three pathways (Table 4.14) including the Wnt Signalling Pathway. With the exception of *PPARD,* all genes associated with the Wnt Signalling Pathway were up-regulated in APs relative to NCs.

Table 4.14: Pathways showing significant differential expression between NCs with APs

Pathway	Number of genes	p-value	Genes Identified
Nitrogen Metabolism	4	5.7E-3	*CA1, CA2, CA4, CA12*
Pathways in Cancer	11	3.7E-2	*CTBP1, E2F2, AXIN2, BCR, FZD9, NCOA4, PPARD, RARA, RXRG, KRAS, VEGFB,*
Wnt Signalling Pathway	6	9.6E-2	*CTBP1, RUVBL1, AXIN2, FZD9, NFATC4, PPARD*

The top 300 differentially expressed genes between NCs and APs identified changes in the three pathways presented in the table. The statistics presented in the table are derived from DAVID and were explained previously in Table 4.8. The genes identified in the pathways show some overlap. For example, the genes identified in the "Thyroid Cancer" pathway are also found in the "Pathways in Cancer" pathway. In addition, four of the genes identified in the "Wnt Signalling pathway" are also present in the "Pathways from Cancer" pathway. This finding is understandable since all of these pathways are associated with cancer.

The data set was analysed in MEV using SAM to compare NC and AP tissues. 29 significant DEGs were identified during the analysis. PCA also showed that the NC and AP samples could be separated into distinct groups (Figure 4.20).

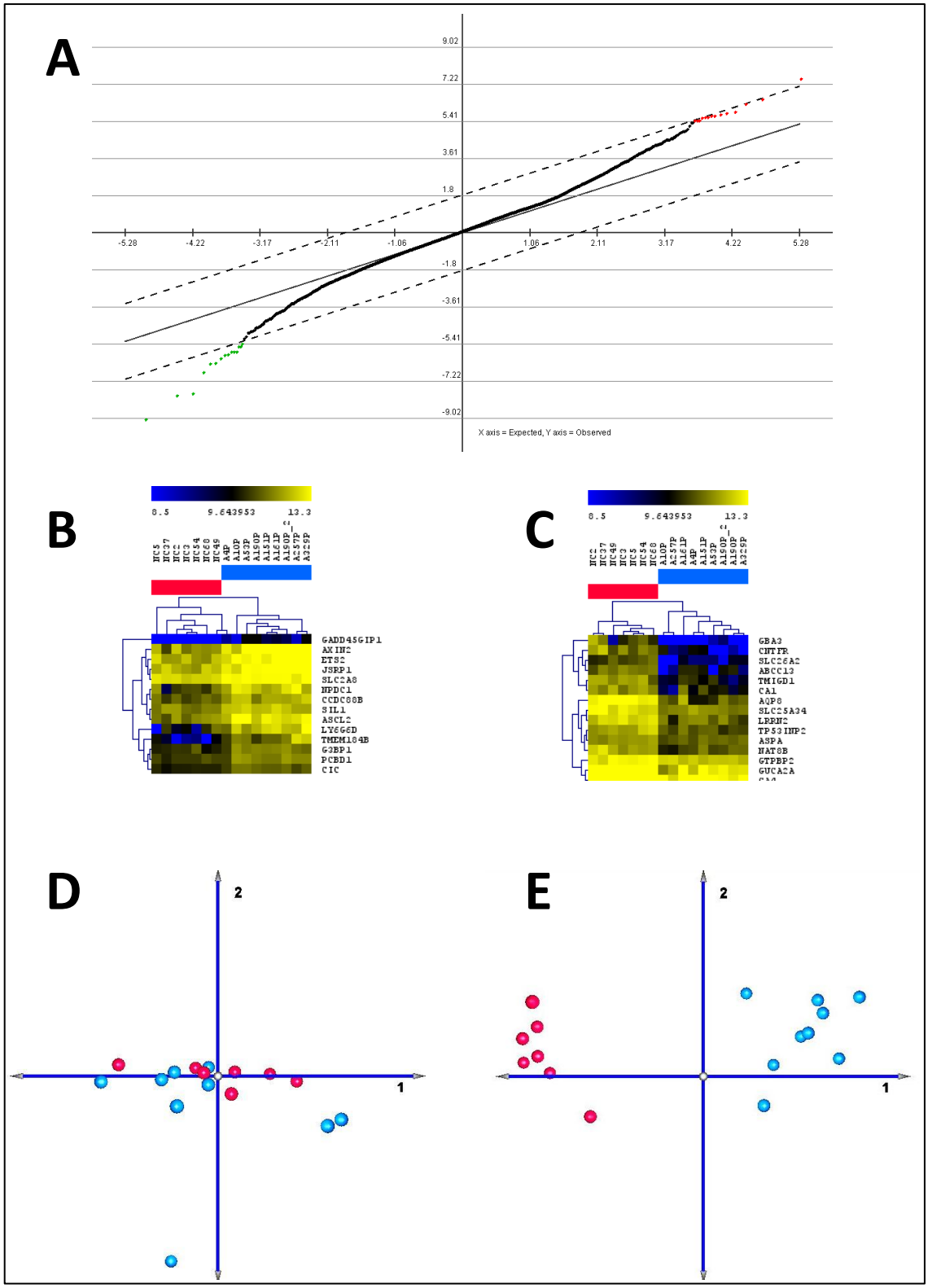

Figure 4.20: SAM and PCA analysis comparing NC and AP samples to identify DEGs
These figures were created during the analysis of NC and AP samples in MEV. A) SAM graph showing the DEGs identified during the SAM analysis. An FDR of zero was used during this analysis. B) HCL showing genes identified during the SAM analysis that are up-regulated in APs relative to NCs. C) HCL showing genes identified during the SAM analysis that are down-regulated in APs relative to NCs. D) A figure showing the results of a PCA on the entire dataset. The sample groups cannot be separated based on the entire gene dataset. E) The results of the PCA when using only the gene list identified during the SAM analysis. Using this gene list the sample groups are clearly distinguishable.

The list of DEGs identified during SAM analysis was up-loaded into DAVID, which identified one pathway; Nitrogen metabolism. *CA1* and *CA4* were both down-regulated in APs relative to NCs.

4.12. GENE EXPRESSION PROFILING OF HYPERPLASTIC AND ADENOMATOUS POLYPS (HP VS AP)

The comparison of HP and AP tissues produced the highest number of DEGs, identifying 1633 DEGs according to the adjusted p-value and providing strong evidence for significant differences in gene expression. Top-tables, MEV and COXPRESSdb have been utilized to investigate further these changes in gene expression.

Using all 1633 genes in DAVID identified 33 pathways and 234 genes (Table 4.15) hence individual genes have not been included in the table.

Of particular interest, a number of pathways associated with other cancers (e.g. non-small cell lung cancer, thyroid cancer, etc) were identified in addition to the "Pathways in Cancer" pathway. 50 genes in this pathway were shown to have differential expression in HP and AP samples. In addition, other pathways such as Cell Cycle, Apoptosis and the MAPK Signalling pathway are important in the development of cancer.

Table 4.15: Pathways showing differential gene expression between HP and AP tissues

Pathway	Number of genes	p-value
Apoptosis	14	1.0E-1
Regulation of Actin Cytoskeleton	29	9.9E-2
Non-small cell lung cancer	10	9.2E-2
Steroid hormone biosynthesis	9	9.0E-2
PPAR Signalling pathway	12	8.7E-2
Epithelial cell signalling in helicobacter pylori infection	12	8.0E-2
Cell Cycle	7	7.7E-2
Nitrogen metabolism	6	7.7E-2
Androgen and Oestrogen metabolism	8	7.6E-2
MAPK Signalling pathway	36	6.7E-2
Oxidative phosphorylation	20	6.5E-2
Fc gamma R-mediated phagocytosis	16	5.5E-2
Oocyte meiosis	18	5.1E-2
Tight junction	21	5.0E-2
Fatty acid metabolism	9	4.5E-2
Chronic myeloid leukaemia	14	3.7E-2
B cell receptor signalling pathway	14	3.7E-2
Citrate cycle (TCA cycle)	8	3.3E-2
Steroid biosynthesis	6	2.3E-2
Thyroid cancer	8	2.3E-2
Bladder cancer	10	2.3E-2
Renal cell carcinoma	14	2.2E-2
Glycolysis/gluconeogenesis	13	1.6E-2
Endometrial cancer	12	1.4E-2
Neurotrophin Signalling pathway	22	1.3E-2
Pancreatic cancer	15	1.2E-2
Valine, leucine and isoleucine degradation	11	1.1E-2
Adherens junction	16	9.6E-3
Propanoate metabolism	10	3.6E-3
Prostate cancer	19	3.1E-3
Pathways in cancer	50	3.1E-3
Long-term potentiation	16	2.8E-3
Pyruvate metabolism	12	1.6E-3

36 pathways were identified when using the 1633 DEGs identified in the limma top-tables. Many of these pathways show associations with cancer. Once again the statistics presented in this table are obtained from DAVID. The statistics used were described in the legend of Table 4.8.

The whole dataset was explored in MEV to identify changes in gene expression between HP and AP tissues. SAM analysis revealed 120 significant DEGs between the two groups (Figure 4.21). PCA showed separation of the samples into two distinct groups based on their gene expression (Figure 4.21). However, when the genes identified in the SAM analysis were uploaded to DAVID for pathway analysis no pathways were identified.

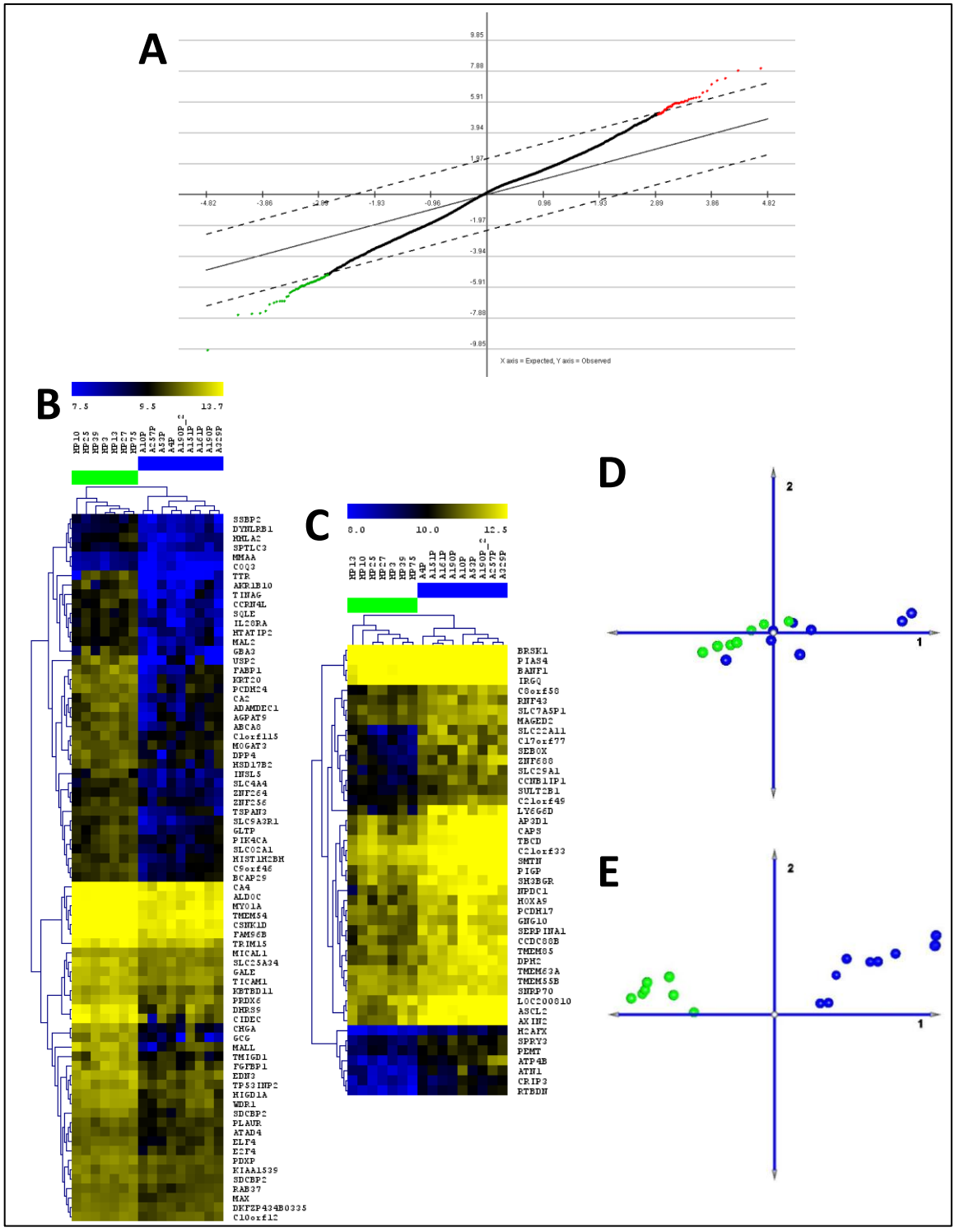

Figure 4.21: Results of SAM and PCA analyses used to compare differential gene expression between HP and AP tissues

The results of the analysis performed in MEV comparing HP and AP gene expression differences is presented in this figure. A) the SAM graph created during the SAM analysis showing the 120 significant DEGs. An FDR of zero was maintained during this analysis. B) HCL created during the SAM analysis representing the list of genes which are up-regulated in HPs relative to APs. C) HCL created during the SAM analysis representing the list of genes which are down-regulated in HPs relative to APs. D) PCA performed on the entire dataset shows how the different sample groups, although showing some overlap, seem to cluster together. Green spheres represent HPs and Blue spheres represent APs E) Following the SAM analysis, PCA was performed on the 120 significant DEGs. The two groups are clearly distinguishable from one another based on this gene list.

4.12.1. THE WNT SIGNALLING PATHWAY SHOWS ALTERED GENE EXPRESSION BETWEEN HP AND AP SAMPLES BY MICROARRAY

A total of 318 Wnt associated genes were identified from a previous study investigating the transformation of colorectal polyps into CRCs (Sabates-Bellver et al., 2007), which examined differences in gene expression between 32 colorectal adenomas and their matched normal tissues. Of the 318 Wnt associated genes noted by Sabates-Bellver *et al*, 296 (93.1%) were present on the Whole-genome DASL microarray.

Of the 296 Wnt associated genes, 121 (41%) showed significant differences in expression in HPs relative to APs: 42 of these were significantly different using the adjusted p-value ($p<0.05$) and 79 according to the raw p-value ($p<0.05$).

SAM analysis performed on the Wnt-associated genes, identified 17 DEGs (Figure 4.22). The Wnt-associated genes deemed significant in MEV were uploaded to DAVID and the Wnt Signalling pathway (p=3.6E-3) was identified as expected.

The significant Wnt-associated genes present in the dataset were used to recreate a KEGG style pathway (Figure 4.23) showing changes in gene expression between HPs and APs.

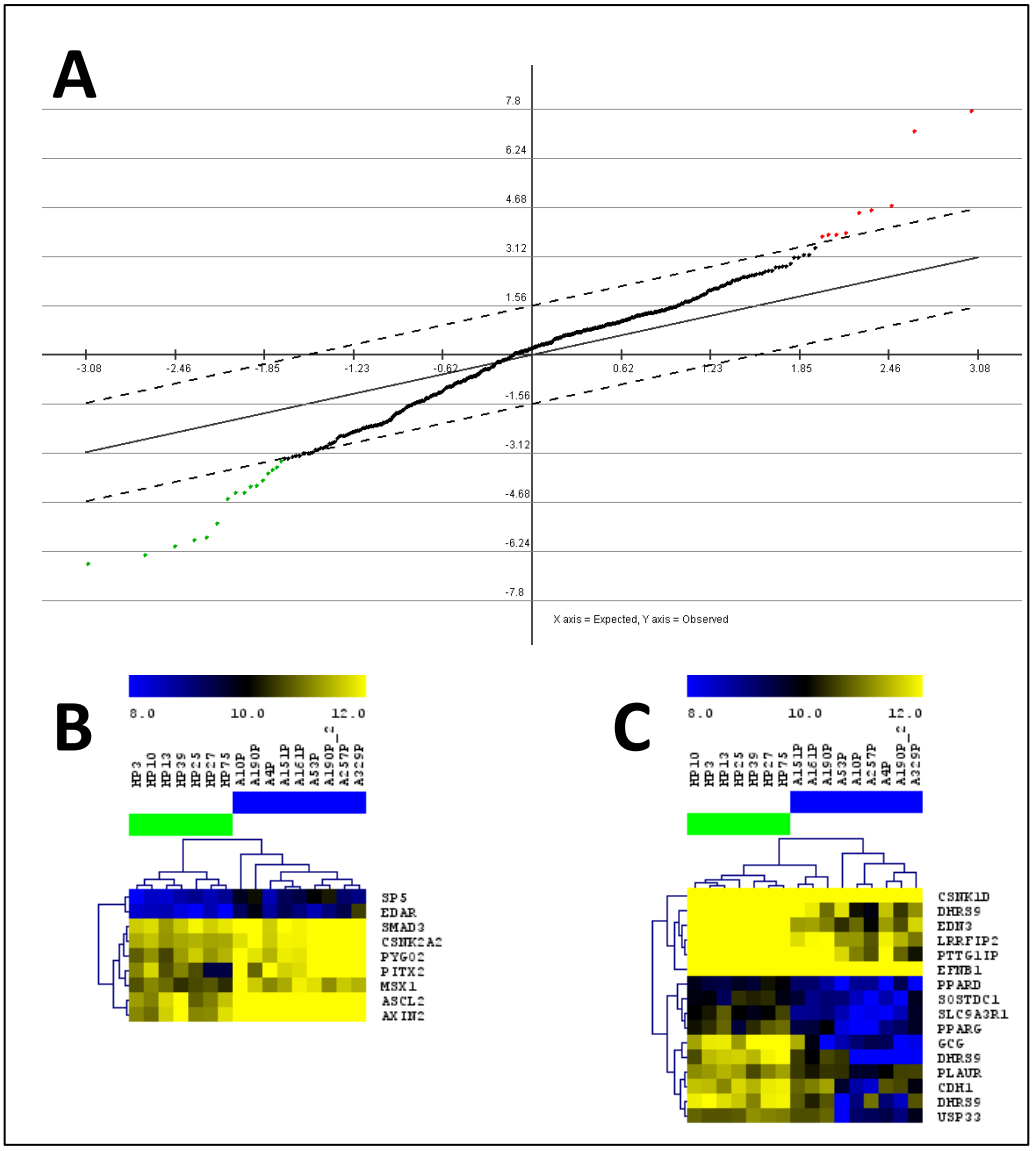

Figure 4.22: SAM graph and HCL showing Wnt-associated DEGs between HP and AP tissues
A) SAM graph showing the up- (red) and down- (green) regulated Wnt-associated genes between HPs and APs. 25 DEGs were identified An FDR of zero was maintained throughout this analysis. B) HCL created during the SAM analysis showing the DEGs which are up-regulated in APs relative to HPs. C) HCL created during the SAM analysis showing the DEGs which are down-regulated in APs relative to HPs.

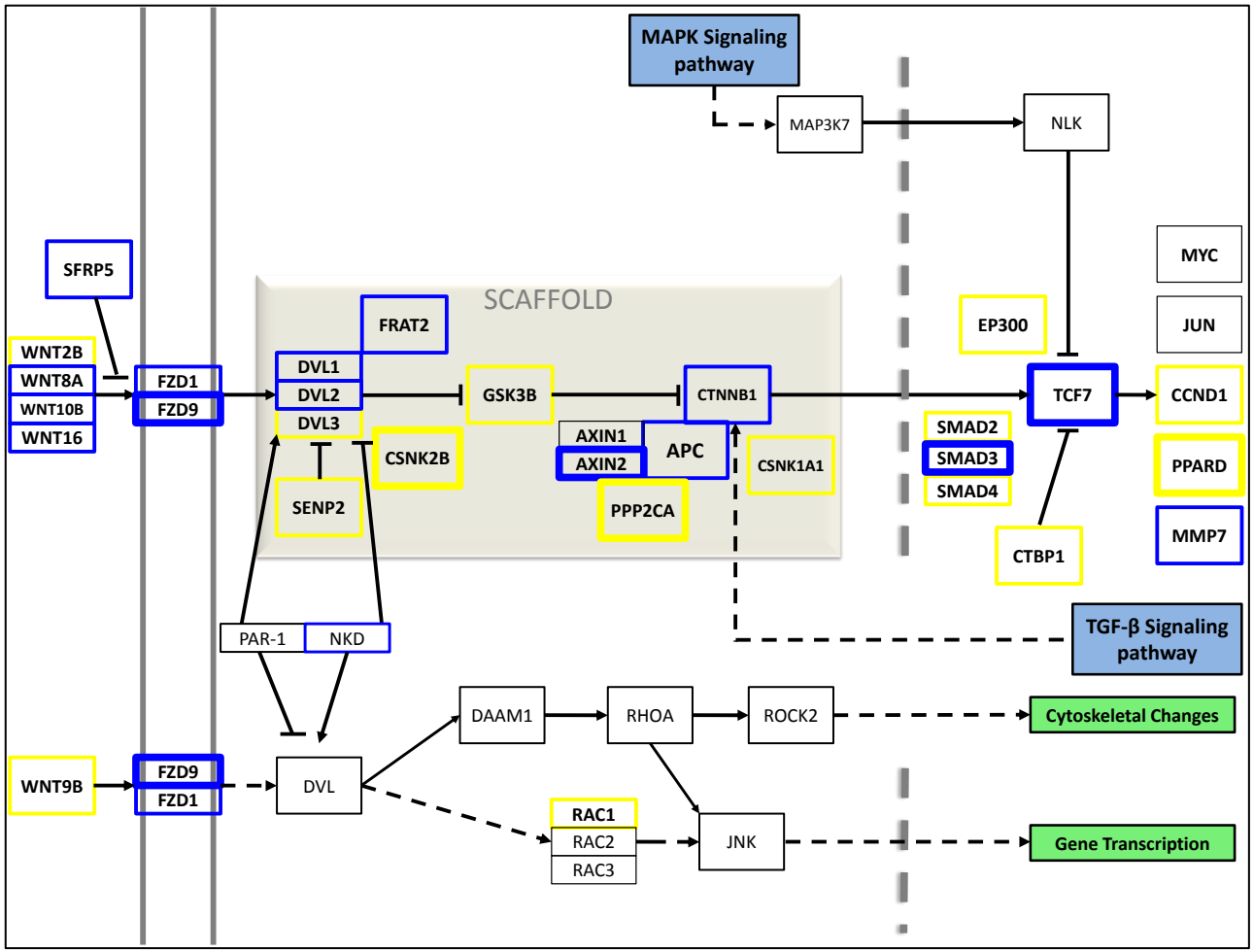

Figure 4.23: Modified Wnt-signalling pathway representing gene expression changes in HPs relative to APs

This figure represents the genes involved in the Wnt signalling pathway and has been adapted from a KEGG diagram produced during analysis on DAVID. Genes highlighted in yellow and blue represent up- and down- regulated gene expression respectively, in HPs relative to APs. For example, yellow genes are up-regulated in HPs relative to APs. Bold genes represent genes identified as having an adjusted p-value <0.05, the remaining highlighted genes represent genes with a raw p-value <0.05.

4.12.2. THE MAPK SIGNALLING PATHWAY IS LARGELY UP-REGULATED IN HYPERPLASTIC COMPARED TO ADENOMATOUS POLYPS

36 DEGs were identified from the MAPK Signalling pathway from analysis of HP and AP samples (Table 4.15). Using the KEGG pathway available via DAVID, a reference gene list of MAPK Signalling associated genes was created and 267 MAPK Signalling associated genes were identified in the top-table.

The normalised dataset was next screened for the 267 MAPK-associated genes: 258/267 (97%) MAPK-associated genes were present on the array. In addition to the 36 MAPK Signalling associated genes showing an adjusted p-value <0.05, a further 30 MAPK Signalling associated genes were identified with a raw p-value <0.05. The MAPK Signalling KEGG pathway produced during the analysis in DAVID was adapted to incorporate the changes in gene expression seen in the top-tables (Figure 4.24), showing the differential expression of MAPK Signalling associated genes in HPs and APs.

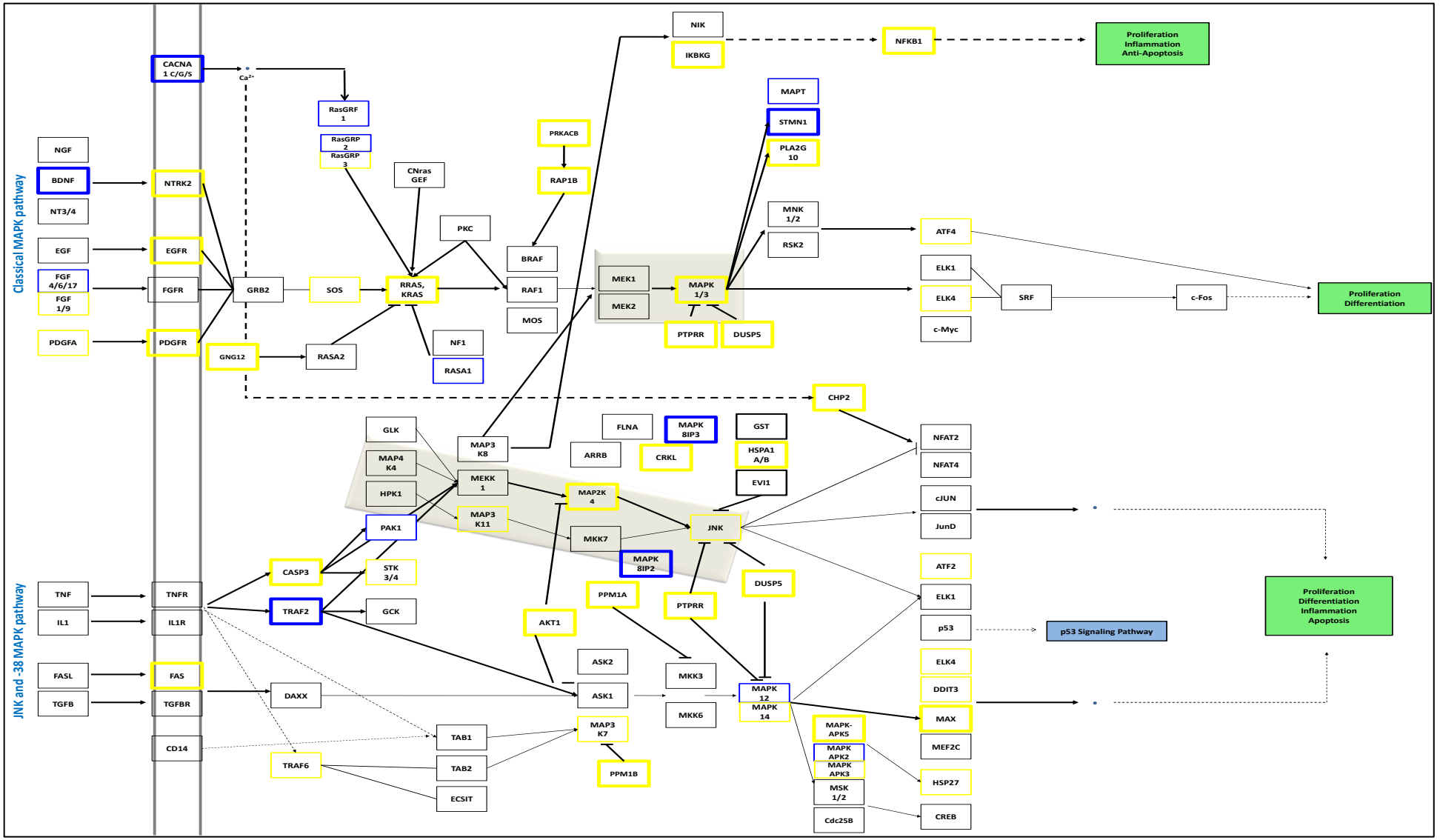

Figure 4.24: Modified MAPK signalling pathway to highlight DEGs between HPs and APs

This figure represents the genes involved in the MAPK Signalling pathway and has been adapted from a KEGG diagram produced during analysis on DAVID. The genes are annotated in the same way as those shown in the Wnt Signalling pathway in Figure 4.23.

4.12.3. GENES INVOLVED WITH P53 SIGNALLING SHOW DIFFERENTIAL EXPRESSION BETWEEN HYPERPLASTIC AND ADENOMATOUS POLYPS

Genes associated with the p53 Signalling pathway were identified as showing differential expression between NC and HP samples (Table 4.12). Analysis of genes associated with this pathway was undertaken between HP and AP tissues since disruption of p53 is seen in CRC.

69 p53 Signalling associated genes were identified using a KEGG pathway available through DAVID. 66/69 p53 Signalling associated genes were present on the DASL microarray, with 28 of these genes being differentially expressed between HPs and APs. 8 genes were significant according to the adjusted p-values, and 20 genes were significant according to the raw p-value.

The KEGG pathway available through DAVID was adapted to represent the changes in gene expression seen between HP and AP samples (Figure 4.25).

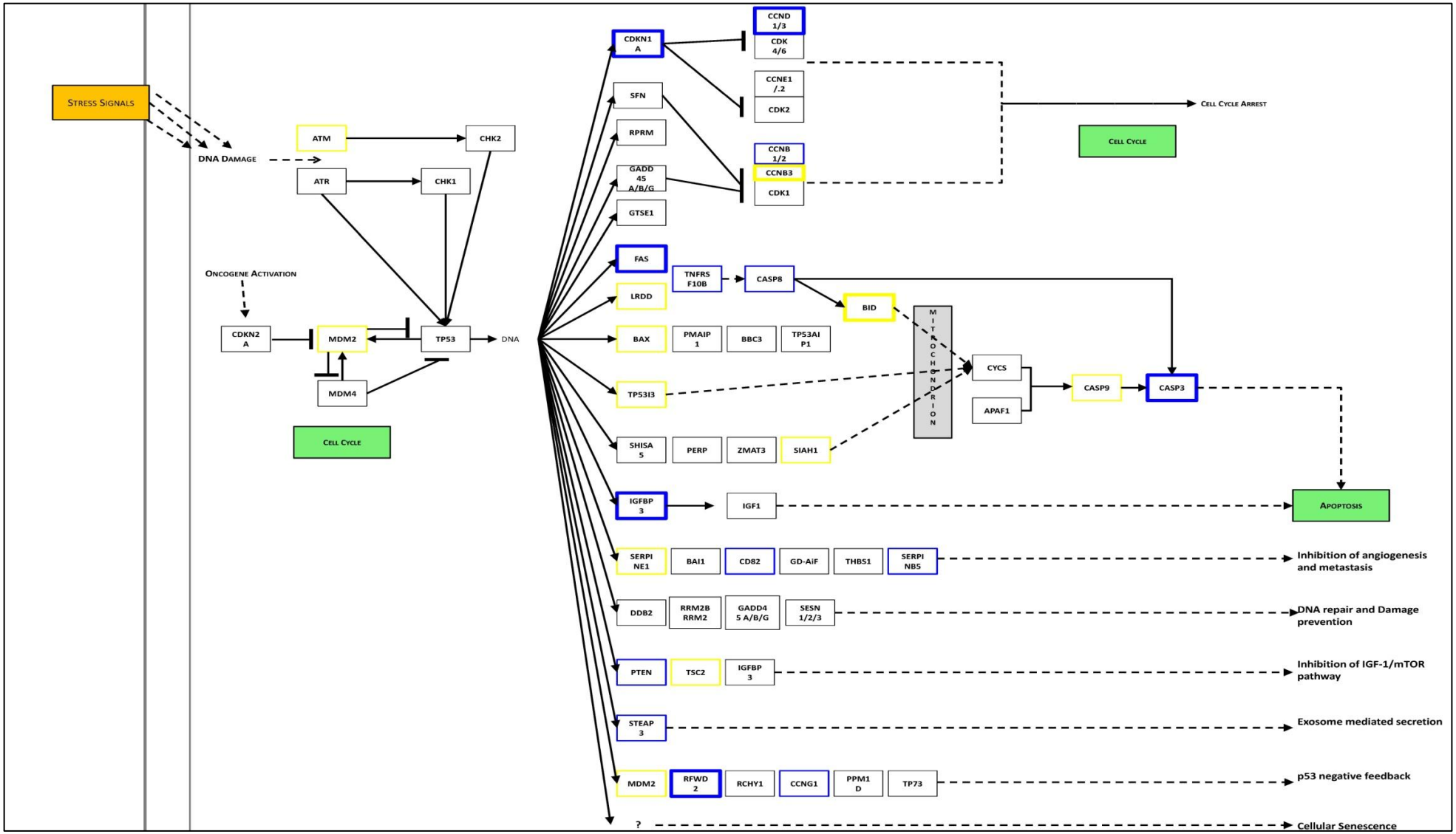

Figure 4.25: Modified p53 signalling pathway reflecting differential gene expression between HP and AP samples

This figure was adapted from the KEGG p53 pathway available from DAVID. The yellow and blue labeled genes represent differentially expressed genes within this pathway, between HPs and APs. The key is similar to that seen in Figure 4.23.

4.12.4. GENES ASSOCIATED WITH CELL CYCLE SHOW DIFFERENTIAL EXPRESSION BETWEEN HYPERPLASTIC AND ADENOMATOUS POLYPS

Changes associated with cell cycle were observed between left and right sided ANs. This pathway was investigated in the HP vs AP comparison due to the importance of cell cycle regulation in relation to cancer development.

124 cell cycle associated genes were identified using the KEGG pathway, which was obtained through DAVID. 119/124 of these genes were present on the DASL microarray. Of these 119 genes, 57 showed differential gene expression between HPs and APs. 17 genes were significant according to the adjusted p-values, and 40 by raw p-values.

A KEGG style pathway was created to reflect the difference in gene expression between HP and AP samples (Figure 4.26).

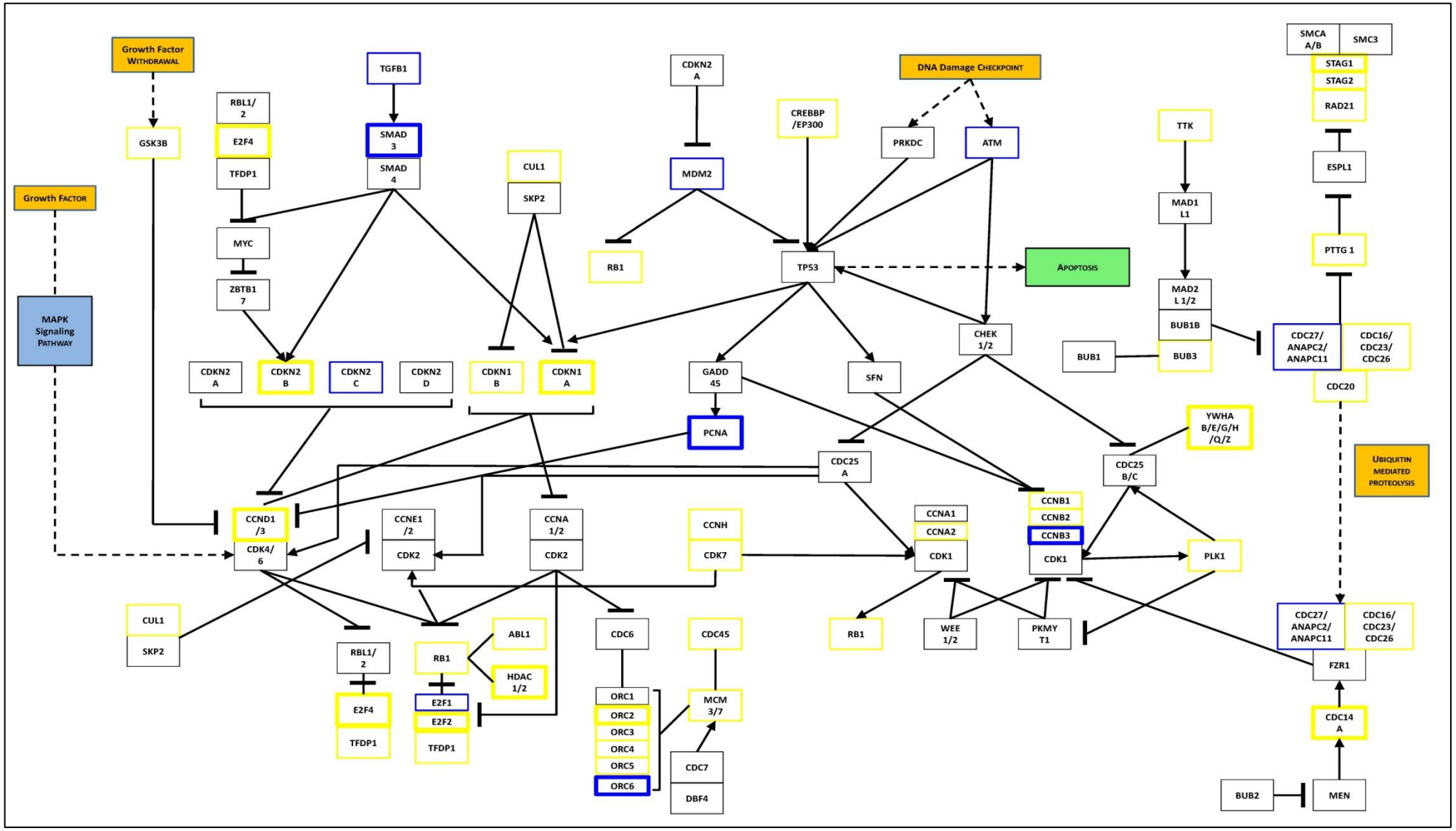

Figure 4.26: Differential expression between HPs and APs associated with the cell cycle

Differential gene expression is seen in genes associated with cell cycle regulation when comparing HPs and APs. The coloured genes are of interest. This figure is annotated in the same way as outlined in Figure 4.23.

4.12.5. DIFFERENTIAL GENE EXPRESSION OF GENES ASSOCIATED WITH APOPTOSIS IS SEEN IN HYPERPLASTIC AND ADENOMATOUS POLYPS

During pathway analysis of HP and AP samples, 14 genes were identified from the Apoptosis pathway in DAVID (Table 4.15). Further analysis was undertaken to identify the genes associated with Apoptosis, which show differential expression between HP and AP samples. Using the KEGG pathway obtained from DAVID, 82 apoptosis-associated genes were identified.

75/82 apoptosis-associated genes were present on the microarray. Of these genes, 35 were significantly differentially expressed between HP and AP samples; 13 according to the adjusted p-value, and 22 according to the raw p-value.

As with previous pathways, a KEGG style pathway was created to compare the differential expression of genes associated with Apoptosis between HP and AP samples (Figure 4.27).

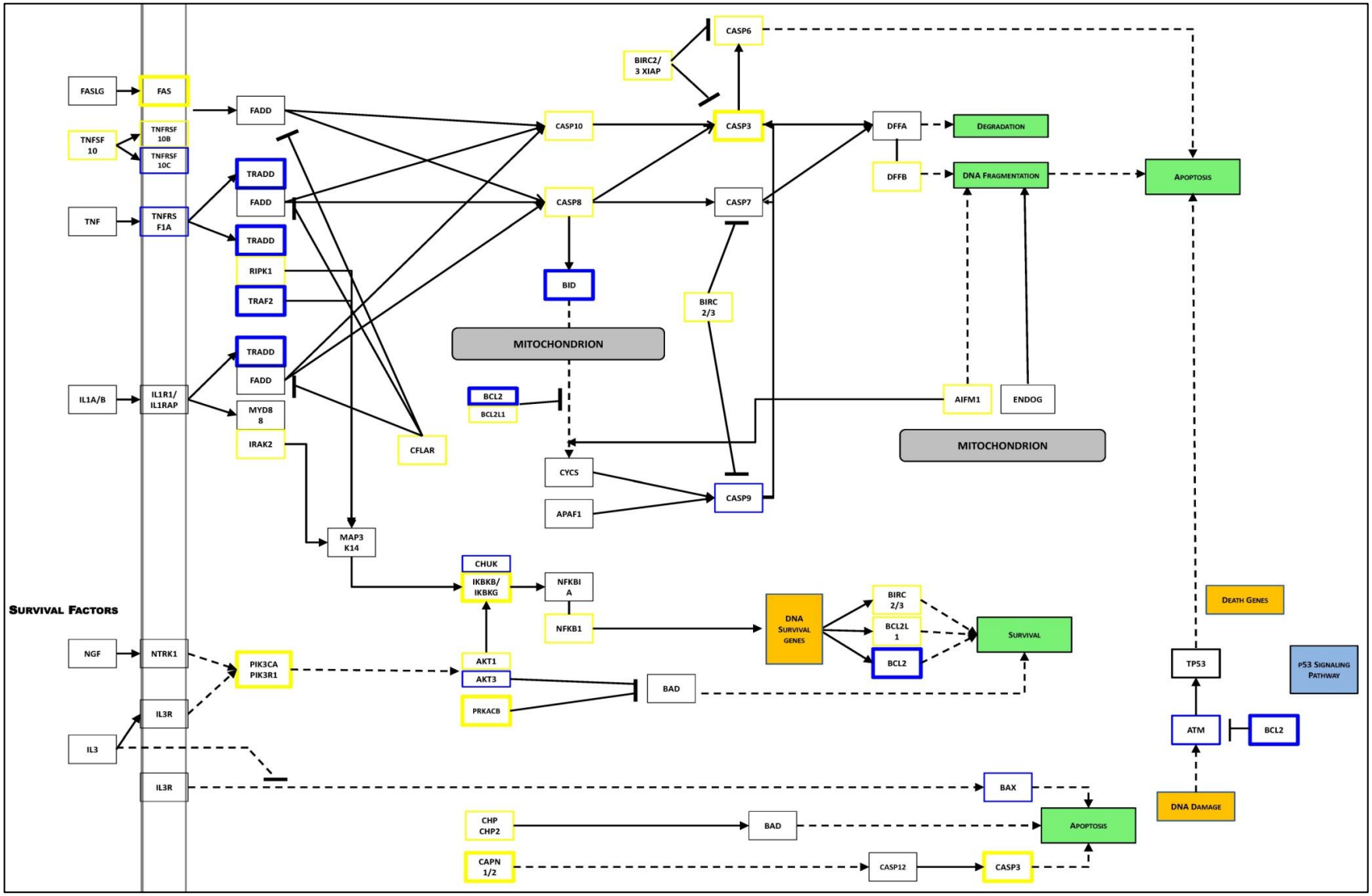

Figure 4.27: Apoptosis pathway

This pathway identifies genes associated with apoptosis, which are differentially expressed between HPs and APs. The annotation of this figure is the same as described previously in Figure 4.23.

4.12.5.1. GENES ASSOCIATED WITH MULTIPLE PATHWAYS ARE DIFFERENTIALLY EXPRESSED BETWEEN HYPERPLASTIC AND ADENOMATOUS POLYPS

During the analysis of the five KEGG pathways presented above, several genes overlap more than one pathway (Figure 4.28). These genes may be important in carcinogenesis due to the influence that they exert over multiple pathways.

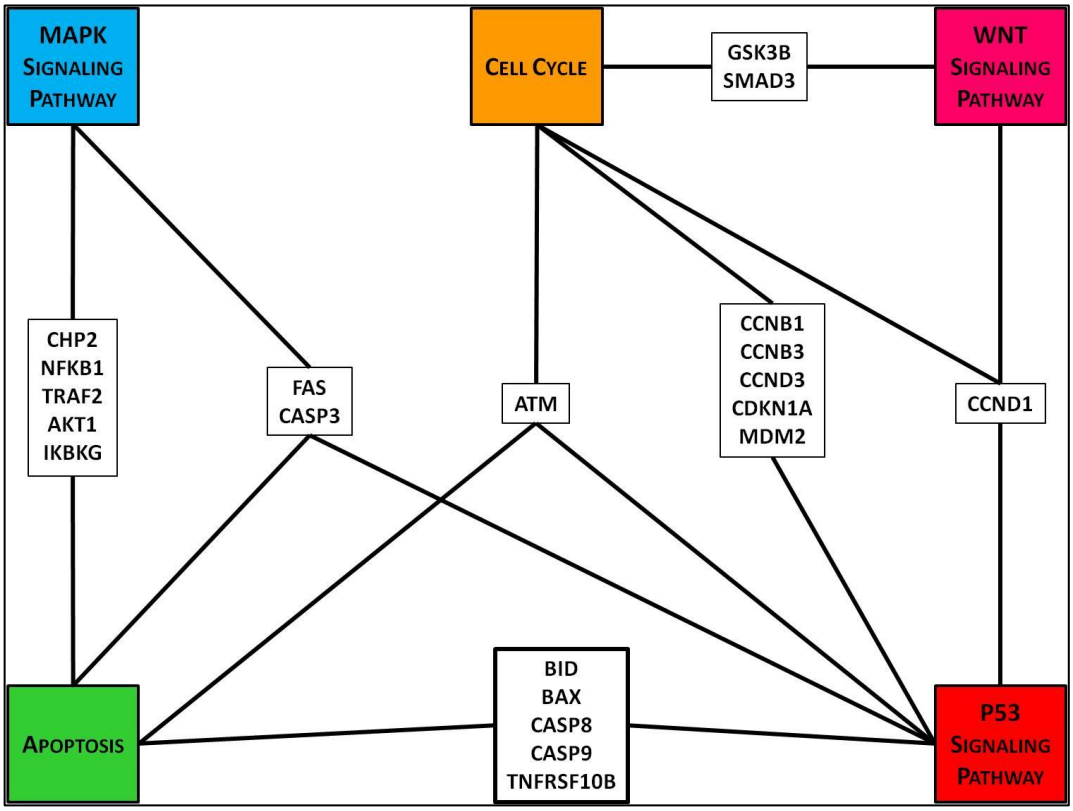

Figure 4.28: Genes showing differential expression in more than one pathway
The genes presented in this figure are differentially expressed between HP and AP samples. In addition, they are associated with more than one of the pathways discussed throughout this chapter. Therefore, these genes may be potential candidates that can be used to differentiate between HPs and APs. In addition they may be useful in understanding the progression of APs to cancer, and the senescence/lack of tumour progression of HPs.

4.12.6. CO-EXPRESSION OF GENES HELPS TO IDENTIFY NOVEL PATHWAYS SHOWING DIFFERENTIAL EXPRESSION BETWEEN HYPERPLASTIC AND ADENOMATOUS POLYPS

The generation of top-tables and their subsequent analysis using MEV, DAVID, COXPRESSdb and PubMed revealed several interesting gene relationships. During the search for candidate genes for qRT-PCR validation, COXPRESSdb was used to identify relationships between different genes. For example Figure 4.29 shows co-expressed genes associated with *MS4A12.* These gene networks were compiled together to provide an overview of the genes within the top table that appear to be related (Figure 4.30). 77 significant genes (raw and adjusted p-values <0.05) were identified and included in the network.

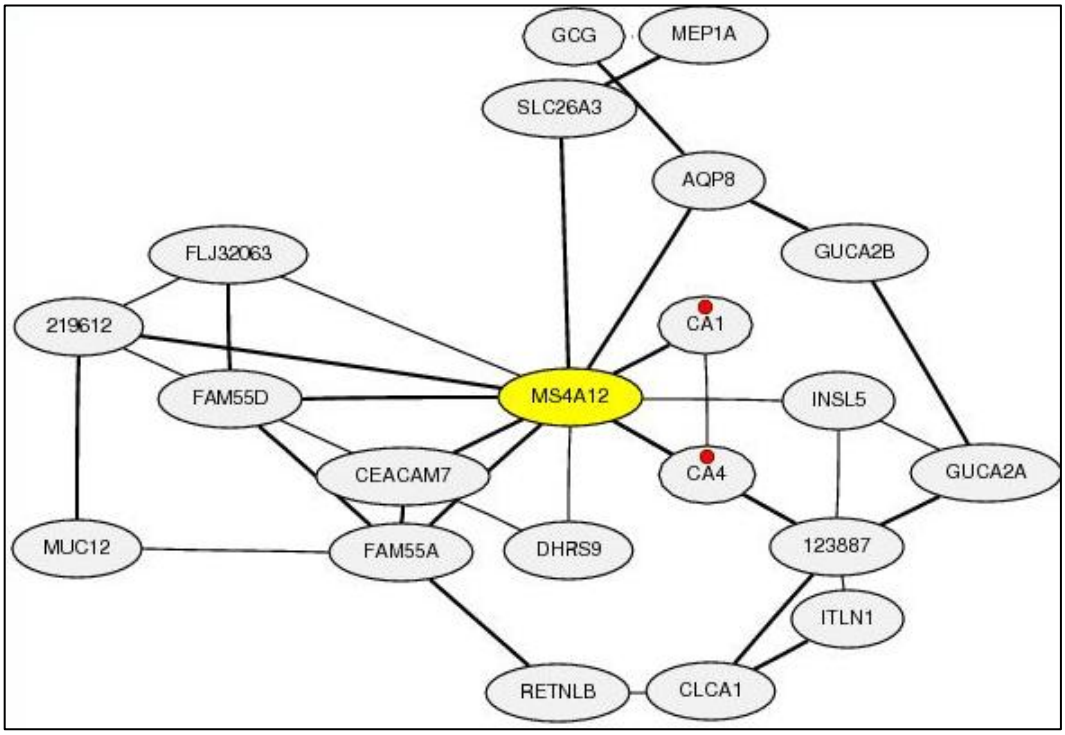

Figure 4.29: COXPRESSdb gene interaction network

This figure shows an example of the COXPRESSdb gene interaction networks produced during the analysis of the top-table comparing HP and AP samples. The gene of interest is highlighted in yellow, in this example it is *MS4A12*. For each gene of interest in the top-table, a COXPRESSdb gene interaction network was produced and compiled where associations were noted.

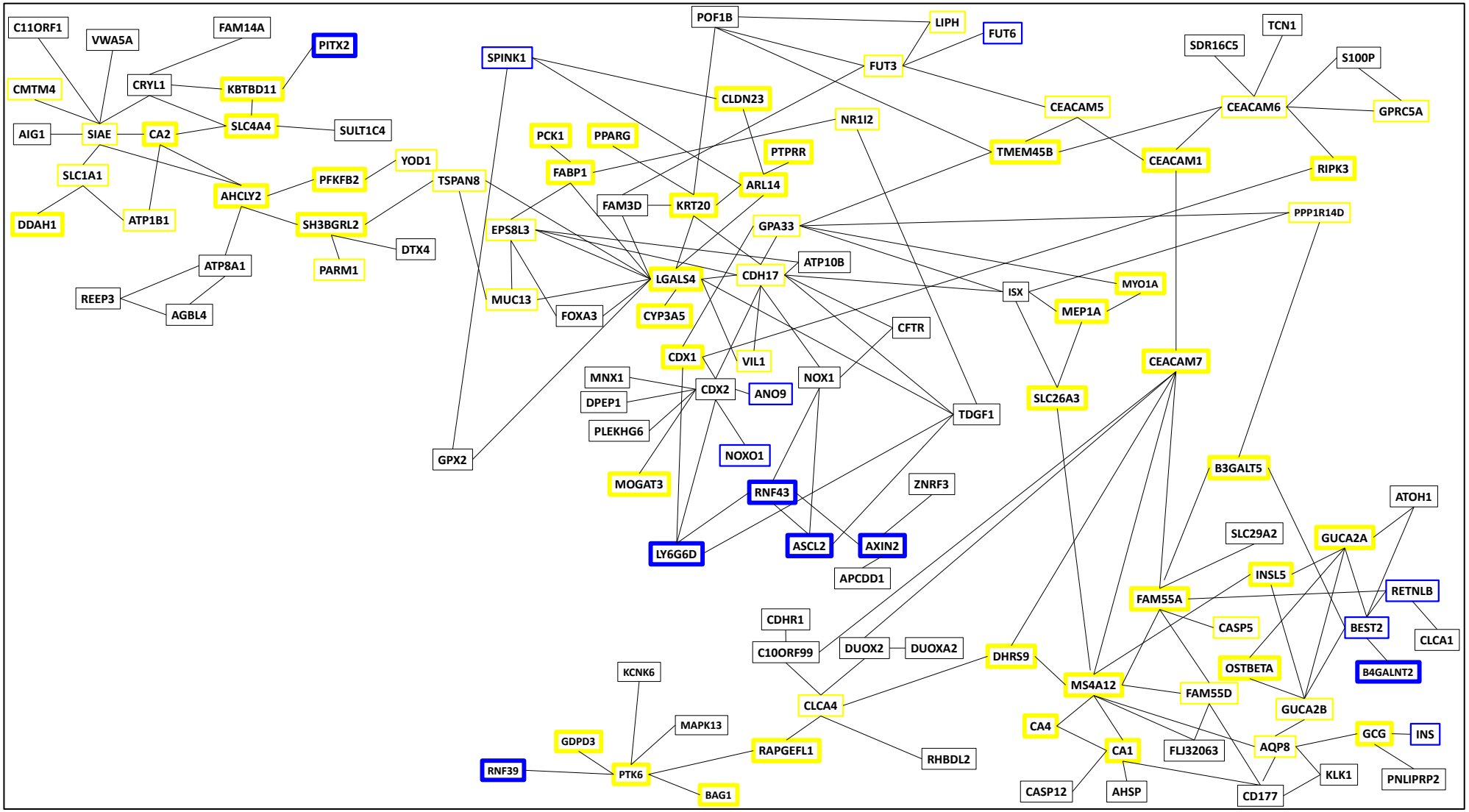

Figure 4.30: Gene network identifying interactions between DEGs in APs relative to HPs

This network was created with the use of the top-table comparing HPs and APs and COXPRESSdb. This network shows how genes identified in the aforementioned top-table are somehow associated with one another. The genes are labeled in the same way as those presented in Figure 4.23. Black outlined genes are not significantly DEGs in the HP-AP comparison, but have been included to allow visualization of how some of the genes are related.

4.13. DISCUSSION

4.13.1. MICROARRAY QUALITY CONTROL AND VALIDATION

Appropriate quality control (QC) is an integral part of a microarray design. Not only are QC steps required in the selection and processing of sample RNA prior to hybridization, they are also required during and after the microarray experiment. The results presented in this Chapter suggest that the Training Set samples were of a sufficient quality to be suitable for microarray analysis (4.3).

The use of lumi in the initial QC stages of the microarray validation was invaluable. Being specifically designed for use with Illumina microarrays, this package provided multiple QC applications that could be used to assess the quality of the array. Without the variety of QC options within lumi, it is possible that some of the samples identified as outliers/failed arrays would not have been identified. Following removal of the three outliers (NC292, A161N, A161HP), the results of the QC analysis suggested that the microarray data were suitable for subsequent analysis.

The data were normalised using the quantile normalisation method available within the lumi package. Quantile normalisation has previously been reported to show advantages over alternative methods of normalisation in relation to speed, bias and variance criteria (Bolstad et al., 2003). The data also underwent Variance Stabilising Transformation (VST) as this method helps to improve the detection of DEGs while limiting the number of false positive genes identified (Lin et al., 2008).

The limma package from R Bioconductor was utilized to produce top-tables of significantly DEGs between specific sample-group comparisons (Smyth, 2004). This

software has been previously described in the literature, with claims that the moderated limma t-test is equal or superior to alternative statistical tests, especially when small sample numbers are present (Jeanmougin et al., 2010, Dondrup et al., 2009).

The incorporation of two technical replicates on the array enabled a correlation of gene expression to be performed in the repeated samples. The high degree of correlation between the replicate samples for both A190N and A190P confirms that there was little technical variability suggesting that the results of the microarray are reproducible.

The validation of the microarray results were assessed following the suggestions by Chuaqui *et al* (2002). Confirmation of gene expression profiles identified in this thesis came from both previously published literature and other microarray data. The use of the GSE4173 dataset was particularly useful as it allowed a direct comparison to be made with the results of the DASL microarray (for NC vs AP). The high concordance of the findings between these two studies provides additional support that the results obtained from the DASL microarray are reliable.

In summary, one important objective was to ensure that the microarray experiment was valid, reliable and reproducible. Using specific QC analysis, replicate analysis and literature review the microarray results appeared to meet the requirements of this aim.

4.13.2. DIFFERENTIAL GENE EXPRESSION IS SEEN BETWEEN NORMAL COLORECTAL TISSUES

Comparative gene expression analysis of different normal-appearing colorectal mucosa could provide insight into genes and pathways that are associated with an increased susceptibility to CRC development (Hong et al., 2007). This chapter presents analysis of three normal appearing colorectal tissues (NCs, ANs and KNs), which show some differential gene and pathway expression. The differences identified between these tissues suggest that widespread changes in gene expression occur throughout the colon of individuals who have polyps and/or cancer.

When comparing NC and AN tissues, differential gene expression was associated with the Calcium Signalling Pathway. Overall, the genes identified in this pathway were down-regulated in AN tissues compared to healthy controls. The role of Calcium is well described, having a growth–restraining and pro-apoptotic role in colorectal mucosa (Lamprecht and Lipkin, 2003, Lamprecht and Lipkin, 2001, Varani, 2011, Bhagavathula et al., 2005), and an increased risk of adenoma development (Peters et al., 2004). The reduced expression of genes associated with this pathway in AN samples suggests a reduced activity of this pathway, which could result in promoting growth and preventing apoptosis. A recent study has postulated that expression of extracellular Ca^{2+}-sensing receptor (CaR) is reduced in CRCs, which is associated with abnormal differentiation and malignant progression (Rey et al., 2010).The findings of the current study support the theory that disturbances in calcium signalling may be associated with adenomagenesis. Altered calcium signalling also exists in other cancers such as breast (Lee et al., 2002b). The specific genes identified in this pathway are also associated with other cancers, including breast (Lee et al., 2002b), bladder (Zaravinos

et al., 2011), and colon (Toyota et al., 1999). *CACNA1G* in particular is down-regulated in CRC due to promoter hypermethylation. This gene appears to be an early marker for carcinogenesis, since differential expression is seen in normal appearing colorectal mucosa.

Six pathways showed differential expression between NC and KN tissues. Several of these are associated with cardiology related pathways, which may be an artifact rather than significant results. However, several of the genes identified in these pathways have been associated with cancer. For example, increased expression of *LMNA* is associated with increased invasion of CRCs and worse prognosis (Willis et al., 2008, Belt et al., 2011), elevated expression of *ITGA5* is associated with increased cell adhesion and migration in breast cancer cells (Qin et al., 2011, Wong et al., 2011), and *ITGA8* is a potential biomarker for ovarian cancer (Cai et al., 2007). These findings suggest that sometimes the genes identified within a pathway, and not the function of the pathway itself, are important to investigate.

Additional pathways identified in the NC vs KN comparison include Nucleotide Excision Repair (NER), ECM-receptor interaction and Haematopoietic cell lineage. NER and Haematopoietic cell lineage have shown DEGs between normal and cancer colorectal tissues (Skrzypczak et al., 2010). NER, a DNA repair pathway is important for correcting abnormalities in the DNA sequence (Leibeling et al., 2006). The genes identified in this pathway show a mixture of up and down-regulation in KN tissues. *XPC* shows up-regulation in KN samples. Since the XPC protein is important in the initial recognition of DNA lesions (Benhamou and Sarasin, 2000), this gene could potentially be up-

regulated in KN tissues due to an increased number of DNA lesions associated with the presence of APs and CRCs in these patients.

ECM-receptor interaction, which is thought to affect cell migration (Ertel et al., 2006) and tumour progression (Krupp et al., 2011), showed consistent up-regulation in KNs with the exception of *CD36*. This pathway is not well described in the literature; however it has been identified as a pathway showing some disruption in several cancers. In addition it is also thought to interact with the Cell Cycle Pathway (Krupp et al., 2011), which was identified as showing differential expression between left and right sided ANs. The genes associated with Cell Cycle were largely down-regulated in right sided ANs, suggesting a different mechanism of adenoma/carcinoma development according to location.

A study by Hong et al identified seven differentially expressed genes that were consistently up-regulated in the mucosa from cancer-associated patients compared to normal mucosa from healthy controls. All of these genes (*KRT24, VIP, FOS, FOSB, EGR1, CYR61* and *UCHL1*) were consistently up-regulated in the NC-KN comparison. These genes are associated with pathways such as Wnt and MAPK, suggesting that a disturbance in cell signalling pathways is essential for carcinogenesis (Hong et al., 2007). Overall the agreement between the results presented in this Chapter and those presented by other researchers provides evidence that the results of the DASL microarray are reliable.

When comparing AN and KN tissues, four genes belonging to the PPAR Signalling Pathway were differentially expressed, showing down-regulation in KNs relative to ANs. This pathway is important in regulating diverse cellular functions such as cellular

differentiation, proliferation and apoptosis (Krupp et al., 2011). *HMGCS2* is one of the genes identified in this pathway as being down-regulated in KN tissues. This gene is a target of *C-MYC*, which has been shown to be down-regulated in moderate and poorly differentiated CRCs (Camarero et al., 2006). An additional gene of interest in this pathway is *PPARG*. *PPARG* function is regulated by the Wnt/β-catenin pathway (Jansson et al., 2005) and it is thought to be down-regulated in CRC (Pancione et al., 2010). Recent studies have used *PPARG* agonists to induce apoptosis in CRC cells and thus suppress CRC development, however the mechanism of action is unknown (Ban et al., 2010, Wang and DuBois, 2010). This finding suggests that the PPAR Signalling Pathway may be associated with carcinogenesis, showing disregulation in early disease.

The comparison between left and right-sided ANs identified two pathways; Cell Cycle and Jak-STAT Signalling pathway. No studies have investigated these pathways in left and right-sided ANs; however the different expression of these pathways has been investigated in CRC and normal colorectal tissues. The Jak-STAT pathway is an important component of signal transduction pathways involved with cellular survival, apoptosis, proliferation and differentiation (Spano et al., 2006). Little literature is available on the expression of this pathway in CRC, however one study implied that genes in this pathway are up-regulated during carcinogenesis, but not in the normal-appearing colorectal mucosa (Corvinus et al., 2005), as found in this study.

Differential expression associated with Cell Cycle was also seen between left and right sided AN samples. The majority of the genes identified in this pathway showed down-regulation in right-sided ANs. Since no studies have investigated left and right sided AN

tissues, confirmation of these results in the literature was not possible. However, *CCNE2* expression has previously been described as up-regulated in human cancers relative to matched normal tissues, suggesting a possible role in carcinogenesis (Gudas et al., 1999).

4.13.3. HYPERPLASTIC AND CANCER-ASSOCIATED POLYPS HAVE DIFFERENTIAL GENE EXPRESSION PROFILES

Within the literature there is no direct comparison between HP and KP type samples, making the results presented in this chapter novel. The top-table analysis identified only two DEGs (*ANXA2* and *HLA-G*), both showing down-regulation in KPs. *ANXA2* will be discussed in detail in Chapter 5. *HLA-G* is a nonclassical major histocompatibility complex (MHC) type 1 antigen, which is thought to cause impairment of the body's immune response against tumour cells (Sheu and Shih Ie, 2010). One paper compared serum levels of *HLA-G* from patients with benign and malignant colorectal tissues, which found a slightly increased expression in polyps relative to HPs (Zhu et al., 2011). Although this finding appears to contradict the findings presented in this chapter, the comparison in the paper by Zhu *et al* was not identical to the one presented herein, since cancer-associated polyps were not included in the study design of the paper. Interestingly, when comparing HPs with right-sided KPs, *HLA-G* expression was increased five-fold in the KPs. This finding supports the findings of Zhu *et al*, and suggests that the expression of *HLA-G* is dependent on the site of the colorectal lesion.

The HCL (Figure 4.14) identified only a few genes that showed differential expression between HP and KP tissues. These genes were *HLA-G, FZD9* and *OSBPL7*. The expression profiles of the remaining genes did not allow easy differentiation between

HP and KP tissues. *FZD9* is member of the frizzled gene family, which encodes receptors required for Wnt Signalling. *FZD9* is a relatively unknown gene with regards to CRC, making it a novel gene that could be validated in future studies. It has been shown to be up-regulated in gastric cancer (Kirikoshi et al., 2001). *OSBPL7* is another relatively novel gene, whose function is currently unknown in CRC.

Several pathways were identified when comparing the expression of left and right sided KPs with HPs. Once again, differential expression of genes associated with the Calcium Signalling Pathway was observed, providing further evidence that this pathway has a role in adenoma/carcinoma formation. An additional pathway identified was the Hedgehog Signalling Pathway, which showed a mixture of up- (*CSNK1D* and *PRKACB*) and down- (*ZIC, BMP8A* and *PTCH2*) regulated genes in HPs relative to right-sided KPs. The Hedgehog Signalling Pathway is thought to be up-regulated in CRCs, and blockade of this pathway could be a potential therapeutic target (Yoshikawa et al., 2009).

Other genes showing differential gene expression between right-sided KPs and HPs were identified during the SAM analysis. Some of the genes show clear differences in gene expression between the two tissues, making them better potential targets to investigate. Examples of such genes include *GCG, CHGA* and *PITX2*. Decreased expression of both *GCG* and *CHGA* in right-sided KPs was seen in the top-table comparison and the SAM analysis in MEV. *GCG* expression is not widely published, however a recent study suggested that this gene could be used to differentiate between HP and AP samples (Galamb et al., 2008b), however cancer-associated polyps were not included in the study design. *CHGA* has previously been investigated as a

potential tumour marker but lack of specificity has been questioned (Molina et al., 2011). *PITX2* encodes a transcription factor, which is under the control of Wnt and Hedgehog Signalling Pathways. Although it is thought to have many roles, the function of *PITX2* in CRC is relatively unknown. One known function of *PITX2* is the activation of Cyclin D2, which is a growth-regulating gene (Baek et al., 2003). *PITX2* showed elevated expression in right-sided KPs relative to HPs as shown in Figure 4.15. A literature review supported the finding that *PITX2* expression is increased in CRC associated tissues, with increased expression being associated with behavior and survival of the cancer cells (Hirose et al., 2011). Given the differential expression of *PITX2* between these two tissues, and the relative lack of literature in this area, *PITX2* appears to be an important gene that warrants further investigation in the future.

Differential expression between left-sided KPs and HPs identified the gluconeogenesis/glycogen pathway. With the exception of *ALDOA*, all of these genes showed up-regulation in left-sided KPs. This finding was confirmed by a recent study, which also identified genes associated with these pathways as being differentially expressed and important in CRC development (Yeh et al., 2008). The mechanisms that these genes employ to bring about an effect on tumorigenesis are still unknown, warranting further investigation.

4.13.4. POLYPS FROM CANCER PATIENTS SHOW SUBTLE DIFFERENCES IN GENE EXPRESSION WHEN COMPARED TO ADENOMATOUS POLYPS

The comparison between AP and KP tissues did not reveal any significant DEGs in the top-table. However, when investigating the top 300 genes in DAVID, five genes from the PPAR Signalling Pathway showed differential expression. This pathway showed

relative up-regulation of all genes (with the exception of *PPARD*) in KP samples relative to APs. This finding contradicts the finding reported previously when discussing the results of the "normal" tissues. However, members of the PPAR Signalling Pathway can act as either tumour suppressor or accelerators, promoting their potential as candidates for therapeutic intervention (Wang and DuBois, 2010). A specific PPAR gene identified during this analysis was *PPARD*, which is thought to be a transcriptional target of the APC pathway and induced by oncogenic *KRAS* (Wang and DuBois, 2010). Mixed findings in terms of gene expression of *PPARD* have been found, and its function in CRC development therefore needs further investigation. An additional gene in this pathway is *FABP1*. This gene showed up-regulation in KPs relative to APs, contradicting a finding by Lee *et al* (2006), which suggests that *FABP1* expression decreases along the adenoma-carcinoma sequence (Lee et al., 2006). However, other studies have suggested that increased levels of *FABP1* are found in the blood of patients with CRC (Lauriola et al., 2010, Smirnov et al., 2005). The expression of *FABP1* along the adenoma-carcinoma sequence may depend on other factors that were not taken into account in either of these studies, such as location, grade and tumour type.

When separating AP and KPs based on their anatomical location, differential gene expression was noted between left sided APs and KPs. Three pathways were identified that showed differential gene expression; PPAR Signalling Pathway, MAPK Signalling Pathway and Haematopoietic Cell Lineage. Three new genes were identified in the PPAR Signalling pathway in addition to *FABP1* and *PPARD*. The expression of this pathway appeared to be in some way linked to the anatomical location of the tissues of interest, which could explain some of the discrepancies seen in the literature. The

MAPK Signalling pathway shows increased expression in the genes encoding calcium channels (*CACNA1E, CACNA1H* and *CACNA2D4*) and *DUSP7* in left-sided KPs, while the remaining genes were down-regulated in left sided KPs. A previous study looking at the mRNA expression of calcium channel genes suggests that they could be a marker of transformation or proliferation since they are up-regulated in cancer tissues (Wang et al., 2000). This theory fits with the results presented, since the calcium channel genes were up-regulated in cancer-associated polyps. *DUSP7* represents a relatively unknown gene, which has previously shown up-regulation in leukaemias (Keyse, 2008, Levy-Nissenbaum et al., 2003a, Levy-Nissenbaum et al., 2003b). Genes showing down-regulation in left-sided KPs relative to APs included *RAC1, FGFR1* and *GADD45A*. Increased expression of *RAC1* is associated with increased rates of tumorigenesis (Espina et al., 2008, Gomez del Pulgar et al., 2007), while increased expression of *FGFR1* is associated with liver metastases (Sato et al., 2009). Increased expression of these two genes is likely to be a later event in the carcinogenic transformation, which could explain the lower levels of expression seen in the DASL microarray data. Low levels of *GADD45A* on the other hand are associated with reduced DNA repair, since *GADD45A* is an important regulator of NER and is important for genomic stability (Jung et al., 2007). Low levels of this gene were seen in left-sided KPs, suggesting that some level of deficient DNA repair, and genomic instability could be present in KP samples.

4.13.5. GENE EXPRESSION PROFILES OF MATCHED TISSUES

Comparing paired AN and AP samples did not reveal any significant genes or pathway associations according to the top-table. A study by Sabates-Bellver *et al* (2007) identified *KIAA1199* as being up-regulated in APs relative to matched ANs. *KIAA1199* is

thought to be important in Wnt Signalling (Sabates-Bellver et al., 2007). This gene was identified in the AN vs AP paired comparison where it was shown to be up-regulated according to the raw p-value (p=0.000835, fold change = 1.3). In addition it was also identified when comparing NC and AP samples, showing up-regulation according to the adjusted p-value (p=0.0147, fold change = 2.00). In addition to *KIAA1199* expression, additional genes were identified by Sabates-Bellver *et al* (2007). The genes mentioned in this paper were identified in the NC vs AP top-table, and gene expression agreement was shown for most of the genes, and was significant according to the adjusted p-value. In particular, genes associated with Transcription (*ASCL2, GRHL1*), cell proliferation, differentiation and apoptosis (*REG4, TGFB1* and *IGFBP2*) were all up-regulated in APs relative to normal tissues (Sabates-Bellver et al., 2007). Cell adhesion genes (*CLDN23* and *CEACAM7*) were down-regulated in APs (Sabates-Bellver et al., 2007). *CEACAM7* is a well known tumour marker, which is down-regulated in APs and CRCs however its function in the development of CRC is unknown. *CA1* and *FRMD6* were both identified during the SAM analysis of paired AN and AP tissues. Down-regulation of *CA1* has been seen in APs and CRCs, with an associated loss of differentiation (Sowden et al., 1993). The role of *FRMD6* is currently unknown.

When investigating the expression of matched tissues based on anatomical location, only one gene was identified. *TRIM29* shows up-regulation in polyps on the right side of the colon relative to matched normal tissues. This gene is used to differentiate between aberrant crypt foci (ACF) and normal mucosa, showing increased expression in ACF (Glebov et al., 2006). No genes showed differential expression between left sided polyps and matched normal tissues.

4.13.6. SIGNIFICANT DIFFERENCES IN GENE EXPRESSION ARE SEEN BETWEEN NORMAL MUCOSA, HYPERPLASTIC AND ADENOMATOUS POLYPS

The analysis of NC, HP and AP tissues identified numerous DEGs and pathway involvement. The HCL (Figure 4.18) revealed "blocks" of genes showing differential expression across the three tissue groups. Of note, expression of genes in HP and AP tissues showed mostly inverse expression, highlighting the fundamental differences in gene expression between these two tissues. The gene list created by SAM enabled the three groups to be differentiated from one another as shown by PCA (Figure 4.18). Genes associated with Nitrogen Metabolism were found to be down-regulated in APs relative to NCs and HPs. This finding confirms a study comparing CRCs and matched normal tissues, where the expression of *CA2* was down-regulated in CRCs (Bianchini et al., 2006).

4.13.6.1. GENES ASSOCIATED WITH P53 SIGNALLING ARE UP-REGULATED IN HYPERPLASTIC POLYPS COMPARED TO NORMAL CONTROLS

When investigating the difference between NC and HP tissues, several pathways were identified; Proteasome, Aminoacyl-tRNA biosynthesis, Axon guidance, p53 Signalling and Pyruvate Metabolism. Genes identified in the axon guidance pathway show up- (*EPHA2* and *SLIT2*) and down- (*SEMA3A*) regulation in breast cancer (Harburg and Hinck, 2011) and HP tissues. Proteasome associated genes have been shown to indicate an increased risk of liver metastasis in CRC (Hu et al., 2008) and are currently under investigation as potential anti-cancer targets for use in rectal cancers (Conrad et al., 2011, Roccaro et al., 2006).

Of interest is the identification of the p53 Signalling pathway, which showed up-regulation in HPs relative to NCs. An additional study also found increased p53

expression in one third of HPs (Pap et al., 2011). p53 Signalling results in increased apoptosis, cell-cycle arrest and cellular senescence (Vigneron and Vousden, 2011, Larsson, 2011). Since HPs represent senescent lesions with increased rates of apoptosis, it is logical that p53 Signalling is increased in these lesions.

Using the SAM generated gene list the NC and HP samples were easily differentiated from one another, as shown by PCA (Figure 4.19). The HCL itself showed clear differential gene expression between the two sample groups, with some genes showing more obvious differential expression. The differences found between NC and HP tissues provide evidence that they are distinct entities, which should not be considered similar.

4.13.6.2. GENES ASSOCIATED WITH WNT-SIGNALLING ARE UP-REGULATED IN ADENOMATOUS POLYPS COMPARED TO NORMAL CONTROLS

The comparison between NC and AP tissues identified numerous DEGs in the top-table. Three pathways were identified during the analysis, which showed differential expression in genes associated with Nitrogen Metabolism, Pathways in Cancer and the Wnt Signalling Pathway. Within these pathways, several genes were identified that have been mentioned previously; *AXIN2, FZD9* and *PPARD*. With the exception of *PPARD*, these genes were up-regulated in APs relative to NCs, supporting the findings of previously presented results. Increased Wnt-Signalling in AP tissues is likely due to loss of *APC*, or mutation of β-catenin (Taipale and Beachy, 2001). Unfortunately, mutations of these genes were not tested in this thesis.

The HCL produced during the SAM analysis of NC and AP tissues identified 29 genes showing differential expression. According to PCA, this gene list allows NC and AP tissues to be distinguished from one another.

4.13.7. HYPERPLASTIC POLYPS SHOW WIDESPREAD DIFFERENCES IN GENE EXPRESSION WHEN COMPARED TO ADENOMATOUS POLYPS

HP and AP tissues represent benign and potentially malignant polyps respectively. The main focus of this thesis has been to investigate the gene expression profiles of these two tissues. Since HPs represent a benign lesion, it was hoped that critical genes/pathways identified in HPs could be targeted in APs to prevent malignant transformation.

The comparison between HP and AP tissues identified over 1000 DEGs according to the adjusted p-value. This alone highlights the diverseness of these two lesions. In addition, when analyzing the DEGs identified between these two tissues, over 30 different pathways were identified (Table 4.15). Once again, the difference between these tissues is apparent.

The result of the SAM HCL shows the differential expression of the 120 genes identified. For a large proportion of these genes, their expression between HPs and APs was inversely related. Genes such as *ASCL2* and *AXIN2,* which will be described in Chapter 5, were identified in the HCL. The results of PCA showed that after limiting the gene list to those present in the HCL, the two groups polarized to separate groups. Interestingly, even before limiting the gene list, the two groups showed a tendency towards opposite directions.

During the analysis of the results it became apparent that these tissues showed distinct patterns of gene expression in important physiological process such as Wnt-Signalling, MAPK Signalling, p53 Signalling, Apoptosis and Cell Cycle. These fundamental processes are disrupted in several cancers, but are important in the development of CRC.

With regards to Wnt-Signalling, *AXIN2, APC* and *CTNNB1* were up-regulated in APs and down-regulated in HPs. *AXIN2* and *APC* are associated with the β-catenin destruction complex. *GSK3B* was up-regulated in HPs. Down-regulation of *GSK3B* is associated with stabilisation and nuclear accumulation of beta-catenin. Nuclear β-catenin activates transcription of *CCND1, MYC, FGF18* and *FGF20*, which are required for cell-fate determination (Katoh, 2006). Therefore, higher levels of *GSK3B* in HPs must result in destabilization and cytoplasmic accumulation/degradation of β-catenin and thus prevent the activation of genes like *C-MYC*. In the polyps tested, no changes were seen in *C-MYC* expression; however these lesions occur early in the carcinogenic pathway.

Additional genes in the Wnt Signalling pathway that showed differential expression between HPs and APs are *FRAT2* and *NKD1*. *FRAT2* is a *GSK3B* binding protein and positive regulator of Wnt Signalling, which showed increased expression in APs relative to HPs. Up-regulation of this gene in tumours has been described (Nguyen et al., 2010). *NKD1* was up-regulated in AP samples relative to HP samples and is thought to contribute to adenoma progression through a β-catenin independent pathway (Nguyen et al., 2010).

The MAPK Signalling pathway is responsible for diverse biological processes such as proliferation, differentiation and apoptosis (Fang and Richardson, 2005). The genes in

this pathway showed increased expression in HPs relative to APs, for the majority of genes. A study has also shown down-regulation of MAPK activity in CRC, supporting the finding of the present study (Gulmann et al., 2009). *PTPRR* is an interesting gene belonging to the protein tyrosine phosphatase family, which is up-regulated in HPs. Silencing of this gene is associated with CRC (Menigatti et al., 2009), and it is currently under review as a potential therapeutic marker (Barr and Knapp, 2006). *STNM1* is one of the few genes in this pathway that showed up-regulation in APs relative to HPs. This gene encodes an oncoprotein, which is associated with uncontrolled cell proliferation, cancer progression, migration and metastasis (Zheng et al., 2010). Several other genes showed up-regulation in APs relative to HPs, such as the calcium channel genes (*CACNA1C*, *CACNA1G* and *CACNA1S*), *BDNF*, *TRAF2*, etc. *BDNF* exerts an anti-apoptotic and increased proliferation effect, making it a desirable ally for cancer cells (Brunetto de Farias et al., 2010, Akil et al., 2011). *TRAF2* is required for cell growth, with increased expression associated with cancer growth suppression (Shitashige et al., 2010). Genes in this pathway showing up-regulation in AP tissues are therefore likely to be involved with increased affinity for carcinogenesis and could represent new markers for therapeutic intervention.

The p53 Signalling pathway showed a diverse pattern of gene expression, which is more difficult to interpret than the Wnt or MAPK Signalling pathways. A study by Ban *et al* (2010) identified several of the genes deemed significant in this pathway and divided them into functional groups. Cyclin B1 and Cyclin D1 were both down-regulated in HPs, and represent cell cycle regulators. *CASP3*, *CASP8* and *BAX* are pro-apoptotic associated proteins, which showed down-regulation in HPs with the

exception of *BAX*. The normal p53 pathway has a protective function to prevent the formation of CRC. The gene expression profiles of genes associated with this pathway were difficult to interpret, however they are likely to be important in distinguishing between HPs and APs. Further analysis of the p53 pathway in these tissues could provide further insight into the differences that exist between HPs and APs.

Correct regulation of Cell Cycle and Apoptosis related pathways is essential for normal tissue maintenance. However, during cancer development these pathways become disrupted and enable cancers to evade apoptosis and continue to proliferate in the absence of cell cycle regulation (Bianchini et al., 2006). These pathways showed disruption between HP and AP samples. Genes associated with Cell Cycle were largely up-regulated in HPs, whereas genes associated with Apoptosis showed a more diverse pattern of gene expression in HPs and APs.

Many of the genes identified in the aforementioned pathways showed overlap between two or more pathways (Figure 4.28). These genes are potentially very important in distinguishing between HP and AP tissues, since they are associated with multiple cellular processes/pathways. Interestingly, for the majority of genes showing overlap between different pathways there is little literature available outlining their roles in CRC. These genes could provide insight into the fundamental differences that exist between HPs and APs and provide new targets for intervention. This is true for all the pathways showing differential gene expression between HPs and APs.

During the background research of potential candidate genes to validate in Chapter 5, it became apparent that the significant genes present in the top-table comparing HP and AP samples were associated with one another. With the use of COXPRESSdb, the

gene network presented in Figure 4.30 was created to outline the links identified between different genes. Although these genes are not associated with a pathway *per se*, it is clear from this analysis that they are connected in some way. A limitation of COXPRESSdb however is the inability to determine the relationship between the connected genes. For example, it is not known whether the genes are activated or inhibited by the associated genes. This pathway demonstrates that there are other pathways and interactions taking place during the development of HPs and APs, which are still unknown. This provides further evidence for the complexity of CRC development, but presents additional candidate genes that can be reviewed and analysed to determine their significance (if any) in polyp formation and differentiation. Further work is required to validate the relationship between these genes.

4.13.8. WEAKNESSES AND LIMITATIONS

Although a large amount of time was spent on the design of the microarray, several limitations were noted while analyzing the data. By using a mixture of both left and right sided APs and KPs, the numbers of DEGs were reduced. This was particularly true in the case of KP samples. It would have been better to include samples that were more homogenous, and omit some variables so that the sample groups were larger. However, despite the limited numbers in some of the comparisons, DEGs were still identified that related to literature findings.

As noted in Table 4.2, only very few DEGs were identified when comparing one group with KPs. In addition, the matched cases (KN vs KP) also revealed very few DEGs, none of which were significant according to the adjusted p-value (<0.05). Although this could represent field effect changes, it could be that the differences between left and right

sided polyps negate each other in this experiment due to the small sample numbers in each group (left and right). In addition, the number of KN and KP samples included in the Validation Set was small and more are required in the future to ascertain the true changes in candidate gene expression in these tissues.

4.14. CONCLUSION

The results presented in this chapter highlight the complexity of colorectal polyp formation. DEGs were identified in all comparisons investigated, with the biggest differences seen between HP and AP samples. Since HPs represent benign lesions, it is these tissues that are of key interest as they may help to determine why some polyps become malignant and others do not. A diverse number of pathways were disrupted between HP and AP tissues, each of which holds a wealth of potential information that could help answer this question. These pathways require further investigation and validation in a larger series of samples, using a qRT-PCR or immunohistochemistry approach.

CHAPTER 5: DIFFERENTIAL GENE EXPRESSION OF CANDIDATE GENES IN COLORECTAL TISSUES

5.1. INTRODUCTION

Chapter 4 presented the primary results of the DASL microarray experiment. This Chapter presents results of qRT-PCR validation of candidate genes that showed differential expression across a panel of colorectal tissues on the microarray. The Validation and Mutation Sets were utilized to allow validation in a larger cohort of samples, and to enable associations between gene expression and mutations of *BRAF* and *KRAS* to be investigated.

5.2. AIMS AND OBJECTIVES

The aim of this chapter was to perform gene expression profiling in different colorectal tissues by qRT-PCR to identify potential genetic markers of CRC development and progression.

Objectives:

- Identify a panel of genes showing differential expression across a range of colorectal tissues that will be selected for qRT-PCR analysis using a range of colorectal tissues.
- Identify associations between candidate gene expression and clinico-pathological parameters such as adenoma type, mutation status and polyp location.

5.3. SELECTION OF CANDIDATE GENES

Criteria from section 4.5.1 were used to select 5 genes (*ANXA2, ASCL2, AXIN2, ETS2* and *G3BP1*) for validation by qRT-PCR. In addition to a significant p-value, all 5 genes had an actual fold change and B-statistic greater than 2. The expression profiles for each gene can be seen in Figure 5. 1.

An additional gene, *TFF2* (Trefoil Factor 2), was also selected for investigation. Although not in the top 20 up or down regulated genes, it had a significant adjusted p-value (p=0.00033) and showed an increased actual fold change (6.82) when comparing NCs and HPs. In addition, the expression of *TFF2* varies across different colorectal tissue types. However, the B-statistic was not above the pre-defined cut off of 2 (B-statistic = 0.223). Therefore this gene was selected to determine the importance of the B-statistic and its ability to predict genes that are differentially expressed. Other studies have been identified that only select genes with a B-statistic >2 as this threshold is thought to identify (more accurately) genes that are differentially expressed (Soon et al., 2009, Hanina et al., 2010, Nguyen et al., 2009).

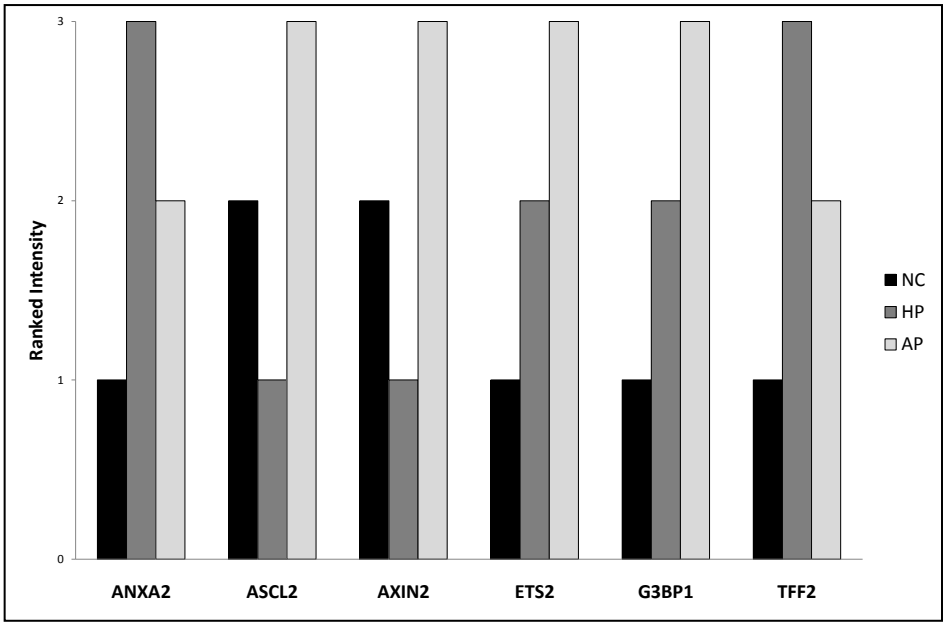

Figure 5. 1: Ranked Intensities of candidate genes

This figure shows the differential gene expression of the six candidate genes in NCs, HPs and APs. Each colorectal tissue was allocated an intensity number according to gene expression of the candidate genes. The scale is between 1 (low expression) and 3 (high expression).For example, ANXA2 has lowest expression in NCs (1) and highest expression in HPs (3).The allocated intensity numbers are not quantitative and are intended to provide a visual trend in gene expression across the tissues, as identified on the DASL microarray.

5.4. VALIDATION OF SIX CANDIDATE GENES

The Validation Set for qRT-PCR comprised 143 colorectal samples: NCs, HPs, APs, KNs and KPs. ANs were not included in this study due to the low abundance of available normal mucosa from patients with APs.

Standard curves were first generated for each primer and probe sets. Each gene was validated separately, but from a common cDNA containing all eight RTPs (*ANXA2, ASCL2, AXIN2, ETS2, G3BP1, TFF2, TPT1,* and *UBC).*

The standard curves showed a range in efficiency from 91.67% (*TPT1*) to 100.32% (*AXIN2*) which is within the optimal range. (Table 5.1 and Figure 5.2).

Table 5.1: Real time qRT-PCR efficiencies of the candidate and EC genes

Gene Name	Intercept	R^2	Efficiency (%)
UBC	34.28	0.99	95.86
TPT1	33.48	0.99	91.67
ASCL2	46.57	0.99	94.29
ANXA2	34.78	0.99	96.17
AXIN2	41.43	0.99	100.32
ETS2	42.14	0.99	94.01
G3BP1	34.28	0.99	93.60
TFF2	46.68	0.99	92.61

The efficiencies of each standard curve are presented within this table. All genes have efficiencies that lie within the optimal range. All R^2 values are greater than 0.98 suggesting that the curves follow a linear trend across the serial dilutions.

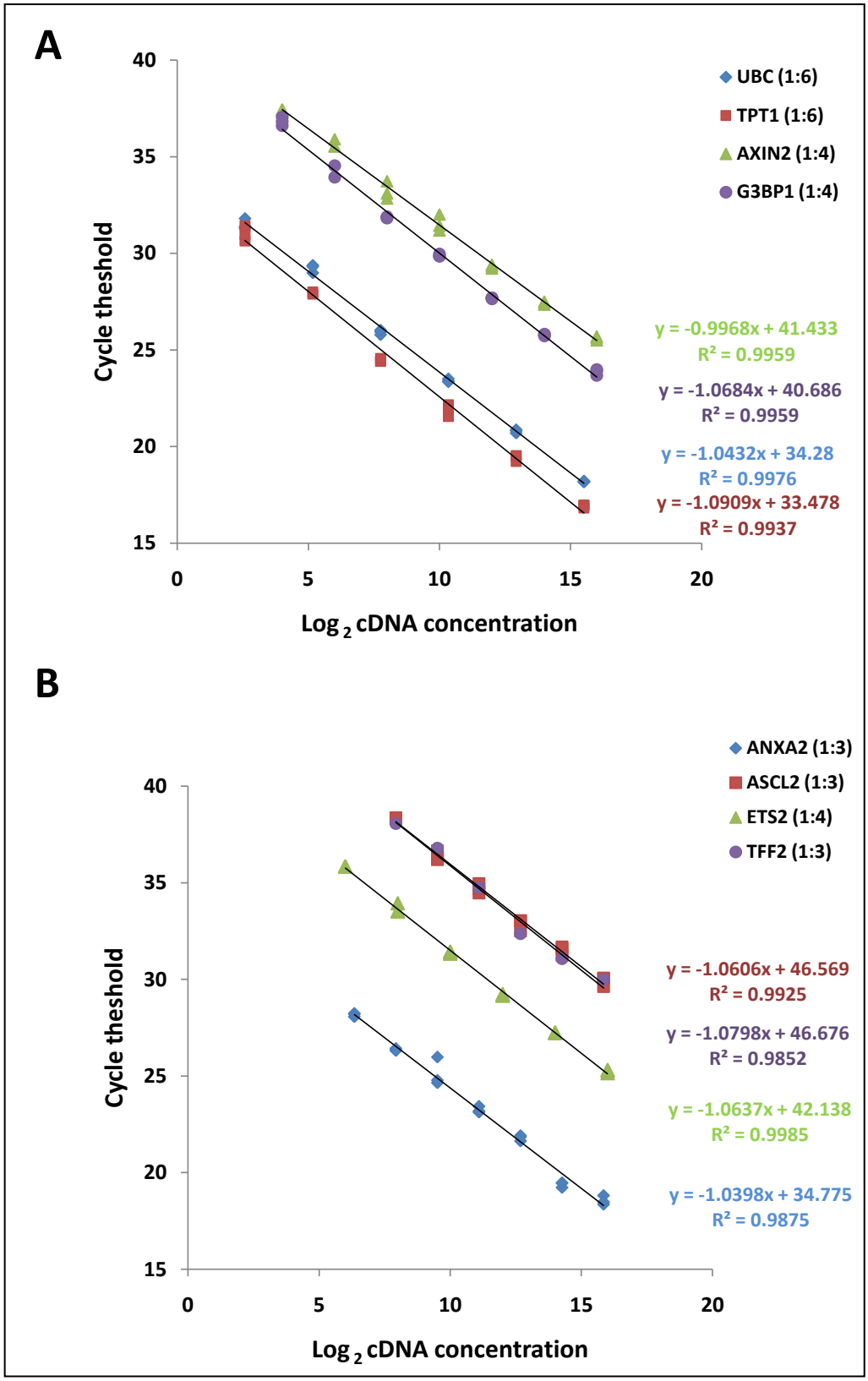

Figure 5.2: Standard curves of candidate and EC genes

Standard curves of the six candidate genes and EC genes. The graphs have been plotted as log2 of the cDNA concentration (x-axis) against the cycle threshold (y-axis). Four graphs have been plotted on each graph for ease of interpretation. The genes are colour-coded and show the slope, intercept and R^2 values in the trendline equation.

Additional validation was performed by repeating RNA isolation and cDNA synthesis in ten samples. For the majority of the replicated samples, there was high agreement between the average Cts for each gene. This suggests that the experiments were reproducible.

To further assess the reproducibility of the ten replicate cases, Bland-Altman tests were performed using the ten dCts obtained for each gene. dCts were created using *UBC,* the EC gene. The results revealed no significant differences between the replicates. The results are shown in Table 5.2.

Table 5.2: Bland-Altman test results for candidate gene qRT-PCR reproducibility

Gene	95% Limits of Agreement	Bias	p-value
TPT1	-2.81-2.86	0.026	0.96
ASCL2	-4.61-5.39	0.39	0.64
AXIN2	-2.68-2.23	-0.23	0.58
ANXA2	-0.91-1.04	0.068	0.67
ETS2	-1.71-1.46	-0.12	0.64
G3BP1	-2.10-1.48	-0.31	0.31
TFF2	-2.81-2.85	0.026	0.96

This table shows the 95% limits of agreement, Bias and p-value for the six candidate genes and the EC gene TPT1. The bias for each gene is close to zero, suggesting that the repeated cases are reproducible in all seven genes. This is confirmed by the lack of significant p-values, suggesting that the repeats are not significantly different.

In addition to the Bland-Altman test, paired t-tests were used to detect significant differences in the replicates across the 7 genes. The results of the paired t-tests were not significant for any of the genes.

Following analysis of the standard curves and replicate analysis, validation of the candidate genes was initiated using NC, HP, AP, KN and KP tissues.

5.5. CANDIDATE GENE EXPRESSION PROFILES IN COLORECTAL TISSUES

5.5.1. QRT-PCR EXPRESSION PROFILES OF THE CANDIDATE GENES IN TRAINING SET SAMPLES MIMIC THOSE SEEN IN THE MICROARRAY

15 of the 24 Training Set NCs (n=6), HPs (n=5) and APs (n=4) had sufficient material for qRT-PCR analysis. Statistical analysis was performed on the qRT-PCR results for the Training Set samples. Due to small sample sizes, the data could not be tested for normality and so non-parametric tests were performed. The results of the Training Set qRT-PCR support the results seen on the microarray, with the exception of *TFF2*. *TFF2* and *TPT1* do not show any differential gene expression (Table 5.3).

Table 5.3: Statistical analysis of Training Set qRT-PCR results of candidate gene expression

Gene	T-tests		
	NC-HP	NC-AP	HP-AP
TPT1	0.43	0.76	0.90
ASCL2	0.08	0.48	0.11
AXIN2	0.05	0.61	0.02
ANXA2	0.004	0.61	0.03
ETS2	0.08	0.91	0.41
G3BP1	0.03	0.02	0.19
TFF2	0.43	0.11	0.56

The results of the Mann-Whitney tests identify, *AXIN2*, *ANXA2*, *ETS2* and *G3BP1* as having differential gene expression across NC, HP and AP tissues. Non-parametric statistical tests were performed as, due to small sample sizes, normalcy could not be evaluated.

5.5.2. SAMPLE DISTRIBUTION OF THE VALIDATION SET

The different groups of colorectal tissues were tested for normalcy (Table 5.4). *AXIN2, ETS2* and *G3BP1* were normally distributed across the groups of colorectal tissue. The remaining genes were not normally distributed. The NC samples were normally distributed across all seven genes and represent the “disease-free” or “control” group in this study.

Table 5.4: Assessment of normal distribution of the candidate genes in different colorectal tissues

Gene	NC	HP	AP	KN	KP
TPT1	Yes	No	No	Yes	Yes
ANXA2	Yes	Yes	Yes	No	Yes
ASCL2	Yes	No	Yes	Yes	Yes
AXIN2	Yes	Yes	Yes	Yes	Yes
ETS2	Yes	Yes	Yes	Yes	Yes
G3BP1	Yes	Yes	Yes	Yes	Yes
TFF2	Yes	Yes	No	Yes	Yes

This table identifies which sample groups are normally distributed for the seven genes of interest. Three tests for normality were investigated using the methods outlined in Chapter 2. Most genes are normally distributed across the majority of the tissue samples and therefore parametric tests are appropriate unless both sample populations are not normally distributed. Under these circumstances, non-parametric tests would be used.

5.5.3. QRT-PCR CONFIRMS DIFFERENTIAL GENE EXPRESSION PROFILES OF CANDIDATE GENES IN NC, HP AND AP TISSUES

Following the qRT-PCR results of the Training Set samples, subsequent analysis of the candidate genes was undertaken using the larger Validation Set. NC (n=16), HP (n=40) and AP (n=60) tissues were analysed using qRT-PCR for changes in gene expression of the candidate and EC genes.

Expression of *TPT1* in the Validation Set contradicted the data presented by Andersen *et al* (2004), which claimed that *TPT1* was a suitable EC gene for the normalisation of qRT-PCR results from colorectal tissues. This “EC gene” showed significant differential expression between NCs and APs (p=0.0416), and HPs and APs (p=0.0017). *TPT1* was subsequently deemed unsuitable for use as an EC gene in this study. Future statistical analysis uses *UBC* as the only EC gene, with *TPT1* re-categorized as an additional candidate gene.

TFF2 did not show any significant differences in gene expression between NC, HP and AP tissues. The remaining candidate genes all showed highly significant gene expression changes across the three tissues (NC vs HP vs AP), especially when

comparing HPs and APs. The gene expression profiles of *ASCL2, AXIN2, ETS2* and *G3BP1* were similar, showing reduced expression in HPs compared to NCs and APs. With the exception of *G3BP1*, the highest expression of these genes was seen in APs. Conversely, *ANXA2* appeared to have highest levels of expression in HP tissues, followed by NCs and APs. These results highlight significant differences in gene expression between HPs and APs (Table 5.5).

Figure 5.3 shows relative difference in gene expression plotted as –dCts as more positive –dCts represent increased expression relative to the EC (*UBC*).

Table 5.5: Summary statistics for gene expression of candidate genes in NC, HP and AP tissues

Gene	Comparison	p-value
ASCL2	NC-HP-AP	<0.0001
	NC vs HP	**
	NC vs AP	NS
	HP vs AP	***
AXIN2	NC-HP-AP	<0.0001
	NC vs HP	**
	NC vs AP	NS
	HP vs AP	***
ANXA2	NC-HP-AP	<0.0001
	NC vs HP	*
	NC vs AP	NS
	HP vs AP	***
ETS2	NC-HP-AP	<0.0001
	NC vs HP	***
	NC vs AP	NS
	HP vs AP	***
G3BP1	NC-HP-AP	<0.0001
	NC vs HP	***
	NC vs AP	**
	HP vs AP	***
TFF2	NC-HP-AP	0.5578
	NC vs HP	NS
	NC vs AP	NS
	HP vs AP	NS
TPT1	NC-HP-AP	0.0011
	NC vs HP	NS
	NC vs AP	*
	HP vs AP	**

The summary statistics show that the difference in gene expression of the candidate genes between HPs and APs is highly significant. In addition, the difference in expression between NC and HPs is also significant for *ASCL2*, *ANXA2*, *AXIN2*, *ETS2* and *G3BP1*. The use of *TPT1* as an EC gene is not possible since it shows a statistically significant difference in gene expression between NCs and APs, and HPs and APs. Statistical tests: three-way comparison (e.g. NC-HP-AP) using one-way ANOVA with Bonferroni Correction to compare pairs of columns (e.g. NC-HP). *, p<0.05; **, p<0.01; ***, p<0.001; NS, not significant.

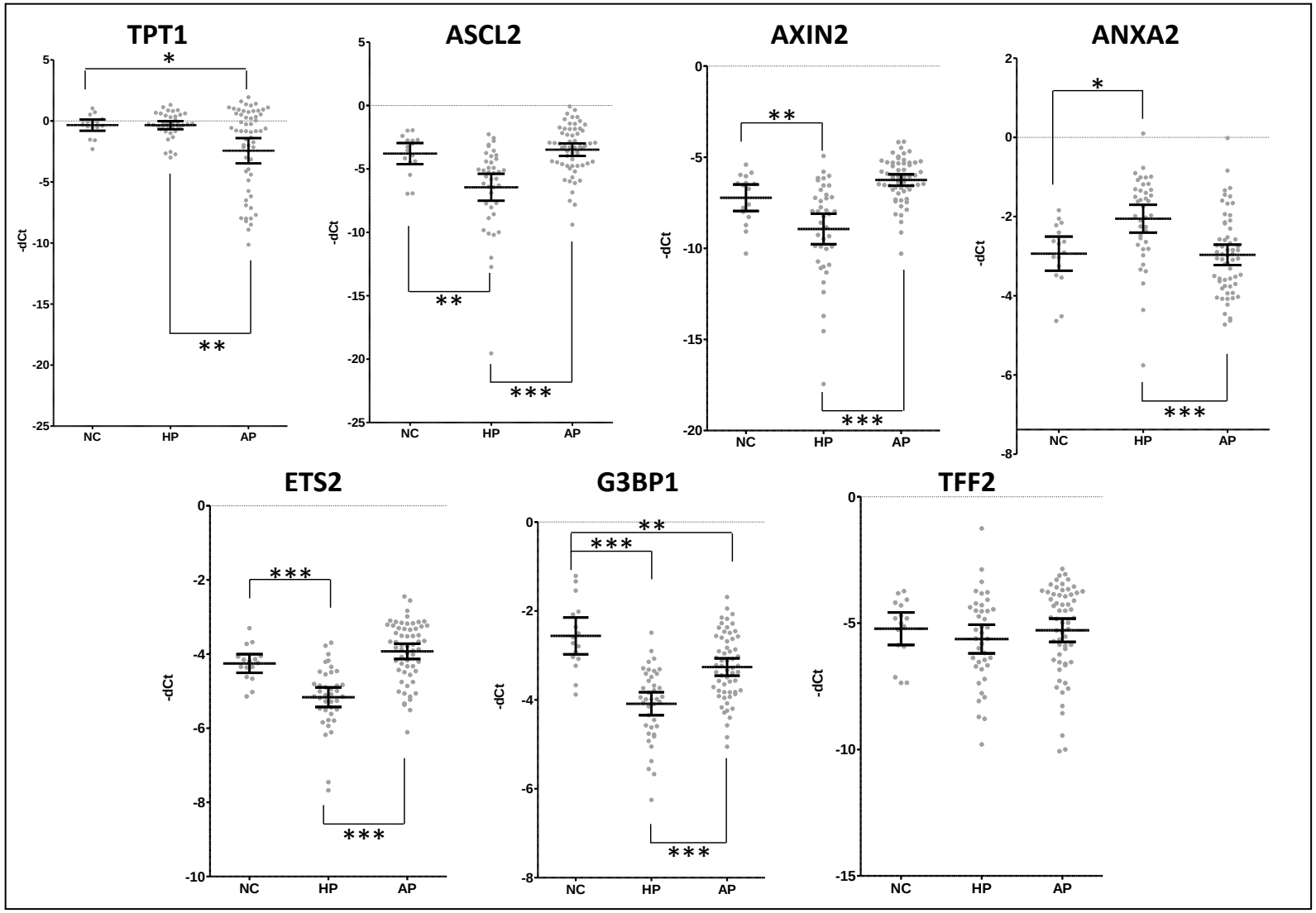

Figure 5.3: Gene expression profiles of the candidate genes in NC, HP and AP tissues

The gene expression profiles for candidate genes in NC, HP and AP tissues from the Validation Set are presented in this figure. The differential pattern of gene expression in these three tissues is noted for all genes, with the exception of *TFF2* which does not show any gene expression difference across the three tissue groups. The inverse expression pattern of HPs and APs is noted for most genes, highlighting the gene expression differences between these two tissue types.*≤0.05, **≤0.01, *≤0.001**

5.5.4. CANDIDATE GENE EXPRESSION PROFILES IN NORMAL COLORECTAL TISSUES

Candidate gene expression was compared in NC and KN samples, identifying significant differences in *ANXA2* (p=0.0232) and *G3BP1* (p=0.0295) expression (Figure 5.4 and Table 5.6).

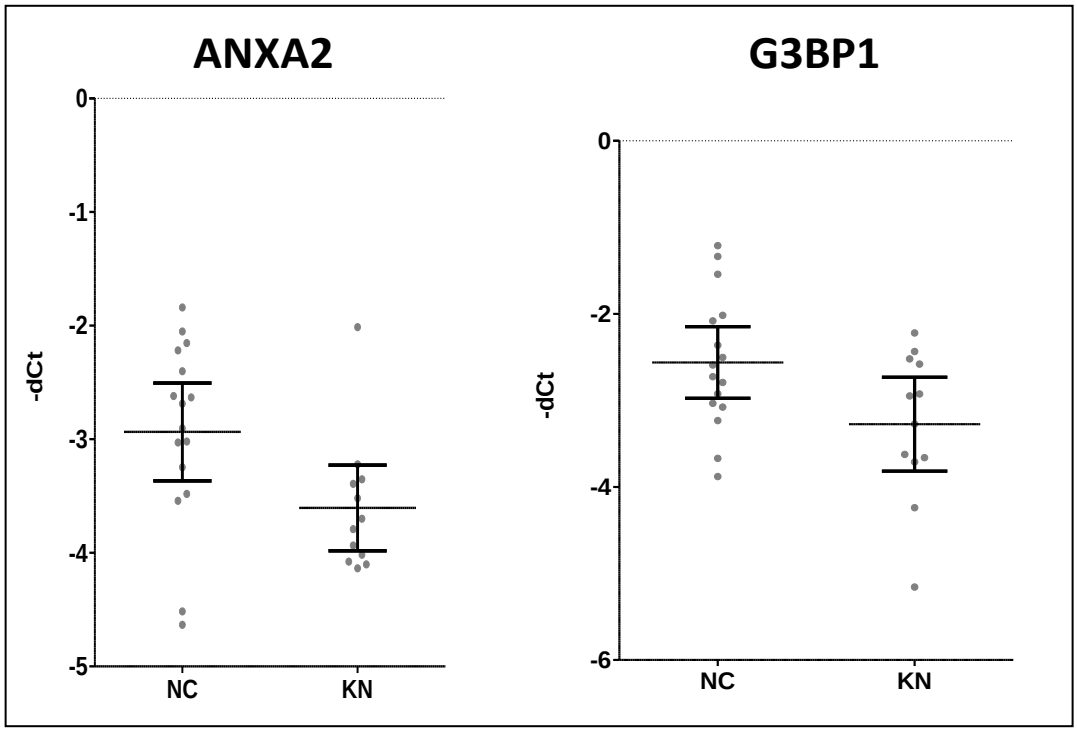

Figure 5.4: Gene expression profiles of *ANXA2* and *G3BP1* in normal colorectal tissues
***ANXA2* shows higher levels of expression in NC samples compared to cancer-associated normal colonic mucosa (KNs). The same pattern of expression is seen for *G3BP1*. No significant difference in gene expression was seen for any of the other candidate genes.**

Table 5.6: Unpaired t-tests comparing NC and KN tissues in the candidate genes

Gene	*p-value*
TPT1	*0.7320*
ASCL2	*0.4731*
ANXA2	*0.0232*
AXIN2	*0.4531*
ETS2	*0.4169*
G3BP1	*0.0295*
TFF2	*0.6567*

This table presents the unpaired t-test results when comparing different normal colorectal tissues (NC and KN), and the effect that these tissue types have on candidate gene expression. *ANXA2* and *G3BP1* are the only two genes showing significant changes in gene expression between different types of colorectal tissues.

5.5.5. GENE EXPRESSION PROFILES OF CANDIDATE GENES BETWEEN NON-CANCER AND CANCER ASSOCIATED COLORECTAL ADENOMAS

APs represent polyps from patients without CRC, while KPs are polyps that were obtained from patients with CRC. Differences in candidate gene expression between these two polyp categories could provide evidence for their association with cancer progression.

The expression of each candidate gene was compared in all AP and KP samples from the Validation Set. However, disappointingly no significant differences in gene expression were seen between the two polyp types for any of the candidate genes.

5.5.6. CANDIDATE GENE EXPRESSION PROFILES IN NORMAL AND POLYP SAMPLES FROM CRC PATIENTS

The Validation Set includes ten matched KN-KP cases. No significant differences in gene expression were observed between the paired KN-KP cases for any of the genes.

The KN-KP cases present in the Validation Set were then separated into anatomical locations (left vs right) to identify any changes in gene expression associated with this. However, no significant changes were observed for any of the candidate genes using paired analysis on the matched cases or unpaired analysis on all KNs and KPs.

5.5.7. CLINICO-PATHOLOGICAL VARIABLES AND THEIR EFFECT ON CANDIDATE GENE EXPRESSION IN COLORECTAL ADENOMAS

Next candidate gene expression in all colorectal adenomas (APs and KPs) was analysed with respect to histology, mutation status and location.

Adenoma histology was divided into TA (n=42), TV (n=17) and VA (n=11) subtypes. None of the candidate genes (including *TPT1*) showed any significant change in gene expression across the three adenoma types.

When performing the analysis with only AP samples (TA = 36, TV = 13, VA = 9) however, *TPT1* showed differential gene expression between TA and TV adenomas (p=0.0027) and TA and VA adenomas (**, p<0.01) (Figure 5.5). No other gene showed differential expression across different polyp types when using only AP samples.

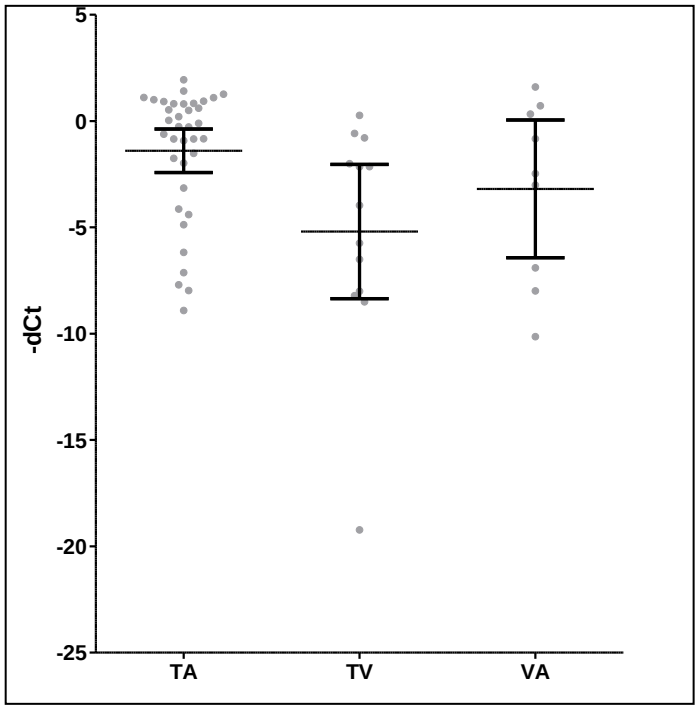

Figure 5.5: The effect of Adenoma type on gene expression of *TPT1*
TPT1 shows higher expression in TAs relative to the other types of adenomas. The –dCts have been plotted for ease of interpretation.

The mutation status of *BRAF* and *KRAS* was evaluated in all colorectal adenomas and related to gene expression data. 61 polyps (APs and KPs) from the Validation Set were analysed for mutations in *BRAF* and *KRAS*. *BRAF* mutations were found in 9.83% (6/61) of polyps (3 were serrated polyps, 2 were TAs and 1 was a TV). *KRAS* mutations were found in 13.11% (8/61) of polyps (3 were TAs with the remaining 5 having villous

architecture (TVs or VAs)). No significant difference was seen in gene expression of the candidate genes in relation to mutation status of *BRAF* or *KRAS*.

55 AP samples were also evaluated for their mutation status. *BRAF* mutations were found in 5.5% (3/55) of polyps, (2 were TAs and 1 was serrated). *KRAS* mutations were found in 10.9% (6/55) of polyps, (2 were TAs, 2 were TVs and 2 were VAs). Again no relationship was seen between mutation status and candidate gene expression.

Polyp location was also investigated to see if candidate gene expression profiles alter according to anatomical location. The location was broadly divided into left and right. Of the 75 available polyps, only 65 had reliable location data. Therefore, the 10 polyps of unknown location were not included in this analysis. No differences in candidate gene expression were seen between Left (n=44) and Right (n=21) sided polyps.

5.6. DISCUSSION

5.6.1. SELECTION OF ENDOGENOUS CONTROL GENES

The identification of suitable EC genes is essential for appropriate normalisation of gene expression data from qRT-PCR experiments. A universally stable gene has yet to be identified and so genes that are stably expressed across a subset of tissues are sought after for normalisation of gene expression data (Kheirelseid et al., 2010, Kubista et al., 2006).

The study by Andersen *et al* (2004) identified *TPT1* and *UBC* as the most stably expressed genes in colon cancer tissues. These results were supported by the NormFinder results presented in this chapter, which looked at gene expression stability across the range of different colorectal tissues present on the DASL

microarray. However the results of the qRT-PCR did not support the use of *TPT1* as an endogenous control and showed that expression of *TPT1* varied significantly between different colorectal tissues. For this reason, *TPT1* was re-classified as an additional candidate gene in this study.

5.6.2. RNA QUALITY IN TRAINING AND VALIDATION SETS

The quality of the RNA in each of the Training and Validation Sets was assessed using the EC genes *RPL13a* and *UBC*. The Training Set was primarily assessed using *RPL13a* and compared to the FFPE samples from the Fixation Set. Further comparisons were made between a subset of the Training Set and the Validation Set using *UBC*.

Differences observed in *RPL13a* expression between the Training and Fixative Sets were observed. The RNA obtained from the Training Set showed a significantly higher average Ct than the FFPE samples from the Fixative Set. These differences could be accounted for by the different length of time that the samples from the different data Sets were exposed to the fixative. As previously mentioned, increased exposure to formalin is associated with increased degradation of nucleic acids. In addition, samples from the Training Set were archived and stored for 2-3 years prior to RNA extraction, compared with 1-12 months for the Fixative Set samples. No significant difference was seen when comparing the different tissue types of the Training Set. Overall the quality of the RNA of the Training Set appeared to be adequate for further investigation, since a Ct<29 was achieved with *RPL13a* in the majority (46/48) of cases.

The comparison between a subset of the Training Set with the Validation Set using *UBC* revealed four samples that were identified as outliers. No difference was seen

between the two Sets. This suggests that further analyses can be performed using the Validation Set.

5.6.3. SELECTION OF CANDIDATE GENES AND VALIDATION USING QRT-PCR

Candidate genes were selected from the top-tables comparing NC, HP and AP samples since these comparisons revealed large numbers of significantly DEGs at both the raw and adjusted p-value levels (Table 4.2). Only one gene, *CARS* (Cysteinyl-tRNA synthetase) was differentially expressed across NC, HP and AP samples (Figure 4.12). *CARS* is thought to be involved with breast cancer (Beyer et al., 2011) however no evidence is available for its role in CRC. *CARS* was not selected as a candidate gene in this analysis as the differences in gene expression were more profound when comparing either HPs or APs to NC tissues. When comparing HPs with APs the differential expression of *CARS* was minimal, preventing the use of *CARS* as a discriminator between these tissue types.

From the top-table of up-regulated genes Table 4.3, *SDC3, REG4, LY6G6D, FGFBP1* and *SPINK4* showed significantly differential expression with large fold changes. However, these genes were not selected as candidate genes for further study. Firstly, *SDC3* and *SPINK4* were unable to differentiate between HP and AP samples based on the microarray. Secondly, *FGFBP1* and *REG4* have previously been described in the literature as having a role in colorectal carcinogenesis. *FGFBP1* is thought to act as an angiogenic switch, providing a growth advantage to tumour cells (Tassi et al., 2006, Tassi and Wellstein, 2006). *REG4* is thought to be an early marker of CRC (Lu et al., 2007), as well as an important predictor of CRC prognosis (Li et al., 2010, Numata et al., 2011). *LY6G6D* represents a relatively unknown gene, with only 7 published

articles available relating to its function. It is thought to belong to a cluster of leukocyte antigen-6 (LY6) genes. Since other genes identified in the top-tables appear to have potential roles in adenoma/carcinogenesis, this gene was not investigated further. However, this gene is of interest and future studies should investigate its role in CRC.

The genes selected (*ASCL2, ANXA2, AXIN2, ETS2, G3BP1* and *TFF2*) all showed differences in gene expression in different colorectal tissues, potentially during both adenomagenesis and carcinogenesis.

The use of eight multiplexed reverse transcription primers enabled limiting patient material to be analysed. Detailed information about each of the candidate genes is presented in the following sections.

5.6.4. TUMOUR PROTEIN, TRANSLATIONALLY CONTROLLED (TPT1)

TPT1 is a multifunctional protein sometimes referred to as *TCTP* (Translationally Controlled Tumour Protein). *TPT1* is thought to be important in cell cycle progression, cell growth and division, apoptosis, calcium binding and malignant transformation. (Bommer et al., 2010, Bommer and Thiele, 2004, Telerman and Amson, 2009, Gnanasekar et al., 2009). In addition *TPT1* is thought to have cytoprotective functions (Bommer and Thiele, 2004).

TPT1 is an anti-apoptotic agent which shows higher levels of expression in cancer cells compared to adjacent normal colorectal tissue (Bommer et al., 2010, Bommer and Thiele, 2004). This pattern of gene expression has been observed in several solid cancers including colon (Ma et al., 2010), contradicting its use as an endogenous

control in colon cancer as suggested by Andersen *et al* (2004). Although *TPT1* expression tends to be elevated in cancer tissues, it is not a tumour specific gene with expression levels varying according to cell or tissue type and external stimuli (Tuynder et al., 2002). Specifically, the expression of *TPT1* is elevated in tissues that are mitotically active, such as the colon, compared to postmitotic tissues (Bommer and Thiele, 2004, Thiele et al., 2000).

TPT1 is down-regulated by the p53 signalling pathway, which is often deregulated in CRC (Bommer and Thiele, 2004). The anti-apoptotic role of *TPT1* provides a growth advantage to cancer cells, which when coupled with its disregulation by the p53 signalling pathway explains its elevated expression in these cells.

Evidence for the role of *TPT1* involvement with tumorigenesis has been found in studies looking to reverse cells from the malignant phenotype, resulting in a reduction in the expression of *TPT1* as the cells are reverted (Tuynder et al., 2002). Additionally, inhibition of *TPT1* has been shown to suppress the malignant phenotype in mice (Tuynder et al., 2002) and LoVo CRC cell lines (Ma et al., 2010). The findings of these studies support the re-allocation of *TPT1* as an additional candidate gene.

5.6.5. ACHAETE-SCUTE COMPLEX HOMOLOG 2 (*ASCL2*)

ASCL2 is a transcription factor responsible for the differentiation of the trophoblast lineage in normal human placenta (Guillemot et al., 1994). In addition *ASCL2* is also a known target of the Wnt Signalling Pathway (Stange et al., 2010, Jubb et al., 2006) and a stem cell maintainer. Located at the Chromosome 11p15.5 region, which is often gained in CRC, *ASCL2* has been shown to be up-regulated in CRC liver metastases with 11p15.5 gain (Stange et al., 2010).

ASCL2 is over-expressed in CRCs and throughout all stages of CRC development. This over-expression is thought to be due in part to the disregulation of the Wnt Signalling Pathway in CRC (Jubb et al., 2006). Increased expression of *ASCL2* was observed in AP samples, supporting this finding.

5.6.6. ANNEXIN 2 (*ANXA2*)

ANXA2 is a member of the Annexin family which has a number of physiological roles such as membrane trafficking, calcium signalling, regulation of ion channels, cell motility, DNA synthesis, proliferation and differentiation (Filipenko et al., 2004, Guzman-Aranguez et al., 2005, Inokuchi et al., 2009). No member of the Annexin family is universally expressed in all tissues. Instead, individual annexin gene expression profiles vary across tissues providing "annexin fingerprints" for each tissue type (Gerke, 2001). This phenomenon suggests that the gene expression of annexins is under tight regulation. Annexins have been linked to solid cancers, including prostate, lung, hepatocellular, and colorectal (Filipenko et al., 2004, Gerke, 2001).

ANXA2 gene expression is growth regulated, showing increased expression under the influence of growth factors (Filipenko et al., 2004, Guzman-Aranguez et al., 2005) and is often differentially expressed during cellular proliferation and differentiation (Guzman-Aranguez et al., 2005). Interestingly, *ANXA2* expression is induced by hypoxia (Gerke, 2001) which frequently occurs during tumour growth. In studies investigating ANXA2 protein expression in CRC, most studies have found that ANXA2 expression appears to be increased in CRC (Emoto et al., 2001), especially in more advanced tumours, and is thus a potential marker of aggressive phenotypes (Duncan et al., 2008,

Chiang et al., 1999, Singh, 2007). Conversely, a study by Wu *et al* (2010) suggests that the ANXA2 protein is actually down-regulated in CRC (Wu et al., 2010).

These findings suggest a possible role for *ANXA2* in the pathogenesis of CRC. Unfortunately, these study findings are limited to protein expression and so the changes in *ANXA2* gene expression can only be assumed to reflect the changes in protein expression. No studies are available to show the gene expression of *ANXA2* in different colorectal tissues or early colorectal lesions and therefore further studies are required to fully understand the expression of *ANXA2* in colorectal tissues.

5.6.7. AXIN 2 (*AXIN2*)

AXIN2 is a negative regulator of the Wnt Signalling Pathway and a tumour suppressor gene (Salahshor and Woodgett, 2005, Hughes and Brady, 2005, Hughes and Brady, 2006). In addition, *AXIN2* is important in controlling the levels of β-catenin, limiting its accumulation and thus preventing its activation of target genes. The expression of the *AXIN2* gene and protein are altered in CRC.

Studies have suggested that up-regulation of *AXIN2* is associated with CIN positive CRCs (Hadjihannas et al., 2006), and increased apoptosis (Salahshor and Woodgett, 2005). Interestingly, it has also been suggested by Hughes *et al* (2006) that an increased level of the AXIN2 protein is sufficient to cause CIN. Conversely the expression of *AXIN2* is down-regulated in MSI positive CRCs, which is thought to be a result of epigenetic silencing of the *AXIN2* gene (Koinuma et al., 2006).

Mutations of *AXIN2* also result in its down-regulation. Down-regulation of *AXIN2* results in the accumulation of β-catenin since it is not sufficiently degraded (Takahashi

et al., 2005). Over-expression of β-catenin potentiates its function as an oncogene, contributing to tumorigenesis.

Hughes *et al* (2006) showed that the expression of *AXIN2* was higher in CRCs than the corresponding normal tissues in CRCs with activating mutations of the Wnt/ β-catenin pathway. The same finding was noted in CRC cell lines (Yan et al., 2001). The expression of the AXIN2 protein however can be elevated in normal colonic cells to maintain low levels of β-catenin (Anderson et al., 2002).

5.6.8. V-ETS ERYTHROBLASTOSIS VIRUS E26 ONCOGENE HOMOLOG 2 (*ETS2*)

ETS2 is a transcription regulator (Ito et al., 2002) of genes involved with cellular proliferation, differentiation, development, transformation and apoptosis (Seth and Watson, 2005). In addition, *ETS2* is a target of *ASCL2* having a similar expression profile (Stange et al., 2010) as seen in the results presented within this chapter.

5.6.9. GTPASE ACTIVATING PROTEIN (SH3 DOMAIN) BINDING PROTEIN 1 (*G3BP1*)

G3BP1 encodes a protein that is thought to participate in signal pathways involved with cellular growth, proliferation, differentiation and apoptosis (Zhang and Shao, 2010, Guitard et al., 2001). In addition, *G3BP1* has been implicated in Ras signalling (Zhang and Shao, 2010, Kim et al., 2007) and p53 regulation (Kim et al., 2007). Elevated expression of the protein is seen in several malignancies, including metastases (Liu et al., 2001) and represents a target for cancer therapy (Zhang and Shao, 2010).

Periodic changes in gene expression of *G3BP1* were identified in a microarray analysis of cell-cycle regulated genes indicating that this gene is a cell cycle regulated transcript (Irvine et al., 2004). Irvine *et al* (2004) also suggested that G3BP1 might function as a growth factor sensor, allowing cell cycle regulators such as *c-myc* to accumulate in stimulated cells, but be degraded in resting cells. The exact mechanism of action of *G3BP1* in cancer development and progression is not fully understood but its involvement with the Ras signalling system, p53 regulation and *c-myc* regulation provides strong evidence for its involvement in tumorigenesis.

5.6.10. TREFOIL FACTOR 2 (*TFF2*)

Literature analysis of *TFF2* suggests that it has multiple functions, which include decreasing cell proliferation, anti-apoptosis and angiogenesis (Kim et al., 2011a). The expression of *TFF2* is thought to be predominantly in the gastric mucosa, whereas other family members (*TFF2* and *TFF3*) are also expressed in the colorectal mucosa (Perry et al., 2008). CRC cell line studies have indicated that *TFF2* has an anti-apoptotic function and that its inhibition could be a potential therapeutic intervention to reduce cancer size (Siu et al., 2004).

TFF2 was selected for validation in this study to determine the significance of the B statistic in the selection of genes that show true altered gene expression profiles. The findings of the qRT-PCR analysis did not confirm the results of the microarray, revealing no significant gene expression differences between any of the colorectal tissues. However, *TFF2* has been identified as a potential therapeutic target in CRC suggesting that it does have a role in the development of CRC, but perhaps in more advanced stages of the disease. On balance, these findings suggest that the B statistic

is important in the selection of valid candidate genes as described by Soon *et al* (2009), Harina *et al* (2010) and Nguyen *et al* (2009). However, adjusted p-value and fold change should also be considered in the selection of candidate genes.

5.6.11. CANDIDATE GENE EXPRESSION PROFILES IN COLORECTAL TISSUES

5.6.11.1. DIFFERENCES IN CANDIDATE GENE EXPRESSION ACROSS NC, HP AND AP TISSUES

The expression profile of the candidate genes seen on the microarray was confirmed using qRT-PCR on available Training Set samples. All genes, with the exception of *TFF2*, showed similar gene expression profiles to those seen on the microarray. This suggests that the gene expression changes seen on the array are valid. The limited number of samples available from the Training Set for qRT-PCR analysis could explain the lack of significant changes in *TFF2* expression. The findings of the Training Set qRT-PCR analysis provided sufficient incentive to continue the validation of the candidate genes in the larger Validation Set.

Within the Validation Set, each candidate gene (with the exception of *TFF2*) was shown to have differential expression across NC, HP and AP tissues. Of particular interest, significant differences were observed between HP and AP tissues. This interesting finding highlights the differential expression seen between these two tissues. Since HPs are considered largely non-neoplastic (Bond, 2000) and APs are considered as pre-cursors to CRC, differences in gene expression could help identify candidate genes for therapeutic interventions.

The candidate genes appear to play a role in adenomagenesis. *ASCL2*, *AXIN2*, *ETS2* and *G3BP1* all show a similar profile in terms of HP and AP expression, with APs showing

higher levels of expression than HPs. Conversely, *ANXA2* and *TPT1* both show higher levels in HPs than APs.

It has been suggested that HPs represent lesions expressing senescent markers (Dhomen et al., 2010). These lesions are thought to have higher frequencies of *BRAF* mutations (Yang et al., 2004, Haferkamp et al., 2009), associated with oncogene-induced cell senescence. The role of the Wnt-pathway in serrated lesions remains conflicting. The balance between the Wnt signalling pathway and cellular senescence is important in tumorigenesis since they have roles in the promotion or prevention of carcinogenesis respectively. *AXIN2*, *ASCL2* and *G3BP1* are all associated with the Wnt-signalling pathway as previously mentioned. It is possible that the Wnt signalling pathway is repressed in HPs, resulting in the reduced expression of these genes.

ETS2 is a direct target of *ASCL2* (Stange et al., 2010), which could explain its expression profile in HPs and APs. As with *ASCL2, ETS2* shows a higher level of expression in APs than HPs. Low expression of *ETS2* in HPs has been identified by other studies (Ito et al., 2002, Seth and Watson, 2005, Flavin et al., 2011).

The findings of the current study show that *ANXA2* gene expression is higher in HPs and lower in NCs and APs. Although this expression profile appears to contradict the findings of most of the literature, it agrees with the study by Wu *et al* (2010), which suggests that *ANXA2* is down-regulated in CRC. It is possible that the expression of *ANXA2* alters throughout the carcinogenesis process, showing elevated expression only in aggressive or advanced tumours. The elevated expression of *ANXA2* in aggressive tumours could be stimulated by tumour associated hypoxia. This hypothesis is based on the finding that in prostate cancer, ANXA2 protein expression is

usually low, however elevation in expression is seen in aggressive phenotypes (Inokuchi et al., 2009).

The results presented in this chapter show reduced expression of *TPT1* in APs compared to NC and HP tissues. This could represent a contradiction to the findings in the literature which show increased expression of *TPT1* in cancers. However, since mutations/loss of *p53* is a later event in the adenoma-carcinoma sequence, perhaps *TPT1* expression is under suppression by p53 during the early stages of cancer development, with elevated expression occurring after loss of *p53*. Additionally, since *TPT1* expression is elevated in tissues that are mitotically active, this may suggest a reason for the higher levels of *TPT1* expression seen in NCs since the colonic mucosa has a high rate of turnover. Currently, no evidence is available on the expression of *TPT1* in early colonic lesions and so further research is required to show the true role of *TPT1* in these lesions.

The candidate gene expression profiles identified in the Validation Set support those seen in the microarray data with the exception of *TFF2*. Overall, these genes show differential expression between HPs and APs suggesting a potential role in adenomagenesis.

5.6.11.2. GENE EXPRESSION CHANGES IN NORMAL COLORECTAL TISSUES

The differential gene expression patterns seen in different types of normal tissues provide insight into the gene expression changes that occur in the early stages of CRC development. When comparing NC and KN tissues, *ANXA2* and *G3BP1* showed lower levels of expression in the KN tissues. The level of expression was similar to that seen in KPs, suggesting that a potential field effect change could be responsible.

Further experiments are required to fully investigate the role of *ANXA2* and *G3BP1* in the development of adenomas and CRC.

5.6.11.3. GENE EXPRESSION CHANGES BETWEEN COLORECTAL POLYPS

When comparing the gene expression of the candidate genes in cancer and non-cancer associated adenomas, no significant differences were identified. This could be due to several reasons. Firstly, only a small number of KPs were investigated when compared to the larger AP cohort. Secondly, although KPs are associated with CRC, they are not actually cancerous tissues. Alternatively, this finding could indicate that the gene expression profile of the candidate genes is altered in earlier stages of CRC, such as during adenomagenesis (NC to AP transition).

In order to fully ascertain the gene expression profile of the candidate genes in these tissues, further KP samples are needed in the analysis.

5.6.11.4. NO SIGNIFICANT DIFFERENCES IN CANDIDATE GENE EXPRESSION WERE FOUND BETWEEN MATCHED SAMPLES

Analysis of paired and unpaired KN and KP samples did not reveal any changes in gene expression of any of the candidate genes. This lack of differential expression could suggest a field effect change in the cancer-associated tissues since for each candidate gene, the expression in KNs and KPs was similar.

5.6.11.5. CLINICO-PATHOLOGICAL VARIABLES AND THEIR EFFECT ON CANDIDATE GENE EXPRESSION

Clinico-pathological variables such as polyp location, type and mutation status are all easily identified within a clinical environment. Gene expression changes associated with these clinical factors can help with the prediction of prognosis and therapeutic

interventions. For example, the presence of *KRAS* mutation is predictive of a poor response to the cetuximab therapy in CRC. These variables were compared to changes in gene expression of the candidate genes in the Validation Set. No significant changes in gene expression occurred in relationship to mutation status or polyp location. Only *TPT1* showed changes in gene expression according to adenoma classification, stable expression was seen in the other candidate genes.

Since the majority of APs present in the microarray are TAs, the results obtained for polyp type and mutation status may be inaccurate. TAs tend to be smaller, with fewer mutations than the larger VAs. Overall, the mutation frequencies were; *BRAF* (9.8%), *KRAS* (13.1%). The literature describes a frequency of *BRAF* mutations between 2.8% and 4.8%, slightly lower than those seen in this study (Yuen et al., 2002, Jass et al., 2006). The frequency of *KRAS* mutations shows more variation, ranging from 3% to 78% (Chen et al., 2011, Barry et al., 2006).

The majority of *BRAF* mutated samples are serrated adenomas associated with CRC. The serrated pathway of CRC has been associated with increased frequencies of *BRAF* mutations, which is supported by the present study (Kambara et al., 2004, Leggett and Whitehall, 2010).

Barry *et al* (2006) describes the reduced mutation rate of *KRAS* in APs which are smaller than 10mm. They find a similar frequency of mutations (9%) in small adenomas to the one described in the present study (13.1%). In addition, the finding that a villous architecture is associated with a higher mutation rate is confirmed by other studies (Chen et al., 2011, Barry et al., 2006, Einspahr et al., 2006, Jass et al., 2006).

Perhaps the lower frequency of *KRAS* mutations presented in this chapter is due in part to the smaller sized adenomas investigated, with the majority being <10mm in diameter, and the limited number of polyps showing a villous architecture. The smaller adenomas with *KRAS* mutations are likely to represent adenomas with a more aggressive phenotype.

5.6.12. WEAKNESSES AND LIMITATIONS

Unfortunately not all of the Training Set samples were available for qRT-PCR validation with the candidate genes. Although this is disappointing, the results obtained still confirmed the results of the microarray. Additionally, since the number of Training Set samples were limited, these samples alone cannot be responsible for the changes seen in the Validation Set. This suggests that the observed changes in gene expression were true.

With regards to the mutation analysis, it would have been preferable to use equal numbers of TAs, TVAs and VAs. TAs dominated in this study due to their relative abundance compared to TVAs and VAs. More TVAs and VAs should have been included in the experiment however there was difficulty in finding these tissues. Future studies should concentrate on screening larger polyp populations, with equal numbers of samples in each group. In addition, the size of the polyps should be noted as this can also affect mutation frequency. This study predominantly used smaller TAs, which are associated with fewer mutations. This could be the reason for the lower *KRAS* mutation rates seen.

A final limitation of this study is the lack of mutation analysis in the HP samples, which would have allowed *BRAF* and *KRAS* mutations to be identified. As already mentioned,

HPs are thought to represent senescent lesions with high frequencies of *BRAF* mutations. The presence of *BRAF* mutations has been associated with oncogene induced senescence. Knowing the mutation status in this group of colorectal tissue could help to explain the low levels of *ASCL2*, *AXIN2*, *ETS2* and *G3BP1* in HPs.

5.7. CONCLUSIONS

The aim of this chapter was to use microarray to identify candidate genes, which could be important biomarkers in the early stages of CRC development. In particular a marked difference in gene expression between HPs and APs was noticed, by both the top-tables of DEGs and the qRT-PCR validation of selected candidate genes. The differences between these "benign" and "pre-malignant" lesions could provide further insight into potential therapeutic interventions.

In summary, the objectives of this chapter have been fulfilled. Future work should focus on increasing the number of NC, KN and KP samples in the Validation Set and determining the mutation status of *BRAF* and *KRAS* in HPs would be useful. Specifically, the differences in gene expression between HPs and APs should be pursued to identify pathway disturbances and differences between the two different groups.

CHAPTER 6: DISCUSSION

6.1. DISCUSSION

Colorectal polyps are a well described colonic lesion with the potential to progress to CRC (Risio, 2010, Markowitz and Winawer, 1997). Whether colorectal polyps arise due to sporadic somatic mutation or as a result of an inherited mutation (*APC),* there is considerable overlap in the molecular and genetic events that take place (Fearon and Vogelstein, 1990, Jass, 2007). To interrogate this further, this thesis investigated differential gene expression between colorectal tissues arising from early stages of CRC development using DASL microarrays and qRT-PCR methods.

6.1.1. COMPARATIVE STUDY OF FIXATION METHODS ON RECOVERY OF NUCLEIC ACIDS FROM COLORECTAL TISSUES

One objective of this thesis was to establish the optimal method of tissue fixation to preserve nucleic acid quality and integrity, whilst maintaining tissue architecture. Three methods were compared (Formalin Fixed Paraffin Embedded (FFPE), Universal Molecular Fixative (UMFIX) and Fresh Frozen (FF)) for preservation of DNA and RNA.

Tissues fixed using FF methods best preserved both DNA and RNA, producing lower average Cts by qPCR and qRT-PCR respectively. However, the tissue architecture of FF tissues was less desirable, showing poorer staining and artefacts of the fixation process. In addition, it is unlikely that FF protocols will be widely used in the clinical setting due to the storage requirements.

Taking into account the preservation of nucleic acids and desired maintenance of tissue architecture for later histological analysis, UMFIX treated tissues appeared to offer the next best method of tissue fixation. DNA analysis by qPCR suggested that UMFIX and FF tissues were comparable, with the lowest average Ct values. Although

the findings for RNA were not as marked, UMFIX tissues also showed a lower average Ct than FFPE treated tissues in both combined and individual data sets. In addition, UMFIX has the advantage of maintaining tissue architecture suitable for histological interpretation, while also being less toxic than formalin.

One major shortcoming of this work is that only eight different cases were compared by the three fixation methods, hence results can only be considered preliminary at this stage. Future studies should therefore increase the size of the cohort of tissue samples investigated in order to either confirm or refute the potential of UMFIX, and protein analysis should also be undertaken, using techniques such as immunohistochemistry and western blotting to ascertain the effectiveness of the three fixation methods on preserving protein integrity.

6.1.2. DASL WHOLE-GENOME EXPRESSION ANALYSIS IN COLORECTAL TISSUES

48 colorectal tissues were included on the DASL Whole-Genome Microarray to investigate differential expression between colorectal tissues from early stages of the adenoma-carcinoma sequence. Differences in gene expression were seen for many of the comparisons made, however the most marked differences noted were between HPs and APs.

APs develop 5-10 years before malignant transformation, and therefore provide a potential therapeutic window enabling early intervention (Leslie et al., 2002, Scholefield, 2000). HPs represent non-neoplastic lesions (Winawer et al., 1990, Bensen et al., 1999, Bond, 2000, Bauer and Papaconstantinou, 2008, Lau and Sung, 2010, Risio, 2010). The differential gene expression between these two types of polyp is of key

interest due to the difference in their malignant potential, as this could enable potential therapeutic targets to be identified, which could be used to prevent AP progression.

Over thirty pathways showed differential expression between these HPs and APs. Of particular interest was differential expression in the Wnt Signalling pathway, MAPK signalling Pathway, p53 Signalling pathway, cell cycle and apoptosis, all of which are important in the development of CRC (Wang et al., 2004, Yen et al., 1998, Fearon and Vogelstein, 1990, Fearnhead et al., 2001, Baker et al., 1990). Differential expression of genes associated with these pathways could provide insight into the fundamental differences relating to the malignant potential of HPs and APs.

In addition, with the use of COXPRESSdb and top-table analysis, a novel network was created. This network showed how significant genes identified in the top-table are potentially related, suggesting that alternative pathways and gene interactions are important in the development of polyps. Unfortunately, a limitation of COXPRESSdb is the inability to determine how genes are related to one another. Further work is required to establish if genes associated with this network are important in the development of HPs and APs, and whether they are potential therapeutic targets.

In hindsight, it would have been useful to include SSA/Ps in the microarray design to identify differences between HPs and SSA/Ps. This could have provided further insight into differences in gene expression between benign and potentially malignant polyps, and potentially reduced the number of candidate genes identified across the three groups (HPs, SSA/Ps and APs).

Several other pathways identified in other comparisons, may also have a role in colorectal carcinogenesis including the calcium signalling pathway and PPAR signalling pathway. These pathways showed differential expression in several comparisons, suggesting that their regulation is altered during neoplastic progression. Other studies have identified genes associated with these pathways as being differentially expressed in cancers, supporting the finding of this thesis (Rey et al., 2010, Peters et al., 2004, Pancione et al., 2010, Krupp et al., 2011).

In the future, the differential gene expression between HPs and APs should be investigated further using larger cohorts to discover their importance in CRC development in the hope that therapeutic targets can be identified. Numerous genes (*ASCL2, ANXA2, AXIN2, ETS2, G3BP1, FZD9, PPARG, PITX2, FABP1, DUSP7,* and *CEACAM7)* and pathways (Wnt Signalling, MAPK Signalling, p53 Signalling, Apoptosis, the Cell Cycle, Calcium signalling and the PPAR Signalling pathway) of interest were identified that can be targeted in the future. Additional microarray, qRT-PCR and protein expression studies can be undertaken to help fully profile the role of candidates in the development of CRC.

6.1.3. VALIDATION OF CANDIDATE GENE EXPRESSION IN COLORECTAL TISSUES

The final chapter of this thesis aimed to validate a panel of 7 candidate genes (*ANXA2, ASCL2, AXIN2, ETS2, G3BP1, TFF2* and *TPT1*) identified by DASL microarray experiment on a range of colorectal tissues. Of the seven candidate genes investigated, 6 (*ANXA2, ASCL2, AXIN2, ETS2, G3BP1* and *TPT1*) showed significant differential expression across a range of colorectal tissues (NC, HP, AP, KN and KP). In addition the 7 genes are

associated with cellular processes such as proliferation, differentiation and apoptosis and *ASCL2, AXIN2* and *G3BP1* are also associated with Wnt Signalling (Jubb et al., 2006, Stange et al., 2010, Hughes and Brady, 2005, Hughes and Brady, 2006, Salahshor and Woodgett, 2005, Kim et al., 2007, Zhang and Shao, 2010) and *G3BP1* with p53 regulation (Kim et al., 2007).

Clinico-pathological parameters were also investigated for association with candidate gene expression. *TPT1* showed changes in gene expression according to adenoma type, however no other genes were associated with any other parameters. Since no difference was seen between adenoma type and expression of the remaining candidate genes, these candidate genes could represent early changes responsible for the initiation of polyp formation rather than the differentiation and progression. Alternatively, the real difference may be masked due to small sample numbers. Some of the comparisons suffered due to small sample numbers (e.g. KN and KP), and it would therefore be beneficial to include additional cases in the Validation Set.

In future it would be useful to determine *APC* mutations in these lesions and relate this to gene expression changes. These mutations (*APC, BRAF* and *KRAS*) should also be investigated in HPs, as previous studies have suggested that HPs have high frequencies of these mutations (Chan et al., 2003) despite having no malignant potential.

6.1.4. CONCLUSION

The results presented in this thesis describe differences in gene expression of a number of colorectal tissues associated with the early stages of CRC development. Differences between hyperplastic and adenomatous polyps were identified, suggesting fundamental differences in genes associated with cancer-associated pathways such as

Wnt-signalling, MAPK signalling, p53 signalling, apoptosis and the cell cycle. The results need validating in a more extensive series before differences such as those identified between HPs and APs, which have different neoplastic potential, can be targeted for screening and therapeutic intervention.

REFERENCES

AKIL, H., PERRAUD, A., MELIN, C., JAUBERTEAU, M. O. & MATHONNET, M. 2011. Fine-tuning roles of endogenous brain-derived neurotrophic factor, TrkB and sortilin in colorectal cancer cell survival. *PLoS One,* 6**,** e25097.

AMOS-LANDGRAF, J. M., KWONG, L. N., KENDZIORSKI, C. M., REICHELDERFER, M., TORREALBA, J., WEICHERT, J., HAAG, J. D., CHEN, K. S., WALLER, J. L., GOULD, M. N. & DOVE, W. F. 2007. A target-selected Apc-mutant rat kindred enhances the modeling of familial human colon cancer. *Proc Natl Acad Sci U S A,* 104**,** 4036-41.

ANDERSEN, C. L., JENSEN, J. L. & ORNTOFT, T. F. 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res,* 64**,** 5245-50.

ANDERSON, C. B., NEUFELD, K. L. & WHITE, R. L. 2002. Subcellular distribution of Wnt pathway proteins in normal and neoplastic colon. *Proc Natl Acad Sci U S A,* 99**,** 8683-8.

BAEK, S. H., KIOUSSI, C., BRIATA, P., WANG, D., NGUYEN, H. D., OHGI, K. A., GLASS, C. K., WYNSHAW-BORIS, A., ROSE, D. W. & ROSENFELD, M. G. 2003. Regulated subset of G1 growth-control genes in response to derepression by the Wnt pathway. *Proc Natl Acad Sci U S A,* 100**,** 3245-50.

BAKER, S. J., PREISINGER, A. C., JESSUP, J. M., PARASKEVA, C., MARKOWITZ, S., WILLSON, J. K., HAMILTON, S. & VOGELSTEIN, B. 1990. p53 gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis. *Cancer Res,* 50**,** 7717-22.

BALDUS, S. E., SCHAEFER, K. L., ENGERS, R., HARTLEB, D., STOECKLEIN, N. H. & GABBERT, H. E. 2010. Prevalence and heterogeneity of KRAS, BRAF, and PIK3CA mutations in primary colorectal adenocarcinomas and their corresponding metastases. *Clin Cancer Res,* 16**,** 790-9.

BALLINGER, A. B. & ANGGIANSAH, C. 2007. Colorectal cancer. *BMJ,* 335**,** 715-8.

BAN, J. O., KWAK, D. H., OH, J. H., PARK, E. J., CHO, M. C., SONG, H. S., SONG, M. J., HAN, S. B., MOON, D. C., KANG, K. W. & HONG, J. T. 2010. Suppression of NF-kappaB and GSK-3beta is involved in colon cancer cell growth inhibition by the PPAR agonist troglitazone. *Chem Biol Interact,* 188**,** 75-85.

BARR, A. J. & KNAPP, S. 2006. MAPK-specific tyrosine phosphatases: new targets for drug discovery? *Trends Pharmacol Sci,* 27**,** 525-30.

BARRY, E. L., BARON, J. A., GRAU, M. V., WALLACE, K. & HAILE, R. W. 2006. K-ras mutations in incident sporadic colorectal adenomas. *Cancer,* 106**,** 1036-40.

BAUER, V. P. & PAPACONSTANTINOU, H. T. 2008. Management of serrated adenomas and hyperplastic polyps. *Clin Colon Rectal Surg,* 21**,** 273-9.

BELT, E. J., FIJNEMAN, R. J., VAN DEN BERG, E. G., BRIL, H., DELIS-VAN DIEMEN, P. M., TIJSSEN, M., VAN ESSEN, H. F., DE LANGE-DE KLERK, E. S., BELIEN, J. A., STOCKMANN, H. B., MEIJER, S. & MEIJER,

G. A. 2011. Loss of lamin A/C expression in stage II and III colon cancer is associated with disease recurrence. *Eur J Cancer,* 47**,** 1837-45.

BENHAMOU, S. & SARASIN, A. 2000. Variability in nucleotide excision repair and cancer risk: a review. *Mutat Res,* 462**,** 149-58.

BENLLOCH, S., PAYA, A., ALENDA, C., BESSA, X., ANDREU, M., JOVER, R., CASTELLS, A., LLOR, X., ARANDA, F. I. & MASSUTI, B. 2006. Detection of BRAF V600E mutation in colorectal cancer: comparison of automatic sequencing and real-time chemistry methodology. *J Mol Diagn,* 8**,** 540-3.

BENSEN, S. P., COLE, B. F., MOTT, L. A., BARON, J. A., SANDLER, R. S. & HAILE, R. 1999. Colorectal hyperplastic polyps and risk of recurrence of adenomas and hyperplastic polyps. Polyps Prevention Study. *Lancet,* 354**,** 1873-4.

BENSON, A. B., 3RD 2007. Epidemiology, disease progression, and economic burden of colorectal cancer. *J Manag Care Pharm,* 13**,** S5-18.

BEYER, S. J., ZHANG, X., JIMENEZ, R. E., LEE, M. L., RICHARDSON, A. L., HUANG, K. & JHIANG, S. M. 2011. Microarray analysis of genes associated with cell surface NIS protein levels in breast cancer. *BMC Res Notes,* 4**,** 397.

BHAGAVATHULA, N., KELLEY, E. A., REDDY, M., NERUSU, K. C., LEONARD, C., FAY, K., CHAKRABARTY, S. & VARANI, J. 2005. Upregulation of calcium-sensing receptor and mitogen-activated protein kinase signalling in the regulation of growth and differentiation in colon carcinoma. *Br J Cancer,* 93**,** 1364-71.

BIANCHINI, M., LEVY, E., ZUCCHINI, C., PINSKI, V., MACAGNO, C., DE SANCTIS, P., VALVASSORI, L., CARINCI, P. & MORDOH, J. 2006. Comparative study of gene expression by cDNA microarray in human colorectal cancer tissues and normal mucosa. *Int J Oncol,* 29**,** 83-94.

BLAKER, H., SCHOLTEN, M., SUTTER, C., OTTO, H. F. & PENZEL, R. 2003. Somatic mutations in familial adenomatous polyps. Nuclear translocation of beta-catenin requires more than biallelic APC inactivation. *Am J Clin Pathol,* 120**,** 418-23.

BOLAND, C. R. & GOEL, A. 2005. Somatic evolution of cancer cells. *Semin Cancer Biol,* 15**,** 436-50.

BOLAND, C. R., SHIN, S. K. & GOEL, A. 2009. Promoter methylation in the genesis of gastrointestinal cancer. *Yonsei Med J,* 50**,** 309-21.

BOLSTAD, B. M., IRIZARRY, R. A., ASTRAND, M. & SPEED, T. P. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics,* 19**,** 185-93.

BOMMER, U. A., HENG, C., PERRIN, A., DASH, P., LOBOV, S., ELIA, A. & CLEMENS, M. J. 2010. Roles of the translationally controlled tumour protein (TCTP) and the double-stranded RNA-dependent protein kinase, PKR, in cellular stress responses. *Oncogene,* 29**,** 763-73.

BOMMER, U. A. & THIELE, B. J. 2004. The translationally controlled tumour protein (TCTP). *Int J Biochem Cell Biol,* 36**,** 379-85.

BOND, J. H. 2000. Polyp guideline: diagnosis, treatment, and surveillance for patients with colorectal polyps. Practice Parameters Committee of the American College of Gastroenterology. *Am J Gastroenterol,* 95**,** 3053-63.

BOYLE, P. & LANGMAN, J. S. 2000. ABC of colorectal cancer: Epidemiology. *BMJ,* 321**,** 805-8.

BOYLE, P. & LEON, M. E. 2002. Epidemiology of colorectal cancer. *Br Med Bull,* 64**,** 1-25.

BRUNETTO DE FARIAS, C., ROSEMBERG, D. B., HEINEN, T. E., KOEHLER-SANTOS, P., ABUJAMRA, A. L., KAPCZINSKI, F., BRUNETTO, A. L., ASHTON-PROLLA, P., MEURER, L., REIS BOGO, M., DAMIN, D.

C., SCHWARTSMANN, G. & ROESLER, R. 2010. BDNF/TrkB content and interaction with gastrin-releasing peptide receptor blockade in colorectal cancer. *Oncology,* 79**,** 430-9.

CAI, L. Y., ABE, M., IZUMI, S., IMURA, M., YASUGI, T. & USHIJIMA, T. 2007. Identification of PRTFDC1 silencing and aberrant promoter methylation of GPR150, ITGA8 and HOXD11 in ovarian cancers. *Life Sci,* 80**,** 1458-65.

CAMARERO, N., MASCARO, C., MAYORDOMO, C., VILARDELL, F., HARO, D. & MARRERO, P. F. 2006. Ketogenic HMGCS2 Is a c-Myc target gene expressed in differentiated cells of human colonic epithelium and down-regulated in colon cancer. *Mol Cancer Res,* 4**,** 645-53.

CARRAGHER, L. A., SNELL, K. R., GIBLETT, S. M., ALDRIDGE, V. S., PATEL, B., COOK, S. J., WINTON, D. J., MARAIS, R. & PRITCHARD, C. A. 2010. V600EBraf induces gastrointestinal crypt senescence and promotes tumour progression through enhanced CpG methylation of p16INK4a. *EMBO Mol Med,* 2**,** 458-71.

CENTER, M. M., JEMAL, A. & WARD, E. 2009. International trends in colorectal cancer incidence rates. *Cancer Epidemiol Biomarkers Prev,* 18**,** 1688-94.

CHAN, T. L., ZHAO, W., LEUNG, S. Y. & YUEN, S. T. 2003. BRAF and KRAS mutations in colorectal hyperplastic polyps and serrated adenomas. *Cancer Res,* 63**,** 4878-81.

CHANDLER, I. & HOULSTON, R. S. 2008. Interobserver agreement in grading of colorectal cancers-findings from a nationwide web-based survey of histopathologists. *Histopathology,* 52**,** 494-9.

CHEN, H., LEFFERTS, J. A., SCHWAB, M. C., SURIAWINATA, A. A. & TSONGALIS, G. J. 2011. Correlation of polypoid colorectal adenocarcinoma with pre-existing adenomatous polyps and KRAS mutation. *Cancer Genet,* 204**,** 245-51.

CHEN, X., HALBERG, R. B., BURCH, R. P. & DOVE, W. F. 2008. Intestinal adenomagenesis involves core molecular signatures of the epithelial-mesenchymal transition. *J Mol Histol,* 39**,** 283-94.

CHENG, L. & LAI, M. D. 2003. Aberrant crypt foci as microscopic precursors of colorectal cancer. *World J Gastroenterol,* 9**,** 2642-9.

CHIANG, Y., RIZZINO, A., SIBENALLER, Z. A., WOLD, M. S. & VISHWANATHA, J. K. 1999. Specific down-regulation of annexin II expression in human cells interferes with cell proliferation. *Mol Cell Biochem,* 199**,** 139-47.

CHUAQUI, R. F., BONNER, R. F., BEST, C. J., GILLESPIE, J. W., FLAIG, M. J., HEWITT, S. M., PHILLIPS, J. L., KRIZMAN, D. B., TANGREA, M. A., AHRAM, M., LINEHAN, W. M., KNEZEVIC, V. & EMMERT-BUCK, M. R. 2002. Post-analysis follow-up and validation of microarray experiments. *Nat Genet,* 32 Suppl**,** 509-14.

CLARK, J. C., COLLAN, Y., EIDE, T. J., ESTEVE, J., EWEN, S., GIBBS, N. M., JENSEN, O. M., KOSKELA, E., MACLENNAN, R., SIMPSON, J. G. & ET AL. 1985. Prevalence of polyps in an autopsy series from areas with varying incidence of large-bowel cancer. *Int J Cancer,* 36**,** 179-86.

CLEARY, T. J., MORALES, A. R., NADJI, M., NASSIRI, M. & VINCEK, V. 2005. Antimicrobial activity of UMFix tissue fixative. *J Clin Pathol,* 58**,** 22-5.

CONRAD, C., PODOLSKY, M. J. & CUSACK, J. C. 2011. Antiproteasomal agents in rectal cancer. *Anticancer Drugs,* 22**,** 341-50.

CORREA, P. 1978. Epidemiology of polyps and cancer. *Major Probl Pathol,* 10**,** 126-52.

CORREA, P., STRONG, J. P., REIF, A. & JOHNSON, W. D. 1977. The epidemiology of colorectal polyps: prevalence in New Orleans and international comparisons. *Cancer,* 39**,** 2258-64.

CORVINUS, F. M., ORTH, C., MORIGGL, R., TSAREVA, S. A., WAGNER, S., PFITZNER, E. B., BAUS, D., KAUFMANN, R., HUBER, L. A., ZATLOUKAL, K., BEUG, H., OHLSCHLAGER, P., SCHUTZ, A., HALBHUBER, K. J. & FRIEDRICH, K. 2005. Persistent STAT3 activation in colon cancer is associated with enhanced cell proliferation and tumor growth. *Neoplasia,* 7**,** 545-55.

COX, M. L., SCHRAY, C. L., LUSTER, C. N., STEWART, Z. S., KORYTKO, P. J., KN, M. K., PAULAUSKIS, J. D. & DUNSTAN, R. W. 2006. Assessment of fixatives, fixation, and tissue processing on morphology and RNA integrity. *Exp Mol Pathol,* 80**,** 183-91.

CRUK 2011. Cancer Research UK. *CancerStats Key Facts - large bowel cancer (colorectal cancer).* Cancer Research UK.

DASH, A., MAINE, I. P., VARAMBALLY, S., SHEN, R., CHINNAIYAN, A. M. & RUBIN, M. A. 2002. Changes in differential gene expression because of warm ischemia time of radical prostatectomy specimens. *Am J Pathol,* 161**,** 1743-8.

DEFAZIO, A., CHIEW, Y. E., SINI, R. L., JANES, P. W. & SUTHERLAND, R. L. 2000. Expression of c-erbB receptors, heregulin and oestrogen receptor in human breast cell lines. *Int J Cancer,* 87**,** 487-98.

DENG, G., BELL, I., CRAWLEY, S., GUM, J., TERDIMAN, J. P., ALLEN, B. A., TRUTA, B., SLEISENGER, M. H. & KIM, Y. S. 2004. BRAF mutation is frequently present in sporadic colorectal cancer with methylated hMLH1, but not in hereditary nonpolyposis colorectal cancer. *Clin Cancer Res,* 10**,** 191-5.

DHOMEN, N., DA ROCHA DIAS, S., HAYWARD, R., OGILVIE, L., HEDLEY, D., DELMAS, V., MCCARTHY, A., HENDERSON, D., SPRINGER, C. J., PRITCHARD, C., LARUE, L. & MARAIS, R. 2010. Inducible expression of (V600E) Braf using tyrosinase-driven Cre recombinase results in embryonic lethality. *Pigment Cell Melanoma Res,* 23**,** 112-20.

DONDRUP, M., HUSER, A. T., MERTENS, D. & GOESMANN, A. 2009. An evaluation framework for statistical tests on microarray data. *J Biotechnol,* 140**,** 18-26.

DONG, X., SEELAN, R. S., QIAN, C., MAI, M. & LIU, W. 2001. Genomic structure, chromosome mapping and expression analysis of the human AXIN2 gene. *Cytogenet Cell Genet,* 93**,** 26-8.

DORRANCE, H. R., DOCHERTY, G. M. & O'DWYER, P. J. 2000. Effect of surgeon specialty interest on patient outcome after potentially curative colorectal cancer surgery. *Dis Colon Rectum,* 43**,** 492-8.

DU, P., KIBBE, W. A. & LIN, S. M. 2008. lumi: a pipeline for processing Illumina microarray. *Bioinformatics,* 24**,** 1547-8.

DUKES, C. E. 1949. The Surgical Pathology of Rectal Cancer. *J Clin Pathol,* 2**,** 95-8.

DUNCAN, R., CARPENTER, B., MAIN, L. C., TELFER, C. & MURRAY, G. I. 2008. Characterisation and protein expression profiling of annexins in colorectal cancer. *Br J Cancer,* 98**,** 426-33.

EDELMANN, L. & EDELMANN, W. 2004. Loss of DNA mismatch repair function and cancer predisposition in the mouse: animal models for human hereditary nonpolyposis colorectal cancer. *Am J Med Genet C Semin Med Genet,* 129C**,** 91-9.

EIDE, T. J. 1986. Risk of colorectal cancer in adenoma-bearing individuals within a defined population. *Int J Cancer,* 38**,** 173-6.

EINSPAHR, J. G., MARTINEZ, M. E., JIANG, R., HSU, C. H., RASHID, A., BHATTACHARRYA, A. K., AHNEN, D. J., JACOBS, E. T., HOULIHAN, P. S., WEBB, C. R., ALBERTS, D. S. & HAMILTON, S. R. 2006. Associations of Ki-ras proto-oncogene mutation and p53 gene overexpression in sporadic colorectal adenomas with demographic and clinicopathologic characteristics. *Cancer Epidemiol Biomarkers Prev,* 15**,** 1443-50.

EMOTO, K., YAMADA, Y., SAWADA, H., FUJIMOTO, H., UENO, M., TAKAYAMA, T., KAMADA, K., NAITO, A., HIRAO, S. & NAKAJIMA, Y. 2001. Annexin II overexpression correlates with stromal tenascin-C overexpression: a prognostic marker in colorectal carcinoma. *Cancer,* 92**,** 1419-26.

ENSARI, A., BOSMAN, F. T. & OFFERHAUS, G. J. 2010. The serrated polyp: getting it right! *J Clin Pathol,* 63**,** 665-8.

ERTEL, A., VERGHESE, A., BYERS, S. W., OCHS, M. & TOZEREN, A. 2006. Pathway-specific differences between tumor cell lines and normal and tumor tissue cells. *Mol Cancer,* 5**,** 55.

ESPINA, C., CESPEDES, M. V., GARCIA-CABEZAS, M. A., GOMEZ DEL PULGAR, M. T., BOLUDA, A., OROZ, L. G., BENITAH, S. A., CEJAS, P., NISTAL, M., MANGUES, R. & LACAL, J. C. 2008. A critical role for Rac1 in tumor progression of human colorectal adenocarcinoma cells. *Am J Pathol,* 172**,** 156-66.

ESTELLER, M. 2005. Aberrant DNA methylation as a cancer-inducing mechanism. *Annu Rev Pharmacol Toxicol,* 45**,** 629-56.

FAKIH, M. & VINCENT, M. 2010. Adverse events associated with anti-EGFR therapies for the treatment of metastatic colorectal cancer. *Curr Oncol,* 17 Suppl 1**,** S18-30.

FANG, J. Y. & RICHARDSON, B. C. 2005. The MAPK signalling pathways and colorectal cancer. *Lancet Oncol,* 6**,** 322-7.

FEARNHEAD, N. S., BRITTON, M. P. & BODMER, W. F. 2001. The ABC of APC. *Hum Mol Genet,* 10**,** 721-33.

FEARON, E. R. & VOGELSTEIN, B. 1990. A genetic model for colorectal tumorigenesis. *Cell,* 61**,** 759-67.

FILIPENKO, N. R., MACLEOD, T. J., YOON, C. S. & WAISMAN, D. M. 2004. Annexin A2 is a novel RNA-binding protein. *J Biol Chem,* 279**,** 8723-31.

FLAVIN, P., REDMOND, A., MCBRYAN, J., COCCHIGLIA, S., TIBBITTS, P., FAHY-BROWNE, P., KAY, E., TREUMANN, A., PERREM, K., MCILROY, M., HILL, A. D. & YOUNG, L. S. 2011. RuvBl2 cooperates with Ets2 to transcriptionally regulate hTERT in colon cancer. *FEBS Lett,* 585**,** 2537-44.

FURLONG, M. T., HOUGH, C. D., SHERMAN-BAUST, C. A., PIZER, E. S. & MORIN, P. J. 1999. Evidence for the colonic origin of ovarian cancer cell line SW626. *J Natl Cancer Inst,* 91**,** 1327-8.

GALAMB, O., GYORFFY, B., SIPOS, F., SPISAK, S., NEMETH, A. M., MIHELLER, P., TULASSAY, Z., DINYA, E. & MOLNAR, B. 2008a. Inflammation, adenoma and cancer: objective classification of colon biopsy specimens with gene expression signature. *Dis Markers,* 25**,** 1-16.

GALAMB, O., SIPOS, F., SOLYMOSI, N., SPISAK, S., KRENACS, T., TOTH, K., TULASSAY, Z. & MOLNAR, B. 2008b. Diagnostic mRNA expression patterns of inflamed, benign, and malignant colorectal biopsy specimen and their correlation with peripheral blood results. *Cancer Epidemiol Biomarkers Prev,* 17**,** 2835-45.

GALAMB, O., SPISAK, S., SIPOS, F., TOTH, K., SOLYMOSI, N., WICHMANN, B., KRENACS, T., VALCZ, G., TULASSAY, Z. & MOLNAR, B. 2010. Reversal of gene expression changes in the colorectal normal-adenoma pathway by NS398 selective COX2 inhibitor. *Br J Cancer,* 102**,** 765-73.

GERKE, V. 2001. Annexins And Membrane Organisation In The Endocytic Pathway. *Cell Mol Biol Lett,* 6**,** 204.

GIOVANNUCCI, E., STAMPFER, M. J., COLDITZ, G., RIMM, E. B. & WILLETT, W. C. 1992. Relationship of diet to risk of colorectal adenoma in men. *J Natl Cancer Inst,* 84**,** 91-8.

GLEBOV, O. K., RODRIGUEZ, L. M., SOBALLE, P., DENOBILE, J., CLIATT, J., NAKAHARA, K. & KIRSCH, I. R. 2006. Gene expression patterns distinguish colonoscopically isolated human aberrant crypt foci from normal colonic mucosa. *Cancer Epidemiol Biomarkers Prev,* 15**,** 2253-62.

GNANASEKAR, M., DAKSHINAMOORTHY, G. & RAMASWAMY, K. 2009. Translationally controlled tumor protein is a novel heat shock protein with chaperone-like activity. *Biochem Biophys Res Commun,* 386**,** 333-7.

GOH, K. L. 2007. Changing trends in gastrointestinal disease in the Asia-Pacific region. *J Dig Dis,* 8**,** 179-85.

GOMEZ DEL PULGAR, T., BANDRES, E., ESPINA, C., VALDES-MORA, F., PEREZ-PALACIOS, R., GARCIA-AMIGOT, F., GARCIA-FONCILLAS, J. & LACAL, J. C. 2007. Differential expression of Rac1 identifies its target genes and its contribution to progression of colorectal cancer. *Int J Biochem Cell Biol,* 39**,** 2289-302.

GOPALAKRISHNAN, S., VAN EMBURGH, B. O. & ROBERTSON, K. D. 2008. DNA methylation in development and human disease. *Mutat Res,* 647**,** 30-8.

GREENE, F. L., STEWART, A. K. & NORTON, H. J. 2002. A new TNM staging strategy for node-positive (stage III) colon cancer: an analysis of 50,042 patients. *Ann Surg,* 236**,** 416-21; discussion 421.

GUDAS, J. M., PAYTON, M., THUKRAL, S., CHEN, E., BASS, M., ROBINSON, M. O. & COATS, S. 1999. Cyclin E2, a novel G1 cyclin that binds Cdk2 and is aberrantly expressed in human cancers. *Mol Cell Biol,* 19**,** 612-22.

GUGIC 2007. Novel tissue preservative and tissue fixative for comparative pathology and animal research. *Journal of Experimental Animal Science,* 43:271-281.

GUILLEMOT, F., NAGY, A., AUERBACH, A., ROSSANT, J. & JOYNER, A. L. 1994. Essential role of Mash-2 in extraembryonic development. *Nature,* 371**,** 333-6.

GUITARD, E., PARKER, F., MILLON, R., ABECASSIS, J. & TOCQUE, B. 2001. G3BP is overexpressed in human tumors and promotes S phase entry. *Cancer Lett,* 162**,** 213-21.

GULMANN, C., SHEEHAN, K. M., CONROY, R. M., WULFKUHLE, J. D., ESPINA, V., MULLARKEY, M. J., KAY, E. W., LIOTTA, L. A. & PETRICOIN, E. F., 3RD 2009. Quantitative cell signalling analysis reveals down-regulation of MAPK pathway activation in colorectal cancer. *J Pathol,* 218**,** 514-9.

GUZMAN-ARANGUEZ, A., OLMO, N., TURNAY, J., LECONA, E., PEREZ-RAMOS, P., LOPEZ DE SILANES, I. & LIZARBE, M. A. 2005. Differentiation of human colon adenocarcinoma cells alters the expression and intracellular localization of annexins A1, A2, and A5. *J Cell Biochem,* 94**,** 178-93.

GYORFFY, B., MOLNAR, B., LAGE, H., SZALLASI, Z. & EKLUND, A. C. 2009. Evaluation of microarray preprocessing algorithms based on concordance with RT-PCR in clinical samples. *PLoS One,* 4**,** e5645.

HADJIHANNAS, M. V., BRUCKNER, M., JERCHOW, B., BIRCHMEIER, W., DIETMAIER, W. & BEHRENS, J. 2006. Aberrant Wnt/beta-catenin signaling can induce chromosomal instability in colon cancer. *Proc Natl Acad Sci U S A,* 103**,** 10747-52.

HAFERKAMP, S., SCURR, L. L., BECKER, T. M., FRAUSTO, M., KEFFORD, R. F. & RIZOS, H. 2009. Oncogene-induced senescence does not require the p16(INK4a) or p14ARF melanoma tumor suppressors. *J Invest Dermatol,* 129**,** 1983-91.

HANINA, S. A., MIFSUD, W., DOWN, T. A., HAYASHI, K., O'CARROLL, D., LAO, K., MISKA, E. A. & SURANI, M. A. 2010. Genome-wide identification of targets and function of individual MicroRNAs in mouse embryonic stem cells. *PLoS Genet,* 6**,** e1001163.

HARBURG, G. C. & HINCK, L. 2011. Navigating breast cancer: axon guidance molecules as breast cancer tumor suppressors and oncogenes. *J Mammary Gland Biol Neoplasia,* 16**,** 257-70.

HARVEY, N. T. & RUSZKIEWICZ, A. 2007. Serrated neoplasia of the colorectum. *World J Gastroenterol,* 13**,** 3792-8.

HEALD, R. J. & RYALL, R. D. 1986. Recurrence and survival after total mesorectal excision for rectal cancer. *Lancet,* 1**,** 1479-82.

HEIJINK, D. M., FEHRMANN, R. S., DE VRIES, E. G., KOORNSTRA, J. J., OOSTERHUIS, D., VAN DER ZEE, A. G., KLEIBEUKER, J. H. & DE JONG, S. 2011. A bioinformatical and functional approach to identify novel strategies for chemoprevention of colorectal cancer. *Oncogene,* 30**,** 2026-36.

HEIJSTEK, M. W., KRANENBURG, O. & BOREL RINKES, I. H. 2005. Mouse models of colorectal cancer and liver metastases. *Dig Surg,* 22**,** 16-25.

HERMAN, J. G., LATIF, F., WENG, Y., LERMAN, M. I., ZBAR, B., LIU, S., SAMID, D., DUAN, D. S., GNARRA, J. R., LINEHAN, W. M. & ET AL. 1994. Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc Natl Acad Sci U S A,* 91**,** 9700-4.

HIGUCHI, T. & JASS, J. R. 2004. My approach to serrated polyps of the colorectum. *J Clin Pathol,* 57**,** 682-6.

HIROSE, H., ISHII, H., MIMORI, K., TANAKA, F., TAKEMASA, I., MIZUSHIMA, T., IKEDA, M., YAMAMOTO, H., SEKIMOTO, M., DOKI, Y. & MORI, M. 2011. The significance of PITX2 overexpression in human colorectal cancer. *Ann Surg Oncol,* 18**,** 3005-12.

HONG, Y., HO, K. S., EU, K. W. & CHEAH, P. Y. 2007. A susceptibility gene set for early onset colorectal cancer that integrates diverse signaling pathways: implication for tumorigenesis. *Clin Cancer Res,* 13**,** 1107-14.

HU, X. T., CHEN, W., WANG, D., SHI, Q. L., ZHANG, F. B., LIAO, Y. Q., JIN, M. & HE, C. 2008. [High expression of proteasome subunit PSMA7 in colorectal cancer is significantly correlated with liver metastasis]. *Zhonghua Zhong Liu Za Zhi,* 30**,** 515-8.

HUANG DA, W., SHERMAN, B. T. & LEMPICKI, R. A. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res,* 37**,** 1-13.

HUANG DA, W., SHERMAN, B. T. & LEMPICKI, R. A. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc,* 4**,** 44-57.

HUANG DA, W., SHERMAN, B. T., TAN, Q., KIR, J., LIU, D., BRYANT, D., GUO, Y., STEPHENS, R., BASELER, M. W., LANE, H. C. & LEMPICKI, R. A. 2007. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res,* 35**,** W169-75.

HUGHES, T. A. & BRADY, H. J. 2005. E2F1 up-regulates the expression of the tumour suppressor axin2 both by activation of transcription and by mRNA stabilisation. *Biochem Biophys Res Commun,* 329**,** 1267-74.

HUGHES, T. A. & BRADY, H. J. 2006. Regulation of axin2 expression at the levels of transcription, translation and protein stability in lung and colon cancer. *Cancer Lett,* 233**,** 338-47.

INOKUCHI, J., NARULA, N., YEE, D. S., SKARECKY, D. W., LAU, A., ORNSTEIN, D. K. & TYSON, D. R. 2009. Annexin A2 positively contributes to the malignant phenotype and secretion of IL-6 in DU145 prostate cancer cells. *Int J Cancer,* 124**,** 68-74.

IRVINE, K., STIRLING, R., HUME, D. & KENNEDY, D. 2004. Rasputin, more promiscuous than ever: a review of G3BP. *Int J Dev Biol,* 48**,** 1065-77.

ITO, Y., TAKEDA, T., OKADA, M. & MATSUURA, N. 2002. Expression of ets-1 and ets-2 in colonic neoplasms. *Anticancer Res,* 22**,** 1581-4.

JANSSEN, K. P., ALBERICI, P., FSIHI, H., GASPAR, C., BREUKEL, C., FRANKEN, P., ROSTY, C., ABAL, M., EL MARJOU, F., SMITS, R., LOUVARD, D., FODDE, R. & ROBINE, S. 2006. APC and oncogenic KRAS

are synergistic in enhancing Wnt signaling in intestinal tumor formation and progression. *Gastroenterology,* 131**,** 1096-109.

JANSSON, E. A., ARE, A., GREICIUS, G., KUO, I. C., KELLY, D., ARULAMPALAM, V. & PETTERSSON, S. 2005. The Wnt/beta-catenin signaling pathway targets PPARgamma activity in colon cancer cells. *Proc Natl Acad Sci U S A,* 102**,** 1460-5.

JASPERSON, K. W., TUOHY, T. M., NEKLASON, D. W. & BURT, R. W. 2010. Hereditary and familial colon cancer. *Gastroenterology,* 138**,** 2044-58.

JASS, J. R. 2007. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology,* 50**,** 113-30.

JASS, J. R., BAKER, K., ZLOBEC, I., HIGUCHI, T., BARKER, M., BUCHANAN, D. & YOUNG, J. 2006. Advanced colorectal polyps with the molecular and morphological features of serrated polyps and adenomas: concept of a 'fusion' pathway to colorectal cancer. *Histopathology,* 49**,** 121-31.

JEANMOUGIN, M., DE REYNIES, A., MARISA, L., PACCARD, C., NUEL, G. & GUEDJ, M. 2010. Should we abandon the t-test in the analysis of gene expression microarray data: a comparison of variance modeling strategies. *PLoS One,* 5**,** e12336.

JOTHY, S., SLESAK, B., HARLOZINSKA, A., LAPINSKA, J., ADAMIAK, J. & RABCZYNSKI, J. 1996. Field effect of human colon carcinoma on normal mucosa: relevance of carcinoembryonic antigen expression. *Tumour Biol,* 17**,** 58-64.

JUBB, A. M., CHALASANI, S., FRANTZ, G. D., SMITS, R., GRABSCH, H. I., KAVI, V., MAUGHAN, N. J., HILLAN, K. J., QUIRKE, P. & KOEPPEN, H. 2006. Achaete-scute like 2 (ascl2) is a target of Wnt signalling and is upregulated in intestinal neoplasia. *Oncogene,* 25**,** 3445-57.

JUNG, H. J., KIM, E. H., MUN, J. Y., PARK, S., SMITH, M. L., HAN, S. S. & SEO, Y. R. 2007. Base excision DNA repair defect in Gadd45a-deficient cells. *Oncogene,* 26**,** 7517-25.

KAMBARA, T., SIMMS, L. A., WHITEHALL, V. L., SPRING, K. J., WYNTER, C. V., WALSH, M. D., BARKER, M. A., ARNOLD, S., MCGIVERN, A., MATSUBARA, N., TANAKA, N., HIGUCHI, T., YOUNG, J., JASS, J. R. & LEGGETT, B. A. 2004. BRAF mutation is associated with DNA methylation in serrated polyps and cancers of the colorectum. *Gut,* 53**,** 1137-44.

KARAPETIS, C. S., KHAMBATA-FORD, S., JONKER, D. J., O'CALLAGHAN, C. J., TU, D., TEBBUTT, N. C., SIMES, R. J., CHALCHAL, H., SHAPIRO, J. D., ROBITAILLE, S., PRICE, T. J., SHEPHERD, L., AU, H. J., LANGER, C., MOORE, M. J. & ZALCBERG, J. R. 2008. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med,* 359**,** 1757-65.

KATOH, M. 2006. Cross-talk of WNT and FGF signaling pathways at GSK3beta to regulate beta-catenin and SNAIL signaling cascades. *Cancer Biol Ther,* 5**,** 1059-64.

KEYSE, S. M. 2008. Dual-specificity MAP kinase phosphatases (MKPs) and cancer. *Cancer Metastasis Rev,* 27**,** 253-61.

KHEIRELSEID, E. A., CHANG, K. H., NEWELL, J., KERIN, M. J. & MILLER, N. 2010. Identification of endogenous control genes for normalisation of real-time quantitative PCR data in colorectal cancer. *BMC Mol Biol,* 11**,** 12.

KIM, H., EUN, J. W., LEE, H., NAM, S. W., RHEE, H. & KOH, K. H. 2011a. Gene expression changes in patient-matched gastric normal mucosa, adenomas, and carcinomas. *Exp Mol Pathol,* 90**,** 201-9.

KIM, H. S., KANG, S. H., PARK, C. H., YANG, W. I., JEUNG, H. C., CHUNG, H. C., ROH, J. K., AHN, J. B., KIM, N. K., MIN, B. S. & RHA, S. Y. 2011b. Genome-wide molecular characterization of mucinous colorectal adenocarcinoma using cDNA microarray analysis. *Oncol Rep,* 25**,** 717-27.

KIM, K. M., CALABRESE, P., TAVARE, S. & SHIBATA, D. 2004. Enhanced stem cell survival in familial adenomatous polyposis. *Am J Pathol,* 164**,** 1369-77.

KIM, K. M. & SHIBATA, D. 2002. Methylation reveals a niche: stem cell succession in human colon crypts. *Oncogene,* 21**,** 5441-9.

KIM, M. M., WIEDERSCHAIN, D., KENNEDY, D., HANSEN, E. & YUAN, Z. M. 2007. Modulation of p53 and MDM2 activity by novel interaction with Ras-GAP binding proteins (G3BP). *Oncogene,* 26**,** 4209-15.

KINZLER, K. W. & VOGELSTEIN, B. 1996. Lessons from hereditary colorectal cancer. *Cell,* 87**,** 159-70.

KIRIKOSHI, H., SEKIHARA, H. & KATOH, M. 2001. Expression profiles of 10 members of Frizzled gene family in human gastric cancer. *Int J Oncol,* 19**,** 767-71.

KITA, H., HIKICHI, Y., HIKAMI, K., TSUNEYAMA, K., CUI, Z. G., OSAWA, H., OHNISHI, H., MUTOH, H., HOSHINO, H., BOWLUS, C. L., YAMAMOTO, H. & SUGANO, K. 2006. Differential gene expression between flat adenoma and normal mucosa in the colon in a microarray analysis. *J Gastroenterol,* 41**,** 1053-63.

KOINUMA, K., YAMASHITA, Y., LIU, W., HATANAKA, H., KURASHINA, K., WADA, T., TAKADA, S., KANEDA, R., CHOI, Y. L., FUJIWARA, S. I., MIYAKURA, Y., NAGAI, H. & MANO, H. 2006. Epigenetic silencing of AXIN2 in colorectal carcinoma with microsatellite instability. *Oncogene,* 25**,** 139-46.

KRUPP, M., MAASS, T., MARQUARDT, J. U., STAIB, F., BAUER, T., KONIG, R., BIESTERFELD, S., GALLE, P. R., TRESCH, A. & TEUFEL, A. 2011. The functional cancer map: a systems-level synopsis of genetic deregulation in cancer. *BMC Med Genomics,* 4**,** 53.

KUBISTA, M., ANDRADE, J. M., BENGTSSON, M., FOROOTAN, A., JONAK, J., LIND, K., SINDELKA, R., SJOBACK, R., SJOGREEN, B., STROMBOM, L., STAHLBERG, A. & ZORIC, N. 2006. The real-time polymerase chain reaction. *Mol Aspects Med,* 27**,** 95-125.

LAMPRECHT, S. A. & LIPKIN, M. 2001. Cellular mechanisms of calcium and vitamin D in the inhibition of colorectal carcinogenesis. *Ann N Y Acad Sci,* 952**,** 73-87.

LAMPRECHT, S. A. & LIPKIN, M. 2003. Chemoprevention of colon cancer by calcium, vitamin D and folate: molecular mechanisms. *Nat Rev Cancer,* 3**,** 601-14.

LARSSON, L. G. 2011. Oncogene- and tumor suppressor gene-mediated suppression of cellular senescence. *Semin Cancer Biol*, 21**,** 367-76.

LAU, P. C. & SUNG, J. J. 2010. Flat adenoma in colon: two decades of debate. *J Dig Dis,* 11**,** 201-7.

LAURIOLA, M., UGOLINI, G., ROSATI, G., ZANOTTI, S., MONTRONI, I., MANARESI, A., ZATTONI, D., RIVETTI, S., MATTEI, G., COPPOLA, D., STRIPPOLI, P., TAFFURELLI, M. & SOLMI, R. 2010. Identification by a Digital Gene Expression Displayer (DGED) and test by RT-PCR analysis of new mRNA candidate markers for colorectal cancer in peripheral blood. *Int J Oncol,* 37**,** 519-25.

LEE, J. H., AHN, J. H., BAHNG, H., KIM, T. W., KANG, Y. K., LEE, K. H., KIM, J. C., YU, C. S., KIM, J. H., AHN, S. D., KIM, W. K., KIM, S. H. & LEE, J. S. 2002a. Randomized trial of postoperative adjuvant therapy in stage II and III rectal cancer to define the optimal sequence of chemotherapy and radiotherapy: a preliminary report. *J Clin Oncol,* 20**,** 1751-8.

LEE, S., BANG, S., SONG, K. & LEE, I. 2006. Differential expression in normal-adenoma-carcinoma sequence suggests complex molecular carcinogenesis in colon. *Oncol Rep,* 16**,** 747-54.

LEE, W. J., ROBERTS-THOMSON, S. J., HOLMAN, N. A., MAY, F. J., LEHRBACH, G. M. & MONTEITH, G. R. 2002b. Expression of plasma membrane calcium pump isoform mRNAs in breast cancer cell lines. *Cell Signal,* 14**,** 1015-22.

LEGGETT, B. & WHITEHALL, V. 2010. Role of the serrated pathway in colorectal cancer pathogenesis. *Gastroenterology,* 138**,** 2088-100.

LEIBELING, D., LASPE, P. & EMMERT, S. 2006. Nucleotide excision repair and cancer. *J Mol Histol,* 37**,** 225-38.

LEIBOVITZ, A., STINSON, J. C., MCCOMBS, W. B., 3RD, MCCOY, C. E., MAZUR, K. C. & MABRY, N. D. 1976. Classification of human colorectal adenocarcinoma cell lines. *Cancer Res,* 36**,** 4562-9.

LESLIE, A., CAREY, F. A., PRATT, N. R. & STEELE, R. J. 2002. The colorectal adenoma-carcinoma sequence. *Br J Surg,* 89**,** 845-60.

LEVY-NISSENBAUM, O., SAGI-ASSIF, O., RAANANI, P., AVIGDOR, A., BEN-BASSAT, I. & WITZ, I. P. 2003a. cDNA microarray analysis reveals an overexpression of the dual-specificity MAPK phosphatase PYST2 in acute leukemia. *Methods Enzymol,* 366**,** 103-13.

LEVY-NISSENBAUM, O., SAGI-ASSIF, O., RAANANI, P., AVIGDOR, A., BEN-BASSAT, I. & WITZ, I. P. 2003b. Overexpression of the dual-specificity MAPK phosphatase PYST2 in acute leukemia. *Cancer Lett,* 199**,** 185-92.

LEWIS, T. S., SHAPIRO, P. S. & AHN, N. G. 1998. Signal transduction through MAP kinase cascades. *Adv Cancer Res,* 74**,** 49-139.

LI, X. H., ZHENG, Y., ZHENG, H. C., TAKAHASHI, H., YANG, X. H., MASUDA, S. & TAKANO, Y. 2010. REG IV overexpression in an early stage of colorectal carcinogenesis: an immunohistochemical study. *Histol Histopathol,* 25**,** 473-84.

LIAO, Y. & HUNG, M. C. 2003. Regulation of the activity of p38 mitogen-activated protein kinase by Akt in cancer and adenoviral protein E1A-mediated sensitization to apoptosis. *Mol Cell Biol,* 23**,** 6836-48.

LIEVRE, A., BACHET, J. B., LE CORRE, D., BOIGE, V., LANDI, B., EMILE, J. F., COTE, J. F., TOMASIC, G., PENNA, C., DUCREUX, M., ROUGIER, P., PENAULT-LLORCA, F. & LAURENT-PUIG, P. 2006. KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. *Cancer Res,* 66**,** 3992-5.

LIN, S. M., DU, P., HUBER, W. & KIBBE, W. A. 2008. Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Res,* 36**,** e11.

LIN, Y. M., FURUKAWA, Y., TSUNODA, T., YUE, C. T., YANG, K. C. & NAKAMURA, Y. 2002. Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas. *Oncogene,* 21**,** 4120-8.

LIU, Y., ZHENG, J., FANG, W., YOU, J., WANG, J., CUI, X. & WU, B. 2001. Identification of metastasis associated gene G3BP by differential display in human cancer cell sublines with different metastatic potentials G3BP as highly expressed in non-metastatic. *Chin Med J (Engl),* 114**,** 35-8.

LU, B., XU, J., ZHU, Y., ZHANG, H. & LAI, M. 2007. Systemic analysis of the differential gene expression profile in a colonic adenoma-normal SSH library. *Clin Chim Acta,* 378**,** 42-7.

LYNCH, H. T. & SMYRK, T. 1996. Hereditary nonpolyposis colorectal cancer (Lynch syndrome). An updated review. *Cancer,* 78**,** 1149-67.

LYNCH, H. T., SMYRK, T. & LYNCH, J. 1997. An update of HNPCC (Lynch syndrome). *Cancer Genet Cytogenet,* 93**,** 84-99.

MA, Q., GENG, Y., XU, W., WU, Y., HE, F., SHU, W., HUANG, M., DU, H. & LI, M. 2010. The role of translationally controlled tumor protein in tumor growth and metastasis of colon adenocarcinoma cells. *J Proteome Res,* 9**,** 40-9.

MAI, V., COLBERT, L. H., BERRIGAN, D., PERKINS, S. N., PFEIFFER, R., LAVIGNE, J. A., LANZA, E., HAINES, D. C., SCHATZKIN, A. & HURSTING, S. D. 2003. Calorie restriction and diet composition modulate spontaneous intestinal tumorigenesis in Apc(Min) mice through different mechanisms. *Cancer Res,* 63**,** 1752-5.

MANN, B., GELOS, M., SIEDOW, A., HANSKI, M. L., GRATCHEV, A., ILYAS, M., BODMER, W. F., MOYER, M. P., RIECKEN, E. O., BUHR, H. J. & HANSKI, C. 1999. Target genes of beta-catenin-T cell-factor/lymphoid-enhancer-factor signaling in human colorectal carcinomas. *Proc Natl Acad Sci U S A,* 96**,** 1603-8.

MARKOWITZ, A. J. & WINAWER, S. J. 1997. Management of colorectal polyps. *CA Cancer J Clin,* 47**,** 93-112.

MASUDA, N., OHNISHI, T., KAWAMOTO, S., MONDEN, M. & OKUBO, K. 1999. Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples. *Nucleic Acids Res,* 27**,** 4436-43.

MATEK, W., HERMANEK, P. & DEMLING, L. 1986. Is the adenoma-carcinoma sequence contradicted by the differing location of colorectal adenomas and carcinomas? *Endoscopy,* 18**,** 17-9.

MCARDLE, C. 2000a. ABC of colorectal cancer: effectiveness of follow up. *BMJ,* 321**,** 1332-5.

MCARDLE, C. 2000b. ABC of colorectal cancer: primary treatment-does the surgeon matter? *BMJ,* 321**,** 1121-3.

MCDONALD, S. A., PRESTON, S. L., LOVELL, M. J., WRIGHT, N. A. & JANKOWSKI, J. A. 2006. Mechanisms of disease: from stem cells to colorectal cancer. *Nat Clin Pract Gastroenterol Hepatol,* 3**,** 267-74.

MECKLIN, J. P., JARVINEN, H. J. & PELTOKALLIO, P. 1986. Cancer family syndrome. Genetic analysis of 22 Finnish kindreds. *Gastroenterology,* 90**,** 328-33.

MEHL, K. A., DAVIS, J. M., CLEMENTS, J. M., BERGER, F. G., PENA, M. M. & CARSON, J. A. 2005. Decreased intestinal polyp multiplicity is related to exercise mode and gender in ApcMin/+ mice. *J Appl Physiol,* 98**,** 2219-25.

MENIGATTI, M., CATTANEO, E., SABATES-BELLVER, J., ILINSKY, V. V., WENT, P., BUFFOLI, F., MARQUEZ, V. E., JIRICNY, J. & MARRA, G. 2009. The protein tyrosine phosphatase receptor type R gene is an early and frequent target of silencing in human colorectal tumorigenesis. *Mol Cancer,* 8**,** 124.

MOLINA, R., ALVAREZ, E., ANIEL-QUIROGA, A., BORQUE, M., CANDAS, B., LEON, A., POYATOS, R. M. & GELABERT, M. 2011. Evaluation of chromogranin A determined by three different procedures in patients with benign diseases, neuroendocrine tumors and other malignancies. *Tumour Biol,* 32**,** 13-22.

MONTICONE, M., BIOLLO, E., MAFFEI, M., DONADINI, A., ROMEO, F., STORLAZZI, C. T., GIARETTI, W. & CASTAGNOLA, P. 2008. Gene expression deregulation by KRAS G12D and G12V in a BRAF V600E context. *Mol Cancer,* 7**,** 92.

MOSER, A. R., PITOT, H. C. & DOVE, W. F. 1990. A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse. *Science,* 247**,** 322-4.

MUNEMITSU, S., ALBERT, I., SOUZA, B., RUBINFELD, B. & POLAKIS, P. 1995. Regulation of intracellular beta-catenin levels by the adenomatous polyposis coli (APC) tumor-suppressor protein. *Proc Natl Acad Sci U S A,* 92**,** 3046-50.

MUTO, T., BUSSEY, H. J. & MORSON, B. C. 1975. The evolution of cancer of the colon and rectum. *Cancer,* 36**,** 2251-70.

NGUYEN, A. V., ALBERS, C. G. & HOLCOMBE, R. F. 2010. Differentiation of tubular and villous adenomas based on Wnt pathway-related gene expression profiles. *Int J Mol Med,* 26**,** 121-5.

NGUYEN, C. Q., SHARMA, A., SHE, J. X., MCINDOE, R. A. & PECK, A. B. 2009. Differential gene expressions in the lacrimal gland during development and onset of keratoconjunctivitis sicca in Sjogren's syndrome (SJS)-like disease of the C57BL/6.NOD-Aec1Aec2 mouse. *Exp Eye Res,* 88**,** 398-409.

NONN, L., ANANTHANARAYANAN, V. & GANN, P. H. 2009. Evidence for field cancerization of the prostate. *Prostate,* 69**,** 1470-9.

NUGENT, K. P., PHILLIPS, R. K., HODGSON, S. V., COTTRELL, S., SMITH-RAVIN, J., PACK, K. & BODMER, W. F. 1994. Phenotypic expression in familial adenomatous polyposis: partial prediction by mutation analysis. *Gut,* 35**,** 1622-3.

NUMATA, M., OSHIMA, T., YOSHIHARA, K., WATANABE, T., TSUCHIDA, K., TAMAGAWA, H., YAMAMOTO, N., SHIOZAWA, M., MORINAGA, S., AKAIKE, M., KUNISAKI, C., RINO, Y., TANAKA, K., MASUDA, M. & IMADA, T. 2011. Relationship between RegIV gene expression to outcomes in colorectal cancer. *J Surg Oncol,* 104**,** 205-9.

O'BRIEN, M. J., WINAWER, S. J., ZAUBER, A. G., GOTTLIEB, L. S., STERNBERG, S. S., DIAZ, B., DICKERSIN, G. R., EWING, S., GELLER, S., KASIMIAN, D. & ET AL. 1990. The National Polyp Study. Patient and polyp characteristics associated with high-grade dysplasia in colorectal adenomas. *Gastroenterology,* 98**,** 371-9.

OBAYASHI, T., HAYASHI, S., SHIBAOKA, M., SAEKI, M., OHTA, H. & KINOSHITA, K. 2008. COXPRESdb: a database of coexpressed gene networks in mammals. *Nucleic Acids Res,* 36**,** D77-82.

OBAYASHI, T. & KINOSHITA, K. 2011. COXPRESdb: a database to compare gene coexpression in seven model animals. *Nucleic Acids Res,* 39**,** D1016-22.

PANCIONE, M., SABATINO, L., FUCCI, A., CARAFA, V., NEBBIOSO, A., FORTE, N., FEBBRARO, A., PARENTE, D., AMBROSINO, C., NORMANNO, N., ALTUCCI, L. & COLANTUONI, V. 2010. Epigenetic silencing of peroxisome proliferator-activated receptor gamma is a biomarker for colorectal cancer progression and adverse patients' outcome. *PLoS One,* 5**,** e14229.

PAP, Z., PAVAI, Z., DENES, L., BRINZANIUC, K. & JUNG, I. 2011. Hyperplastic polyps and serrated adenomas: precancerous lesions with mixed immunophenotype. *Rom J Morphol Embryol,* 52**,** 797-802.

PAPADOPOULOS, N., NICOLAIDES, N. C., WEI, Y. F., RUBEN, S. M., CARTER, K. C., ROSEN, C. A., HASELTINE, W. A., FLEISCHMANN, R. D., FRASER, C. M., ADAMS, M. D. & ET AL. 1994. Mutation of a mutL homolog in hereditary colon cancer. *Science,* 263**,** 1625-9.

PARFITT, J. R. & DRIMAN, D. K. 2007. Survivin and hedgehog protein expression in serrated colorectal polyps: an immunohistochemical study. *Hum Pathol,* 38**,** 710-7.

PERRY, J. K., KANNAN, N., GRANDISON, P. M., MITCHELL, M. D. & LOBIE, P. E. 2008. Are trefoil factors oncogenic? *Trends Endocrinol Metab,* 19**,** 74-81.

PETERS, U., CHATTERJEE, N., YEAGER, M., CHANOCK, S. J., SCHOEN, R. E., MCGLYNN, K. A., CHURCH, T. R., WEISSFELD, J. L., SCHATZKIN, A. & HAYES, R. B. 2004. Association of genetic variants in the calcium-sensing receptor with risk of colorectal adenoma. *Cancer Epidemiol Biomarkers Prev,* 13**,** 2181-6.

PHELPS, R. A., CHIDESTER, S., DEHGHANIZADEH, S., PHELPS, J., SANDOVAL, I. T., RAI, K., BROADBENT, T., SARKAR, S., BURT, R. W. & JONES, D. A. 2009. A two-step model for colon adenoma initiation and progression caused by APC loss. *Cell,* 137**,** 623-34.

PINO, M. S. & CHUNG, D. C. 2010. The chromosomal instability pathway in colon cancer. *Gastroenterology,* 138**,** 2059-72.

PINTO, D. & CLEVERS, H. 2005. Wnt, stem cells and cancer in the intestine. *Biol Cell,* 97**,** 185-96.

POTTEN, C. S., BOOTH, C. & PRITCHARD, D. M. 1997. The intestinal epithelial stem cell: the mucosal governor. *Int J Exp Pathol,* 78**,** 219-43.

PRESTON, S. L., WONG, W. M., CHAN, A. O., POULSOM, R., JEFFERY, R., GOODLAD, R. A., MANDIR, N., ELIA, G., NOVELLI, M., BODMER, W. F., TOMLINSON, I. P. & WRIGHT, N. A. 2003. Bottom-up histogenesis of colorectal adenomas: origin in the monocryptal adenoma and initial expansion by crypt fission. *Cancer Res,* 63**,** 3819-25.

PRITCHARD, C., CARRAGHER, L., ALDRIDGE, V., GIBLETT, S., JIN, H., FOSTER, C., ANDREADI, C. & KAMATA, T. 2007. Mouse models for BRAF-induced cancers. *Biochem Soc Trans,* 35**,** 1329-33.

PROLLA, T. A., PANG, Q., ALANI, E., KOLODNER, R. D. & LISKAY, R. M. 1994. MLH1, PMS1, and MSH2 interactions during the initiation of DNA mismatch repair in yeast. *Science,* 265**,** 1091-3.

QIN, L., CHEN, X., WU, Y., FENG, Z., HE, T., WANG, L., LIAO, L. & XU, J. 2011. Steroid receptor coactivator-1 upregulates integrin alpha expression to promote breast cancer cell adhesion and migration. *Cancer Res,* 71**,** 1742-51.

RACHLIN, J., DING, C., CANTOR, C. & KASIF, S. 2005. MuPlex: multi-objective multiplex PCR assay design. *Nucleic Acids Res,* 33**,** W544-7.

REINHOLZ, M. M., ECKEL-PASSOW, J. E., ANDERSON, S. K., ASMANN, Y. W., ZSCHUNKE, M. A., OBERG, A. L., MCCULLOUGH, A. E., DUECK, A. C., CHEN, B., APRIL, C. S., WICKHAM-GARCIA, E., JENKINS, R. B., CUNNINGHAM, J. M., JEN, J., PEREZ, E. A., FAN, J. B. & LINGLE, W. L. 2010. Expression profiling of formalin-fixed paraffin-embedded primary breast tumors using cancer-specific and whole genome gene panels on the DASL(R) platform. *BMC Med Genomics,* 3**,** 60.

REY, O., YOUNG, S. H., JACAMO, R., MOYER, M. P. & ROZENGURT, E. 2010. Extracellular calcium sensing receptor stimulation in human colonic epithelial cells induces intracellular calcium oscillations and proliferation inhibition. *J Cell Physiol,* 225**,** 73-83.

RICKERT, R. R., AUERBACH, O., GARFINKEL, L., HAMMOND, E. C. & FRASCA, J. M. 1979. Adenomatous lesions of the large bowel: an autopsy survey. *Cancer,* 43**,** 1847-57.

RIDGWAY, P. F. & DARZI, A. W. 2003. The role of total mesorectal excision in the management of rectal cancer. *Cancer Control,* 10**,** 205-11.

RISIO, M. 2010. The natural history of adenomas. *Best Pract Res Clin Gastroenterol,* 24**,** 271-80.

ROCCARO, A. M., VACCA, A. & RIBATTI, D. 2006. Bortezomib in the treatment of cancer. *Recent Pat Anticancer Drug Discov,* 1**,** 397-403.

ROZEN, S. & SKALETSKY, H. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol,* 132**,** 365-86.

SABATES-BELLVER, J., VAN DER FLIER, L. G., DE PALO, M., CATTANEO, E., MAAKE, C., REHRAUER, H., LACZKO, E., KUROWSKI, M. A., BUJNICKI, J. M., MENIGATTI, M., LUZ, J., RANALLI, T. V., GOMES, V., PASTORELLI, A., FAGGIANI, R., ANTI, M., JIRICNY, J., CLEVERS, H. & MARRA, G. 2007. Transcriptome profile of human colorectal adenomas. *Mol Cancer Res,* 5**,** 1263-75.

SAEED, A. I., BHAGABATI, N. K., BRAISTED, J. C., LIANG, W., SHAROV, V., HOWE, E. A., LI, J., THIAGARAJAN, M., WHITE, J. A. & QUACKENBUSH, J. 2006. TM4 microarray software suite. *Methods Enzymol,* 411**,** 134-93.

SAEED, A. I., SHAROV, V., WHITE, J., LI, J., LIANG, W., BHAGABATI, N., BRAISTED, J., KLAPA, M., CURRIER, T., THIAGARAJAN, M., STURN, A., SNUFFIN, M., REZANTSEV, A., POPOV, D., RYLTSOV, A., KOSTUKOVICH, E., BORISOVSKY, I., LIU, Z., VINSAVICH, A., TRUSH, V. & QUACKENBUSH, J. 2003. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques,* 34**,** 374-8.

SALAHSHOR, S. & WOODGETT, J. R. 2005. The links between axin and carcinogenesis. *J Clin Pathol,* 58**,** 225-36.

SALAMA, P. & PLATELL, C. 2009. Colorectal cancer stem cells. *ANZ J Surg,* 79**,** 697-702.

SANSOM, O. J., MENIEL, V., WILKINS, J. A., COLE, A. M., OIEN, K. A., MARSH, V., JAMIESON, T. J., GUERRA, C., ASHTON, G. H., BARBACID, M. & CLARKE, A. R. 2006. Loss of Apc allows phenotypic manifestation of the transforming properties of an endogenous K-ras oncogene in vivo. *Proc Natl Acad Sci U S A,* 103**,** 14122-7.

SATO, T., OSHIMA, T., YOSHIHARA, K., YAMAMOTO, N., YAMADA, R., NAGANO, Y., FUJII, S., KUNISAKI, C., SHIOZAWA, M., AKAIKE, M., RINO, Y., TANAKA, K., MASUDA, M. & IMADA, T. 2009. Overexpression of the fibroblast growth factor receptor-1 gene correlates with liver metastasis in colorectal cancer. *Oncol Rep,* 21**,** 211-6.

SCHOFIELD, P. F. & JONES, D. J. 1992a. ABC of colorectal diseases. Colorectal neoplasia--II: Large bowel cancer. *BMJ,* 304**,** 1561-3.

SCHOFIELD, P. F. & JONES, D. J. 1992b. ABC of colorectal diseases. Colorectal neoplasia--III: Treatment and prevention. *BMJ,* 304**,** 1624-7.

SCHOLEFIELD, J. H. 2000. ABC of colorectal cancer: screening. *BMJ,* 321**,** 1004-6.

SEBOLT-LEOPOLD, J. S. 2000. Development of anticancer drugs targeting the MAP kinase pathway. *Oncogene,* 19**,** 6594-9.

SENDA, T., IIZUKA-KOGO, A., ONOUCHI, T. & SHIMOMURA, A. 2007. Adenomatous polyposis coli (APC) plays multiple roles in the intestinal and colorectal epithelia. *Med Mol Morphol,* 40**,** 68-81.

SETH, A. & WATSON, D. K. 2005. ETS transcription factors and their emerging roles in human cancer. *Eur J Cancer,* 41**,** 2462-78.

SHEU, J. & SHIH IE, M. 2010. HLA-G and immune evasion in cancer cells. *J Formos Med Assoc,* 109**,** 248-57.

SHINYA, H. & WOLFF, W. I. 1979. Morphology, anatomic distribution and cancer potential of colonic polyps. *Ann Surg,* 190**,** 679-83.

SHITASHIGE, M., SATOW, R., JIGAMI, T., AOKI, K., HONDA, K., SHIBATA, T., ONO, M., HIROHASHI, S. & YAMADA, T. 2010. Traf2- and Nck-interacting kinase is essential for Wnt signaling and colorectal cancer growth. *Cancer Res,* 70**,** 5024-33.

SINGH, P. 2007. Role of Annexin-II in GI cancers: interaction with gastrins/progastrins. *Cancer Lett,* 252**,** 19-35.

SIU, L. S., ROMANSKA, H., ABEL, P. D., BAUS-LONCAR, M., KAYADEMIR, T., STAMP, G. W. & LALANI EL, N. 2004. TFF2 (trefoil family factor2) inhibits apoptosis in breast and colorectal cancer cell lines. *Peptides,* 25**,** 855-63.

SKRZYPCZAK, M., GORYCA, K., RUBEL, T., PAZIEWSKA, A., MIKULA, M., JAROSZ, D., PACHLEWSKI, J., OLEDZKI, J. & OSTROWSKI, J. 2010. Modeling oncogenic signaling in colon tumors by multidirectional analyses of microarray data directed for maximization of analytical reliability. *PLoS One,* 5(10).

SMIRNOV, D. A., ZWEITZIG, D. R., FOULK, B. W., MILLER, M. C., DOYLE, G. V., PIENTA, K. J., MEROPOL, N. J., WEINER, L. M., COHEN, S. J., MORENO, J. G., CONNELLY, M. C., TERSTAPPEN, L. W. & O'HARA, S. M. 2005. Global gene expression profiling of circulating tumor cells. *Cancer Res,* 65**,** 4993-7.

SMITS, R., VAN DER HOUVEN VAN OORDT, W., LUZ, A., ZURCHER, C., JAGMOHAN-CHANGUR, S., BREUKEL, C., KHAN, P. M. & FODDE, R. 1998. Apc1638N: a mouse model for familial adenomatous polyposis-associated desmoid tumors and cutaneous cysts. *Gastroenterology,* 114**,** 275-83.

SMYTH, G. K. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol,* 3**,** Article3.

SNOVER, D. C. 2011. Update on the serrated pathway to colorectal carcinoma. *Hum Pathol,* 42**,** 1-10.

SOH, J. W., MAO, Y., LIU, L., THOMPSON, W. J., PAMUKCU, R. & WEINSTEIN, I. B. 2001. Protein kinase G activates the JNK1 pathway via phosphorylation of MEKK1. *J Biol Chem,* 276**,** 16406-10.

SOON, P. S., GILL, A. J., BENN, D. E., CLARKSON, A., ROBINSON, B. G., MCDONALD, K. L. & SIDHU, S. B. 2009. Microarray gene expression and immunohistochemistry analyses of adrenocortical tumors identify IGF2 and Ki-67 as useful in differentiating carcinomas from adenomas. *Endocr Relat Cancer,* 16**,** 573-83.

SOWDEN, J., LEIGH, S., TALBOT, I., DELHANTY, J. & EDWARDS, Y. 1993. Expression from the proximal promoter of the carbonic anhydrase 1 gene as a marker for differentiation in colon epithelia. *Differentiation,* 53**,** 67-74.

SPANO, J. P., MILANO, G., RIXE, C. & FAGARD, R. 2006. JAK/STAT signalling pathway in colorectal cancer: a new biological target with therapeutic implications. *Eur J Cancer,* 42**,** 2668-70.

SPRADLING, A., DRUMMOND-BARBOSA, D. & KAI, T. 2001. Stem cells find their niche. *Nature,* 414**,** 98-104.

STANGE, D. E., ENGEL, F., LONGERICH, T., KOO, B. K., KOCH, M., DELHOMME, N., AIGNER, M., TOEDT, G., SCHIRMACHER, P., LICHTER, P., WEITZ, J. & RADLWIMMER, B. 2010. Expression of an ASCL2 related stem cell signature and IGF2 in colorectal cancer liver metastases with 11p15.5 gain. *Gut,* 59**,** 1236-44.

STEIN, C. J. & COLDITZ, G. A. 2004. Modifiable risk factors for cancer. *Br J Cancer,* 90**,** 299-303.

STRAND, M., PROLLA, T. A., LISKAY, R. M. & PETES, T. D. 1993. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature,* 365**,** 274-6.

STRATHDEE, G. & BROWN, R. 2002. Aberrant DNA methylation in cancer: potential clinical interventions. *Expert Rev Mol Med,* 4**,** 1-17.

SUZUKI, H. I., YAMAGATA, K., SUGIMOTO, K., IWAMOTO, T., KATO, S. & MIYAZONO, K. 2009. Modulation of microRNA processing by p53. *Nature,* 460**,** 529-33.

TAIPALE, J. & BEACHY, P. A. 2001. The Hedgehog and Wnt signalling pathways in cancer. *Nature,* 411**,** 349-54.

TAKAHASHI, M., NAKAMURA, Y., OBAMA, K. & FURUKAWA, Y. 2005. Identification of SP5 as a downstream gene of the beta-catenin/Tcf pathway and its enhanced expression in human colon cancer. *Int J Oncol,* 27**,** 1483-7.

TAKAKU, K., OSHIMA, M., MIYOSHI, H., MATSUI, M., SELDIN, M. F. & TAKETO, M. M. 1998. Intestinal tumorigenesis in compound mutant mice of both Dpc4 (Smad4) and Apc genes. *Cell,* 92**,** 645-56.

TAKETO, M. M. & EDELMANN, W. 2009. Mouse models of colon cancer. *Gastroenterology,* 136**,** 780-98.

TANG, H., GUO, Q., ZHANG, C., ZHU, J., YANG, H., ZOU, Y. L., YAN, Y., HONG, D., SOU, T. & YAN, X. M. 2010. Identification of an intermediate signature that marks the initial phases of the colorectal adenoma-carcinoma transition. *Int J Mol Med,* 26**,** 631-41.

TASSI, E., HENKE, R. T., BOWDEN, E. T., SWIFT, M. R., KODACK, D. P., KUO, A. H., MAITRA, A. & WELLSTEIN, A. 2006. Expression of a fibroblast growth factor-binding protein during the development of adenocarcinoma of the pancreas and colon. *Cancer Res,* 66**,** 1191-8.

TASSI, E. & WELLSTEIN, A. 2006. The angiogenic switch molecule, secreted FGF-binding protein, an indicator of early stages of pancreatic and colorectal adenocarcinoma. *Semin Oncol,* 33**,** S50-6.

TATEYAMA, H., LI, W., TAKAHASHI, E., MIURA, Y., SUGIURA, H. & EIMOTO, T. 2002. Apoptosis index and apoptosis-related antigen expression in serrated adenoma of the colorectum: the saw-toothed structure may be related to inhibition of apoptosis. *Am J Surg Pathol,* 26**,** 249-56.

TAUPIN, D. & PODOLSKY, D. K. 1999. Mitogen-activated protein kinase activation regulates intestinal epithelial differentiation. *Gastroenterology,* 116**,** 1072-80.

TELERMAN, A. & AMSON, R. 2009. The molecular programme of tumour reversion: the steps beyond malignant transformation. *Nat Rev Cancer,* 9**,** 206-16.

THIELE, H., BERGER, M., SKALWEIT, A. & THIELE, B. J. 2000. Expression of the gene and processed pseudogenes encoding the human and rabbit translationally controlled tumour protein (TCTP). *Eur J Biochem,* 267**,** 5473-81.

TOYOTA, M., AHUJA, N., OHE-TOYOTA, M., HERMAN, J. G., BAYLIN, S. B. & ISSA, J. P. 1999. CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A,* 96**,** 8681-6.

TROPPMAIR, J., BRUDER, J. T., MUNOZ, H., LLOYD, P. A., KYRIAKIS, J., BANERJEE, P., AVRUCH, J. & RAPP, U. R. 1994. Mitogen-activated protein kinase/extracellular signal-regulated protein kinase activation by oncogenes, serum, and 12-O-tetradecanoylphorbol-13-acetate requires Raf and is necessary for transformation. *J Biol Chem,* 269**,** 7030-5.

TUSHER, V. G., TIBSHIRANI, R. & CHU, G. 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A,* 98**,** 5116-21.

TUYNDER, M., SUSINI, L., PRIEUR, S., BESSE, S., FIUCCI, G., AMSON, R. & TELERMAN, A. 2002. Biological models and genes of tumor reversion: cellular reprogramming through tpt1/TCTP and SIAH-1. *Proc Natl Acad Sci U S A,* 99**,** 14976-81.

VAKIANI, E. & SOLIT, D. B. 2011. KRAS and BRAF: drug targets and predictive biomarkers. *J Pathol,* 223**,** 219-29.

VARANI, J. 2011. Calcium, calcium-sensing receptor and growth control in the colonic mucosa. *Histol Histopathol,* 26**,** 769-79.

VELHO, S., MOUTINHO, C., CIRNES, L., ALBUQUERQUE, C., HAMELIN, R., SCHMITT, F., CARNEIRO, F., OLIVEIRA, C. & SERUCA, R. 2008. BRAF, KRAS and PIK3CA mutations in colorectal serrated polyps and cancer: primary or secondary genetic events in colorectal carcinogenesis? *BMC Cancer,* 8**,** 255.

VIGNERON, A. M. & VOUSDEN, K. H. 2011. An indirect role for ASPP1 in limiting p53-dependent p21 expression and cellular senescence. *EMBO J,* 31**,** 471-80.

VINCEK, V., NASSIRI, M., NADJI, M. & MORALES, A. R. 2003. A tissue fixative that protects macromolecules (DNA, RNA, and protein) and histomorphology in clinical samples. *Lab Invest,* 83**,** 1427-35.

VOGELSTEIN, B., FEARON, E. R., HAMILTON, S. R., KERN, S. E., PREISINGER, A. C., LEPPERT, M., NAKAMURA, Y., WHITE, R., SMITS, A. M. & BOS, J. L. 1988. Genetic alterations during colorectal-tumor development. *N Engl J Med,* 319**,** 525-32.

VOUSDEN, K. H. & PRIVES, C. 2009. Blinded by the Light: The Growing Complexity of p53. *Cell,* 137**,** 413-31.

WADDELL, N., COCCIARDI, S., JOHNSON, J., HEALEY, S., MARSH, A., RILEY, J., DA SILVA, L., VARGAS, A. C., REID, L., SIMPSON, P. T., LAKHANI, S. R. & CHENEVIX-TRENCH, G. 2010. Gene expression profiling of formalin-fixed, paraffin-embedded familial breast tumours using the whole genome-DASL assay. *J Pathol,* 221**,** 452-61.

WANG, D. & DUBOIS, R. N. 2010. Therapeutic potential of peroxisome proliferator-activated receptors in chronic inflammation and colorectal cancer. *Gastroenterol Clin North Am,* 39**,** 697-707.

WANG, X., WANG, Q., HU, W. & EVERS, B. M. 2004. Regulation of phorbol ester-mediated TRAF1 induction in human colon cancer cells through a PKC/RAF/ERK/NF-kappaB-dependent pathway. *Oncogene,* 23**,** 1885-95.

WANG, X. T., NAGABA, Y., CROSS, H. S., WRBA, F., ZHANG, L. & GUGGINO, S. E. 2000. The mRNA of L-type calcium channel elevated in colon cancer: protein distribution in normal and cancerous colon. *Am J Pathol,* 157**,** 1549-62.

WARRINGTON, J. A., NAIR, A., MAHADEVAPPA, M. & TSYGANSKAYA, M. 2000. Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. *Physiol Genomics,* 2**,** 143-7.

WEITZ, J., KOCH, M., DEBUS, J., HOHLER, T., GALLE, P. R. & BUCHLER, M. W. 2005. Colorectal cancer. *Lancet,* 365**,** 153-65.

WHEELER, J. M., BODMER, W. F. & MORTENSEN, N. J. 2000. DNA mismatch repair genes and colorectal cancer. *Gut,* 47**,** 148-53.

WILLIS, N. D., COX, T. R., RAHMAN-CASANS, S. F., SMITS, K., PRZYBORSKI, S. A., VAN DEN BRANDT, P., VAN ENGELAND, M., WEIJENBERG, M., WILSON, R. G., DE BRUINE, A. & HUTCHISON, C. J. 2008. Lamin A/C is a risk biomarker in colorectal cancer. *PLoS One,* 3**,** e2988.

WINAWER, S. J., O'BRIEN, M. J., WAYE, J. D., KRONBORG, O., BOND, J., FRUHMORGEN, P., SOBIN, L. H., BURT, R., ZAUBER, A. & MORSON, B. 1990. Risk and surveillance of individuals with colorectal polyps. Who Collaborating Centre for the Prevention of Colorectal Cancer. *Bull World Health Organ,* 68**,** 789-95.

WINAWER, S. J., ZAUBER, A. G., HO, M. N., O'BRIEN, M. J., GOTTLIEB, L. S., STERNBERG, S. S., WAYE, J. D., SCHAPIRO, M., BOND, J. H., PANISH, J. F. & ET AL. 1993. Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. *N Engl J Med,* 329**,** 1977-81.

WINDER, T. & LENZ, H. J. 2010. Molecular predictive and prognostic markers in colon cancer. *Cancer Treat Rev,* 36**,** 550-6.

WONG, A. W., PAULSON, Q. X., HONG, J., STUBBINS, R. E., POH, K., SCHRADER, E. & NUNEZ, N. P. 2011. Alcohol promotes breast cancer cell invasion by regulating the Nm23-ITGA5 pathway. *J Exp Clin Cancer Res,* 30**,** 75.

WU, J. S., PAUL, P., MCGANNON, E. A. & CHURCH, J. M. 1998. APC genotype, polyp number, and surgical options in familial adenomatous polyposis. *Ann Surg,* 227**,** 57-62.

WU, X. Y., FU, Z. X., WANG, X. H. & SHEN, W. 2010. Identification of differential proteins in colon cancer SW480 cells with HIF1-alpha silence by proteome analysis. *Neoplasma,* 57**,** 299-305.

YAN, D., WIESMANN, M., ROHAN, M., CHAN, V., JEFFERSON, A. B., GUO, L., SAKAMOTO, D., CAOTHIEN, R. H., FULLER, J. H., REINHARD, C., GARCIA, P. D., RANDAZZO, F. M., ESCOBEDO, J., FANTL, W. J. & WILLIAMS, L. T. 2001. Elevated expression of axin2 and hnkd mRNA provides evidence that Wnt/beta -catenin signaling is activated in human colon tumors. *Proc Natl Acad Sci U S A,* 98**,** 14973-8.

YANG, K., EDELMANN, W., FAN, K., LAU, K., LEUNG, D., NEWMARK, H., KUCHERLAPATI, R. & LIPKIN, M. 1998. Dietary modulation of carcinoma development in a mouse model for human familial adenomatous polyposis. *Cancer Res,* 58**,** 5713-7.

YANG, S., FARRAYE, F. A., MACK, C., POSNIK, O. & O'BRIEN, M. J. 2004. BRAF and KRAS Mutations in hyperplastic polyps and serrated adenomas of the colorectum: relationship to histology and CpG island methylation status. *Am J Surg Pathol,* 28**,** 1452-9.

YEH, C. S., WANG, J. Y., CHUNG, F. Y., LEE, S. C., HUANG, M. Y., KUO, C. W., YANG, M. J. & LIN, S. R. 2008. Significance of the glycolytic pathway and glycolysis related-genes in tumorigenesis of human colorectal cancers. *Oncol Rep,* 19**,** 81-91.

YEN, A., ROBERSON, M. S., VARVAYANIS, S. & LEE, A. T. 1998. Retinoic acid induced mitogen-activated protein (MAP)/extracellular signal-regulated kinase (ERK) kinase-dependent MAP kinase activation needed to elicit HL-60 cell differentiation and growth arrest. *Cancer Res,* 58**,** 3163-72.

YOSHIKAWA, K., SHIMADA, M., MIYAMOTO, H., HIGASHIJIMA, J., MIYATANI, T., NISHIOKA, M., KURITA, N., IWATA, T. & UEHARA, H. 2009. Sonic hedgehog relates to colorectal carcinogenesis. *J Gastroenterol,* 44**,** 1113-7.

YOUNG, A. & REA, D. 2000. ABC of colorectal cancer: treatment of advanced disease. *BMJ,* 321**,** 1278-81.

YUEN, S. T., DAVIES, H., CHAN, T. L., HO, J. W., BIGNELL, G. R., COX, C., STEPHENS, P., EDKINS, S., TSUI, W. W., CHAN, A. S., FUTREAL, P. A., STRATTON, M. R., WOOSTER, R. & LEUNG, S. Y. 2002. Similarity of the phenotypic patterns associated with BRAF and KRAS mutations in colorectal neoplasia. *Cancer Res,* 62**,** 6451-5.

ZARAVINOS, A., LAMBROU, G. I., BOULALAS, I., DELAKAS, D. & SPANDIDOS, D. A. 2011. Identification of common differentially expressed genes in urinary bladder cancer. *PLoS One,* 6**,** e18135.

ZHANG, H. & SHAO, R. G. 2010. [G3BP: a promising target for cancer therapy]. *Yao Xue Xue Bao,* 45**,** 945-51.

ZHENG, P., LIU, Y. X., CHEN, L., LIU, X. H., XIAO, Z. Q., ZHAO, L., LI, G. Q., ZHOU, J., DING, Y. Q. & LI, J. M. 2010. Stathmin, a new target of PRL-3 identified by proteomic methods, plays a key role in progression and metastasis of colorectal cancer. *J Proteome Res,* 9**,** 4897-905.

ZHU, C. B., WANG, C. X., ZHANG, X., ZHANG, J. & LI, W. 2011. Serum sHLA-G levels: a useful indicator in distinguishing colorectal cancer from benign colorectal diseases. *Int J Cancer,* 128**,** 617-22.