DOCTOR OF PHILOSOPHY

Identification and characterisation of germline-associated genes as potential human cancer biomarkers

Sammut, Stephen

Award date:
2015

Awarding institution:
Bangor University

PRIFYSGOL
BANGOR
UNIVERSITY

Identification and characterisation of germline-associated genes as potential human cancer biomarkers

Ph.D. Thesis 2015

Stephen John Carmelo Sammut

Declaration and Consent

Details of the Work

I hereby agree to deposit the following item in the digital repository maintained by Bangor University and/or in any other repository authorized for use by Bangor University.

Author Name: ……..

Title: ……..…….

Supervisor/Department: .....

Funding body (if any): .....

Qualification/Degree obtained: …….

This item is a product of my own research endeavours and is covered by the agreement below in which the item is referred to as "the Work". It is identical in content to that deposited in the Library, subject to point 4 below.

Non-exclusive Rights

Rights granted to the digital repository through this agreement are entirely non-exclusive. I am free to publish the Work in its present version or future versions elsewhere.

I agree that Bangor University may electronically store, copy or translate the Work to any approved medium or format for the purpose of future preservation and accessibility. Bangor University is not under any obligation to reproduce or display the Work in the same formats or resolutions in which it was originally deposited.

Bangor University Digital Repository

I understand that work deposited in the digital repository will be accessible to a wide variety of people and institutions, including automated agents and search engines via the World Wide Web.

I understand that once the Work is deposited, the item and its metadata may be incorporated into public access catalogues or services, national databases of electronic theses and dissertations such as the British Library's EThOS or any service provided by the National Library of Wales.

I understand that the Work may be made available via the National Library of Wales Online Electronic Theses Service under the declared terms and conditions of use (http://www.llgc.org.uk/index.php?id=4676). I agree that as part of this service the National Library of Wales may electronically store, copy or convert the Work to any approved medium or format for the purpose of future preservation and accessibility. The National Library of Wales is not under any obligation to reproduce or display the Work in the same formats or resolutions in which it was originally deposited.

Statement 1:

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree unless as agreed by the University for approved dual awards.

Signed …….. (candidate)

Date ……..

Statement 2:

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).

All other sources are acknowledged by footnotes and/or a bibliography.

Signed ……. (candidate)

Date …….

Statement 3:

I hereby give consent for my thesis, if accepted, to be available for photocopying, for inter-library loan and for electronic storage (subject to any constraints as defined in statement 4), and for the title and summary to be made available to outside organisations.

Signed ……. (candidate)

Date …….

NB: Candidates on whose behalf a bar on access has been approved by the Academic Registry should use the following version of **Statement 3:**

Statement 3 (bar):

I hereby give consent for my thesis, if accepted, to be available for photocopying, for inter-library loans and for electronic storage (subject to any constraints as defined in statement 4), after expiry of a bar on access.

Signed …… (candidate)

Date ……

Statement 4:

Choose **one** of the following options

a) I agree to deposit an electronic copy of my thesis (the Work) in the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University and where necessary have gained the required permissions for the use of third party material.	
b) I agree to deposit an electronic copy of my thesis (the Work) in the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University when the approved **bar on access** has been lifted.	
c) I agree to submit my thesis (the Work) electronically via Bangor University's e-submission system, however I **opt-out** of the electronic deposit to the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University, due to lack of permissions for use of third party material.	

Options B should only be used if a bar on access has been approved by the University.

In addition to the above I also agree to the following:

1. That I am the author or have the authority of the author(s) to make this agreement and do hereby give Bangor University the right to make available the Work in the way described above.
2. That the electronic copy of the Work deposited in the digital repository and covered by this agreement, is identical in content to the paper copy of the Work deposited in the Bangor University Library, subject to point 4 below.
3. That I have exercised reasonable care to ensure that the Work is original and, to the best of my knowledge, does not breach any laws – including those relating to defamation, libel and copyright.
4. That I have, in instances where the intellectual property of other authors or copyright holders is included in the Work, and where appropriate, gained explicit permission for the inclusion of that material in the Work, and in the electronic form of the Work as accessed through the open access digital repository, *or* that I have identified and removed that material for which adequate and appropriate permission has not been obtained and which will be inaccessible via the digital repository.
5. That Bangor University does not hold any obligation to take legal action on behalf of the Depositor, or other rights holders, in the event of a breach of intellectual property rights, or any other right, in the material deposited.
6. That I will indemnify and keep indemnified Bangor University and the National Library of Wales from and against any loss, liability, claim or damage, including without limitation any related legal fees and court costs (on a full indemnity bases), related to any breach by myself of any term of this agreement.

Signature: …… Date : ……

Summary

It is increasingly apparent that cancer cells exhibit major intraclonal and interclonal heterogeneity that impacts on disease behaviour, response to treatments as well as clinical outcomes for patients. Understanding this diversity is critical to gain insights that ultimately influence treatment decisions in the era of personalised medicine. There is an increasing need to identify unique and relevant diagnostic and prognostic protein biomarkers to allow better stratification. Identification of aberrant gene expression patterns is one way in which additional insights into cancer causation can be discovered and can also inform novel therapeutic targets.

In this study we sought to focus on a group of proteins that are predominantly present in the germline but not in somatic cells. It is now well established that cancer cells exhibit aberrant expression of germline factors. Cancer testis antigens (CTAs) are one such family of proteins that exhibit increased presence in a variety of cancers and are also potentially immunogenic. As meiotic cell division is restricted to germline cells such as the testis, we hypothesised that identification and characterisation of genes that govern this process may yield additional clinically useful biomarkers that are also relevant to disease biology and behaviour.

With the aid of a pre-existing computational and bioinformatics approach, we first identified a subset of genes, some of which were likely to be involved in meiosis, and interrogated normal and cancerous cells to ascertain their expression pattern. This allowed us to identify a distinct cohort of candidate genes that appeared to be predominantly expressed in the germline and cancer. Immunohistochemistry using normal and (predominantly colorectal) cancer tissues was performed for further evaluation. Significantly, we identified two novel proteins (C20orf201 and TEX19) as putative meiosis-associated proteins that may have enhanced presence in cancer and the potential to be immunogenic.

In summary, this research into germline restricted genes in cancer *vs* normal tissues has identified two previously uncharacterised proteins that are likely to be relevant for the biology, behaviour or therapy of some cancers and has also identified a cohort of further genes that warrant further scientific exploration.

Acknowledgements

I am very grateful to many people who have helped me during the course of my studies. Firstly I thank my supervisors, Dr Ramsay McFarlane and Dr Jane Wakeman for their constant encouragement, wise and constructive advice and rigorous academic guidance. Nobody could have wished for better supervision. I am also especially indebted to Professor Nicholas Stuart for his mentorship and supervision. I am very grateful to Professor Michael Rees for his sound advice and firm support at the beginning of my studies and Professor Bharat Jasani for his mentorship and encouragement in the later stages of my research. Dr Steve Man gave very helpful input from an immunological perspective and provided friendly supervision of my work at Cardiff University. I am grateful for the expert input from several Pathologists, in particular Dr Mark Atkinson but also Dr Shawki Howida, Dr Mark Lord and Dr Susan Andrew, as well as Dr Anthony Caslin who sadly passed away during the course of this research.

My thanks go to all members of D7 lab at Bangor University, both past and present, who have shared scientific knowledge as well as good company with me along the way; particularly to Dr Natalia Gomez Escobar and Dr Ellen Vernon for experimental advice as well as Dr David Pryce who was always ready to offer friendly and generous discussion. A special mention goes to Vicente Planells Palop for being available for a second opinion, numerous discussions and much good humour; my thanks also go to Atif Oyouni for his good company whilst writing up. I would also like to acknowledge Thuraya Alzubaidi and Rui de Sousa in particular, for their assistance with processing of the pathological specimens.

I am grateful for the North West Cancer Research charity, as well as the Welsh Clinical Academic Training scheme, without whose funding this project would not have been possible.

And most importantly, I would like to thank my wife Pui and my children, for their love and support, not least when I needed it most.

List of abbreviations

ADC	Antibody drug conjugate
ALT	Alternative lengthening of telomeres
APC	Antigen presenting cell
BCA	Bicinchoninic acid assay
BLAST	Basic local alignment search tool
BTB	Blood-testis barrier
CDK	Cyclin-dependent kinase
cDNA	Complimentary DNA
CE	Central element
cfDNA	Cell free DNA (or circulating free DNA)
CIMP	CpG island methylator phenotype
CIN	Chromosomal instability
CNS	Central nervous system
CRC	Colorectal cancer
CSC	Cancer stem cell
CT	Cancer/testis
CTA	Cancer/testis antigen
CT-X	Cancer/testis genes encoded by the X-chromosome
DAB	3,3'-Diaminobenzidine
DAPI	4',6-diamidino-2-phenylindole
dH_20	Distilled H_2O
DMSO	Dimethyl sulfoxide
DSB	Double-strand break
DSBR	Double-strand break repair
EBV	Epstein-Barr virus
ECL	Enhanced chemiluminescence
ECM	Extraceullar matrix

EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
EMT	Epithelial-mesenchymal transition
EST	Expressed sequence tag
FACS	Fluorescence activated cell sorting
FAP	Familial adenomatous polyposis
FBS	Fetal bovine serum
FH	Follicle-stimulating hormone
GI	Gastrointestinal
GnRH	Gonadotropin-releasing hormone
HGNC	HUGO Gene Nomenclature Committee
HLA	Human leucocyte antigen
HNPCC	Hereditory non-polyposis colorectal cancer
HPA	The Human Protein Atlas
HPV	Human papilloma virus
H.pylori	*Helicobacter pylori*
HR	Homologous recombination
IHC	Immunohistochemistry
iPSC	Induced pluripotent stem cell
LDS	Lithium dodecyl sulfate
LE	Lateral element
LH	Luteinizing hormone
MA	Microarray
MAGE	Melanoma associated antigen
meiCT	Meiosis-associated cancer/testis genes
MET	Mesenchymal-epithelial transition
MFI	Mean fluorescence intensity
MHC	Major histocompatibility complex
MMR	Mismatch repair

M-PER	Mammalian protein extraction reagent
MRN	Mre11, Rad50 and Nbs1 protein complex
mRNA	Messenger RNA
MSI	Microsatellite instability
NAT	Normal adjacent tumour
NCBI	National center for biotechnology information
NHEJ	Non-homologous end joining
PBS	Phosphate buffered saline
piRNA	PIWI-interacting RNA
PVDF	Polyvinylidene fluoride
qPCR	Quantitative polymerase chain reaction
RQI	RNA quality indicator
RT-PCR	Reverse transcription polymerase chain reaction
SDSA	Synthesis-dependent strand annealing
SC	Synaptonemal complex
SSC	Spermatogonial stem cell
ssDNA	Single-stranded DNA
TAA	Tumour-associated antigen
TCR	T-cell receptor
TE	Transposable elements
TGF-β	Transforming growth factor-β
TIL	Tumour-infiltrating lymphocyte
TLDA	Taqman low density array
TMA	Tissue microarray
TNM	Tumour-node-metastasis
UNG	Uracil-N glycosylase
WCE	Whole cell extract

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1. Cancer

1.1.1. The burden of cancer

Cancer is one of the most prevalent causes of human mortality and the proportion of deaths from the disease is projected to increase to an extent it may collectively become the commonest cause (Mathers and Loncar, 2006; Seigel *et al*., 2015). Cancer is predominantly but not exclusively a disease of old age and as human longevity improves so the levels of cancer increase but the burden of disease varies across different regions of the World, impacted to some degree by economic prosperity (Mathers *et al.,* 2009; Siegel *et al.,* 2015). There are important gender differences in cancer incidence but, overall, colorectal cancer is the second most common cancer diagnosed in the UK and ovarian cancer is the fifth most common cause of cancer-related death in women after breast, lung, colorectal and pancreatic cancer (Hossack, 2012; Seigel et al 2015).

1.1.2. Cancer definition and oncogenesis

The diverse functions within the human body require the coordinated expression of distinct genes in time and place; a process governed by homeostatic mechanisms occurring throughout the body. Cancer, as with other diseases, can be seen as a failure of homeostasis, where cells divide uncontrollably and can develop the ability to invade and metastasise to other regions of the body. Although cancer usually arises from a primary site or group of cells, the process of oncogenesis (or carcinogenesis) is a drawn out process, where the cells acquire a sequential series of changes (e.g., genetic and epigenetic alterations) that ultimately lead to the cancer phenotype (Stratton *et al.*, 2009). This process typically takes place over a period of years and is the main reason cancer becomes more common as we get older. Exceptions to this general 'rule' are the few cancers which can be considered to arise from a single (genetic) event: for example, NUT midline carcinomas and many paediatric cancers fall into this group (French, 2012; Horton *et al.*, 2007; Pui *et al.*, 2011). In addition the occurrence of chromothripsis has also been suggested to occur in up to 3% of human malignancies, whereby the vast array of genetic alterations can be accounted for by a single catastrophic genetic event, rather than a progressive series of alterations (Stephens *et al.*, 2011).

1.1.3. Hallmarks of cancer

Several distinct phenotypes have been described for cancers, which are considered necessary for cancer formation. In a seminal review six 'hallmarks' of cancer were described – see Figure 1.1 (Hanahan and Weinberg, 2000). Four additional characteristics have now been added to this list – see Figure 1.2 (Hanahan and Weinberg, 2011). Cancer cells must be able to: evade apoptosis, growth suppressive signals and immune attack; sustain proliferative signals and replicate indefinitely; invade surrounding tissues and induce angiogenesis in order to grow, spread and metastasize; and adjust their metabolic cellular machinery to serve their cause.

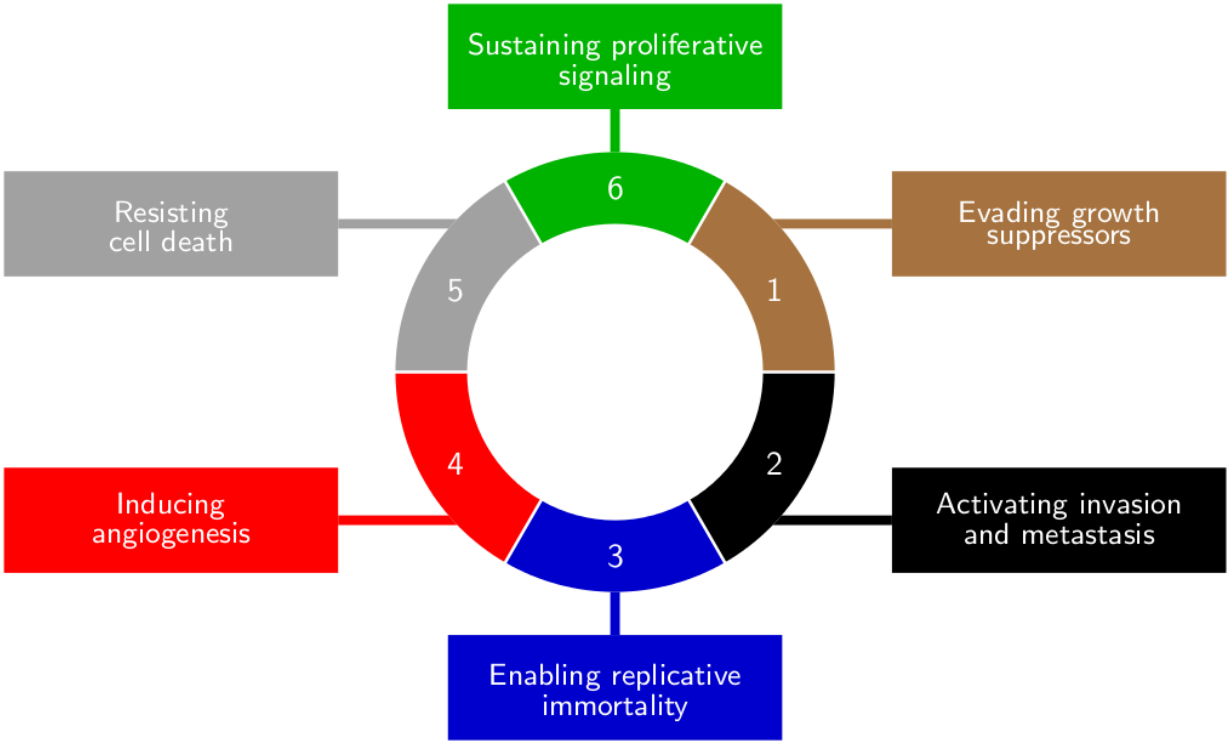


Figure 1.1. The original six hallmarks of cancer proposed in 2000.
These hallmarks can be seen to arise from either increased or abnormal function of genes that directly promote these attributes, or reduction in the function of genes that guard against their development. Adapted from Hanahan and Weinberg (2011).

The process of metastasis, whereby a cell or group of cells escape from the primary tumour and travel to distant sites within the body where they set up 'colonies' is a leading cause of death from cancer. Understanding the key drivers or events that lead to development of these attributes and/or developing treatments which target multiple hallmarks is important. Two key drivers of the oncogenic process, which can also be considered additional 'hallmarks', are depicted in Figure 1.2 – these are genomic instability/mutability and tumour-promoting inflammation (Hanahan and Weinberg, 2011; Mantovani *et al.*, 2008; Negrini *et al.*, 2010).

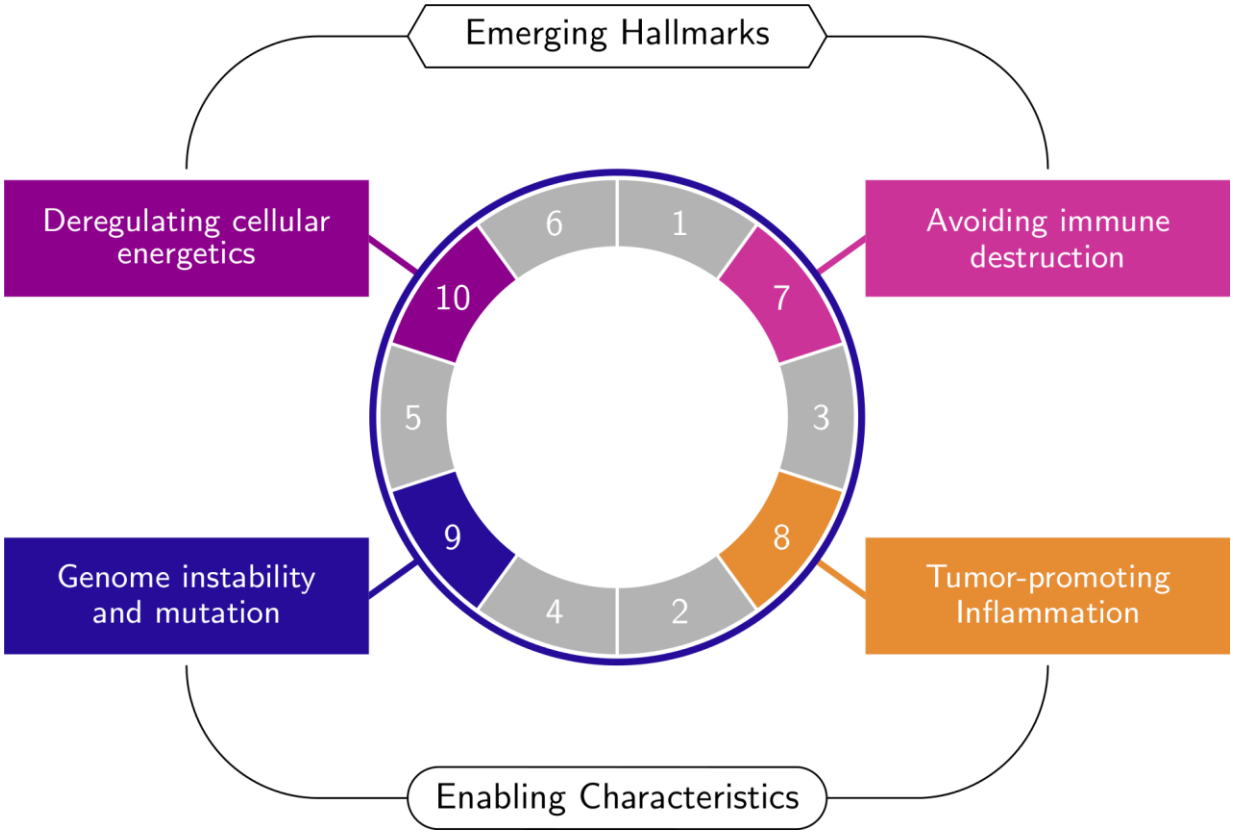


Figure 1.2. The additional hallmarks and enabling characteristics of cancer.
Four new features (7-10 in the above figure) of cancer were added to the original six hallmarks (labelled 1-6 in the above figure) that had previously been described (see Figure 1.1). Deregulation of cellular metabolism and avoiding immune destruction are considered as additional hallmarks; whereas genomic instability and cancer-promoting inflammation are considered as mechanisms to promote or enable the other hallmarks to develop. Adapted from Hanahan and Weinberg (2011).

Many of these attributes or hallmarks of cancer can be ascribed to the activation of tissue-specific genes; a fact that should aid the development of new treatment strategies (Wang *et al.*, 2014). There is little doubt that evading immune attack is of fundamental importance for most, if not all, cancers. There are various immunotherapeutic approaches, which aim to reverse what can been seen as the immune system's failure in preventing cancer formation; some of these will be discussed later in this chapter. Such approaches are often targeted at the products of tissue specific genes, which are activated out of context and in tissues in which there are not normally found (Hunder *et al.*, 2008; Robbins *et al.*, 2015; Wang *et al.*, 2014).

1.1.4. Classification

The most common type of cancers are the carcinomas. These are cancers that arise from epithelial cells, or the cells lining the internal viscera. Epithelial cells are not only a common cell type, but as they line the viscera, they are frequently exposed to compounds that could potentially initiate genetic/epigenetic change. Two illustrative examples of this are: lung cancer, which is more common in people who smoke; and skin cancer, whose risk is increased by increased exposure to the Sun's ultraviolet ionising (i.e. free-radical promoting and DNA damaging) radiation. Other types of cancer are the sarcomas, which arise from the mesenchymal tissue. The mesenchyme, or more specifically the mesenchymal stem cells, can differentiate into muscle, fat, nerves, cartilage, tendons, bone, blood vessels and the fibrous connective tissue that supports other cells. Sarcomas only make up 1% of all cancers but over 10% of cancers in children and young adults (Bleyer *et al.*, 2008). Haematological malignancies make up the final group of cancers and include: myeloma (which arises from the plasma cells in the blood), the myeloproliferative disorders (overproduction of mature blood cells), leukaemias (which arise from the white blood cells) and lymphomas (which arise from lymphoid tissue).

1.1.5. Genes and cancer development

The differentiation and division of cells is tightly controlled to maintain tissue homeostasis and normal cellular and as a result visceral, structure and function. Three classes of genes (tumour suppressor, oncogenes and genomic stability genes) can be considered as contributing to cancer development, through genetic mutations, chromosomal rearrangements or overexpression, which disrupts their normal function. Firstly, tumour suppressor genes (or anti-oncogenes) encode proteins that inhibit cell proliferation and/or induce apoptosis. It can be seen that loss or reduction of tumour suppressor gene function can lead to deregulated and uncontrolled proliferation of cells, which is a hallmark of cancer. Proto-oncogenes are normal genes that produce proteins that promote cellular proliferation; they can through a genetic change become an oncogene. Oncogenes have the potential to also promote uncontrolled cellular proliferation by overriding the inhibitory signals (i.e. of tumour suppressor gene products) in the cellular milieu. A third class of genes, the genomic stability genes, can be viewed as a subcategory of tumour suppressor genes. These genes are involved in DNA repair and chromosome recombination and contribute to tumourigenesis by failure to maintain genome integrity; their

function is also altered by genetic or epigenetic change. So, oncogenes promote cancer development, and tumour suppressor genes along with genome stability genes guard us against its formation (Hanahan and Weinberg, 2011; Vogelstein and Kinzler, 2004). Table 1.1 summarises examples of well-characterised tumour suppressor genes and Table 1.2 shows examples of some oncogenes.

Table 1.1. Examples of tumour suppressor and genome stability genes and their associated cancer types.

Gene	Function	Common cancer types
APC	Promotes beta-catenin degradation, thus influencing cell adhesion and transcriptional activation of other genes (Aoki and Taketo, 2007)	Colorectal carcinomas; Desmoid tumour
BRCA1	DNA repair and maintaining genome stability (Silver and Livingston, 2012)	Breast and ovarian carcinomas
RB1	G1 checkpoint control inhibiting cellular proliferation (Giacinti and Giordano, 2006)	Retinoblastoma; sarcoma; breast, lung and bladder carcinomas
TP53	Cell cycle arrest; autophagy regulation; induction of apoptosis (Ryan, 2011)	Broad range of cancer types – mutation present in ~50% of cancers
WT1	Transcriptional regulator, influences RNA metabolism and MET (Hohenstein *et al.*, 2015)	Wilms' tumour
PTEN	Acts as a phosphatase to influence cell signalling and thus influence cellular proliferation, migration and apoptosis; also genome stability (Milella *et al.*, 2015)	Melanoma; brain tumours; lung endometrial; prostate and kidney carcinomas

MET – mesenchymal-to-epithelial transition

Knudson's two hit hypothesis was proposed over 40 years ago in regard to the *RB1* gene, whereby there needs to be a loss of function in both alleles to promote cancer development (Knudson, 1971). Such loss of function mutations can be inherited, that is through the germline, or acquired after birth and can be referred to as a loss of heterozygosity. It remains a useful way of seeing how inherited genetic factors can combine with environmental factors to result in cancer.

Table 1.2 Examples of oncogenes, their broad functions and associated cancer types.

Gene	Function	Common cancer types
SHH	Evading apoptosis and promoting proliferation (Ng and Curran, 2011)	Medulloblastoma and numerous carcinomas
*KRAS**	GTPase signal transduction promoting cell proliferation (Malumbres and Barbacid, 2003; Vale *et al.*, 2012)	Lung, pancreatic and intestinal carcinomas
MYCBP	Transcription factor that promotes cell proliferation (Vafa *et al.*, 2002)	Numerous types of carcinomas and central nervous system tumours
NOTCH1	Inter-cellular communication; evasion of apoptosis and cell proliferation. Also, tumour suppressor functions (Radtke and Raj, 2003)	Skin cancer among other cancer types
FOS	Forms part of Activator Protein-1 complex; acts as a transcription factor to promoting cell proliferation and evasion of apoptosis (Milde-Langosch, 2005)	Sarcomas and numerous types of carcinoma.
*ERBB2**	Cell surface receptor that triggers cell proliferation through tyrosine kinase activity and other mechanisms (Moasser, 2007)	Breast and ovarian carcinomas

*Currently in use as biomarkers. ERBB2 (also known as HER2) is used as a prognostic indicator for breast cancer treatment (Constantinidou and Smith, 2011) and likewise *KRAS* mutation testing guides which patients will benefit from anti-EGFR therapy in colorectal cancer (Vale *et al.*, 2012).

1.1.6. The heterogeneous nature of cancer

It has been known for some time that different forms of cancer, despite common characteristics, are very varied in their biological make-up. However, it has only been relatively recently that we have fully appreciated the extensive heterogeneity of human cancers (Swanton *et al.*, 2010). There can be huge differences, both genetic and epigenetic, not only between cancers of the same type but even within individual tumours. Such intra-tumour or intraclonal heterogeneity has far reaching implications for cancer diagnosis and treatment. In terms of predictive biomarkers for instance, do we try and predict the outcome for patients from a single biopsy? New approaches to cancer diagnosis, such as biopsy techniques and/or resected specimen processing, will most likely become increasingly important and more widely used in the future to meet the demands of personalised medicine and deal with the analysis of tumour heterogeneity that will be required. The use of functional imaging is likely to become of increasing importance and may prove a useful way of targeting biopsies as well as characterising tumours pre-operatively (Schmitz *et al.*, 2016).

The ideal cancer biomarker would be easily detectable and present either exclusively or to a vastly different degree than in normal tissues, making it both sensitive and specific for disease. There is an additional advantage of biomarkers that are present exclusively in the cancer, in that they may also be the target for (immune-based) therapies. If the targets of such treatments are not present in healthy tissues in the body and specific targeting occurs, then no or minimal side effects would be expected. Histological characteristics, and cell-surface markers, commonly used to stratify patients into 'risk' categories at present often do not adequately reflect the degree of tumour heterogeneity. New biomarkers are required to help overcome this problem. Another, perhaps surprising fact, is that the gene expression profiles of tumours remain fairly stable from early to advanced stages of disease (Visvader, 2011). Tumour heterogeneity and immune-based cancer treatments will be discussed again later, with emphasis on colorectal cancer, but the importance of cancer stem cells and how they relate to tumour heterogeneity will now be highlighted in the following section.

1.1.6.1. The importance of cancer stem cells

In normal physiology, stem cells are specialist cells which have the ability to give rise to cells that constitute an organ. Stem cells can be quiescent but their activation following injury is of great importance in restoring homeostasis and organ function following injury or damage due to disease. When a sperm cell fertilises an ovum, the initial cell (and those of the first few cell divisions) is totipotent; that is these cells can differentiate to produce an entire organism and the extra-embryonic tissue of a trophoblast (Jaenisch and Young, 2008). Adult stem cells are often capable of differentiating into different cell types that make up an organ or tissue (i.e. they are multipotent) but they can be bipotent or unipotent (Visvader and Clevers, 2016). Stem cells and their progenitor cells act in a hierarchical manner to produce differentiated cells that constitute an organ. For example, an intestinal stem cell will differentiate into the various cells of the intestinal crypt, or a liver stem cell will help reconstitute the liver parenchyma. The same concept is believed to apply to cancer; that is a small proportion of the cells within a given tumour have the ability to maintain and regenerate the cancer. These cells are termed cancer stem cells (CSCs) or cancer stem-like cells (Plaks *et al.*, 2015).

The function of adult stem cells is incompletely understood in normal physiology; the situation in cancer and CSC function is even more complex and indeed continues to be debated. The cell which initially undergoes oncogenic genetic change (i.e. the 'cell of origin'), may not be related to the CSC: indeed, this is believed to be the case for the majority of cancers (Visvader, 2011).

The cell of origin can be viewed as the cancer initiating event, whereas the CSCs are the cancer-sustaining cells. Any cell within the cellular hierarchy (of stem cells and their progenitors) can potentially act as a cell of origin for cancer. As reviewed by Visvader, the majority of the proposed cells of origin for different cancers are progenitor cells (Visvader, 2011). There is a rapid turnover of intestinal crypt cells, thus the progenitor cells may not exist for long enough to acquire the mutations necessary to result in cancer. Consistent with this, the cells of origin in intestinal cancer are in fact stem cells, which acquire *APC* and Wnt signalling gene mutations (Visvader, 2011). Difficulties in establishing the cells of origin in cancer are in part due to methodological challenges but there may also be numerous cells of origin that lead to a situation where a self-sufficient tumour exists. There may also be numerous CSCs within a tumour and this can contribute to tumour heterogeneity and at the same time pose great challenges for successful cancer treatment (see Figure 1.3).

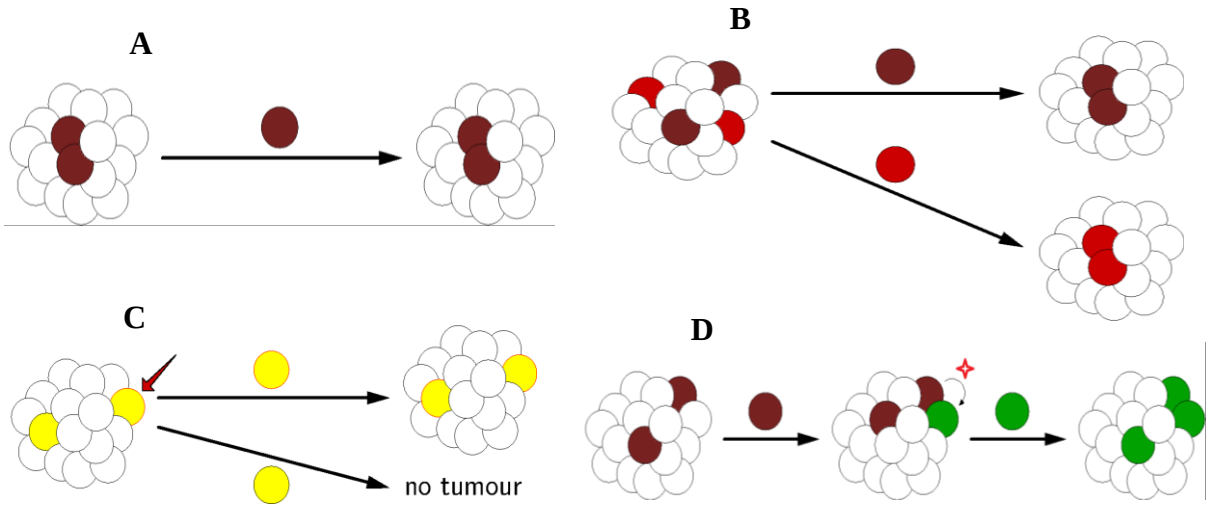


Figure 1.3. The contribution of cancer stem cells to tumour generation and heterogeneity.
There numerous ways in which cancer stem cells (CSCs) can contribute to tumour heterogeneity and tumour generation. There may be a single CSC subset and non-CSCs are not capable of tumour generation (A). There may be several distinct CSCs within the tumour, each capable of clonal expansion and independently able to generate the tumour (B). There may be CSCs which have been dormant for some time which become reactivated following an activation signal depicted by red arrow (C). These reactivated CSCs could result in local or distant tumour recurrence following activation, potentially many years following initial treatment. Finally, as the tumour grows and the cancer progresses, existing CSCs may undergo further genetic or epigenetic change (indicated by red star) resulting in distinct subsets of CSCs (D). Clonal expansion of the more aggressive CSC (depicted by green circle) will become the dominant driver for tumour generation and resistance to further therapy. In addition, changing of cell surface markers may result in a change of CSC phenotype. Adapted from, Visvader and Lindeman (2012).

As the vast majority of chemo-therapeutic strategies are aimed at rapidly dividing cells, this causes a major problem for successful cancer treatment as CSCs may be quiescent or slow-cycling and protected by the niche in which they reside (Borovski *et al.*, 2011; Plaks *et al.*, 2015; Visvader and Lindeman, 2012). The agents used are often successful at reducing the tumour burden (i.e. killing the rapidly dividing cells which make up the bulk of many solid tumours) but the CSCs may still survive and retain the ability to regenerate the cancer following therapy. Recurrent tumours are often altered in that they are more resistant to the initial therapy as cellular selection or tumour adaptation, via genetic or epigenetic change of an existing CSC for example, has taken place (see again Figure 1.3). Thus, targeting CSC is of vital importance in successful cancer treatment, especially if they have recurred and/or are refractory to surgical removal.

The mechanism of CSC formation remains poorly understood. There are two principal possibilities: either a differentiated cell dedifferentiates and regains stem-like properties, or a normal adult stem cell is transformed into an abnormal (cancerous) one. The dedifferentiating model is more widely accepted, not least as pluripotent stem cells can be induced from terminally differentiated cells via epigenetic manipulation (Friedmann-Morvinski and Verma, 2014). Such induced pluripotent stem cells (iPSC) have many potential applications, not only in understanding cancer biology but also in the emerging field of regenerative medicine (Matsa *et al.*, 2014). Others have argued more strongly in favour of the latter model of carcinogenesis in which oncogenes are seen as the cellular 'reprogammers' (Vicente-Duenas *et al.*, 2013). It is possible that there is an overlap between the theories of initiating events or differences between cancer types but targeting CSCs is undoubtedly important if dramatic improvements in cancer survival are to be achieved. In addition to understanding and therapeutically overcoming the mechanisms by which CSC survive (Flemming, 2015; Pattabiraman and Weinberg, 2014; Plaks *et al.*, 2015), further complexity may be revealed by the existence of multiple CSCs within any given tumour.

Before going on to introduce cancer biomarkers and how they are linked to the identification of novel targets for cancer therapy, the concept of field cancerization will be mentioned. This concept has implications for biomarker discovery (e.g., when comparing 'normal' to cancerous tissue) and relevance to tissue sampling, which formed a substantial part of the work of this research. The concept of field cancerization is also inextricably linked to CSCs and tumour heterogeneity through the clonal expansion theory.

1.1.6.2. Field Cancerization

The concept of field cancerization (or field carcinogenesis) was first proposed over half a century ago (Slaughter *et al.*, 1953) and has evolved considerably over the years (Frede *et al*., 2014; Lochhead *et al*., 2015). It is believed to result from the clonal expansion of a cell that has acquired mutations in tumour suppressor and/or proto-oncogenes. The resulting cell population from this 'rogue' cell can cover a large area. Although histologically the tissue looks the same as normal tissue, there is a (genetic) predisposition within it to develop cancer. The concept can apply to an area around a solid established tumour as well as a premalignant state before cancer has developed. It helps explain the existence of synchronous tumours that can occur in many organs and it may also be the cause of cancer recurrence following surgery even when the surgical margins were 'clear' of cancer. There is growing evidence that field cancerization is important in the development of CSCs (Frede *et al.*, 2014). For example, in the intestine, crypt fission events may be tipped in favour of cancer-promoting lineage (containing *KRAS* mutations) through mutations in the stem-cell compartment, leading to clonal expansion across a potentially large area of mucosa (Snippert *et al.*, 2014). When considering colorectal cancer, field changes have been seen over 17 cm from the primary tumour (Dawson *et al.*, 1987) but there is probably a tapering effect in that more mutations are present, or the genotype more closely resembles that of the cancer, the closer you move towards the cancer (Hawthorn *et al.*, 2014). The evidence for this tapering is limited and requires clarification by future research. If clonal expansion occurs from mutations in the stem cell compartment, then there would need to be multiple 'fields' within the area of field change for the tapering effect to take place.

1.1.7. Cancer prevention, stratification and the importance of biomarkers

Understanding the causative factors for different cancers is key to prevention. Reducing environmental exposure to known carcinogens forms the basis for many health prevention strategies, such as the "Stop Smoking Wales" campaign. Vaccination against viruses known to be carcinogenic is another approach, which can be considered a form of anti-cancer immunotherapy (Cohen *et al.*, 2013; Frazer, 2004). In addition to preventative strategies, cancer mortality rates can be further reduced by early diagnosis. This requires tests that can be used to diagnose cancer.

A biological marker, or biomarker, has been defined as: *"A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic intervention"* (Biomarkers Definitions Working Group., 2001).

Cancer-specific biomarkers can be used to facilitate diagnostic testing. Other cancer biomarkers may be better for monitoring response to treatment and/or predicting the likely natural course of the disease and thus guide which treatment strategy is most appropriate. A clinical endpoint is a characteristic or variable that reflects how a patient feels, functions or survives (Biomarkers Definitions Working Group., 2001). Some cancer biomarkers will have application as surrogates for clinical endpoints, which can be used to assess their suitability in the clinical setting or how effective new treatment strategies are. This is particularly important when the actual clinical endpoint, e.g., mortality or cancer recurrence, may occur after a long period of time. Developing good surrogate markers is tremendously useful in the realm of clinical translational research and also forms the basis of personalised treatment strategies.

Good diagnostic biomarkers also form the basis of many cancer screening programmes, which are aimed at diagnosing cancer in the stage before it has become symptomatic. Screening tests can be aimed at high risk patient groups or an entire subset of the population (e.g., within a given age range). A good screening test should be reproducible, preferably non-invasive, cost-effective and both sensitive and highly specific for the disease (Smith *et al.*, 2015). Even a slight reduction in specificity could lead to treatment and/or tests at significant costs, as well as anxiety and potential harm to the patient.

So, in addition to diagnosing and screening patients for the presence of cancer, cancer biomarkers can also be used for staging, prognostication and guiding therapy. Despite the efforts to establish useful cancer biomarkers, relatively few are in common clinical use and the path to discovery is typically a long one (Ludwig and Weinstein, 2005; Pavlou *et al.*, 2013). Table 1.3 provides a list of some biomarkers that are in current clinical use; there remain fewer than 30 cancer biomarkers in common clinical use. The stages of biomarker development can be split into pre-analytical, analytical, statistical or bioinformatics input, validation and then finally commercialisation and approval (Diamandis, 2014; Ludwig and Weinstein, 2005; Pavlou *et al.*, 2013). Reasons for the failure to develop more biomarkers are multifactorial and lie at every stage of the biomarker development process. The significant costs throughout this process are also a hindrance. The search for new biomarkers is as important as ever, especially

as we move increasingly towards an era of personalised cancer treatments (Aris and Barrio, 2015; Kraus *et al.*, 2014). Indeed, predictive biomarkers form the basis for the development of personalised treatments; at present breast cancer treatment is probably the best example of where treatment is tailored depending on tumour characteristics (Cho *et al.*, 2012).

1.1.7.1. Sources and development of cancer biomarkers

Tumour biopsy, urine, sputum, ascitic fluid, faeces, buccal mucosa or airways brushings and even exhaled breath can all be used as a source of biomarker detection (Hensing and Salgia, 2013; Srinivas *et al.*, 2001). Blood is often the ideal source and many of the existing biomarkers in clinical use are proteins detected in the serum – see Table 1.3. Protein biomarkers can be cell surface receptors, tumour antigens, proteins with post-translational modifications, as well as peptides released or excreted into or from the body (Ludwig and Weinstein, 2005).

Table 1.3. Examples of biomarkers in current clinical use

Biomarker	Type	Biological source, cancer type and clinical use
AFP	Glycoprotein	Serum test for staging of testicular cancer
BRAF	DNA mutation analysis	Pathological tumour specimen. Predicts response to BRAF inhibitors in patients with melanoma.
CA19-9	Protein	Serum test for monitoring pancreatic cancer
CA125	Glycorotein	Serum test for monitoring and stratifying risk of ovarian cancer
CEA	Protein	Serum test for monitoring colorectal cancer
ER + PR	Proteins	Pathological tumour specimen. Prognostic indicator and therapy selection for patients with breast cancer
HER2 (*ERBB2*)	Protein (IHC) + DNA (FISH)	Pathological tumour specimen. Prognostic indicator and therapy selection for patients with breast cancer
KRAS	DNA mutation analysis	Pathological tumour specimen. Prognostic indicator and therapy selection for patients with colorectal cancer
PSA	Protein	Serum test for screening and monitoring prostate cancer

AFP - α-fetoprotein; CA – cancer antigen; CEA – carcinoembryonic antigen; PSA – prostate specific antigen; HER2 – human epidermal growth factor receptor 2; *KRAS* – gene that encodes the KRAS protein; ER – oestrogen receptor; PR – progesterone receptor; *BRAF* – gene that encodes the B-Raf protein. (Gonzalez de Castro *et al.*, 2013; Ludwig and Weinstein, 2005; Pavlou *et al.*, 2013).

The rapid development of various "-omics" techniques has expanded the possibilities of developing other types of cancer biomarkers; though few have come to clinical fruition they still hold great promise. Such alternate biomarkers include: DNA- and RNA-based changes (including microRNAs and epigenetic alternations such as DNA methylation), autoantibody detection, and circulating tumour cells or free DNA (cfDNA) (Gonzalez de Castro *et al.*, 2013; Hensing and Salgia, 2013; Ludwig and Weinstein, 2005; Yoon *et al.*, 2014). Some have advocated 'multiplexing' or using pattern-based biomarkers, whereby several biomarkers are combined to increase the diagnostic or stratification accuracy (Ludwig and Weinstein, 2005;

Ullah and Aatif, 2009). In a similar vein, multiple tumour antigens, which can themselves be biomarkers, can be targeted simultaneously to improve the efficacy of immunotherapeutic approaches.

Tumour antigens and immunotherapy will now be discussed. It can be seen how tumour antigens often overlap with biomarker discovery and research, as cell surface markers that enable cancer to be detected and recognised via screening or diagnostic molecular tests, may also be the very same targets for anti-cancer pharmaceuticals.

1.2. Tumour antigens and immunotherapy

Abnormal proteins arising and/or occurring in increased abundance as a result of cancer that are capable of eliciting an immune response are collectively termed tumour-associated antigens (TAAs). Genetic or epigenetic changes in any of the three tumourigenic gene classes described above may result in a TAA. Peptides of TAAs have the ability to induce the immune system as they may not be recognised as 'self' antigens. Harnessing the body's own immune-system to fight cancer is a very attractive anti-cancer treatment modality (Mellman *et al.*, 2011). To be useful for cancer immunotherapy, TAAs should ideally be: 1) restricted to the tumour and not present in healthy tissues; 2) present throughout the relevant cancer or cancer type predictably (i.e. not in a sporadic or hap-hazard fashion); and 3) targeted by cytotoxic T lymphocytes (Krishnadas *et al.*, 2013). The identification and characterisation of TAAs has become a large area of research in the field of cancer immunotherapy (Mellman *et al.*, 2011). The third attribute of TAA, being able to stimulate an immune response by activating T-cells, is crucially important when it comes to developing immunotherapies to target cancer (Coulie *et al*., 2014; Krishnadas *et al*., 2013).

1.2.1. Types of tumour-associated antigens

Human TAAs can be classified into different types according to their presence or patterns of gene expression within the body (Neller *et al.*, 2008; Srinivasan and Wolchok, 2004):

- Overexpressed (and universal antigens such as telomerase and survivin) are peptides derived from over-expressed genes that are found to be associated with T-cell responses. The low level of expression in normal cells should not stimulate the immune system to recognise these antigens.

- Oncofetal antigens are present during embryonic development but not present in adult tissues.

- Differentiation antigens are present in specific tissues and associated with a particular state of differentiation.

- Cancer/testis antigens (CTAs) are found in the normal testis and malignant tumours. Because of their unique expression pattern, CTAs are becoming useful biomarkers as well as targets for immunotherapy such as cancer vaccines (Krishnadas *et al.*, 2013; Whitehurst, 2014). This class of TAA will be discussed in more detail below.

- Some gene mutations result in the translation of proteins with an altered amino-acid sequences that result in cancer-specific antigenic peptides capable of eliciting an immune response. The genetic change may or may not be in a tumour suppressor gene or proto-oncogene; the important point being that the immune system will not have 'seen' the protein before. These unique antigens are also referred to as neoantigens (Schumacher and Schreiber, 2015).

- Viral antigens or antigens occurring as a result of an infectious agent that can cause cancer are also considered an additional class of cancer antigens.

1.2.2. Immunotherapeutic approaches

Globally, infections are thought to result in around 2 million cases of cancer each year (de Martel *et al.*, 2012). Many of these cases are related to viral infections, such as human papilloma virus (HPV), Epstein-Barr virus (EBV) and viral induced hepatitis. Vaccination against such viruses is a highly effective method of cancer prevention and a vaccination programme against HPV has recently been introduced in the UK to protect women against cervical cancer. Therapeutic vaccines directed against TAAs have had limited success in treating cancer. It is believed part of this limited success is due to targeting patients with advanced and metastatic disease; vaccines may hold the greatest benefit at halting the progression of pre-malignant or early stage disease (Gray *et al.*, 2008). However, patients with early stage disease can often be treated successfully by alternate means and immunotherapeutic approaches are unfortunately often a last and/or experimental resort.

Allogeneic bone marrow transplantation is a well-established and effective treatment of certain haematological cancers. This can be considered a form of immunotherapy as it is thought to involve an immediate immune reaction against the residual cancer through the action of antibodies within the donor cells (Mellman *et al.*, 2011). Adoptive therapy, or adoptive cell transfer, follows a similar principal but here autologous cells are used (Essand and Loskog, 2013; Gattinoni *et al.*, 2006). The patient's own lymphocytes are cultured and expanded *ex vivo* before being transferred back into the patient to permit auto-immune destruction of the cancer – see Figure 1.4. Dramatic results have been achieved in some instances, most notably in malignant melanoma (Hunder *et al.*, 2008) but also in subsets of patients with other tumour types, for example, cervical cancer (Stevanovic *et al.*, 2015).

Monoclonal antibodies targeting TAAs have been widely investigated, with some agents gaining regulatory body approval for specific patient groups (Luke and Hodi, 2013; Mellman *et al.*, 2011). Recombinant cytokines have also been shown to be effective in treating cancer in patients with immunogenic tumours and who are predisposed to autoimmunity. Predicting or identifying which patients will respond best to this form of immunotherapy, however, has proved not to be straightforward (Mellman *et al.*, 2011). Antibody drug conjugates (ADCs) are an additional example of a personalised immunotherapeutic approach (Leal *et al.*, 2014). Here, a drug conjugate containing a toxic payload is attached using a linker molecule to an antibody that specifically recognises cancer cells. The toxic payload of the drug conjugate is internalised only by the (cancer) cells which contain cell surface antigens recognised by the antibody. This limits potentially unpleasant systemic toxic side effects and in fact enables the use of substances that it would be too harmful if administered via traditional routes. An example of a toxic substance that has been used in an ADC approach with varying success in a number of cancer types is monomethyl auristatin E (MMAE) (Bendell *et al.*, 2014; Ott *et al.*, 2014; Younes *et al.*, 2012).

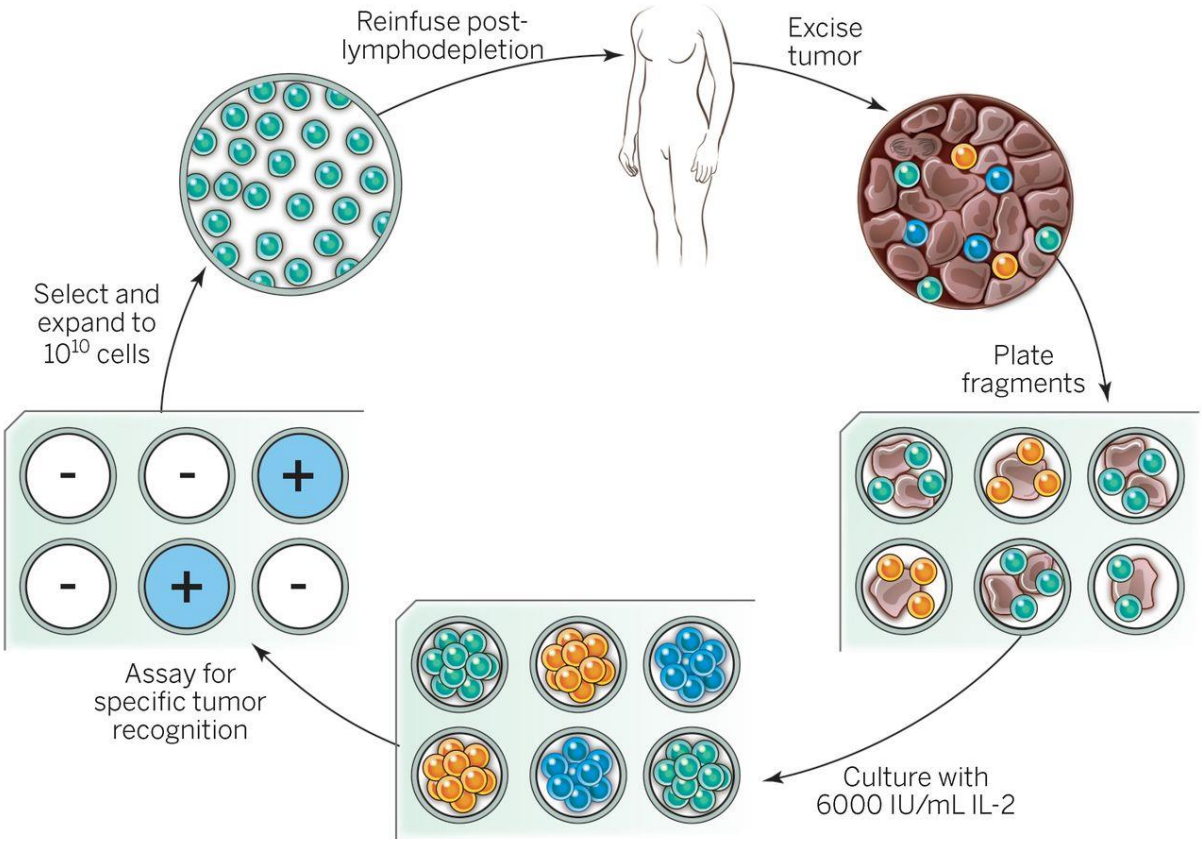


Figure 1.4. Schematic representation of adoptive therapy.
Tumour infiltrating lymphocytes are cultured and expanded *ex vivo* and transferred back into the patient after they have received non-myeloablative chemotherapy (Gattinoni *et al.*, 2006; Rosenberg and Restifo, 2015). IL-2 – Interleukin-2.

1.2.3. Immunosuppression within cancer

For an effective immune response to occur there needs to be antigen presentation by antigen presenting cells (APCs) such as dendritic cells, followed by a T-cell response against this antigen – see Figure 1.5. In addition the immunosuppressive signals, produced by the cancer, need to be overcome. These signals are varied and multifactorial, resulting in highly effective evasion of the immune system's surveillance of the tumour and its microenvironment. For example, paracrine mediators and cytokines released by cancer may directly suppress the APCs or inhibit access of the T-cells to certain regions of the tumour (Mellman *et al.*, 2011). Immune checkpoints also exist on the surface of T-cells, which have the effect of limiting their proliferation or cytotoxic capabilities (Nirschl and Drake, 2013; Topalian *et al.*, 2015). Overcoming this immunosuppression is a major challenge in the development of immunotherapies against cancer.

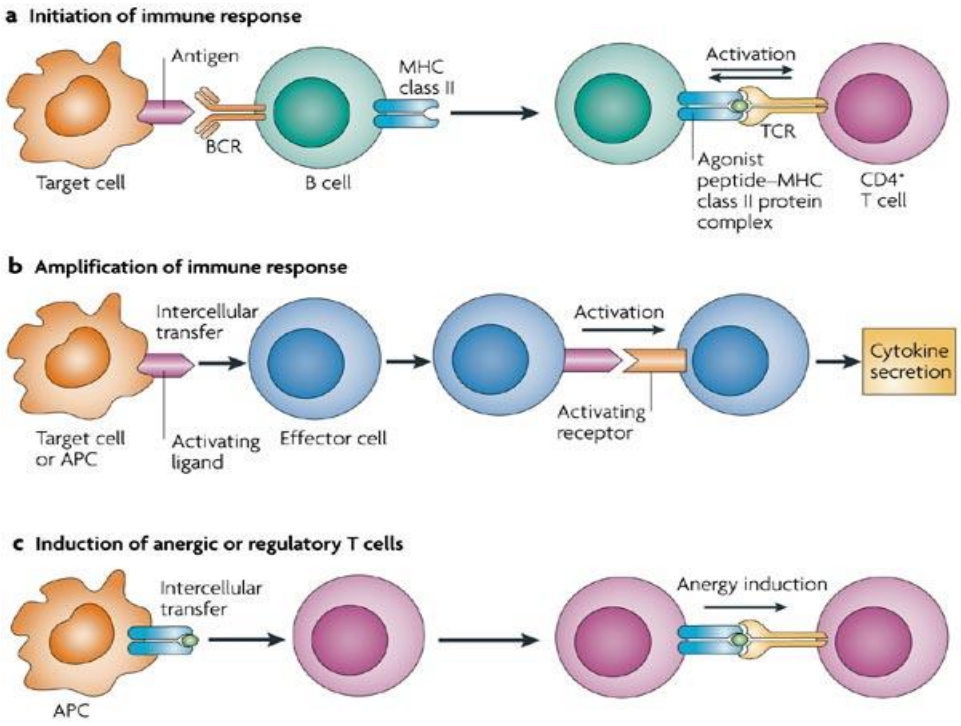


Figure 1.5. Proposed responses to cell surface antigens.

The transfer of peptides from cell surface proteins can influence numerous stages and facets of the immune response. Aberrant cell behaviour and desensitisation of effector cells, not depicted above, has also been proposed as additional responses. The influence of regulatory T-cells is important in establishing an immunosuppressive tumour microenvironment. Adapted from (Davis, 2007).

Immunotherapeutic approaches, such as the use of monoclonal antibodies, in the treatment of melanoma have by and large witnessed the greatest success in recent years (Luke and Hodi, 2013). For example, Ipilimumab is a monoclonal antibody that has been approved for use in both the UK and USA in the treatment of advanced melanoma. Rather than targeting cancer cells, this form of immunotherapy targets a protein, CTLA-4, that has an inhibitory effect on cytotoxic T-cells. So, this is a way of enhancing the ability of the T-cells to then attack the cancer cells. Trials are ongoing for this use of this antibody in other cancer types.

A large part of the work presented within this thesis is targeted at the identification of new genes which may encode CTAs – one important class of TAAs introduced above. CTAs will be discussed in more detail below but before this it is important to review the process of spermatogenesis, as well as highlight the differences between meiotic and mitotic cellular division. CTAs obtained their name because of their discovery in normal testis as well as cancer (Chen *et al.*, 1997a). Meiosis takes place only in the testis and there are particular genetic processes, some discussed below, which enable this type of cell division but functionally are of

potential significance if genes that govern these processes are reactivated in cancer (for example, see MacFarlane *et al.*, 2014; Simpson *et al.*, 2005; Whitehurst, 2014). Interpretation of immunohistochemistry (IHC) of normal testis is aided by a basic understanding of the structure and function of the testis: IHC experiments are presented later in this thesis. Furthermore, some basic inferences as to the possible functions of proteins present in a certain region of the testis or seminiferous tubules can be made, which can direct avenues for further research into particular genes. It will also be seen how the blood-testis barrier is of relevance in terms of CTAs as immunotherapeutic targets and in addition the relationship to other tissue barriers forms part of the basis for the subclassification of different classes of CTAs.

1.3. Spermatogenesis

Spermatogenesis is the process by which the male gametogenic cells produce spermatozoa, or sperm (Lie *et al.*, 2013). The process begins in puberty and after an early peak gradually reduces but unlike in females, where gametogenesis discontinues at the menopause, it does continue into old age (Eskenazi *et al.*, 2003). The functional structures of the testis in which spermatogenesis occurs are the seminiferous tubules.

Spermatogenesis is under direct influence from androgens such as testosterone, at least in part mediated by paracrine signalling from the Sertoli cells within the seminiferous tubules (Chen and Liu, 2015; Smith and Walker, 2014; Yan *et al.*, 2008). Androgen production itself is controlled by the gonadotropins: follicle-stimulating hormone (FSH) and luteinizing hormone (LH), which are released from the anterior pituitary. LH acts via receptors on Leydig cells, which are located between the seminiferous tubules, to stimulate them to produce testosterone.

1.3.1. Spermatozoa maturation and sub-compartmentalisation

The spermatogonia which reside in the basal compartment of the seminiferous tubules are the progenitor cells for spermatogenesis. The undifferentiated spermatogonia form the spermatogonial stem cells (SSCs). SSCs are thought to divide into either a new SSC, thus resulting in self-renewal of the stem cell compartment (see Figure 1.6) or result in a type A1 spermatogonial cell. The A1 spermatogonia then undergo a further mitotic division before producing primary spermatocytes via a meiotic division (Chen and Liu, 2015; Lui and Cheng, 2012; Zhang and Wu, 2015). When the primary spermatocytes have undergone the first meiotic

division (see Section 1.4.3) they become secondary spermatocytes. Following the second meiotic division, four haploid cells will result: initially round spermatids. These spermatid cells will then elongate to generate mature spermatids that are released into the lumen of the tubules as spermatozoa, or sperm. The process of release of spermatozoa from the adluminal compartment into the lumen is called spermiation. Thousands of sperm are produced every hour and millions are stored, primarily within the epididymis, prior to ejaculation. The acrosome is an organelle that forms a cap-like structure in the front of mature spermatozoa. A complex biological process known as the acrosome reaction, which enables invasion of an ovum, is an essential process for fertilization. Some genes discussed later in this thesis are involved in the acrosome reaction.

1.3.2. The Blood-Testis Barrier

The regional areas of the seminiferous tubules are split into basal and adluminal compartments by the blood-testis barrier (BTB). The BTB is formed by the tight junction between adjacent Sertoli cells (see Figure 1.6) and provides the developing germ cells with a site of immune privilege (Lie *et al.*, 2013). The spermatogonia and SSCs reside in the basal compartment, whereas meiosis and subsequent development of the spermatocytes occur within the adluminal compartment.

The Sertoli cells form a physical barrier to prevent large molecules, such as proteins, moving between the tubules and neighbouring blood vessels. However, even spermatogonial cells which reside in the basal compartment 'outside' the BTB do not result in the production of self-antigens. This is because the germline cells (each side of the BTB) lack the major histocompatibility complex (MHC) molecules required for the ability to present peptides produced within them for recognition by the immune system (Guillaudeux *et al.*, 1996). The BTB is a dynamic structure that is constantly reformed and dismantled; in part this is believed to be the mechanism which permits germ cells to progress towards the lumen between the supporting Sertoli cells (Lie *et al.*, 2013). Testosterone along with cytokines such as transforming growth factor-β (TGF-β) regulate the kinetics of protein endocytosis and recycling in the Sertoli cells that is necessary for this dynamic process to occur (Yan *et al.*, 2008). So, the BTB is achieved not only by there being tight junctions between adjacent Sertoli cells creating an anatomical or physical barrier but also through physiological and immunological means (Mital *et al.*, 2011).

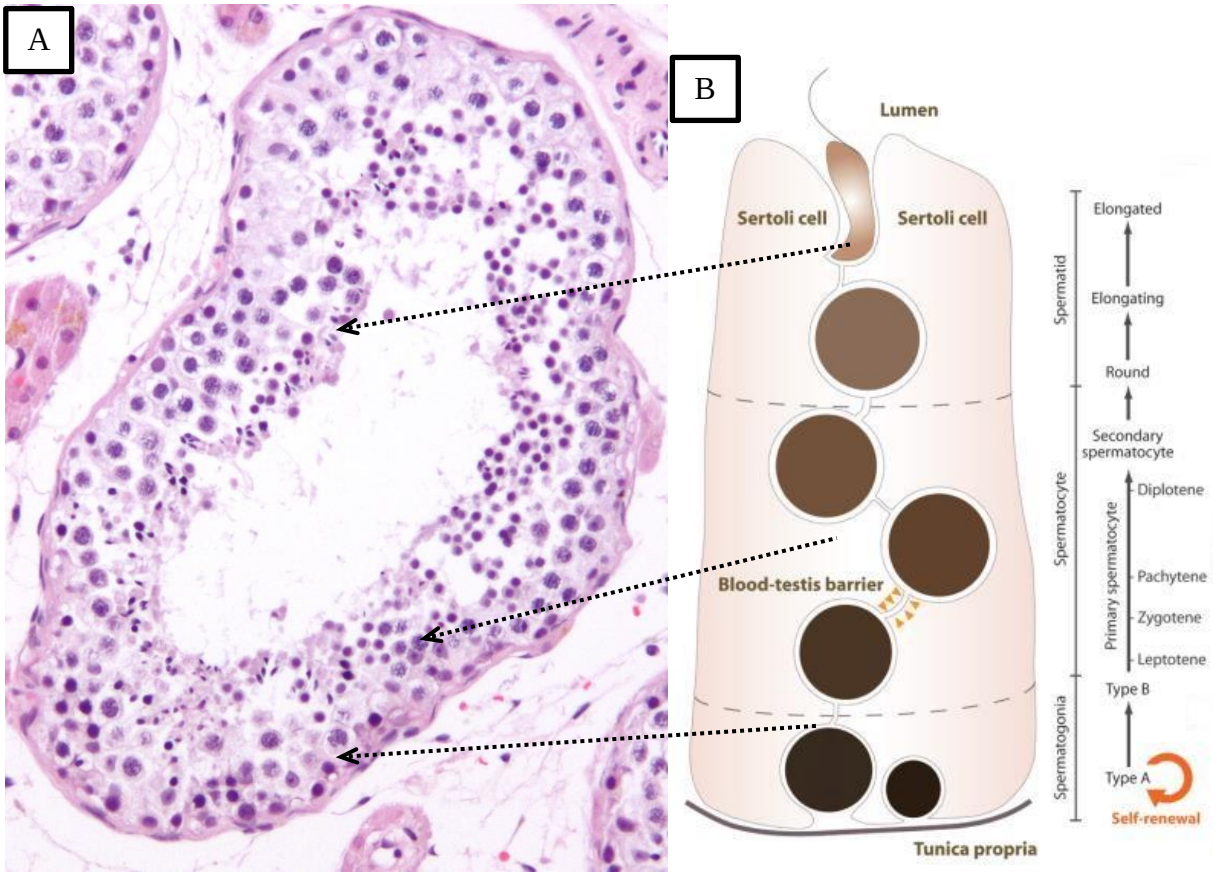


Figure 1.6. Cross section through the seminiferous tubule.

(A) A haemotoxylin and eosin stain of a seminiferous tubule in cross section indicating the structure of the seminiferous tubules and the morphological differences between developing gametogenic cells. (B) Vertical representation of this cross section – sperm develop and migrate between Sertoli cells within the seminiferous tubules. The spermatogonial cells are seen furthest from the lumen of the tubule in the basal compartment (lowermost arrow). As the sperm develop there is considerable change in the morphology of the developing cells as they undergo first of all mitosis and then meiosis within the adluminal compartment to produce mature spermatids that are capable of fertilisation. Note also the existence of the blood-testis barrier immediately above the basal compartment, which is formed by adjacent Sertoli cells but mediated via several mechanisms – see main text for details. A primary function of this barrier is to prevent the developing germ cells from autoimmune attack. Adapted from (Lui and Cheng, 2012).

1.3.2.1. Lack of HLA molecules potentially increases immunogenicity in cancer

The lack of MHC (or human leucocyte antigen – HLA) molecules within the gametogenic cells is brought about at least in part through post-transcriptional means (Guillaudeux *et al.*, 1996). The lack of such auto-antigen presentation by germ cells of the testis is of fundamental importance in protecting developing germ cells from immune attack but also of fundamental importance when one considers the CTA class of tumour antigens. It means that if the antigen is produced elsewhere in the body then the immune system will, or at least should, recognise them as 'foreign' and mount an immune response. Even if this immune response is weak or absent, it at the very least opens up the possibility of harnessing the immune system's capability of destroying cells with these 'non-self' TAAs. CTAs, and the relevance of the BTB in CTA biology, will be discussed again below in Section 1.5.

1.3.2.2. Similarities with other blood-tissue barriers

Tight junctions form at the interface between any selective epithelial-endothelial interface; they allow, for example, the selective transfer of nutrients and ions from the gut lumen into the blood stream, or excretion and reabsorption of ions and waste products within the kidney. Highly specialised blood-tissue barriers occur at certain regions of the body, most notably in the eye, placenta, central nervous system and testis. The blood-testis barrier has unique features but also shares similarities to the other specialised blood-tissue barriers in the body (Li *et al.*, 2012). For instance, connexins form the basic structural unit of the tight junctions whether these are in the eye, central nervous system or testis, and P-glycoprotein, which plays a crucial role in maintaining the blood-brain barrier, also acts at the BTB (Fromm, 2004; Li *et al.*, 2012; Mruk *et al.*, 2011). Interestingly, the efflux pump activity of P-glycoprotein, which has a physiological role at these blood-tissue barriers, can be hijacked by cancers to increase their resistance to chemotherapy and lead to a worse prognosis (Mao *et al.*, 2015; Sharom, 2011). Inhibitors of the action of these proteins are being developed in the hope of improving outcome for patients with multi-drug resistant cancers (Kapse-Mistry *et al.*, 2014). Due to the existence of the blood-brain barrier this makes the brain and central nervous system (CNS) a region of the body with relative immune privilege and this fact forms the rationale for the classification of CTAs, as described in Section 1.5.

1.4. Cell cycle and division in eukaryotes

1.4.1. Cell division in eukaryotes

Two types of cell division occur in eukaryotic cells: mitosis and meiosis. Both types of cell division occur during spermatogenesis as mentioned above. Meiosis is unique to spermatogenesis in males and so a more extensive overview will be provided.

Mitosis occurs throughout the human body as part of normal tissue homeostasis in which the chromosome complement is maintained in its diploid state producing identical daughter cells (Civelekoglu-Scholey and Cimini, 2014). Meiosis, however, only occurs in specialist (reproductive) cells of the germline and the cells become haploid in relation to their chromosome complement (Duro and Marston, 2015). Both types of cell division are tightly regulated and involve DNA replication followed by chromosomal segregation. Meiosis occurs in the foetal ovary but in adults is confined, as an entire process, to the testis.

1.4.2. Mitotic cell cycle

This type of cell division produces two identical daughter cells from a single parent cell (Civelekoglu-Scholey and Cimini, 2014). It is essential for maintaining homeostasis and normal regional cellular function. At any one time, the majority of cells are not undergoing mitosis; that is, they are quiescent. This is known as the G_0 phase and the duration of this can vary greatly between cell types. Mitogenic signals promote exit from G_0 and progression towards mitosis. Once a cell progresses out of this phase it is committed to completing at least one cell cycle and producing two identical diploid daughter cells; assuming, that is, it passes the various cell cycle checkpoints that safeguard against errors in DNA processing (Malumbres and Barbacid, 2009). The cell cycle is split into interphase and M-phase – see Figure 1.7. Oncogenic signals may also promote exit from G_0 and cause uncontrolled cell proliferation, which is a fundamental hallmark of cancer (Hanahan and Weinberg, 2011). Cyclin dependent kinases (CDKs) control progression through the normal cell cycle. Mutations in the genes that activate or inhibit CDKs, or indeed the genes that encode them, can lead to not only uncontrolled cellular proliferation but also changes in chromosome number (i.e. chromosomal instability) as well as genetic instability (Malumbres and Barbacid, 2009).

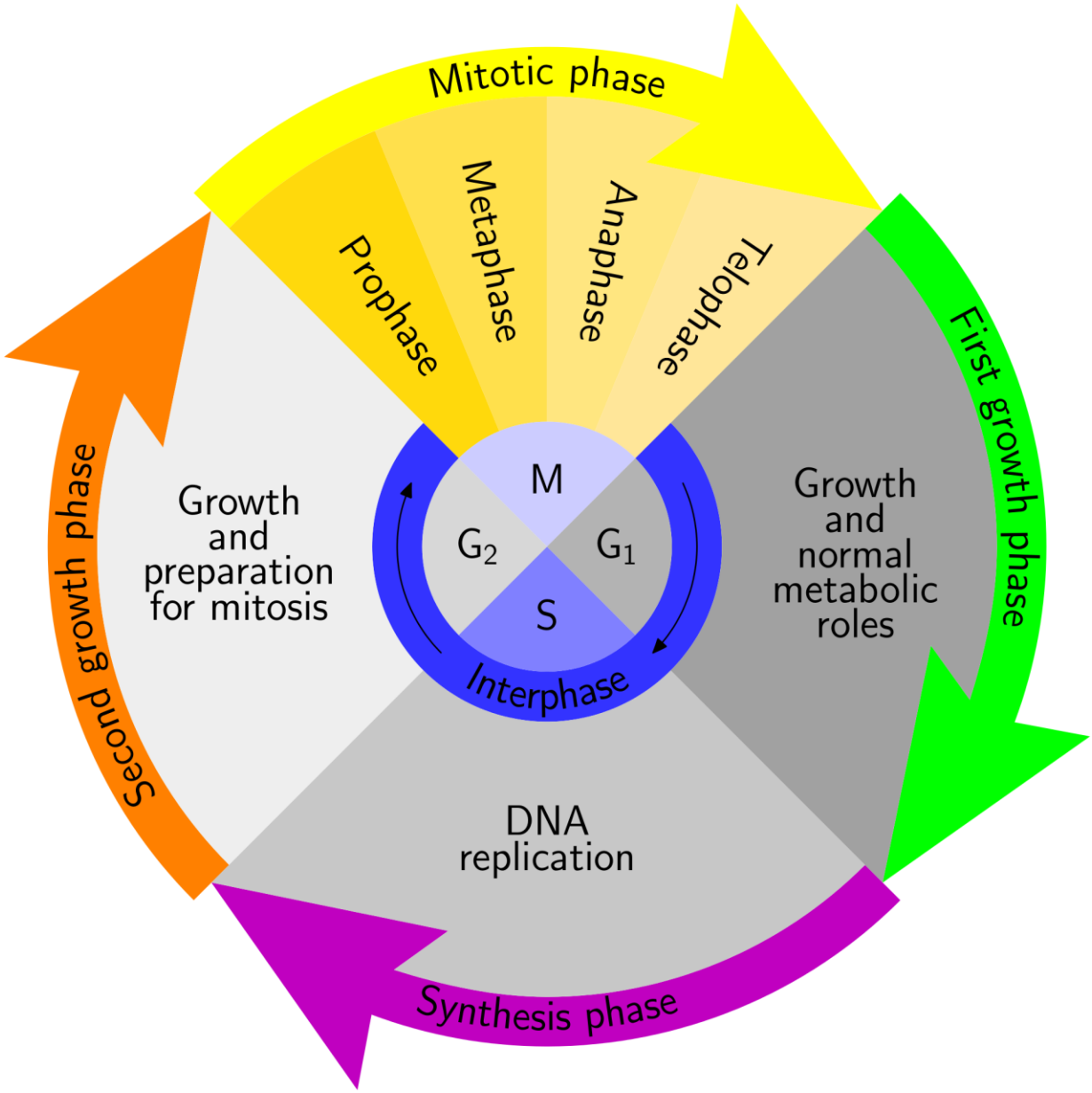


Figure 1.7. The cell cycle.

The adult cell cycle is split into interphase and an M-phase or division phase. Interphase is split into three separate phases: G1 where the cell grows and prepares for DNA replication; S-phase where DNA synthesis occurs; and G2 which is a second growth phase where the cell prepares for the division phase. M-phase is itself subdivided into five phases, which retain the same terminology in both meiosis and mitosis. These five phases of cellular division are: prophase; prometaphase (not depicted above); metaphase; anaphase; and telophase. Meiosis involves two separate divisions to produce four haploid daughter cells, whereas mitosis produces two identical daughter cells. Mitotic daughter cells will 'exit' the cell cycle and enter a quiescent G_0 phase and await a mitogenic signal to re-enter G1.

1.4.3. Meiosis

Meiosis only occurs in gametogenic cells of the body. Meiosis results in cells (i.e. ova or spermatozoa) that are haploid – this is essential for sexual reproduction in mammals and many other organisms. These specialist cells are found only within the ovaries and testis. When the two gametes fuse during the fertilisation process they produce a zygote, which once again becomes a diploid cell. The resulting offspring, which develops from the zygote, has thus inherited one distinct copy of each chromosome, known as a homologue, from both parents. There are important differences in when meiosis occurs in males and females – see Figure 1.8. Most notably, in men the process is not initiated until puberty. In females the ova are produced in the foetus and meiosis is arrested in prophase 1 before birth until puberty.

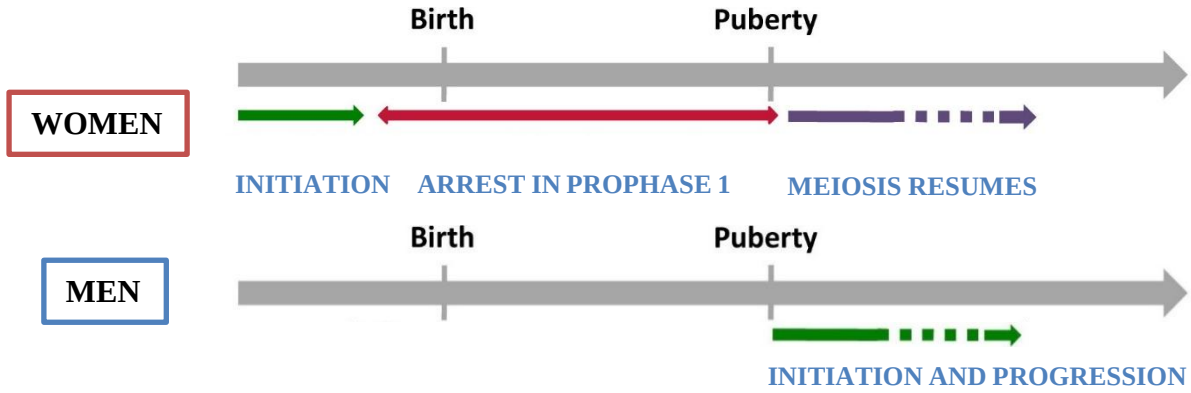


Figure 1.8. Differences in meiosis between males and females.
In a woman all the ova she will ever produce are produced in the foetus. After this and until puberty, meiosis is arrested in prophase 1 at which point meiosis will resume until she reaches the menopause. For men, however, meiosis is initiated at puberty and they will continue to produce sperm for the rest of their lives, although the rate of sperm production declines gradually at they become older.
Adapted from, Baillet and Mandon-Pepin (2012).

1.4.4. The stages of meiosis

Meiosis is a complex and tightly controlled process (Schvarzstein *et al.*, 2010). Comparison to mitosis of the stages involved is depicted in Figure 1.9. Both mitosis and meiosis are split into five phases and meiosis occurs in two separate divisions termed meiosis I and meiosis II. As is the case during mitosis, before cellular division occurs the DNA must first replicate and the cell prepares for the division. Significant cellular structural changes occur, such as breakdown of the nuclear membrane and cytoskeletal rearrangement, during both mitosis and meiosis but the

chromosome segregation and rearrangements that occur in meiosis are unique and more complex resulting in genetically distinct daughter cells (Civelekoglu-Scholey and Cimini, 2014; Duro and Marston, 2015; Schvarzstein *et al.*, 2010). Following DNA replication two rounds of chromosome segregation occur during meiosis: during meiosis I reductional chromosome segregation occurs, and during meiosis II equational segregation of the chromosomes takes place (Duro and Marston, 2015).

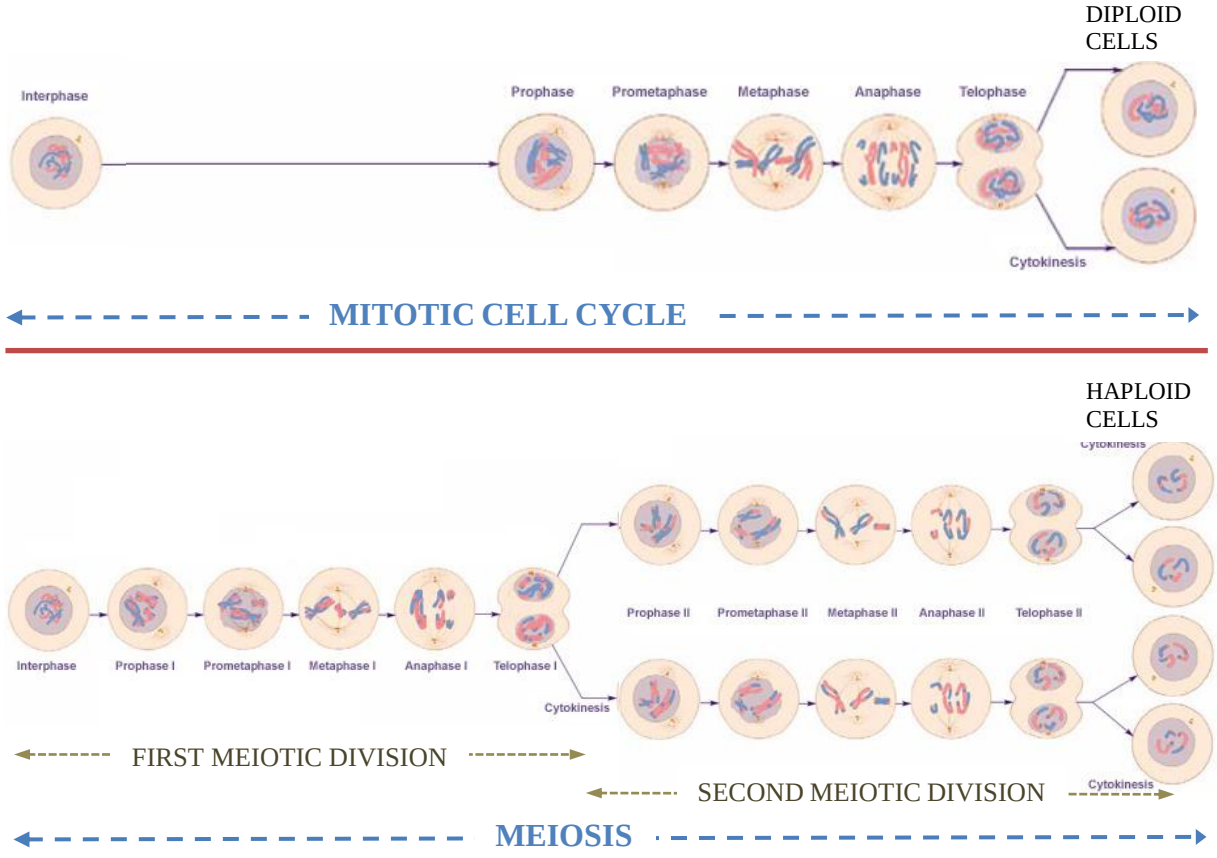


Figure 1.9. Comparison between the mitotic and meiotic cell cycles.
Both meiosis and mitosis involve growth phases occurring before and after the S phase in which DNA synthesis/replication occurs prior to cellular division. Mitotic cells spend the majority of their time in this interphase and/or outside of the cell cycle in a phase called G_0 when they are not dividing. In meiosis, pictured here below mitosis, four daughter cells have been produced with half as many chromosomes as the parent cell. These haploid cells will be able to fuse with haploid cells from the other parent and produce a zygote through the process of fertilisation. The process of meiosis is split into two stages: in meiosis I, or the first meiotic division, the chromosome number is maintained; in meiosis II chromosome segregation this time results in four haploid daughter cells. The stages of each cell division (for both meiosis and mitosis) are split into five distinct phases: interphase, prophase, prometaphase, metaphase, anaphase, and telophase. Adapted from (Saltsman, 2005).

1.4.5. DNA double-strand breaks and meiotic recombination

Formation of DNA double-strand breaks (DSBs) occurs early on during meiosis and is necessary for meiotic recombination to occur, which is essential for correct homologue alignment and bivalent formation (Baudat *et al.*, 2013; Longhese *et al.*, 2009). The most marked differences to mitosis occur during prophase I. During this phase homologous chromosomes pair and then synapse. The pre-meiotic S-phase during which chromosome replication takes place is prolonged; during this time inter-homologue interactions take place to subsequently permit recombination events and chromosome segregation. The synapsis of homologous chromosomes during prophase I is brought about by a large proteinaceous structure called the synaptonemal complex (see Figure 1.10). The pairing of homologous chromosomes prior to this synapsis is promoted by the polarised and condensed chromosomal arrangement. When visualised microscopically the chromosomes are thought to resemble a bouquet of flowers and so this is referred to as the bouquet formation (Harper *et al.*, 2004; Klutstein and Cooper, 2014). This formation is caused by the transient clustering of telomeres on the inside of the nuclear envelope. Only recently the function of the telomeres as important tethering factors for the centromeres during meiosis has been elucidated (Klutstein *et al.*, 2015).

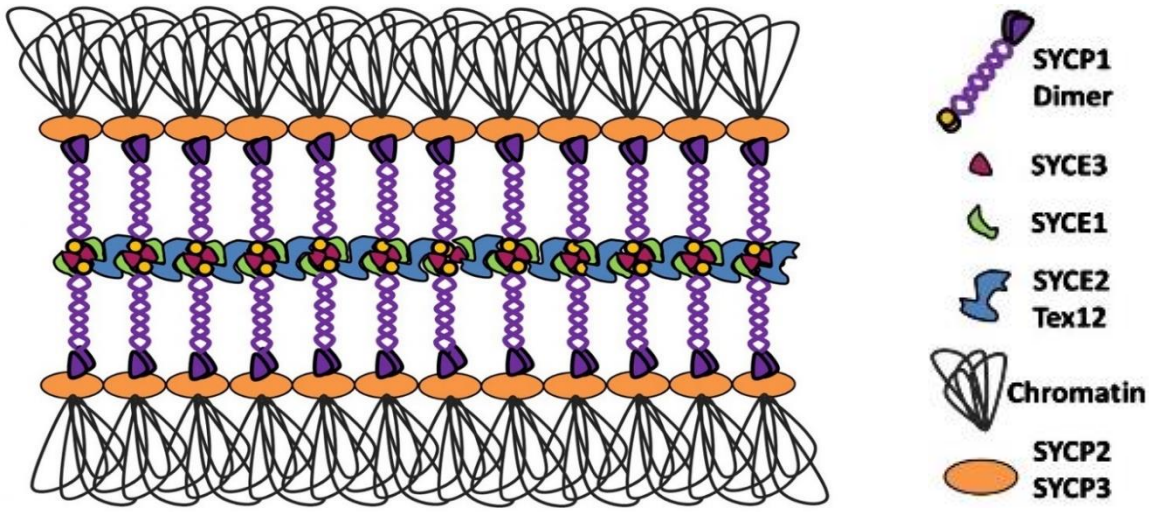


Figure 1.10. The synaptonemal complex.

There are seven known constituent meiosis-specific proteins that form the synaptonemal complex (SC). The formation of the SC, arranged in a ladder-like formation, begins early in prophase I. The central element (CE) is formed by four proteins: SYCE1 coalesces with SYCE3, and TEX12 with SYCE2. SYCE1 and SYCE3 have the ability to self-assemble and are thought to act as the structural framework binding other SC proteins (Fraune *et al.*, 2012). The transverse filaments are composed of parallel homodimers of SYCP1; the N-terminal domain associates with the proteins of the CE and the C-terminal domain with the lateral element (LE) proteins. The LE proteins are SYCP2 and SYCP3. Adapted from, (Fraune *et al.*, 2012).

The primary purpose of meiotic recombination is to generate inter-homologue connections, which, together with sister centromere monopolarity, permit the reductional division. However, crossover recombination events also serve as the mechanism for creating unique, genetically distinct, gametes and as a consequence contribute to genetic and thus biological diversity. Recombination events do not occur randomly but are concentrated in specific regions of the genome, known as hotspots. The PRDM9 protein has been identified as being able to bind to some of these hotspots and thus regulates the site for some meiotic recombination initiation (Baker *et al.*, 2014; Baudat *et al.*, 2013). DSBs are generated in meiosis by a topoisomerase II-like enzyme SPO11. SPO11 is a highly conserved protein without known physiological roles outside of meiosis/meiotic recombination. It acts as a dimer to catalyse the breakage of both DNA strands and is removed from the DNA by endonucleolytic cleavage by the MRN complex of proteins (de Massy, 2013; Lam and Keeney, 2014). It will be seen that several of these genes will be mentioned again later in the thesis and it is useful here to outline the context of their roles in normal physiology.

RAD51 is a recombinase that is central to mitotic recombination events; it is a homologue of the prokaryotic protein RecA (San Filippo *et al.*, 2008). It is also required for homologous recombination (HR) during meiosis but another RecA homologue, DMC1, is also critical. DMC1 is meiosis-specific and the precise mechanism of action is unclear but it acts in concert with RAD51 to perform HR during meiosis. Both proteins catalyse the exchange of DNA strands from one DNA duplex to another (Handel and Schimenti, 2010). DMC1, together with two other meiosis-specific proteins Hop2 and Mnd1, helps drive the bias that exists in meiosis to increase the probability of choosing a homologous chromatid over a sister chromatid with which to exchange strands and thus ensure correct bivalent formation (Brown and Bishop, 2014; Kang *et al.*, 2015). The recombinase proteins interact with single-stranded DNA (ssDNA) to from a nucleoprotein structure known as the presynaptic filament. The presynaptic filament engages a neighbouring strand of DNA and searches for homology in it – this is referred to as homology search. Successful homology search is thus required for homologous recombination following DSB formation – see Figure 1.11. It has been suggested that the presynaptic filament formed by DMC1 drives the initial homology search in preference of a DNA strand from the homologous chromatid and RAD51 performs the second end capture but various models have been proposed (Brown and Bishop, 2014). Other accessory proteins, such as the Hop2-Mnd1 heterodimer in meiosis, are also critical for successful HR (Brown and Bishop, 2014; Kang *et al.*, 2015; San Filippo *et al.*, 2008).

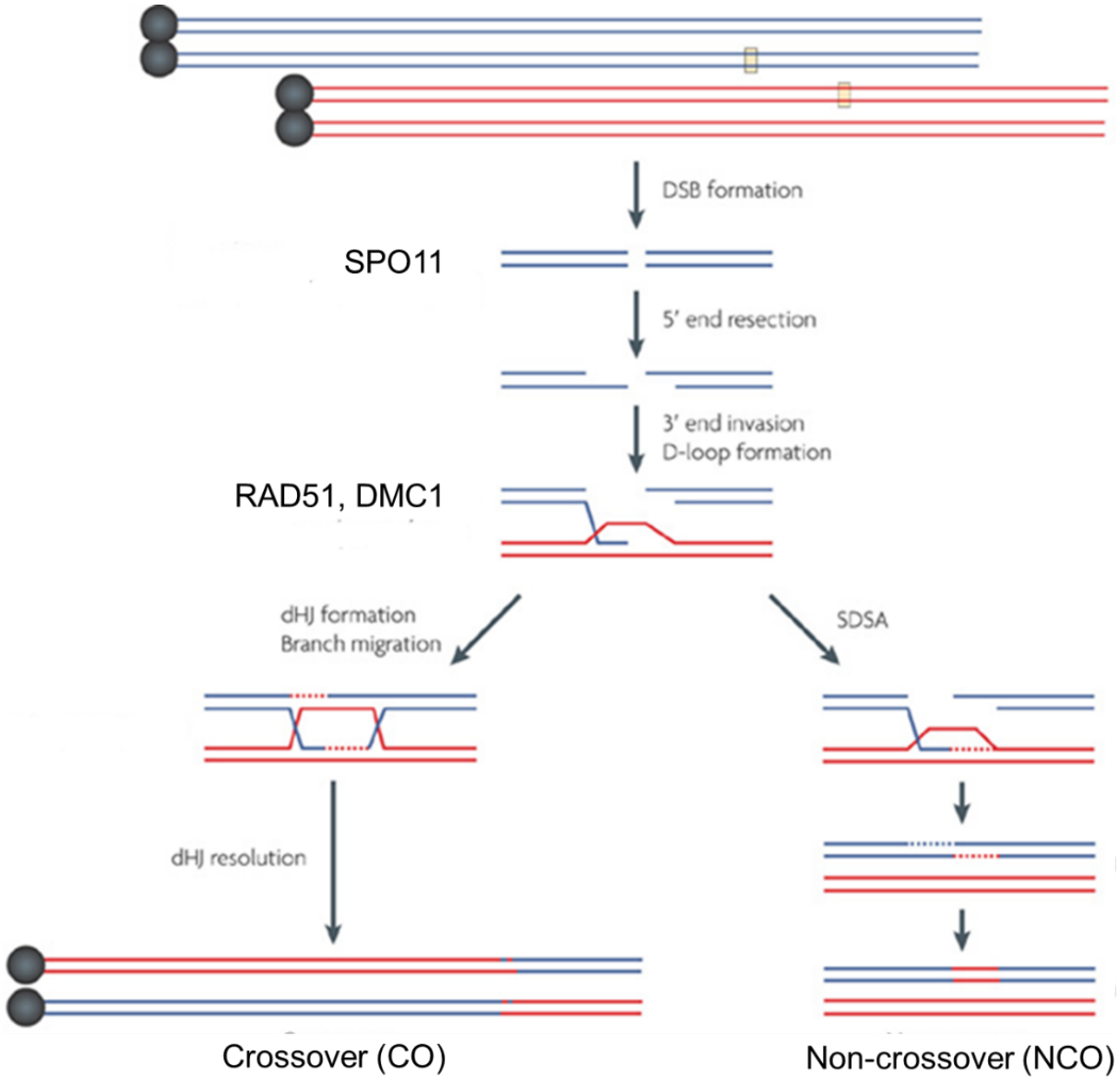


Figure 1.11. Repair of DSBs by homologous recombination in meiosis.

Homologous recombination in meiosis can result in crossover (CO) and non-crossover (NCO) events. SPO11 cleaves the DNA to form a DSB; this is the trigger for the initiation of homologous recombination. Single-stranded DNA (ssDNA) is formed at the 3' end following resection of the DNA in a 5' to 3' fashion. This overhanging ssDNA is the substrate for the HR protein machinery which form a nucleoprotein filament known as the presynaptic filament; permitting invasion of the 3' end into its homologue. DNA synthesis ensues using the homologue as a template. The DSBs can be repaired via the double-strand brake repair (DSBR) or synthesis-dependent strand annealing (SDSA) pathway. DSBR involves the capture of the second overhanging ssDNA strand to form two Holliday junctions. This can result in either a crossover or a non-crossover recombination event. Gap-filling DNA synthesis and ligation are required to complete the process in both types of repair. There are in fact four distinct repair pathways for DSB repair in meiosis as inter-sister, as well as inter-homologue recombination events can occur, although there is a preference for the latter.
Adapted from (Handel and Schimenti, 2010).

1.4.6. DNA repair and human disease

DNA is under constant insult from different sources and there are numerous and highly effective pathways for repairing damaged DNA (Abbotts *et al.*, 2014). Defects in these repair pathways, either inherited or acquired, can lead to an increased susceptibility to cancer as they will lead directly to genetic instability. There are also specific repair pathways for repairing DSBs as seen in meiosis; including non-homologous end joining (NHEJ) and HR (Chapman *et al.*, 2012; Longhese *et al.*, 2009; Lord and Ashworth, 2012). HR, which occurs in the S and G2 phases of the cell cycle, is less error prone than NHEJ.

Certain inherited genetic disorders carry an increased risk of developing cancers. Fanconi anaemia is an uncommon but interesting example as links to various DNA repair pathways have been discovered through a drive to better understand the disease (Longerich *et al.*, 2014). Although patients can succumb to bone marrow failure in childhood, patients with Fanconi anaemia have a dramatically increased risk of developing acute myeloid leukaemia as well as other forms of cancer (Longerich *et al.*, 2014). Lynch syndrome is another example of an inherited disorder that leads to an increased risk of developing cancer. Here, mutations in various mismatch repair genes result in suboptimal DNA repair, thus contributing to genomic instability and an increased risk of developing many types of cancer (Al-Sohaily *et al.*, 2012). This condition will be mentioned again below in connection to colorectal cancer – see Section 1.6. Mutations in the same genes can also be acquired and lead to chromosomal instability and cancer in the same fashion.

BRCA2 is an example of a tumour suppressor gene; when a mutation is inherited or acquired patients have an increased risk of developing cancer (most notably breast and ovarian cancer, as is the case for its sister gene *BRCA1* - see Table 1.1). BRCA2 is a large protein that when a gene mutation results in a defective form, contributes greatly to genomic instability. This is largely due to the fact that healthy BRCA2 interacts directly with the recombinase RAD51 via its BRC motif and assists with HR repair events in somatic cells by promoting the binding of RAD51 to ssDNA (Thorslund *et al.*, 2010). Thus, an increased probability of cancer results over time as suboptimal repair mechanisms exist to cope with the exogenous DNA damaging insults that occur throughout life. BRCA2 also interacts, via a separate binding site, to DMC1 and it is thought to be required for the co-localisation of both recombinases in meiosis (Thorslund and West, 2007). The RAD51-associated protein 1, RAD51AP1, is also thought to be involved in the co-localisation of these recombinases (Dunlop *et al.*, 2011). PALB2

(encoded by another important tumour suppressor gene, *PALB2*) is required for the localisation of BRCA2 to damaged regions of chromosomes (Xia *et al.*, 2006) and also acts in synergy with RAD51AP1 to promote D-loop formation by the recombinase (Dray *et al.*, 2010). Thus, it can be seen how there may be interactions between numerous factors that ultimately result in genetic instability due to a failure of adequate DNA repair and as a consequence can contribute to oncogenesis. Some of these DNA repair genes will be mentioned again later in the thesis. I will now go on to discuss an important class of TAAs, the cancer/testis antigens.

1.5. Cancer-testis antigens

1.5.1. Discovery and classification

The term cancer/testis antigen (CTA) was coined in 1997 to describe a class of TAA that had limited presence in somatic tissues (Chen *et al.*, 1997a). At this point several cancer/testis (CT) genes had already been discovered; including the first members of the melanoma-associated antigen family (e.g. *MAGE-A1* and *MAGE-A3*) as well as *PAGE* and two members of the *GAGE* gene family (Boel *et al.*, 1995; Gaugler *et al.*, 1994; Van den Eynde *et al.*, 1995; van der Bruggen *et al.*, 1991). It is now known that CT genes are more often present as large paralogous families of antigens such as the *MAGE* genes and the *SSX* class of CT genes that were discovered soon afterwards (Chomez *et al.*, 2001; Gure *et al.*, 1997). Many of the early discoveries were in patients with melanoma, which is a cancer that has many tissue specific gene activations associated with it leading to more prominent adaptive immune responses *in vivo* (Blankenstein *et al.*, 2012). *NY-ESO-1* encodes one of the most immunogenic TAA that shares little homology with other known proteins: it was discovered in a patient with oesophageal squamous carcinoma (Chen *et al.*, 1997b) and later used for a successful immunotherapeutic approach in a patient with metastatic melanoma (Hunder *et al.*, 2008).

The CTAs are a diverse class of TAAs (Fratta *et al.*, 2011; Whitehurst, 2014). The genes which encode these antigens, termed CTA genes or cancer/testis (CT) genes, are expressed in the testis and malignant tissues. *Bona fide* CT genes have expression silenced in normal tissues other than the testis, but then become activated in cancer (Hofmann *et al.*, 2008; Scanlan *et al.*, 2004). The expression of CT genes has been shown to vary greatly between different types of cancer: melanoma, ovarian and lung cancer for instance have generally higher expression levels than colorectal or haematological cancers (Almeida *et al.*, 2009; Caballero and Chen, 2009). They

have attracted a great deal of interest in the field of oncology and tumour immunology (Cheng *et al.*, 2011; Costa *et al.*, 2007; Drake *et al.*, 2014; Fratta *et al.*, 2011; Simpson *et al.*, 2005). CTAs have attracted such interest owing to their expression profiles, coupled with the fact that germ cells do not have HLA molecules enabling them to present antigens for immune recognition (as mentioned in Section 1.3). This increases their immunogenicity (i.e. ability to stimulate an adaptive immune response) making them particularly attractive targets for both cancer vaccines and adoptive therapy (Hunder *et al.*, 2008; Valmori *et al.*, 2007). Their expression profile also makes them highly attractive as cancer-specific biomarkers that could improve diagnosis, stratification or current screening methods (see Section 1.1.7).

Current knowledge of CT gene expression patterns is summarised on 'CTDatabase' (Almeida *et al.*, 2009). Over 250 genes from over 100 separate gene families have been identified although not all of these meet the strict CT gene expression criteria (Hofmann *et al.*, 2008). CT genes can be subclassified based on their chromosomal location or alternatively how restricted their expression profile is. Many of the known CT genes are encoded by the X-chromosome and are expressed in the basal layer of the seminiferous tubules, i.e. in the spermatogonial cells (Chomez *et al.*, 2001). CT genes located on the autosomes are known as non-X-CT genes. As discussed above, meiosis is a special type of cell division limited to the germ cells. It has been known for some time that meiosis-specific genes (e.g. *SCP-1*, *HORMAD1*) can encode a CTA (Chen *et al.*, 2005; Tureci *et al.*, 1998a). A subcategory of CT genes, termed meiCT genes, that are specifically associated with meiosis have more recently been described (Feichtinger *et al.*, 2012b; Sammut *et al.*, 2014). Due to transcriptional silencing of the X-chromosome during meiosis in males (Turner, 2007) meiCT genes would be expected to be autosomally encoded. This is indeed the case for the majority of the meiCT genes described.

It has been observed that some cancers co-express numerous CT genes, whereas others from the same or different cancer types do not (Burgdorf *et al.*, 2008; Choi and Chang, 2012; Daudi *et al.*, 2014). The promotor regions of many germline genes are demethylated in the testis allowing them to become active during spermatogenesis. Moreover, the methylation of these genes in somatic tissues is believed to be a primary silencing mechanism (Shen *et al.*, 2007). Global hypomethylation of DNA is a common feature of cancer (Feinberg and Vogelstein, 1983); a feature that could lead to the activation of CT genes outside of their normal site of expression. This is believed to be the reason for the co-expression of CT-genes observed in some tumours. Furthermore, treatment of cancer cell lines with the demethylating agent, 5-aza-

2'-deoxycytidine, has been shown to lead to increased expression of many CT genes adding further support for this idea (Adair and Hogan, 2009; Almatrafi *et al.*, 2014). Other epigenetic influences, such as histone deacetylation, are likely to also be important mechanisms for ectopic activation of CT genes as demethylation treatment does not consistently result in expression (Almatrafi *et al*., 2014; Fratta *et al*., 2011).

1.5.1.1. Subclassification based on somatic tissue expression profile

In addition to being classified as CT-X and non-X-CT genes based on their chromosomal locations, CT genes can also be subclassified into three broad groups based on their expression profiles in normal tissues (Hofmann *et al.*, 2008):

- As mentioned above a *bona fide* CT gene would be expressed in cancer and testis only: these CT genes are termed testis restricted. However, as the placenta is a specialised tissue that is both a site of immune-privilege and also contains germline cells/genes it too can be considered as a restricted site. That is, if a gene is expressed in the testis, the placenta and cancer it would still be considered as belonging to this tightly restricted class. It has been proposed that CTAs may play a role in various phases of placenta development (Jungbluth *et al.*, 2007). Extravillous trophoblast invasion and implantation, as well as angiogenesis, are fundamental processes involved in placental development (Pollheimer *et al.*, 2014). It is possible that the same functions (of cancer-germline gene products) that promote placental development also promote oncogenesis through similar mechanisms (see Section 1.5.3).

- A second group, termed testis/brain (or testis/CNS) restricted are present in the testis, CNS and cancer only, with no expression in other normal somatic tissues.

- The third group has predominant expression in testis (and found in cancer) but limited expression in somatic tissues. 'Limited expression' could be considered as very low (e.g., 100-fold lower) expression compared to testis and/or limited to two other normal tissue types (Caballero and Chen, 2009; Feichtinger *et al.*, 2012b; Hofmann *et al.*, 2008). Genes with such low expression titres in normal tissue might still be excellent targets for immunotherapeutic approaches and remain viable biomarkers for cancer but clearly their potential strength in both these regards would be less than the truly restricted category of CT genes. As the ovaries do not offer a site of immune privilege in the same way (as the placenta, testis, or even CNS do) then genes expressed here are

generally considered the same as other somatic tissues in terms of categorisation. However, some germline genes would be expected to be expressed exclusively in the ovaries in females and they have been considered a third site of restricted expression by some (Scanlan *et al.*, 2002). Many of the meiCT genes were shown to have significant upregulations in ovarian cancer when clinical data-sets were meta-analysed (Feichtinger *et al.*, 2012b; Sammut *et al.*, 2014), though the reasons for this are not clear as the cellular origin of ovarian cancer is believed to be extra-ovarian in many cases (Berns and Bowtell, 2012; Sorensen *et al.*, 2015).

1.5.2. CTAs as cancer biomarkers

As is the case with numerous other new biomarkers, despite their promise no CTAs have been widely adopted and/or endorsed, for example, by the regulatory authorities in the UK or USA. However, examples of CTAs that have been proposed as useful biomarkers and/or shown to be predictive of prognosis or treatment response include: ATAD2, MAGE-A3, -A6 and -A9 for breast cancer (Ayyoub *et al.*, 2014; Kalashnikova *et al.*, 2010; Xu *et al.*, 2014); NY-ESO-1 in non-small cell lung cancer, hepatocellular carcinoma and gastrointestinal stromal tumours (John *et al.*, 2013; Perez *et al.*, 2011; Xu *et al.*, 2012); MAGE-C1 in multiple myeloma (de Carvalho *et al.*, 2012). It may remain the case that such biomarkers do establish themselves within personalised treatment strategies, rather than for internationally used markers for diagnosis or screening. It has recently been shown that expression of a panel of germline genes can be used to predict a worse outcome and guide therapy in lung cancer patients (Rousseaux *et al.*, 2013b). Importantly activation of this specific cohort of germline genes was an independent predictor of a poor outcome, even in patients that would have otherwise (i.e., by existing stratification methods such as TNM staging) not been predicted to have such a poor prognosis. Thus, such approaches have tremendous potential to be of real clinical benefit to patients by guiding the need for more aggressive treatment strategies and/or informing the clinician that the treatment is not going to work and consequently saving money, as well as the potential unpleasant side effects to the patient.

1.5.3. Possible roles in oncogenesis

As mentioned above, the targeting of testis-restricted CTAs for cancer treatment would be expected to have no or minimal deleterious side effects in non-cancerous healthy tissue, as the

proteins are not found there. However, there is a second distinct reason why the study of CTAs and CT genes is important. This is that they, and in particular the meiCT genes and their products, could be having a powerful oncogenic influence on the cancer. The products of some of these genes would normally serve to drive the meiotic chromosome dynamics and differentiation of developing gametes; related functions if aberrantly expressed in somatic tissues could be driving significant genetic (and epigenetic) changes, a possibility that remains largely unexplored (McFarlane *et al.*, 2015). For example, it has been suggested that aneuploidy, which is a very common feature in cancer, could be brought about in somatic cells by the action of these normally silent germline genes (Fratta *et al.*, 2011; Rousseaux *et al.*, 2013a; Wang *et al.*, 2014; Whitehurst, 2014). In a seminal study in *Drosophila melanogaster* genome wide expression profiling in a brain tumour model system revealed that a significant proportion of upregulated genes were germline in origin, and inactivation of some members suppressed tumour growth (Janic *et al.*, 2010). This provided direct evidence that these germline factors were promoting oncogenesis. Moreover, it was subsequently shown that several of these genes displayed statistically significant upregulations when clinically-derived (cancerEST and cancerMA) datasets were meta-analysed, thus adding support to the idea that a soma-to-germline transition is a hallmark of many cancers (Feichtinger *et al.*, 2014). Although the precise mechanisms for the oncogenic-promoting effects of germline genes is unknown, it is believed that many of the effects result from changes in broader genome epigenetics that occur as a direct result of germline- or testis-specific gene activation in cancer (Wang *et al.*, 2011; Whitehurst, 2014).

Possible and identified functions of CTAs and/or roles in oncogenesis include:

- Contributing to genetic and chromosomal instability. Conserved meiotic functions, such as meiotic recombination, could lead to chromosomal rearrangements leading directly to such instability in a self-perpetuating manner (Cheng *et al.*, 2011; Fratta *et al.*, 2011; McFarlane *et al.*, 2015; Simpson *et al.*, 2005). The synaptonemal complex (see Figure 1.10) is a large protein structure that forms a stable connection between two homologous chromosomes during prophase in meiosis 1. SCP1 is an integral member of this complex and was first identified as a CTA in 1998 and since found to be present in several cancer types (Haffner *et al.*, 2002; Lim *et al.*, 1999; Tureci *et al.*, 1998a). Other examples of CTAs found in cancer that could be contributing to unstable genetic alterations by driving meiotic-like chromosome dynamics include: the meiotic regulator

STRA8, the metalloprotease ADAM2 known to play a role in spermatogenesis, the cohesin subunits SMC1β and RAD21L, the meiotic hot spot activator PRDM9, and SPO11 which creates DSBs as mentioned above (Feichtinger *et al*., 2012b; Koslowski *et al*., 2002; Litvinov *et al*., 2014; Sammut *et al*., 2014). HORMAD1 is a meiosis-specific protein that is needed for formation of the synaptonemal complex and has been implicated in directing the site of DSB formation involving MEI4 as well as promotion of homologue alignment (Daniel *et al.*, 2011; Kumar *et al.*, 2015). Recently, it has been shown that overexpression of HORMAD1, which localised to the nucleus of breast cancer cells, suppresses RAD51-dependent HR, thus promoting genomic instability and a bias towards NHEJ as a means of DNA repair (Watkins *et al.*, 2015). Initial data also suggest that HORMAD1 expression will be a useful way of tailoring effective treatment options (Watkins *et al.*, 2015).

- Alteration of chromosome number. Creation of aneuploidy, which is a cancer-wide phenomenon, would be promoted by activation of germline genes through the same functions that cause large chromosomal rearrangements and/or altered cohesion (i.e., sister monopolarity) during gametogenesis (Fratta *et al.*, 2011; Simpson *et al.*, 2005; Whitehurst, 2014). Genome amplification provides cancer cells with a survival advantage (for example by providing a 'buffer' against damaging gene mutations) and can explain many of the hallmarks of cancer as well as resistance to anti-cancer therapy (Coward and Harding, 2014). It has been suggested that cancer cells hijack the gametogenic programme by expressing meiotic factors to create a polyploid state that promotes immortality and genome plasticity (Erenpreisa *et al.*, 2014).

- Acting as transcriptional regulators or activators. It is known that some CTAs act as transcriptional activators, often for other germline genes but also for Myc-mediated oncogenes and genes involved in epithelial-mesenchymal transition (EMT) (Daudi *et al.*, 2014; Kim *et al.*, 2009; Minges *et al.*, 2013; Whitehurst, 2014). Activation of oncogenes or disrupted function of tumour suppressor genes would be directly oncogenic. The co-activation of other germline genes serves as an additional mechanism (to DNA hypomethylation described above), for how some cancers co-express multiple CT genes, whereas others none or very few. LUZP4 has recently been shown to interact with a major mRNA export receptor and act as an adaptor for mRNA export (Viphakone *et al.*, 2015). This would provide a mechanism for the transcription

activation of other genes and as it is commonly expressed in melanoma may be acting as an important oncogenic driver in this cancer type. Testis-specific factors can also have other reprogramming influences on the epigenome of somatic cells to promote a malignant phenotype (reviewed in, Wang *et al*., 2011).

- Promotion of mitotic activity, thus enhancing proliferation which is seen as a fundamental hallmark of cancer. For example, TEX14 and CASC5 have been important for kinetochore assembly and so may be actively driving mitotic division when expressed in the tumour (Whitehurst, 2014). CAGE has also been shown to promote proliferation of cancer cells by increasing levels of cyclins involved in progression to S phase of the mitotic cell cycle (Por *et al.*, 2010). On the other hand, another CTA BORIS has been shown to have growth inhibitory effects and may have tumour-suppressor functions *in vivo* (Tiffen *et al.*, 2013); although, contrarily, in an acute leukaemia cell line the opposite effect was seen and it was thought this was mediated through epigenetic changes influencing NOTCH signalling (Zampieri *et al.*, 2014).

- Influencing cellular signalling, inducing EMT and maintenance of cancer stem-like cell characteristics. The localisation of certain CTAs within the CSC compartment is of particular interest (Wen *et al.*, 2014; Yamada *et al.*, 2013); though whether they actively support the CSC niche or whether they are activated as a consequence of other genetic/epigenetic changes in that region of the tumour remains uncertain. In any case, it raises the possibility of targeting these proteins to kill the resistant CSCs that are believed to be a primary cause of cancer recurrence following standard treatments. However, the fact that CTAs have been shown to be able to induce EMT makes a putative role in carcinogenesis possible and certainly worthy of further exploration (Yang *et al.*, 2015). The gastric carcinogenic bacterium *Helicobacter pylori* (*H. pylori*), known to be important in EMT, can lead to activation of the CT gene *MAGE-A3* raising the possibility of an oncogenic mechanism (Fukuyama *et al.*, 2012). Increased levels of *PAI-1*, a plasminogen activator inhibitor gene, have been shown to be associated with both a worse outcome in gastric cancer and *H. pylori* infection in healthy individuals (Norsett *et al.*, 2011). Downregulation of *PAI-1* was seen when the CT-gene *CAGE* was downregulated (Kim *et al.*, 2013). It has not been shown whether *H. pylori* activates *CAGE* in the gastric mucosa but would be worth investigating in order to assess a possible mechanism promoting EMT and gastric carcinogenesis.

- Alteration of cellular adhesion and proteolytic activity. This putative function of CTAs, which would enhance cell motility and migration, may have important implications for the invasive properties of a tumour. For example, ADAM2 which has been shown to be upregulated in prostate cancer (Romanuik *et al.*, 2009), belongs to a family of proteolytic enzymes implicated in extravillous trophoblast invasion (Pollheimer *et al.*, 2014) and also specifically has a role in fertilisation (Evans, 2001; Pollheimer *et al.*, 2014). Proteins involved in implantation of the embryo and development of the placenta which is a highly vascular structure might also be expected to promote angiogenesis if expressed in a tumour. Indeed, in a chorioallantoic membrane assay, the application of recombinant CAGE protein was shown to lead to the induction of angiogenesis (Kim *et al.*, 2013) providing direct support for this possibility. CAGE has also been shown to induce the activity of matrix metalloproteinases and promote the motility and invasion of cancer cells in both culture and in xenograft models of carcinogenesis (Kim *et al.*, 2009; Kim *et al.*, 2013). Overexpression of *SSX2* in a breast cancer cell line enhanced proliferation and invasive potential of the cells and it was suggested this effect resulted via an influence on oestrogen receptor signalling pathways (Chen *et al.*, 2012). NY-ESO-1 has also been shown to induce the migration but not proliferation of a hepatocellular cell line, though the mechanism for this is uncertain (Xu *et al.*, 2012). It was proposed that NY-ESO-1 may promote metastasis of tumour cells due to this positive effect on cell migration.

- Evasion of apoptosis, another hallmark of cancer, may also occur as a result of CT gene expression in cancer. For example, overexpression of *MAGE-A4* has been shown to induce apoptosis in spontaneously transformed normal keratinocytes (Bhan *et al.*, 2012). A testis-specific serine/threonine kinase (TSSK4) has also been shown to induce apoptosis in various cell lines (Wang *et al.*, 2015). CAGE, a CTA found to be present in many cancers, has been shown to inhibit p53 activity which has broad and potent tumour suppressive activity including the induction of apoptosis (Kim *et al.*, 2010). Similarly, MAGE proteins have been shown to be negative regulators of p53 function (Ladelfa *et al.*, 2012). It has also been shown that downregulation of *CAGE* expression led to enhanced apoptotic activity and increased susceptibility in an *ex vivo* model to chemotherapeutic agents (Kim *et al.*, 2010). In a xenograft model of tumourigenesis, CAGE also displays invasive and angiogenic responses (Kim *et al.*, 2013).

- Promoting immortality by maintenance of telomere length. Telomeres are specialised ribonucleoprotein structures that exist on the ends of chromosomes. Reduction in telomere length can render a cell senescent but also contributes to genomic instability and in this way promotes cancer formation (Bertorelle *et al.*, 2014; O'Sullivan and Karlseder, 2010). Telomeres shorten very slightly in somatic cells with each round of cell division but some cell types (e.g., stem cells and germ cells) are particularly active in maintaining telomere length (e.g., through increased telomerase activity). Cancer cells by their very nature are usually rapidly dividing; this should make them susceptible to cellular senescence induced by a reduction in telomere length. However, many cancers actively maintain telomere length by increasing telomerase activity in a similar fashion to that seen in stem cell populations (Bertorelle *et al.*, 2014). There is another pathway independent of telomerase activity employed in a proportion of human cancers, termed alternative lengthening of telomeres (ALT). It has recently been shown that ALT telomeres are dependent on meiotic factors for their maintenance via a mechanism inextricably linked to their normal function in meiotic recombination (Cho *et al.*, 2014). This study provides clear evidence of how cancers can become 'addicted' to germline genes as had been proposed previously (Janic *et al.*, 2010; McFarlane *et al.*, 2014; Rousseaux *et al.*, 2013a; Simpson *et al.*, 2005; Wang *et al.*, 2011) and moreover, provides direct evidence of a link to cellular biology seen in meiosis.

Thus, it can be seen that despite relatively little being known about the detailed function of most CTAs, many of the hallmarks of cancer can be explained by the proposed and known functions that have been put forward. Focusing on the normal function of the CTAs, for example, their role within the placenta or during gametogenesis, may lead to a greater understanding of their role in cancer and to new avenues in which to focus further research.

We had particular interest in colorectal cancer (CRC) as it is a common and important cancer type with varying but generally low expression of these antigens. Could it be that a subset of CRC does express some germline genes that could potentially be of benefit either diagnostically or therapeutically? Ovarian cancer is an important cause of morbidity and mortality in women and we were interested in this cancer type as well, which acts as a useful comparison. The following discussion will introduce colorectal and ovarian cancer, and briefly outline what is known about CT gene expression in these cancer types but deals with CRC in more detail as more data is presented later in the thesis on this cancer type.

1.6. Colorectal cancer and CT genes

Colorectal cancer (CRC) is among the most common malignancies in both men and women worldwide and a leading cause of cancer related mortality (Mathers *et al.*, 2009; Siegel *et al.*, 2015). When treated early by surgical or endoscopic resection, the disease carries a relatively good prognosis, but when it presents in a more advanced stage the prognosis is generally poor. All too often, because of the frequent lack of symptoms in the early stages, the disease presents at a late stage. This latter group of patients are often offered chemotherapy to either palliate symptoms or increase chances of cure when patients are undergoing surgical resection. The introduction of bowel screening programmes (e.g., faecal occult blood or immunochemical testing) attempts to reduce the proportion of patients presenting with advanced stages of disease by picking it up before symptoms occur (Smith *et al.*, 2015; Stracci *et al.*, 2014). Radiotherapy is commonly given as adjuvant treatment for tumours within the rectum.

Before going on to discuss the pathogenesis of colorectal cancer, it is important to briefly outline the normal structure and histology of the colon. This is of relevance later in the thesis when immunohistochemistry results, of both normal as well as cancerous colonic tissue, are presented.

1.6.1. Structure of the normal colon

The embryonic hindgut gives rise to left-sided colon structures; that is, the distal transverse colon, descending colon, sigmoid colon and rectum. The midgut gives rise to the right colon (i.e. caecum, ascending and proximal transverse colon) as well as the small bowel distal to the ligament of Treitz (Stranding, 2015). The right and left colon merge at the splenic flexure and the left colon merges with the rectum in the pelvis at the rectosigmoid junction. Cancers most commonly occur in the sigmoid colon and rectum, each accounting for approximately 30% of cases of colorectal cancer (Albuquerque *et al.*, 2011).

The layers of the gastrointestinal tract conform to the same broad histological features throughout the gastrointestinal (GI) tract – see Figure 1.12. On the luminal surface of the GI tract is a mucosal epithelial layer; it is from this layer that the majority of cancers (i.e. all carcinomas) arise. Beneath this is the supporting lamina propria, then a thin layer of muscle termed the muscularis mucosae. Deep to this is the submucosa, which contains blood vessels and lymphatics, before a deeper and larger muscular layer termed the muscularis propria.

Finally, beneath this is an adventitial layer of serosa or peritoneum (Stranding, 2015). Large mucosal folds or ridges, called haustra, are formed by contraction of the muscularis propria. There is also an outer longitudinal layer of muscle, which is organised into three bands termed *taeniae coli*; these run from the rectum to the caecum and all three terminate on the appendix.

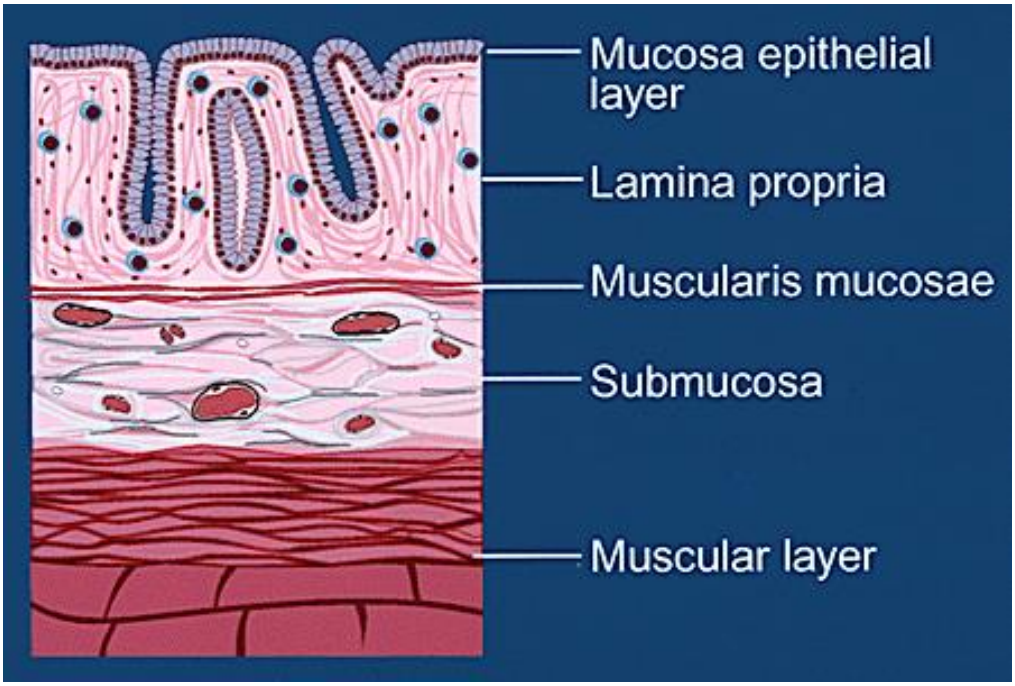


Figure 1.12. Layers of the GI tract.
The colonic mucosa is composed of a single layer of columnar epithelium. One of the primary functions of the colon is to absorb water and the surface area of the colon is increased by involutions of the surface epithelium into crypts. The normal crypt architecture, roughly similar to a "rack of test tubes" is maintained throughout the colon. The crypts contain a more heterogenous mix of cells than the surface epithelium, including the stem cell compartment. Taken from (Strong and de la Motte, 2015).

1.6.2. Colorectal cancer pathogenesis

There are several pathways to the development of CRC that have been characterised in some detail over the past three decades (Al-Sohaily *et al.*, 2012). The majority of CRCs are sporadic and most of these conform broadly to the Vogelstein model of carcinogenesis (see Figure 1.13); that is they arise from successive acquired mutations in genes such as *APC* and *KRAS* and lead to CRC via the chromosomal instability (CIN) pathway (Al-Sohaily *et al.*, 2012; Fearon and

Vogelstein, 1990). Sporadic cases of CRC usually occur in later life, with an average age at diagnosis of around 70 years and the vast majority of patients being diagnosed over the age of 50 years. The microsatellite instability (MSI) pathway and CpG Island Methylator Phenotype (CIMP) pathway are the other pathways described, together accounting for approximately one third of all cases of CRC (Al-Sohaily *et al.*, 2012; Kang, 2011). So, at least three distinct 'routes' to CRC exist but in reality there is a complex overlap between genetic and environmental factors that leads to disease development.

Although over half of CRCs develop within the left (or distal) colon and rectum, tumours that develop via the MSI pathway have a propensity to develop within the proximal colon. The exact reasons for this phenomenon are not known but may be related to the level of β-catenin signalling within the colon (Albuquerque *et al.*, 2011).

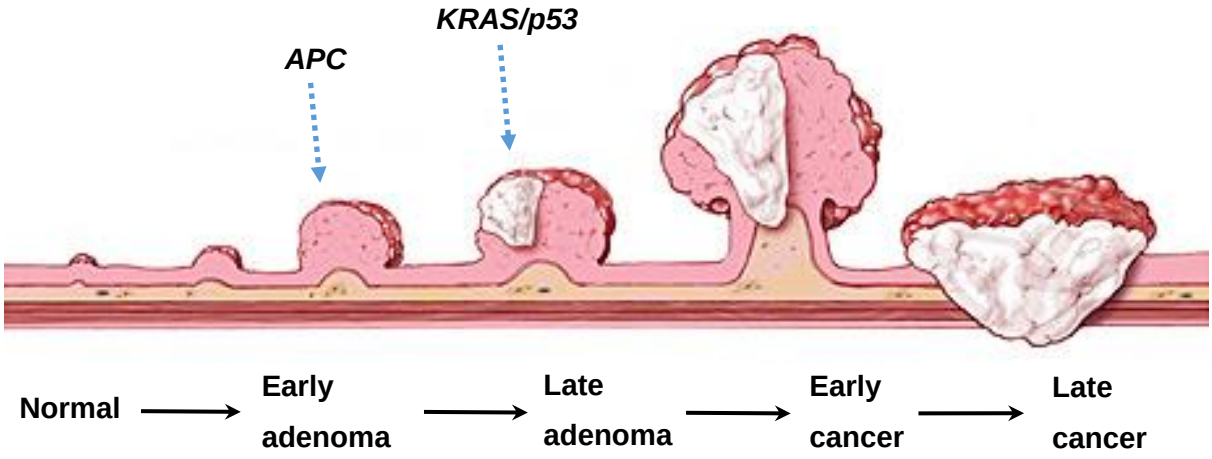


Figure 1.13 The Vogelstein model of carcinogenesis or "Vogelgram".
An initial insult to the colonic epithelium leads to genetic change; mutation in the *APC* gene is often an early event. Subsequent acquired genetic (e.g., mutations in genes such as *p53* and *KRAS*) and epigenetic changes (e.g., demethylation of DNA sequences) leads to greater and greater genetic instability. Small polyps grow is size, develop severely dysplasia and eventually become cancerous, when they have the ability to invade and metastasize. Adapted from (Hopkins Colon Cancer Center, 2015) and (Rao and Yamada, 2013).

1.6.2.1. Genetic predisposition

A minority of CRC cases are due to inherited genetic defects that lead to an increased likelihood of developing the disease. Collectively the inherited syndromes associated with CRC account for at least 5% of all cases of the disease, but as yet ill-defined inherited factors may explain a much larger proportion of the overall disease burden (Gala and Chung, 2011; van Wezel *et al.*, 2012). Identifying patients who have an inherited disorder predisposing them to CRC is often not easy, and relies on obtaining a detailed family history. However, the diagnosis of an inherited disorder is important as it not only has implications for the management of the disease but also for relatives of the individuals who may for instance be offered regular colonoscopic screening and/or genetic counselling/testing. There are two main types of inherited disorder that lead to an increased risk of colorectal cancer:

- Familial adenomatous polyposis (FAP) is an autosomal dominant condition caused by mutation of the *APC* gene that results in loss of normal function leading to the development of hundreds of colonic polyps and the almost inevitable development of CRC before the age of 40 years. Patients known to have this condition usually undergo surgery to remove their entire large bowel before their third decade of life (Gala and Chung, 2011).

- The most common inherited disorder associated with CRC is Lynch Syndrome or Hereditary Non Polyposis Colorectal Cancer (HNPCC) (Al-Sohaily *et al.*, 2012; Hewish *et al.*, 2010). Lynch syndrome is also inherited in an autosomal dominant fashion and is caused by germline mutations in DNA mismatch repair (MMR) genes.

There are numerous MMR genes that are implicated in Lynch syndrome but common examples of genes that are mutated include: mutL homolog 1 (*MLH1*), mutS homolog 2 (*MSH2*), mutS homolog 6 (*MSH6*) and postmeiotic segregation increased 2 (*PMS2*) (Hewish *et al.*, 2010; Poulogiannis *et al.*, 2010; Soreide *et al.*, 2006). Failure of MMR through such gene mutations leads to a state termed microsatellite instability (MSI). Somatic mutations in the same or other MMR genes can lead to CRC via the MSI pathway in much the same way that germline mutations do (Boland and Goel, 2010; Soreide *et al.*, 2006; Vilar and Gruber, 2010). Mismatches in nucleotide base pairing are common during DNA replication and the MMR system is of vital importance in recognising and repairing these mistakes. Microsatellites are repetitive sequences of DNA that occur throughout the genome and are particularly susceptible

to errors during DNA replication (Jansen *et al.*, 2012; Soreide *et al.*, 2006). If the errors are not corrected in genes involved, for example, in cell growth or replication then this can lead to a phenotype that promotes tumourigenesis (Shah *et al.*, 2010). MSI leads to mutations in *β-catenin* and *TGFβRII* more commonly than genes such as *KRAS* and *p53* that are frequently mutated in the Vogelstein model/CIN pathway described above (Hsieh and Yamane, 2008).

Overall 15-20% of all CRCs develop via the MSI pathway, of which a proportion of patients will come from a family with Lynch syndrome. Failure of MMR, as seen in Lynch Syndrome, leads to an increased susceptibility to cancer more generally. Indeed, Lynch Syndrome is associated with an increased risk of various malignancies, including: endometrial, ovarian, brain, skin as well as tumours of the upper gastrointestinal and urological systems (Al-Sohaily *et al.*, 2012). The Amsterdam criteria and Bethesda guidelines are the two commonly used tools to identify patients at high risk of an inherited genetic disorder from their family history; specific genetic testing is then performed to reach a diagnosis (Weissman *et al.*, 2011). The term HNPCC is no longer favoured, largely because it ignores the fact that the condition is associated with an increased risk of many other malignancies and also because these patients can develop polyps within the colon.

Once the diagnosis of CRC is established, the treatment of colorectal cancer is similar regardless of the genetic background with surgical resection when possible affording the best chance of cure. Treatment is, however, becoming more personalised. Although biological agents are used for selected patients with advanced disease, immunotherapy is not as commonly utilised for the treatment of CRC as it is for other types of cancer (Boghossian *et al.*, 2012; Dalerba *et al.*, 2003; Noguchi *et al.*, 2013a). Recent advances in our understanding of tumour immunology have reignited interest in such therapeutic strategies across a range of tumour types including CRC for both early at late stage disease (Boncheva *et al.*, 2013; Hanna, 2012; Markman and Shiao, 2015; Rosenberg and Restifo, 2015; Schumacher and Schreiber, 2015; Sondak and McArthur, 2015; Topalian *et al.*, 2011). As CTAs contain amongst the most immunogenic antigens, there is a drive to establish whether germline genes are activated in colorectal and other cancers. Some of what is known about CT gene expression will now be outlined - see (Sammut *et al.*, 2013) for a more comprehensive review.

1.6.3. CT gene expression in colorectal cancer

Previous studies have indicated that CT genes are not widely or universally expressed in all tumour types (Almeida *et al.*, 2009; Caballero and Chen, 2009; Cheng *et al.*, 2011; Costa *et al.*, 2007; Fratta *et al.*, 2011; Hofmann *et al.*, 2008; Simpson *et al.*, 2005; Viphakone *et al.*, 2015). CRC falls within the group of tumours that have limited CTA gene expression but considerable variability is seen between studies. Moreover, this generalised classification does not account for the possibility that there is a sub-group of CRCs that do have a high incidence of CTA gene expression that could provide valuable clinical targets that currently remain unexplored. Given the patient-specific tailored therapy that advanced immunotherapy approaches represent it could be the case that a cohort of patients might benefit from the therapeutic potential of CTA genes, assuming that CTA gene expression is associated with antigen production that can be therapeutically harnessed. Some tumours have been found to co-express CTA genes (Wang *et al.*, 2002), and thus, it may be the case that the majority of CRCs do not express any CTA genes, but rather a minority express several. Two possible reasons for co-expression of CT genes were outlined above (see Section 1.5). The *MAGE* family of genes serve as an example of both co-expression and the variability of expression profiles seen between studies. These facts will be revisited later in the thesis and further discussion is provided in Chapter 5 (Section 5.1.3).

1.6.4. CTA correlation with clinical outcome

Choi and Chang found the expression of *MAGE* and *SSX* genes to be correlated with the presence of liver metastases in CRC (Choi and Chang, 2012). In this study they used primer sets that were complimentary to several *MAGE* (*MAGEA1-A6*) and *SSX* (*SSX1-9*) genes; from 37 samples tested a combined expression frequency of 51.4% was found for *MAGE* genes and 32.4% for *SSX* genes. Protein levels of another CTA, OY-TES-1, were found to correlate with tumour grade and invasion (Luo *et al.*, 2013). Interestingly, in this study low levels of gene activation (but not protein) were seen in normal adjacent colonic tissue but not normal control samples. This fact has implications when considering the results of gene expression profiles and immunohistochemistry data, which will be presented later in this thesis.

1.6.5. Therapeutic use targeting CTAs in CRC

A phase II trial assessing the use of a therapeutic vaccine based on dendritic cells that have been pulsed with a tumour cell lysate (prepared from a *MAGE*-expressing melanoma cell line)

showed some promising results and low toxicity in patients with metastatic CRC (Burgdorf *et al.*, 2008). However, in an earlier study a very weak overall clinical response rate of less than 1% was found for the use of cancer vaccines to treat CRC, despite induction of an immune response in around half of patients (Noguchi *et al.*, 2013b). However, these vaccines were not all CTA derived and so CTA based vaccines may result in distinct and possibly higher clinical response rates, but this requires formal testing. A partial clinical response with no major side effects of treatment to a cancer vaccine based on MAGE-A4 has also been observed in a single patient with metastatic colorectal cancer (Takahashi *et al.*, 2012).

1.7. CT gene expression in ovarian cancer

1.7.1. Ovarian cancer overview

Ovarian cancer is one of the leading causes of cancer deaths in women and the commonest cause of gynaecological cancer deaths (Siegel *et al.*, 2015; Sung *et al.*, 2014). There have only been modest improvements in survival since the 1980s: in part this is due to a lack of effective screening methods as well as the vast heterogenous nature of the disease (Berns and Bowtell, 2012; Erickson *et al.*, 2013; Menon *et al.*, 2014; Vaughan *et al.*, 2011). Although grouped as "ovarian cancer" due to the anatomical location of the tumours, there are five distinct subtypes of the disease and it is now widely accepted that many tumours do not in fact originate from the ovaries (Berns and Bowtell, 2012; Erickson *et al.*, 2013; Nelson, 2015; Sorensen *et al.*, 2015). The different molecular subtypes of ovarian carcinomas are: high-grade serous, clear-cell, endometrioid, mucinous and low-grade serous. Serous carcinomas for instance are thought to arise from the ovary, fallopian tube or peritoneum and mucinous carcinomas are often metastatic from the gastrointestinal (GI) tract (Erickson *et al.*, 2013; Micci *et al.*, 2014; Vaughan *et al.*, 2011).

Given the insidious nature of the presentation, coupled with the poor prognosis associated with the disease, there has been a drive to establish new biomarkers for the disease. This has gone hand in hand with an improved understanding of the molecular events that occur in ovarian cancer pathogenesis (Berns and Bowtell, 2012; Micci *et al.*, 2014). CA-125 (see Table 1.3) has been used for many years to aid diagnosis but HE4 is the only new biomarker to establish itself

in recent years (Simmons *et al.*, 2013). HE4, or human epididymis protein-4, has too broad an expression in normal tissues to classify it as a testis-selective CTA.

1.7.2. CT genes in ovarian cancer

Numerous studies have shown CT genes to be expressed in ovarian cancer (reviewed in Mirandola *et al.*, 2011(Mirandola *et al.*, 2011). Some studies have indicated that expression of CT genes predicts a less favourable clinical outcome in ovarian cancer (Daudi *et al.*, 2014; Hasegawa *et al.*, 2004; Sharma *et al.*, 2005; Tammela *et al.*, 2004; Tureci *et al.*, 1998b; Yakirevich *et al.*, 2003; Zimmermann *et al.*, 2013). It is certainly not a universal finding that the presence of CTAs or CT gene expression is associated with a worse prognosis. Using a bioinformatics approach it has been shown that an immunoreactive subgroup of CTAs is associated with a favourable prognosis in ovarian cancer (Eng and Tsuji, 2014). Immune responses to certain CTAs may result in the body's ability to (partially) fight cancer and impact positively on prognosis; this has also been speculated as a possible reason contributing to the favourable prognosis of CRC that develops via the MSI pathway (Sammut *et al.*, 2013). The presence of tumour-infiltrating lymphocytes (TILs) has been associated with a better prognosis in various solid tumours, including ovarian cancer (Sato *et al.*, 2005). In this study by Sato and colleagues they did not see an association between $CD8^+$ TILs (that were associated with the favourable prognosis) and the presence of three CTAs in the tumour. Equally, however, they did not identify which TAAs were responsible for the TIL response.

It could be that certain CTAs are able to predict a worse outcome in specific subsets of ovarian cancer; differentiating which CTAs are best for which subtypes requires further resolution. It may well be that combining various biomarkers will prove of greatest clinical potential. The presence of certain CTAs was found to accurately differentiate between borderline tumours and tumours of higher grade (Zimmermann *et al.*, 2013) – such clinically relevant predictive power could be very useful when planning surgical intervention to remove tumours and how aggressive the resection should be. It was over a decade ago that MAGEA4, detected using IHC, was found to be an independent predictor of survival in ovarian cancer (Yakirevich *et al.*, 2003). More recently MAGEA4 has been suggested to be a central player, regulating the expression of other MAGE family members (Daudi *et al.*, 2014).

Combining CTA detection with conventional methods of disease stratification could conceivably increase the prognostic accuracy. For instance, an even worse outcome has been

shown when looking at tumours that are both of higher grade and containing CTAs (Yakirevich *et al.*, 2003). Although not found to be an independent predictor of prognosis, sperm protein 17 (SP17) was found in nearly half of the 70 ovarian cancer specimens examined in one study (Li *et al.*, 2009). More interestingly it was found in all cells isolated from the ascitic fluid of patients and in addition increased migration and resistance to chemotherapeutic agents of ovarian cancer cells *in vitro*. Ovarian cancer often presents with clinically evident ascites. Isolating mRNA from peritoneal fluid samples and then using standard PCR amplification of different CT genes proved a very sensitive method for predicting that the fluid was a malignant exudate in another study (Hofmann and Ruschenburg, 2002). Thus, using ascitic fluid as the source of biomarker development could potentially prove more useful than testing blood samples from patients and should be investigated further. A similar approach could also be used for differentiating between benign and malignant causes of pleural effusions and appears to not have been extensively investigated.

There has been a significant and increasing interest in the use of immunotherapy to treat ovarian cancer in the last decade (Tse *et al.*, 2014; Wefers *et al.*, 2015). Immunotherapy has experienced resurgence itself (see Section 1.2) but our increased understanding of the pathogenesis of ovarian cancer coupled with the encouraging findings of CTAs being present has contributed to this increased interest. Disappointingly, however, the results of early phase trials using immunotherapy have collectively shown little benefit, though the majority have not used CTAs as the target for antibody-based approaches (Leffers *et al.*, 2014). Recently, using xenograft and *in vitro* models of ovarian cancer, the use of a demethylating agent has been shown to not only induce CT gene expression but also enhance the effectiveness of an immunotherapeutic approach (Srivastava *et al.*, 2015). Such immunomodulatory effects of epigenetic manipulation may enhance the effectiveness of immunotherapy in the clinical setting in the coming years. It has also been shown that the expression of some CT genes is associated with immunoreactive markers that are associated with a favourable prognosis in serous ovarian carcinoma (Eng and Tsuji, 2014). The fact that other CT genes are expressed but not associated with this favourable 'immunoreactive' subgroup underlines the complex interplay between genes expression, antigen production and recognition, and anti-cancer immunomodulatory effects *vs.* oncogenic effects of CTAs.

1.8. CTAs and cancer treatment

As discussed above, germline genes are a rich source of potential cancer-specific biomarkers. Clinically relevant biomarkers can be tremendously useful in developing and establishing treatment strategies in the clinical setting and tailoring them in a patient-specific fashion. Germline genes can also encode CTAs, which have been the target for effective immune-based treatments. The most notable examples are in melanoma, where dramatic clinical responses have been seen in an adoptive therapeutic approach targeted at the CTA NY-ESO-1 (Hunder *et al.*, 2008). Another immunotherapeutic drug, Ipilimumab, has also shown considerable efficacy in the treatment of melanoma (Luke and Hodi, 2013) and mixed benefit in other tumour types such as ovarian cancer (Tse *et al.*, 2014). This drug does not directly target CTAs but has been shown to increase T-cell responses to NY-ESO-1 (Yuan *et al.*, 2008) so this may partly explain the mixed clinical benefit observed in other tumour types with varying degrees of *NY-ESO-1* expression. Given the increasing evidence of germline factors being important in oncogenesis together with current knowledge of CTAs as both biomarkers and therapeutic targets, it is likely that in the coming years genes and proteins normally found in the germline will be of increasing importance in various aspects of cancer treatment.

1.9. Project aims

The overarching aim of this project was to identify potential novel CT genes that could be used in therapeutics, diagnostics and/or disease stratification. A number of related approaches were used in an attempt to fulfil the following subsidiary project aims.

Firstly, we aimed to establish if there were any additional meiosis-associated cancer testis (meiCT) genes that could potentially act as cancer-specific biomarkers. The MacFarlane group had set up and designed a computational pipeline which had already proved a useful method of identifying putative cancer-specific biomarkers (Feichtinger *et al*., 2012a; Feichtinger *et al*., 2012b; Feichtinger *et al*., 2014). There was a cohort of genes from this panel of genes which remained unvalidated. We aimed to define the expression profile of these genes in a range of normal and cancerous tissues as a means of filtering out less promising candidates. The genes as a group were loosely termed meiCT genes as it was predicted that some at least may be functionally related to meiosis.

Secondly, we aimed to establish if there were any antigens (of meiCT genes) produced and present in normal testicular tissue. This would serve as a route to establishing if the proteins were also produced in cancerous tissue. Attempting to answer this question was an important next step with regards to potential use as an immunotherapeutic or drug target, as well as potential cell-surface marker for disease stratification purposes based on tissue and not only gene expression profile analysis. We also aimed to investigate the potential immunogenicity of these proteins.

Additionally, we aimed to draw-together information from the published literature to compile a list of candidate germline genes. This is something that hitherto had not been done. From this list, we aimed to go some way to establishing if there was a cohort of particularly promising candidates. To fulfil this aim, the gene expression profiles for a larger cohort of genes would be tested in normal and cancerous tissue. Further, a small biobank of tissue would be collected in which the expression or not of these germline genes could be compared directly in cancerous tissue as well as matched normal tissue from the same patients. By focusing on CRC, we also aimed to establish if there was a subset of colorectal tumours that did express some of the germline-associated genes more widely, which would counter the generally held dogma that CTAs were infrequently found in CRC.

2. Materials and Methods

2.1. Source of human cell lines

The NTERA-2 (clone D1) cell line was gifted by Prof. P.W. Andrews (University of Sheffield) and the A2780 cell line was gifted by Prof. P. Workman (Cancer Research UK Centre for Cancer Therapeutics, Surrey, UK). The following cell lines were purchased from the European Collection of Cell Cultures (ECACC); 1321N1, COLO800, COLO857, G-361, HCT116, HT29, LoVo, MCF7, MM127, SW480 and T84. H460 and MDA-MB-453 were purchased from the American Type Culture Collection (ATCC), and two ovarian adenocarcinoma cell lines, PEO14 and TO14, were obtained from Cancer Research Technology Ltd. The T2 cells were gifted by Dr Steven Man (Cardiff University, Wales, UK). Cell lines were regularly tested using the LGC Standards Cell Line Authentication service.

2.2. Cell culture

The cells were grown at 37˚C in a humidified incubator, in a CO_2 enriched environment. The media used for each cell line was supplemented with foetal bovine serum (FBS) from GIBCO, Invitrogen (Catalogue number; 10270, Lot 41Q6208K). The cell lines and their growth conditions are detailed in Table 2.1. All cell lines were routinely checked for mycoplasma contamination using the LookOut® Mycoplasma PCR Detection kit (Sigma Aldrich; MP0035) as per the manufacturer's instructions. Cell lines were authenticated following the key experiments through a short tandem repeat profiling technique, using LGC Standards Cell Line Authentication service – reports are provided in Appendix D (on CD).

2.3. Preparation of frozen cell stocks

Confluent cells were trypsinised using 1x trypsin-EDTA (Sigma-Aldrich; T3924) and collected. The cells were counted and then centrifuged for 5 minutes at 400 r.c.f. The cells were then resuspended in freezing media (10% DMSO, 90% FBS) and transferred into a cryotube and placed at -80˚C for 24 hours before being transferred to liquid nitrogen for long term storage.

Table 2.1. Description of the cells lines and their growth conditions

Cell Line	Cell Line Description	CO_2	Media
1321N1	Human brain astrocytoma	5%	Dubeco's modified Eagle's medium (DMEM) + GLATAMAX™ (Invitrogen; 61965) + 10% FBS
NTERA-2 (clone D1)	Caucasian pluripotent embryonal carcinoma	10%	
SW480	Human colon adenocarcinoma	5%	
A2780	Human ovarian carcinoma	5%	DMEM + GLATAMAX™ + 10% FBS and 1xNEAA (non-essential amino acids)
MCF7	Human Caucasian breast adenocarcinoma	5%	
COLO800	Human melanoma	5%	Roswell Park Memorial Institute 1640 medium (RPMI 1640) + GLUTAMAX™ (Invitrogen; 61870) + 10% FBS
COLO857	Human melanoma	5%	
H460	Large cell lung carcinoma	5%	
PEO14*	Ovarian Adenocarcinoma, peritoneal ascites	5%	RPMI 1640 + GLUTAMAX™ + 10% FBS and 2 mM sodium pyruvate
TO14*	Ovarian Adenocarcinoma, solid metastasis	5%	
MM127	Human malignant melanoma	5%	RPMI 1640 + GLUTAMAX™ + 10% FBS and 25 mM HEPES
T2 Cells	T cell hybrid expressing HLA-A2	5%	RPMI 1640 + GLUTAMAX™ + 10% FBS + 100 units penicillin and 100 µg streptomycin
G-361	Human Caucasian malignant melanoma	5%	McCoy's 5A medium + GLUTAMAX™ (Invitrogen; 36600) + 10% FBS
HCT116	Human colon carcinoma	5%	
HT29	Human Caucasian colon adenocarcinoma	5%	
T84	Human colon carcinoma	5%	Ham's F12 + DMEM (1:1) + GLUTAMAX™ (Invitrogen; 31331) + 10% FBS
DA-MB-453	Human breast carcinoma	0%	Leibovitz's (L-15) medium + GLUTAMAX™ (Invitrogen; 31415) + 10% FBS

* These cells lines were from the same patient, collected prior to treatment.

2.4. Total RNA extraction

Total RNA was extracted from human tissue samples using the RNeasy Mini Kit (Qiagen; #74104) according to the manufacturer's instructions. Briefly, 15-30 mg of RNAlater® (Ambion™; #AM7020) stabilised tissue was added to 600 µl of Buffer RLT with 6 µl of 14.3 M β-mercaptoethanol (β-ME) in a gentleMACS™ M-Tube (Miltenyi Biotec; #130-093-236).

The tissue was homogenised in a gentleMACS^{TM} rotor-stator homogeniser (Miltenyi Biotec; #130-093-235) using the RNA_02 setting. An optional DNase digestion step was routinely performed by adding 10 µl of DNase I stock solution (Qiagen; #79254) directly to the RNeasy spin column and leaving on the benchtop to incubate for 15 minutes at room temperature. RNA was eluted from the RNeasy spin column into a fresh 1.5 ml collection tube by adding 30 µl of RNase-free water directly to the spin column membrane and centrifuging at 13,000 r.c.f. This elution step was repeated once to produce a final volume of RNA of ~60 µl, which was then stored at -80˚C prior to use. Total RNA extracts were prepared for all of the cancer cell lines grown in the lab (see Table 2.1) using a standard protocol (as described in Feichtinger *et al.* 2012a). The concentration and quality of RNA extracted from the cell lines was measured using a NanoDrop (ND_1000).

2.5. cDNA synthesis

2.5.1. cDNA synthesis for RT-PCR screening of mei-CT candidates

Total RNA preparations from the human tissue panel (Clontech; #636643) and a range of tumour tissues and cell lines were purchased from Clontech and Ambion – refer to Appendix C (on CD) for details supplied by the company on this RNA. The quantity of RNA was assessed as necessary using Nanodrop technology; 1.0 µg of total RNA was reverse transcribed into cDNA using the SuperScript III First Strand synthesis system (Invitrogen; #18080-051) as per the manufacturer's instructions and the cDNA was then diluted eight times. The quality of the cDNA was tested using a control RT-PCR using primers for *ACTB.*

2.5.2. cDNA synthesis for qPCR screening on TLDA cards

Total RNA preparations were from the human tissue panel (Clontech; #636643), a range of tumour samples purchased from Origene (see Appendix C, on CD) or extracted directly from tumour and adjacent normal tissue. The quantity and quality of the RNA was assessed as necessary using BioRad Experion RNA analysis chip according to the manufacturer's instructions. RNA with RQI values lower than 7.5 were not used for TLDA card qPCR analysis. 1.0 µg of total RNA was reverse transcribed into cDNA using the SuperScript III First Strand synthesis system (Invitrogen; #18080-051 and #11752-250) as per the manufacturer's instructions and the cDNA was then diluted five times. The quality of the cDNA was tested using a control RT-PCR using primers for *ACTB* prior to use of the TLDA cards.

2.6. Reverse transcription (RT-) PCR

The sequences for each of the genes analysed were obtained from the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/) and primers specific to each of the genes were designed to span introns where possible. An oligonucleotide properties calculator (OligoCalc, available from: simgene.com/OligoCalc) and the NCBI primer-BLAST and primer designing tool (available from: http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome) were used to aid primer design.

2.6.1. Method 1

A volume of 2 µl diluted cDNA was used in the PCR with a final volume of 50 µl. BioMix™ Red 2x (Bioline; BIO-25006) was used for the PCR amplification as per the manufacturer's instructions. Samples were amplified with a pre-cycling hold at 96°C for 5 minutes, followed by 40 cycles of denaturing at 96°C for 30 seconds, annealing for 30 seconds and extension at 72°C for 40 seconds, followed by a final extension step at 72°C for 5 minutes. For *ACTB* amplification a total of 35 reaction cycles was used.

2.6.2. Method 2

My Taq Red x2 (Bioline: BIO-25043) was also used for PCR amplification as per the manufacturer's instructions. Samples were amplified with a pre-cycling hold at 95°C for 1 minute, followed by 40 cycles of denaturing at 95°C for 15 seconds, annealing for 15 seconds and extension at 72°C for 10 seconds, followed by a final extension step at 72°C for 5 minutes. Again, for *ACTB* amplification a cycle length of 35 was used instead of 40.

Alternatively half the reaction volume was used (i.e. 1ul of cDNA to produce a final PCR volume of 25 µl). RT-PCR for *ACTB* and *NYESO-1* (a known X-CT gene) was carried out for all cancer and noncancer cDNA samples as controls. The products were separated on 1% agarose gels stained with ethidium bromide or peqGREEN (Peqlab; 37-5000).

2.6.2.1. Annealing temperatures used and RT-PCR cycling conditions

Two reagents were used (BioMix™ Red 2x and MyTaq™ Red Mix, Bioline) for RT-PCR as described above. The annealing temperatures used are shown in Table 2.2 – the letters

correspond to those used in Table 1 in Section 2 of Appendix, where the list of primers used is provided.

Table 2.2. Cycling conditions and annealing temperatures used for RT-PCR

Bioline Mix	Annealing temperature	Cycling Code*
BioMix™ Red 2x	52˚C	A
BioMix™ Red 2x	54˚C	B
BioMix™ Red 2x	55˚C	C
BioMix™ Red 2x	56˚C	D
BioMix™ Red 2x	60˚C	E
MyTaq™ Red	54˚C	F
MyTaq™ Red	55˚C	G
MyTaq™ Red	56˚C	H
MyTaq™ Red	58˚C	I
MyTaq™ Red	60˚C	J

*See Table 1 in Appendix for which conditions used for individual primer sets

2.6.3. Purification of RT-PCR products using the Roche purification kit

PCR products were run on a 1% agarose gel containing ethidium bromide and the band(s) were cut out of the gel using a sterile blade. The products were then purified using the High Pure PCR Product Purification Kit (Roche Applied Science; #11732676001), as per the manufacturer's instructions. The purified PCR product was eluted from the column in nuclease free water.

2.6.4. Sequencing RT-PCR products

A minimum amount of 5 ng/µl DNA was sent at room temperature in a clean Eppendorf tube to Eurofins MWG, with 15 pmol of the corresponding forward and/or reverse primers. The sequencing results were blasted and aligned against the expected PCR product sequence, using the NCBI Basic Local Alignment Search Tool.

2.6.5. RT-PCR screening using the QIAxcel system

PCR products were screened using QIAxcel DNA screening kit (Qiagen). The QIAxcel was run as per the manufacturer's instructions using the AM320 method. The 15 bp – 1 Kb alignment marker and the 50-800 bp v2.0 size marker were used.

2.7. Taqman Low Density Array

After confirming RNA quality with BioRad Experion™ analysis kit and performing cDNA synthesis as above, the cDNA was diluted approximately 1:5 with RT-PCR grade water (Ambion®, AM9935), producing a final volume of 110 µl. RT-PCR was performed using 1 µl of this cDNA and *ACTB* primers to confirm reverse transcription was successful. 101-102 µl of the cDNA was mixed with an equal volume of Taqman® Universal Mastermix II, with UNG (Applied Biosystems™, #4440038) after allowing the reagent to equilibrate to room temperature for 30 minutes. 100 µl of this sample-specific PCR reaction mix was dispensed into the fill reservoir of the TLDA card. The card was then spun down at 1200 r.p.m. for 1 minute (Sorvall® Legent T centrifuge; Heraeus™ rotor 75006445) and repeated for a further 1 minute spin. The card was sealed using the Life Technologies mechanical sealing device; the ports were then cut off using scissors before loading on the Applied Biosystems® 7900HT Real-Time PCR instrument. Results were analysed using Life Technologies ExpressionSuite software.

2.8. Western blot analysis

2.8.1. Cell preparation

Whole cell extracts (WCEs) were prepared from confluent or sub-confluent cell cultures. The cells were trypsinised and the flask washed with approximately 10 ml of complete media to inactivate the trysin. The pooled cells were transferred into Falcon tubes and spun at 1500 r.p.m. for 3 minutes (Jouan C3 centrifuge: approximately 426 r.c.f.). The supernatant was discarded and the cells were resuspended in approximately 10 ml of 1xPBS and spun again at 1500 r.p.m. for 3 minutes. This step was repeated once more.

2.8.2. Cell lysis

The M-PER reagent was prepared (e.g., for 1 ml, 10 µl of x100 HALT protease inhibitor cocktail and 10 µl of x100 EDTA solution to 980 µl of M-PER® reagent). The cell pellet from above was resuspended in 1-1.5 ml of cold 1xPBS and transferred into a 1.5 ml microcentrifuge tube. This was centrifuged at 1500 r.p.m. (Thermoscientific Heraeus™ PICO 17 centrifuge) for 5 minutes at 4˚C and the supernatant removed carefully. The pellet was weighed (subtracting the weight of an empty microcentrifuge tube) and 10µl of prepared M-PER reagent for every 1mg of dry pellet added. The microcentrifuge tube was then placed on a rotary incubator and incubated for 10 minutes at room temperature. After 'full lysis' the cell debris was pelleted by centrifugation at maximum speed (13,300 r.c.f./17,000 r.p.m. Heraeus™ PICO 17 centrifuge) for 15 minutes. The supernatant was then carefully removed and transferred into a new microcentrifuge tube. The protein concentration was estimated by BCA (Thermo Fisher Scientific; #23277) method. The lysis was flash frozen in liquid nitrogen and stored at -80˚C.

2.8.3. Lysis sample preparation and running gel

The samples were prepared as follows:

- Lysate – correct volume for 30 µg
- Bolt™ LDS Sample Buffer (4X) – 5 µl
- Bolt™ Reducing Agent (10X) – 2 µl
- Deionized water to make a final volume of 20 µl

The samples were first incubated at 70˚C for 10 minutes and the 20 µl aliquots were then separated on SDS-polyacrylamide pre cast gels (Bolt™ 4-12% Bis-Tris Plus Gels, 15 well, Life Technologies, BG04125BOX). 8 µl of Precision Plus Protein™ Dual Color Standard (Bio-Rad, #161-0374) was used as a molecular weight marker. The gels were run at 100V for approximately 60 minutes and subsequently electro-blotted onto a methanol-soaked PVDF membrane (Immobilon-P, Millipore, IPVH00010) at 500 mA for 120 minutes in 2x Towbin buffer (380 mM Glycine, 50 mM Tris).

2.8.4. Blocking, antibody and ECL

The membranes were blocked in milk solution (10% milk/1xPBS/0.2% Tween) for 60 minutes at room temperature. The blotted membranes were then probed with the primary antibody (diluted in milk solution) at 4˚C overnight. The membranes were rinsed in milk solution, followed by two 10 minute washes before applying the secondary antibody (diluted in milk solution) for 60 minutes at room temperature. Antibody detection was performed using Pierce ECL Plus/2 Western Blotting substrate (Thermo Scientific, #32132) and CL-Xposure^{TM} film (Thermo Scientific, 34091) in an enhanced chemiluminescence detection system according to the manufacturer's instruction. For GAPDH loading control, standard ECL substrate was used.

2.9. siRNA (small interfering RNA) knockdown

Subconfuent cells were seeded at $5x10^5$ cells per well in 6-well plates that contained fresh complete media. The cells were incubated under normal growth conditions. The transfection mixture for each well was prepared as a mastermix in a clean Eppendorf tube and 106.6 µl was added to each well containing the cells. The 106.6 µl contained 100 µl of serum free media, 6 µl of HiPerFect transfection reagent (Qiagen; #301705) and 0.6 µl of the siRNA (Qiagen). The list of the siRNAs used are shown in Table 2.3. The mixture was incubated at room temperature for 20 minutes to allow formation of transfection complexes. The transfection mixture was then added to the cells in the 6 well plate, which were then shaken gently within the incubation hood. Untreated cells and negative-control siRNA (Qiagen) were used as controls. Cells were treated with siRNA on day 0, with two further 'hits' at 24 and 48 hours, before harvesting the cells at 72 hours post transfection.

Table 2.3. siRNAs used for gene knockdown of *C20orf201*

siRNA	Name	Target Sequence (5' to 3')	Catalogue number
1	Hs_LOC198437_2	ACCGCCAAGAGGTGCAGACAA	Sl00485135*
2	Hs_LOC198437_5	CCCGTGGACGCAGTCGCTCGA	Sl03186386*
3	Hs_LOC198437_6	CCAGCCTCCCACATAAAGTTA	Sl04258772*
4	Hs_LOC198437_7	TCCCGCGGTGACGGCGACTGA	Sl04319574*

*From Qiagen

2.10. Immunohistochemistry (IHC) protocols

2.10.1. Ventana automated IHC procedure

Samples of human colon cancer and adjacent normal tissue was obtained from individuals undergoing colonic resection, following written informed consent. The tissue samples were fixed in formalin and embedded in paraffin. IHC was performed on 4 µm sections of this tissue. Staining was automated on a Ventana Benchmark XT IHC machine (Roche) using a standard IHC protocol (see Appendix A on CD). 3,3'-Diaminobenzidine (DAB) was used as the chromogenic substrate and the sections were counterstained with haemotoxylin. Negative control was provided by addition of the antibody diluent without the primary antibody. Images were obtained on Zeiss Axio Scan.Z1 digital slide scanner.

Antibody optimization steps included both protease and heat methods of antigen retrieval. Standard Roche protocols were used in this regard – for full details see Appendix A (on CD). Heat retrieval produced the most consistent and reliable results both for performing IHC by hand and using the Ventana automated machine. In addition to this a range of concentrations of the primary antibody were used, again both when performing IHC by hand and for the Ventana automated procedure to establish the optimum concentration. For example, 1:50 to 1:2000 dilutions of the primary antibody were used. Preliminary work using TEX19 at a 1:500 dilution had already been undertaken (Planells-Palop, unpublished data). It was discovered that lowering the TEX19 concentration from a 1:500 dilution to 1:800 still produced consistent results, so 1:800 dilution was used to minimise the chance of background staining.

2.10.2. IHC by hand

The tissue sections were cut on a microtome set at 4 microns thickness and mounted onto Leica sticky coat slides. The slides were then baked in a dry oven at 60˚C for 30 minutes and allowed to cool to room temperature. Deparaffinization with xylene was performed for 5 minutes and repeated three times in a fume hood. The slides were then hydrated in 96% ethanol for 5 minutes three times before washing them in dH_2O prior to antigen retrieval. 450ml of dH_2O was added to 50 ml of 10x Dako^{TM} sodium citrate buffer (S1700; pH 6.0). The 1x buffer solution was heated in a glass container on a heating platform to 95-99˚C before immersing slides (held in metal slide rack) into the buffer. The buffer was maintained at a sub-boiling

temperature (98˚C +/- 2˚C) for 10 minutes. The slides were then allowed to cool on the bench top (still immersed in the citrate buffer) for 20-30 minutes. Alternative antigen retrieval with proteinase K was performed for 10 minutes. The purpose of antigen retrieval was to unmask epitopes that were potentially hidden from the primary antibody and immunostaining without antigen retrieval was not attempted. The slides were then immersed in Dako™ wash buffer (S3006) for 10-15 minutes on a gentle rotating platform; changing the wash buffer once during this time. Enough blocking solution to cover the tissue section (e.g., 200-500 µl) was added and left to incubate in a humidified chamber. The blocking solution used was either 5% normal serum/0.3% Triton™ X-100/1xPBS, or more commonly Dako mouse or rabbit serum (e.g., 995 µl of the diluent mixed with 5 µl of rabbit serum, X0902) for at least 10 minutes at room temperature. Overnight incubation with the primary antibody was at 4˚C in the humidified chamber, or alternatively at 30˚C for two hours. Dako™ antibody diluent with background reducing components (S3022) was used to dilute the primary antibody. Overnight incubation of the primary antibody at 4˚C was found to produce more consistent results so was used for the majority of IHC procedures performed by hand. The slides were then washed for 10-15 minutes in Dako™ wash buffer; changing the buffer once or twice over this time. An optional step not usually carried out was to rinse briefly in methanol to remove excess water as a separate container was used. In order to quench endogenous peroxidase, 3% hydrogen peroxide mix was made up (e.g., by adding 8 ml of 30% hydrogen peroxide to 72 ml of methanol). The coplin jar was rinsed with methanol (to get rid of carryover water) if any doubt as to whether it was dry but a different container was usually used for quenching as opposed to washing so this was not generally carried-out. The slides were then immersed (back to back) into the methanol/3% hydrogen peroxide mix for at least 7 minutes. Enough secondary antibody to cover the specimen (e.g., 250 µl) at 1:1000 dilution was then added being careful not to allow the tissue to dry. Incubation was for 30-60 minutes at room temperature. The slides were then washed in Dako™ wash buffer for 10-15 minutes, again changing the wash buffer once or twice over this time. The secondary antibodies used were purchased from Abcam (goat anti-mouse; ab47827 and goat anti-rabbit; ab6721). After adding the secondary antibody the DAB substrate working solution was made up by adding 1 drop (or 20 µl) of the DAB Chromogen (Dako K3467) per ml of Substrate buffer. The working solution was left at room temperature for at least 30 minutes before use. The DAB solution was added to the slides and incubated for typically 4-5 minutes; or sometimes less according to the positive control sample on the day (i.e. DAB added to positive control first when appropriate and then visualised in real time under a light microscope). The slides were placed on a rack and rinsed in a container with running

tap water for around 5 minutes before counterstaining with Haematoxylin (Dako™; CS70030). Using a dropper the haemotoxylin was added and incubated for 10-15 seconds. The slides were again rinsed in running tap water for 5 minutes before mounting. Finally, the slides were dehydrated in 96% ethanol for 5 minutes three times (alternatively sequentially in 80%, then 95% and then 100% ethanol for 2-5 minutes each). The slides were then cleared by rinsing in xylene three times, leaving the slides in the final xylene solution before mounting for around 5 minutes. The coverslips were mounted with DPX mountant (Sigma; #06522) and allowed to dry on bench for at least 30 minutes before imaging.

2.10.3. Variations to IHC protocol by hand

2.10.3.1. Deparaffinization and hydration

Rather than three 5 minute washes in xylene, the slides were dewaxed in fresh xylene for 60 minutes in a fume hood. The slides were then rinsed in a sequential series of 100%, 95%, 80%, 70%, 50% ethanol for 2 minutes each. The slides were then washed in dH_2O prior to antigen retrieval as before.

2.10.3.2. Antigen retrieval

The Dako™ sodium citrate buffer was heated to 80˚C on a heating platform and the slides were immersed in the solution, which was then maintained at 80˚C (+/- 2˚C) for 60 minutes. The slides were then allowed to cool on the bench top (still immersed in the citrate buffer) for at least 20 minutes or until less than 50˚C.

2.11. Immunofluorescence protocol

The immunofluorescence (IF) protocol followed a similar approach to the IHC by hand method just described. Tissue sections were cut on a microtome set at 4 microns thickness and mounted onto Leica sticky coat slides. The slides were then baked in a dry oven at 60˚C for 30 minutes and allowed to cool to room temperature. Slides were deparaffinised in xylene for 5 minutes three times, or alternatively for 60 minutes, in a fume hood. The slides were then washed twice in 100% ethanol for 5 minutes and then twice in 95% ethanol for five minutes. Alternatively, the same procedure as for IHC by hand version 2 above was used (i.e. 2 minute washes in 100%, 95%, 80%, 70%, and then 50% ethanol). The slides were then washed in dH_2O prior to antigen retrieval. Antigen retrieval was performed using heat-induced epitope retrieval by submerging slides in 10 mM sodium citrate buffer (Dako™) pH 6.0 and maintaining at a sub-boiling temperature (95-99˚C on a heating platform for 10 minutes. Alternatively, the slides for held at 80˚C (+/- 2˚C) for 60 minutes in the same solution. The slides were then allowed to cool on the bench top (still immersed in the citrate buffer) for 20-30 minutes, or until a temperature less than 50˚C had been reached. The slides were then immersed in Dako™ wash buffer (S3006) for 10-15 minutes on a gentle rotating platform; changing the wash buffer once during this time. After rinsing with dH_2O, the sections were incubated with 5% normal serum/0.3% Triton™ X-100/1xPBS, or Dako mouse/rabbit serum, for 1 hour at room temperature to block nonspecific binding and permeablize membranes, using a Dako pen (S2002) to draw around the tissue sections. The slides were then incubated overnight with primary antibodies, diluted in Dako™ antibody diluent with background reducing components (S3022), or alternatively 1xPBS/1% BSA/0.3% Triton™, at 4˚C in a humidified chamber. Following washing for 10-15 minutes in 1xPBS, the tissue sections were incubated with Alexa Fluor secondary antibodies, diluted in 1xPBS/1% BSA/0.3% Triton™ X-100, for 2 hours in a humidified chamber in the dark. A list of the primary and secondary antibodies used is provided in Table 2.4 and Table 2.5. The slides were washed again for 10-15 minutes in 1xPBS. Prolong® Gold Antifade Reagent with DAPI (Cell Signalling; #8961) was used to mount the coverslips on the slides, which were then sealed with nail varnish. The slides were left overnight in the dark (e.g., in fridge) prior to visualisation on Zeiss confocal microscope or slide scanner. Alternatively, DAPI was added to the slides and incubated in the dark for 10 minutes prior to coverslipping with Prolong® Gold Antifade Reagent (#9071). Immunofluorescence was detected and imaged using a Zeiss LSM 710 confocal microscope and ZEN lite software (Zeiss).

Table 2.4. Primary Antibodies used for IHC and IF Procedures

Antibody	Clone	Host	Dilution	Source	Cat. No.
Anti-C20orf201	Polyclonal	Rabbit	1:250-1:500	Abcam	ab108142*
Anti-C20orf201	Polyclonal	Rabbit	1:15-1:20	Abcam	ab170783
Anti-MAGEA1	Monoclonal	Mouse	1:20	LSBio	LS-C87868
Anti-PIWIL1	Monoclonal	Mouse	1:1000-1:2000	Sigma	SAB4200365
Anti-PIWIL2	Monoclonal	Mouse	1:20-1:50	Abnova	MAB0843
Anti-SPO11	Polyclonal	Rabbit	1:1000-1:2000	Abcam	ab81695
Anti-TEX19	Polyclonal	Rabbit	1:400-1:800	Abcam	ab185547

*Lot: GR191628-1 and GR125175-1

Table 2.5. Secondary Antibodies used for IF

Antibody	Host	Species Reactivity	Label	Dilution	Source	Cat. No.
Anti-Mouse IgG	Goat	Mouse	Alexa Fluor® 488	1:400	Life Technologies	A11029
Anti-Mouse IgG	Goat	Mouse	Alexa Fluor® 568	1:400	Life Technologies	A11031
Anti-Mouse IgG	Donkey	Mouse	Alexa Fluor® 647	1:400	Life Technologies	A31571
Anti-Rabbit IgG	Goat	Rabbit	Alexa Fluor® 488	1:400	Life Technologies	A11034
Anti-Rabbit IgG	Goat	Rabbit	Alexa Fluor® 568	1:400	Life Technologies	A11011

2.12. T-cell binding assay

2.12.1. Peptide design

T2 cells were donated by Dr Stephen Man (Cardiff University). They are a TxB cell hybrid which express unstable HLA-A2 molecules and have been extensively used in biological assays to measure peptide binding to human leucocyte antigen (HLA) (Trojan *et al.*, 2001). Peptides were designed using different online tools including: Bioinformatics and Molecular Analysis Section (BIMAS) of the Center for Information Technology, National Institute of Health (http://www-bimas.cit.nih.gov/molbio/hla_bind/); SYFPEITHI Database (www.syfpeithi.de) (Rammensee *et al.*, 1999); Immune Epitope Database (IEDB) Analysis Resource (http://tools.immuneepitope.org/mhci/result/); Zhiping's Lab (http://www.umassmed.edu/zlab/ and http://zlab.bu.edu) (Peters *et al*., 2003). The top ranking peptides and/or peptides that featured on multiple prediction results were chosen to take forward for the commonest HLA alleles for Caucasians.

2.12.2. Peptide preparation and storage

The peptides were ordered from Mimotopes (The Peptide Company) and 1mg of peptide was provided at a minimum of 94% purity. Each peptide was dissolved in 20 µl of dimethyl sulfoxide (DMSO). Alternatively 40 µl of DMSO was used if the peptide appeared not to dissolve well but this was only required for one peptide. The peptide stock at a concentration of 50 mg/ml was stored at -80˚C. 2 µl of this stock peptide solution was added to 8 µl of RPMI 1640 (+ Glutamax/2mM L Glutamine + 100 units per ml penicillin and 100 µg per ml of streptomycin) for a 1 in 5 dilution. The positive control peptide was M1 (supplied by Stephen Man, Cardiff University) and was not diluted 1 in 5 as above for the custom peptides because the stock was at a lower concentration (10 mg/ml).

2.12.3. Peptide incubation

The T2 cells were cultured in complete RPMI 1640 media (+ 2mM L Glutamine + Penicillin/Streptomycin + 10% foetal calf serum) at 37˚C with 5% CO_2. The cultured cells were counted and $1x10^5$ cells were added to the wells of a 96-well culture plate in a final volume of 200 µl media. 1 µl of the resuspended and diluted peptides were added separately to the

individual wells containing the T2 cells, which were then covered and incubated overnight at 37˚C with 5% CO_2.

2.12.4. Addition of anti-human HLA-A2 antibody and flow cytometric read-out

The T2 cells were harvested from the 96-well plate by pipetting up and down several times and transferred to separate 5 ml Falcon tubes. The cells were washed by adding a similar volume (200-400 µl) of Flow Cytometry Staining (FACS) Buffer (1xPBS + FCS). The tubes were then centrifuged at 1200 r.p.m. for 4 minutes in a Heraeus Megafuge 1.0R centrifuge. The supernatant was removed from the cell pellet by inverting the tube. 5 µl of FITC anti-human HLA-A2 antibody (Biolegend, Clone: BB7.2, 400 µg/ml, #343304) was added to the resuspended cells (approximately 100 µl of supernatant remained on the cells after inverting the Falcon tube). The cells were incubated with the antibody for 30-40 minutes at 4˚C in the dark. Following this incubation, the cells were washed again with FACS buffer and the tubes then centrifuged at 1200 r.p.m. for 4 minutes. The supernatant was removed in the same fashion by inverting the Falcon tubes, leaving approximately 100 µl of supernatant and cell suspension within the tube. The Falcon tubes were tapped gently to dislodge the cell pellet before measuring on a FACS machine. The viable T2 cells were identified on the basis of forward and side scatter, then gated using the negative control sample (T2 cells only). HLA-A2 expression was determined for 20,000 cells.

3. Analysis of testis-specific genes as potential novel CT-genes

3.1. Introduction

In contrast to mitotic cell division that results in two identical daughter cells, meiosis results in four genetically distinct haploid cells termed gametes. Mitosis occurs throughout our bodies to generate, renew and remodel tissues. This allows organised growth during a child's development, as well as enables tissue homeostasis and repair. Meiosis and gametogenesis, on the other hand, are restricted to the germline. The production of gametes in the cells of the germline is essential for successful reproduction, whereby ploidy is maintained in the offspring as the haploid cells combine during fertislisation to re-establish the diploid state in the zygote. Both mitosis and meiosis are tightly regulated processes orchestrated by a variety of proteins (Civelekoglu-Scholey and Cimini, 2014; Duro and Marston, 2015). Some proteins are unique to meiosis.

Dysregulation of mitosis can result in uncontrolled proliferation of cells that is a fundamental hallmark of cancer (Hanahan and Weinberg, 2000; Hanahan and Weinberg, 2011). Equally errors that occur during cell division could lead to chromosome segregation defects and aneuploidy or polyploidy. The chromosome dynamics that occur during meiotic cell division are particularly complex, such as meiotic recombination, bouquet formation and sister centromere monopolarity (some of these details have been discussed in the Introduction). If the genes regulating these processes become activated in somatic cells they could contribute greatly to genetic instability and in this way contribute to oncogenesis (McFarlane *et al*., 2015). Indeed, genetic instability is now considered a hallmark, or enabling characteristic, of cancer (Hanahan and Weinberg, 2011).

A prognostic correlation in cancers has been shown for the expression of many CT genes (Choi and Chang, 2012; Luo *et al*., 2013; Yakirevich *et al*., 2003). However, it remains unclear whether the (worse) clinical outcome observed for patients who have tumours which express CT genes are caused by their expression. It is conceivable that the activation of CT genes is a consequence of increased genetic instability seen in certain tumours. Despite the strong mechanistic reasons that would result, in theory, from the expression of testis- or meiosis-specific genes, experimental evidence for a direct mechanistic link to cancer has been lacking. However, recent evidence from independent sources has added greatly to evidence in support

of a direct link between activation of such genes and cancer formation (Cho *et al*., 2014; Janic *et al*., 2010; Watkins *et al*., 2015).

Proteins known to be essential for homologous chromosome conjoining in meiosis have been shown to be required for the alternative lengthening of telomeres (ALT) (Cho *et al*., 2014). Telomerase is an enzyme that enables elongation and maintenance of the telomere length. This is the most common method for promoting immortalisation in cancer cells by preventing telomere shortening (Bertorelle *et al*., 2014; O'Sullivan and Karlseder, 2010). There is an alternative method for the lengthening of telomeres in the absence of telomerase activity, termed ALT. It has recently been shown that a heterodimer of two meiotic proteins interacts with RAD51 to promote directional movement of ALT telomeres and recombination of non-sister telomeres to maintain telomere length and thus promote immortality (Cho *et al*., 2014).

HORMAD1 is a meiosis-specific protein that was identified as a potential CTA over a decade ago (Chen *et al*., 2005). HORMAD1 is involved in partner bias during homologous recombination (HR) events in meiosis. More recently is has been shown that HORMAD1 suppresses RAD51-dependent HR in triple-negative breast cancers, further supporting a role for meiotic proteins in oncogenesis (Watkins *et al*., 2015).

Janic and colleagues identified that germline factors were oncogenic drivers in a *Drosophila melanogaster* brain tumour model (Janic *et al*., 2010). It was subsequently shown that the orthologues of these genes were widely activated in human cancers and proposed that a 'soma-to-germline' transition may be a key feature in many cancer types (McFarlane *et al*., 2014). These latest studies add to the evidence that tumour cells induce a gametogenic programme to promote the development of cancer hallmarks, as has been postulated (McFarlane *et al*., 2015; Simpson *et al*., 2005). Meiosis-specific genes may be important drivers of this transition and understanding their function may help us understand important oncogenic mechanisms. Moreover, the same genes and protein products could be clinically useful biomarkers and/or targets for immunotherapeutics in oncology. For these reasons, we set out to identify potential meiosis-specific genes as a source of novel CT genes. Manual curation of the literature to identify candidate genes is both time-consuming and likely to miss meiosis-associated genes that have not yet been characterised in humans. Thus, we set out to establish a more systematic approach as outlined below using a bioinformatics pipeline (developed in the McFarlane group).

3.2. Selection of candidate genes from a bioinformatics pipeline

A bioinformatics pipeline was previously established to identify putative human meiosis-specific genes that could potentially encode CTAs (Feichtinger *et al*., 2012b). This was based initially on a cohort of mouse genes predicted to be associated specifically with spermatocyte development (Chalmel *et al*., 2007). This study by Chalmel and colleagues identified 744 genes in mice, many of which would be expected to have roles during meiosis and potentially be meiosis and/or spermatocyte-specific. The human orthologues of these genes were then identified with a high degree of stringency, revealing 408 candidate genes. Figure 3.1 summarises the pipeline specific to this research, which has been adapted from the pipeline described in detail elsewhere (Feichtinger *et al.*, 2012b).

The 408 human orthologue genes that had been identified were checked for expression in mitosis using MitoCheck (www.mitocheck.org). MitoCheck was a project to identify genes and proteins involved in the cell division and chromosome segregation in humans – for more information see website. Genes identified in this project will be expected to be expressed widely in somatic cells. Thus, if any of our candidate genes were included in the MitoCheck list they were excluded at this stage as we wanted to identify genes restricted to meiosis or the germline. This filtering step excluded 33 genes, leaving a list of 375 candidates (see Figure 3.1). These genes were then challenged against a bioinformatics tool, based on expressed sequence tag (EST) identifiers from the Unigene database (Feichtinger *et al*., 2014a). ESTs are short nucleotide sequences (up to 500 bp) that are clustered and counted following a single-pass nucleotide sequence read (Nagaraj *et al*., 2007). EST databases have been utilised to identify potential new CTAs (Bettoni *et al*., 2009; Chen *et al*., 2005a; Chen *et al*., 2005b; Hofmann *et al*., 2008).

If a candidate gene was represented in a non-testis/non-central nervous system (CNS) normal tissue EST library, then it was excluded. The remaining genes were assessed further to see if they were represented in cancer EST libraries. From the original 375 potential meiosis-specific genes the EST analysis identified 177 candidate genes that fulfilled these criteria. These 177 genes were sub-categorised into different classes based on the EST expression profiles. Class 1 genes, which contained 9 candidates, were found in cancer EST libraries and normal testis. Class 2 genes, which contained 75 candidates, had EST signatures seen in normal testis only. Class 3 genes had EST signatures seen in normal testis and central nervous system (CNS) tissues as well as being present in cancer EST libraries – 21 candidate genes fell into this

category. Class 4 gene candidates, for which there were 72, were expressed in normal testis and CNS tissues but not found in the cancer EST libraries. The 105 genes within class 1, 2 and 3 formed the basis for the initial study from our group (Feichtinger *et al*., 2012b). The aim of the work reported in this chapter was to validate the expression profile of a cohort of the remaining class 4 genes as potential CT genes (Sammut *et al*., 2014).

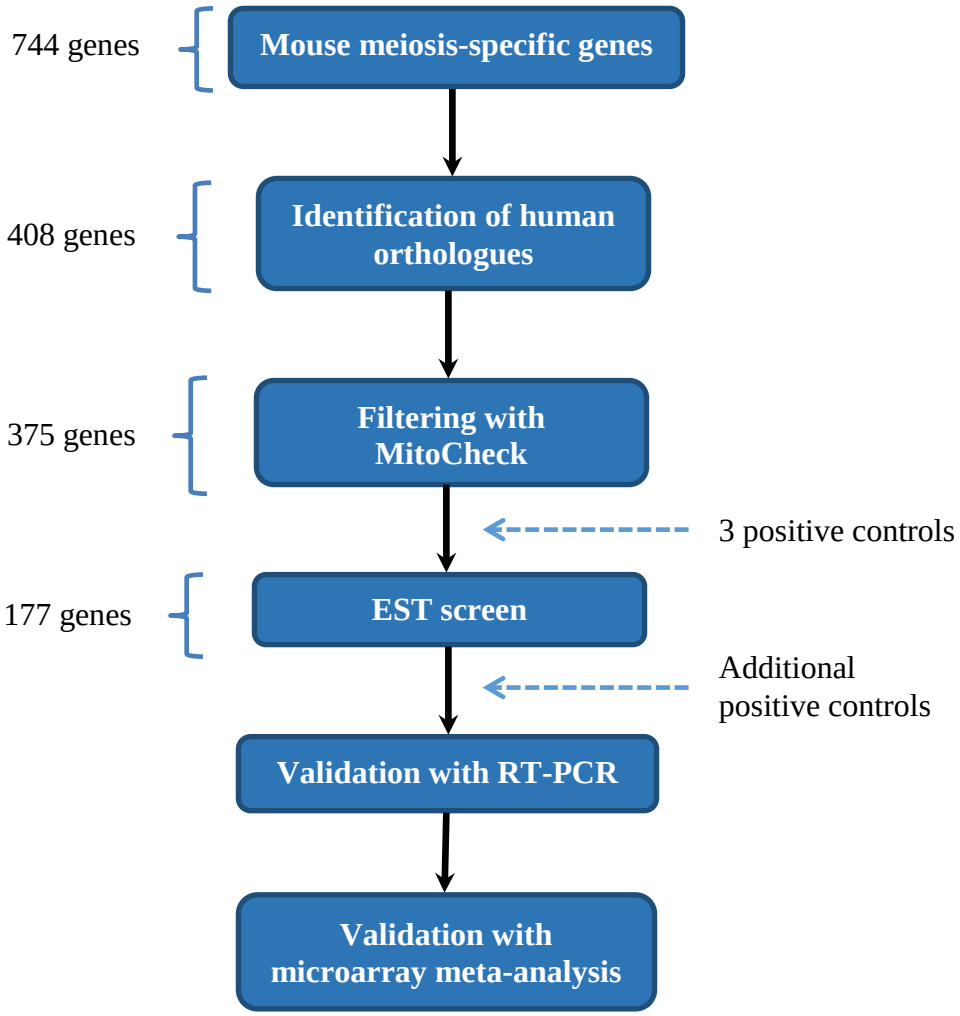


Figure 3.1 Schematic flow diagram of the bioinformatics pipeline used to identify potential novel CT genes.

The initial pipeline was designed by McFarlane and Feichtinger (Fechtinger *et al.* , 2012b). 744 genes were identified as being meiosis-specific in mice; 408 human orthologues were then identified. 33 genes were excluded by MitoCheck, leaving 375 candidate human meiosis-specific genes. All these genes were fed into an EST analysis pipeline, along with 3 positive control genes known already to encode CTAs. The positive controls were included to assess whether known CTAs were not excluded by the pipeline and displayed appropriate expression profiles using the samples that were used for PCR validation. This returned 177 candidate genes; 72 which fell into a testis/CNS restricted category. A cohort of these genes were then subject to RT-PCR expression validation, which formed the basis of the work presented in this chapter (Sammut *et al.*, 2014). Adapted from (Feichtinger *et al.*, 2012b).

3.3. Identification and validation of potential novel CT genes predicted through an EST pipeline

The candidate meiosis-specific genes were subjected to an EST filtering pipeline as outlined above. This process categorised the genes into 4 groups. Our group had previously validated the expression profiles of the candidate genes that fell into categories 1-3 of the EST pipeline (Feichtinger *et al*., 2012b). Within the initial class 1-3 predicted gene sets RT-PCR validation revealed that a number were actually expressed in extensive somatic tissues (Feichtinger *et al*., 2012b). Given this, we re-analysed our predicted class 4 genes using updated CancerEST and CancerMA pipelines (Feichtinger et al., 2012a; Feichtinger et al., 2014a) and from this we identified 54 putative class 4 genes, those with expression signatures only in the testis and CNS of healthy tissue (see Table 3.1). This cohort of genes totalled approximately two thirds of all the class 4 genes identified.

The tissue specificity of these candidate genes was assessed by analysing the expression profiles of these genes in 20 different normal human tissue types. The cDNA was generated from total RNA preparation from normal tissue samples, including testicular RNA, obtained post-mortem from Clontech and Ambion (see Appendix C, on CD, for further details). The cDNA quality was assessed using PCR amplification of *ACTB*. *ACTB* was chosen as a simple and practical way of assessing that the cDNA synthesis had worked and there was not massively varying quantities across our samples. We were not undertaking quantitative PCR in these experiments, so a range of normalisation or 'housekeeping' control genes were not used. *NY-ESO-1* was used as a control CT gene: this is a well characterised CT gene that encodes one of the most immunogenic CTAs. Where possible, intron-spanning primer sets were designed to avoid amplification of possible contaminating genomic DNA. 40 cycles of PCR was used with the aim of detecting low levels of gene expression within the tested samples (except for *ACTB* where 35 cycles were used).

The PCR validation acted as a crude but important step to try and identify genes with expression profiles consistent with germline predominant expression and equally to exclude genes that were expressed widely in somatic tissues. Not all the PCR data is presented here but a sample of the genes from each of the following groups, as well as some of the genes that were excluded from further analysis. It does not detract from the overarching message and conclusions drawn, which have inherent limitations as discussed again later. A summary of the positive PCR data

is provided in Figure 3.6. Based on the RT-PCR expression profile from the normal somatic tissues, the candidate genes were classified into different groups:

- If the RT-PCR indicated that there was expression in more than two normal tissues (excluding testis and CNS tissues) then the genes were dismissed (see Figure 3.2). These genes were not subjected to further analysis within the cancer panel of tissues/cell lines.

- The gene was classified as testis-restricted if expression appeared limited to the testis within normal tissues (see Figure 3.3).

- If the RT-PCR indicated expression within the testis and normal CNS tissues only then the gene was classified as testis/CNS-restricted (see Figure 3.4).

- Testis-selective genes were those which were seen in the testis, as well as one or two normal tissues. Testis/CNS-selective genes were considered an additional category of genes (see Feichtinger *et al.*, 2012b). These were genes which the PCR indicated were expressed in testis and CNS tissues, as well as one or two other normal somatic tissues; none of the genes tested here exhibited this type of expression profile (see Figure 3.6).

Table 3.1. List of the 54 potential meiCT genes.

Gene name	Ensembl ID	Unigene cluster ID	Chromosomal location with GRC coordinates
ACTL9	**ENSG00000181786**	**Hs.209206**	**19: 8,807,747-8,809,172 in GRCh37 coordinates**
ADAM2	**ENSG00000104755**	**Hs.177959**	**8: 39,601,254-39,695,808 in GRCh38 coordinates**
ASB17	**ENSG00000154007**	**Hs.125423**	**1: 75,918,873-75,932,431 in GRCh38 coordinates**
BOLL	**ENSG00000152430**	**Hs 692026**	**2: 197,726,879-197,786,762 in GRCh38 coordinates**
C1orf141*	**ENSG00000203963**	**Hs.666621**	**1: 53,916,574-53,945,929 in GRCh38 coordinates**
C2orf61	ENSG00000239605	Hs.531575	2: 47,045,538-47,155,378 in GRCh38 coordinates
C3orf22	**ENSG00000180697**	**Hs.178210**	**3: 126,245,842-126,277,808 in GRCh37 coordinates**
C3orf30	ENSG00000163424	Hs.271580	3: 119,146,150-119,160,042 in GRCh38 coordinates

C6orf81	ENSG00000157343	Hs.533066	6: 35,704,809-35,716,856 in GRCh37 coordinates
C8orf74*	**ENSG00000171060**	**Hs.371776**	**8: 10,530,147-10,558,103 in GRCh37 coordinates**
C9orf153	ENSG00000187753	Hs.632073	9: 88,835,180-88,874,572 in GRCh37 coordinates
C10orf67	ENSG00000179133	Hs.522360	10: 23,267,195-23,344,845 in GRCh38 coordinates
C12orf50 *	**ENSG00000165805**	**Hs.112930**	**12: 88,373,812-88,427,814 in GRCh37 coordinates**
C16orf78	**ENSG00000166152**	**Hs.125875**	**16: 49,373,823-49,399,431 in GRCh38 coordinates**
CAPZA3*	**ENSG00000177938**	**Hs.131288**	**12: 18,891,035-18,892,121 in GRCh37 coordinates**
CCDC116	**ENSG00000161180**	**Hs.131615**	**22: 21,632,716-21,637,327 in GRCh38 coordinates**
CCDC63	ENSG00000173093	Hs.437141	12: 111,284,573-111,345,339 in GRCh37 coordinates
CCDC73	ENSG00000186714	Hs.706808	11: 32,602,246-32,794,658 in GRCh38 coordinates
CCIN	ENSG00000185972	Hs.115460	9: 36,169,392-36,171,332 in GRCh38 coordinates
DDI1	ENSG00000170967	Hs.591941	11: 104,036,580-104,039,194 in GRCh38 coordinates
DNAJC5G	**ENSG00000163793**	**Hs.116303**	**2: 27,498,289-27,504,367 in GRCh37 coordinates**
DYDC1	ENSG00000170788	Hs.407751	10: 82,095,861-82,116,511 in GRCh37 coordinates
FAM170A*	**ENSG00000164334**	**Hs.713304**	**5: 119,629,559-119,635,822 in GRCh38 coordinates**
FAM194A	ENSG00000163645	Hs.147128	3: 150,659,885-150,703,971 in GRCh38 coordinates
FAM71B*	**ENSG00000170613**	**Hs.666099**	**5: 157,161,846-157,166,264 in GRCh38 coordinates**
FIGLA	ENSG00000183733	Hs.407636	2: 70,777,310-70,790,643 in GRCh38 coordinates
GK2	**ENSG00000196475**	**Hs.98008**	**4: 80,327,506-80,329,447 in GRCh37 coordinates**
H2AFB1	ENSG00000198082	Hs.592246	X: 154,113,247-154,113,833 in GRCh37 coordinates
HEATR7B*	**ENSG00000**	**Hs.97714**	**5: 40,998,017-41,071,342 in GRCh38 coordinates**
HMGB4	**ENSG00000176256**	**Hs.568628**	**1: 33,860,475-33,864,791 in GRCh38 coordinates**
IQCF1	**ENSG00000173389**	**Hs.671210**	**3: 51,928,892-51,937,351 in GRCh37 coordinates**
KLF17*	**ENSG00000171872**	**Hs.567674**	**1: 44,118,850-44,135,140 in GRCh38 coordinates**
LYZL6	**ENSG00000161572**	**Hs.97477**	**17: 35,934,518-35,943,699 in GRCh38 coordinates**
PDHA2	**ENSG00000163114**	**Hs.131361**	**4: 95,840,088-95,841,474 in GRCh38 coordinates**
PDILT	ENSG00000169340	Hs.376025	16: 20,359,170-20,404,737 in GRCh38 coordinates

PPP3R2	**ENSG00000188386**	**Hs.151167**	**9: 101,591,615-101,595,001 in GRCh38 coordinates**
PRPS1L1	**ENSG00000229937**	**Hs.169284:**	**7: 18,026,782-18,027,863 in GRCh38 coordinates**
RBM44	**ENSG00000177483**	**Hs.720233**	**2: 238,707,032-238,751,451 in GRCh37 coordinates**
RNF133	ENSG00000188050	Hs.126730	7: 122,337,766-122,339,210 in GRCh37 coordinates
SATL1*	**ENSG00000184788**	**Hs.640783**	**X: 85,092,287-85,109,048 in GRCh38 coordinates**
SHCBP1L*	**ENSG00000157060**	**Hs.497034**	**1: 182,899,865-182,953,525 in GRCh38 coordinates**
SLC25A31*	**ENSG00000151475**	**Hs.149030**	**4: 127,730,378-127,774,293 in GRCh38 coordinates**
SPDYA	ENSG00000163806	Hs.511956	2: 29,005,383-29,073,477 in GRCh37 coordinates
SPZ1	**ENSG00000164299**	**Hs.519403**	**5: 80,319,625-80,321,842 in GRCh38 coordinates**
TBC1D21*	**ENSG00000167139**	**Hs.124512**	**15: 74,165,949-74,181,555 in GRCh37 coordinates**
TGIF2LX	**ENSG00000153779**	**Hs.592220**	**X: 89,176,881-89,177,882 in GRCh37 coordinates**
TMEM146*	**ENSG00000174898**	**Hs.631842**	**19: 5,720,688-5,778,745 in GRCh37 coordinates**
TRIM42*	**ENSG00000155890**	**Hs.343487**	**3: 140,396,881-140,419,992 in GRCh37 coordinates**
TSGA13	ENSG00000213265	Hs.592266	7: 130,668,648-130,687,432 in GRCh38 coordinates
TSSK2	ENSG00000206203	Hs.694070	22: 19,130,808-19,132,621 in GRCh38 coordinates
UBL4B	ENSG00000186150	Hs.374027	1: 110,112,440-110,113,947 in GRCh38 coordinates
UBQLN3	ENSG00000175520	Hs.189184	11: 5,507,300-5,509,985 in GRCh38 coordinates
UMODL1	ENSG00000177398	Hs.242520	21: 42,062,959-42,143,453 in GRCh38 coordinates
ZSWIM2	**ENSG00000163012**	**Hs.375054**	**2: 187,692,562-187,713,935 in GRCh37 coordinates**

Note: emboldened genes are those for which RT-PCR on a range of normal human tissues indicated that gene expression was present in two or fewer somatic tissues other than testis and CNS tissues.
*Signifies genes that had no expression in cancer cells/tissues as assessed by RT-PCR validation using 34 cancer cell/tissue RNA sets.

Of the 54 genes, 21 were expressed in more than two non-testis/non-CNS normal tissues and were therefore dismissed at this stage. Of the remaining 33 genes (bold in Table 3.1), 30 had expression limited to the testis in normal tissue, 2 had expression limited to the testis and normal CNS tissues and 1 further gene had expression in two normal tissues in addition to testis. These 33 genes, were then analysed by RT-PCR in a range of cancer tissues/cell lines – see Figures 3.5 and 3.6. Following this analysis, 14 of these genes were shown to have no expression in any of the cancerous material/cells. For a further 16 genes expression was indicated in at least one cancerous tissue/cell line and no expression in normal tissues other than the testis (Figure 3.6, class B). Of the remaining 3 genes, 2 were cancer/testis/CNS restricted [i.e. expressed in at least one cancer cell type, in addition to the testis and normal CNS tissues (Figure 3.6, class C)] and 1 was cancer/testis-selective [i.e. expressed in one or two normal tissues other than CNS, as well as the testis and at least one cancer type (Figure 3.6, class D)].

The SW480 cell line, which is a colorectal cancer cell line, stands out as expressing several of the candidate genes though weakly. Although intron-spanning primer sets were used where possible it may be that genomic contamination accounts for some of these positive results. Further, it was identified by a member of our group that there was genomic contamination within this cell line cDNA (Jones-Hutchins, unpublished data). Intron-spanning primer sets could not always be designed – the sets that were not intron spanning are highlighted in Table 1 in the Appendix. On this basis the positive results for *ACTL9*, *PDHA2*, *PRPS1L1*, *SPZ1* and *TGIF2LX* in the SW480 cell line should be viewed with caution. However, we do not feel genomic contamination was a wide-spread problem in our samples and largely mitigated through the use of intron-spanning primer sets as described. There was no expression in normal tissues for these primer sets and limited expression elsewhere suggesting a lack of genomic DNA contamination in these samples. In addition, genomic DNA controls proved negative when these were tested on a limited group of samples where cDNA had been synthesised following human tissue collection for data presented elsewhere in this thesis. These possible false-positive results in the SW480 cell line do not detract from the overall message and the relevant genes would still have been taken forward had these findings been shown to be negative. The primary aim of the PCR validation was to exclude the genes that were widely expressed in somatic tissues.

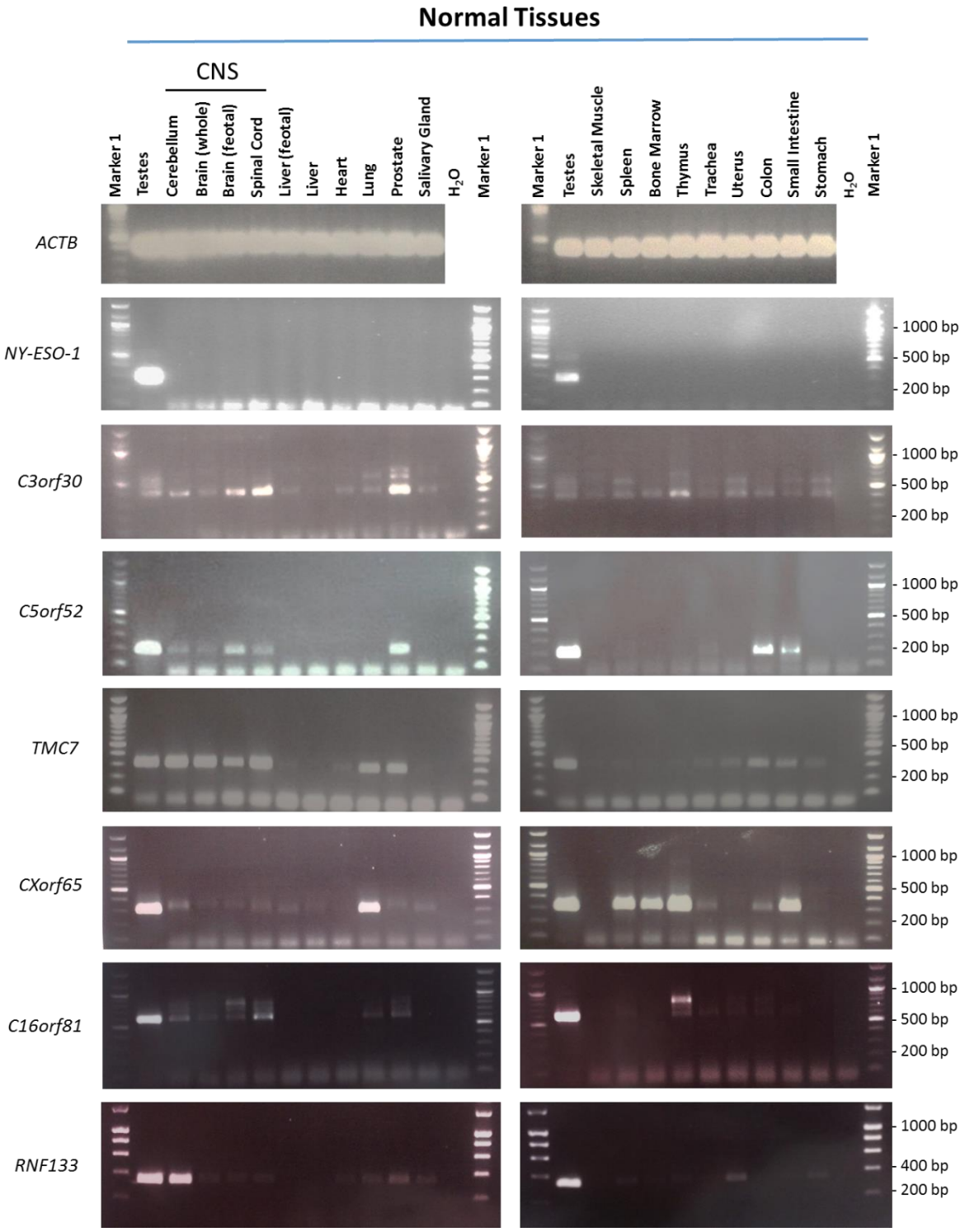


Figure 3.2. RT-PCR analysis of candidate genes that were expressed in somatic tissues.

Agarose gels showing the expression profile of the candidate genes in a range of normal tissue samples. cDNA was generated from the RNA extracted from these tissues *post mortem*. A number of the candidate genes, examples of which are given here, were found to be expressed in a range of somatic tissues. If expression was indicated in more than two normal tissues (other than CNS tissues) then the genes were excluded from further analysis. All the examples here fell into this category. *ACTB* gene was used as a positive control for the cDNA samples and *NY-ESO-1* as a comparison to a known CTA. The DNA marker used in the agarose gel for *RNF133* was hyperladder 1 (Bioline), rather than hyperladder 2 (Bioline) which was used in the other gels.

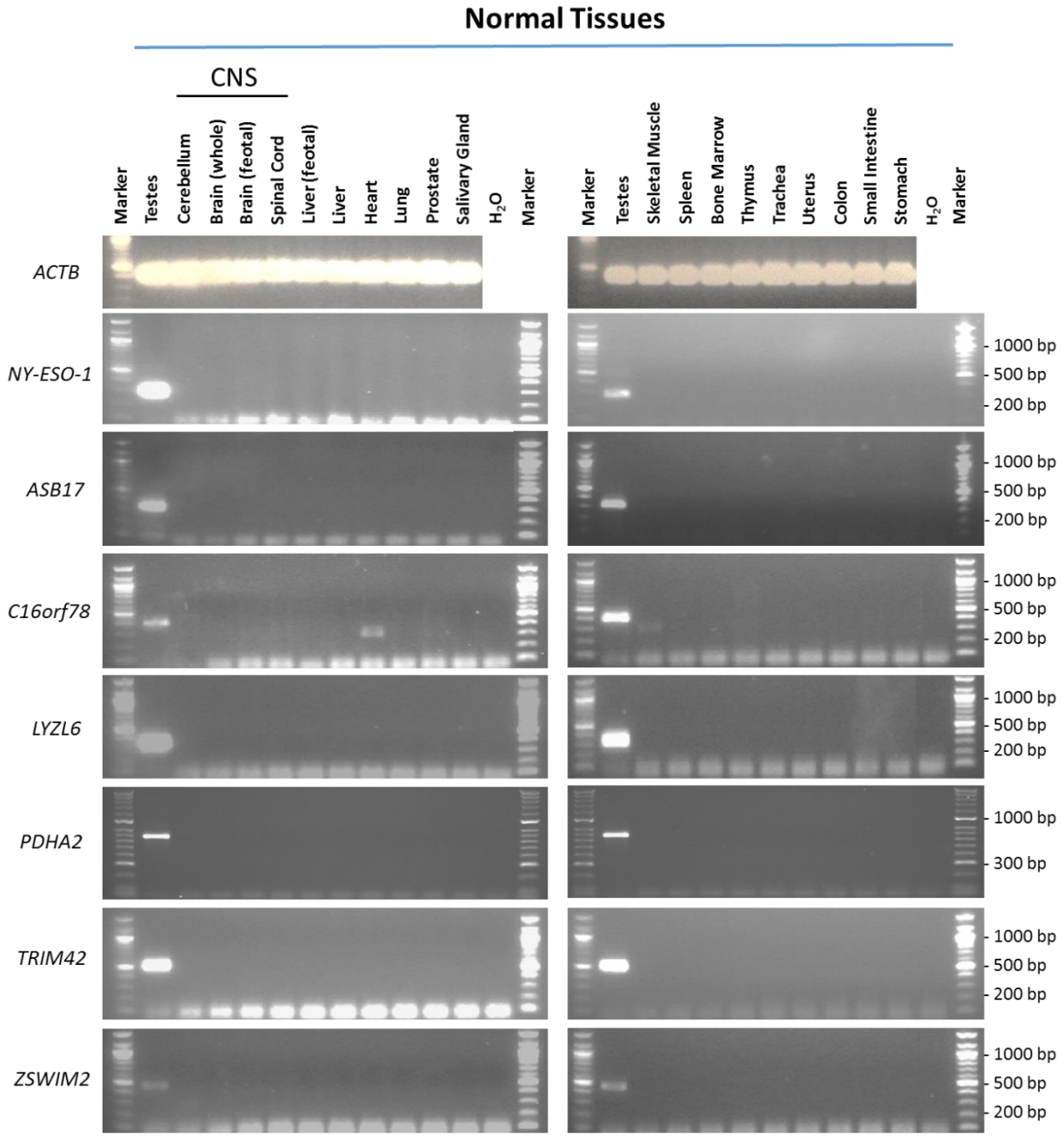


Figure 3.3. RT-PCR validation of putative meiosis-associated genes that were restricted to testis in normal tissues.

Agarose gels showing the expression profile of the candidate genes in a range of normal tissue samples. cDNA was generated from the RNA extracted from these tissues *post mortem*. *NY-ESO-1* is a known CT-gene which has a testis-restricted expression in normal tissues. A number of the candidate genes were found to be expressed only in the testis and categorised as testis-restricted. *ACTB* gene was used as a positive control for the cDNA samples. *C16orf78* appeared to express a band, though slightly shorter than the expected product size, in normal heart tissue but this was sequenced and found to show no significant similarity to this gene. The DNA marker used in the agarose gel for *PDHA2* is a more recent version of the Hyperladder 2, displaying intense marking bands at 300bp and 1000bp, rather than 500bp and 1000bp.

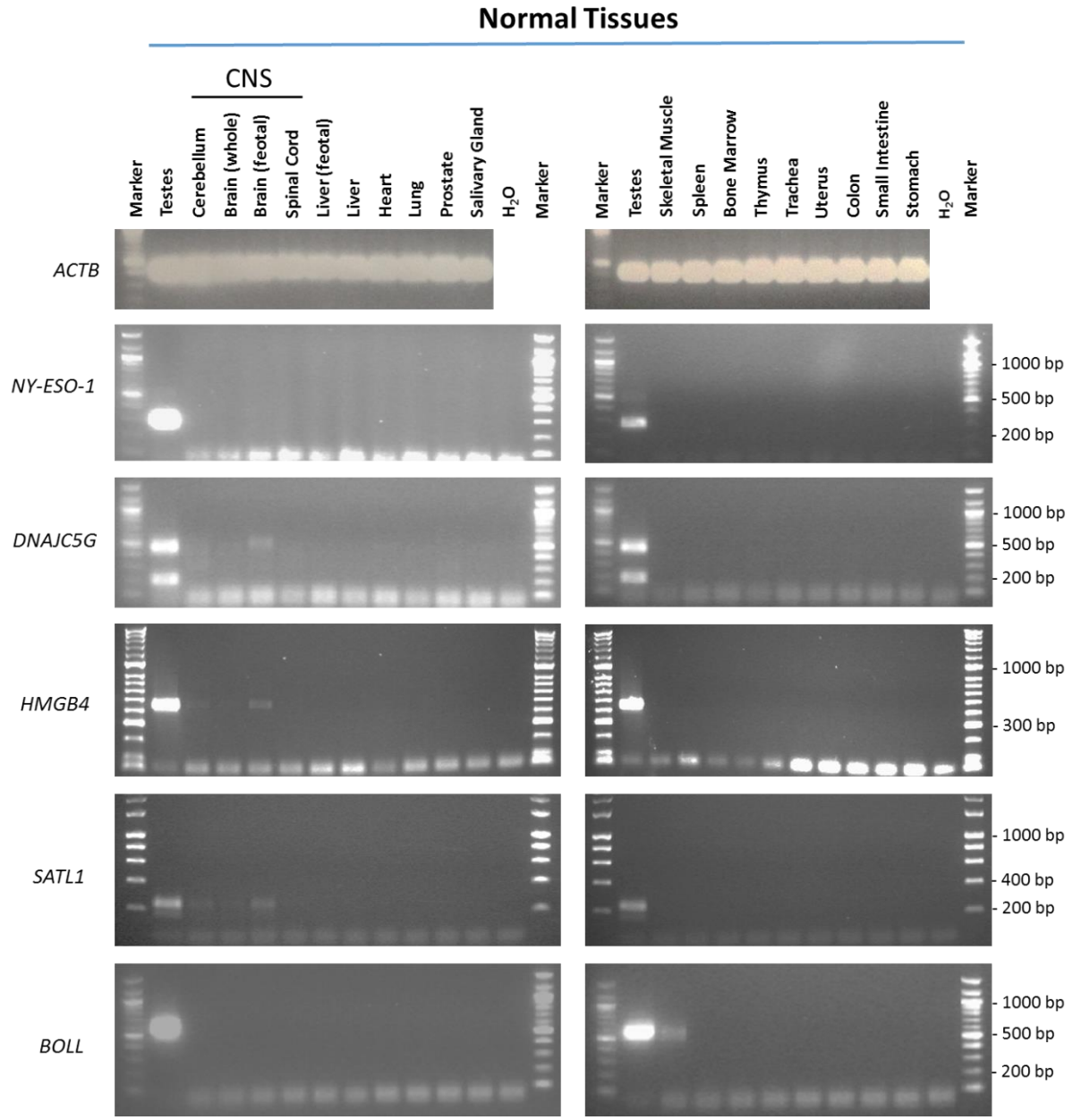


Figure 3.4. RT-PCR analysis of putative CT genes that indicated limited expression in somatic tissues other than testis.

cDNA was generated from the RNA extracted from a range of normal tissues *post mortem*. Agarose gels were then run to visualise the expression profile of the candidate genes. A number of the candidate genes (e.g., *HMGB4* and *SATL1*) indicated that there was testis/CNS-restricted expression; that is expression in normal testis and CNS tissues but none of the other normal tissues tested. Other candidate genes were classed as testis-selective if it was indicated that there expression profile was limited to two or less somatic tissues other than CNS – for example, *BOLL* above. A*CTB* gene was used as a positive control for the cDNA samples and *NY-ESO-1* as a comparison to a known CTA. The DNA marker used in the agarose gel for *SATL1* was hyperladder 1, rather than hyperladder 2 which was used in the other gels and on the gel assessing *HMGB4* above a more recent version of hyperladder 2 was used which has a strong band at 300bp rather than 500bp.

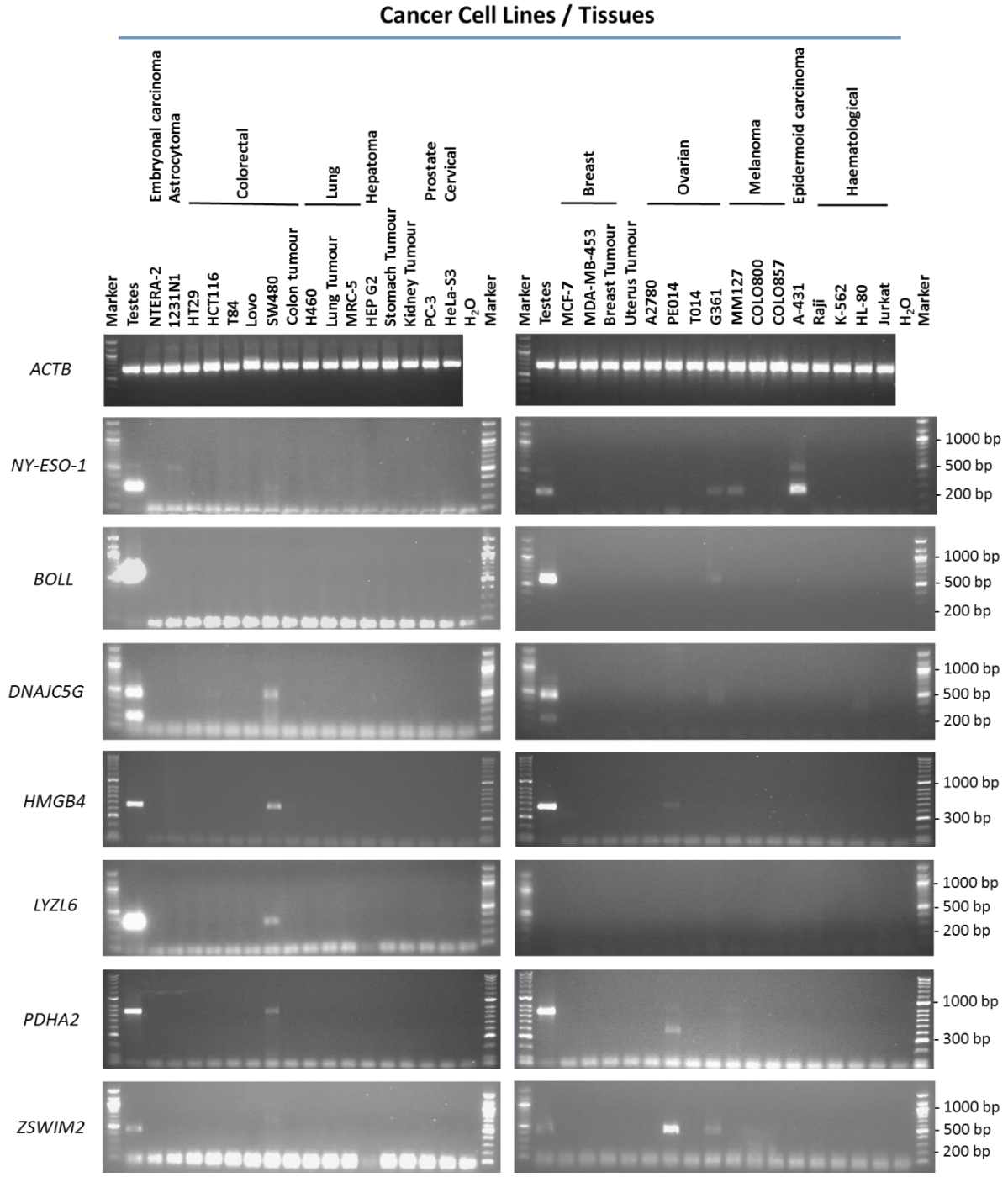


Figure 3.5. RT-PCR analysis of candidate genes in cancer cell lines and tissues.

Genes which passed the screening of gene expression and displayed limited activation in a panel of normal tissues, were subjected RT-PCR analysis in a range of cancer tissues and cell lines. cDNA was generated from the RNA extracted from a range of cancer tissues obtained *post mortem* or extracted from various cancer cell lines. Agarose gels were then run to visualise the expression profile of the candidate genes. A number of the candidate genes displayed here showed no expression in any of the cancer samples. Others, examples of which are given here, indicated limited expression in some cancer types. A*CTB* gene was used as a positive control for the cDNA samples and *NY-ESO-1* as a comparison to a known CTA. The DNA marker used in the agarose gel for *PDHA2* and *HMGB4* was a more recent version of hyperladder 2 that has a strong band at 300bp rather than 500bp.

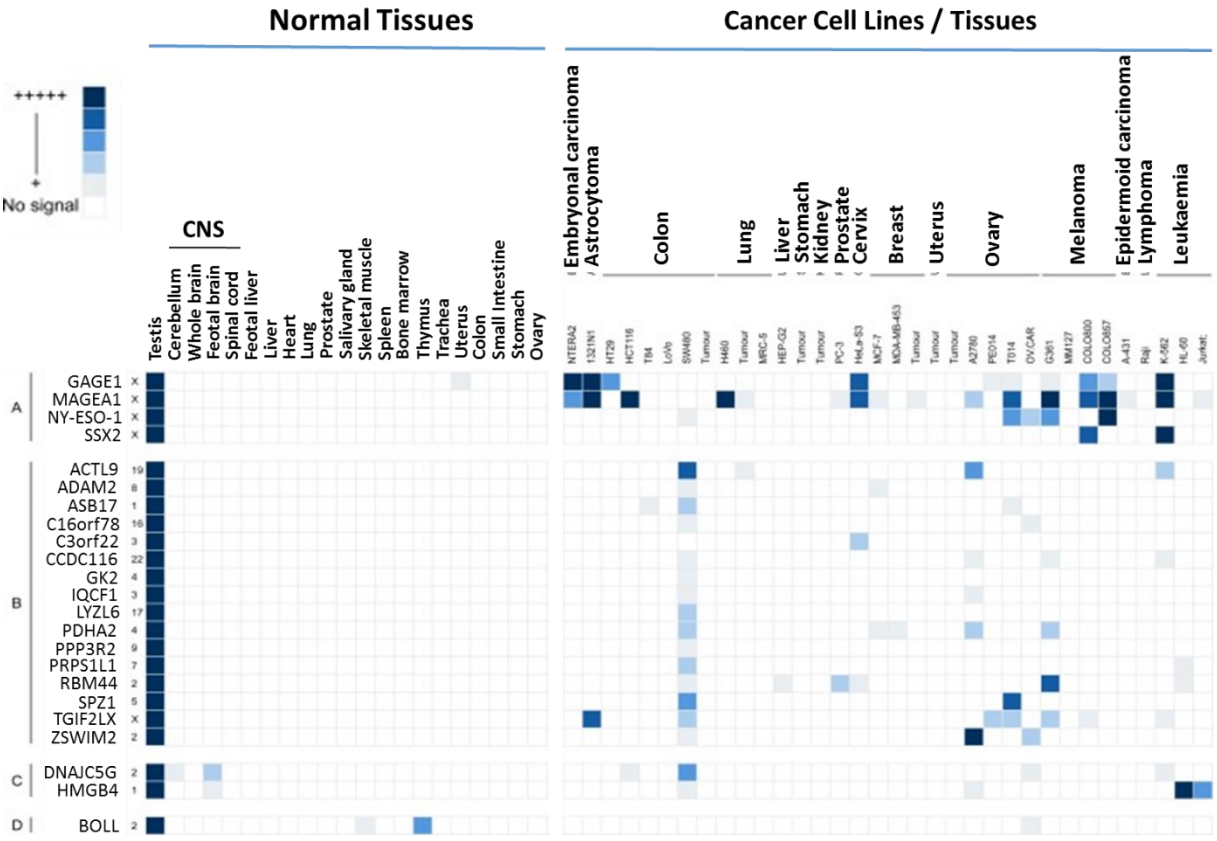


Figure 3.6. Grid representation of the gene expression profiles for the 19 genes for which RT-PCR indicated expression in at least one cancer type.

Each gene has a row allocation on the grid; the presence of a filled square in any column indicates that a positive RT-PCR signal was obtained. The shade of the blue indicates the intensity of the visualised signal seen on the agarose gel relative to the intensity of the band seen for the testis sample; the shade is not a reflection of the relative expression levels, rather it is a simple qualitative representation of PCR product intensity. The columns represent the tissue/cell type tested: normal tissues in left hand panel and cancer tissues/cells on the right side of the grid. The expression profiles for four previously characterised X chromosome encoded CT genes are shown as a positive control (set A: *GAGE1*, *MAGEA1*, *NY-ESO-1*, *SSX2*). The majority of the newly identified putative CT genes had expression restricted to the testis in normal tissues. Three of the genes had expression profiles consistent with limited expression in normal tissues as well as testis and at least one cancer type (sets C and D). The chromosomal location of the genes is given in the column to the right of the gene names; only one of the novel genes is X-encoded.

3.4. Meta-analysis of candidate genes expression profiles

In order to explore the possible clinical relevance of the newly identified genes, we conducted meta-analyses using patient-derived cancer microarray data, including 13 cancer types in a range of 80 microarray data sets (Feichtinger *et al*., 2012a). Full details of the arrays included in this pipeline and how they were chosen are provided in the original publications from the McFarlane group (Feichtinger *et al.* 2012a; Feichtinger *et al.* 2012b). It is a limitation of this study that more cancer types and more arrays were not included but the selection of high-quality arrays that was conducted previously is a trade-off here and has been discussed further in the relevant publication describing the design of this tool (Feichtinger *et al.*, 2012a).

Microarray data was collated from the various array sets for which a given gene was present and the combined data was then meta-analysed to investigate whether there was a statistically significant upregulation in gene expression. This meta-analysis was conducted using a freely available resource (CancerMA) that stemmed from work previously conducted by the McFarlane group (Feichtinger *et al.* 2012a). An example of a significant meta-upregulation of an individual gene (*SPZ1*) is given in the Forest plot profile for ovarian cancer in Figure 3.7. The additional Forest plots are provided in the Appendix (Figures 14-19).

First, we investigated the expression profiles of 18 of the 19 genes that exhibited expression in at least one of the cancer cell types tested by RT-PCR (one gene, *TGIF2LX*, was not present on the microarrays). Of these, 9 (50%) of the genes showed meta-upregulation in either ovarian and/or prostate cancers (see Figure 3.8). The circos plots, like those displayed in Figure 3.8, summarize the Forest plot information for multiple genes. The microarray datasets were chosen as they provided good quality information on the gene expression in cancerous tissue compared to normal tissue. If a gene displayed statistically significant upregulation in a given cancer type compared to normal tissue, as summarised in the Forest plots, then a line linking that gene to that cancer type is displayed within the Circos plot. The thickness of the connecting line is representative of the strength of this meta-upregulation (i.e. the thicker the line, the greater the strength of the statistical increase in gene expression in the cancer) but the thickness of the connecting line is also relative to the number of genes contained within the plot (i.e. the more genes displayed the thinner the lines become). A maximum of approximately 300 genes can be challenged against cancerMA in a single 'job' and it becomes difficult to read the circos

plots if too many genes are contained within the plots, which is why additional plots are provided in the Appendix for data presented in Chapter 5 (see Figures 1-3 in the Appendix).

Of the 33 genes with meiCT gene potential (based on expression patterns in normal tissue), 14 genes did not appear to be expressed in any of the cancer cell types analysed by RT-PCR (see Table 3.1). To further explore the possibility that these genes are CT genes, we used 11 of the 14 genes for meta-analyses using the 80 cancer gene expression microarray data sets (3 of the genes, *C1orf141*, *HEATR7B* and *SATL1*, were not present on the microarrays) and found expression profiles for 5 of these genes were indicative of a cancer type marker for ovarian and prostate cancers (45.5%; Figure 3.9). So, nearly half of the genes that were present on the arrays appeared to be potentially relevant biomarkers for these cancer types. A further 5 (10 genes in total) were expressed in at least one single cancer data set (see Figure 3.10), indicating the potential to mark a specific sub-group of tumours within a cancer type. Only *C8orf74* exhibited no measurable expression in cancer cells / tissues (although *C1orf141*, *HEATR7B* and *SATL1* could not be analysed via meta-analysis due to their absence on the arrays).

Whilst the meta-analysis revealed 9 genes to be upregulated for two given cancer types (ovarian and prostate cancer), analysis of single cancer data sets from the 80 cancer data sets used reveals evidence for activation of a total of 15 of the 18 candidate genes in at least one patient-derived sample set (83.3%; Figure 3.10). So, the range of possible cancer types for which these genes are potentially relevant is broader than the meta-analysis indicates. For the genes that lack a statistically significant upregulation it suggests that they are less likely to be good biomarkers for the cancer type as a whole (e.g., for screening purposes) but this does not mean that there may be a subset of tumours within the cancer type for which individual genes could be clinically relevant.

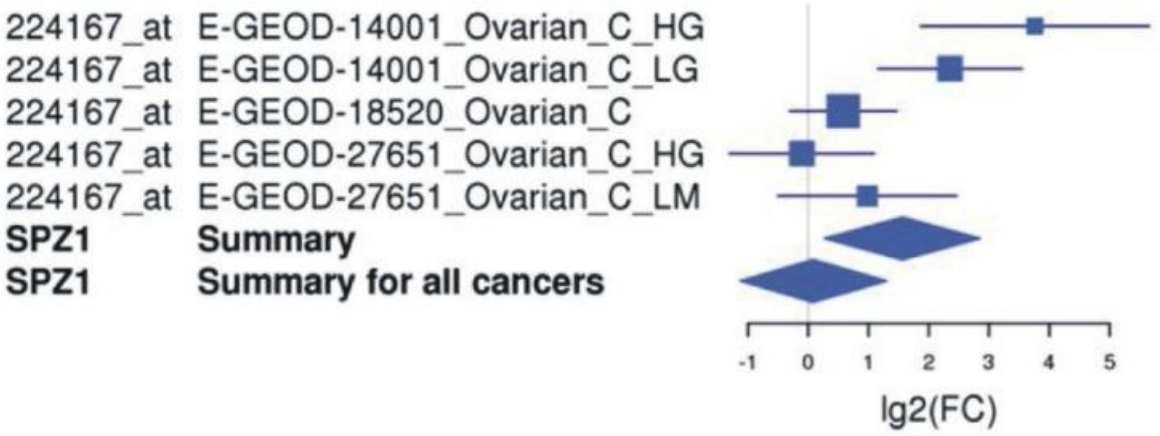


Figure 3.7. An example of a Forest plot for one of the identified meiCT genes, *SPZ1*.

The Forrest plot indicates that *SPZ1* is upregulated in ovarian cancers with an approximate 1.8 log-fold change over matched normal *vs.* cancer array sets. The upper five squares illustrate the individual microarray studies, with the confidence intervals for the individual studies represented by the horizontal lines. The size of the squares is proportional to the weight assigned to the individual study. The upper diamond is a summary of all five data-sets showing significant upregulation for ovarian cancer. When all the cancer-types on the arrays were meta-analysed there was no significant change in expression over normal tissues (lower diamond). The additional Forest plots for the genes displaying significant meta-upregulation are provided in the Appendix (Figures 14-19).

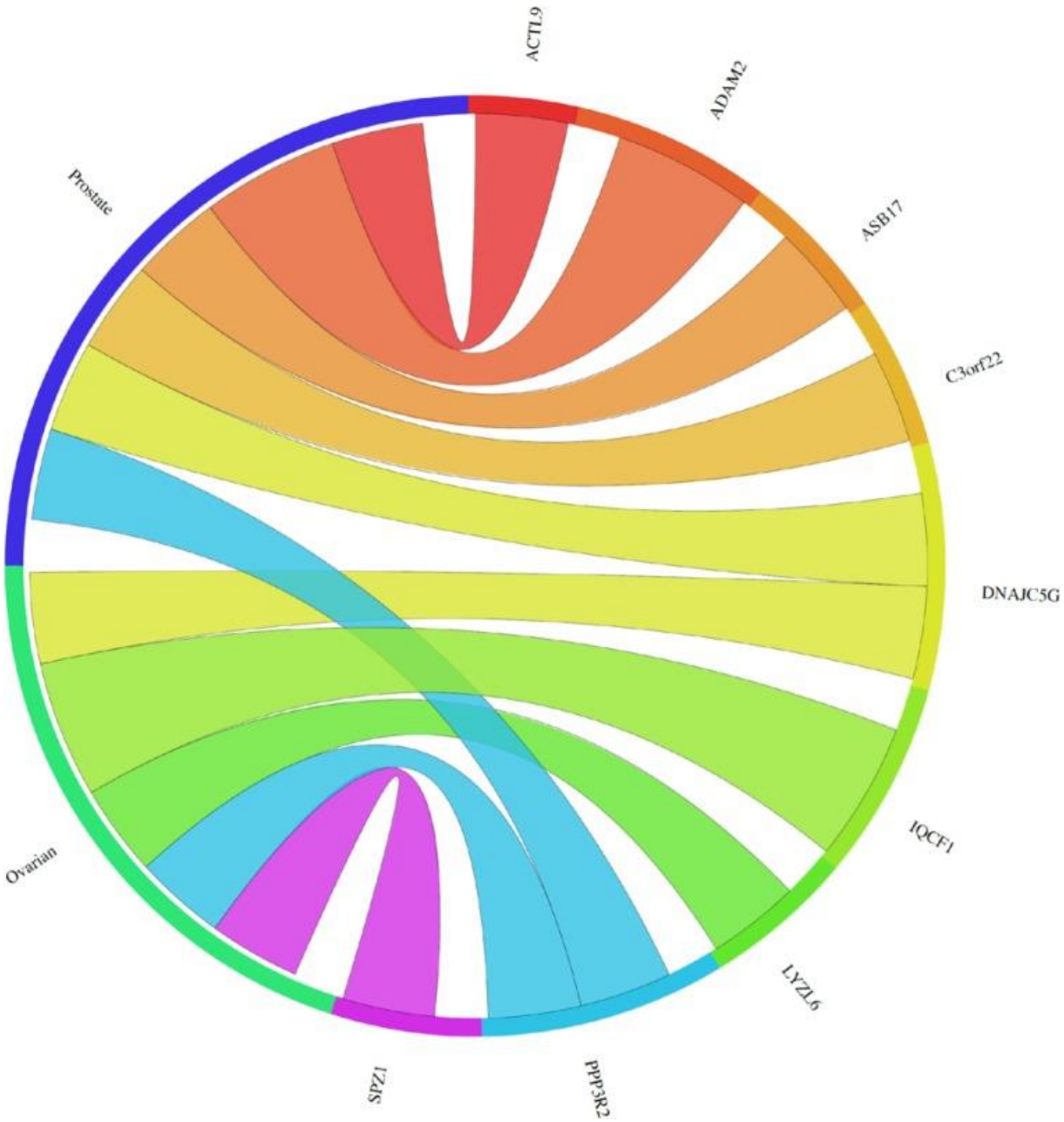


Figure 3.8. Circos plot showing the meta-change in gene expression by cancer type.

The Circos plot links the individual genes which displayed a significant upregulation in expression when combined microarray data-sets were meta-analysed to the corresponding cancer type (ascribed by tissue type). Only the identified putative CT genes that were indicated to have expression in at least one cancer cell type as assessed by RT-PCR (sets B-D in Figure 3.6) are displayed here. One of the 19 genes that fell into this category, *TGIF2LX*, was not on the arrays and as such not subject to the meta-analysis. The plot shows 9 genes exhibit meta-upregulation in ovarian and/or prostate cancers. The weight of the connection corresponds to the magnitude of the meta-change in gene expression.

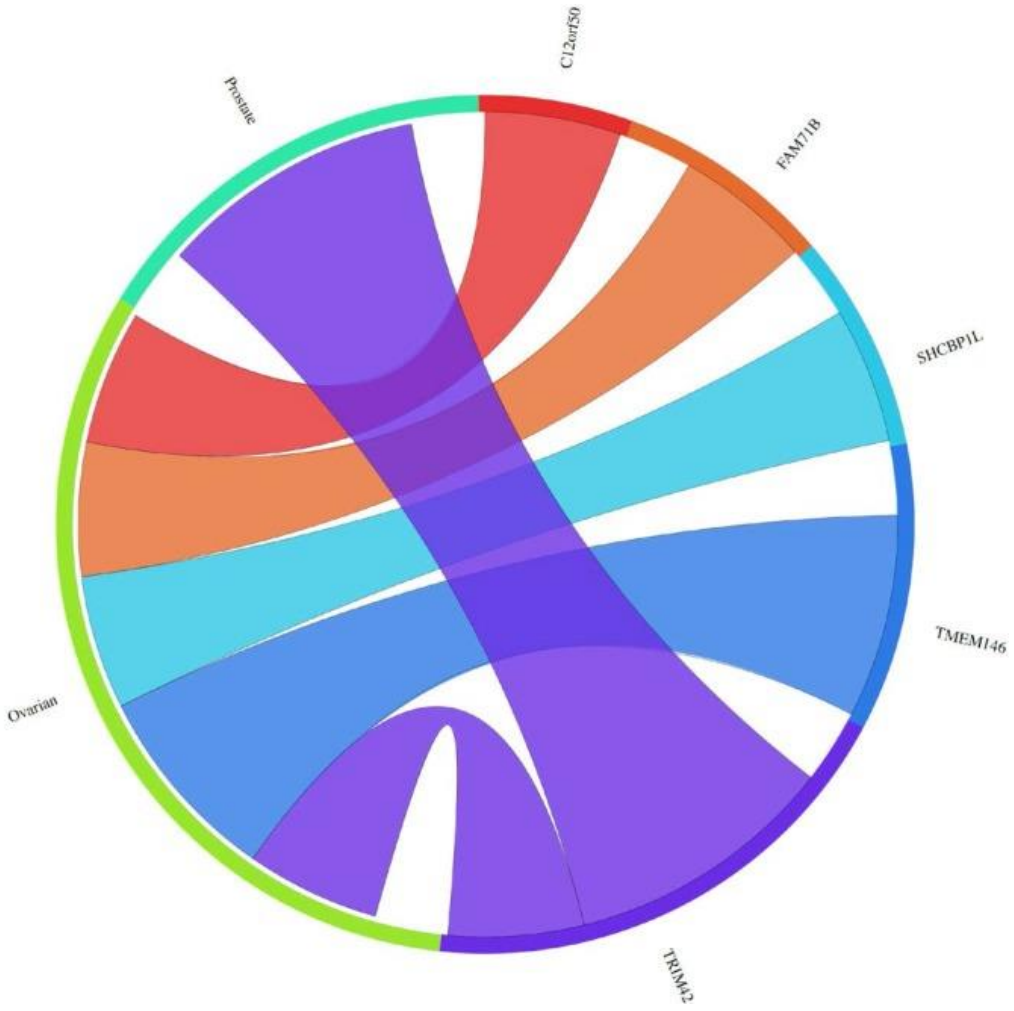


Figure 3.9. Circos plot showing the meta-change in gene expression in relation to corresponding cancer types (ascribed by tissue type) for identified genes that did not exhibit expression in any of the cancer cells or tissues assessed by RT-PCR.

Three of these 14 genes, *C1orf141*, *HEATR7B*, *SATL1*, were not subjected to the meta-analysis as they were not present on the arrays. The plot shows that 5 genes exhibit meta-upregulation in ovarian and/or prostate cancers. The weight of the connection corresponds to the magnitude of the meta-change in gene expression.

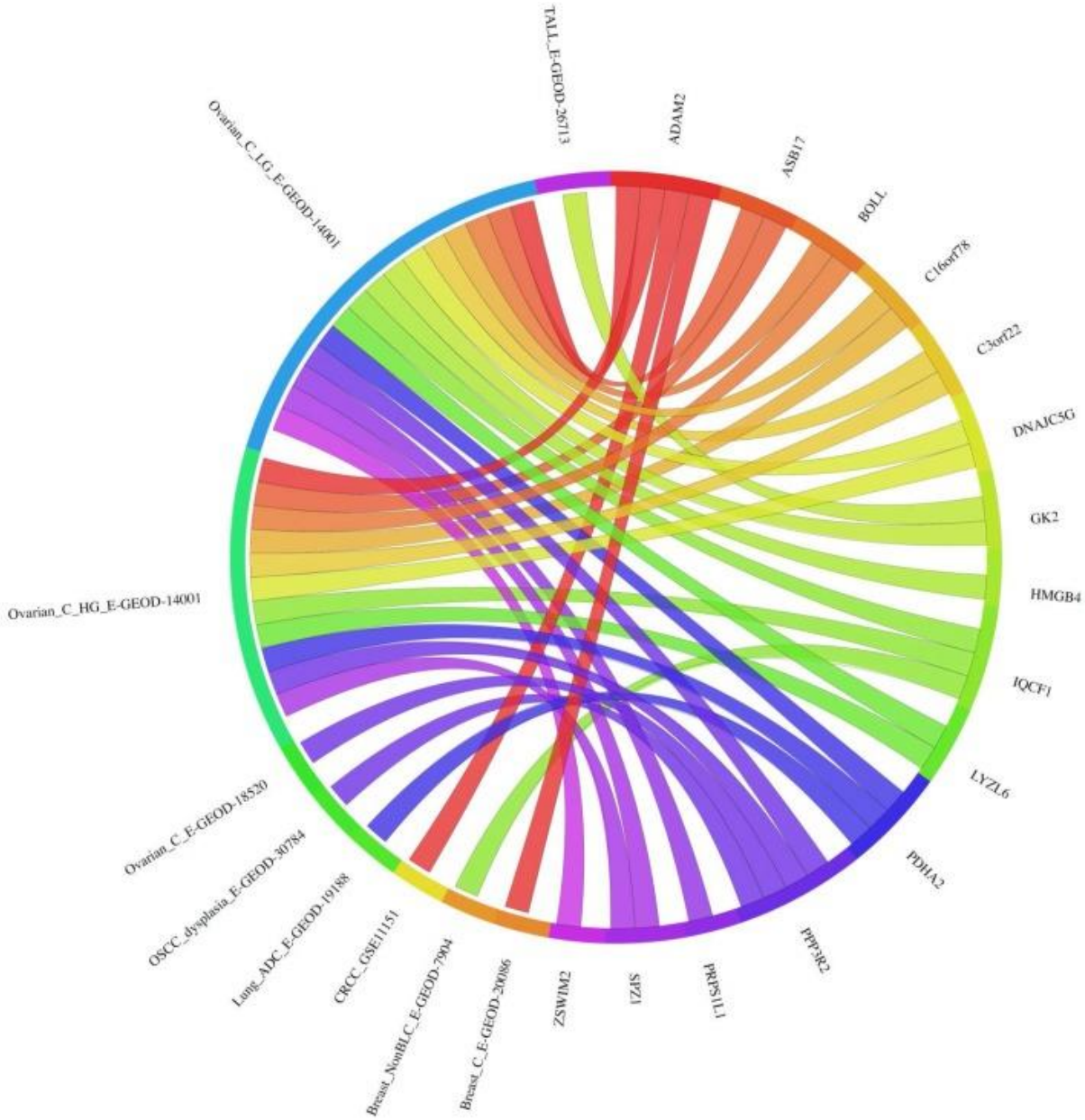


Figure 3.10. Circos plot of single microarray analysis for identified meiCT genes exhibiting expression in cancer cells/tissues as revealed by RT-PCR.

The expression of the genes is given corresponding to the data set in which a statistically significant upregulation was observed for the tumour *vs*. normal array analysis. For data set designations see (Feichtinger *et al*., 2012a; Feichtinger *et al*., 2012b). *TGIF2LX* was not present on the arrays. The plot shows 15 of the remaining 18 genes analysed are statistically significantly upregulated in at least one cancer data set. Three genes (*ACTL9, CCDC116, RBM44*) show no significant upregulation in any of the array sets and so are not shown on the Circos plot.

3.5. Discussion

CTAs are cancer-specific biomarkers with considerable potential as prognostic and diagnostic markers as well as being therapeutic targets. The current classification system for CTA genes continues to be based on that put forward by Hoffman and colleagues (Hofmann *et al*., 2008). It has since been proposed that a sub-category of CT genes that are associated with meiosis exists. These genes, termed meiCT genes, were identified from an in-silico pipeline originating with putative meiotic genes (Feichtinger *et al*., 2012b; Sammut *et al*., 2014). A further 29 meiCT-genes have now been identified (Sammut *et al*., 2014). As was the case for previously characterised meiCT genes, most of these genes are autosomally encoded (28 out of 29; see Table 3.1). This finding is consistent with the transcriptional inactivation of the X chromosome during male meiosis (Turner, 2007). Many of the previously identified CT genes are encoded by the X chromosome and expressed in spermatogonial mitotic cells (Almeida *et al*., 2009).

An additional commonality with the previously characterised meiCT genes was the fact that many of this new cohort were shown to be upregulated as a general marker for ovarian cancers: 10 of the new cohort displaying a meta-change increase in gene expression were in ovarian cancer. This again raises the possibility of using the meiCT genes to improve the diagnosis of this diverse and pernicious cancer type. It may be the case that genes that have a normal biological function (i.e. meiotic role) in the foetal ovary are preferentially reactivated in cancers of this tissue type. However, a significant proportion of ovarian cancers are believed to be extra-ovarian in origin (Berns and Bowtell, 2012; Erickson *et al*., 2013; Sorensen *et al*., 2015). So the picture is complex and the reason for this finding is not entirely clear. The possibility of stimulating meiotic gene expression via oncogenic paracrine signalling mechanism, even if the tumour originates outside of the germline tissue, still exists.

Ovarian cancers are currently most frequently treated with cytoreductive surgery and chemotherapy, although these types of tumours are immunoreactive and there is currently extensive work ongoing to explore the application of immune-based therapies for their treatment (Nelson, 2015; Tse *et al*., 2014). Thus, the identification of ovarian cancer-specific biomarkers such as these is of considerable potential clinical importance.

Recent work has demonstrated that sub-groups of 26 germline and placental specific genes can be used to delineate aggressive metastasis-prone lung cancers (Rousseaux *et al*., 2013b). The

ability of these 26 germline genes to act as surrogates for a worse clinical outcome was independent of other traditional methods of disease stratification. This indicates that small sub-groups of tissue-specific genes can serve as accurate biomarkers in the stratification of complex and heterogeneous cancers. The clinical implications of this are far reaching as they offer extensive potential in establishing optimal treatment strategies and both preventing unnecessary or unefficacious treatment as well as identifying patients who would benefit from a particular strategy when they would not be offered the treatment according to established methods of guiding therapeutic decisions. This work provides a paradigm for how germline gene expression in cancers can be applied to clinical stratification of complex disease. Having a definitive on/off expression profile, as observed with many of the meiCT genes, greatly enhances the potential simplicity of the application of these genes in novel prognostics technologies.

A number of studies have now specifically explored the potential of expression of human germline genes as cancer biomarkers. Interestingly, whilst common genes have been identified, the various studies have all identified additional distinct genes indicating that the full mining of data sets of this magnitude require multiple and diverse approaches. For example, this current study has identified 29 new genes with tight germline and germline/CNS tissue-specific expression restrictions; however, a recent seminal and extensive study of human male germline/placental genes only identified 16 of the 29 (55.1%) genes reported here as germline/placental-specific (Rousseaux *et al*., 2013b). A similar trend is seen when analysing previously reported meiCT genes (Feichtinger *et al*., 2012b). This highlights how the investigative approach used will impact on the research findings and the need to establish a definitive list of germline genes. The issue of creating a definitive list of germline genes will be revisited later in this thesis.

In addition to serving as cancer biomarkers the meiCT genes may serve as therapeutic targets via a variety of routes. Firstly, the immunogenicity of the gene products of the meiCT genes remains very poorly characterised. Their highly stringent tissue specificity infers that their gene products could potentially serve as tumour-specific immunotherpeutic targets. Given the heterogeneity of cancer, both intra- and inter-tumour, the development of a large bank of biomarkers will be of increasing importance in the development of personalised treatment strategies (Jamal-Hanjani *et al*., 2015; Mendelsohn, 2013; Seoane and De Mattos-Arruda, 2014).

Germline genes in *Drosophila melanogaster* have been shown to serve to drive the oncogenic programme (Janic *et al*., 2010). The human orthologues of these genes are also widely activated in human tumours (Feichtinger *et al*., 2014b) supporting the notion that a soma-to-germline transition is a hallmark of many cancers. Additionally, other CT genes have been demonstrated to be required for cancer cell proliferation and their depletion can serve to sensitise cancer cells to standard therapeutic agents (Cappell *et al*., 2012; Whitehurst *et al*., 2007). Thus, not only can germline gene products potentially offer direct targets for drug therapies, but depletion of their activity can also serve to enhance the efficacy of existing therapies potentially enabling reduced dose regimens, which will limit undesired drug toxicities. The cancer-specific nature of meiCT gene expression makes these genes exceptionally attractive for further exploration in drug targeting and drug sensitisation.

The suggestion that germline genes are oncogenic infers that some of the genes identified here could play a tumour initiating and/or progression role. One of the genes validated here is a member of the *ADAM* gene family - *ADAM2*. The ADAM proteins often exhibit proteolytic activity and have emerging roles in the invasive properties of specialist cells within the placenta (Pollheimer *et al*., 2014). Such proteolytic functions may promote invasion and metastasis of solid tumours. A putative role for other members of the adamalysins in the aetiology and pathology of colorectal cancer and melanoma has been proposed (Moro *et al*., 2014; Przemyslaw *et al*., 2013). This might indicate that not only are germline genes required for oncogenesis, they might drive metastasis and thus offer cancer-specific intervention points to stop the lethal spread of tumours.

SPZ1 was shown to be testis-specific in our normal tissue panel, a finding previously shown by Hsu and colleagues (Hsu *et al*., 2001). In their study, they further showed that the gene was expressed both in the testis and epididymis. We found a positive expression signal in a colon and ovarian cancer cell line and on meta-analysis there was a significant upregulation in ovarian cancer. It has since been shown by Hsu and colleagues that *SPZ1*, which encodes a transcription factor, acts as a proto-oncogene to promote cellular proliferation and tumour formation in a mouse model (Hsu *et al*., 2005). Despite this, *SPZ1* has not previously been recognised as a CT gene (Almeida *et al*., 2009).

It has been suggested that another of the novel meiCT genes identified here, *SHCBP1L*, encodes a protein with strong homology to a mouse protein present in proliferating cells and may have similar physiological effects (Sood *et al*., 2001). It remains unexplored whether this protein indeed acts through similar signal transduction pathways to promote proliferation but as the gene has shown a statistically significant meta-change upregulation, again in ovarian cancer, makes this possibility worthy of further exploration. The protein product of this gene has been found in normal testis but not normal ovarian tissue and through an interaction with HSPA2 contributes to spindle stability during meiosis (Liu *et al*., 2014). Thus, it would be worth investigating whether the protein is indeed present in ovarian cancer.

The fact that the cDNA used in the PCR validation was synthesized from RNA extracted *post mortem* has potential limitations. The RNA may have degraded prior to extraction. Limited information is provided from the companies about the patients from which the RNA was extracted – details that were provided by the company are supplied in the Appendix C (on CD). In a study aimed at identifying novel CT genes using a massively parallel signature sequencing technique Chen and co-workers showed that distinct sources of 'normal' tissues produced sometimes widely varying results in terms of whether a gene was expressed in that tissue type (Chen *et al.*, 2005). RNA extracted *post-mortem* will often be from the elderly population, in which there will be increased likelihood of an occult cancer that had not been diagnosed prior, or contributing, to the death of the individual. Even if the tissue that had occult neoplasia had not been sampled, given what is known about field cancerization, an occult cancer could cause genetic changes in cells distant from the site of the tumour. Some of the commercially sourced RNA was pooled from tissues from a number of individuals; this would increase the likelihood of containing cancerous material within the sample. It is a major limitation of this work that the experiments were in general not repeated. Ideally experiments should be repeated at least three times to improve the validity of the results and when dealing with *post mortem* samples this becomes even more important. However, our screen was crude and basic and limited by the practicalities of cost and time. Some genes may have been included that would have not passed more vigorous validation steps but equally we may have been too quick to exclude other candidates.

The results of the meta-analysis (using the cancerMA tool) should not be over-interpreted. There are limitations of this approach and the fact that many of the upregulations in gene expression may be random should not be ignored. As can been seen in the circos plots shown

in Appendix Section 3 (Figure 20), when a random selection of genes are chosen a similar proportion display upregulations – this suggests that the upregulations presented here may also be random. Further, when PCR control ('housekeeping') genes are chosen these also display meta-upregulations in the cancer datasets included (see Figures 21 and 22, in Section 3 of Appendix). Cancer is, however, a very heterogeneous disease and the fact that 'housekeeping' genes that are supposed to be present at roughly equal levels are not, is not unduly surprising. This fact has indeed been highlighted by others (e.g., Cicinnati *et al*., 2008). Not all control genes will be present in equal amounts in all cancer types or sample groups. Additionally, we selected some genes known to be upregulated in cancer – the circos plot for this list of genes is also shown in Section 3 of the Appendix (Figure 23). Many of the genes do display upregulations but the information is limited by the number of cancer types included on cancerMA. The data produced by cancerMA is further limited by the quality of the raw data of the original research on which the meta-analysis is based on. The findings and conclusions drawn must not be over-interpreted.

The concentration and purity of the RNA was assessed using NanoDrop (ND-1000) spectrophotometry technology prior to cDNA synthesis. This is a widely used and accepted technique but there are limitations in terms of the limited information about the possible degradation of the RNA and this is a limitation of this study. RNA extracted from human tissue samples that were collected for this research project (data presented later in thesis) were subject to more robust analysis using the Experion^{TM} RNA analysis kit (BioRad), which does provide a greater degree of reassurance that the RNA is of good quality and not degraded. The fact that the RNA was not subjected to more robust analysis prior to cDNA synthesis is a further limitation of the results presented in this chapter.

Data on The Human Protein Atlas (HPA) is freely available on www.proteinatlas.org (Uhlen *et al.*, 2005) and provides a useful reference point for what has been established about gene expression and protein levels in certain tissues. The data does have limitations, for example, in that some antibodies have not been fully validated but our results will be compared to the data presented on HPA for the most promising candidate genes put forward here. *ADAM2* and *SPZ1* have been mentioned already above.

Data on HPA indicates that mRNA is present in normal testis but not in other tissue types for *ACTL9*. The antibodies used are polyclonal and the website states that they have not been

externally validated. Immunohistochemistry (IHC) indicates that the protein is present in normal renal tubular cells as well as cytoplasmic staining within hepatocytes. The website acknowledges that the presence of the protein within the normal kidney and liver is not consistent with the gene expression data. *ACTL9* has recently been implicated as playing a possible role in eczema (Hirota *et al*., 2012). It is acknowledged that the function of ACTL9 remains obscure.

ADAM2 is expressed in normal testis and this is consistent with the strong cytoplasmic staining seen in developing spermatocytes/spermatids within the seminiferous tubules of normal testis. The presence of mRNA or the protein was not demonstrated in other normal tissues according to HPA, which is consistent with our findings and those of other groups (Cho, 2012, Choi *et al.*, 2016, Pollheimer *et al*., 2014). Thus, it would appear likely that the antibody used for HPA is specific. The protein is believed to play a role in sperm-egg interactions and may additionally be present within the placenta (Cho *et al.*, 2012, Pollheimer *et al*., 2014) and as postulated above could conceivably play a role in tumour cell migration and/or invasion if expressed in cancer. However, the function of *ADAM2* may be different in primates compared to mice and the precise function of this gene in humans remains unclear (Choi *et al.*, 2016).

Consistent with our findings *ASB17* is expressed in normal testis only according to HPA, though no information on protein expression is provided. There are no further references to this gene in humans and its function remains obscure.

Again, consistent with our data the expression of *BOLL* is restricted to testis in normal tissues according to HPA. The polyclonal antibody used also displayed presence only in the testis (developing spermatocytes) consistent with the gene expression data. Additionally, there was weak staining on IHC of thyroid cancer. Despite the fact that the HPA antibody appears to display specificity for the protein and be present in thyroid cancer, no further reference to the possible role of this gene in thyroid cancer has been identified – it is however, worthy of further scientific exploration. *BOLL* has been implicated in lung and colorectal cancer and as such can be considered a CTA due to its germline-restricted expression pattern (Kim *et al*., 2011; Tessema *et al*., 2009). As *BOLL* has been implicated in a rare syndrome in which there is marked hypomethylation in specific regions of the genome (Simo-Riudalbas *et al*., 2015), it could be predicted that this was the mechanism for reactivation in cancer. However, hypermethylation of its promotor region in cancer appears more common (Tessema *et al*., 2009;

Kang *et al*., 2015). The encoded proteins may be involved in cell cycle progression in normal testis and though a functional role in cancer has not been elucidated, it has been suggested it may be a useful biomarker for the detection of early CRC (Kang *et al*., 2015).

C3orf22 is an uncharacterised gene for which no published scientific papers proposing a function or linking it to cancer have been identified. The gene expression data present on HPA is consistent with the data presented here, in that it does appear to be germline restricted. Moreover, a polyclonal antibody which has not been independently validated but does display consistent results in that the protein is only found in normal testis according to HPA. Moreover, HPA include IHC data indicating that the protein is found in colorectal, head and neck, lung, urothelial, testicular cancers and melanoma. So, a possible oncogenic role for this gene would certainly appear possible.

C16orf78 is another uncharacterised gene. HPA mRNA data is consistent with our assertion that expression is limited to the testis in normal tissues. A polyclonal antibody, as yet to be externally validated indicates that the protein is present in many cancer types as well as normal testis. However, given the wide presence in cancer, as well as strong staining throughout the testis, the antibody may lack specificity. Very little is known about this gene and it was not shown to have any upregulations in cancer *vs.* normal tissue according to CancerMA (Feichtinger *et al.*, 2012a). However, it has been shown that the protein may interact with other proteins known to be associated with prostate cancer, so this would be worthy of further exploration to establish a possible oncogenic link (Tao *et al.*, 2012).

The gene expression data included on HPA for *CCDC116* is broadly in agreement with our data indicating a testis-selective expression pattern with very low levels of mRNA in some other normal tissues. A polyclonal antibody used by HPA displays strong cytoplasmic staining in pancreatic islet cells. The majority of cancer samples tested using IHC displayed negative or weak staining. The strongest staining was seen in carcinoid tumours. Thus, a possible relevance for this gene in neuroendocrine tumours of the gut is worthy of further exploration, especially in light of a possible normal physiological role in specialised cells of the pancreas. Tsolakis and colleagues have in fact further characterised the antibody in a range of pancreatic tumours, finding it to be present quite widely (Tsolakis *et al.*, 2012). The polyclonal antibody used by HPA has not been externally validated; though some external validation was carried out by Tsolakis and colleagues (2012), they did not correlate the presence of the protein with

gene expression data in the specific tissues in which it was shown to be present. More recently, CCDC116 has been shown to be a constitutive part of the centrosome of sperm cells (Firat-Karalar *et al.*, 2014). Further knowledge about its normal physiological role may help shed light on possible functional roles in cancer.

DNAJC5G displays a testis selective expression with a much higher abundance of mRNA in the testis according to data on HPA. The HPA polyclonal rabbit antibody against the protein displayed moderate staining within the seminiferous tubules of the testis as well as occasional positive staining in follicular adenocarcinoma of the thyroid, urothelial and some lung cancers. This antibody has not been externally validated. *DNAJC5G* remains largely uncharacterised, though it belongs to a wider gene family of heat shock proteins. These proteins are implicated in numerous disease processes and are important with regards to 'protein stress' that has the potential to lead to somatic mutations (Dekker *et al.*, 2015). The gene has been shown to be expressed in a neuroblastoma cell line and upregulated following ethanol treatment (Ferns *et al.*, 2012). The possible relevance of an upregulation or reactivation of this gene in cancer remains obscure.

The IHC data included on HPA is inconsistent with the gene expression data for *GK2* indicating normal expression restricted to the testis; the latter result being consistent with our findings. A polyclonal antibody lacking external validation was used displaying moderate, predominantly cytoplasmic staining in several normal tissue types (strongest in hepatocytes). There was also positive staining in numerous cancer types but most strongly and widely in endometrial, hepatocellular, prostate and thyroid cancer. Given the widely different results of the IHC to gene expression results it is likely that the antibody lacks specificity for the protein and this requires clarification. *GK2* has recently been identified through a serological analysis of recombinant cDNA expression libraries (a technique used previously to identify many CT genes – for example, Mashino *et al.*, 2001) as a potential biomarker and therapeutic target for early breast cancer (Zuo *et al.*, 2016). Thus, although the HPA antibody may lack specificity for the antigen, further exploration of this gene and its potential relevance in a subset of cancers would appear worthwhile. According to cancerMA, however, the gene is not significantly upregulated in any cancer type when the included array data are meta-analysed.

IHC data for HMGB4 is consistent with the gene expression data indicating the gene is only switched on in the testis. Normal testis but no other normal tissue or indeed cancer types stained

positively. Additionally, the protein has been validated, showing consistent single band on Western Blotting. Very little research has linked this gene to cancer, though one important paper has suggested the protein may be the binding site for cisplatin and one of the main reasons that testicular germ cell tumours display such high levels of response to cisplatin chemotherapy (Park and Lippard, 2012). Our preliminary PCR data indicate that *HMGB4* is expressed at moderate levels in leukaemia cell lines, so further investigation of a link between this gene and blood cancers would be worth investigating. The presence of the antigen in a subset of tumours may pave the way to targeted cisplatin-based treatment for these tumours, though initial data suggest few tumours express this gene.

Data from HPA is consistent with ours showing that *IQCF1* has expression limited to the testis in normal tissues. No antibody has been tested and protein analysis for this gene is lacking on HPA. Few studies have investigated the function of this gene and no link to cancer has yet been formally identified. However, a functional role in sperm capacitation has been postulated (Fang *et al.*, 2015).

LYZL6 has been stained positively in normal testis using IHC using two separate polyclonal antibodies, consistent with gene expression data indicating testis restriction expression. Weak to moderate cytoplasmic staining was seen in some ovarian, pancreatic and stomach cancers according to HPA. No formal link to cancer for this gene has been identified. Functionally it may have immune and antibacterial properties within the testis (Wei *et al.*, 2013). A possible role in cancer remains obscure.

PDHA2 has been shown to be testis restricted both at an mRNA and protein level on HPA. The gene expression data is consistent with our data showing expression limited to the testis in normal tissues. HPA used three separate polyclonal antibodies, none of which have been externally validated. These antibodies displayed varying cytoplasmic localisation in developing spermatocytes within the seminiferous tubules as well as Leydig cells located between the seminiferous tubules. IHC staining was generally negative in the cancer samples for one of the antibodies but staining for the other two was moderate both cytoplasmic and nuclear across many cancer types. So, the picture for this gene and its relationship to cancer is not clear and some of the antibodies may lack specificity for the antigen. Given that this gene is intronless, our PCR data in the cancer samples, where expression was indicated in a colorectal, ovarian and melanoma cancer cell line, needs to be treated with caution. It is a major

limitation of this work that genomic DNA contamination controls were not used but intron-spanning primer sets were used where possible. Functionally PDHA2 may be involved in the acrosome reaction and sperm capacitation (Kumar *et al.*, 2008). It has also been shown that demethylation of the promotor region for this gene can induce expression in somatic tissues (Pinhiero *et al.*, 2012), so this could provide a route to aberrant gene expression in cancer.

PPP3R2 forms part of a regulatory subunit of calcineurin and is testis specific; it is important for the acrosome reaction as well as sperm motility (Miyata *et al.*, 2015). HPA state that the polyclonal antibody they used targeted antigen from more than one gene and the IHC data should be treated with caution; the antibody has not been externally validated. Gene expression data for *PPP3R2* presented on HPA is consistent with ours in that the gene is testis restricted and this is also consistent with the findings of previous studies (Liu *et al.*, 2005; Miyata *et al.*, 2015). We did not find convincing expression in any cancer cell type tested and no further studies have linked this gene to cancer. Thus, the HPA data displaying moderate to strong cytoplasmic as well as nuclear immunoreactivity widely in endometrial and renal cancers in additional to less common but positive staining in several other cancer types, requires further exploration to establish a definite link to cancer. Unrelated to cancer but of interest to note, it has been proposed PPP3R2 may form part of a drug target for a reversible male contraceptive (Miyata *et al.*, 2015).

No antibody against PRPS1L1 has been tested by HPA and no other reference to cancer has been identified in the literature. However, HPA do provide gene expression data consistent with our own that expression is restricted to the testis in normal tissues. The coding region of this gene is intronless and the suggested expression in a colorectal cancer cell line should be treated with caution for this reason; there may have been genomic DNA contamination within the SW480 cell line as stated previously. There were no meta-upregulations in any cancer types included on cancerMA and a possible link to cancer for this gene is at present very tenuous.

RBM44 is a highly conserved intercellular bridge protein that interacts with TEX14 in the testis (Iwamori *et al.*, 2011). Gene expression and IHC data presented on HPA is consistent with our data also indicating testis restriction in normal tissues. However, the polyclonal antibody used by HPA has not been externally validated and presumed off-target binding was disregarded. Within cancer samples, moderate cytoplasmic immunoreactivity was observed in carcinoid tumours and gliomas, as well as some thyroid cancers. Occasional melanoma samples also

tested positive, which is perhaps consistent with gene expression indicated in one melanoma cell line in our PCR experiment.

Overexpression of Spz1 has been shown to stimulate cellular proliferation and induce infertitilty in mice (Hsu *et al.*, 2004). No studies linking *SPZ1* to cancer in humans has been identified, however as a transcription factor that can induce cellular proliferation if the gene were switched on it could have significant adverse consequences. HPA gene expression data, like ours, indicates that the gene is expressed only in the testis. Our data indicated expression in a colorectal and ovarian cancer cell line. Positive IHC staining was seen in some colorectal and ovarian cancer samples tested on HPA but stronger staining was observed for melanoma, glioma, breast and cervical cancers. The polyclonal antibody used by HPA has not been externally validated. Moreover, it is stated that expert annotation of the IHC results in normal tissues, where immunoreactivity was demonstrated quite widely, could not be performed due to inconsistent results. In light of the gene expression data, it is likely that the HPA antibody lacks specificity.

TGIF2LX is a testis-specific gene present on the X-chromosome and testis restricted expression in normal tissues is shown by our data and that presented on HPA. IHC data included on HPA for TGIF2LX shows nuclear staining in developing spermatocytes. There was very occasional but strong nuclear or cytoplasmic observed in colorectal, head and neck and urothelial cancers. It may be of clinical relevance in a limited subset of cancers but this requires further clarification. Further discussion about related genes is provided in Chapter 5 (Section 5.3.5.1).

The function of ZSWIM2 remains obscure but it is thought to be a testis-restricted protein, which may regulate apoptosis during germ cell differentiation via paracrine or autocrine means (Nishito *et al.*, 2006). Gene expression data on HPA is, like ours, also consistent with *ZSWIM2* being a testis-restricted gene. Our preliminary PCR data suggest the gene is activated in two separate ovarian cancer cell lines, though the gene was not meta-upregulated when ovarian cancer microarray datasets were metaanalysed. HPA used two polyclonal antibodies, which both displayed positive staining in some normal tissues inconsistent with the expected result. Serous ovarian cancer was the most common cancer type displaying positive staining on IHC, with very occasional positivity in endometrial and lung cancer, and negative staining in other cancer types. The HPA antibodies may lack specificity for ZSWIM2 but it is interesting to note that the IHC staining in the cancer samples broadly correlates with our PCR data. Thus, it is

possible that ZWIM2 is a relevant biomarker for certain ovarian adenocarcinomas but this requires substantiation through further focussed research.

3.6. Closing remarks

Germline genes are emerging as important cancer-specific factors. The extent of their clinical importance is only now starting to come into focus, either as biomarkers for stratification and diagnosis, as oncogenic activators and as drug or immunotherapeutic targets. What is becoming increasingly apparent is the heterogeneity of tumour cell populations. This is the driver for the need to identify a large cohort of cancer cell markers that individually or in combination can target and mark a large number of tumour types and cell populations. The list of genes identified in this chapter contributes to the growing catalogue of cancer-specific biomarkers.

4. Characterisation of two novel CT antigens and evaluation of their immunogenic potential

4.1. Introduction

So far I have presented work looking at the gene expression of germline and CT genes. As has been demonstrated by Rousseaux and colleagues, such analysis alone can serve as a powerful stratification or predictive tool for patients with cancer (Rousseaux *et al*., 2013b). Gene expression analysis has also been used to categorise cancers into clinically relevant subgroups (for example, Marisa *et al*., 2013; Sadanandam *et al*., 2014; Tan *et al*., 2013). Such an approach may be increasingly used across a wide array of cancer types to help guide therapy from an early stage. However, as anti-cancer therapies are often targeted at a biological function/antigen, it is the protein products that are in general of the most interest. Immunotherapeutic approaches also rely on (tumour-associated) antigen production. Indeed, one of the most attractive attributes of CT, and many germline, genes is that they may encode antigens that are highly specific for the cancer which produces them and in so targeting these antigens (e.g., with immunotherapies), the treatment should not result in deleterious side effects in healthy tissues as the antigen will, so far as the body's immune system is concerned, be unique to the cancer.

So, gene expression experiments are very useful but generally, in terms of therapeutics, there is a need to establish that a protein is produced by the cancer cells which express the genes. This makes it important and relevant to establish whether CT genes do result in antigen production. If this is the case for the novel CT genes which have been identified, then this would make them *bona fide* novel CTAs. This is of fundamental importance as the proteins may have a functional (i.e. oncogenic) role in the cancer. Additionally the antigens may be recognised by the immune system. Thus, the two subsequent questions following confirmation of antigen production are: is there a functional role for these proteins, and are they immunogenic? Either, or both, of these questions may reveal an antigen with therapeutic potential. For example, the biological function of the protein may be disrupted to halt the progression of cancer or result in regression of disease. Tumour associated antigens (TAAs) can also be the target for immunotherapeutic approaches if they are recognised by the immune-system. The function of the immune system can then be harnessed to kill cancer cells that present such antigens when they are recognised as non-self (Mellman *et al*., 2011). This also

requires a deeper understanding of the mechanisms that create a site of immune-privilege within the tumour microenvironment, which prevent the innate function of the immune system from achieving this.

Antibodies are commonly used to detect the presence of antigens within biological material. The presence of the antigen can also be inferred indirectly using enzyme-linked immunosorbant assay (ELISA), which detects circulating antibodies in the serum that have been produced by the patient in response to a particular antigen. ELISA tests also suggest the antigen may be of interest as an immunotherapeutic target, as the antigen must have been able to evoke an immune response in order to test positive using this technique.

Previously in this thesis I have presented work predominantly looking at the expression of germline or CT genes. Here, I took forward two novel candidate CT genes (*C20orf201* and *TEX19*) to see if the antigens were produced and present in cancer tissue. These genes were chosen as our group had identified them as promising cancer-specific markers (Feichtinger *et al*., 2012) and we also had antibodies that had been validated for detecting the antigen of interest. An additional CT gene, *SPO11,* was also investigated as this had been identified as a candidate meiCT gene as well as being previously described as a CTA but there remained uncertainty about its function or indeed presence within cancer (Almeida *et al*., 2009; Hofmann *et al*., 2008; Litvinov *et al*., 2014). SPO11 is also an interesting protein given its known physiological functions during meiosis (see Section 1.4.5). Germline specific augonaute proteins were also investigated as one member, *PIWIL1*, was shown to be of potential interest from the gene expression analysis conducted on CRC samples. Finally, the potential immunogenicity of the novel CTAs were investigated using a T-cell binding assay.

4.2. C20orf201 as a potential biomarker in ovarian cancer

C20orf201 was described as a meiCT gene, which had a testis/CNS restricted expression profile in normal tissues (Feichtinger *et al*., 2012b). It was additionally shown in this study to be clearly expressed in ovarian cancer and leukaemia cell lines. The data presented in Chapter 5 are broadly consistent with these previous findings but weak expression in foetal liver, bone marrow and trachea were shown in addition to testis and CNS tissues, questioning how tightly restricted the expression of this novel CT gene is (see Figure 5.16). However, it remained a promising candidate as it was expressed in a colon cancer sample, one melanoma sample, both lung cancer samples and more strongly in ovarian cancer samples. In addition to this, expression of this gene displayed a significant upregulation in gene expression when ovarian cancer *microarray* datasets were meta-analysed (Feichtinger *et al*., 2012).

Thus, it appeared a promising novel CTA candidate but it was unclear if it was present at the protein level. Using an antibody purchased from Abcam® I tested to see if the protein was present in various cancer cell lines, which appeared to be the case (see Figure 5 in Appendix). I went on to test the specificity of this antibody by using siRNA knockdown targeting this gene. As shown in Figure 4.1, four different siRNAs reduced the intensity of the band detected on a Western Blot, indicating that the antibody was detecting the C20orf201 protein. There was a further band detected at ~50kDa, which was not affected by the siRNA treatment.

To look into the presence of C20orf201, I then used normal testis as a positive control for detection using immunohistochemistry (IHC). Although, the function of this protein is unknown, it was predicted to play a role during spermatogenesis and shown to be expressed strongly in the testis. Moreover, a requirement of its classification as a CTA would require presence of the protein in the normal testis. This initial data revealed positive staining in the testis, and although the staining pattern varied depending on the experimental variables and tissue section used, stronger staining was present in the spermatogonial layer of the testis and appeared predominantly nuclear (see Figure 4.2 and Figure 6 in Appendix). Positive staining was also revealed using immunofluorescence (IF) in the testis using a different antibody (see Figure 7 in Appendix) but the staining was not localised to the spermatogonial layer of the seminiferous tubules but more dispersed throughout the tissue section and developing germ cells. MAGEA1 is a spermatogonial protein with very tight expression limited to basal layer of the testis in healthy tissues and was used as a control to co-stain in the IF experiments.

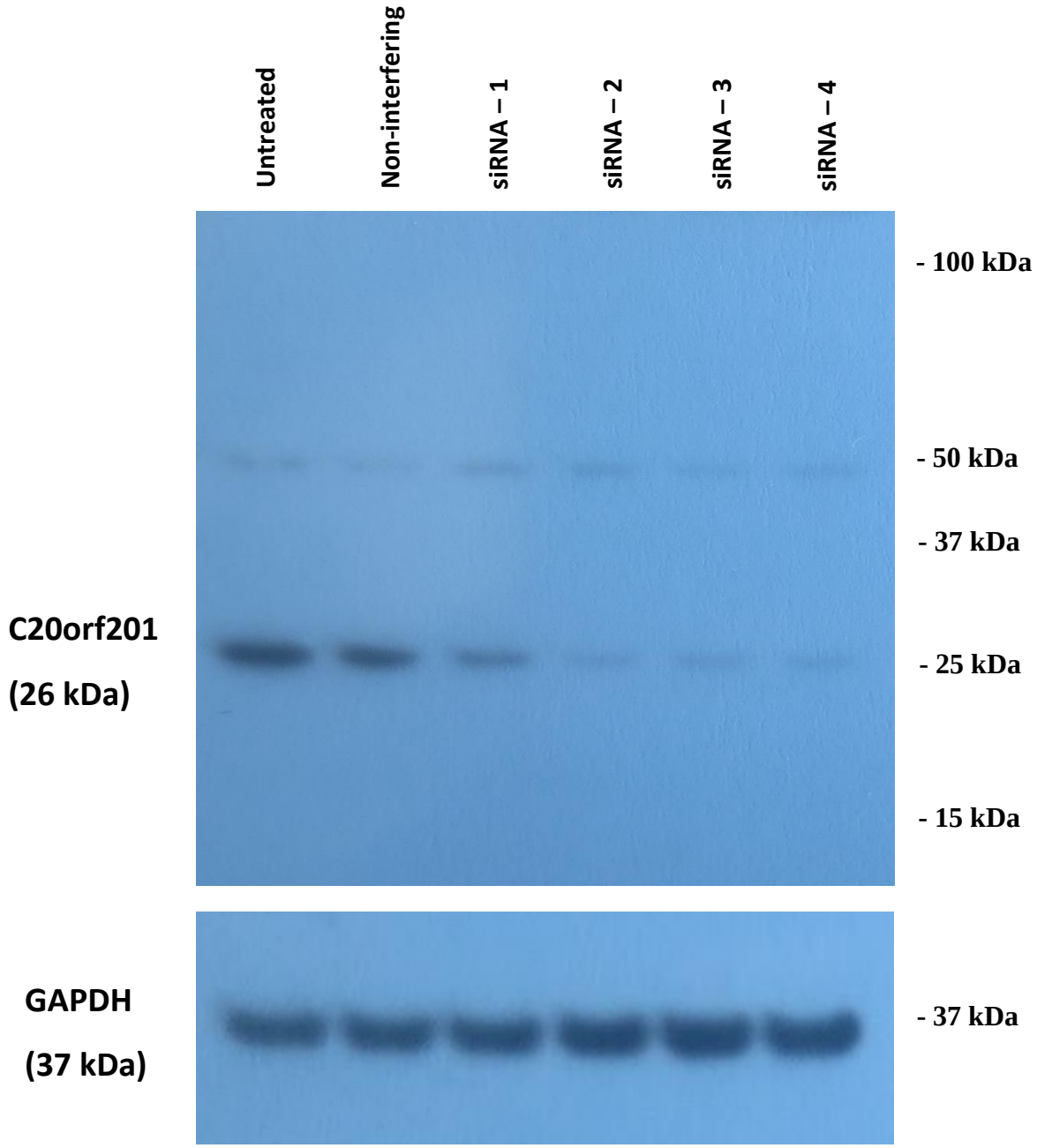


Figure 4.1 Western Blot indicating the presence of C20orf201 in A2780 cells following siRNA treatment.

The expected size was 26kDa and a clear reduction in band intensity was seen using four different siRNA treatments that were carried-out over three days before conducting cell lysis. ab108142 Abcam antibody (Lot. GR191628-1; 1:1000 dilution) for detection of C20orf201. Secondary antibody: anti-rabbit IgG, HRP-linked (Cell Signaling, #7074; 1:3000 dilution). A band is also visible at around 50 kDa but this did not respond to siRNA knockdown. GAPDH is shown as a loading control.

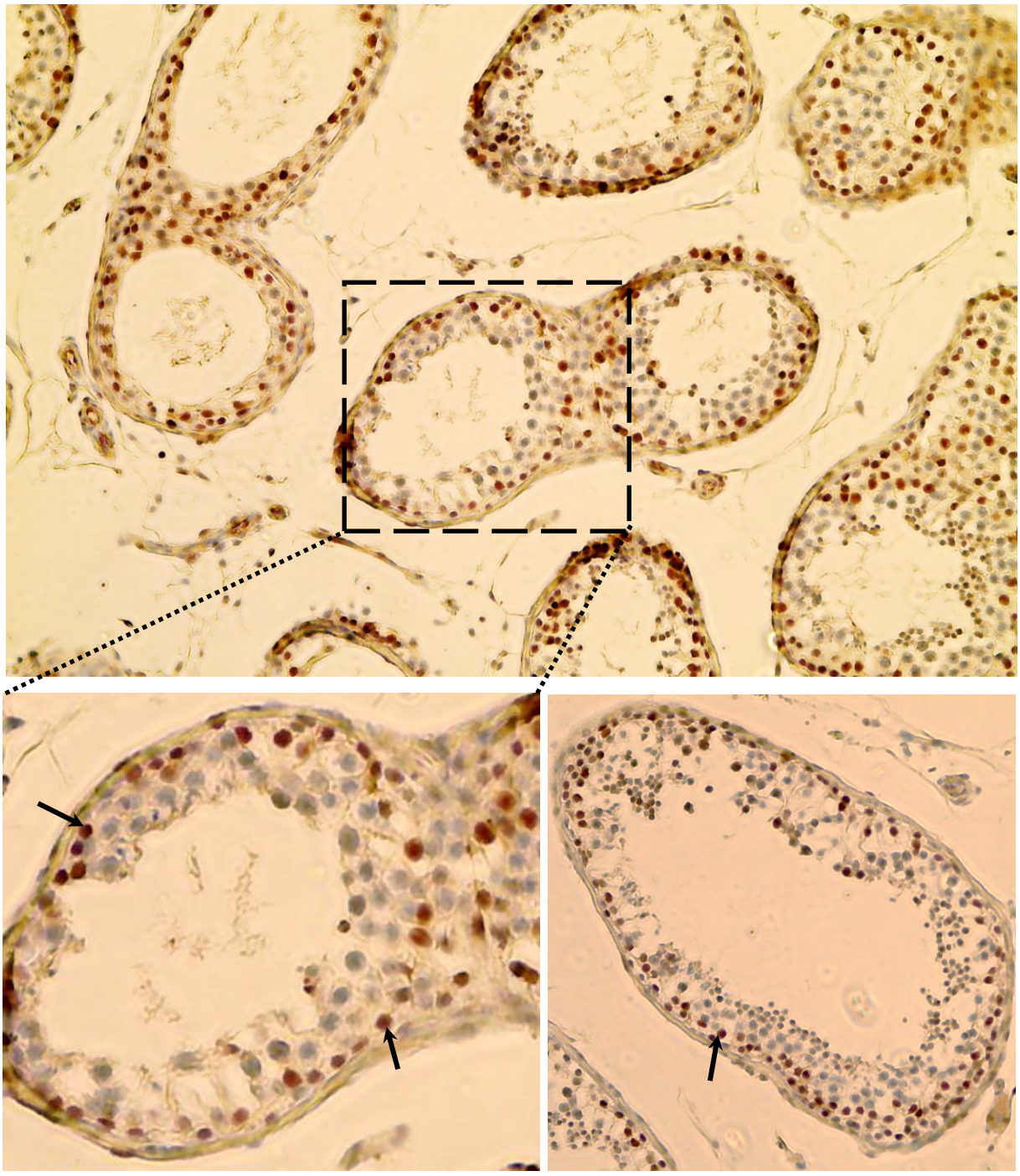


Figure 4.2. IHC staining of C20orf201 in normal testis.

IHC was performed on the semi-automated Ventana machine (Roche) using the rabbit polyclonal antibody to C20orf201 (ab108142, Abcam). Positive staining in these examples was stronger in the basal or spermatogonial compartment of the seminiferous tubules and appeared nuclear (indicated by arrows). An example of a different tissue section is shown in Figure 6 in Appendix. Images acquired on an AMG EVOS® xl Core microscope.

Given that there was an indication from combined sources that *C20orf201* may be expressed strongly in ovarian cancer, I tested for the presence of this protein in ovarian cancer samples. Ten ovarian cancer samples were purchased from ProteoGenex, which were supplied with matched normal surrounding tissue from the same patient. On purchase, Proteogenex indicated that the normal tissue was also ovarian but subsequent analysis of H+E staining of the tissue sections revealed that the majority of the samples contained non-ovarian tissue. The pathological and clinical information supplied by the company is supplied in Appendix C (on CD). The H+E stains for each specimen were assessed by two consultant pathologists. Most of the cancer samples tested displayed clear positive staining for C20orf201, though in general the staining pattern was classified as weak to moderate when assessed by a trained pathologist. A selection of these examples are shown in Figures 4.3 to Figure 4.5. There was weak staining in the normal ovarian tissue that was tested, but there was in general clear differential staining compared to normal adjacent tissue that was negative for the majority of the samples tested. A summary of the staining patterns seen, independently verified by two consultant pathologists (Dr M Atkinson and Dr H Shawki), are provided in Table 4.1.

Next, I went on to test tissue microarray (TMA) slides, which contained a variety of normal tissue. The aim of this was to show the staining pattern was limited within the normal tissue, as was expected from the gene expression analysis that had suggested this gene was a novel CT gene. The staining of the TMA section proved problematic as a large proportion of the tiny cores that were present were displaced by the staining process. I attempted staining the TMAs by hand, again with limited success and could not reproduce the positive staining pattern previously obtained using the Roche Ventana® automated machine in the testis sections.

I obtained TMAs from three separate sources and the best results were obtained using the TMA supplied by the Co-operative Human Tissue Network (CHTN). The results indicated that there was positive staining using this antibody against some normal tissues, including colon, as well as breast cancer (see Table 4.2). Some examples of the staining patterns obtained on the TMA is provided in the Appendix (see Section 3, Figure 8). Separate staining of the colon tissue that I had collected further corroborated this finding of positive staining within normal colon. There was no indication from the gene expression analysis previously conducted that *C20orf201* was expressed in normal colon (see Figure 4.18).

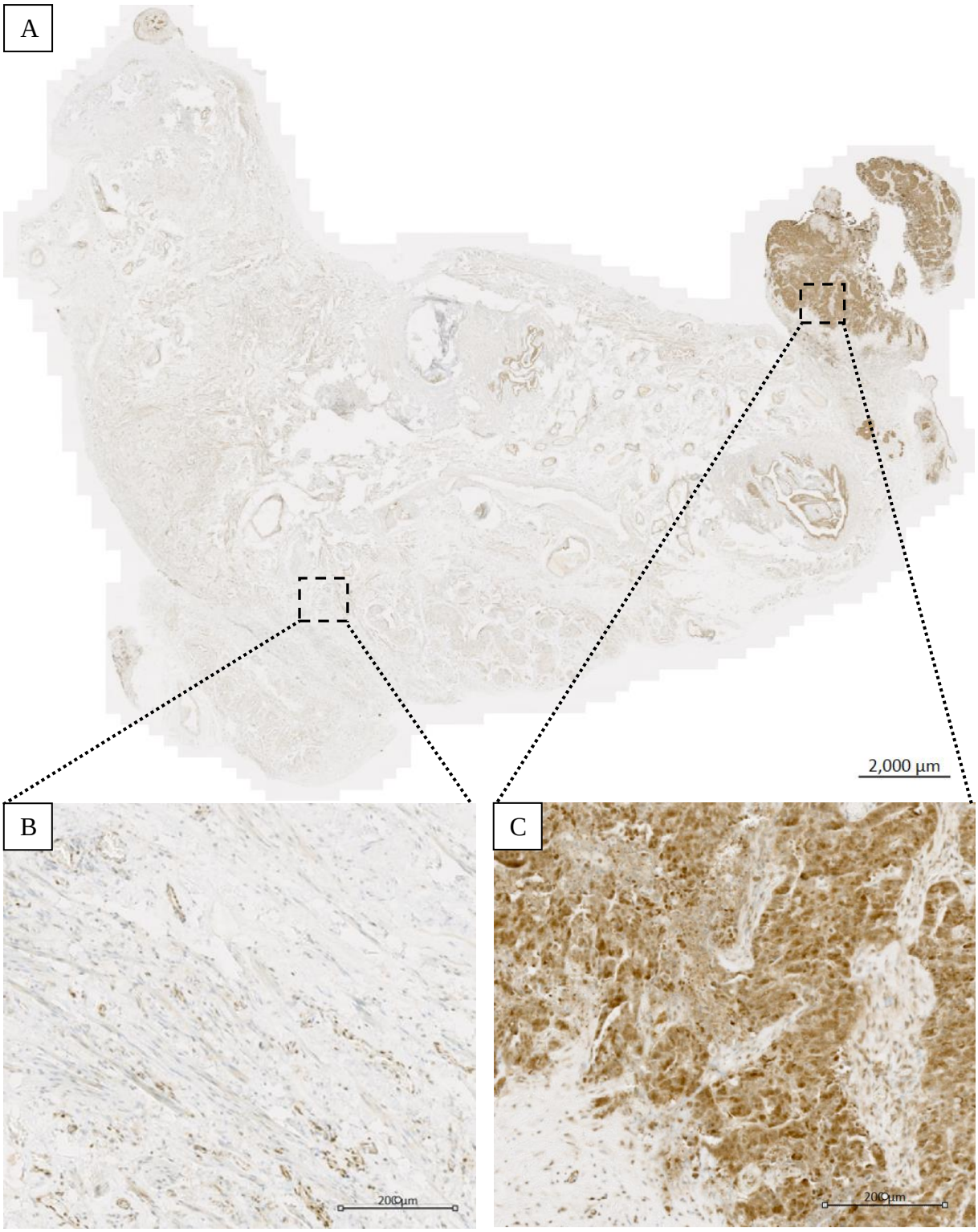


Figure 4.3. C20orf201 staining in ovarian carcinoma and adjacent normal tissue.

(A) Ovarian sample 032036 (see Table 4.1) contained a small focus of serous ovarian carcinoma and is shown here to depict the staining pattern seen in the cancer (top right of image A) compared to the normal para-tubal tissue surrounding the cancer. Staining for C20orf201 was following heat retrieval, using Abcam antibody (ab108142). A higher power magnification of the paratubal tissue (B) and a higher power magnification of the carcinoma is shown in (C). It can be seen that there is stronger, predominantly nuclear staining in the section containing cancer. Images obtained on a Zeiss Axio Scan.Z1 digital slide scanner.

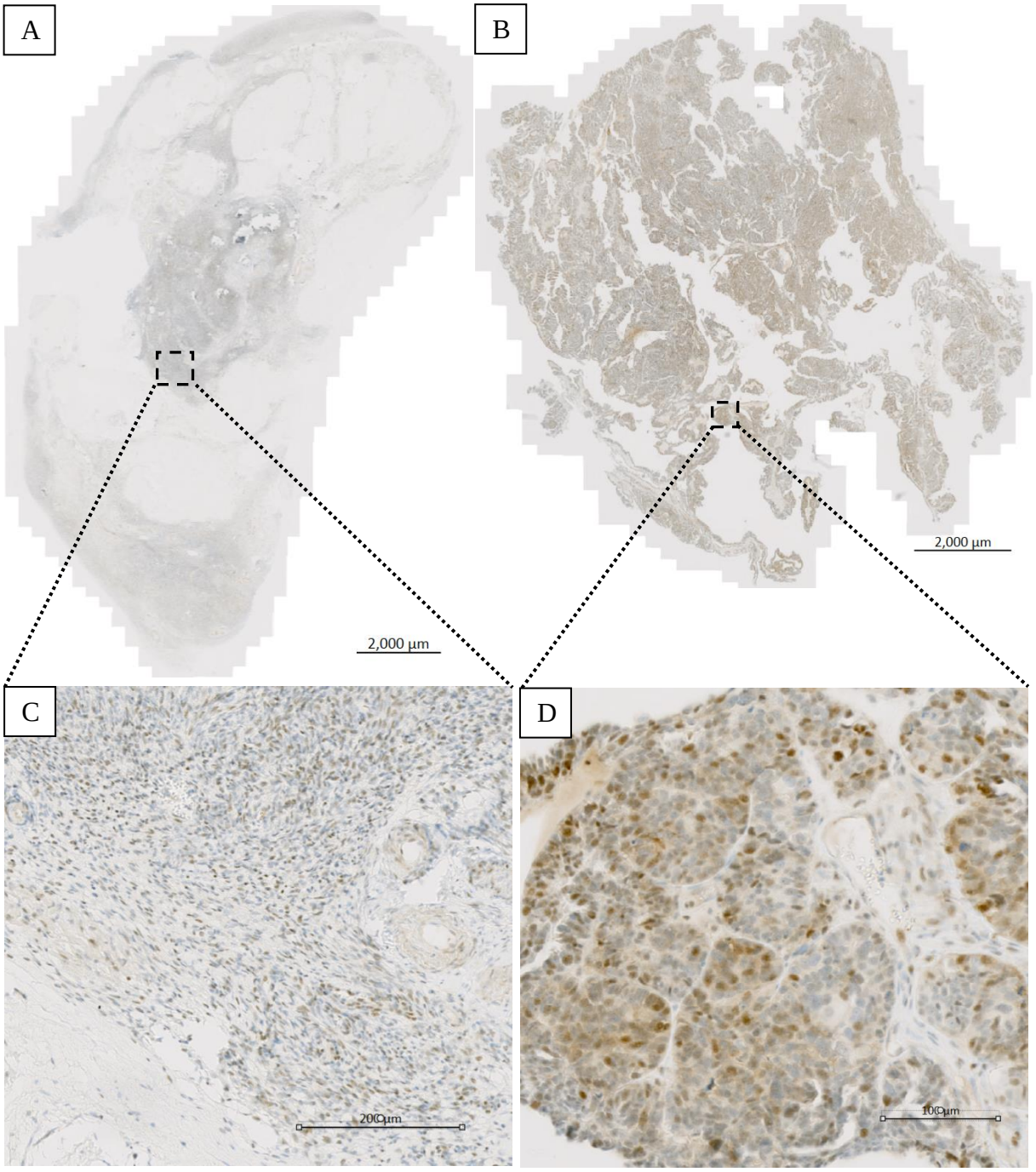


Figure 4.4. C20orf201 staining in normal ovary and ovarian cancer tissue.

(A) Shows negative or very weak staining of C20orf201 in normal ovarian tissue (sample 031890T1 in Table 4.1). The staining pattern seen in the serous ovarian carcinoma from the same patient is shown in (B) – sample 031890T2 in Table 4.1. Very weak nuclear staining can be seen on higher magnification in the normal ovary shown in higher magnification (C) but this was considered negative by the pathologist who looked at this independently. Weak cytoplasmic staining in the high grade ovarian cancer and stronger nuclear staining seen in the carcinoma, shown in higher magnification in (D). Staining following heat retrieval for C20orf201 (Abcam, AB108142). Images obtained on a Zeiss Axio Scan.Z1 digital slide scanner.

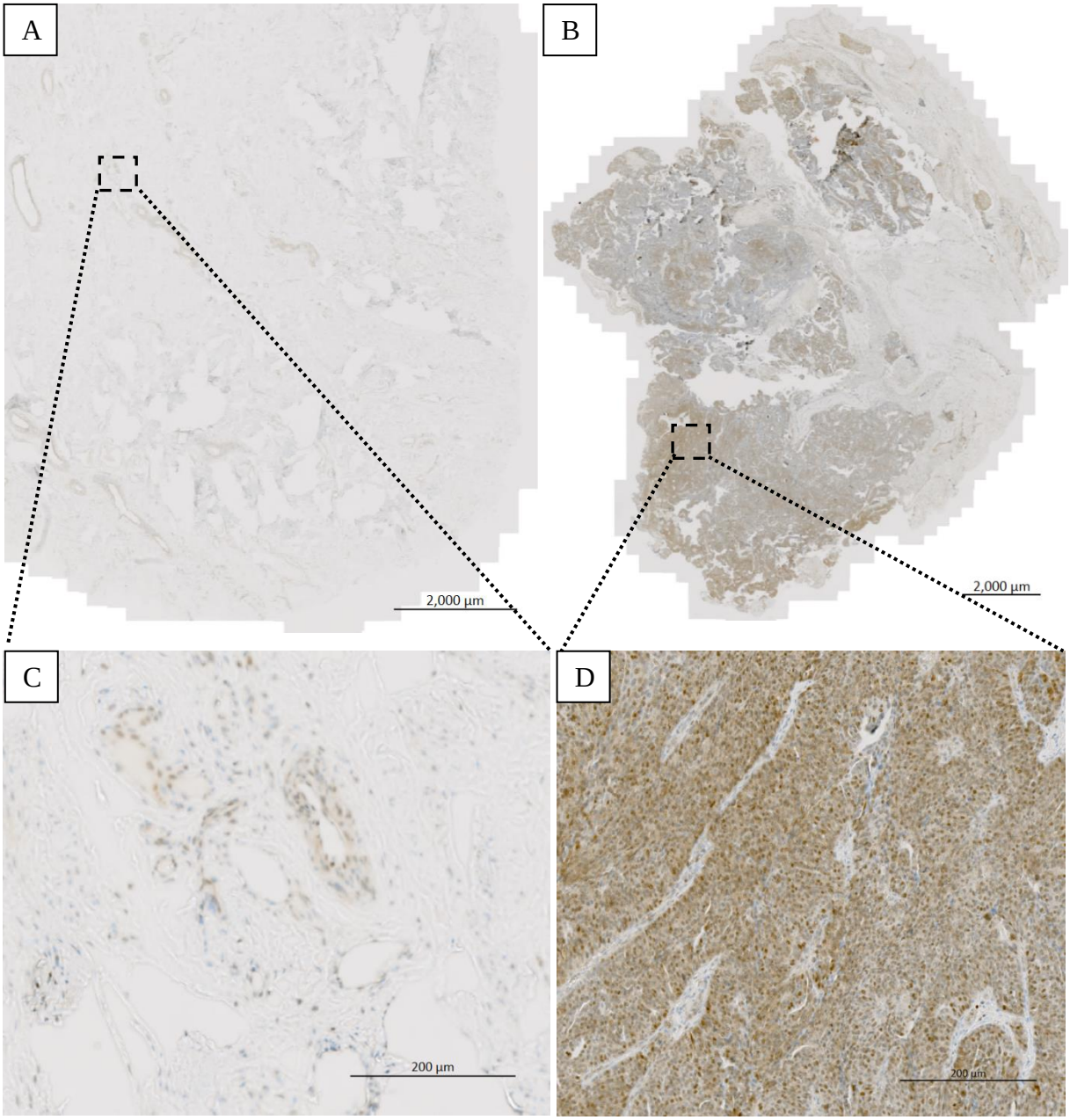


Figure 4.5. C20orf201 staining in ovarian carcinoma and matched normal adjacent tissue.

Here the staining of C20orf201 in a high-grade serous carcinoma of the ovary (B) is shown together with normal surrounding connective tissue (see samples 031603T2 and 031603T1 in Table 4.1). Staining was with the rabbit polyclonal antibody raised against C20orf201 (Abcam, ab108142). Negative staining was seen in the surrounding connective tissue, shown in higher power magnification in (C). Weak cytoplasmic and patchy/focal moderate nuclear staining seen in the ovarian cancer, shown in higher magnification in (D). Images obtained on a Zeiss Axio Scan.Z1 digital slide scanner.

Table 4.1. C20orf201 staining in serous ovarian cancer and matched normal tissue.

Sample description (sample code*)	Staining pattern	Matched normal description (code*)	Staining pattern
Adenocarcinoma (possibly high grade) (Sample: 032050T2)	Focal weak nuclear staining	Normal cervix (Sample: 032050T1)	Weak to moderate nuclear staining in epithelial cells
High grade serous carcinoma (Sample: 032036T2)	Moderate nuclear and cytoplasmic staining in cancer	Paratubal tissue and small focus of ovarian carcinoma (Sample: 032036T1)	Weak to moderate predominantly nuclear staining in carcinoma only
Serous carcinoma (possibly low grade) (Sample: 031940T2)	Moderate but variable nuclear and cytoplasmic staining	Normal cervix (Sample: 031940T1)	Weak to moderate nuclear staining in epithelium
High grade serous carcinoma (Sample: 031985T2)	Moderate nuclear and weak cytoplasmic staining	Myometrium and endometrium (Sample: 031985T1)	Negative
High grade serous carcinoma (Sample: 031890T2)	Focal moderate positive nuclear and weak cytoplasmic staining	Normal ovary (Sample: 031890T1)	Negative
High grade serous carcinoma (Sample: 031827T2)	Weak cytoplasmic and focal moderate nuclear staining	Smooth muscle and connective tissue (Sample: 031827T1)	Focal weak staining in some endothelial cells
High grade serous carcinoma (Sample: 031772T2)	Very sparse and weak nuclear staining	Normal cervix (Sample: 031772T1)	Patchy positive nuclear staining
High grade serous carcinoma (Sample: 031722T2)	Moderate nuclear staining	Myometrium (Sample: 031722T1)	Very weak focal nuclear staining
Serous carcinoma (possibly high grade) (Sample: 031666T2)	Weak nuclear and moderate cytoplasmic staining	Normal ovary (Sample: 031666T1)	Very weak scanty nuclear staining in stromal cells
Serous carcinoma (Sample: 031603T2)	Moderate nuclear and cytoplasmic	Normal Fallopian tube (Sample: 031603T1)	Weak cytoplasmic and patchy nuclear staining

*Details of the individual samples supplied by Proteogenex are provided in Appendix C (on CD).

Table 4.2. Summary of TMA staining for C20orf201 using IHC

Tissue	Staining Pattern*
Normal Prostate	Scanty nuclear staining in myoepithelial cells. Negative in epithelium
Normal Liver	Negative
Smooth Muscle	Negative
Breast Cancer	Patchy (<30%) positive nuclear staining
Breast Cancer	Patchy (15-20%) positive nuclear staining
Smooth Muscle	Negative
Normal Liver	Negative
Normal Prostate	Very weak and scanty nuclear staining (consider negative)
Normal Colon	Positive predominantly cytoplasmic staining of glandular epithelium
Normal Colon	Cytoplasmic staining in colonic crypt. Weak staining of lymphocytes but negative central region of Peyer's patch
Normal Prostate	Scanty nuclear staining in myoepithelial cells only.
Normal Liver	Negative
Smooth Muscle	Negative
Breast cancer	Very weak and scanty nuclear staining
Breast cancer	Nuclear staining in 30-50% of cancerous tissue
Smooth Muscle	Negative
Normal Liver	Negative
Normal Prostate	Negative
Normal Colon	Very weak cytoplasmic staining.
Normal Spleen	Scanty nuclear staining in lymphocytes
Normal Spleen	Very occasional positive staining of lymphocytes
Normal Spleen	Negative
Normal Spleen	Negative
Normal Spleen	Very occasional weak positive
Normal Colon	Very weak surface and cytoplasmic staining in glandular cells
Normal Prostate	Very weak and scanty nuclear staining in stroma (epithelium negative)
Normal Liver	Negative
Myometrium	Negative
Normal Prostate	Negative
Normal Colon	Very weak cytoplasmic staining in glandular epithelium
Breast Cancer	Patchy (~10%) focal nuclear staining
Breast Cancer	Patchy (5-10%) nuclear staining
Normal Colon	Weak cytoplasmic staining in glandular epithelium and weak nuclear staining in stroma
Normal Spleen	Occasional positive staining in some lymphocytes
Normal Spleen	Negative

*Example IHC images are provided in the Appendix (see Figure 8). The IHC findings were confirmed with a consultant pathologist (Dr Mark Atkinson).

4.3. TEX19 localisation in normal testis and colonic tissue

Our group previously identified *TEX19* as a novel mei-CT gene with expression seen in the majority of cancer samples tested but restricted to the testis and thymus in normal tissues (Feichtinger *et al*., 2012b). *TEX19* appeared to be a promising candidate on my TLDA analysis also, with expression restricted to the testis but also seen in colon and lung cancer (see Figure 4.18). Little is known about the function or localisation of TEX19 in humans, though it is known to be expressed in the germline as well as pluripotent stem cells (Kuntz *et al*., 2008). Deletion of *TEX19* resulted in impaired or absent spermatogenesis in mice (Ollinger *et al*., 2008).

Argonaute proteins interact with the major sub-classes of small RNAs (Watanabe and Lin, 2014). In the germline members of the PIWI group of argonaute proteins interact with PIWI-interacting RNA (piRNA). Together these complexes functionally interact to repress the expression of retroelements via epigenetic silencing (Crichton *et al*., 2014). Although TEX19 has not been linked directly to the PIWI-piRNA pathway (Crichton *et al*., 2014) it remains possible that there is an interaction between these proteins. Deletion of *TEX19* resulted in activation of retroelements and it was thought that *TEX19* may function to maintain genomic stability during sperm production and development (Ollinger *et al*., 2008). *PIWIL1* is a known germline gene: it also was shown to be a potential biomarker in colon cancer on the TLDA analysis. Of the four colon cancer samples that expressed *TEX19*, three also expressed *PIWIL1* at moderate to high levels (see Figures 5.18 and Figure 5.22). Both genes had limited expression in normal tissue samples, though *PIWIL1* expression was seen more commonly than *TEX19*. The expression of *PIWIL1* was limited to normal colon tissue that had been obtained adjacent to colon cancer, so it is possible that *PIWIL1* expression was influenced in these cells by the neighbouring tumour.

In order to investigate the localisation of TEX19 within the testis and possible association with PIWI proteins, IF experiments were performed – see Figures 4.6-4.8. First of all TEX19 was co-stained with MAGEA1 that is known to localise to the spermatogonial layer of the testis and the monocoloncal antibody produced a very clean and consistent signal, thus acting as a positive control. As shown in Figure 4.6, TEX19 associated with MAGEA1 in and close to the spermatogonial compartment but the two proteins did not exactly co-localise. TEX19 appeared to localise around the stem cell and developing spermatocytes with a tapering of less intense

staining towards the adluminal compartment and region where post-meiotic cells would reside. Thus, staining for TEX19 appeared to be in an ill-defined matrix between the spermatogonial cells or perhaps more likely within the cytoplasm of the Sertoli cells that envelope developing spermatocytes.

As can be seen in Figures 4.6 and 4.8 (see also, Figure 9 in Section 3 of the Appendix), this staining pattern for TEX19 was consistent. In some cases there was weak staining seen in the cytoplasm of developing spermatocytes in the adluminal compartment (e.g., Figure 4.7) but in general the most consistent appearance was that just described above. In Figure 4.7 there were some cells, in the adluminal compartment of the seminiferous tubule, in which TEX19 and PIWIL1 did co-localise. In a second example provided in Appendix (Figure 9) there did not appear to be co-localisation of the proteins and the presence of TEX19 within the cytoplasm of some spermatocytes within the adluminal compartment was not seen. However, the regions of the tubule that expressed the proteins more strongly were the same, suggesting that a functional interaction may still exist. A way of confirming the cellular localisation of TEX19, or any other candidate germline protein, would be to co-stain with other known spermatocyte (and Sertoli) cell markers and observe for any co-localisation.

As can be seen in Figure 4.7, as was the case with MAGEA1, TEX19 also did not co-localise exactly with PIWIL1. PIWIL1 appeared to be present throughout the seminiferous tubules but stronger staining was seen in the adluminal compartment where many of the cells will be undergoing meiosis at various stages of sperm cell development. The staining pattern for PIWIL2 was less consistent but more of this protein appeared to be present in the spermatogonial layer of the seminiferous tubules (see Figure 4.8). An additional example of the staining seen for PIWIL2 in the testis is shown in Figure 10 in Section 3 of the Appendix. The negative controls for these and subsequent IF experiments are shown in the Appendix (Figures 11-12). Co-localisation of the PIWI proteins with each other, or with MAGEA1, could not be performed as all three of the antibodies used were raised in mice.

The presence of TEX19 was also assessed using IHC and a consistent staining pattern was seen in the basal area of the seminiferous tubule. As with the IF experiments, the staining appeared to surround the germ cells close to the spermatogonial compartment and staining was weaker towards the adluminal compartment (see Figure 4.9).

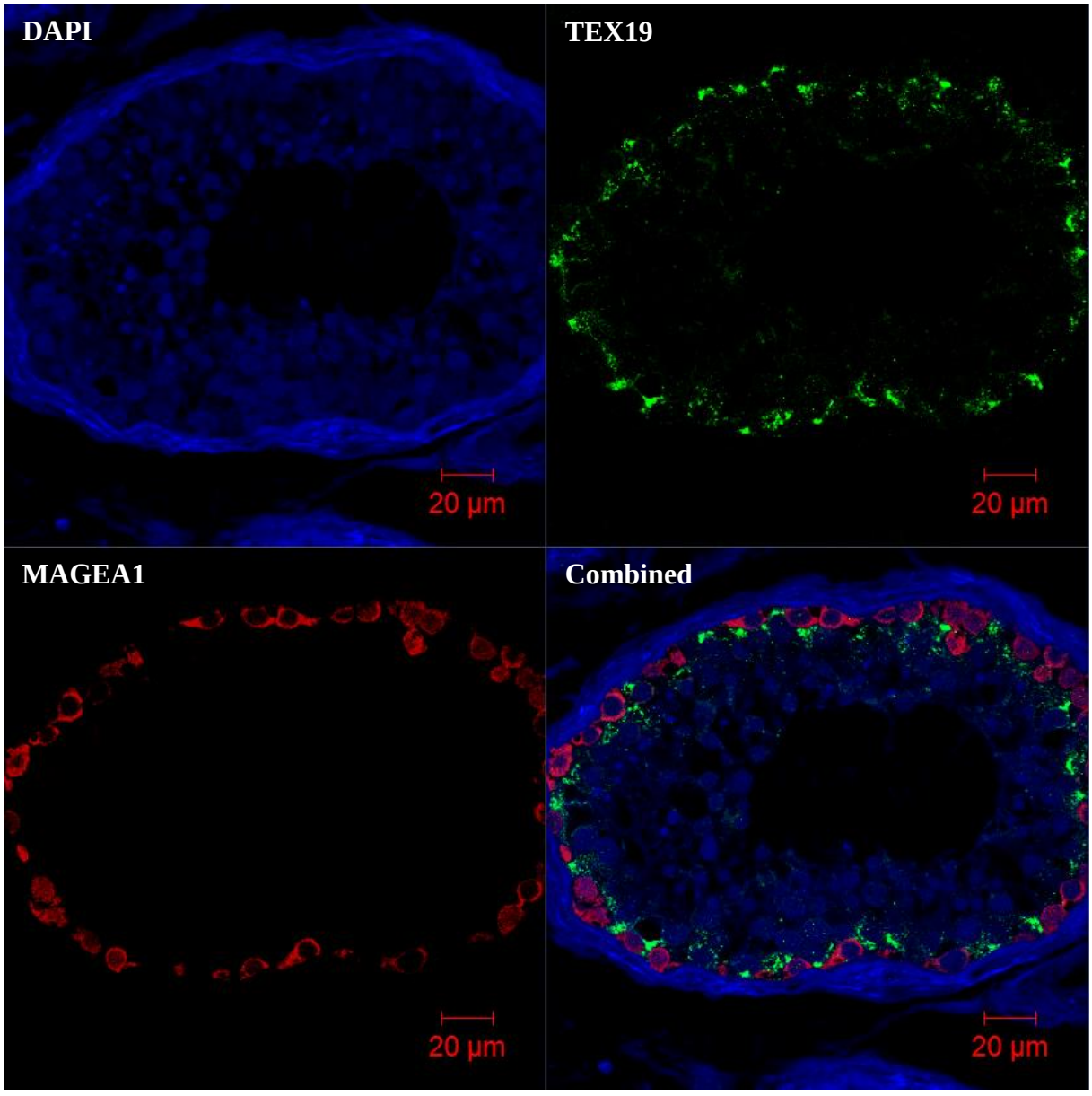


Figure 4.6. IF of TEX19 and MAGEA1 staining in normal testis (x40).

IF was carried out on FFPE tissue blocks of normal testis showing MAGEA1 (red) staining detected by the mouse monoclonal antibody (LSBio, LS-C87868) localised to the spermatogonial layer of the seminiferous tubule. The staining for TEX19 (green) was with the rabbit polyclonal antibody (Abcam, ab185547). Although there was a strong correlation between the regions of the seminiferous tubule that stained positive for the two proteins, both being towards/in the basal or spermatogonial compartment, there was no exact co-localisation. Images acquired on Zeiss LSM 710 confocal microscope.

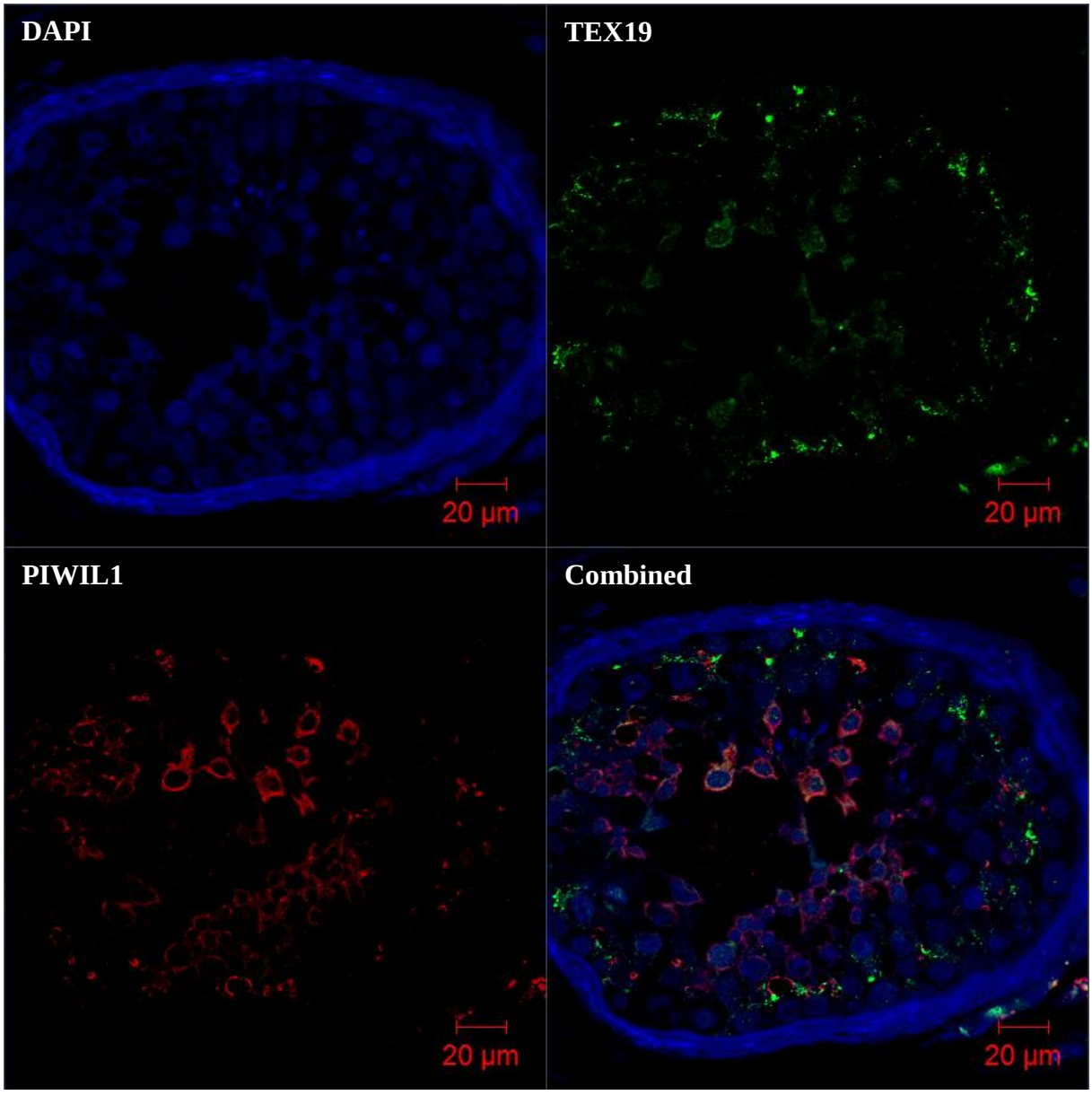


Figure 4.7. TEX19 and PIWIL1 staining in normal testis (x40).

IF was carried out on FFPE tissue blocks of normal testis showing TEX19 (green) staining detected by the rabbit polyclonal antibody (Abcam, ab185547), which localised between developing spermatocytes but towards the basal compartment of the seminiferous tubule. The co-staining for PIWIL1 was with the mouse monoclonal antibody (Sigma, SAB4200365). The staining for PIWIL1 appeared stronger in the adluminal compartment where a higher proportion of cells would be undergoing meiosis and once again there was no exact co-localisation with TEX19, except in the cytoplasm of the occasional germ cell in the adluminal compartment. The cytoplasmic staining of TEX19 within developing germ cells was not consistently reproducible and another example of IF staining for TEX19 and PIWIL1 in normal testis is shown in Figure 9 of Appendix. Images acquired on Zeiss LSM 710 confocal microscope.

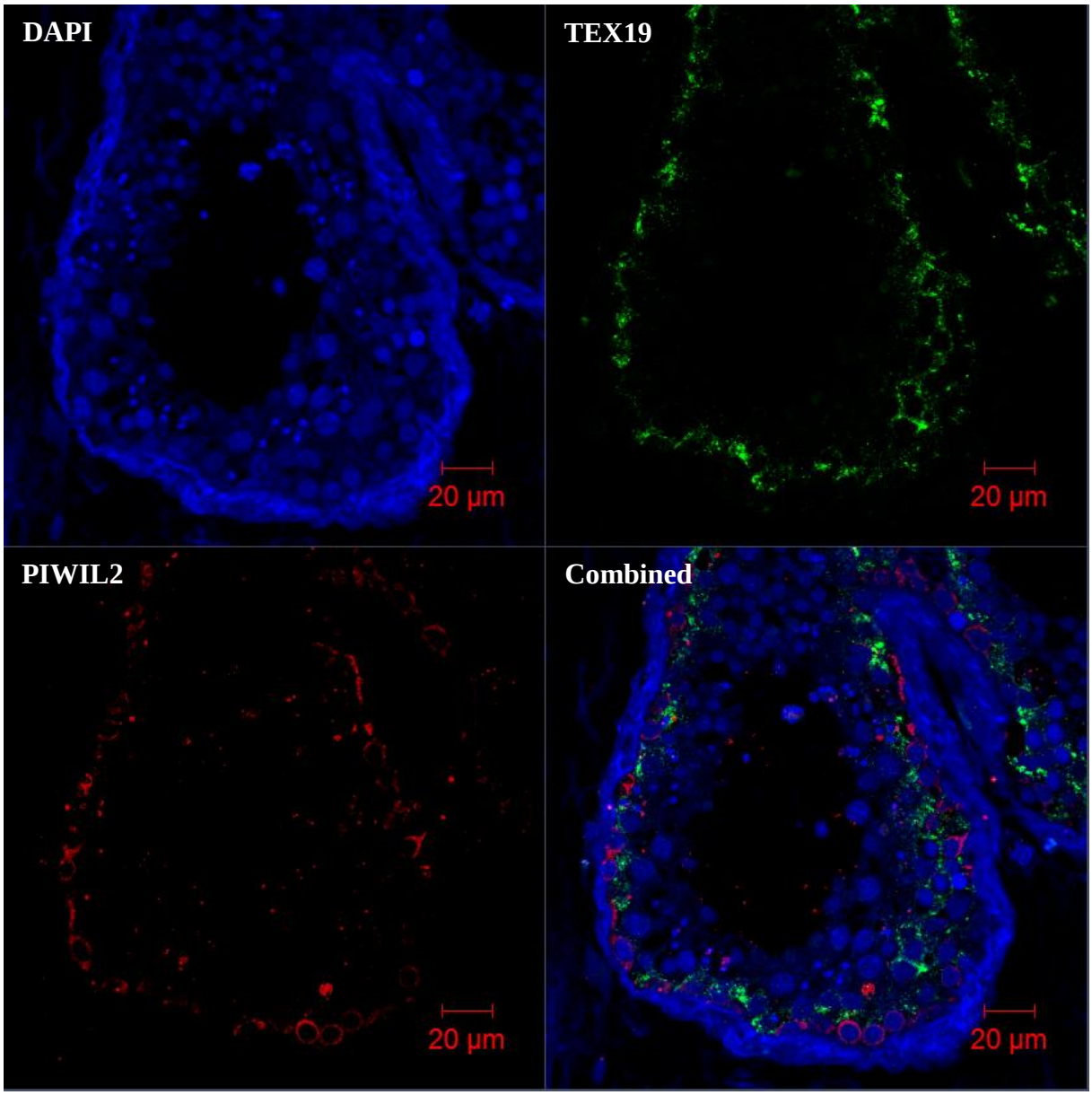


Figure 4.8. TEX19 and PIWIL2 staining in normal testis (x40).

IF was carried out on FFPE tissue blocks of normal testis showing PIWIL2 (red) staining detected by the mouse monoclonal antibody (Abnova, MAB0843). Weak and patchy staining was present throughout the seminiferous tubule and occasional surrounding cells but stronger, predominantly cytoplasmic, staining was present in the spermatogonial layer of the testis. TEX19 (green) appeared to stain a region surrounding the germ cells predominantly towards the spermatogonial layer of the seminiferous tubule. See Figure 10 in Section 3 of the Appendix for a repeat experiment, where the staining in the spermatogonial layer of the seminiferous tubule was more distinct for PIWIL2 but on that occasion co-staining for TEX19 was not performed. Images acquired on Zeiss LSM 710 confocal microscope.

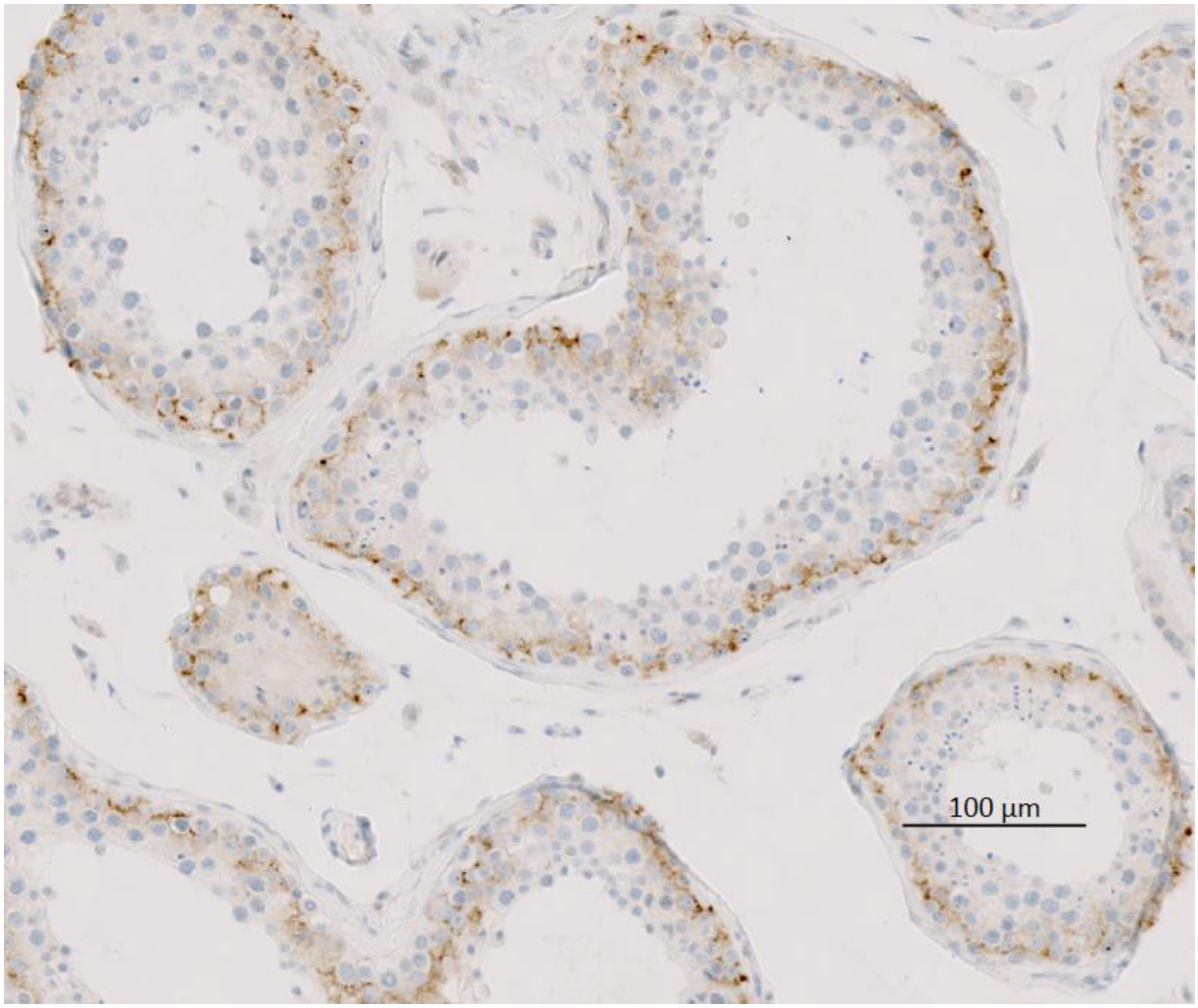


Figure 4.9. IHC staining of TEX19 in normal testis.

IHC was performed on 4 µm tissue sections cut from FFPE tissue blocks of normal testis (Abcam, ab185547). Stronger staining was seen towards the spermatogonial layer of the seminiferous tubules but the staining appeared predominantly between the spermatogonial cells; at least no nuclear positivity was present within the germ cells. The IHC findings were consistent with the IF experiments shown above. Image obtained on a Zeiss Axio Scan.Z1 digital slide scanner.

I went on to test for the presence of the protein in colon cancer samples. Here, a common finding in the normal adjacent tumour (NAT) tissue was limited cytoplasmic staining in the adluminal columnar epithelial cells (for example, see Figure 4.10 and Figure 4.11). Expression was seen more widely in the cancer samples, an example of which is shown in Figure 4.10. In the region of cancer, TEX19 appeared to display nuclear as well as cytoplasmic staining, raising the possibility that the protein shuttles to the nucleus to have differing consequences when present in cancer. The staining pattern was seen with some consistency across the normal colonic mucosa and appeared cytoplasmic and predominantly confined to the epithelial cells immediately adjacent to the bowel lumen. Although the example shown in Figure 4.10 was directly adjacent to a tumour (on the same slide), it was fairly typical of the staining seen in other (though more distant) NAT colonic mucosa tissue sections that I tested. All the normal samples stained were from patients with colonic tumours. A summary of the staining pattern seen is provided in Table 4.3.

The staining pattern was not strongly positive in any of the colon cancer samples, though there was a trend to stronger staining seen compared to the NAT tissue. This was particularly true with regards to the nuclear staining. One cancer sample (C40) had an area of mucinous carcinoma that displayed stronger cytoplasmic staining for TEX19 but the adjacent normal colonic mucosa also displayed areas of moderate staining adjacent to the lumen (see Table 4.3). Also, sample 45 which was a poorly differentiated tumour displayed widespread nuclear positivity that was not seen in the adjacent normal tissue. Both these tumours were also found to express *TEX19* (see Figure 5.16) and suggest a possible functional relevance in these tumours that carry a poor prognosis. However, somewhat confusingly two other tumours that displayed stronger differential staining in the cancer (samples 26 and 35) did not clearly express *TEX19* (see again Figure 5.16). This could be accounted for by the heterogeneity of expression of the gene within individual tumour samples. The weak cytoplasmic staining seen in the majority of the NAT colonic samples tested is not entirely inconsistent as very weak/borderline expression was seen in some of the NAT colon samples tested on the TLDA cards, though this generally fell below the threshold I had set for determining definite gene expression. The localisation of TEX19 within the adluminal cells is of uncertain significance and it is possible that this also represents an artefact or ‘edge effect’ of the IHC procedure. It should be remembered that all the normal colon samples stained were from patients with colonic tumours, so in certain respects they may not be representative of normal colonic mucosa in patients without cancer.

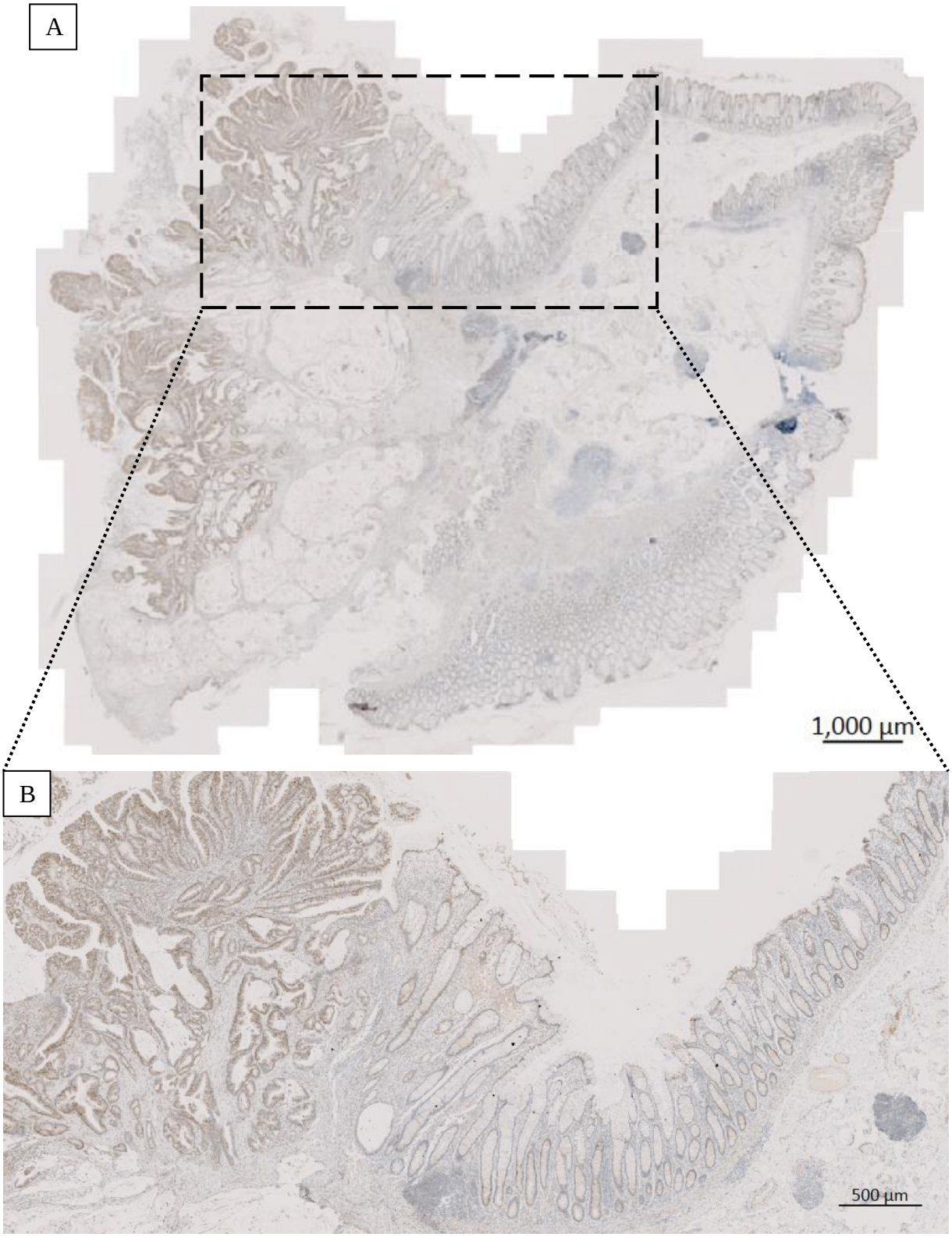


Figure 4.10. IHC staining of TEX19 on colon cancer.

IHC was performed on 4 µm tissue sections of sample C26. (A) Displays the entire tissue section and (B) shows a higher-power magnification of the transition from cancer to normal epithelium, which becomes progressively more normal from a left to right direction. It is clear from (A) that stronger staining was present in the portion of the tissue section containing cancer. Images obtained on a Zeiss Axio Scan.Z1 digital slide scanner.

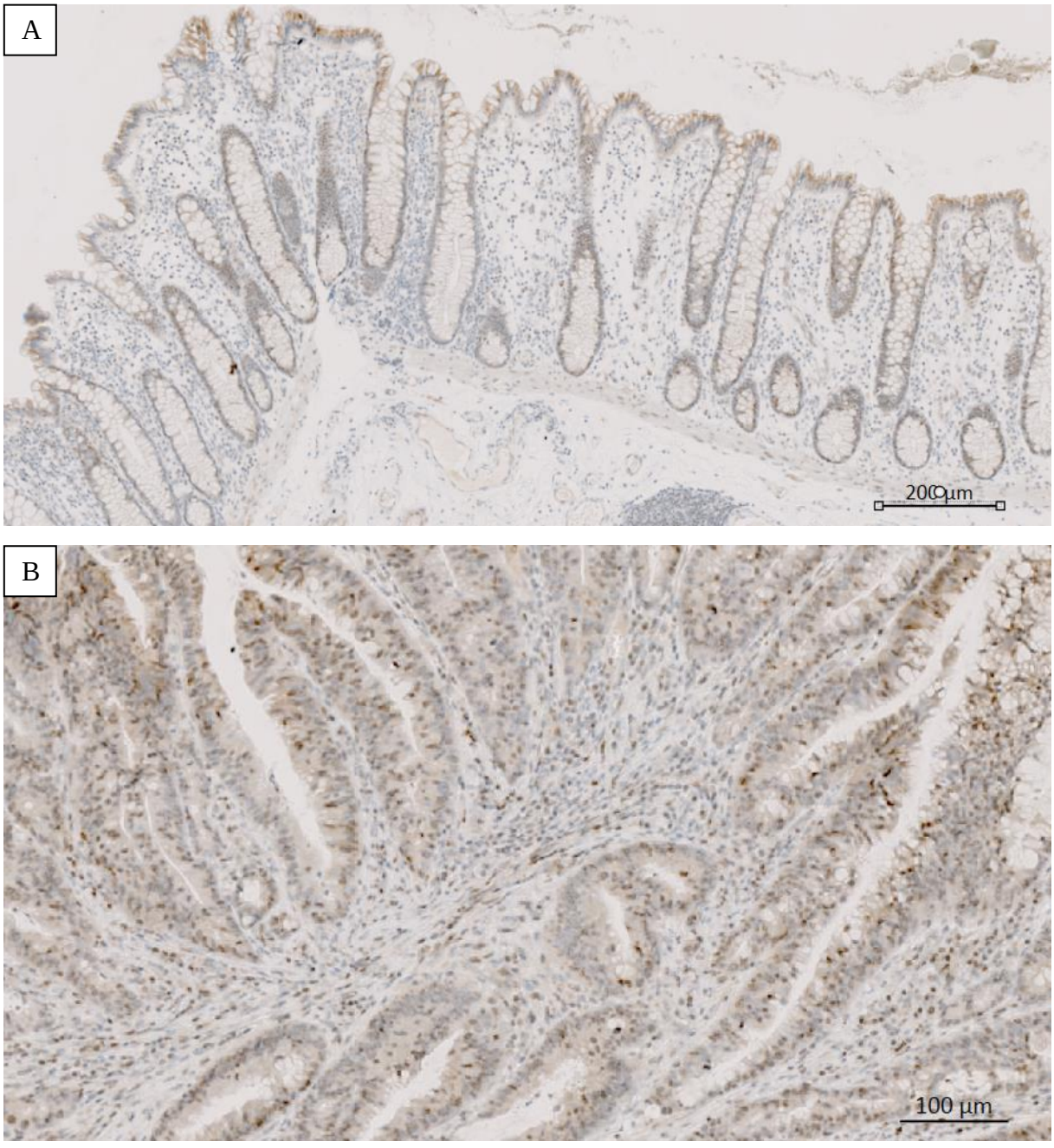


Figure 4.11. TEX19 staining in colon cancer and adjacent normal mucosa assessed by IHC.

Higher power magnification of the same sample shown in Figure 4.10. (A) Shows a histologically normal area displaying weak positive cytoplasmic staining in the glandular cells closest to the luminal surface of colon. (B) Higher power magnification of the area of cancer: here again cytoplasmic staining is seen but there is also weak to moderate nuclear staining of the cancer cells. Nuclear staining was present in the abnormal epithelial cells but also the stromal cells of the tumour, neither of which were seen in the normal areas of colon adjacent to the cancer. Images obtained on a Zeiss Axio Scan.Z1 digital slide scanner.

Table 4.3. Summary of the IHC staining results for TEX19 in colon cancer and adjacent normal tissue.

Sample	Tumour staining pattern	Normal adjacent tissue staining pattern
C7*	Weak cytoplasmic and patchy weak nuclear staining	Weak cytoplasmic and nuclear staining in glandular epithelium
C9	Weak cytoplasmic staining	Weak cytoplasmic in adluminal epithelium
C12*	Weak to moderate focal nuclear and weak cytoplasmic staining	Scanty weak nuclear staining in epithelial and stromal cells; weak to moderate cytoplasmic staining in adluminal epithelium
C22	Weak focal nuclear staining	Weak cytoplasmic staining in adluminal epithelium; scanty focal mucosal nuclear
C26	Moderate cytoplasmic and moderate focal nuclear staining in adenoma and focal nuclear staining in stroma	Moderate cytoplasmic staining in adluminal epithelium and patchy weak nuclear staining
C27	Scanty moderate nuclear staining and weak cytoplasmic staining	Weakly cytoplasmic positive adluminal epithelial cells
C31*	Scanty weak predominantly nuclear staining	Very weak and patchy nuclear and cytoplasmic adluminal glandular staining
C32*	Patchy moderate nuclear staining and weak cytoplasmic staining	Very weak cytoplasmic adluminal epithelial staining
C33	Weak cytoplasmic staining with patchy areas of both weak/moderate nuclear and cytoplasmic staining	Weak cytoplasmic staining in adluminal epithelium and scanty focal nuclear positivity
C35	Moderate nuclear and weak cytoplasmic staining	Weak adluminal cytoplasmic epithelial staining and very weak focal nuclear staining in stroma
C36*	Weak and patchy nuclear and cytoplasmic staining	Weak to moderate cytoplasmic staining in adluminal epithelium
C40	Weak nuclear and cytoplasmic staining with focal areas of moderate cytoplasmic staining in area of tumour with mucinous component	Weak to moderate cytoplasmic staining in adluminal epithelium
C45	Weak to moderate nuclear staining and weak cytoplasmic staining	Weak to moderate cytoplasmic staining in adluminal epithelium

Sample names in bold denote cancer samples for which the TLDA analysis indicated that *TEX19* was expressed. *TEX19 gene expression not assessed on the TLDA cards for these samples. Please refer to Table 5.2 for details relating to the cancer from which the samples were taken.

To further validate the presence of TEX19 in normal tissues, I also conducted IHC using the same experimental conditions, on a TMA containing numerous types of normal tissues. As was the case with my previous attempts at staining TMA slides, there was some loss of the cores during the process. However, the majority of the cores were intact and most of these displayed negative staining for TEX19. A summary of the positive staining seen is provided in Table 4.4 and example images are displayed in the Appendix (see Figure 13 in Section 3). Consistent with my findings in the NAT colon samples that I had collected, there was similar staining in the cytoplasm of the intestinal epithelial cells present on the array. Thus, the possibility of the staining being the result of the mucosa being close to a tumour is very unlikely.

Table 4.4. Summary of positive IHC staining patterns seen for TEX19 on a normal TMA

Tissue type	Positive staining pattern description
Amniotic membrane	Weak cytoplasmic
Aorta (smooth muscle)	Very weak and scanty cytoplasmic
Appendix (MALT)	Weak cytoplasmic staining in cells adjacent to lumen
Bladder	Very weak and scanty cytoplasmic staining (adluminal)
Colon	Weak cytoplasmic staining (adluminal)
Endometrium (secretory)	Very weak cytoplasmic staining in glandular cells
Eosophagus	Very weak and scanty cytoplasmic staining
Epididymis	Weak to moderate adluminal cytoplasmic positivity
Fallopian Tube	Very weak glandular cytoplasmic staining
Gallbladder	Weak cytoplasmic staining
Intestinal smooth muscle	Very weak cytoplasmic staining
Kidney (cortex)	Very weak cytoplasmic
Lung (bronchial epithelium)	Very weak and scanty cytoplasmic staining
Pancreas	Focal weak nuclear and cytoplasmic staining
Placenta	Very weak and scanty cytoplasmic positivity
Seminal vesicle	Very scanty cytoplasmic staining
Small Intestine	Weak to moderate cytoplasmic in adluminal epithelium
Spleen	Very weak and scanty cytoplasmic staining
Stomach (oxyntic mucosa)	Weak cytoplasmic staining
Testis (seminiferous tubules)	Moderate cytoplasmic in basal compartment

MALT – mucosa associated lymphoid tissue. Other tissue types included on the TMA, for which TEX19 staining was considered negative, included: adrenal gland, breast, heart, parathyroid gland, cervix, ovary, prostate, liver, lymph node, thymus, mesothelium, brain, peripheral nerve, salivary gland, tonsil, skin (squamous epithelium), synovium, skeletal muscle, uterine smooth muscle, kidney (medulla). There was significant loss of tissue limiting the interpretation for: breast, oesophagus, ovary, prostate, skin and synovium; in addition, there was complete loss of all 3 cores of articular cartilage.

4.4. SPO11 and PIWIL1 localisation in testis and colonic tissue

DSBs are generated in meiosis by the topoisomerase II-like enzyme SPO11. SPO11 is a highly conserved protein without known physiological roles outside of meiosis/meiotic recombination; it is thus widely believed to have a testis/germline restricted expression profile (de Massy, 2013). The TLDA analysis revealed that *SPO11* was only expressed in the testis, with no expression seen in any other normal or cancer samples tested using two different assays. However, SPO11 is considered a CTA and has been described as such by at least two groups (Almeida *et al*., 2009; Litvinov *et al*., 2014). The specificity of *SPO11* to the testis has been called into question by others (Hofmann *et al*., 2008). We also remained interested in this gene given its important physiological roles during meiosis and work had already been conducted by other group members to validate a polyclonal antibody that detected SPO11. Thus, the presence of SPO11 was investigated as an example of a known CTA but also to reveal its localisation within the testis and possible presence in cancerous tissue.

The antibody for SPO11 produced a very strong signal and proved difficult to optimise using IHC. Thus, IF was used as the main technique to establish where the protein could be detected in normal testis. As shown in Figure 4.12 and Figure 4.13, SPO11 appeared to be present in a large proportion of developing spermatocytes consistent with its known physiological roles during meiosis. Co-staining with MAGEA1 is displayed in Figure 4.12 and although there appeared to be some focal positivity in the spermatogonial compartment, SPO11 staining was strong throughout the germ cells within the seminiferous tubules. In Figure 4.13, co-localisation with PIWIL1 is shown – it can be seen that many of the same cells tested positive but SPO11 appeared to have predominantly nuclear localisation, whereas PIWIL1 was largely cytoplasmic. PIWIL1 staining using IHC in normal testis is shown in Figure 4.17, which again showed stronger staining in the adluminal compartment of the seminiferous tubule. Nuclear localisation of SPO11 in developing spermatocytes would be expected given the known function of SPO11 in initiation of meiotic recombination by causing DNA DSBs. In the colon specimens tested, there appeared to be very weak staining in the normal adjacent cancer specimens (see Figure 4.14 and 4.16). In Figure 4.14, nuclear localisation was limited to a specific cell within the colonic crypt – this was at the interface between the stem cell and progenitor zone. In the cancer samples tested there was stronger staining but this was non-specific. Similarly for PIWIL1 there was positive but non-specific staining and no clear co-localisation of these proteins (see Figure 4.15 and 4.16).

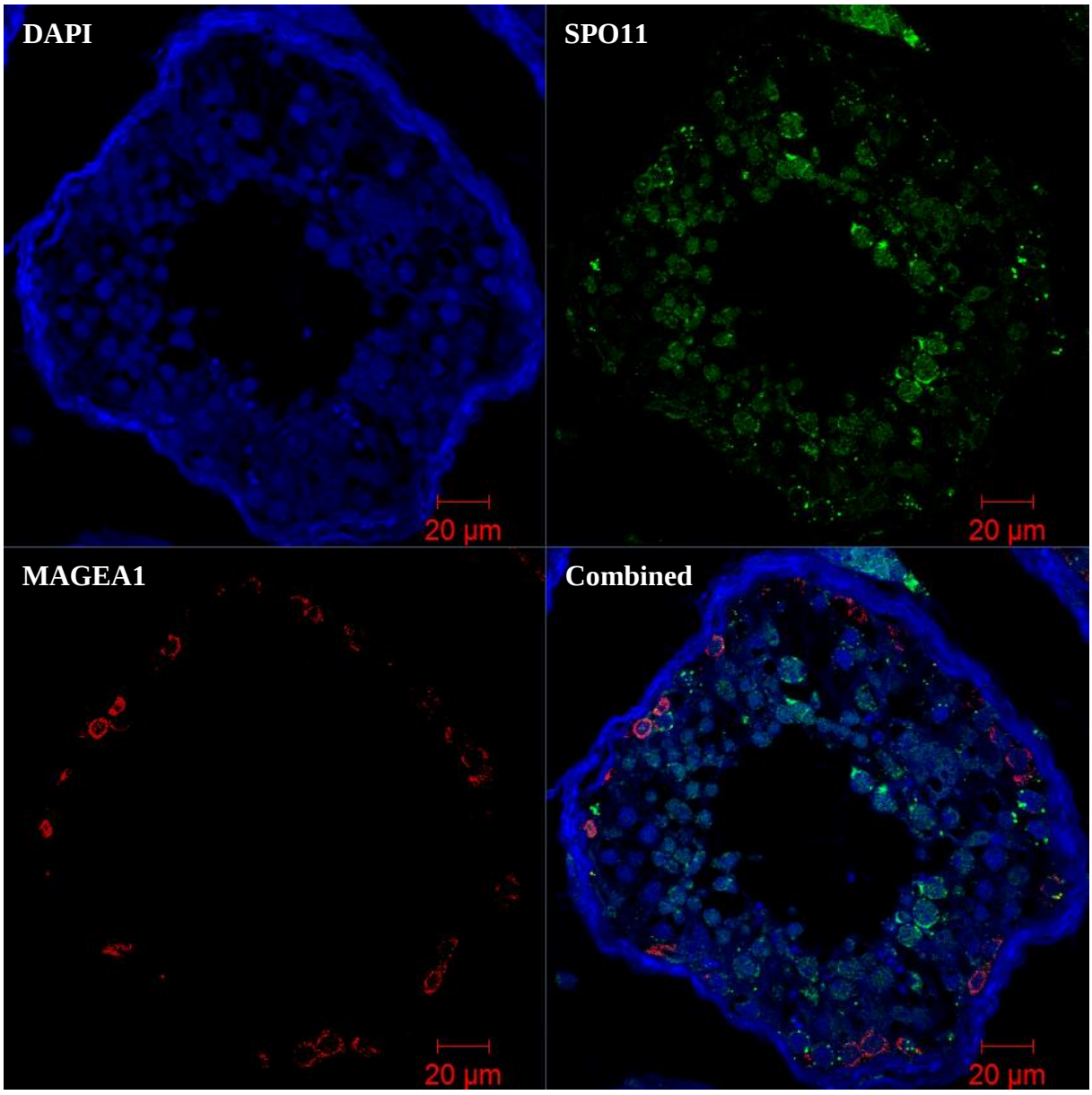


Figure 4.12. SPO11 and MAGEA1 staining in normal testis (x40).

IF was carried out on FFPE tissue blocks of normal testis showing MAGEA1 (red) staining detected by the mouse monoclonal antibody (LSBio, LS-C87868) localised to the spermatogonial layer of the seminiferous tubule. The staining for SPO11 (green) was with the rabbit polyclonal antibody (Abcam, ab81695). Although there was some clear focal positivity for SPO11 within the basal compartment of the seminiferous tubule, there were very few cells which displayed exact co-localisation with MAGEA1. Moreover, staining for SPO11 was stronger within developing germ cells in the adluminal compartment. Images acquired on Zeiss LSM 710 confocal microscope.

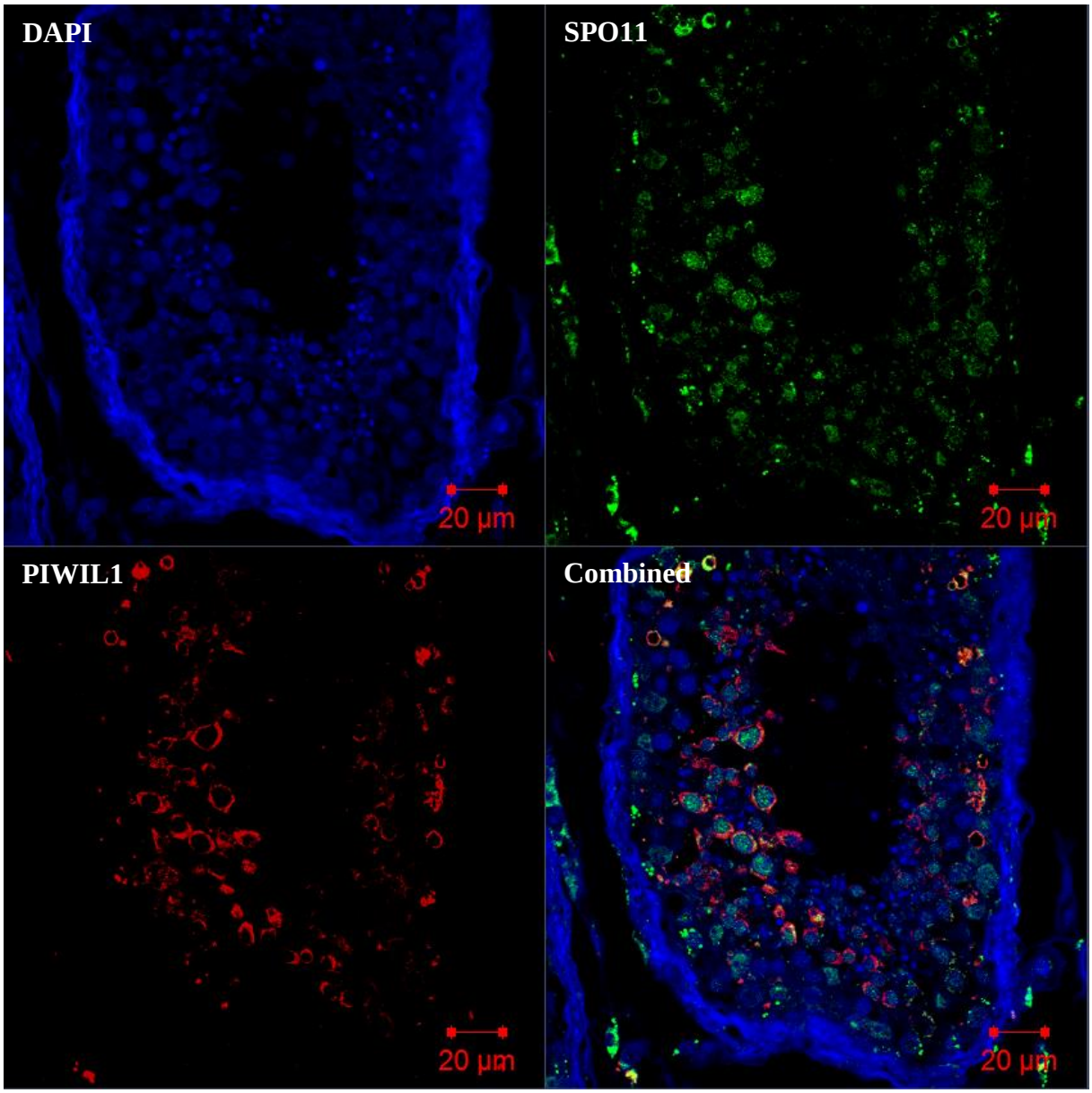


Figure 4.13. SPO11 and PIWIL1 staining in normal testis (x40).

IF was carried out on 4 µm sections of FFPE tissue blocks of normal testis showing SPO11 (green) staining detected by the rabbit polyclonal antibody (Abcam, ab81695). The staining for PIWIL1 (red) was with the mouse monoclonal antibody (Sigma, SAB4200365). The staining pattern was similar in distribution but SPO11 appeared to have a predominant nuclear localisation within developing spermatocytes, whereas PIWIL1 as before demonstrated stronger cytoplasmic staining. There were some cells that stained SPO11 only but for PIWIL1 positive cells, they appeared also to stain positive for SPO11. Some mature spermatids are also seen, which are weakly positive for SPO11 only. Images acquired on Zeiss LSM 710 confocal microscope.

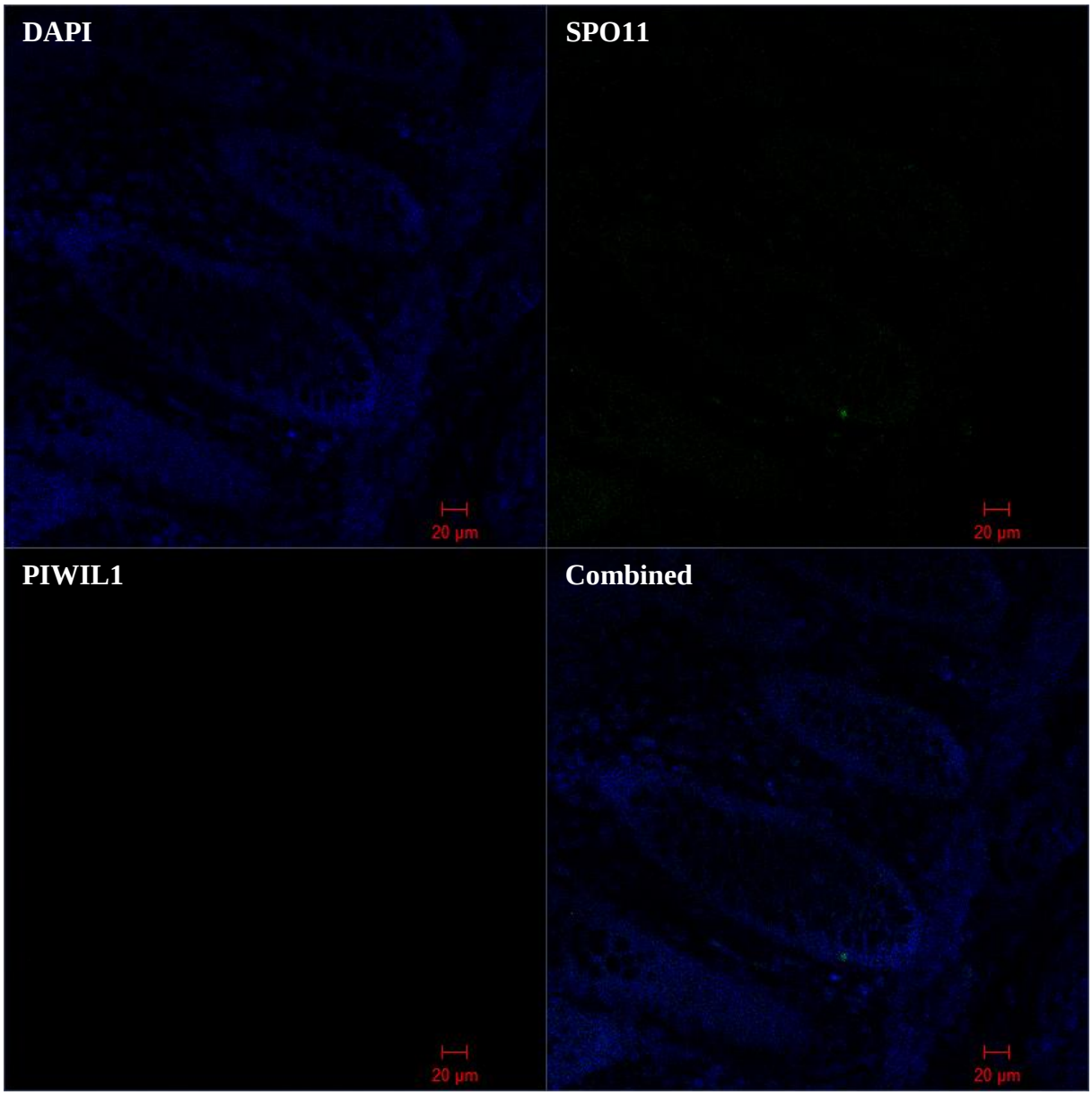


Figure 4.14. IF staining of SPO11 and PIWIL1 in normal adjacent colonic mucosa.

IF was carried out on 4 µm sections of FFPE tissue blocks of normal colon sample B9 showing SPO11 (green) staining detected by the rabbit polyclonal antibody (Abcam, ab81695). The staining for PIWIL1 (red) was with the mouse monoclonal antibody (Sigma, SAB4200365). Staining for PIWIL1 was negative and there was weak positive nuclear staining for SPO11 in a cell at the interface between the progenitor and stem cell zone. In other regions of normal adjacent colonic tissue there was patchy staining for SPO11 in occasional cells just beneath the basement membrane of the colonic crypt. Images acquired on Zeiss LSM 710 confocal microscope.

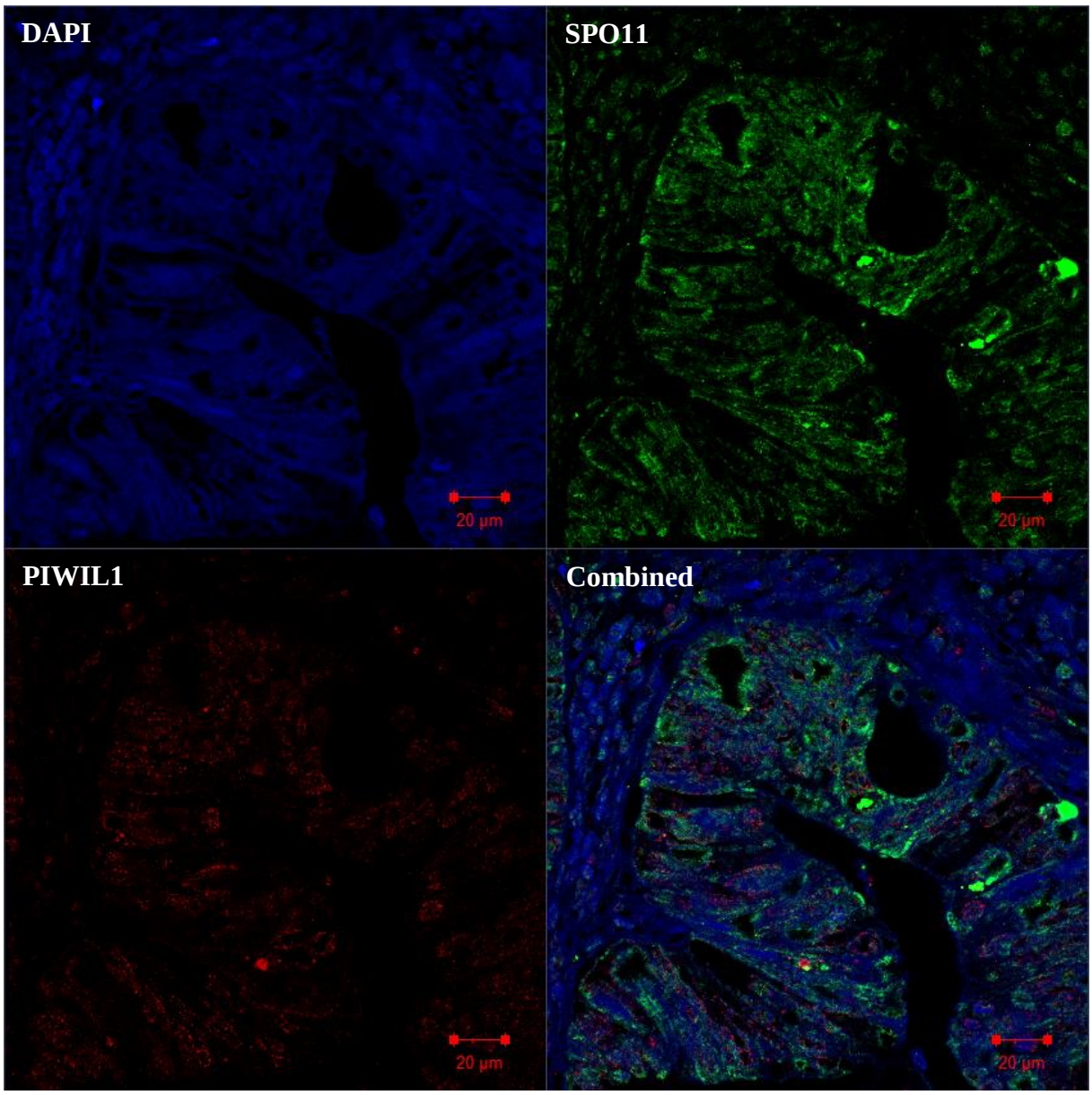


Figure 4.15. IF staining of SPO11 and PIWIL1 in colon cancer.

IF was carried out on FFPE tissue blocks of colon cancer sample C9 (matched adjacent tumour to Figure 5.14) showing SPO11 (green) staining detected by the rabbit polyclonal antibody (Abcam, ab81695). The staining for PIWIL1 (red) was with the mouse monoclonal antibody (Sigma, SAB4200365). Staining for PIWIL1 was weakly positive in the adenomatous crypt as well as tumour stroma. SPO11 staining appeared stronger in the adenomatous tissue and weaker in the stroma but there was likewise a lot of non-specific staining as seen for PIWIL1. There was no clear co-localisation for SPO11 and PIWIL1 but some overlap in staining that was not surprising given the high amount of non-specific staining seen. Images acquired on Zeiss LSM 710 confocal microscope.

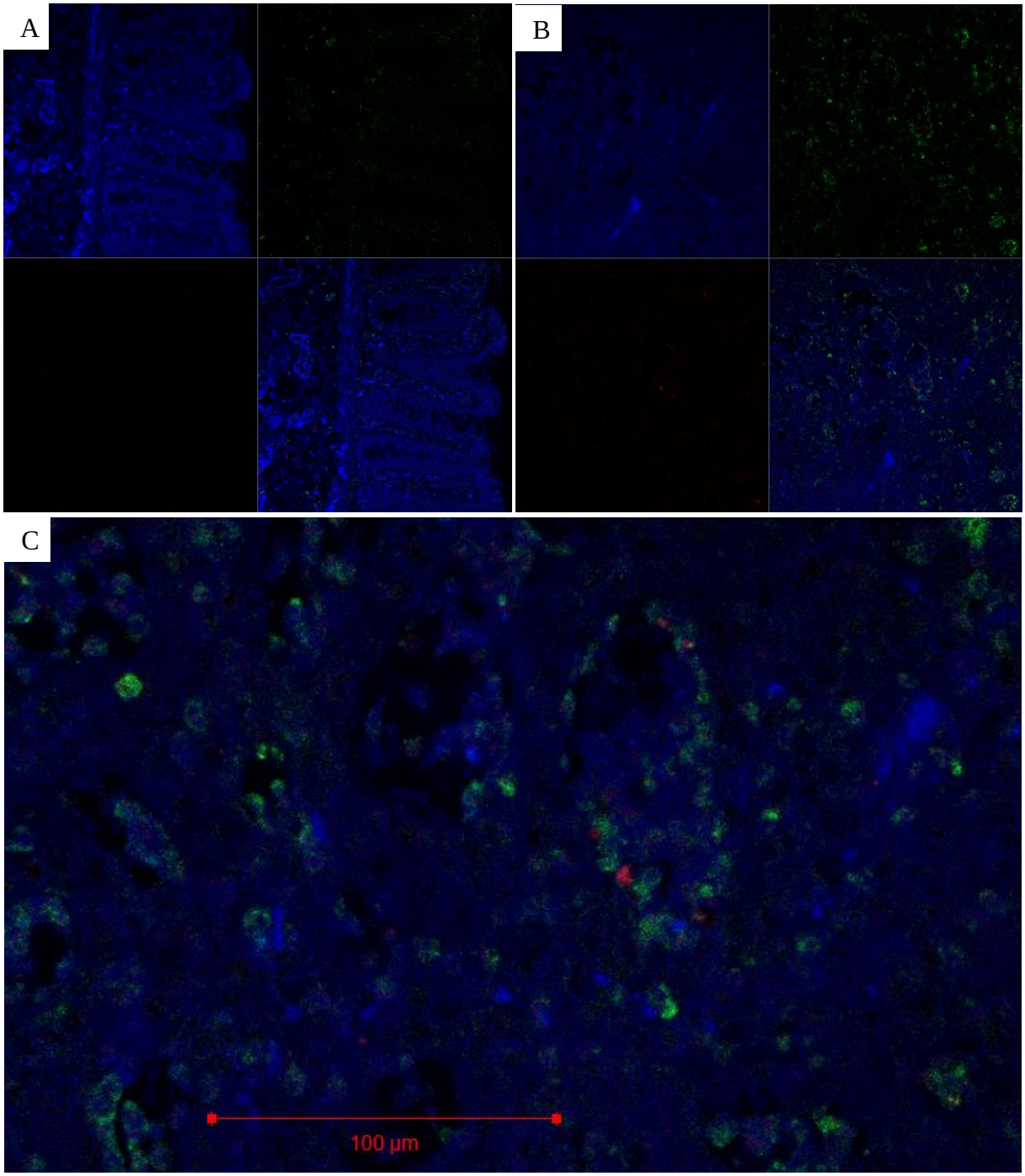


Figure 4.16. IF staining of SPO11 and PIWIL1 in colon cancer and normal adjacent colonic tissue.
IF was carried out on 4 µm sections of FFPE tissue blocks of colon cancer sample C45 and adjacent normal tissue. The normal colonic tissue section in shown in (A) and the colon cancer sample shown in (B). A higher power magnification of the maximum intensity projection image for the poorly differentiated adenocarcinoma is shown in (C). SPO11 (green) staining was detected using the rabbit polyclonal antibody (Abcam; ab81695). The staining for PIWIL1 (red) was with the mouse monoclonal antibody (Sigma; SAB4200365). Staining for PIWIL1 was weak and appeared to be have nuclear localisation but this was not clear. SPO11 staining was slightly stronger but patchy and appeared to localise to both the cytoplasm and nucleus. Again there was no clear co-localisation for SPO11 and PIWIL1. Images acquired on Zeiss LSM 710 confocal microscope.

The weak and suggested nuclear staining for PIWIL1 in colon cancer using IF (as displayed in Figure 4.16) was investigated for confirmation by IHC. Attempts to stain for SPO11 using IHC displayed a lot of background staining and I did not manage to produce interpretable results due to the strong staining seen. However, for PIWIL1 there was more consistent results using the monoclonal antibody used in the IF experiments (Sigma; SAB4200365, Lot. 111M4781). As was seen on IF, IHC on normal testis produced positive and relatively clean staining in the adluminal compartment of the seminiferous tubules (see Figure 4.17). A selection of the colon cancer and NAT colon samples that I had collected were also tested: I chose to stain some based on the strength of expression indicated by the gene expression analysis in the matched samples (see Chapter 5 for details). The colon samples appeared negative, though there was perhaps some focal weak nuclear positivity in a poorly differentiated adenocarcinoma which the PCR results had indicated expressed *PIWIL1* strongly (see Figure 4.18). This result was broadly consistent with what the IF experiments had shown on the same tissue sample (see Figure 4.16). It may however be the case that despite the *PIWIL1* gene being expressed in several colon cancers tested, the protein is not translated or the experimental conditions used were unable to detect the presence of the antigen. PIWIL1 did appear to be present in breast cancer samples that were included on a TMA slide used to test the PIWIL1 antibody – see Figure 4.18.

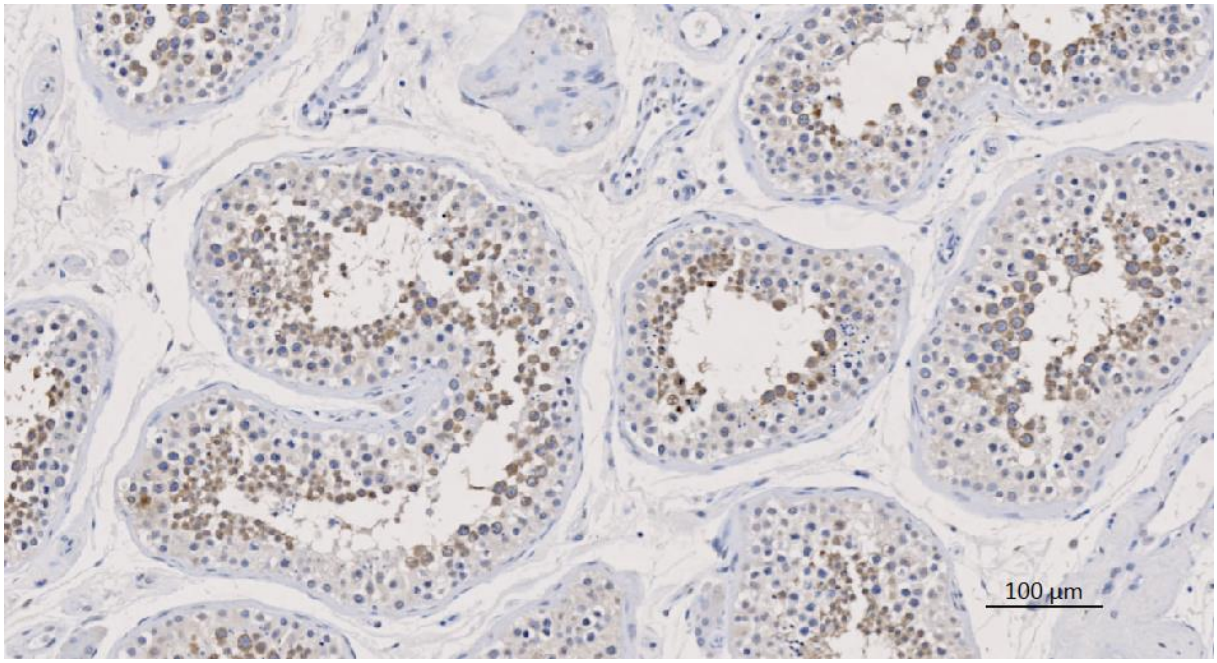


Figure 4.17. IHC staining for PIWIL1 in normal testis.
IHC was performed on 4 µm section of normal testis using the mouse monoclonal antibody targeting PIWIL1 (Sigma, SAB42000365). Image obtained on a Zeiss Axio Scan.Z1 digital slide scanner.

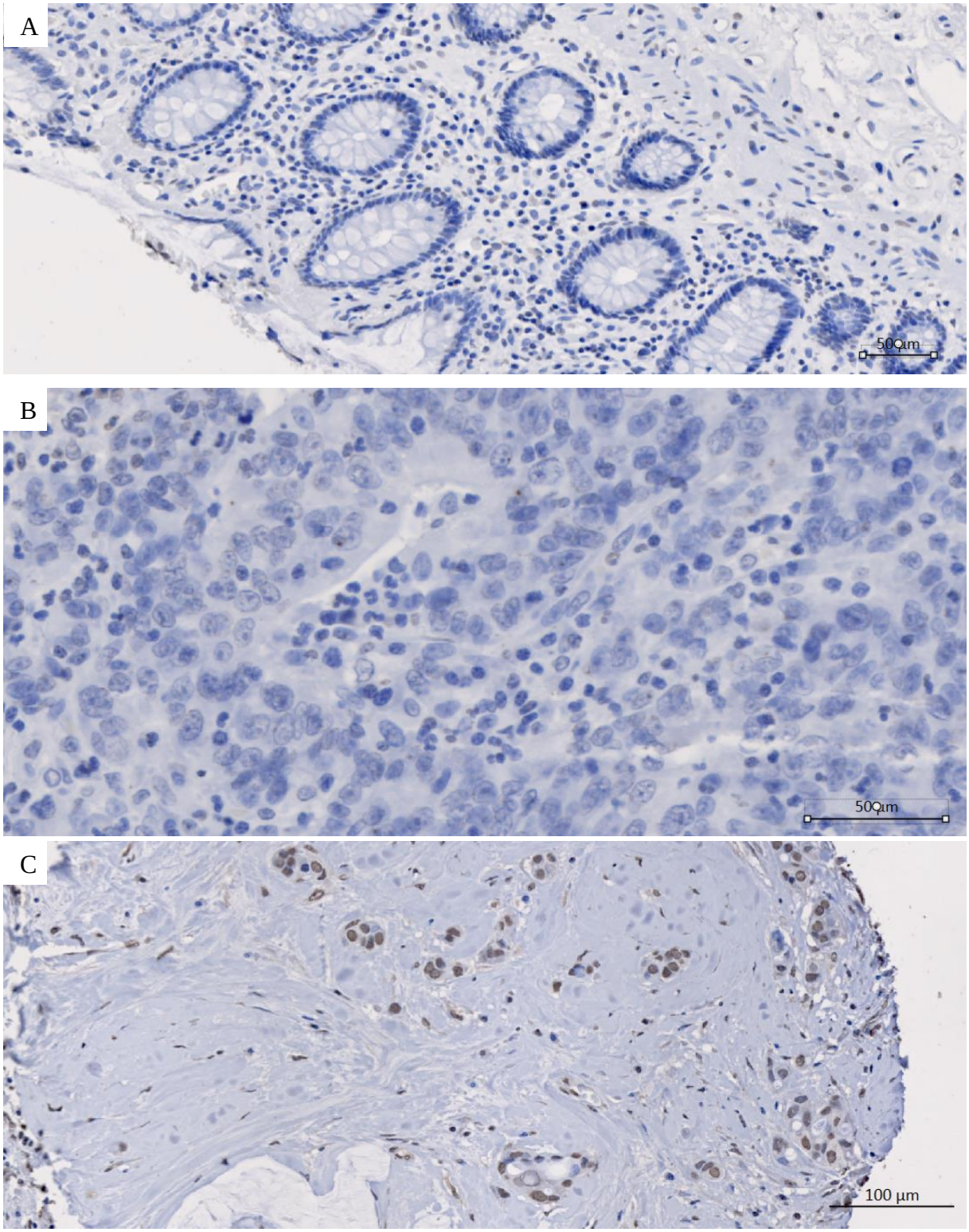


Figure 4.18. IHC staining of PIWIL1 in normal colon, colon cancer and breast cancer.

IHC was performed on 4 µm tissue sections of colon cancer sample 45 and adjacent normal tissue from the same patient (A). There was a suggestion of some weak scanty nuclear positive staining in this poorly differentiated adenocarcinoma shown in (B) but the majority of the tissue sections stained negative, despite previous results indicating that *PIWIL1* was expressed at reasonable levels in some tissue samples. There was more obvious patchy nuclear positivity in a breast cancer sample. Images obtained on a Zeiss Axio Scan.Z1 digital slide scanner.

4.5. Assessment of the potential immunogenicity of TEX19 and C20orf201

From the gene expression analysis I conducted and previous work to identify novel meiosis-associated CT genes, TEX19 and C20orf201 appeared two of the most promising candidates. The attempts to establish the localisation and presence of the proteins in normal and cancer tissues produced mixed but broadly encouraging findings. To establish the immunogenic potential of these two proteins, customised peptides from each protein were produced and used to see if they could bind to T-cells by utilising a cell binding assay in specialised T2 cells. An example of the flow cytometry result is shown in Figure 4.19 and a summary of all the results is displayed in Table 4.5. Three repeat experiments were performed.

If the peptide binds to the T cell, then there will be a change in the mean fluorescence intensity (MFI). It can be seen in Table 4.5 that several peptides produced a positive result – an increase of at least 30% above the MFI for the control is significant. Five of the TEX19 peptides appeared of potential immunogenic interest, with the most significant change in intensity seen for two peptides that produced in excess of a 100% change in the MFI value: these were LEDAGLDPHF and YLYASWMYQL that were both predicted to bind to HLA-A2 molecules strongly.

Nine of the C20orf201 peptides appeared to be potentially immunogenic (i.e. resulted in an at least 30% increase in the MFI). Three of these peptides produced an increase in the MFI of over 100% over the control cell population. It should be noted that the T-cell binding assay does not prove that the peptide will be immunogenic *in vivo* and the 'strong binders' (i.e. the peptides which produce the largest change in MFI) do not always turn out to be the most immunogenic (Nunes *et al*., 2011). It does, however, give a good indication of which peptides would be good candidates to take forward for a given human population.

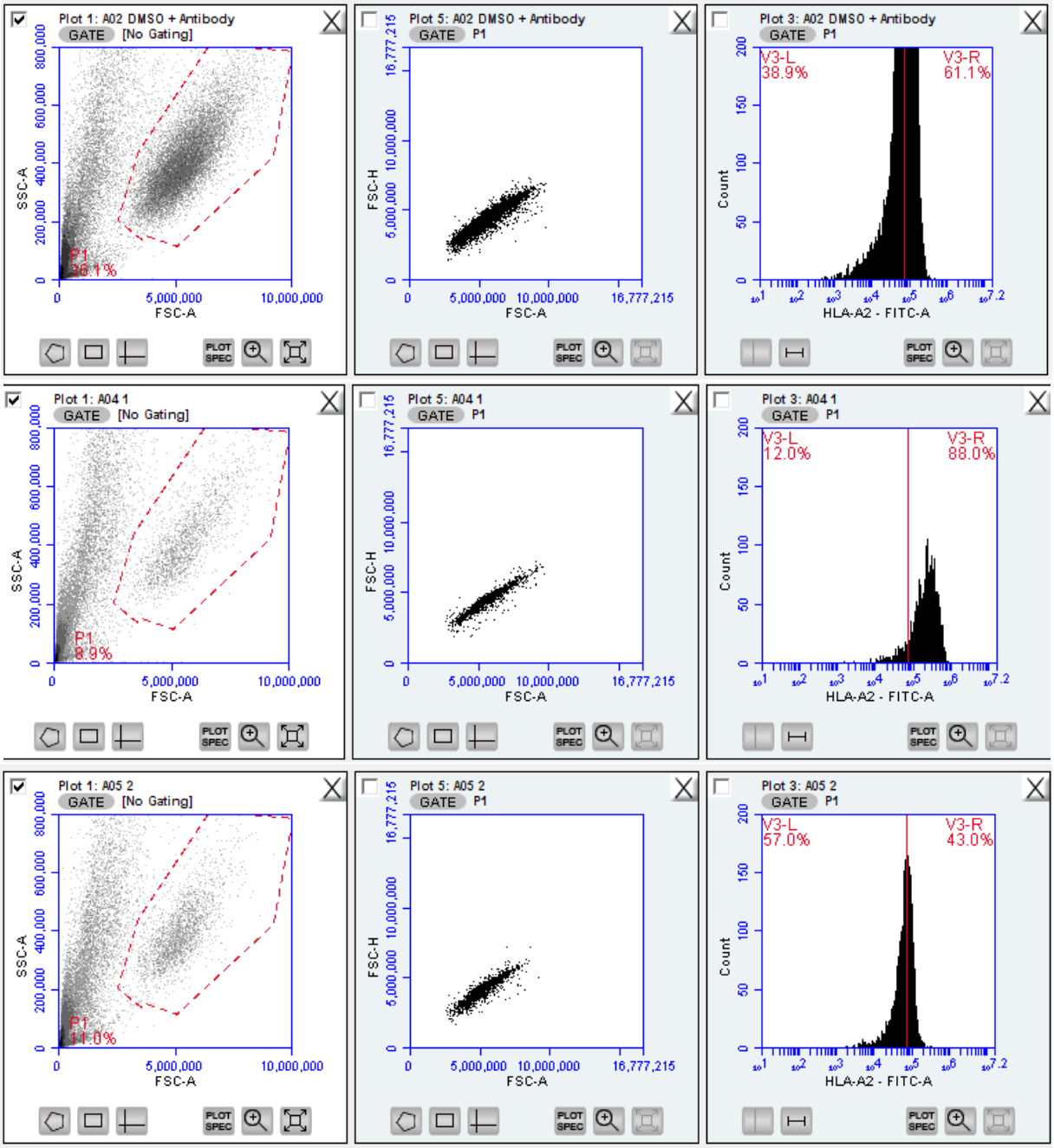


Figure 4.19. Flow cytometry result for T-cell binding assay.

T2-cells were cultured overnight with dimethyl sulfoxide (DMSO) only as a control or with customised peptides dissolved in DMSO. A flow cytometric readout was performed – the vertical red line indicates the mean fluorescence intensity (MFI) for the DMSO only treated cells. The red line is indicated in the same position for the two peptides below this: for the middle peptide there is a shift in the fluorescence intensity to the right and the bottom peptide the MFI is very similar to that of the control. The middle peptide (YLYASWMYQL) for TEX19 can be thus seen to bind strongly to the T2 cells and result in a shift in the MFI. Results analysed using FCS Express 4 Flow Research Edition software.

Table 4.5. Summary of T-cell responses to the tested peptides

Peptide	Sequence	Protein	MFI-1	MFI-2	MFI-3	% mean increase
1	YLYASWMYQL	TEX19	253037.67	635064.04	1239791.90	395.25
2	QLSICFTCFK	TEX19	71341.99	116117.48	287597.33	18.25
3	ELWPQEAVPL	TEX19	99752.60	97660.75	350629.75	41.19
4	LPWRFEELL	TEX19	82306.63	112726.43	220123.91	1.43
5	WPQEAVPLGL	TEX19	128830.05	104999.47	375714.32	54.33
6	YEEEGMSYLY	TEX19	83743.93	79100.43	282640.05	14.00
7	EEEGMSYLY	TEX19	65562.28	91430.40	284992.98	15.84
8	LEDAGLDPHF	TEX19	151888.31	108245.03	376706.04	58.28
9	GLPWRFEEL	TEX19	217289.02	298953.73	374186.61	103.53
10	QLQHGDQLSI	TEX19	93827.72	104864.29	288924.77	23.12
11	RMPDPRPWT	C20orf201	91464.79	88398.82	333698.49	33.21
12	LLFGDLLEDV	C20orf201	124848.21	343677.00	1519956.95	456.73
13	DPRPWTQSL	C20orf201	71370.35	85626.57	352472.44	36.14
14	EARGRVRAL	C20orf201	157674.73	161374.24	337502.72	57.25
15	FLPPQLKPA	C20orf201	444807.22	595682.71	1060447.52	405.33
16	RLLGVLKAA	C20orf201	241931.96	482995.83	505697.68	185.48
17	ATEPPKPGW	C20orf201	136002.64	148423.50	335763.63	51.52
18	RPWTQSLEL	C20orf201	71109.13	77902.70	264482.76	7.44
19	EENVDGTIF	C20orf201	88822.55	91225.62	300866.52	22.84
20	LPAQRHRHL	C20orf201	116007.56	110243.22	334121.90	40.78
21	ELFLPPQLK	C20orf201	79095.62	87739.02	287653.44	17.75
22	LPAQRHRHLL	C20orf201	127779.12	121689.59	219097.34	8.74
23	AEARGRVRAL	C20orf201	153480.42	139289.74	274059.40	32.09
24	Positive control	M1	79882.06*	193124.10	265955.81	48.02*
N/A	Negative control	N/A	87734.06	112603.19	213555.36	N/A

MFI – Mean Flourescence Intensity. * - MFI1 result excluded as M1 peptide diluted inappropriately. % mean increase refers to the percentage mean increase in MFI taken from the three repeat experiments.

4.6. Discussion

The search for more precise methods of cancer stratification is as important as ever. As we enter an era of more personalised treatment strategies, a vast array of biomarkers will be required to meet the clinical need of accurate stratification, which may use a combination of markers and approaches to reach the desired degree of predictive accuracy. The work presented in this chapter makes the initial steps towards a translational research outcome at the antigenic level for some novel and poorly characterised germline genes. Despite the encouraging initial findings of the staining pattern in ovarian cancer samples and the validation of the antibody using siRNA knockdown, the antibody against C20orf201 (Abcam, ab108142) did appear to produce positive staining in tissue for which we did not expect to see positive staining. This may relate to a lack of specificity of this polyclonal antibody. Alternatively, very low levels of gene expression may produce a protein that is translated and stable *in vivo.* Weak expression was in fact seen on the TLDA card analysis in a few normal tissue samples, which was more than indicated by the previous RT-PCR analysis by our group in which expression was limited in normal tissues to the testis and CNS (Feichtinger *et al*., 2012b). So, the gene may not have such tight tissue specificity as we originally thought and this is reflected in the positive staining patterns seen on IHC.

As far as developing a novel therapy (e.g., immune-based) targeting C20orf201 is concerned, the fact that weak staining was seen in normal ovarian tissue is perhaps not of particular significance. If the surrounding tissues do not contain the protein, as the IHC experiments indicated, then damage to normal ovarian tissue is unlikely to be of concern to the patient, especially if they are of post-menopausal age as is often the case. However, given that immunotherapeutics are given systemically, an antigen present anywhere in the body could potentially be targeted by such an approach and result in adverse/unwanted consequences. Given that staining of the normal TMAs revealed staining in some other normal tissues outside of the ovaries, this would potentially prevent the protein being used as an immunotherapeutic target. However, a TAA that is present at very low levels in normal healthy tissues may still be a potential target of novel immunotherapies if present to a far higher degree in the cancer being targeted. Given the differential staining seen in serous ovarian cancer samples tested in this, it may be possible to use the presence of C20orf201 as a predictive marker to help guide certain therapies, even if low level presence in certain somatic tissues could obviate its use as an immunotherapeutic target.

Also, I experienced some reproducibility problems when ordering a new batch of the C20orf201 antibody, which highlights the problems and importance of reproducibility in experimental research. This is particularly true when attempting to translate the findings into the clinical setting. Another issue with antibody production is the specificity of antibodies. Although polyclonal antibodies can be 100% specific for the target protein, there is an increased likelihood (over monoclonal antibodies) that other proteins could be detected. To deal with the reproducibility issues of using animals to produce antibodies, some have urged a shift-change in the way antibodies are produced with the antigenic sequences being defined at the DNA level and produced within engineered cells (Bradbury and Pluckthun, 2015). The process and issues surrounding antibody validation have recently been discussed (Baker, 2015; Bordeaux *et al*., 2010). One reasonable conclusion is that, given staining was seen in tissues not expected to contain the protein, the antibody should be "abandoned", as specified in the final validation step described by Bordeaux and colleagues (Bordeaux *et al*., 2010). C20orf201 may still be an interesting TAA with therapeutic potential but the antibody I have been using may lack the rigour required to take the findings forward into the clinical setting. Since we published our initial findings highlighting *C20orf201* as a potential meiCT-gene (Feichtinger *et al*., 2012b) another group have provided evidence that the antigen is produced in patients with prostate cancer (Kamata *et al*., 2013). I did not test for the expression of the gene in prostate cancer. Together these combined findings suggest the *C20orf201* may be of relevance in certain cancer types and warrants continued experimental exploration.

The other novel meiCT gene that I investigated for antigen production was *TEX19.* Deletion of *TEX19* has been shown to result in activation of retroelements, or transposable elements (TE), in mice (Ollinger *et al*., 2008). The PIWI proteins, which form complexes with piRNAs, together repress the expression of TEs in the germline and guard against unwanted genetic alteration as a result of TE activation (Siomi *et al*., 2011; Watanabe and Lin, 2014). Our group has shown that knockdown of *TEX19* in a colorectal cancer cell line resulted in upregulation of *PIWIL1* (Planells Palop, unpublished data). IF data here has shown both proteins to be present in the seminiferous tubules without significant co-localisation. It may be the case that in cancer, if TEX19 is present it could promote activation of the TEs that promote oncogenic change. However, in the testis there is competition from piRNAs and PIWI proteins to protect against this effect. The physiological effects may also depend on the localisation and magnitude to which the protein is present in the cell. For example, cytoplasmic staining appeared to be present in the seminiferous tubules of the testis and weak cytoplasmic staining also present in

some normal colonic epithelial cells adjacent to the luminal surface. Whereas, in colon cancer samples, stronger cytoplasmic as well as nuclear staining for TEX19 was seen in some instances (for example, see Figure 4.11).

Somewhat confusingly, the IHC example showing perhaps the clearest differential staining within the same specimen, did not appear to display expression of *TEX19* in the TLDA analysis (see Figure 5.18). This may be because the RNA was extract from a non-cancerous portion of the specimen. This highlights a deficiency of the sampling and experimental technique. It can be seen from Figure 4.10 that under half of the specimen contained cancer, so it is quite possible given that not the entire sample was homogenised for RNA extraction that the invasive cancer, and region where *TEX19* may have been expressed more strongly, was inadvertently excluded.

TEX19 displayed consistent staining pattern on both IHC and IF experiments and appeared to stain a 'matrix' between the germ cells. An alternative possibility is that the staining pattern seen was in the cytoplasm of specific spermatocytes close to the BTB; nuclear staining was never observed. There was more intense staining towards the spermatogonial or basal compartment of the seminiferous tubules. One distinct possibility is that TEX19 is not in fact present within the germ cells and may be present within the cytoplasm of Sertoli cells which are known to be critical for regulating spermatogenesis. A staining seen predominantly outside of the germ cells, does raise questions about what functional role it may play during germ cell development. Perhaps it acts via interaction with cell surface receptors to regulate spermatogenesis but does not play a direct role in meiosis. The stronger staining towards the basal compartment where the stem cells of the germline reside, would suggest a possible link to pluripotency for which a functional role has already been suggested (Kuntz *et al*., 2008). Co-staining for known cytoplasmic markers of Sertoli cells should be performed to investigate if TEX19 does co-localise with these markers. Similarly, IF experiments with known spermatocyte subtype-specific cell markers may reveal an association with specific germ cells within the testis but I believe the staining pattern observed and presented here is less consistent with this possible cellular localisation.

SPO11 is responsible for the formation of DNA DSBs and the initiation of meiotic recombination. The protein was shown to have predominantly nuclear localisation within germline cells in the testis (see Figure 4.13). IF experiments also indicated that very low levels of protein expression may be present in normal colonic tissue. The nuclear staining of a

possible stem cell in the region of a normal colonic crypt (as shown in Figure 4.14) was not a consistent finding and no co-localisation with known stem cell markers was performed. It was found some time ago that the *Drosophila* orthologue of SPO11, mei-W68, was present in somatic cells and appeared to have a mitotic role (McKim and Hayashi-Hagihara, 1998). This finding has not been corroborated in humans and it is widely regarded as a meiosis-specific protein where it regulates other aspects of chromosome dynamics in addition to forming DNA DSBs (Celerin *et al*., 2000; Murakami and Keeney, 2008). SPO11 may also play a role in check-point signalling and the targeting of cells with mitotic errors for destruction following entry into meiosis (Stevens *et al*., 2013). SPO11 is considered a CTA and may be present in limited amounts in cervical cancer, melanoma, colorectal cancer and lymphoma (Almeida *et al*., 2009; Eldai *et al*., 2013; Koslowski *et al*., 2002; Litvinov *et al*., 2014). In general there has been limited confirmation of the presence of either mRNA or protein in human cancers; for example, for squamous cell carcinoma of the head and neck region and acute myeloid leukaemia the gene has been found not to be expressed in patient samples, and the only sample that did contain *SPO11* in these studies was a single leukaemia cell line (Atanackovic *et al*., 2006; Atanackovic *et al*., 2011). In the samples I tested I found the gene to have a testis-restricted expression profile (see Figure 5.20).

Despite the apparent lack of expression of *SPO11* in the cancer samples I had tested, I went on to test a selection of colon cancer samples using IF. SPO11 had been described as potentially relevant in CRC (Eldai *et al*., 2013) and I co-stained for the presence of PIWIL1 in some colon cancer samples (see Figure 4.15 and Figure 4.16). *PIWIL1* did appear to be expressed in an increased number of colon cancer samples, compared to the adjacent normal tissue (see Figure 5.23). PIWIL1 was not present in NAT colonic tissue when assessed using IF, whereas SPO11 displayed possible and very limited staining as just described. In the cancer samples there was non-specific staining for both proteins. This could be due either to low levels of protein present in the cancer or non-specific binding of the antibody to other antigens within the tissue. The presence of SPO11 appeared stronger in general in the cancer samples compared to PIWIL1, despite an indication from the TLDA analysis that *PIWIL1* was activated in some of the colon cancer samples, whereas *SPO11* was not. One explanation for this, and for the lack of confirmation of *SPO11* expression more generally in cancer, could be the lack of polyadenylation in the mRNA encoding SPO11. Oligo-dT was used as the complementary primary for all cDNA synthesis reactions in my experiments.

4.7. Conclusions

C20orf201 and *TEX19* are germline associated genes that appear to be upregulated in certain cancers. It has been shown that they have potential as immunogenic biomarkers and thus may serve as targets for immune-based treatments against cancer. The function of these genes is yet to be established, although TEX19 may interact with or influence the function of PIWI proteins in the germline and could via a gene activating role cause the upregulation of other oncogenic genes when present in cancer. C20orf201 has been shown to be a potential biomarker in ovarian cancer and shown to be present in the majority of serous cancers tested by IHC. TEX19 likewise appears to be present in increased amounts in colon cancer, though not clearly associated with a particular tumour grade or stage.

Both proteins appear to be present in some normal tissues not expected from their gene expression profiles. We have validated the antibodies used by siRNA knockdown but a common feature seen with these polyclonal antibodies may be the detection of multiple antigens, leading to a lack of specificity and detection of a different protein in the tissue samples when using IHC and IF. This is an historical and ongoing issue in the field of cancer research and molecular biology more widely (Bordeaux *et al*., 2010). Validating and profiling the behaviour of antibodies on an appropriate platform that can be easily translated into the clinical setting is important. PIWIL1 appears to be upregulated in some colon cancer and could potentially via epigenetic mechanisms influence expression of other genes in cancer. There are emerging roles for PIWI-proteins and piRNAs in cancer, including colorectal cancer (Li *et al*., 2010; Siddiqi and Matushansky, 2012; Tan *et al*., 2015); these should be investigated further along with a possible role for TEX19 via related, opposing or independent mechanisms.

5. Germline gene expression in colon and other cancers

5.1. Introduction

5.1.1. Classification and staging of colorectal cancer

At least three recognised 'routes' to the development of CRC exist (i.e. the CIN, MSI and CIMP pathways) although there is some degree of overlap between these various pathways (Al-Sohaily *et al*., 2012; Fearon and Vogelstein, 1990; Tamas *et al*., 2015). Most cases of CRC are sporadic and generally occur in middle-aged people and the elderly. The Vogelstein model for carcinogenesis, which remains representative of the most common CIN pathway (see Figure 1.13), served as a paradigm for our understanding of molecular pathways in other cancer types.

Traditional staging was based on the Duke's staging system, which groups CRC in 4 groups (A to D). This system remains in use today, although it has been largely superseded by the tumour-node-metastasis (TNM) system that is used more widely across a variety of solid tumours internationally (Compton and Greene, 2004). The tumour stage is numbered I to IV according to the TNM staging system and this correlates to the Duke's staging A to D. A summary of these staging systems is provided in Table 5.1, along with the approximate associated survival rates. This does not take into account the importance of the degree of differentiation, as well as other molecular characteristics related to intra- and inter-tumour heterogeneity (Compton and Greene, 2004; Hu *et al*., 2011). Undoubtedly the staging system will continue to evolve and undergo more accurate refinement; survival rates will also change as treatment efficacy improves over time. Indeed, there have been recent studies based on the genetic signature of the disease, which attempt to classify the disease more accurately into clinically meaningful subgroups (Marisa *et al.*, 2013, Sadanandman *et al.*, 2014). These alternative classification systems are inextricably linked to our better understanding of tumour heterogeneity.

Table 5.1. Staging and survival in colorectal cancer

Stage	Tumour	Node	Metastasis	Duke's	1 year survival	5 year survival
I	T1 or T2	N0	M0	A	~95%	~90%
IIA	T3	N0	M0	B	~90%	~85%
IIB	T4	N0	M0	B	~80%	~65%
IIIA	T1 or T2	N1	M0	C	~90%	~80%
IIIB	T3 or T4	N1	M0	C	~80%	~65%
IIIC	Any T	N2	M0	C	~80%	~55%
IV	Any T	Any N	M1	D	~40%	~10%

T1 – Tumour invades submucosa; T2 – Tumour invades muscularis propria; T3 – Tumour invades through the muscularis propria; T4 – Tumour invades other organs or perforates the visceral peritoneum; N0 – no lymph node metastases; N1 – Metastasis in one to three lymph nodes; N2 – Metastasis in four or more lymph nodes; MX – presence of distant metastasis cannot be assessed; M0 – No distant metastasis; M1 – Distant metastasis present. Adapted from American Cancer Society (www.cancer.org – accessed 30th July 2015) and (Compton and Greene, 2004).

5.1.2. Alternative classification and heterogeneity of colorectal cancer

Recent studies have characterised CRC into distinct molecular subtypes based on their gene expression profiles, which correlated with known signalling pathways and clinical outcomes (Marisa *et al*., 2013; Sadanandam *et al*., 2013; Sadanandam *et al*., 2014). As the cost and speed of transcriptomic profiling continues to improve it would appear likely that such classification systems will become widely used to complement existing stratification methods such as TNM staging. A major challenge in developing such stratification methods in a clinical practice will be dealing with intratumour heterogeneity, as well as inter- and intra-metastatic heterogeneity (Jamal-Hanjani *et al*., 2015). The analysis of circulating factors may prove to be the optimal strategy to overcome this problem, as they could be representative of the entire cancer landscape for a given patient. However, the detection of circulating tumour cells and cell-free (cfDNA), which occur at low concentrations provides an additional level of analytical challenge. The extent of tumour heterogeneity certainly raises questions as to the validity of existing biopsy methods to accurately stratify patients to optimal treatment approaches. Although there is concordance in some important genetic mutations (e.g., *KRAS)*, there are also genetic differences between CRC metastases and the primary tumours from which they originate (Sylvester & Vakiani, 2015). Thus, despite the fact metastatic deposits and primary tumours are clonally related to each other, the evolution and complexities of the cancer stem cell models of tumour formation and metastatic spread are increasingly apparent (Visvader & Lindeman,

2012). Moreover, tumour heterogeneity is not purely determined by genetic difference between and within tumours; there are important additional factors to consider. The impact of epigenetics, the tumour microenvironment and the inherent random variability of cellular biological processes all contribute to tumour heterogeneity (Caiado *et al.*, 2015).

5.1.3. Melanoma associated antigen genes in colorectal cancer

The melanoma associated antigen (MAGE) family of germline genes serve as a good example of the variability of CT gene expression in CRC. It can also be seen that some CRCs do co-express several genes, which suggests these tumours may respond to specific therapeutic approaches. The MAGE family of CTAs were the first to be described (van der Bruggen *et al.*, 1991). They have tightly restricted expression limited to the testis in normal tissues and are among the most immunogenic of the known CTAs (Chomez *et al.*, 2001; Meek and Marcar, 2012; van der Bruggen *et al.*, 1991). They are encoded by a large paralogous gene family of over 60 genes that are located on the X chromosome (Chomez *et al.*, 2001), and whilst their function in testis germ cells is poorly understood, they are known to possess oncogenic activity (Bhan *et al.*, 2012; Ladelfa *et al.*, 2012). Several of the MAGE gene family members are expressed in CRC with expression frequencies generally ranging between 5-20% (reviewed in, Sammut *et al.*, 2013). Others genes, such as *MAGEB1* and *MAGEB2* may not be expressed at all in CRC but this has not been extensively corroborated (Lurquin *et al.*, 1997). Burgdorf *et al.* tested the expression of various *MAGE* genes in liver biopsy specimens from nineteen patients with metastatic CRC (Burgdorf *et al.*, 2008). In contrast to a previous study showing no *MAGEA12* expression in primary tumours (Dakshinamurthy *et al.*, 2008), Burgdorf and colleagues found this gene to be expressed in many of the specimens tested. Although the exact frequency of *MAGEA12* expression was not stated, 47% of the specimens were shown to express all six of the *MAGE* genes tested for. Overall 79% of liver biopsy specimens were found to express at least one of the *MAGE* gene family members (Burgdorf *et al.*, 2008). In a separate study a subset of CRC samples were found to express *MAGEA3* with relatively high frequency compared to other known CT genes (Shantha Kumara *et al.*, 2012). This could imply that more aggressive or advanced forms of CRC (i.e. that have metastasised) express *MAGE* genes more widely and that they may even be contributing to the metastatic process. Taken together these findings illustrate the great variability seen between studies that have assessed CT gene expression in tumours. Differences in the site and technique for tissue sampling, post sampling tissue handling and subsequent analytical techniques all potentially impact on the

expression frequency detected. The further complicating factor in gene expression profiling is the extent of (intra-)tumour heterogeneity (Swanton *et al.*, 2010); CT gene expression profiles and differences between studies should also be viewed in this light.

The extent of *MAGE* gene co-expression and collective expression of at least one family member within a cohort of primary tumours has not been extensively investigated for comparison to the Burgdorf study (2008) that assessed CRC liver metastases, therefore limiting the conclusions that can be drawn from these findings. The largest set of samples tested found 90 out of 250 CRC specimens to express at least one of six *MAGE-A* genes (Chen *et al.*, 2010). Li *et al.* also tested for the expression of ten well characterised CTA genes, including four members of the *MAGE* family, in a cohort of 121 CRC patients (Li *et al.*, 2005). Over half (56.2%) of patients expressed at least one CT gene and around one quarter co-expressed two or more of the CT genes; the four assessed *MAGE* genes were relatively widely expressed in this cohort. These figures are not dramatically different to the study investigating liver metastatic specimens, which importantly looked at a different set of *MAGE* genes making a direct comparison of limited value (Burgdorf *et al.*, 2008). An even higher co-expression frequency (approaching 90%) of one or more of ten *MAGE* genes was seen in a study of eighty CRC specimens (Hasegawa *et al.*, 1998). It is interesting to note that *MAGEA8* was found to have the highest expression frequency (44%) in this study and this gene has not been as widely investigated as some of the other *MAGE* genes.

The highly stringent tissue specificity of germline-specific and meiCT genes implies that their gene products could potentially serve as tumour-specific immunotherpeutic targets. Equally as important could be their ability to enhance the accuracy of transcriptomic profile to predict disease outcome. Given the heterogeneity of cancer, both intra- and inter-tumour, the development of a large bank of biomarkers will be of increasing importance in personalised treatment strategies (Jamal-Hanjani *et al*., 2015; Mendelsohn, 2013; Seoane and De Mattos-Arruda, 2014). In this chapter some preliminary work on expression profiling of genes associated with the germline, principally focussing on colon cancer samples, in presented. Our approach started from the initial tissue collection of colon cancer samples and adjacent normal tissue. The data provide some resolution of which genes may be useful biomarkers; we included both meiCT genes that had been identified as possible biomarkers previously by our group, as well as other germline genes identified and reported in the literature. There is a huge variability in the expression of CT genes in CRC and it is generally considered a disease of low germline

gene expression levels (Almeida *et al.*, 2009; Caballero and Chen, 2009; Cheng *et al.*, 2011; Costa *et al.*, 2007; Fratta *et al.*, 2011; Hofmann *et al.*, 2008; Simpson *et al.*, 2005; Viphakone *et al.*, 2015). The variability, as illustrated above with regards to the *MAGE* gene family, is poorly understood and a principal hypothesis in which we were interested was that there may be a subgroup of CRCs with higher levels of germline gene expression contrary to this generally held view.

5.2. Part I - Results

5.2.1. Establishment of biobank and examples of tissues collected

In order to obtain good quality RNA for gene expression screening, an application was made for research and development and ethical approval. Ethical approval was sought from the North West Wales Research Ethics Committee following this application (Ethical approval reference number 12/WA/0042). In brief the study outline involved:

- Attending multidisciplinary team meetings to identify potential patients for the study. I liaised with nurse specialists and other members of the clinical team to distribute information leaflets to the patients in advance of their operation, giving them time to consider if they wished to take part.
- Patients were then approached and consented individually, usually when they were admitted to hospital for the operation. Written informed consent was obtained.
- Prior to surgery (usually in the anaesthetic room) a blood sample was obtained. This was spun the same day at 3600 r.p.m. for 3 minutes and the serum extracted. This was stored at -20˚C before transferring for long term storage at -80˚C.
- Following surgical resection, the specimen was inspected and a sample of the tumour and adjacent normal tissue at least 5 cm distant from the tumour was sampled with a sterile scalpel. This was split into three small pieces: one placed directly in formalin; a second placed in RNAlater®; and a third was flash frozen in isopentane that had been cooled in liquid nitrogen. Some sections were frozen in OCT embedding media as well for approximately 20 seconds that could subsequently be used for frozen sectioning and immunofluorescence if required.
- The tissue in RNAlater® solution was left in the fridge for 24 hours before transferring to -20˚C or -80˚C for long-term storage.

Below is a selection of haemotoxylin and eosin (H+E) stains for the colon cancer and normal adjacent tumour (NAT) samples collected in this study. The H+E stains shown are for some of the samples tested for gene expression of various germline genes and there are also comparative IHC data presented in Chapter 4. A full list of the pathological classification and related clinical variables is provided in Table 5.2. The mean age of the patients for the cohort of samples collected for this research was 71.2 years old and the male to female ratio was 1.67:1. This is

comparable to the known statistics of CRC in this region, with an overall male to female ratio of 1.43:1 (White *et al.*, 2015). The figures depict the heterogeneity, even at the macroscopic level between tumour samples (see Figures 5.1 to 5.7). Although the sampling method provided assurance and knowledge of the tissue processing and RNA quality for example, it did not always guarantee that invasive cancer had been sampled. However, overall the tissue collection provided a reasonable selection of cases for which to perform the screening of potentially relevant germline genes.

The importance of EMT in cancer is well established; epithelial-derived cancer cells that acquire mesenchymal phenotype have a greater propensity to invade and metastasize. For instance, the gene expression methods of CRC classification both identified a subtype with worse prognosis that was associated with stem-cell or mesenchymal-like properties (Marisa *et al*., 2013; Sadanandam *et al*., 2013). Greater resolution of how the tumour stroma is crucially important in contributing to oncogenesis has recently been provided in a study that isolated different subpopulations of cells in CRC (Calon *et al*., 2015). They provided evidence that genes were expressed in the tumour stroma that were responsible for the unfavourable gene expression-classified subtype and could be the main drivers of oncogenic processes.

It can be appreciated in some of the following H+E-stained examples how a reasonable proportion of some solid tumours samples/biopsies are made up of stromal cells. This has implications in the development of novel therapeutic approaches targeting these gene products and also in creating valid predictive gene expression signatures to guide therapy. An alternative approach would be to view the tumour stroma and epithelial cancer cells as a single entity and use both for either of these translational research aims. As long as the genes are differentially expressed (i.e. either in normal stroma *vs.* tumour stroma and/or normal epithelial *vs.* carcinoma cells), and important in oncogenesis, then the diagnostic or therapeutic endpoint is unchanged.

It can also be seen that many of the NAT colonic cancer samples contained a high proportion of non-epithelial/mucosal cells. So, in this study, it cannot be confirmed to what extent the genes are expressed in the supporting tissues. It would be interesting to investigate whether there is differential expression of germline genes in the stroma surrounding the tumour that may be contributing to tumorigenesis. The sampling method and RNA extraction technique adopted in this study did not permit this degree of resolution.

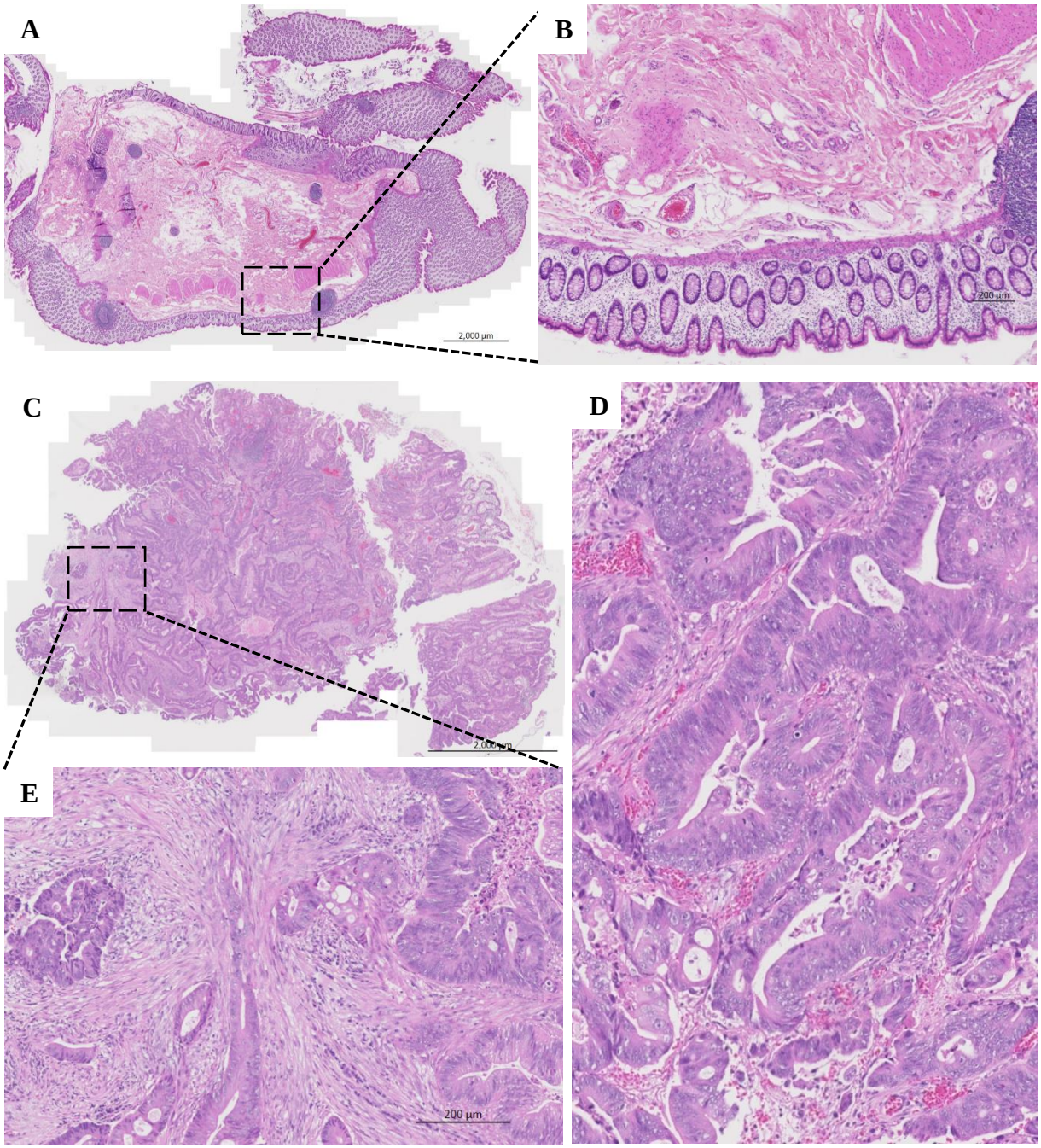


Figure 5.1. Haemotoxylin and eosin stain of colorectal sample 9.

(A) Normal colonic tissue section displaying normal crypts mainly in cross section, submucosa and a small region of muscularis propria. (B) Higher magnification of the normal colonic mucosa, with the muscularis mucosa immediately beneath this and then the submucosa. A small region of muscularis propria is also visible as well as part of a Peyer's patch. (C) Shows the tissue section obtained from the cancer from the same patient – no normal mucosa is visible. The section is a high grade tubular adenoma, and a higher power magnification is shown in (D). Although the tissue section is not deep enough to display invasion, there is a prominent desmoplastic reaction seen on the left side of the tissue section, which is pathognomonic of an underlying cancer. A higher power magnification of this desmoplastic reaction is shown in (E). Thus, it can be concluded that the high grade adenoma is part of an adenocarcinoma.

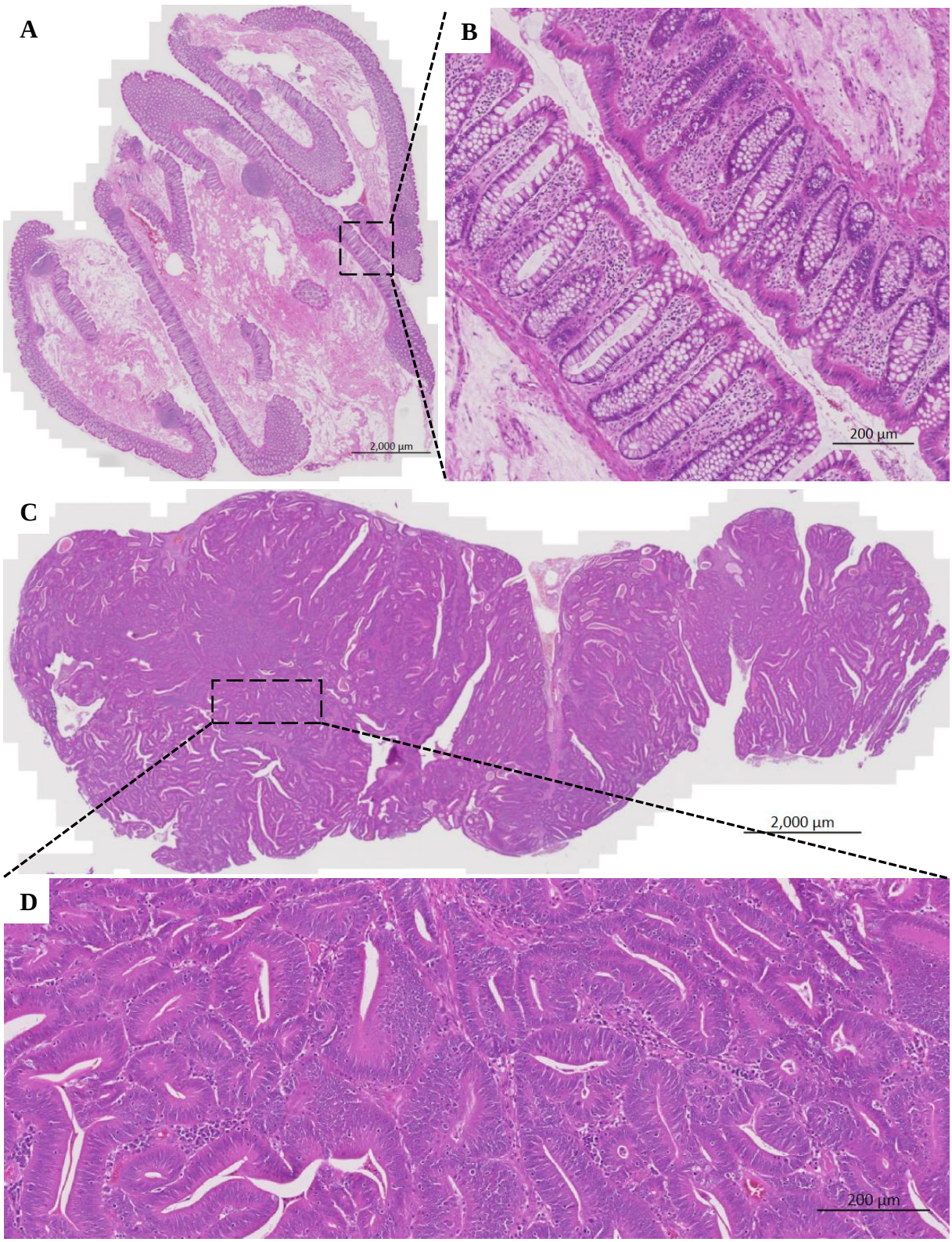


Figure 5.2. Haemotoxylin and eosin stain of colorectal sample 28.

(A) Normal tissue section with mucosa and submucosa visible; no muscularis propria is contained in this section. The crypts of the mucosa are seen predominantly in cross section – a higher power magnification is shown in (B). The tissue sample from the tumour from the same patient is shown in (C). There is no normal mucosa contained within this section but no features that are diagnostic of invasive cancer. A higher power magnification of the tubular adenoma is shown in (D).

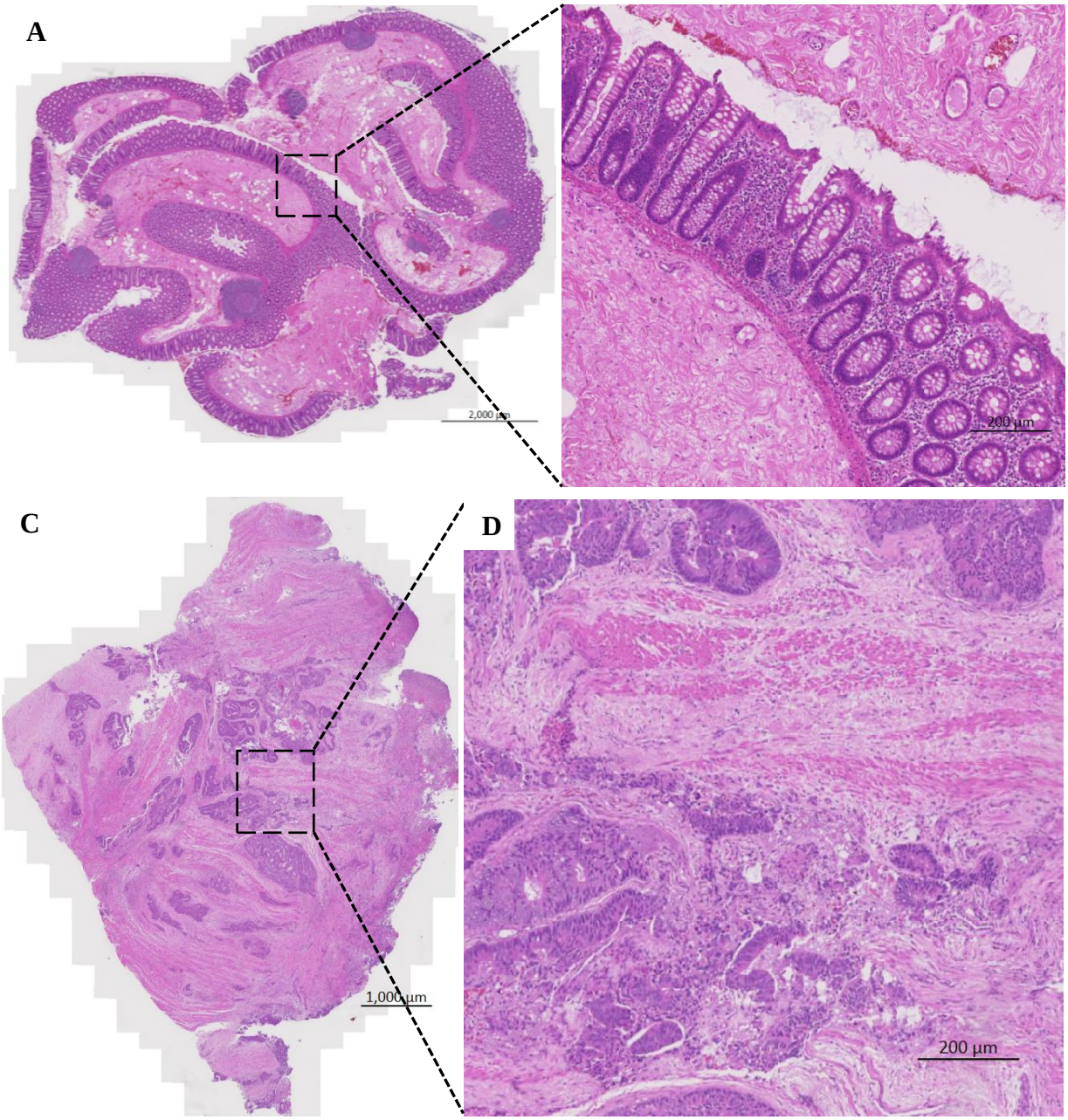


Figure 5.3. Haemotoxylin and eosin stain of colorectal sample 33.

Normal colonic mucosa with submucosa but no muscularis propria is seen in (A). A higher power magnification of some of the colonic crypts is shown in (B); the muscularis mucosa is seen immediately beneath the crypts and the submucosa beneath this – an area of submucosa is also seen in the top right of this image, due the tissue being folded on itself. Tissue from the tumour of the same patient is shown in (C). This section contains entirely invasive adenocarcinoma with no normal colonic mucosa. The cancer can be seen invading the muscularis propria; a high power magnification of the invasive carcinoma is shown in (D).

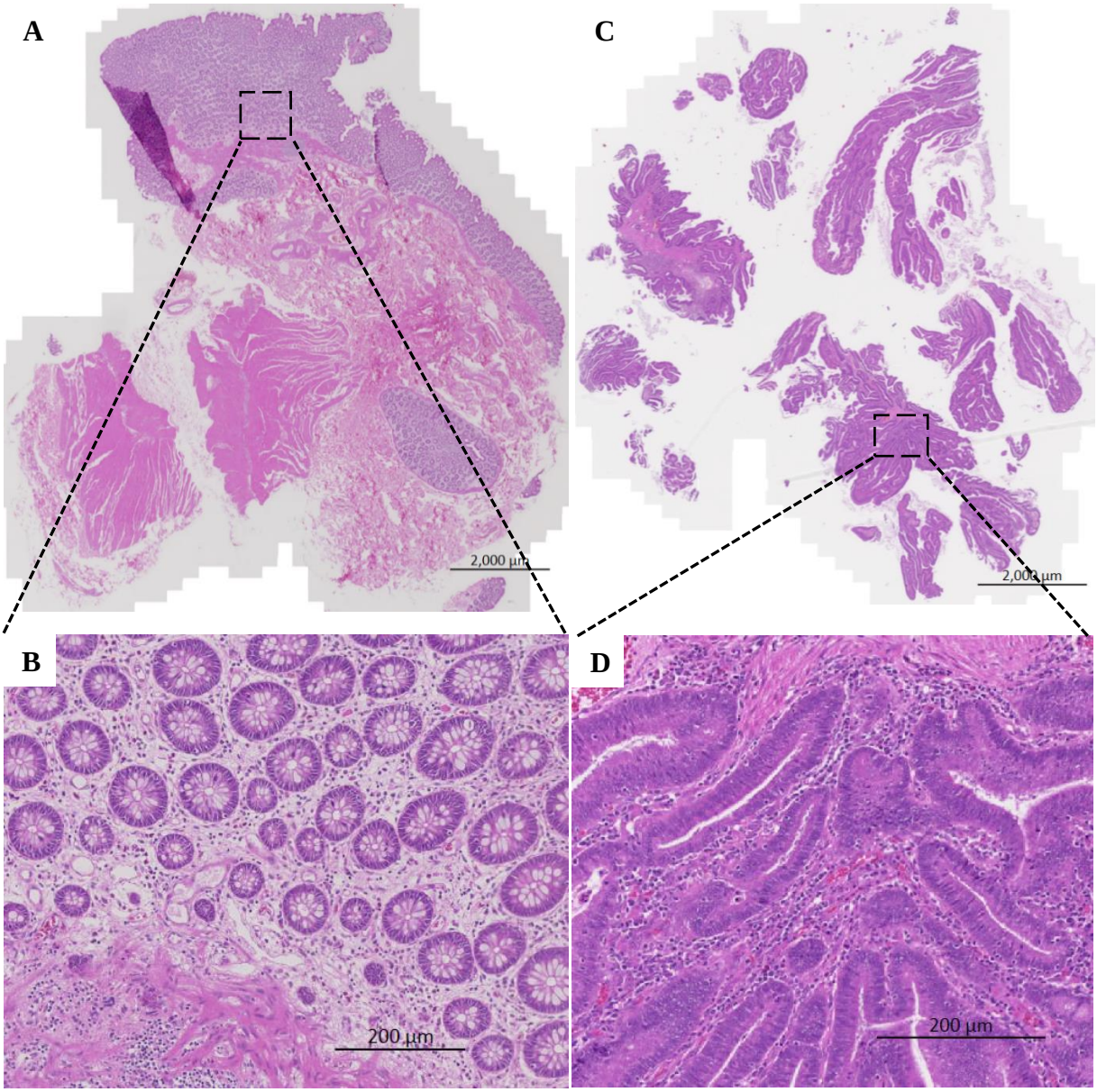


Figure 5.4. Haemotoxylin and eosin stain of colorectal sample 35.

(A) Normal tissue section displaying normal colonic crypts in cross section and an area of muscularis propria in the bottom left of the image. A higher power magnification of the normal mucosa is shown in (B), which also contains muscularis mucosa and a small lymphoid aggregate. (C) Displays tissue obtained from the tumour of the same patient. No normal colonic mucosa is seen but there is also no evidence of invasive cancer in this specimen. Higher power magnification of the fragmented low grade villous adenoma displayed in (B) is shown in (D). The adenoma was part of a caecal tumour that was a moderately differentiated adenocarcinoma (see Table 5.2).

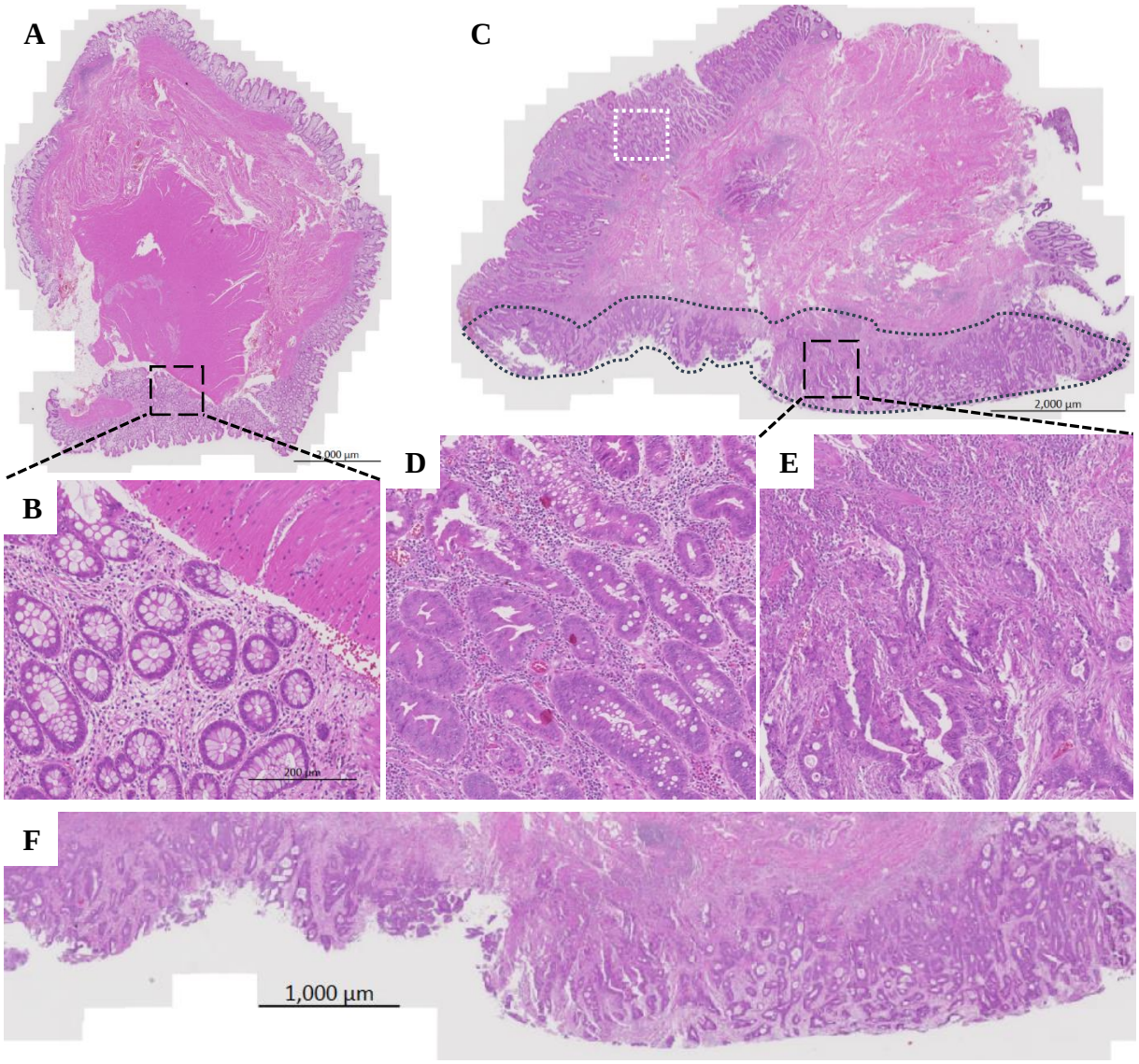


Figure 5.5. Haemotoxylin and eosin stain of colorectal sample 38.

(A) Normal colonic tissue section showing a large rim of normal colonic mucosa and a central section of muscularis propria. A higher power magnification of the normal colonic crypts seen in cross section adjacent to the muscularis propria (B): note, this is not the normal anatomical arrangement but rather a result of the tissue processes technique. (C) Displays a matched cancer tissue sample from the same patient. No normal colonic mucosa is seen but regions of low grade tubular adenoma are seen interspersed by lymphocytes and muscle in the upper portion of the tissue section; a higher power magnification of the adenoma is shown in (D) which corresponds to the white dotted square within (C). The region outlined by a dark blue dotted line is shown in higher magnification in (F) and contains invasive cancer; a higher power magnification of this invasive cancer is shown in (E).

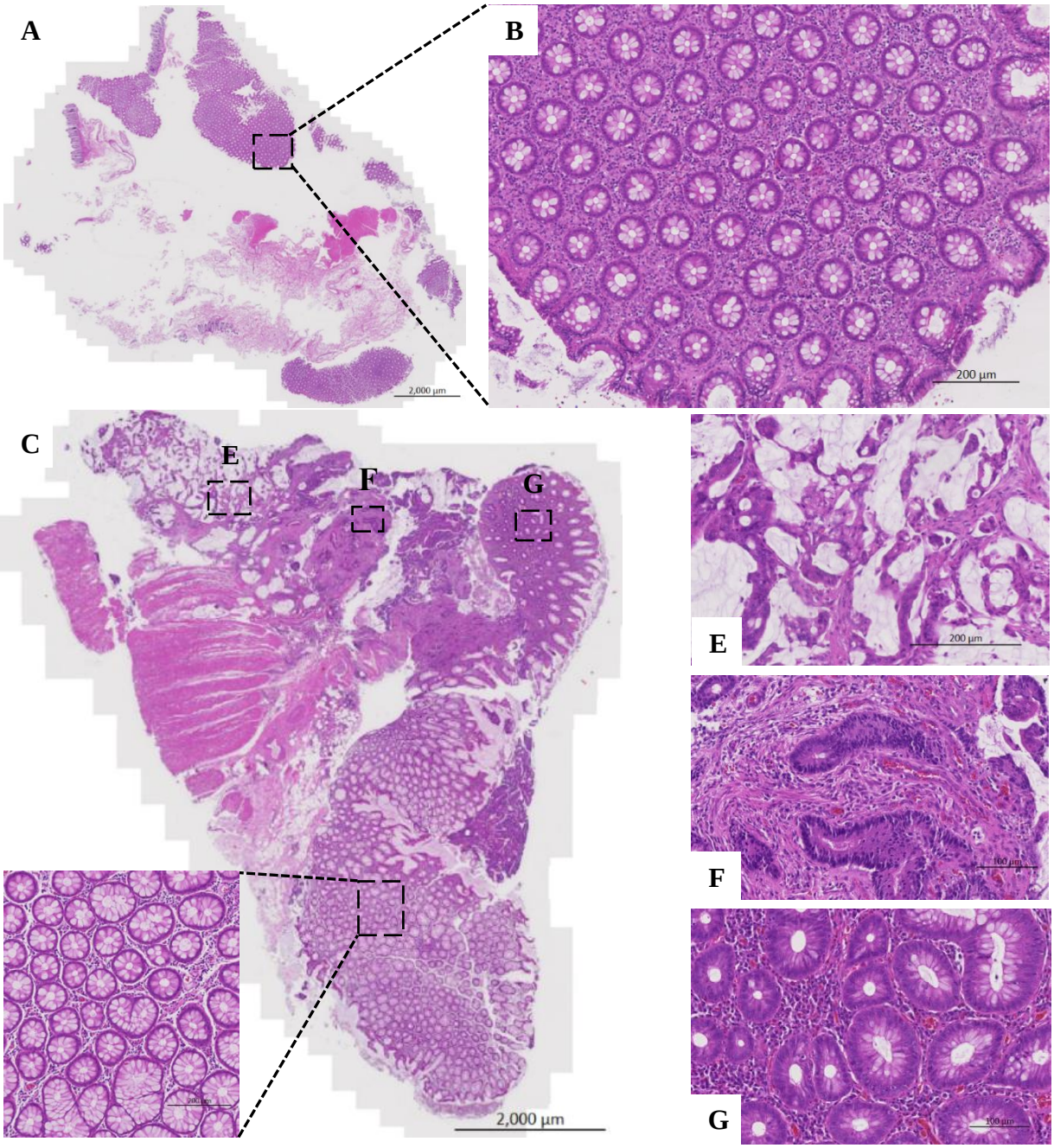


Figure 5.6. Haemotoxylin and eosin stain of colorectal sample 40.

(A) Entirely histologically normal colon sample displaying colonic mucosa, some submucosa and a small amount of muscularis propria in the left central region of this section. (B) High power magnification of the normal colonic mucosa from the tissue section in (A); the normal colonic crypts seen here are shown in cross-section. (C) Matched sample of the colorectal cancer from the same patient: an area of normal colonic mucosa is seen in the lower portion of this tissue section and a higher power magnification is shown in (D). Two distinct areas of carcinoma are shown in the upper central and left portion of the section: mucinous type on the left with a region containing muscularis propria below this; and intestinal type carcinoma in the central portion, both shown in higher magnification in (E) and (F), respectively. An area of low-grade tubular adenoma is seen in the upper right portion of the cancer sample, shown in higher magnification in (G).

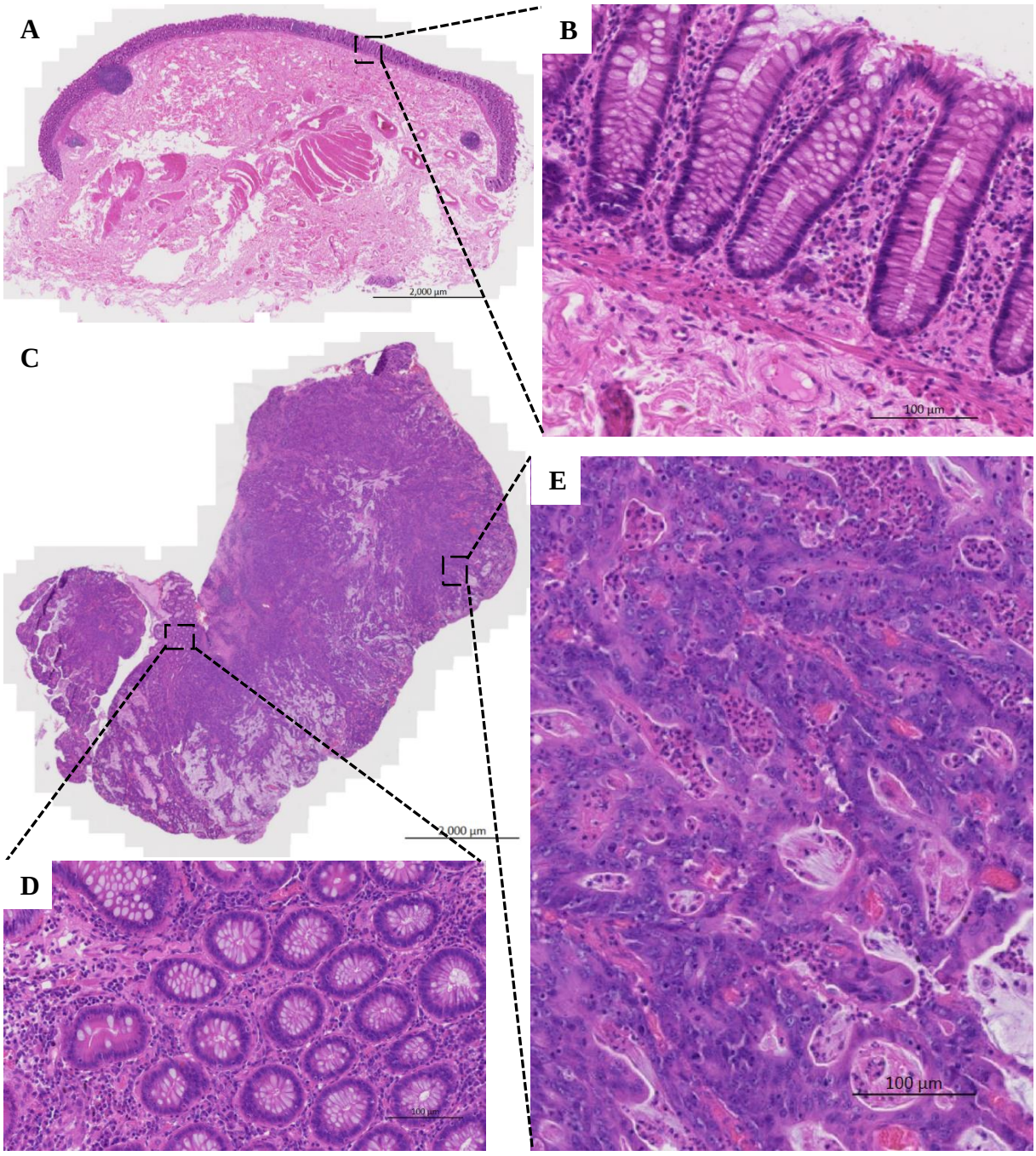


Figure 5.7. Haemotoxylin and eosin stain of colorectal sample 45.

(A) Normal colon tissue section, displaying a rim of normal colonic mucosa shown in higher magnification in (B) and a larger area of submucosa and muscularis propria beneath this. A Peyer's patch is seen in the upper left of the tissue section in (A). A matched section of carcinoma from the same patient is shown in (C). Only a very small area of histologically normal colonic mucosa is seen, displayed in higher magnification in (D). The vast majority of the tissue section in (C) is taken up by an adenocarcinoma of high grade (poorly differentiated), a representative section of which is shown in higher magnification in (E). This cancer is of particularly high grade, displaying few characteristics to guide the tissue of origin (i.e. if the small section of normal colon was removed from the section).

Table 5.2. Pathological and demographic characteristics of sampled colonic tumours

Age	Gender	Case	Site	Differentiation	Duke's Stage	T	N	Positive/ identified	M	EVA
71	M	C1	Ascending	Moderate	B	3	0	(0/10)	X	No
63	M	C2	Ascending	Moderate	C	3	1	(1/16)	X	Yes
67	M	C3	Caecum	Moderate	C	3	1	(2/10)	X	Yes
76	M	C4	Caecum	Moderate	C	3	1	(1/12)	X	No
85	F	C5	Caecum	Moderate	B	3	0	(0/16)	X	No
65	M	C6	Ascending	TVA*	N/A	N/A	0	(0/8)	N/A	N/A
74	M	C7	Ascending	Moderate	B	3	0	(0/7)	X	Yes
85	F	*C8*	Caecum	Moderate	C	3	2	(7/21)	X	No
61	F	C9	Descending	Moderate	B	3	0	(0/4)	X	No
75	F	C10	Sigmoid	No tumour	N/A	N/A			X	N/A
54	F	C11	Descending	Moderate	B	3	0	(0/17)	X	No
69	F	C12	Transverse	Moderate	B	2	0	(0/10)	X	No
78	M	C13	Caecum	X	X	X	X	X	1	X
76	M	C15	Transverse	Moderate	B	3	0	(0/10)	X	Yes
64	F	C16	Ascending	Moderate	A	1	0	(0/16)	X	X
60	M	C17	Transverse	Moderate	B	3	0	(0/13)	X	No
X	M	C18	Caecum	Moderate	C	3	1	(2/20)	X	No
71	M	C19	Sigmoid	Moderate	C	3	2	(7/18)	X	Yes
53	M	C20	Recto-sigmoid	Moderate	C	2	4	(4/12)	X	No
79	F	C21	Ascending	Poor	B	3	0	(0/6)	X	No
72	M	C22	Ascending	Moderate	B	3	0	(0/7)	X	No
78	F	C23	Recto-sigmoid	Moderate	B	2	0	(0/50)	X	No
73	M	C24	Caecum	Moderate	C	3	1	(3/14)	X	No
79	M	C25	Caecum	Moderate	B	3	0	(0/14)	X	No
76	F	*C26*	Ascending	Moderate	B	3	0	(0/6)	X	No
60	F	C27	Sigmoid	Moderate	B	3	0	(0/8)	X	Yes
74	M	C28	Descending	Moderate	A	2	0	(0/11)	X	No
74	M	C29	Caecum	VA**	N/A	N/A	0	(0/2)	N/A	N/A
85	M	C30	Sigmoid	Poor	C	3	1	(1/34)	X	No
81	M	*C31*	Ascending	Moderate	A	2	0	(0/12)	X	No
84	M	*C32*	Ascending	Moderate	B	3	0	(0/7)	X	No
79	M	C33	Ascending	Moderate	B	3	0	(0/13)	X	No
78	M	C34	Caecum	Moderate	B	3	0	(0/3)	C	No
77	F	C35	Caecum	Moderate	B	3	0	(0/18)	X	No
84	F	C36	Descending	Moderate	C	3	1	(3/8)	1	Yes
65	M	C38	Sigmoid	Moderate	B	3	0	(0/8)	X	No
81	F	C39	Appendix	Moderate	B	4	0	(0/9)	X	X
67	M	C40	Transverse	Moderate	B	3	0	(0/4)	X	No
70	F	C44	Caecum	Moderate	B	2	0	(0/14)	X	No
90	M	C45	Ascending	Poor	B	3	0	(0/18)	X	No

*TVA – Tubulovillous Adenoma (high degree of atypia); **VA – Villous Adenoma (severe dysplasia); T – Tumour; N – Node; M – Metastasis; EVA – Extramural Vascular Invasion; X – denotes unknown.

5.2.2. Expression of germline genes in colon cancer investigated by RT-PCR

RNA was extracted from the colonic tissue that had been preserved in RNAlater® solution. This was quantified and an equal amount was reverse transcribed into cDNA. The expression of germline genes, mainly members of the meiCT gene group that we had previously identified were tested using a 40-cycle RT-PCR. As shown in Figure 5.8 the germline genes tested were not measurably expressed in either the normal or cancerous colonic tissue. Other than several of the meiCT genes appearing to be expressed in the SW480 colorectal cancer cell line, there was no strong indication that the genes were transcribed in this cancer type.

Qaixcel gels were used in part as they provided a slightly higher throughput screening for gene expression as compared to RT-PCR screening using agarose gels. They also provide a higher degree of resolution for smaller product sizes. There appeared to be very limited expression of germline genes in the colon cancer samples that I had obtained; a finding which is perhaps not unexpected given what is known about CT gene expression and CRC in the literature (Sammut *et al*., 2013). CRC is considered to have generally low expression of CT genes as opposed to other cancer types such as melanoma or lung cancer (Dakshinamurthy *et al*., 2008; Sammut *et al*., 2013). The melanoma sample tested (see Figure 5.9) did not express either of the germline genes investigated but this was only one melanoma sample and two germline genes not known to be strongly associated with melanoma.

The screening of samples using RT-PCR was time consuming and it was becoming clear that a new approach was needed to increase the throughput of the screening process in the hope of identifying potentially relevant germline genes associated with this cancer type – see Part II of this chapter. *PIWIL1* did emerge as one possible marker in colon cancer as it appeared to be expressed in 6 out of the twelve cancer samples tested but only 2 of the adjacent normal tissues (see Figure 5.9).

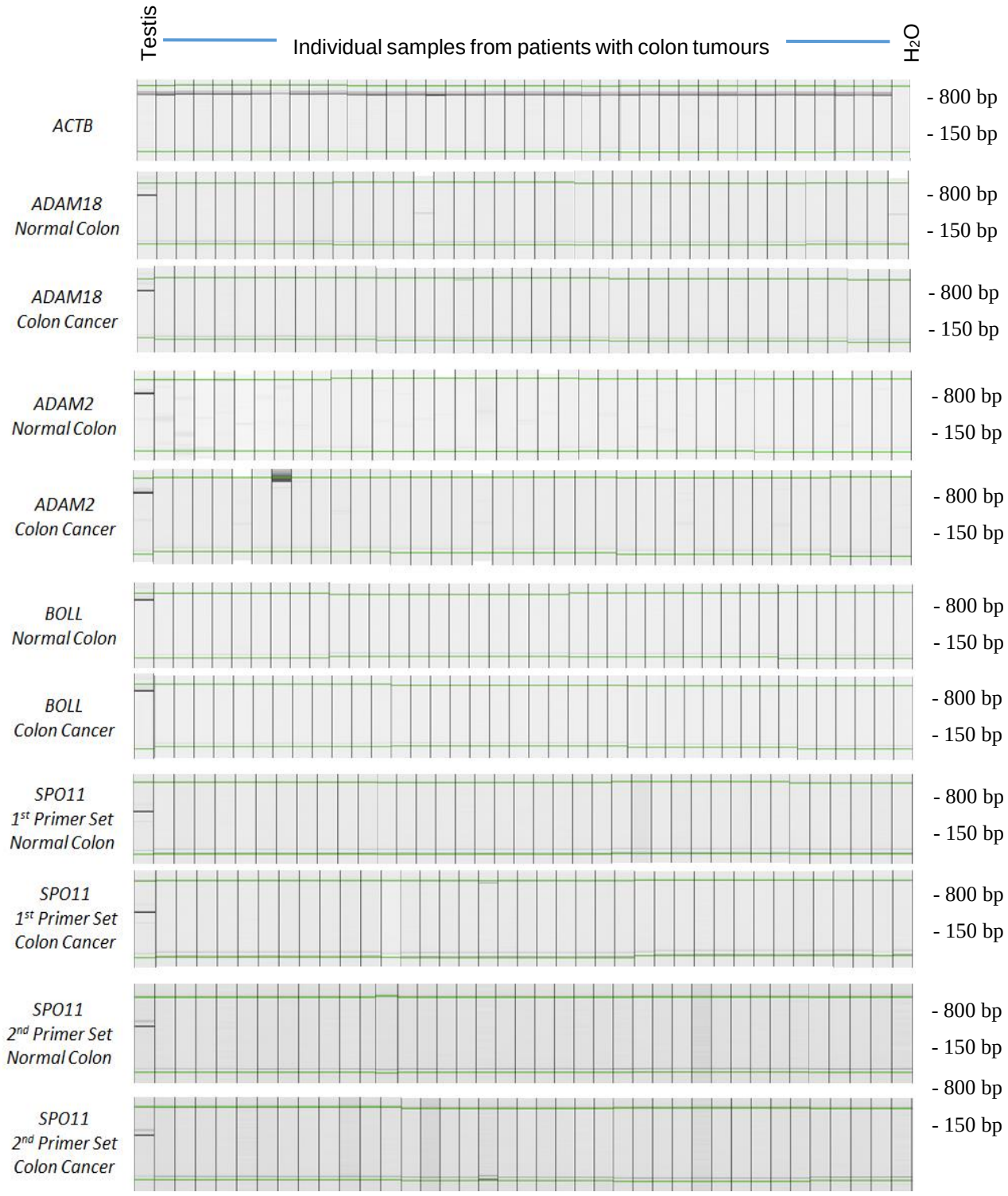


Figure 5.8. RT-PCR of meiCT genes in colon tissues.

RT-PCR was performed for different germline genes using cDNA generated from RNA extracted from colon tumour and normal adjacent tumour tissue samples from the patients listed in Table 5.2. Qaixcel gels shown here were used as a slightly higher throughput method of screening a larger number of samples at the same time. cDNA was generated from RNA extracted from colonic tissue obtained immediately post-operatively. Various meiCT genes were tested and were not shown to be expressed using a 40-cycle PCR amplification. Two separate primer sets were used for *SPO11*. The rows alternate between normal adjacent cancer colonic samples and colon cancer samples. Testis, in the first column, was used as a positive control sample.

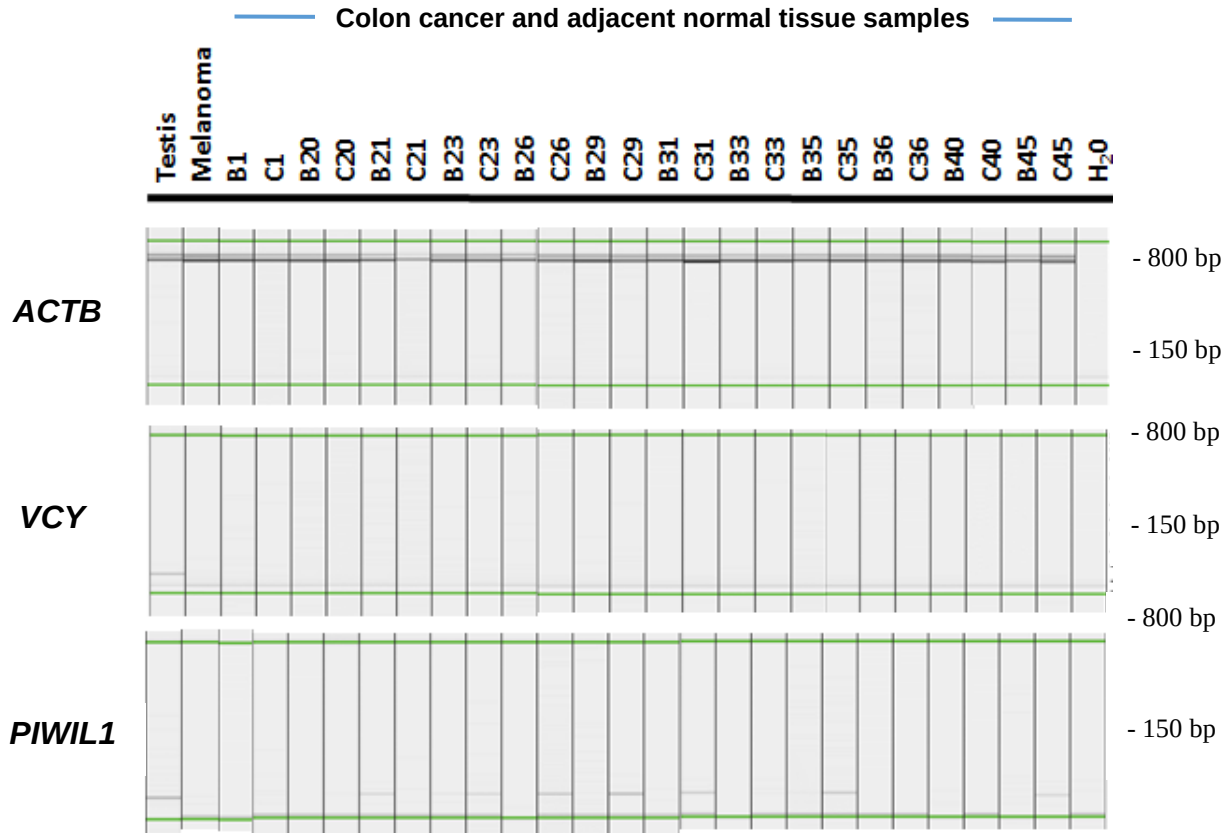


Figure 5.9. RT-PCR of germline genes in colon cancer samples and melanoma.

Qaixcel gels were used to explore the expression, using a 40-cycle PCR amplification. Here two germline genes identified as predictive markers in lung cancer (Rousseaux *et al*., 2013a) were tested using the primers sets used in the paper (see Table 1 in Section 2 of Appendix). Testis, was used as a positive control samples in column 1. cDNA was synthesised from RNA obtained commercially from a melanoma sample and from a selection of the colonic samples I had obtained. *VCY* was found to be expressed in testis only. *PIWIL1* was found in testis but not the melanoma sample. However, some of the colonic samples did express this gene. Most of these samples were from the cancer samples (e.g., C23, C26, C29, C31, C35 and C45) but two of the normal adjacent cancer samples also appeared to express this gene (B21 and B22). Interestingly, in the corresponding cancer sample a PCR product could not be detected. This may be due to lower cDNA content in this sample as the *ACTB* control band was weaker in C21 compared to B21 and other samples. The lower green marker line is at 15 bp and the upper green line corresponds to 1000 bp.

5.2.3. Developing a definitive list of germline-associated genes

Knochbin and colleagues (Rousseaux *et al*., 2013b) provided the first description of how germline genes can be used for identifying patients who had a worse outcome and may benefit from particular forms of therapy. In this paper they identified genes which had expression limited to the testis or placenta. The crucial finding of this work was that the gene expression signature profile alone was enough to identify patients with a higher likelihood of suffering an early demise from lung cancer. The gene expression profile of a limited panel of germline genes (26 in total) could predict an unfavourable outcome independently of accepted stratification methods, such as TNM staging, grade, or cancer type. This has considerable implications for patient management.

We had adopted a different approach to identify genes associated with spermatogenesis, which may have been meiosis-specific or at least be expected to be germline in origin (Feichtinger *et al*., 2012b; Sammut *et al*., 2014). Although there was considerable overlap in the approaches used, there were several of our genes that were not present on the Knochbin group panel of germline genes and *vice versa* (see Appendix B, on CD for complete lists).

These differences, together with the increasing evidence supporting a major role of germline factors in cancer both as predictive markers and oncogenic drives, provided a rational to develop a definitive list of germline genes. There is no agreed or definitive list of germline genes in the literature. Some of the genes will be expressed and have roles in somatic cells as well as those of the germline, whereas others will have expression (and function) limited entirely to the germ cells. When expression is limited in somatic cells, this forms the basis for the identification of CT genes, which can also be considered cancer/germline genes.

With this aim in mind, five sources from the literature were compiled to produce what we believe at present to be the closest to a definitive list of germline-associated genes. These five sources were:

- We used the original list of gene candidates that had been identified as orthologues of mouse spermatocyte specific genes (Feichtinger *et al*., 2012b). This list comprised 400 genes, many of which had been tested for expression in a range of normal tissues as well as cancer tissue/cell lines and from which the meiCT genes, of which 85 have now been

described, were identified (Feichtinger *et al*., 2012b; Sammut *et al*., 2014). Rather than limit ourselves to the 85 meiCT genes, for the purpose of compiling the 'definitive' list I included all the orthologues that had originally been identified, although we knew that a significant proportion of these genes were expressed widely in somatic tissues and perhaps only had upregulation in the germline.

- The second list of genes was the testis and/or placental genes that were described by the Knochbin group (Rousseaux *et al*., 2013b). This comprised 506 genes, of which 439 were testis-specific and 67 were associated with the placenta.

- At the time of compiling the list 285 CT genes had been accepted and compiled on 'CTDatabase' (Almeida *et al.*, 2009). I did not attempt, by manually searching the literature, to include other genes that had been proposed as CT genes but not corroborated by others.

- Janic and colleagues reported 49 germline genes to be over-expressed in *Drosophila melanogaster l(3)mbt* tumors and may in fact be oncogenic drivers for these tumors (Janic *et al*., 2010). 28 human orthologues of these genes were identified; the list of genes totalled 48 when all paralogs were included (Feichtinger *et al*., 2014a). This paper provided support to the notion that a soma-to-germline transition exists during the development of many human cancers.

- Finally, Hoffman and colleagues described 53 genes as CT genes based on their expression profile which they validated as restricted to the testis and brain in normal tissues; the majority of these genes were X-encoded (Hofmann *et al*., 2008).

When presenting the data on gene expression below (see Section 5.3.1) I have categorised the final list of germline genes tested according to the sources just described. These are the groupings used – it should be noted that several of the genes selected for gene expression analysis appeared in more than one of these groups:

- Group A – genes previously categorised as meiCT genes (Feichtinger *et al*., 2012b; Sammut *et al*., 2014)

- Group B – CT genes identified as testis/brain restricted (Hofmann *et al*., 2008)

- Group C – all human orthologues of the mouse spermatocyte/meiosis-associated genes (Chalmel *et al*., 2007; Feichtinger *et al*., 2012b)

- Group D – all human orthologues of the *Drosophila melanogaster* germline genes, which drove oncogenesis in a brain tumour model (Feichtinger *et al*., 2014a; Janic *et al*., 2010)

- Group E – known CT genes according to CTdatabase (Almeida *et al*., 2009)

- Group F – all the germline-specific genes, identified as testis-specific or placenta-specific by the Knochbin (Rousseaux *et al*., 2013b)

- Group G – the cohort of 26 genes taken from group F that were predictive of a poor prognosis in lung cancer patients (Rousseaux *et al*., 2013b).

We attempted to conduct a systematic review but this did not identify additional relevant papers. Moreover, not all the above papers included for the list generation were identified through this approach. We acknowledge that our approach here is open to selection bias but it remains as close to a definitive list of potential germline-specific genes as we are aware of in the literature. Further details are provided in Appendix B in the folder entitled "Germline gene lists with source", on the CD.

5.2.3.1. Validation of germline gene expression of a panel of genes identified as predictive markers in lung cancer

These five sources of germline genes produced a list of 1292 genes; this number was reduced to 1060 once duplications had been removed. The degree of overlap between the different sources varied somewhat – full details are provided in Appendix B (supplied on CD). Just over half of the meiCT genes which we had identified appeared in the germline gene list compiled by Knochbin's group and all were from the testis-specific, not placental gene list (Rousseaux *et al*., 2013b). There was a much higher degree of overlap proportionally between the meiCT list and the Knochbin group gene list (see Appendix B, on CD) than with the original list from which the meiCT genes were derived. There was, however, no overlap between the final 26 genes which they identified as independent predictors of a worse outcome in lung cancer patients and the meiCT genes. In fact only 4 of the 26 genes appeared on the original list of genes which our group had identified as orthologues of mouse spermatocyte-specific genes. Partly for this reason, I validated the expression profile of the 26 Rousseaux genes using RT-PCR in the normal panel of tissue that we had used to validate the expression profiles for the meiCT genes (i.e. as in Chapter 3). In their paper they described the genes as having tightly restricted expression profiles to the germline (Rousseaux *et al*., 2013b). As shown in Figure 5.10 and Figure 5.11, not all the genes appeared to have such tightly restricted tissue expression in our panel of normal tissues.

We would have excluded over one quarter (7 of the 26) of the Knochbin group panel of genes from further analysis. This shows how varying degrees of stringency can impact on the findings as we may have been excluding clinically relevant genes due to our more stringent approach. It also perhaps raises questions about the normality of our 'normal' panel. The tissues were obtained post-mortem and it is possible that the individuals had an undiagnosed cancer prior to their death, though given how extensively the 7 Knochbin group genes were represented in the normal panel, this possibility is unlikely to affect the overall conclusion. It would be interesting to see if the 7 genes were excluded from the Rousseaux list, whether the remaining 19 genes would still hold predictive power of a worse outcome in lung cancer.

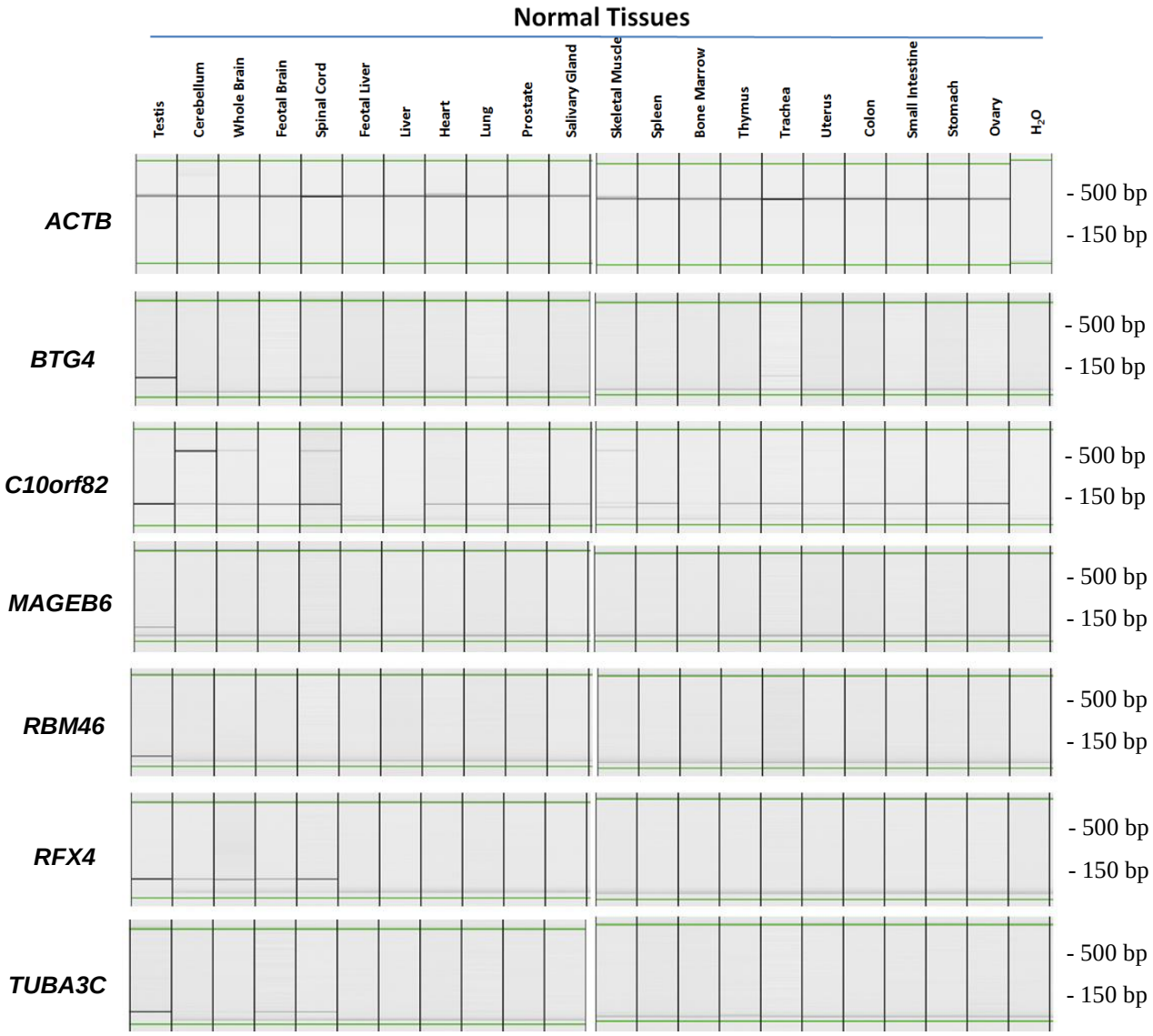


Figure 5.10. RT-PCR of genes identified as predictive markers in lung cancer in a panel of normal tissues.

A 40-cycle amplification in a panel of normal tissues was conducted using the PCR primers used in the study by Knochbin and colleagues (Rousseaux *et al*., 2013b). The higher resolution of the Qaixcel gels used for this allowed identification of the PCR bands in the majority of cases; testis was used as a positive control (far left column). Several genes, for examples *MAGEB6* and *RMB46* shown here were restricted to the testis only. *RFX4* was restricted to the testis and central nervous system and would have passed through the validation/screening process that we employed for the meiCT genes but classified as testis/CNS-restricted and not testis-restricted. The testis/CNS-restricted expression was also confirmed using a separate assay/primer set on the TLDA analysis – see Figure 5.22. Other genes, for example *C10orf82*, would have been excluded as it appeared to be widely expressed in the normal tissue panel and again this was confirmed on the TLDA analysis.

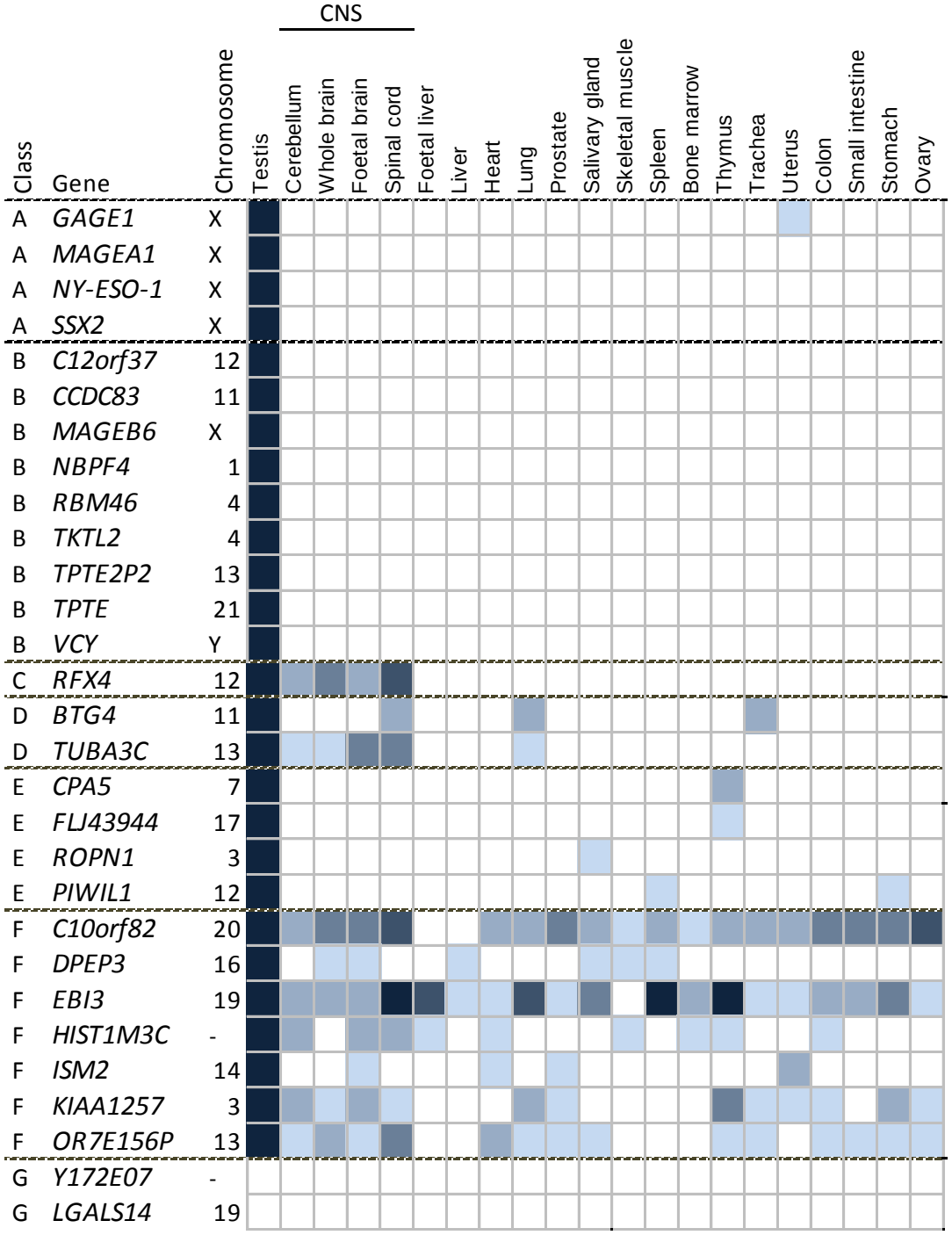


Figure 5.11. Grid representation of the gene expression profiles for germline genes that had been shown to be associated with a worse outcome in lung cancer patients.

The primers used for PCR amplification were taken from the paper identifying this cohort of genes (Rousseaux *et al.*, 2013b). RNA was obtained from commercial sources for a range of normal tissue types and reverse transcribed; the cDNA was subject to 40-cycle PCR amplification. Two of the genes did not produce a visible product on the Qiaxcel gel (category G). 7 genes appeared to be widely expressed across the tissue panel (category F). A further 7 were testis-selective or testis/CNS-restricted (see categories C-E) and the remaining genes appeared to have expression restricted to the testis (category B). The category A genes above are known CT genes that had previously been screened in the same samples and included here for comparison purposes. Results can be compared directly to Figure 5.22, which acts as a further validation of these findings through similar experimentation using a different approach and assays.

Despite my finding that some of the predictive germline genes identified by the Knochbin group (Rousseaux *et al*., 2013b) appear to be expressed in somatic tissues, we did not exclude them from consideration and kept them all within the list of germline-associated genes at this stage. Using the CancerMA pipeline we investigated how many genes displayed a meta-upregulation in any cancer type included in the pipeline. Around 20% of the germline genes were not present on the arrays included for this analysis. The definitive list of germline genes included genes with Hs.xxx and LOCxxx identifiers (these are generally transcribed regions of the genome but without known gene function or assigned name). The data from which CancerMA was based did not consider genes with these number identifications so it is not surprising that a significant proportion were not present on the arrays. CancerMA can only deal with ~300 genes of interest at a time, so challenging of array data with the complete list produced 4 separate circos plots (for example, see Figure 5.12). See also Figures 1-3 in Section 3 of the Appendix for the additional circos plots analysed in the same fashion. Brain and ovarian cancers were the most common cancer types displaying statistically significant meta-upregulations. When considering only the 'meiCT' genes identified by the McFarlane group (Feichtinger *et al*., 2012b; Sammut *et al*., 2014), ovarian cancer was by far the most common cancer type displaying meta-upregulation of these genes (see Figure 4 in Section 3 of Appendix).

Over 250 CT genes, from over 100 separate gene families, have been identified although not all of these meet the strict CT gene expression criteria (Almeida *et al*., 2009; Hofmann *et al*., 2008). This is consistent to some degree with the findings presented here (i.e. from TLDA results). We chose not to include genes which Hofmann (2008) reported as not testis or testis/CNS restricted, though many of the testis-selective group were included from the list taken from 'CTDatabase' (Almeida *et al*., 2009).

Around half the germline genes displayed at least one statistically significant upregulation in gene expression when the microarray datasets were meta-analysed (for examples, see Figure 5.12). After feeding all the germline genes through CancerMA, I then challenged the genes displaying meta-upregulations in any cancer type against CancerEST. I filtered for EST signatures restricted to immunologically privileged tissues (i.e. testis, placenta, brain and other CNS tissues) and then additionally cancer. This resulted in a list of 78 genes – see Figure 5.13 and Appendix B (on CD). Further explanation of the degree of overlap and how the list of candidate genes was generated is provided in the folder "germline gene lists with source" in

Appendix B, on the CD. No claim is made with regards to the clinical or biological significance of the gene lists generated (e.g., in Figures 5.12 and 5.13) – indeed as highlighted in Chapter 3 and Figure 20 in Appendix, some of the upregulations may be stochastic in nature. One must be careful not to over-interpret the findings but the selected genes provided a rational approach on which to focus down on a more practical and manageable list on which to base initial gene expression analyses. Ideally, all genes would be taken forward but this was not deemed practical for the purpose of this research, limited in part by the costs of screening over 1000 gene targets. A further limitation of this research and the results of the TLDA analyses is that the results were not duplicated (i.e. n=1 for the majority of the following experiments). We accept this is a major limitation but the purpose of this research was a 'first look' exploratory study. We included control genes based on previous experiments and no internal validation was performed for our sample population. This is also a weakness to some degree but the quantitative nature of the TLDA experiments was an added bonus rather than purpose of the research. We were interested primarily in "on/off" answers (i.e. was a given gene switched on in a certain cancer but not expressed in normal tissue). Such "on/off" gene signatures may ultimately prove to be useful therapeutic targets or in disease stratification due to lack of presence in normal tissues. We included testis as a positive control for the experiments, as all genes should have been switched on in normal testis. As shown in the Appendix B (on CD – file entitled "TLDA repeat for Testis Sample"), the single repeat performed on a separate day for the positive control sample, displayed very consistent results. This is reassuring and suggests there is not a high rate of false positive or false negative results. However, it is important that additional repeat experiments are conducted before further exploratory work is conducted.

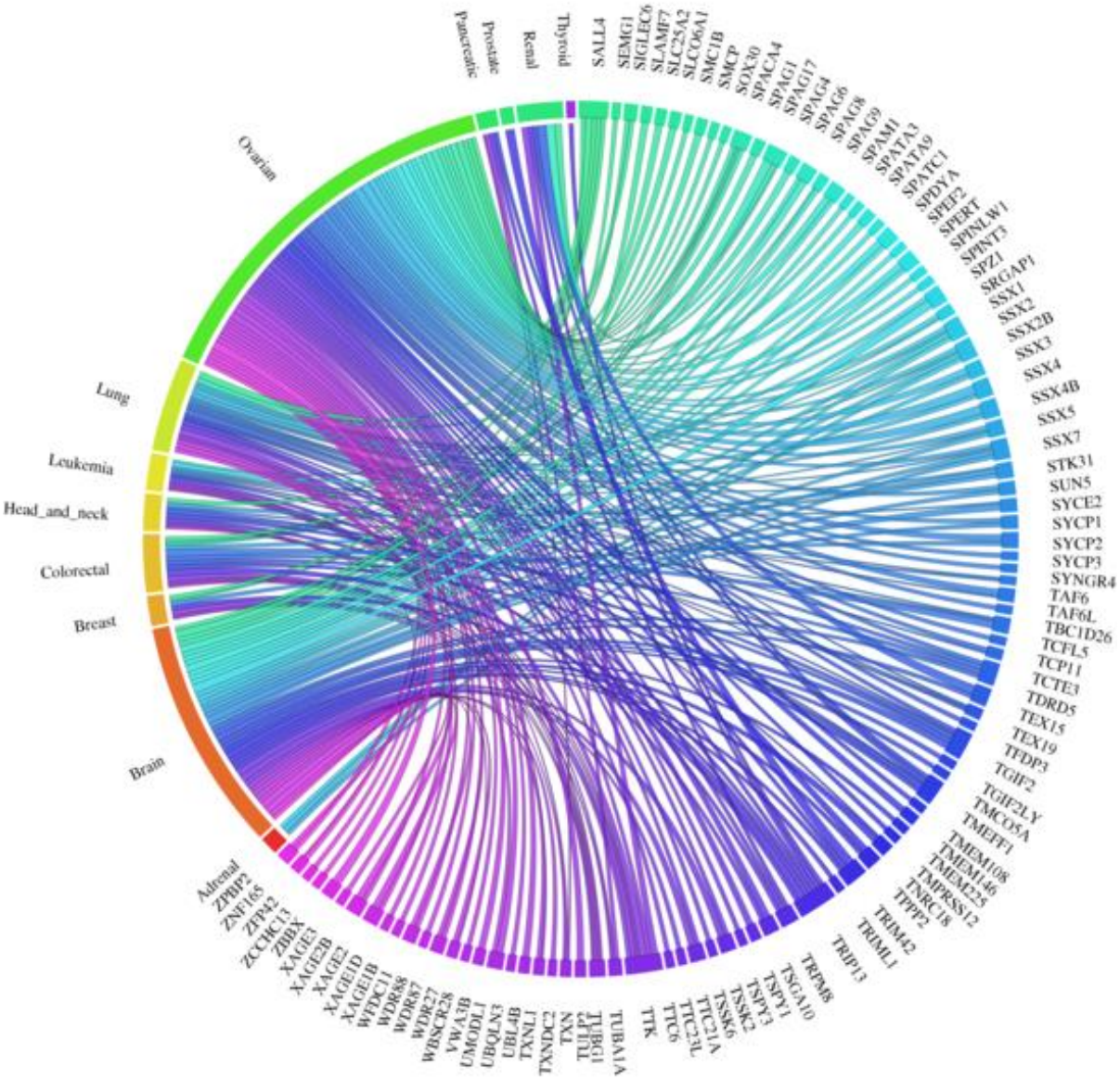


Figure 5.12. Germline genes displaying meta-change upregulation in cancer.

The definitive list of germline genes (comprising over 1000 genes) were assessed using CancerMA (Feichtinger *et al*., 2012a). The website is limited to ~300 targets on a single run, so the results for a selection of the germline genes are shown here; genes beginning with the letter S through to Z are included. The picture was broadly similar for the other germline genes when challenged against the same pipeline (see Figures 1-3 in Section 3 of the Appendix). Brain cancer and ovarian cancer feature most commonly as cancer types displaying a statistically significant upregulation in gene expression when the microarray datasets were meta-analysed. Lung cancer was the third most common cancer type displaying meta-upregulations and colorectal cancer the fourth most common.

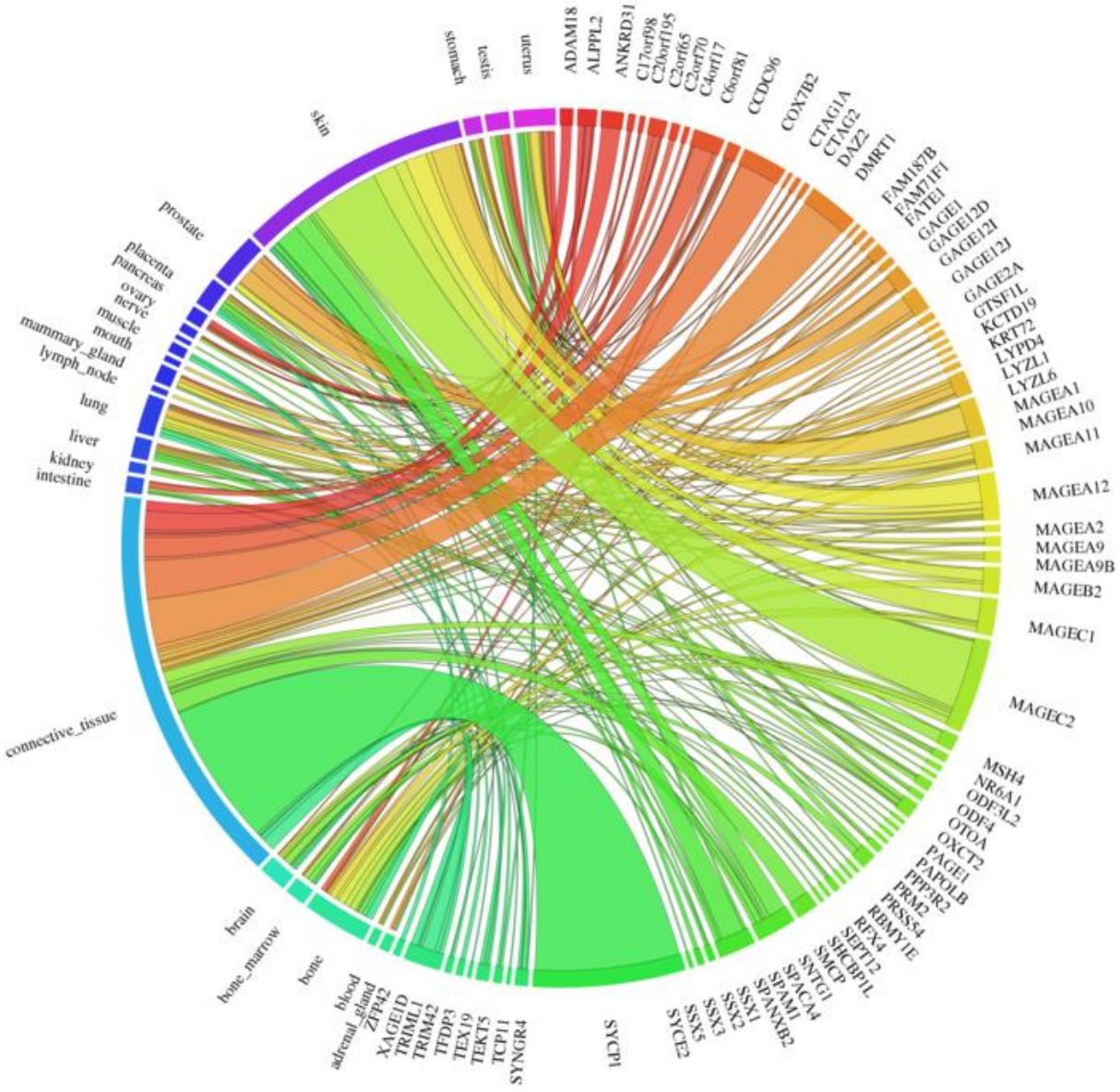


Figure 5.13. Circos plot showing genes with EST signatures restricted to immunological privileged tissues and cancer.

The germline genes that were meta-upregulated in CancerMA were then challenged against CancerEST and the genes with signatures limited to immunologically privileged tissues as well as cancer are displayed here. The thickness of the lines is representative of the magnitude of the expression found for a given gene in a given tissue.

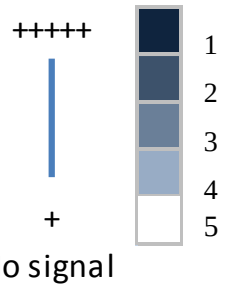
5.3. Part II – Results and Discussion

5.3.1. TLDA analysis of germline gene expression in human tissues

Using a combination of approaches I identified 192 targets that could be analysed using Taqman Low Density Array (TLDA) technology. The full list of targets and the TLDA assays used for each target is provided in Appendix B (on CD). If a gene appeared on more than one germline gene list (see Section 5.2.3 and Appendix B on CD) then this increased the likelihood of selection. I also attempted to include the majority of the meiCT genes as well as the 26 targets that Knochbin and colleagues had identified as predictive markers in lung cancer (Feichtinger *et al*., 2012b; Rousseaux *et al*., 2013b; Sammut *et al*., 2014).

Based on the Ct values for each sample I assigned the gene expression as followed:

- Group 1 – highest expression (Ct value <23)
- Group 2 – moderate-strong expression (Ct value 23-26)
- Group 3 – moderate expression (Ct value 26-29)
- Group 4 – weak-moderate expression (Ct value 29-32)
- Group 5 – not expressed (Ct value >32)



The Ct value cut-off of 32 was based on advice given by Life Technologies on when gene expression can be considered positive. Due to the TLDA microfluidics, a Ct value of 32 is equivalent to a Ct value of 35 seen during most standard qPCR experiments. Such borderline results should be subject to more rigorous testing to establish if the genes are truly expressed. To provide as clear a representation as possible of whether the gene was expressed we chose the more stringent cut-off of 32 recommended by the company, especially as the samples were not run in triplicate as would normally be the case for qPCR experiments. The colours in Figures 5.15 to 5.20 depict these different levels of expression, as shown in the key above – the darker the shade of blue, the stronger the level of gene expression. An example of the readouts from the TLDA cards from which these Ct values are based is provided in Figure 5.14. The list of Ct values is provided in Appendix B (on CD). Figure 5.15 is a heatmap of all the gene expression data obtained in a range of normal and cancer tissues. RNA extracted from melanoma, lung and ovarian cancers were obtained from Origene – information supplied by the company for these samples is provided in Appendix C (on CD).

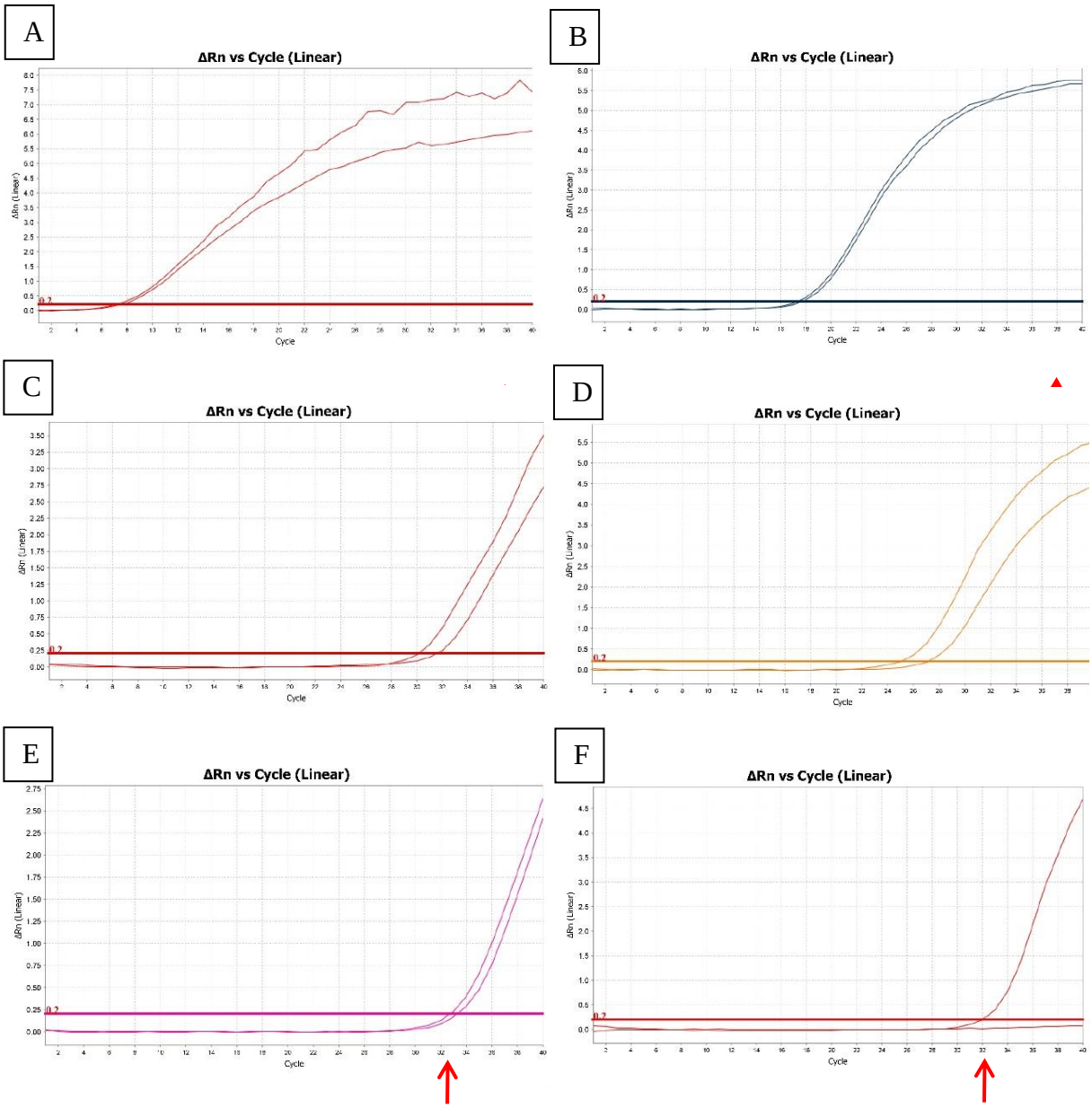


Figure 5.14. Gene expression of six target genes in colon cancer and adjacent normal tissue.

This is an example of the TLDA analysis for colon sample 26. The red arrows indicate the cutoff Ct value indicating definite gene expression. The upper two plots are two endogenous control genes: (A) *18s*; (B) *RPS13*, which are both highly expressed. The Ct values for both control genes are very similar indicating equal amounts of cDNA were loaded onto the card. The amplification plot for *C20orf195* is shown in (C); in this case the curve shifted to the left for the cancer sample compared to the normal sample. However, both Ct values fell within the same range: categoriesed as weak expression (Ct between 29 and 32). (D) is the amplification plot for *CEP55*, which again indicates higher expression in the cancer sample but this gene is expressed more strongly than for *C20orf195*: both samples in this case are categorised as moderate expression (i.e. Ct between 26 and 29). (E) is the amplification plot for *C20orf201* – here the Ct value for the cancer sample falls just below the cutoff but for the normal sample the Ct value is above the cutoff, which meant in this case it was categorised as not expressed. (F) is the amplification plot for *IQCF1* – here the gene has weak expression in the cancer sample but is not expressed in the normal adjacent tissue and the curve fails to cross the threshold even up to 40 cycles.

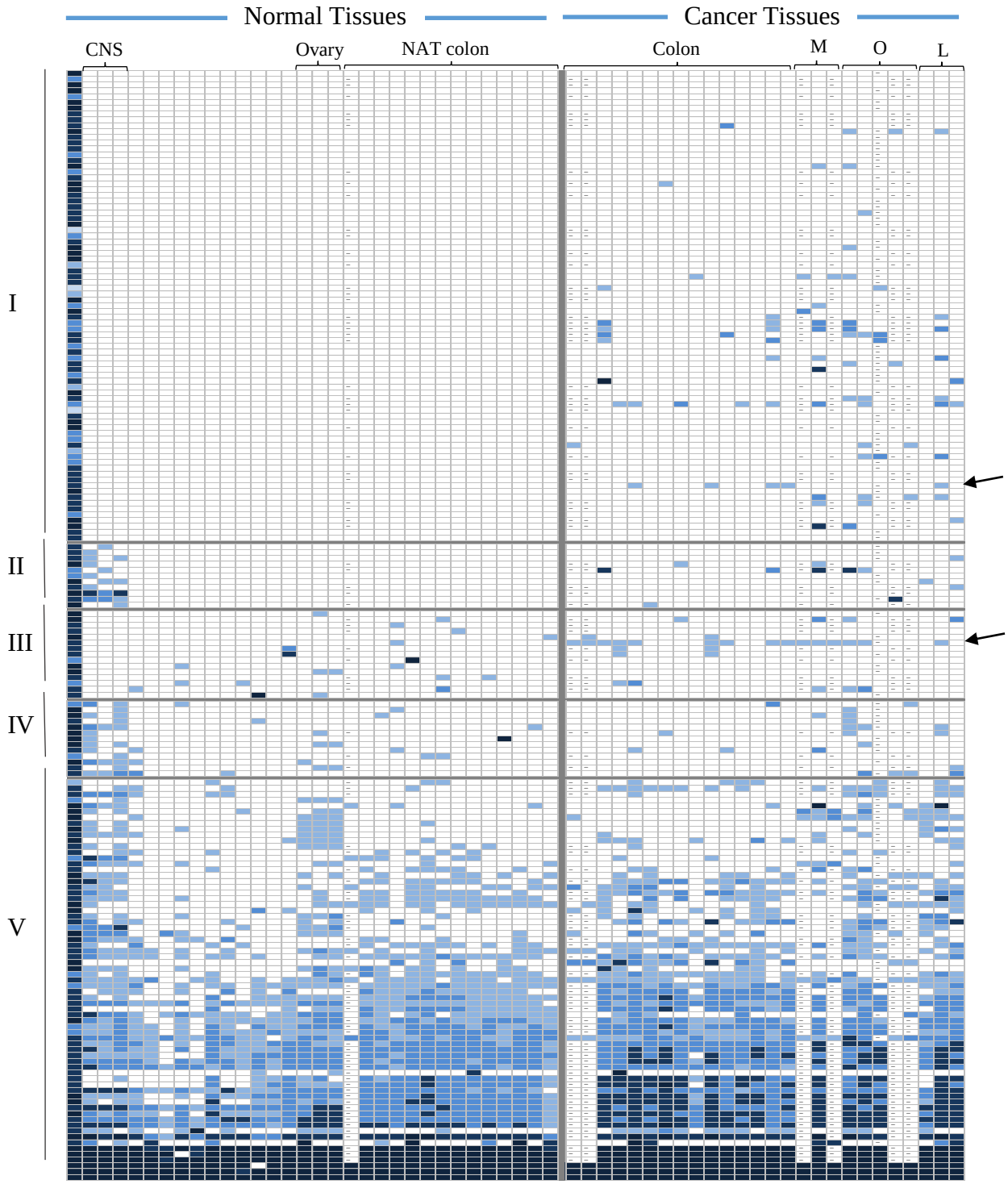


Figure 5.15. Heatmap of the expression of germline-associated genes in human tissues.

The thicker grey lines divide the normal from cancerous tissues and/or the categories the genes were grouped into. The genes generally and as expected had moderate-high levels of expression (darker blue) in the testis (far left column). The heatmap is ordered according to expression levels in normal tissues. Around half had expression restricted to the testis in normal tissues: group I. Group II genes had testis/CNS restricted expression; group III genes had testis-selective expression pattern; group IV genes had testis/CNS-selective expression; whereas group V were expressed in three or more normal tissues excluding the testis and CNS tissues. These groups are ordered according to the strength of gene expression seen and/or alphabetically. A "-" within a cell means that gene expression for that sample was not tested for the corresponding gene. The arrows mark results for two *TEX19* assays – see main text for discussion. NAT – Normal adjacent tumour; M – Melanoma; O – Ovarian; L – Lung.

5.3.2. Co-expression of germline genes in colorectal cancer samples

Colon cancer samples 20, 28 and 40 (see Table 5.2) expressed several of the germline-associated genes. When considering group I-IV genes in Figure 5.15 (i.e. the more promising CT gene candidates), sample 20 expressed 8 of the germline genes and colon cancer sample 40 expressed 10 of the genes. Typically, the colon cancer samples only expressed 2 or 3 genes from these groups and in a couple of cases none of these genes appeared to be expressed. Sample 20 was from a patient with more advanced stage disease that had already metastasized to the regional lymph nodes (Duke's C cancer). Although sample 40 was not a metastatic tumour, interrogation of the H+E stain for the sample collected from this patient revealed a region within this adenocarcinoma to have a mucinous component (see Figure 5.6). Mucinous adenocarcinomas of the colon generally carry a worse prognosis and are resistant to certain forms of cancer therapy (Lee *et al*., 2013). Demethylation of DNA may be a common feature leading to the activation of multiple genes in these tumours (Adair and Hogan, 2009; Shen *et al*., 2007) or alternatively the activation of a certain oncogene (or genes) acts as a catalyst for the activation of others. Whether the activation of these germline genes is correlated with an unfavourable outcome in CRC would need to be established by a more focussed investigation in a larger cohort of samples.

Colon cancer sample 28 expressed 5 of the germline-selective genes (i.e. groups I-IV); this cancer also expressed *PIWIL1* strongly. This was an early stage cancer (Duke's A) and there was no evidence of either metastasis to the regional lymph nodes on the pathological staging or distant metastasis at the time of surgery on the radiographic staging. Patients with Duke's A tumours have a 5-year survival of over 90% and their long-term prognosis is good. There was no indication on any of the pathological characteristics that this patient should have an adverse outcome. However, on a follow-up computerised tomography scan the patient was discovered to have an adrenal metastasis. Two years after their initial surgery, the patient remains alive. It is possible that the expression of these half-dozen germline genes could have predicted this unforeseen outcome, in a similar fashion to how 3 or more of the panel of 26 germline genes predicted an unfavourable outcome in lung cancer patients in the Knochbin study (Rousseaux *et al*., 2013b). It is possible that adjuvant (e.g., chemo- or immuno-) therapy would have reduced the chances of such recurrence/metastasis occurring and the germline genes could be the basis of tailored personalised therapy for such patients.

5.3.3. General comments and differences to previously published work

With the exception of the normal adjacent tumour (NAT) samples, the other normal tissues were the same as those used in the PCR experiments in Chapter 3 and subject to the same limitations of using post-mortem samples as highlighted previously. All the genes were selected from a cohort that was associated with the germline. It may have been expected that most should have displayed testis-restricted or testis-selective expression. One of the striking features when looking at the heatmap above (see Figure 5.15) was the fact that a significant cohort of the genes did not exhibit a testis-specific expression pattern; approximately half fell into the tightly restricted category. Around one third of the germline genes selected were expressed in more than 2 normal tissue samples excluding the CNS, although a couple of these genes were restricted to two or less normal tissue types. This is due to testing three normal ovarian tissues and several NAT colon samples, which collectively only represent two tissue types. We would have discarded the majority of these genes from further analysis if we had seen this expression profile on RT-PCR. The final 20% were testis-selective in their expression profiles (see groups II-IV in Figure 5.15). Group II genes were expressed in CNS tissues only in the normal panel as well as normal testis. Group III genes were expressed in up to 2 normal tissue samples and not in the CNS. Finally, group IV genes were expressed in testis and CNS and in addition up to a maximum of 2 other normal tissue samples.

5.3.3.1. Germline-associated genes with expression in normal tissues

Some of the findings were somewhat contradictory to what has been described previously in the literature. For example, *CCDC79* has been described as a meiosis-specific gene with an expression profile limited to male and female germ cells (Daniel *et al*., 2014). Moreover, we validated this gene previously as a meiCT gene and it was included in the germline gene list developed by the Knochbin group (Feichtinger *et al*., 2012b; Rousseaux *et al*., 2013b). Although this gene was not widely expressed across the tissue samples tested, it was shown to have high levels of expression in a selection of normal and cancer tissues, including normal liver, ovary and colon samples (see Figure 5.16, category ACF). It may be that the primers lacked specificity for this gene, or alternatively the gene is not as germline-restricted in its expression profile as previously thought. In general, however, it was reassuring to see that a large proportion of the genes (~70%) did have a limited expression profile in the normal tissues. Moreover, some of the genes that did not have a testis-restricted or selective expression pattern in normal tissues are known to be expressed in somatic tissues.

Rousseaux and colleagues described 506 genes that were said to be germline-specific: 439 were testis-associated and 67 were placenta-associated (Rousseaux *et al*., 2013b). As depicted in Figure 5.21 the majority of these genes were not widely expressed in normal tissues, though *EBI3* stood out as a gene that was widely expressed. This gene we had shown separately to be expressed widely in normal tissues using the primers that Rousseaux *et al.* provided in the supplementary materials for this study (see Figure 5.10), confirming a degree of consistency between the experiments. Thus, it may be that the different experimental approach adopted by Rousseaux and colleagues did not adequately filter out all somatically expressed genes.

TGIF2 and related genes are discussed below in Section 5.3.5.1. Some other examples of genes that had high and/or frequent expression within the normal tissues tested on the array were *TUBA1A* and *RPS5*. These genes were mapped as orthologues of mouse and fly genes respectively, being associated with spermatogenesis in mice or brain tumour development in *Drosophila melanogaster* (Feichtinger *et al*., 2012b; Feichtinger *et al*., 2014a). Both genes have recently been described as useful normalisation genes for qPCR in specific tissues (Mehta *et al*., 2015; Park *et al*., 2013), which is consistent with the expression profile displayed across all the tissues tested in this study. Again and as for *TGIF2*, these genes may have been inappropriately mapped as human orthologues of the respective *Drosophila melanogaster* germline genes.

TRIP13 was identified as an orthologue of a mouse spermatocyte-associate gene (Feichtinger *et al*., 2012b). Although it is required for HR in meiosis and interacts with other meiosis-specific proteins such as HORMAD1, the gene is also expressed in somatic tissues (Wang *et al*., 2014a; Wojtasz *et al*., 2012). This is again consistent with the expression profile demonstrated in this study (see Figure 5.17, group C). There was generally higher expression of the gene in cancer samples, which is also consistent to what has been shown in various malignancies (Banerjee *et al*., 2014; Larkin *et al*., 2012). The expression levels of *TRIP13* were able to predict recurrence of prostate cancer when used in combination with established methods of disease stratification (Larkin *et al*., 2012). Interestingly, *TRIP13* promotes error prone NHEJ, rather than HR, as a form of DNA repair, as well as malignant transformation in head and neck cancer (Banerjee *et al*., 2014). Wang and colleagues have identified a mitotic checkpoint-silencing function of TRIP13 and suggested this may be an important mechanism promoting chromosomal instability in cancer (Wang *et al*., 2014a). Thus, this gene has complex

and different impacts on DNA repair and chromosome stability in the normal and malignant state, which appears to depend to some degree on the level of expression. It would be interesting to establish whether similar mechanisms occur to promote oncogenesis and/or treatment resistance in CRC. Although *TRIP13* may be germline-associated, it should not be considered a germline-specific gene or a CT gene but does appear to be an important oncogene when it is upregulated. It is unlikely to be useful as an immunotherapeutic target, though it is possible that as yet ill-defined oncogenic functions could be targeted in other forms of cancer therapy.

CEP55 is considered a CTA according to CTDatabase, though this gene did not appear on any of the other germline gene lists we identified (Almeida *et al*., 2009). It has been shown in numerous cancer types to be associated with survival and/or more aggressive forms of disease (reviewed in, Jeffery *et al.*, 2015). The gene was widely expressed in both the normal and cancerous tissues tested in this study, though to a greater degree in the cancerous tissue where it was expressed in moderate-high levels in all but one tested cancer sample (an ovarian cancer). It is known to be expressed at much higher levels in the germline, which is presumably why is has been classified as a CTA (perhaps due to a fold-increase in expression over normal). However, there are now well established roles during mitosis (Jeffery *et al*., 2015), so the classification of this gene even as a testis-selective CTA should be revisited. This does not mean it does not have important oncogenic functions and/or therapeutic potential, though again unlikely to be a useful target for immunotherapies.

5.3.3.2. Meiosis-associated genes in normal and cancerous tissue

The TLDA analysis served as a 'revalidation' of the expression profiles of genes which we had previously identified as possible cancer-specific biomarkers and termed meiCT genes given their putative association with meiosis (Feichtinger *et al*., 2012b; Sammut *et al*., 2014) but went further to investigate the expression of these genes in clinical tissue samples. The results, summarised in Figure 5.16, in general were consistent with our previous findings and RT-PCR validation of gene expression profiles. Exceptions to this appeared to be *TDRD1* and *CCDC79*, though in terms of tissue the highest number of separate tissue types for which any of the meiCT genes were expressed was 4 (excluding the testis and CNS).

We previously validated *IQCF1* as a meiCT gene which had weak expression in a colorectal and ovarian cancer cell line (Sammut *et al*., 2014). Here it was shown that this gene had a

testis/CNS restricted expression profile in normal tissues. Additionally, there was expression in one colon cancer sample and clear expression in one melanoma sample (see Figure 5.16, group AC). This gene has recently been implicated in sperm capacitation and the acrosome reaction that are both essential processes for successful fertilisation (Fang *et al*., 2015). A functional role in cancer has not been established, though it can be postulated that the protein could promote invasion if present aberrantly in cancer, as some acrosomal proteins are involved with 'invasion' of the zona pellucida to permit fertilisation.

MBD3L1 has been shown previously to have upregulated expression in the testis and acts as a transcriptional repressor in a methylation dependant fashion (Jiang *et al*., 2004). The protein product of this gene forms a complex that has high affinity for methylated DNA and may be involved in a targeted mechanism of tumour suppressor gene inactivation in cancer (Rauch *et al*., 2006). We found the gene to be expressed in a melanoma and ovarian cancer sample, with expression restricted to the testis in normal tissues (see Figure 5.16, group ACF). The McFarlane group had previously validated this gene as a meiCT with a testis-selective expression profile and expression also seen in an endometrial and ovarian cancer cell line (Feichtinger *et al*., 2012b). The gene was also identified independently by Rousseaux and colleagues as a testis-associated germline gene, though it was not a member of the final panel of genes used to predict a worse outcome for lung cancer patients (Rousseaux *et al*., 2013b). It would be interesting to investigate whether this gene does have a functional role in cancer, for example, by reducing the expression of tumour suppressor genes.

The topoisomerase encoding meiosis-specific gene *SPO11* was not shown to be clearly expressed in any tissue other than the testis and was tested using two separate assays (see Figure 5.17, group CE); very weak/borderline expression was seen in foetal liver and a colon cancer sample only. The colon cancer sample which appeared to express *SPO11* at very low levels was an aggressive mucinous histo-type (see Figure 5.6). However, additional experimentation would be needed to confirm whether this protein, which causes DNA DSBs during meiosis, is in fact present in a subset of colon cancers, though it is a recognised CTA (Almeida *et al*., 2009). It has recently been shown that SPO11, along with other meiosis-specific proteins may be present in cutaneous T-cell lymphoma (Litvinov *et al*., 2014).

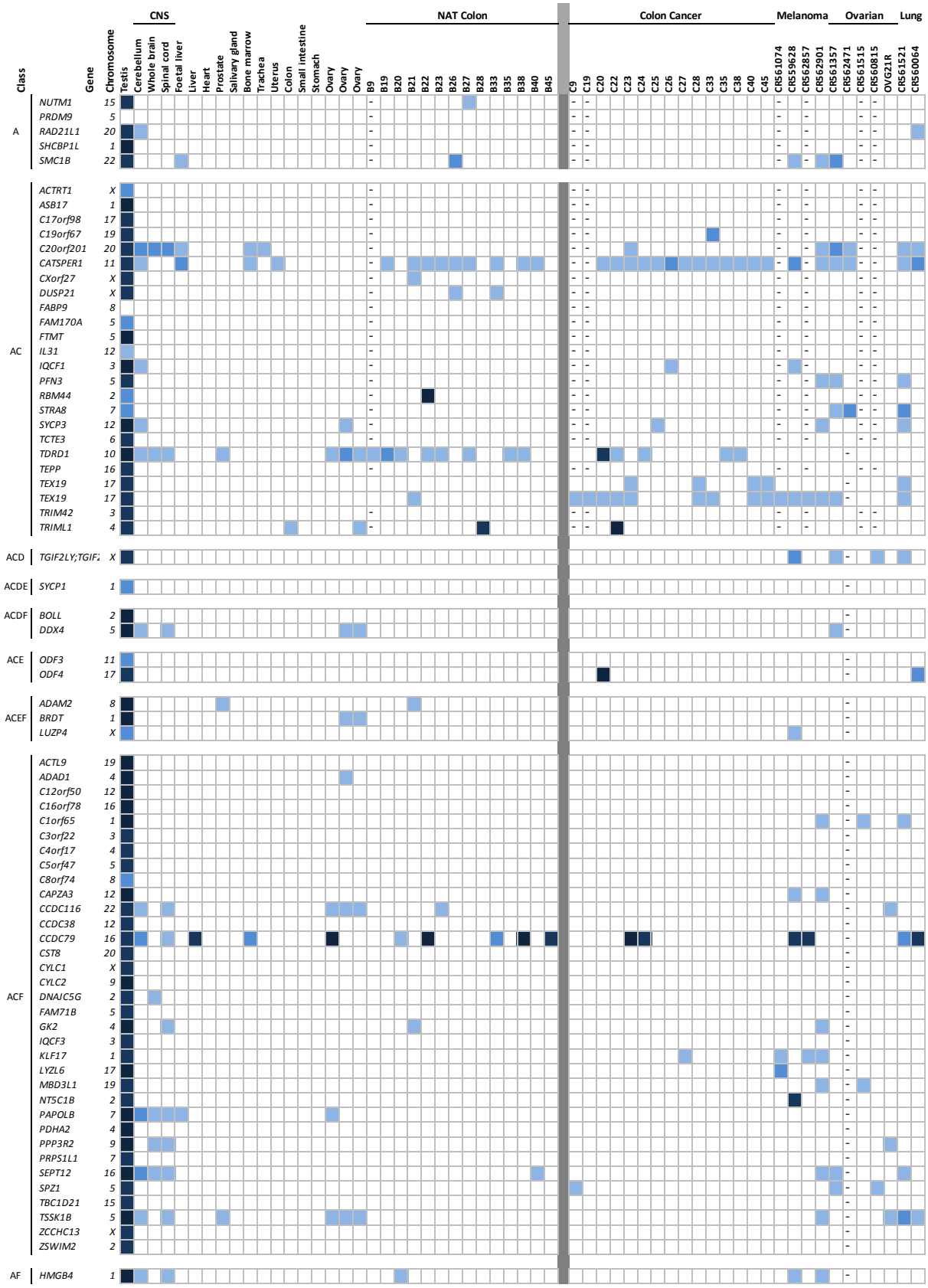


Figure 5.16. Gene expression of germline genes in normal and cancerous tissues previously categorised as meiCT genes (group A genes).

The letters on the far left indicate which groups the genes were included in and depict overlap between these groups – see Section 5.2.3. The grey vertical line divides the normal from cancerous samples.

LUZP4 has been identified in multiple myeloma cell lines and enriched specifically in sub-populations of these cells with stem-cell characteristics (Wen *et al*., 2014). It is possible that *LUZP4* expression was already high in the primary tumour before the melanoma recurred in this patient, although the study by Wen *et al.* hints at an association with stem cells in support of the theory postulated above, though this was seen in a different cancer type. Wen and colleagues additionally demonstrated that knockdown of *LUZP4* enhanced the sensitivity of myeloma cells to chemotherapeutic agents and reduced their colony-forming ability prior to treatment with these agents (Wen *et al*., 2014). *LUZP4* was expressed in the only recurrent malignant melanoma sample that we analysed; indeed this was the only sample of either cancer or normal tissues other than testis that expressed this gene (see Figure 5.16, group ACEF). This gene was identified as a testis-specific germline gene, though not included in the 26 predictive genes for lung cancer which they narrowed down to (Rousseaux *et al*., 2013b). We additionally identified and validated it as a meiCT gene that had displayed a meta-upregulation in ovarian cancer (Feichtinger *et al*., 2012b). *LUZP4* is considered a CT gene where expression has been shown most commonly in melanoma, lung, brain and ovarian cancers (Tureci *et al*., 2002). Given that expression was seen on the TLDA analysis in a recurrent malignant melanoma sample only, it may be the case that the CSC population expressed this gene which then became enriched in the tumour cell population when the cancer recurred. It is possible that the presence of this gene confers a survival advantage to the stem cell population or contributes to resistance to treatment, though a mechanism for this would need to be established and this remains purely speculative. Recently, it has been shown that LUZP4 can bind to RNA and interacts with other known mRNA export receptors and associated proteins (Viphakone *et al*., 2015). It was further shown that LUZP4 could reconstitute the RNA export deficiency produced when a known mRNA adaptor was repressed and consequently promote proliferation of melanoma cells. Thus, there is growing evidence *LUZP4* functions as an oncogenic driver in cancer.

Also included on the TLDA cards we designed were a cohort of genes that were present on the original list of human orthologues of the mouse genes associated with meiosis/spermatogenesis (Chalmel *et al*., 2007; Feichtinger *et al*., 2012b). The results for this group of genes are shown in Figure 5.17 – it is immediately apparent when comparing to Figure 5.16 that the expression profile for these genes was more widely seen in somatic tissues, which confirms that our previous experimental validation process was sound (Feichtinger *et al*., 2012b; Sammut *et al*., 2014). It is interesting to observe that the category C genes in Figure 5.17 (i.e. the genes that

did not overlap with any other groups and were restricted to group C) had the most extensive expression in normal tissues; some of these genes had been excluded as meiCT candidates previously during our validation process (Feichtinger *et al*., 2012b). A similar observation was also seen for the group D genes, where 4 out of the 5 genes, which did not overlap with any other groups/sources of germline genes, were widely expressed in the normal samples – see Figure 5.20. These observations also display how if candidate genes did overlap between the groups they do generally appear to be more promising targets to take forward and investigate further as potential cancer-specific biomarkers.

5.3.3.3. Testis-expressed (*TEX*) genes

TEX101 has been accepted as a CTA gene but not widely investigated across different cancer types (Almeida *et al*., 2009). It was identified as a human orthologue of a mouse gene associated with meiosis/spermatogenesis but not validated as a meiCT gene (Feichtinger *et al*., 2012b). It was identified as being associated with small cell lung cancer but not seen in cell lines derived from non-small cell tumour types (Tajima *et al*., 2003). Both the lung cancer samples we tested were of a non-small cell type (one squamous cell and the other an adenocarcinoma) and consistent with this previous finding, which found it to be present in small cell lung cancer only, we found this gene not to be expressed in the tested samples. The gene was found to be expressed but no antibodies against the protein could be detected in the serum of patients with chronic myeloid leukaemia (Ghafouri-Fard *et al*., 2012). *TEX101* expression was not detected in breast cancer samples tested in a further study (Dianatpour *et al*., 2012). It has not been described as a CT gene in colorectal cancer but a novel finding here is that two of the colon cancer samples expressed this gene (see Figure 5.17, group CE). However, one of the NAT colonic samples also expressed the gene, although the gene was otherwise restricted to the testis in the normal tissue samples. Interestingly, it has been shown that NYD-SP8 (an alias for TEX101) was shown to suppress the activity of various proteases involved in extracellular matrix (ECM) degradation, and have a stronger association with earlier stages of hepatocellular carcinoma (Yin *et al*., 2009). The effect of suppressing ECM degradation took place despite the fact that this protein shares homology with the urokinase plasminogen activator receptor, which is known to promote ECM breakdown (Smith and Marshall, 2010). The association with early stage or better prognostic tumours would also be at odds with its expression being associated with small cell lung cancers, which carry a very poor prognosis. However, consistent with the suppression of tumour invasion (Yin *et al*., 2009), the presence

of the protein in squamous cell carcinomas of the head and neck was associated with a lower rate of lymph node metastasis (Yoshitake *et al*., 2012). Thus, there may be a differing role of this gene in different tumours and at different stages of tumour development.

In addition to *TEX101* other TEX genes were included on the TLDA cards, namely *TEX14*, *TEX15* and *TEX19*. *TEX14* and *TEX15* were both expressed in more than two of the normal tissues other that testis. Stronger expression was seen in melanoma for both genes and for *TEX15* in an ovarian cancer sample also (see Figure 5.17, group CE). *TEX15* has been proposed previously as a CT gene that is preferentially expressed in cancer stem-like cells (Yamada *et al*., 2013). Knockout experiments in mice have revealed a functional role for TEX15 in DSB repair and chromosomal synapsis during meiosis (Yang *et al*., 2008). Both genes appear to have expression in normal tissue to an extent which may hinder their use as immunotherapeutic targets, though it may be the case that in the instances of low gene expression seen in normal tissue, translation of the protein does not take place. In the samples tested, however, if there was upregulation of *TEX14* or *TEX15* in the cancer samples, it appeared to be minimal.

TEX19 displayed a tighter restriction of its expression in normal tissues. We used two separate assays for this gene and it was only expressed in one NAT sample in one of these assays (see Figure 5.16). Consistent with our previous findings but where we used predominantly cancer cell lines (Feichtinger *et al*., 2012b), the gene was expressed in several cancer samples and in all cancer types tested. We used two separate primer sets for this revalidation and although there was a slightly wider expression pattern for one of the primer sets, the results were broadly similar between the two (see Figure 5.16 as well as the rows marked by black arrows in Figure 5.15). There was also an indication of very weak expression in more cancer samples, as well as some normal tissues; though it is worth pointing out that this very weak/borderline expression *TEX19* in normal tissues was almost entirely limited to the NAT samples. In mice, TEX19 is a placenta and germline specific protein, which associates with the stem cell compartment and is essential for spermatogenesis (Kuntz *et al*., 2008; Tarabay *et al*., 2013). It has not been described as a CTA other than its potential highlighted by our group (Feichtinger *et al*., 2012b) and although other TEX genes were, it was not present on the list of ~500 germline genes (group F) identified by the Knochbin group as either testis- or placenta-specific (Rousseaux *et al*., 2013b).

Figure 5.17. Gene expression of group C genes, which had not been confirmed as meiCT genes previously, in normal and cancerous tissues.

This grid represents all the genes that were identified as orthologues of mouse spermatocyte-specific that were included on the arrays and not already included in Figure 5.16. Again, the letters on the left hand column indicate which groups the gene also appeared in. Most of the genes overlapped with another group but some were restricted to this group and these genes displayed the highest frequency of expression in the normal tissues (category C above).

One of the genes which stood out as a good candidate CT gene was *C20orf195* (see Figure 5.17, group C). This gene was expressed in testis and none of the normal tissues tested but was also expressed in two melanoma samples and one ovarian cancer sample. The gene encodes an uncharacterised protein with unknown function. It was identified as an orthologue of a mouse spermatocyte/meiosis-associated gene but it was not validated as a meiCT. The same gene also appeared on the EST filter for meta-upregulated genes in cancer (see Figure 5.13). Together these findings provide a strong indication this gene may be a clinically relevant gene that should be investigated further. The gene was not expressed in any of the CRC samples collected for this research, so appears not to be candidate marker for this cancer type but it could be for other cancer types such as melanoma or ovarian cancer as the data suggests.

Another interesting gene was *C16orf46*, which had a testis-selective expression profile but appeared to be expressed in several of the cancer samples tested (see Figure 5.17, group C). In fact, 14 of the 20 cancer samples tested expressed this gene. Despite this no previous association with cancer has been described in the medical literature and the gene has unknown function. Further investigation of this gene would certainly appear warranted.

SYNGR4 has unknown physiological functions, though is a member of the synaptogyrin gene family which encode proteins that are involved in exocytosis (Shin, 2014). The gene was identified as an orthologue of a mouse gene characterised as being associated with spermatogenesis (Feichtinger *et al*., 2012b). The gene has a statistically significant upregulation in expression when ovarian cancer microarray datasets were meta-analysed (see Figure 5.12). The gene had a testis/CNS-selective expression pattern in normal tissues, being expressed commonly at low levels in NAT colon samples and having higher expression in normal brain and testis (see Figure 5.17). There was also higher expression in 3 colon cancer samples and the gene was also expressed in melanoma, lung and ovarian cancer (in 4 out of the 7 samples tested for these cancer types). It may well be worth exploring the function of this gene and expression pattern further in cancer and normal tissues as well as investigating the normal function of this gene.

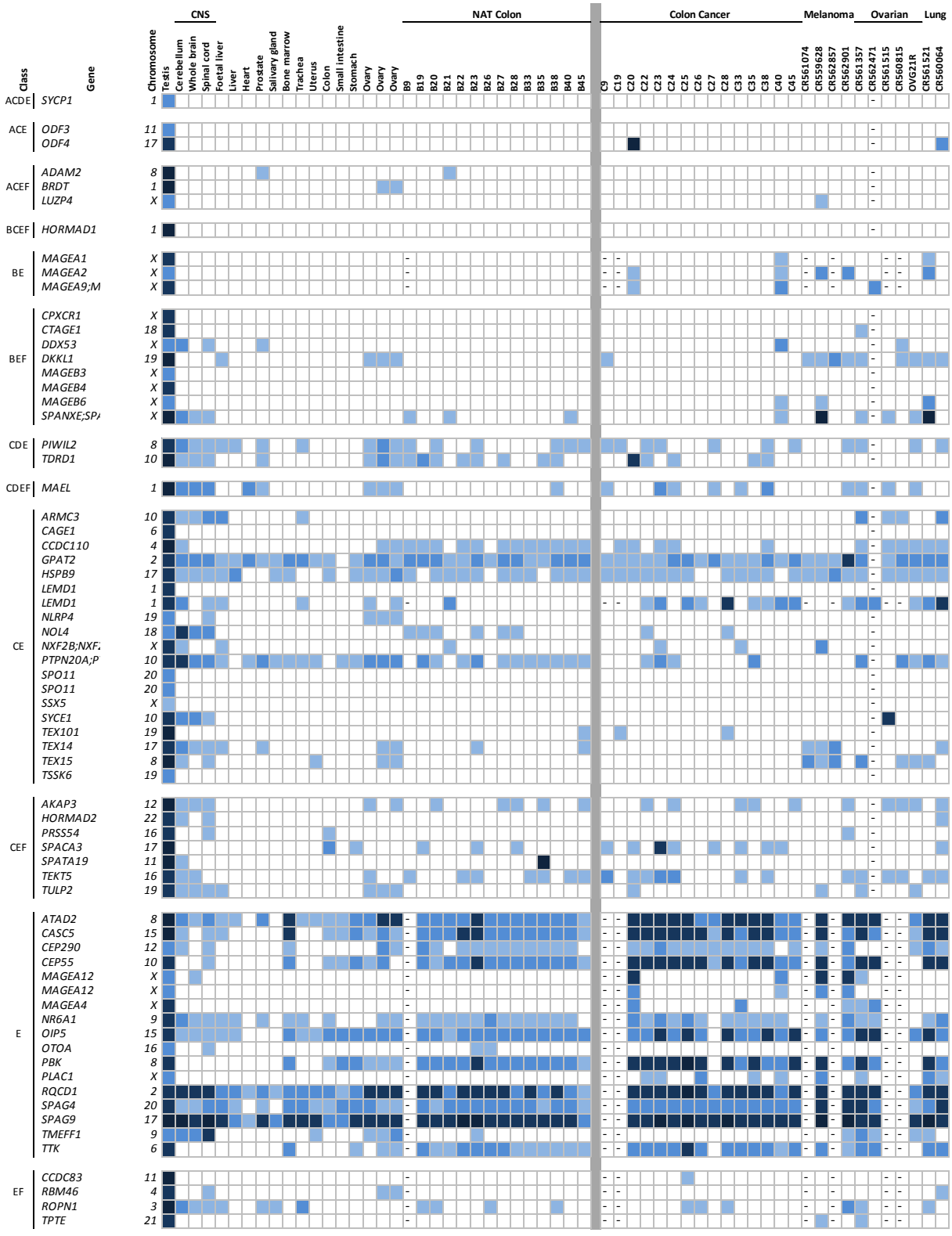


Figure 5.18. Gene expression of group E genes in normal and cancerous tissues.

All these genes had previously been identified as CT genes according to 'CTdatabase' (Almeida *et al*., 2009). As Hofmann and colleagues (2008) had previously discovered several CT genes are selective in their expression profile with a preference for the testis/germline and upregulated in cancer. Some genes, for example *GPAT2*, appeared to be equally strongly expressed in normal tissues questioning its categorisation as even a testis-selective CTA though it was strongly expressed in an ovarian cancer sample.

5.3.4. Analysis of genes previously described as CT genes

Several genes included on the TLDA cards had previously been described as or suggested to be CTA genes (Almeida *et al*., 2009). The expression profiles for these genes are shown in Figure 5.18. Discussion of specific examples is detailed below (see also Section 1 of Appendix for additional discussion).

5.3.4.1. PLAC1 and PLAC1-like genes

PLAC1 is a known CT gene and it has been suggested that it may be a marker of stemness in epithelial ovarian cancer (Almeida *et al*., 2009; Tchabo *et al*., 2009). The gene is considered a trophoblast-associated gene where it is thought to play an important role in placental development but low levels of expression have been seen in testis and cerebellum (Fant *et al*., 2010). The suggested roles this gene plays in cancer relate to the processes involved with invasion of trophoblastic cells that normally occur during placental development (Devor *et al*., 2014; Fant *et al*., 2010; Tchabo *et al*., 2009). Expression of this gene has been linked to many cancers, for example, ovarian, endometrial, colorectal and breast cancer; PLAC1 has also been recommended as a potential target for immunotherapies (Devor *et al*., 2014; Liu *et al*., 2014; Liu *et al*., 2012b; Shantha Kumara *et al*., 2012; Tchabo *et al*., 2009). *PLAC1* expression has been shown to be induced by Epstein-Barr virus (EBV) (Wang *et al*., 2014b), providing further evidence for a link between infections/inflammation, CT genes and cancer. *PLAC1* was expressed in a proportion of each cancer type tested in the TLDA cards but not clearly in any normal tissue other than the testis (see Figure 5.18), confirming its existence as a CT gene across many cancer types. The gene is believed to be held in a transcriptionally-repressed state in somatic cells due to its promotor region being in a heterchromatic state (i.e. outside the germline/placenta). Mutations in *TP53* and *RB* have been identified as possible routes to the transcriptional upregulation of *PLAC1* in tumour cells (Chen *et al*., 2013). Although high levels of *PLAC1* are found in many cancer cell lines, the model *in vitro* system used in that study was transformed primary fibroblasts. Given the increasing realisation of the importance of the tumour stroma in driving oncogenesis (Calon *et al*., 2015; Isella *et al*., 2015), it could be that fibroblasts associated with tumours experience an upregulation in *PLAC1*, which is then able to exert tumorigenic effects. This would require in-depth analysis of gene expression in tumour cell subpopulations as has been recently described in CRC (Calon *et al*., 2015).

PLAC1L encodes the PLAC1-like protein. It shares no significant homology with the known CT gene *PLAC1* and very little is known about this gene's function. *PLAC1L* was identified as a testis-associated germline-specific gene, though interestingly *PLAC1* was not identified as a germline gene despite being classified as a placental-associated gene (Rousseaux *et al*., 2013b). PLAC1L was identified as an interacting partner with the BRCA1 C terminus (BRCT) domain (Woods *et al*., 2012) so potentially could impact on DNA repair pathways if upregulated in cancer. It would be interesting to investigate if this gene is expressed in cancer but I did not include this gene on the TLDA cards.

5.3.4.2. Previously validated tissue-specific CTA-genes

A cohort of the 'known CTA-genes' (see Figure 5.18) were separately validated by Hofmann and colleagues as germline restricted or selective (Hofmann *et al*., 2008). It can be seen in Figure 5.19 that all but 3 of these genes conformed to the testis-restricted or testis/CNS-restricted expression profile in normal tissues as had been previously described (Hofmann *et al*., 2008). None of the genes were found to be widely expressed in normal tissues, and all would still be classified as CT genes though some with a testis-selective expression pattern. One of the genes, *HORMAD1*, appeared on our list of spermatocyte-associated genes (group C) but it was excluded as a meiCT gene as weak expression was seen in a range of normal tissues in previous work by the McFarlane group (Feichtinger *et al*., 2012b). A greater proportion of the genes from this group overlapped with group F genes, which were germline tissue-restricted genes identified by the Knochbin groups (Rousseaux *et al*., 2013b). The consistency of our results with the cohort put forward by Hofmann and colleagues (2008) is reassuring that our experiments, although not repeated, have a certain degree of validity. The experiments should however be repeated.

DKKL1 is expressed during spermatogenesis and during placental development, where it plays a role in implantation of the blastocyst as well as promoting the ability of sperm to penetrate the zona pellucida during fertilisation (Kohn *et al*., 2010). Thus, it may well interact with proteolytic factors and could conceivably play a role in tumour invasion if ectopically expressed in cancer. Due to the fact that microarray data suggested this gene was upregulated in certain brain tumours, Sibbe and Jarowyj (2013) investigated whether the protein was present in the brain of mice. They found that it was present in low abundance, with highest levels in cortical neurons and postulated a role in oncogenesis as well as neurodegenerative disorders (Sibbe and

Jarowyj, 2013). The gene was categorised as testis-selective given limited expression in normal tissues (see Figure 5.19, group BEF). Borderline or very low expression was seen in CNS tissues but this did not meet the criteria we set (i.e. Ct value <32) to say the gene was definitely expressed in this tissue type but the previous findings of others suggest this low expression in the CNS may be real (Sibbe and Jarowyj, 2013). It was observed that foetal liver, as well as normal ovarian tissue did express low-moderate levels of this gene. None of the NAT colonic samples clearly expressed the gene but one of the colon cancer samples did. Moreover, the gene was widely expressed in the melanoma, ovarian and lung cancer samples, with only one of the tested ovarian cancer sample not clearly expressing this gene. It is possible that this gene plays an important oncogenic role in many cancers and more widely than is currently appreciated despite it being recognised as a CTA; it may also be a good therapeutic target for these cancers.

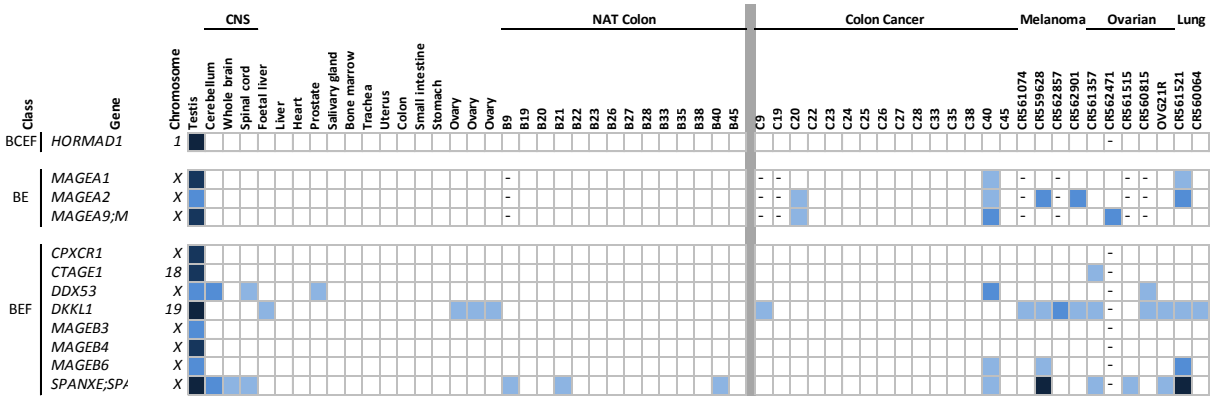


Figure 5.19. Gene expression of germline genes previously categorised as CT genes and validated by Hofmann and colleagues as testis or testis/CNS-restricted.

All these genes overlapped with category E genes (known CT genes), which is not surprising as Hofmann *et al.* (2008) were analysing genes thought to belong to this category of genes. Consistent with their acceptance as CT genes, none of the genes analysed fell outside what would classify them as such, although some were testis-selective in their expression profile in normal tissues.

HORMAD1, for which a functional role in breast cancer has recently been proposed (Watkins *et al.*, 2015), displayed a tight testis-restricted expression pattern in normal tissues but was likewise not found to be expressed in any of the cancer samples we tested (e.g., see Figure 5.19 and Figure 5.21, group BCEF). *HORMAD2*, displayed a testis/CNS-restricted expression pattern and was shown to be expressed reasonably strongly in a lung adenocarcinoma sample (e.g., see Figure 5.18, group CEF). HORMAD2, along with its sister HORMA domain protein HORMAD1, play critical roles during meiosis and meiotic recombination (Kim *et al.*, 2014).

HORMAD2 has recently been characterised as a CT gene in lung cancer patients (Liu *et al*., 2012a). This is consistent with the findings from our PCR experiment, where lung cancer and not colon, ovarian and melanoma cancer samples expressed this gene. It has further been shown that single-nucleotide polymorphisms of this gene impact significantly on the clinical outcome in a Chinese population with lung cancer (Zhang *et al*., 2014). Thus, it is possible that such genetic testing of this CT gene could facilitate disease stratification and guide therapy accordingly for this cancer type.

5.3.5. Analysis of germline genes previously identified as oncogenic drivers in *Drosophila melanogaster*

The paper by Janic and colleagues (2010) was very important in that it provided evidence for a direct role for germline factors acting as oncogenic drivers in a brain tumour model in *Drosophila melanogaster*. It was subsequently shown that the human orthologues of these genes were frequently upregulated in human cancers and speculated that a soma-to-germline transition may be a hallmark of many cancer types (Feichtinger *et al*., 2014a). Several of these human orthologues were included on the TLDA cards and the genes overlapped with several other germline gene lists (see Figure 5.20). All but one of the genes that did not display an overlap with other gene lists (category D in Figure 5.20) were widely expressed in both the normal and cancerous tissue samples tested; some of these are discussed below in Section 5.3.5.1 and Section 5.3.6.

It has recently been shown that the meiosis-associated gene, *FKBP6*, that is known to play a role in male fertility, is also associated with cervical cancer (Brebi *et al*., 2014). Through genome wide DNA methylation analysis this gene was identified as being associated with cervical cancer and in particular precancerous lesions (Brebi *et al*., 2014). It was shown in the TLDA analysis that the gene was expressed in the testis and a melanoma sample only, categorising it as a novel CT gene (see Figure 5.20, group DF). A role for this gene in both melanoma and cervical cancer is thus worthy of further exploration. The gene was identified as a testis-specific germline gene by the Knochbin group but not included in the predictive panel of 26 genes for lung cancer patients (Rousseaux *et al*., 2013b). *FKBP6* was not identified as an orthologue of mouse spermatocyte/meiosis-associated gene but it should be noted that

orthologues for every mouse gene listed by Chalmel and colleagues could not be identified (Chalmel *et al*., 2007; Feichtinger *et al*., 2012b).

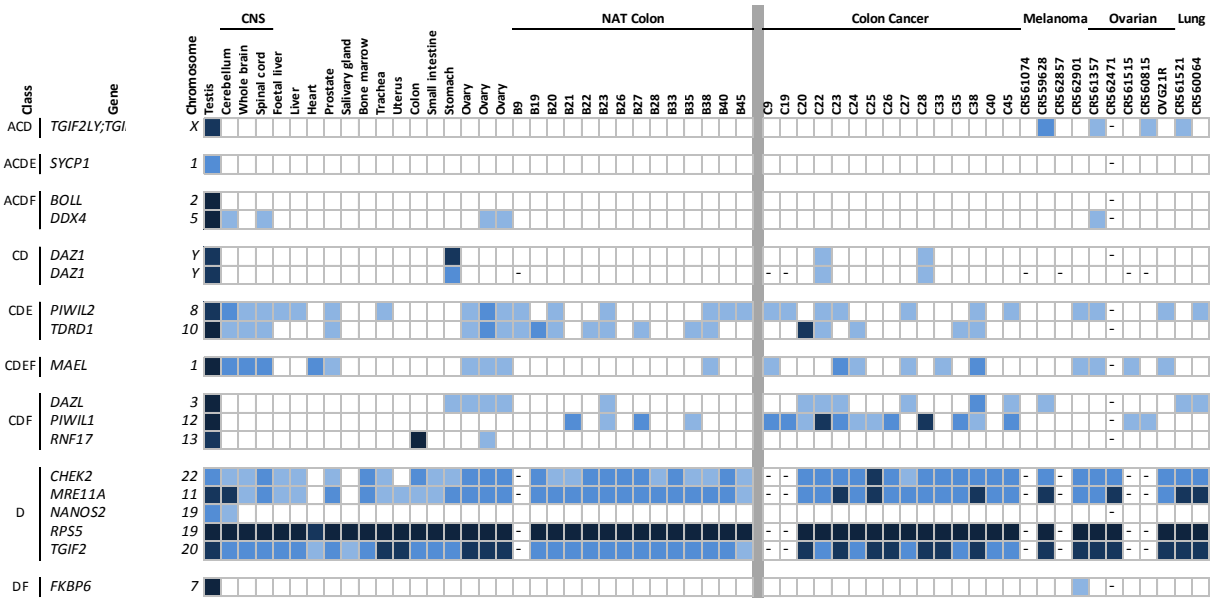


Figure 5.20. Gene expression of group D genes in normal and cancerous tissues.

These genes had all previously been identified as ortholgues of germline genes that drove oncogenesis in brain tumours in *Drosophila melanogaster* (Feichtinger *et al*., 2014a). In a similar observation to Figure 5.17, the genes that were only found in this group (category D) were more widely expressed in the normal panel, although *NANOS2* was an exception but this gene was not expressed in any of the cancer samples.

5.3.5.1. *TGIF2* and related genes

TGIF2 was identified as the human orthologue of the *Drosophila melanogaster* germline gene *zpg* that had been implicated in contributing to oncogenesis in a brain tumour model in *Drosophila melanogaster* (Feichtinger *et al*., 2014a; Janic *et al*., 2010). It was originally identified as a transcription factor upregulated in ovarian cancer cell lines (Imoto *et al*., 2000). It was shown in this earlier study that the gene was expressed across all normal tissue tested but at a lower levels than the ovarian cancer samples; consistent with the finding that the gene was expressed widely in the normal tissues analysed on the TLDA cards (see Figure 5.20, group D). *TGIF2* has more recently been revealed as a target for microRNAs (miRNAs) and implicated in metastasis formation (Lujambio *et al*., 2008; Suzuki and Yoshino, 2008). *TGIF2* has also been identified as one of the most highly upregulated genes when comparing rectal adenomas to carcinomas (Lips *et al*., 2008). Again, consistent with this finding we found generally higher expression in the colon cancer samples compared to the matched normal tissue. High levels of

expression were also seen in all the melanoma, ovarian and lung cancer samples tested, though similar levels of expression were also seen in normal ovarian tissue. *TGIF2* is undoubtedly important in the pathogenesis of many cancer types but it is one gene to exclude from further analysis if trying to identify cancer-specific germline genes with a clear on-off signal. The mapping of this gene as a human orthologue of a *Drosophila melanogaster* gene may not have been entirely accurate.

TGIF2LX and *TGIF2LY* on the other hand are testis/germline-specific genes and the expression profile seen on the TLDA array was consistent with this finding. The assay used on the array did not differentiate between these two genes. It was further shown that the gene was expressed clearly in a melanoma sample, with weaker expression in two ovarian cancer samples and the metastatic lung cancer sample tested (e.g., see Figure 5.16 and Figure 5.20, group ACD). The gene did not appear to be expressed in any of the colon cancer samples. Interestingly the melanoma sample displaying the strongest expression among the cancer types was a recurrent cancer, raising the possibility of expression of this gene playing a role in drug resistance or cancer-stem-like cell characteristics. Clearly this would need to be corroborated by further studies and is very speculative on the basis of the current evidence provided. However, it has been suggested to have a functional role in regulating spermatogonial stem cells in the testis (Aarabi *et al*., 2008). *TGIF2LX* has been shown to display similar functions in a fly model to that of *TGIF2*, although expression of both genes inhibited rather than promoted cell growth and proliferation (Wang *et al*., 2008). Interestingly, *TGIF2LY* did not display the same effect on fly tissues despite its close relation to *TGIF2LX*. Thus, *TGIF2LX* may have tumour suppressor functions but given the functional overlap with *TGIF2*, which is implicated in various oncogenic processes, it also raises the possibility of as yet uncharacterised functions. Little is known about its function within the testis, though it is believed to play important roles in spermatogenesis and is implicated in infertility (Aarabi *et al*., 2008; Wang *et al*., 2008).

We identified *TGIF2LX* as a meiCT gene (Sammut *et al*., 2014) and it was also shown, along with *TGIF2LY*, to be an orthologue of a *Drosophila melanogaster* gene that had been implicated as having oncogenic functions in brain tumours (Feichtinger *et al*., 2014a; Janic *et al*., 2010). Adding to the possible role of these genes in driving oncogenesis in the brain, *TGIF2LY* was shown to have statistically significant upregulation in brain tumours when clinical datasets were meta-analysed (see Figure 5.12). *TGIF2LX* was not present on the arrays (i.e. that contributed

to CancerMA) so it cannot be stated that this upregulation is specific for *TGIF2LY* (Feichtinger *et al*., 2012a). The only other reference specifically linking these two related genes to cancer in the literature that we are aware of is a study showing that more aggressive forms of prostate cancer commonly expressed either *TGIF2LX* or *TGIF2LY* (Ousati Ashtiani *et al*., 2009). Given its expression in three separate cancer types in this study, this raises the possibility of either or both of these genes playing an oncogenic role more widely than is currently appreciated.

5.3.5.2. Expression profiles of genes that contribute to the synaptonemal complex

SYCP1 and SYCP3 are constituent members of the synaptonemal complex (SC) (Fraune *et al*., 2012). The genes encoding these proteins were both identified as meiCT genes; *SYCP3* displayed a significant upregulation in ovarian cancer on the microarray datasets and *SYCP1* in ovarian and brain cancer (Feichtinger *et al*., 2012b). *SYCP1* is additionally considered a CT gene and was also identified as an orthologue of germline oncogenic driver gene of brain tumours in *Drosophila melanogaster* (Almeida *et al*., 2009; Feichtinger *et al*., 2014a; Janic *et al*., 2010). On the TLDA analysis *SYCP1* was only expressed in the testis (e.g., see Figures 5.16 and 5.20, group ACDE). *SYCP3* on the other hand was expressed strongly in the testis, with weak to moderate expression in normal brain and ovary as well as a melanoma, colon and lung cancer sample (e.g., see Figure 5.16, group AC). Interestingly neither of these genes appeared to be expressed in ovarian cancer, which was the cancer type displaying significant upregulation in expression on the microarrays (Feichtinger *et al*., 2012b), although both genes did display very weak/borderline expression when the lower threshold for expression was considered. *SYCP3* has been shown to inhibit *BRCA2* function and consequently HR, lead to increased DSB formation and aneuploidy *in vitro*, providing a mechanism for contributing to chromosomal instability in cancer (Hosoya *et al*., 2011). The expression of *SYCP3* has also recently been shown to correlate with a worse outcome in lung and cervical cancers (Cho *et al*., 2014; Chung *et al*., 2013). Such gene expression analysis could be a relatively simple way of identifying patients who would benefit from more aggressive forms of treatment at an earlier point in their treatment, before the disease had progressed to a stage where cure cannot be achieved. Rousseaux and colleagues (2013b) were the first to demonstrate the potential power of combining a relatively small number of germline genes to predict poor outcome in lung cancer patients. Similar cohorts of genes, for which *SYCP1* and *SYCP3* may well be useful members, could be identified and prove useful in different types of cancer. Such predictive groups of genes would be of considerable clinical benefit at relatively low cost. For example,

gene expression profiling using both germline, including *SYCP1*, and somatic genes has been shown to predict disease progression in cutaneous T-cell lymphoma, once again displaying the power of relatively simple stratification approaches that could have wider clinical applications (Litvinov *et al*., 2015).

SYCE1 also forms part of the central element of the SC – see Section 1.4.5 (Fraune *et al*., 2012). It is considered a CTA (Almeida *et al*., 2009) and the gene encoding this protein was identified as a candidate meiCT gene but found to be expressed in several normal tissues, thus excluding it being classified as a meiCT gene (Feichtinger *et al*., 2012b). In the 'revalidation' of this genes tissue expression using the TLDA cards, the gene was shown to have testis/CNS restricted expression pattern, so on this basis should not have been excluded, possibly suggesting the negative 40-cycle PCR validation was in this case too stringent. It should be noted that although we tested the expression in more normal tissues, the expression was in fact tested in slightly fewer individual tissue types than the McFarlane group had originally performed. The gene appeared only to be expressed in one NAT tissue sample. Thus, the gene may indeed be a (meiosis-associated) CT gene and our previous classification was perhaps over-stringent in some instances. A benefit of this 'over-stringency' should be the identification of highly-specific cancer biomarkers. Others have recently provided evidence suggesting *SYCE1*, along with *VCX/Y* and other CT genes, do encode CTAs of potential importance in lung adenocarcinomas (Taguchi *et al*., 2014). We found strong expression in one of the ovarian cancer samples in this study (e.g., see Figure 5.18, group CE), suggesting the gene may be clinically relevant in a subset of patients with this cancer type.

5.3.6. Previously identified tissue-specific germline genes

CCDC83 appeared on the extended list of germline genes identified by Rousseaux and colleagues and also in the list of known CT genes (Almeida *et al*., 2009; Rousseaux *et al*., 2013b). This gene was shown to have restricted expression to the testis in the normal tissues tested and one of the colon cancers tested also displayed moderate expression levels (e.g., see Figure 5.21, group EF). These findings were consistent with *CCDC83* being a CT gene. In support of this it was shown through an immunoscreening technique using sera obtained from patients with CRC, some displayed IgG antibodies against this protein (Song *et al*., 2012). They also confirmed a gene expression profile consistent with it being a CT gene. Together with our findings this adds further support to this protein being of real significance for a subset of CRC

patients. In addition to the clear expression of the gene in one of the colon cancer samples, three other colon cancer samples had very weak/ borderline expression. Borderline expression of this gene was also seen in two of the NAT sample (though interestingly the corresponding cancer sample did not display borderline expression) and two ovarian cancer samples that were tested. In the study by Song *et al.* they also found a weak IgG response to the protein in a small cohort of the normal patients tested (Song *et al*., 2012). This may limit the use of this CTA as a target for immunotherapeutic approaches in the clinical setting.

5.3.6.1. The genes previously identified as predictive markers in lung cancer

The genes depicted in Figure 5.22 all belong to the group of gene identified by the Knochbin group as predictive of a worse outcome in lung cancer (Rousseaux *et al*., 2013b). The genes were not widely expressed in the two lung cancer samples tested on the TLDA cards. However, the prediction of a poor outcome was based on 3 or more of any of the 26 genes being expressed in a given tumour. So in fact, on this basis, both the tumours from which the RNA was extracted and used in this study would be expected to carry a poor prognosis given the fact that 5 and 3 of the 26 gene were expressed respectively. The lung cancer specimen expressing 5 of the 26 predictive genes was a metastatic tumour.

In comparison to my RT-PCR findings (see Figures 5.10 and 5.11) the results are broadly similar, displaying consistency between the different assays/primer sets, which is reassuring. *ROPN1* was shown on the RT-PCR to be testis-selective but on the TLDA analysis showed more expression in the normal tissues to the point that its classification as a CT gene is called into question. The other example of a discrepancy is for *OR7E156P*, for which it was suggested as having a weak expression pattern in normal tissues on RT-PCR but here the gene was expressed in the testis only. Given that the strength of expression in the testis was weaker than many of the other germline genes tested in this study, this might suggest the primers/assay is not efficient in this instance and as a consequence of this perhaps failed to detect weak expression that may be present in other tissues. It would be worth repeating the analysis with an alternative assay and confirming expression with DNA sequencing.

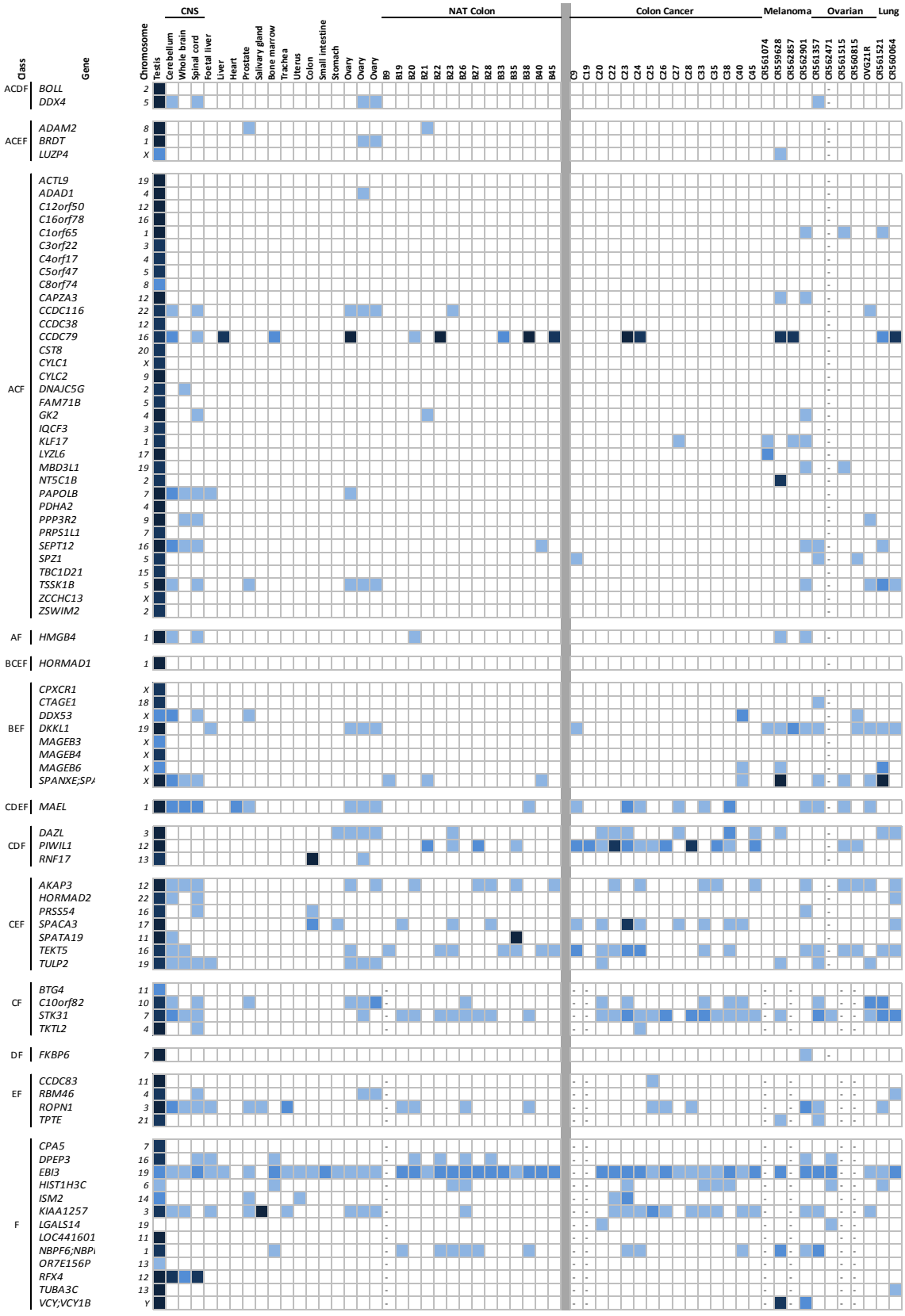


Figure 5.21. Gene expression of group F genes in normal and cancerous tissues.

All these genes had previously been identified as tissue-specific germline (placental or testis-associated) genes (Rousseaux *et al*., 2013b). *EB13* was the only gene from this group that was widely expressed in the normal samples; a finding that concurred with the previous RT-PCR analysis (see Figures 5.10 and 5.11). Interestingly, this was one of the predictive genes of a poor outcome in lung cancer.

LGALS14 is a placental-associated gene that was also identified as a predictive marker in lung cancer patients (Rousseaux *et al*., 2013b). It did not appear on other germline lists (perhaps due to its placental and not testis-association) but was meta-upregulated in cancer, as well as being EST restricted when the germline list was challenged against these pipelines (see Figure 5.13). It did display a restricted expression profile in the normal tissues and also displayed only very weak/borderline expression in the testis. It was additionally expressed in one colon cancer sample and one ovarian cancer sample (see Figure 5.22, group F). Interestingly it was not shown to be expressed in lung cancer samples for which it was identified as a predictive marker, though it should be noted that it was not universally expressed in lung cancer or even the group with the worst prognosis in the Knochbin group study (Rousseaux *et al*., 2013b).

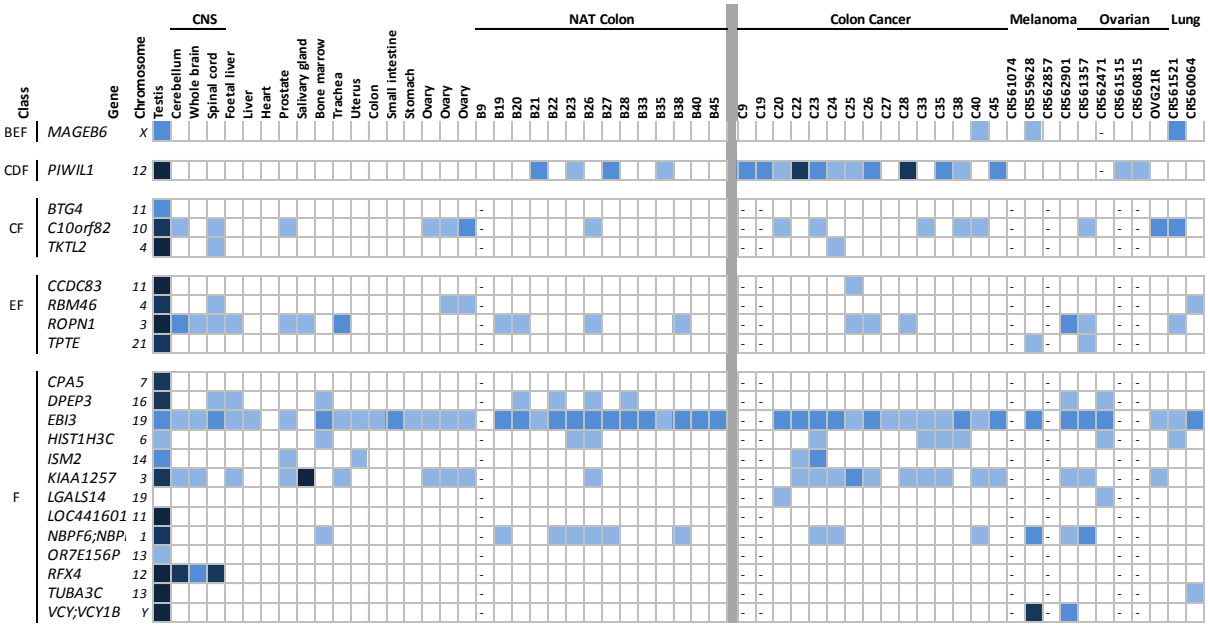


Figure 5.22. Gene expression of group G genes in normal and cancerous tissues.

These genes belong to the group of 26 that Rousseaux and colleagues identified as predictive markers of a poor prognosis in lung cancer patients (Rousseaux *et al*., 2013b); the genes are also included in Figure 5.21 as they were taken from the category (F) genes. Four of targets were not included on the TLDA cards: these were *TPTE2P2*, *FLJ43944*, *Hs.601545* and *C12orf37*. When considering the NAT samples as a single tissue type, the majority of these genes did display an expression pattern consistent with testis-restricted or testis-selective CT genes.

HIST1H3C was shown to be testis-selective in normal tissues; expression was seen in bone marrow as well as one of the NAT colonic samples. This gene had previously been described as having a testis only expression profile in normal tissues and it was one of 26 genes that predicted a poor prognosis in lung cancer (Rousseaux *et al*., 2013b). In the cancer samples tested, four of the colonic cancer samples indicated that the gene was expressed and in addition

one of the ovarian samples and one of the lung cancer samples also expressed this gene (see Figure 5.22). Little is known about this gene's association with cancer but it has been shown that methylation of this gene was associated with overall survival in neuroblastoma patients (Decock *et al*., 2012). It was further shown that DNA methylation of this gene together with three other genes could predict which patients were more likely to survive from the cohort of patients collectively grouped as already being in a high risk, more aggressive disease category (Decock *et al*., 2012). This highlights not only the important of epigenetic factors but also how traditional methods of stratifying disease are not entirely accurate. In a similar way to that which Rousseaux and colleagues discovered (Rousseaux *et al*., 2013b), ectopic activation of tissue specific genes together with analysis of epigenetic alterations may prove a useful discriminator and selector for specific forms of therapy. Both studies also highlight how combining factors (for example, gene expression and DNA methylation status) increases the accuracy of, or even enables, accurate stratification. Validating appropriate combinations of such factors to accurately predict response to treatments or survival across the range of cancer types will be an ongoing challenge for scientists and clinicians into the future. The methylation status of *HIST1H3C* was not investigated in the Knochbin group study (Rousseaux *et al*., 2013b).

5.4. Conclusions

With reference to colorectal cancer *C19orf67*, *CCDC83*, *DAZ1*, *IQCF1*, *ISM2*, *KLF17*, *LGALS14*, *ODF4*, *PLAC1*, *SPZ1*, *TEX19*, *TEX101*, *TKTL2*, and various members of the *MAGE* family of CT genes appear to be potentially relevant CT genes. All these genes were expressed in at least one colon cancer sample and displayed an expression profile in normal tissues consistent with a CT gene. Many of these are not widely recognised as CT genes in colon cancer. In addition to this, *PIWIL1* and *C16orf46* emerged as interesting candidate genes with limited expression in normal tissues but expression seen across many cancer samples tested. *PIWIL1* displayed weak to moderate expression in some NAT colonic samples but expression was otherwise restricted to the testis in normal tissues, including normal non-adjacent cancer colonic tissue. This is consistent with earlier findings using RT-PCR (see Figure 5.9). The importance of PIWIL1 as a stem cell marker in colorectal cancer has recently been described (Litwin *et al*., 2015). TEX19 and PIWIL1 were discussed previously in Chapter 4. Consistent with the findings of previous studies two of the colon cancer samples were shown to co-express

several members of the *MAGE* family, suggesting a common mechanism of transcriptional activation. It should be noted that this is a large gene family and we included relatively few members on the TLDA cards. The two colon cancer samples that expressed the most candidate CT/germline genes were of an aggressive histological type, or had displayed the ability to metastasise (to the regional lymph nodes). By focusing on a smaller cohort of genes, it may be possible to predict a worse outcome and/or guide therapy from biopsies taken early on during the diagnostic and staging process.

Over 100 genes appeared as potential candidates for the development of immunotherapies given their limited expression in normal tissues. All these genes could also potentially be utilised to improve cancer diagnostic and/or stratification techniques; though many of them display a very limited or no expression in the cancer samples tested in this study. Additional discussion for some of these candidate genes is provided in the Appendix (Section 1). As already mentioned, a major limitation of this work is that the experiments were not repeated and this is something that should be done before further exploratory work is carried out but the consistency seen between the findings of others and the separate RT-PCR and TLDA experiments is reassuring that the results can be trusted to some degree.

The most promising of the genes selected and tested in this study are listed in the Table 5.3 below. These genes were selected on the basis that they displayed an expression profile consistent with a CT gene (i.e. tightly restricted to testis in the normal tissues tested or were testis-selective) and were also expressed in at least one of the cancer samples tested in this study. Some well characterised CT genes are included (e.g., members of the *MAGE* family). This list provides a more focussed list of genes, which could be taken forward and tested on a larger cohort of cancer samples. Statistical analysis could then be applied to test the possible predictive power of these genes in different cancer types but the smaller number of genes will both decrease the costs of gene expression profiling and reduce the problems related to 'statistical noise' when dealing with a larger cohort of genes. It should be remembered that potentially relevant genes have not been included in this list, as the limited number of samples, and indeed cancer types, tested may have failed to reveal expression of certain genes that does exist. The cost of screening a large cohort of genes in a large cohort of cancers would be prohibitive, so the list provides a more manageable number to investigate further.

Some genes that display limited expression in normal tissues (perhaps excluding them as putative CT genes) but appear to be expressed with increased frequency in cancer have also been highlighted. For example, *C16orf46* is an uncharacterised gene but appears to be expressed frequently in cancers and is certainly worthy of further investigation. Likewise, other genes (e.g. *DAZL* and *SPACA3*) appear to be expressed with increased frequency in colon cancer samples but their expression in normal tissues means they may not serve well as genes that have an "on-off" cancer detection signal and/or as targets for immunotherapy.

Table 5.3. Possible germline genes to focus further research on

Germline-restricted	Germline-selective*	Colon cancer germline-restricted	Colon cancer germline-selective*
ARMC3	*CCDC110*	*C19orf67*	*C20orf195*
C1orf65	*DDX4*	*CCDC83*	*DAZ1*
CAPZA3	*GK2*	*IQCF1*	*DDX53*
CTAGE1	*HMGB4*	*KLF17*	*DKKL1*
FKBP9	*ISM2*	*LGALS14*	*HIST1H3C*
HORMAD2	*NBPF6*	*MAGEA1*	*MAGEA12*
LUZP4	*PRSS54*	*MAGEA2*	*PIWIL1*
LYZL6	*RBM46*	*MAGEA4*	*SPANX*
MBD3L1	*SEPT12*	*MAGEA9*	*STK31*
NFX2	*SMC1B*	*MAGEB6*	*SYCP3*
NT5C1B	*TEX15*	*ODF4*	*SYNGR4*
PFN3		*PLAC1*	*TEKT5*
PPP3R2		*SPZ1*	*TEX101*
RAD21L1		*TKTL2*	*TEX19*
STRA8			
SYCE1			
TGIF2LX			
TPTE			
TUBA3C			
VCY			

*Expression in up to two normal tissue types (i.e. testis-selective or testis/CNS-selective). NAT colonic samples are considered collectively as a single tissue type for this purpose. The genes in the third and fourth columns are more promising potential biomarkers for colorectal cancer based on increased frequency or level of expression in the colon cancer samples tested compared to the normal samples in this study.

Several candidate germline-associated genes that may be of relevance in cancer have been identified and the findings here with regards to previously identified cancer-associated genes are generally consistent, which confirms the validity of this exploratory study. Some genes have been associated with cancers that have not been described previously. For example,

LYZL6 with melanoma, *TEX101* with colon cancer, and *TPTE* with melanoma and ovarian cancer. The findings provide a platform on which to conduct further focussed research and the suggested list of germline genes on which to focus on is provided in Table 5.3. This research should attempt to clarify the functional relevance of these often poorly characterised genes, as well as investigate their predictive power, for example, in transcriptomic profiling of cancer, which can potentially improve the classification/stratification of tumours with real clinical benefit.

6. General Discussion

6.1. A move towards more precise personalised treatment in medicine

The emergence and expansion of omics technologies together with the increased computational power of bioinformatical techniques is heralding what could be a paradigm shift in the way medicine is practiced in the modern world. Historically medicine is driven by the symptoms of the patient to reach a diagnosis and then formulate a treatment plan. However, the increased predictive power of combining various disease stratification approaches is likely to lead to personalised treatment strategies for patients and particularly so in the field of oncology (Duffy, 2015; Vockley and Niederhuber, 2015). This so called 'precision medicine' or 'personalised medicine' approach will clearly be expensive but the reduction in waste it could bring about, coupled with an improvement in the efficacy of treatments, could well make it not only affordable but in fact desirable (Duffy, 2015). The same predictive tools that indicate treatment will not be effective for a certain group of patients, will at the same time reserve treatment strategies for patients that it is known will obtain clinical benefit. Such an approach would reduce the morbidity associated with treatments and promote rational decision making in cases where the therapy is of marginal or no benefit. In order to reach the goal of tailoring therapy effectively for the many different cancer types, it is likely that a considerable array of biomarkers will be needed. Work presented here will contribute to the growing list of putative cancer-specific biomarkers that may help form part of personalised treatment strategies. Close collaboration between clinical statisticians and bioinformaticians will be essential, as numerous sources of data are screened for this information, if we are to fulfil the promise of precision medicine.

One problem of screening such vast and expanding data sources is being able to see through the 'haze' to reach clinically relevant predictive endpoints. Rousseaux and colleagues (2013b) provided an example of how such problems can be overcome. In this study they focussed on a limited number of genes that ultimately led to identifying a manageable list of 26 genes that were of considerable predictive value in lung cancer patients. Similar gene lists for different cancer types may prove equally as powerful. The cost of such an approach could also be widely affordable – it is conceivable that 'omic-chips' that analyse expression (and mutation) profiles for a limited number of genes could be developed and analysed relatively cheaply. For example, if a diagnosis is suspected one 'chip' could be used (e.g., using a blood sample) or a

different one selected to screen a biopsy sample once a provisional diagnosis has been made. As the costs and accuracy of next generation sequencing using both liquid and FFPE samples improves (Pasquale et al., 2015; Sakai et al., 2015), it is conceivable that disease or cancer-type specific gene lists will not in fact be required. Companies are already trying to cash-in using such liquid biopsy techniques (for example, Biocept Inc.) highlighting that the approach provides the most up-to-date information to clinicians about the state of the patients' cancer. Mutations can evolve over the duration of disease and differ when cancer recurs following treatment – liquid biopsies can monitor this in real-time and help both guide therapy during treatment as well as pick-up on recurrences at an earlier stage. Such approaches could even be used for screening certain at-risk populations.

6.2. Germline genes in cancer

Work here has focussed on germline-associated genes. First I investigated a novel cohort of putative testis-specific genes and using clinically-derived cancer microarray datasets confirmed that some were significantly upregulated in prostate and ovarian cancer. More genes were upregulated in individual datasets from a wider degree of cancer types, indicating that they could be relevant markers for sub-categories of cancers. Dissecting the complexities of tumour subtypes together with intratumour heterogeneity is an ongoing challenge in the field of oncology and precision medicine (Blagden, 2015; Jamal-Hanjani et al., 2015; Seoane and De Mattos-Arruda, 2014). The use of liquid biopsies to detect circulating free DNA and/or circulating tumour cells may help provide some degree of resolution without the need for detailed analysis of numerous areas of a given tumour. Given the often very low frequency of expression of many germline genes in cancer makes it very challenging to attempt to reliably detect these genes from a liquid biopsy.

It is possible that relevant genes were excluded in the validation process adopted in this research. This is highlighted by the fact that there was a relatively poor overlap between genes that have been identified by the McFarlane group (Feichtinger et al., 2012; Sammut et al., 2014) and germline gene lists identified by the Knochbin group (Rousseaux et al., 2013). It was hoped that a more stringent approach would enable us to identify genes that had a clear on/off signal that could be used in a relatively simple test enabling more ready translation into the clinical setting. I attempted to collate a definitive list of germline genes and from this selected a cohort to investigate in normal and cancerous (predominantly CRC samples) human tissues. A clear

subpopulation of CRC did not emerge expressing germline genes more widely. This is possibly a limitation of the relatively small numbers of samples included in this study. However, it does act as an indicator of how similar technology can potentially be utilised. This list of genes proposed (see Table 4.8) may prove to be useful as a predictive panel of genes in different cancer types. This would require screening of a much larger number of cancer samples. This smaller list of ~50 targets makes such screening less financially constrained and provides a rational basis from which to take the exploratory study further.

There is mounting evidence for a functional role of germline and meiotic factors in cancer biology. Janic and colleagues (2010) were the first to establish that germline factors were oncogenic drivers in *D. melanogaster*. Specific proteins known to play important roles in meiosis have now been shown to have a cancer-phenotype promoting effect when present in tumours: Hop2 and Mnd1 promote maintenance of telomere length via the ALT pathway and HORMAD1 promotes genetic instability through inhibition of the most effective DNA repair pathway (Cho et al., 2014; Watkins et al., 2015). It has also been demonstrated that the human orthologues of the genes identified by Janic *el al.* (2010) are widely upregulated in human cancers suggesting a broader role of germline genes in cancer that is currently realised (Feichtinger et al., 2014). Two novel germline proteins, TEX19 and C20orf201, have been shown here to be present and potential immunological targets for cancer therapy. A functional role for these proteins in cancer, or indeed in normal physiology, remains to be clarified. It is possible that TEX19 contributes to genetic instability when present in cancer and although a putative meiosis-associated protein, it may not play a direct role in meiosis.

6.3. Concluding remarks and future directions

Precision, or personalised, medicine is based on the predictive power of existing biomarkers and stratification techniques. Gene expression is one important facet on which to focus stratification efforts and germline factors are emerging as useful in this regard. Future work should screen a larger cohort of cancer samples and attempt to link germline gene expression to clinical outcomes in a range of cancer types. A putative list of genes is proposed for further investigation based on their limited expression in the cancer samples tested in this research. An alternative approach would be to screen a larger number of genes but the cost of such an approach is likely to be prohibitory.

The function of many germline and CT genes remains poorly characterised despite the interest in CTAs as targets for immunotherapy for some time, some of which have already been targeted (Hunder et al., 2008; Simpson et al., 2005; Whitehurst, 2014). Establishing physiological roles for certain genes (for example, *TEX19* and *C16orf46*) which appear to be widely expressed in human cancers should be a priority. This may hint at important oncogenic roles. Two novel putative CTAs (C20orf201 and TEX19) with immunogenic potential have been investigated. The tissue-specificity of the presence of these antigens in human cancers and normal tissues should be tested through the development of monoclonal antibodies, which may facilitate their transition to use within the clinical setting either as predictive biomarkers and/or therapeutic targets.

1. References

Aarabi, M., Ousati-Ashtiani, Z., Nazarian, A., Modarressi, M.H. & Heidari, M. 2008, "Association of TGIFLX/Y mRNA expression with azoospermia in infertile men", *Molecular reproduction and development*, vol. 75, no. 12, pp. 1761-1766.

Abbotts, R., Thompson, N. & Madhusudan, S. 2014, "DNA repair in cancer: emerging targets for personalized therapy", *Cancer management and research*, vol. 6, pp. 77-92.

Adair, S.J. & Hogan, K.T. 2009, "Treatment of ovarian cancer cell lines with 5-aza-2'-deoxycytidine upregulates the expression of cancer-testis antigens and class I major histocompatibility complex-encoded molecules",*Cancer immunology, immunotherapy : CII*, vol. 58, no. 4, pp. 589-601.

Akiyama, Y., Komiyama, M., Miyata, H., Yagoto, M., Ashizawa, T., Iizuka, A., Oshita, C., Kume, A., Nogami, M., Ito, I., Watanabe, R., Sugino, T., Mitsuya, K., Hayashi, N., Nakasu, Y. & Yamaguchi, K. 2014, "Novel cancer-testis antigen expression on glioma cell lines derived from high-grade glioma patients", *Oncology reports*, vol. 31, no. 4, pp. 1683-1690.

Albrecht, M. 2009, "Insights into the nature of human testicular peritubular cells", *Annals of Anatomy = Anatomischer Anzeiger : Official Organ of the Anatomische Gesellschaft*, vol. 191, no. 6, pp. 532-540.

Albuquerque, C., Bakker, E.R., van Veelen, W. & Smits, R. 2011, "Colorectal cancers choosing sides", *Biochimica et biophysica acta*, vol. 1816, no. 2, pp. 219-231.

Almanzar, G., Olkhanud, P.B., Bodogai, M., Dell'agnola, C., Baatar, D., Hewitt, S.M., Ghimenton, C., Tummala, M.K., Weeraratna, A.T., Hoek, K.S., Kouprina, N., Larionov, V. & Biragyn, A. 2009, "Sperm-derived SPANX-B is a clinically relevant tumor antigen that is expressed in human tumors and readily recognized by human CD4+ and CD8+ T cells", *Clinical cancer research : an official journal of the American Association for Cancer Research*, vol. 15, no. 6, pp. 1954-1963.

Almatrafi, A., Feichtinger, J., Vernon, E.G., Escobar, N.G., Wakeman, J.A., Larcombe, L.D. & McFarlane, R.J. 2014, "Identification of a class of human cancer germline genes with transcriptional silencing refractory to the hypomethylating drug 5-aza-2'-deoxycytidine", *Oncoscience*, vol. 1, no. 11, pp. 745-750.

Almeida, L.G., Sakabe, N.J., deOliveira, A.R., Silva, M.C., Mundstein, A.S., Cohen, T., Chen, Y.T., Chua, R., Gurung, S., Gnjatic, S., Jungbluth, A.A., Caballero, O.L., Bairoch, A., Kiesler, E., White, S.L., Simpson, A.J., Old, L.J., Camargo, A.A. & Vasconcelos, A.T. 2009, "CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens", *Nucleic acids research*, vol. 37, no. Database issue, pp. D816-9.

Al-Sohaily, S., Biankin, A., Leong, R., Kohonen-Corish, M. & Warusavitarne, J. 2012, "Molecular pathways in colorectal cancer", *Journal of gastroenterology and hepatology*, vol. 27, no. 9, pp. 1423-1431.

Ansari, D., Andersson, R., Bauden, M.P., Andersson, B., Connolly, J.B., Welinder, C., Sasor, A. & Marko-Varga, G. 2015, "Protein deep sequencing applied to biobank samples from patients with pancreatic cancer", *Journal of cancer research and clinical oncology,* vol. 141, no. 2, pp. 369-380.

Aoki, K. & Taketo, M.M. 2007, "Adenomatous polyposis coli (APC): a multi-functional tumor suppressor gene", *Journal of cell science,* vol. 120, no. Pt 19, pp. 3327-3335.

Aris, M. & Barrio, M.M. 2015, "Combining immunotherapy with oncogene-targeted therapy: a new road for melanoma treatment", *Frontiers in immunology,* vol. 6, pp. 46.

Aruga, A., Takeshita, N., Kotera, Y., Okuyama, R., Matsushita, N., Ohta, T., Takeda, K. & Yamamoto, M. 2013, "Long-term Vaccination with Multiple Peptides Derived from Cancer-Testis Antigens Can Maintain a Specific T-cell Response and Achieve Disease Stability in Advanced Biliary Tract Cancer", *Clinical cancer research : an official journal of the American Association for Cancer Research,* vol. 19, no. 8, pp. 2224-2231.

Atanackovic, D., Blum, I., Cao, Y., Wenzel, S., Bartels, K., Faltz, C., Hossfeld, D.K., Hegewisch-Becker, S., Bokemeyer, C. & Leuwer, R. 2006, "Expression of cancer-testis antigens as possible targets for antigen-specific immunotherapy in head and neck squamous cell carcinoma", *Cancer biology & therapy,* vol. 5, no. 9, pp. 1218-1225.

Atanackovic, D., Luetkens, T., Kloth, B., Fuchs, G., Cao, Y., Hildebrandt, Y., Meyer, S., Bartels, K., Reinhard, H., Lajmi, N., Hegewisch-Becker, S., Schilling, G., Platzbecker, U., Kobbe, G., Schroeder, T., Bokemeyer, C. & Kroger, N. 2011, "Cancer-testis antigen expression and its epigenetic modulation in acute myeloid leukemia", *American Journal of Hematology,* vol. 86, no. 11, pp. 918-922.

Ayyoub, M., Scarlata, C.M., Hamai, A., Pignon, P. & Valmori, D. 2014, "Expression of MAGE-A3/6 in primary breast cancer is associated with hormone receptor negative status, high histologic grade, and poor survival",*Journal of immunotherapy (Hagerstown, Md.: 1997),* vol. 37, no. 2, pp. 73-76.

Baillet, A. & Mandon-Pepin, B. 2012, "Mammalian ovary differentiation - a focus on female meiosis", *Molecular and cellular endocrinology,* vol. 356, no. 1-2, pp. 13-23.

Baker, C.L., Walker, M., Kajita, S., Petkov, P.M. & Paigen, K. 2014, "PRDM9 binding organizes hotspot nucleosomes and limits Holliday junction migration", *Genome research,* vol. 24, no. 5, pp. 724-732.

Baker, M. 2015, "Reproducibility crisis: Blame it on the antibodies", *Nature,* vol. 521, no. 7552, pp. 274-276.

Banerjee, R., Russo, N., Liu, M., Basrur, V., Bellile, E., Palanisamy, N., Scanlon, C.S., van Tubergen, E., Inglehart, R.C., Metwally, T., Mani, R.S., Yocum, A., Nyati, M.K., Castilho, R.M., Varambally, S., Chinnaiyan, A.M. & D'Silva, N.J. 2014, "TRIP13 promotes error-prone nonhomologous end joining and induces chemoresistance in head and neck cancer", *Nature communications,* vol. 5, pp. 4527.

Baudat, F., Imai, Y. & de Massy, B. 2013, "Meiotic recombination in mammals: localization and regulation", *Nature reviews.Genetics,* vol. 14, no. 11, pp. 794-806.

Behnen, M., Murk, K., Kursula, P., Cappallo-Obermann, H., Rothkegel, M., Kierszenbaum, A.L. & Kirchhoff, C. 2009, "Testis-expressed profilins 3 and 4 show distinct functional characteristics and localize in the acroplaxome-manchette complex in spermatids", *BMC cell biology,* vol. 10, pp. 34-2121-10-34.

Bendell, J., Saleh, M., Rose, A.A., Siegel, P.M., Hart, L., Sirpal, S., Jones, S., Green, J., Crowley, E., Simantov, R., Keler, T., Davis, T. & Vahdat, L. 2014, "Phase I/II study of the antibody-drug conjugate glembatumumab vedotin in patients with locally advanced or metastatic breast cancer", *Journal of clinical oncology : official journal of the American Society of Clinical Oncology,* vol. 32, no. 32, pp. 3619-3625.

Bentz, S., Cee, A., Endlicher, E., Wojtal, K.A., Naami, A., Pesch, T., Lang, S., Schubert, P., Fried, M., Weber, A., Coy, J.F., Goelder, S., Knuchel, R., Hausmann, M. & Rogler, G. 2013, "Hypoxia induces the expression of transketolase-like 1 in human colorectal cancer", *Digestion,* vol. 88, no. 3, pp. 182-192.

Berns, E.M. & Bowtell, D.D. 2012, "The changing view of high-grade serous ovarian cancer", *Cancer research,* vol. 72, no. 11, pp. 2701-2704.

Bettoni, F., Filho, F.C., Grosso, D.M., Galante, P.A., Parmigiani, R.B., Geraldo, M.V., Henrique-Silva, F., Oba-Shinjo, S.M., Marie, S.K., Soares, F.A., Brentani, H.P., Simpson, A.J., de Souza, S.J. & Camargo, A.A. 2009, "Identification of FAM46D as a novel cancer/testis antigen using EST data and serological analysis", *Genomics,* vol. 94, no. 3, pp. 153-160.

Bhan, S., Chuang, A., Negi, S.S., Glazer, C.A. & Califano, J.A. 2012, "MAGEA4 induces growth in normal oral keratinocytes by inhibiting growth arrest and apoptosis", *Oncology reports,* vol. 28, no. 4, pp. 1498-1502.

Biomarkers Definitions Working Group. 2001, "Biomarkers and surrogate endpoints: preferred definitions and conceptual framework", *Clinical pharmacology and therapeutics,* vol. 69, no. 3, pp. 89-95.

Blagden, S.P. 2015, "Harnessing Pandemonium: The Clinical Implications of Tumor Heterogeneity in Ovarian Cancer", *Frontiers in oncology,* vol. 5, pp. 149.

Blankenstein, T., Coulie, P.G., Gilboa, E. & Jaffee, E.M. 2012, "The determinants of tumour immunogenicity", *Nature reviews.Cancer,* vol. 12, no. 4, pp. 307-313.

Bleyer, A., Barr, R., Hayes-Lattin, B., Thomas, D., Ellis, C., Anderson, B. & Biology and Clinical Trials Subgroups of the US National Cancer Institute Progress Review Group in Adolescent and Young Adult Oncology 2008, "The distinctive biology of cancer in adolescents and young adults", *Nature reviews.Cancer,* vol. 8, no. 4, pp. 288-298.

Boel, P., Wildmann, C., Sensi, M.L., Brasseur, R., Renauld, J.C., Coulie, P., Boon, T. & van der Bruggen, P. 1995, "BAGE: a new gene encoding an antigen recognized on human melanomas by cytolytic T lymphocytes",*Immunity,* vol. 2, no. 2, pp. 167-175.

Boggon, T.J., Shan, W.S., Santagata, S., Myers, S.C. & Shapiro, L. 1999, "Implication of tubby proteins as transcription factors by structure-based functional analysis", *Science (New York, N.Y.)*, vol. 286, no. 5447, pp. 2119-2125.

Boghossian, S., Robinson, S., Von Delwig, A., Manas, D. & White, S. 2012, "Immunotherapy for treating metastatic colorectal cancer", *Surgical oncology*, vol. 21, no. 2, pp. 67-77.

Boland, C.R. & Goel, A. 2010, "Microsatellite instability in colorectal cancer", *Gastroenterology*, vol. 138, no. 6, pp. 2073-2087.e3.

Boncheva, V., Bonney, S.A., Brooks, S.E., Tangney, M., O'Sullivan, G., Mirnezami, A. & Guinn, B.A. 2013, "New targets for the immunotherapy of colon cancer-does reactive disease hold the answer?", *Cancer gene therapy*,vol. 20, no. 3, pp. 157-168.

Bordeaux, J., Welsh, A., Agarwal, S., Killiam, E., Baquero, M., Hanna, J., Anagnostou, V. & Rimm, D. 2010, "Antibody validation", *BioTechniques*, vol. 48, no. 3, pp. 197-209.

Borovski, T., De Sousa E Melo, F., Vermeulen, L. & Medema, J.P. 2011, "Cancer stem cell niche: the place to be", *Cancer research*, vol. 71, no. 3, pp. 634-639.

Bradbury, A. & Pluckthun, A. 2015, "Reproducibility: Standardize antibodies used in research", *Nature*, vol. 518, no. 7537, pp. 27-29.

Brebi, P., Maldonado, L., Noordhuis, M.G., Ili, C., Leal, P., Garcia, P., Brait, M., Ribas, J., Michailidi, C., Perez, J., Soudry, E., Tapia, O., Guzman, P., Munoz, S., Van Neste, L., Van Criekinge, W., Irizarry, R., Sidransky, D., Roa, J.C. & Guerrero-Preston, R. 2014, "Genome-wide methylation profiling reveals Zinc finger protein 516 (ZNF516) and FK-506-binding protein 6 (FKBP6) promoters frequently methylated in cervical neoplasia, associated with HPV status and ethnicity in a Chilean population", *Epigenetics : official journal of the DNA Methylation Society*, vol. 9, no. 2, pp. 308-317.

Brown, M.S. & Bishop, D.K. 2014, "DNA strand exchange and RecA homologs in meiosis", *Cold Spring Harbor perspectives in biology*, vol. 7, no. 1, pp. a016659.

Burgdorf, S.K., Fischer, A., Myschetzky, P.S., Munksgaard, S.B., Zocca, M.B., Claesson, M.H. & Rosenberg, J. 2008, "Clinical responses in patients with advanced colorectal cancer to a dendritic cell based vaccine", *Oncology reports*, vol. 20, no. 6, pp. 1305-1311.

Caballero, O.L. & Chen, Y.T. 2009, "Cancer/testis (CT) antigens: potential targets for immunotherapy", *Cancer science*, vol. 100, no. 11, pp. 2014-2021.

Caiado F, Silva-Santos B & Norell K. 2015, "Intra-tumour heterogeneity – going beyond genetics", *The FEBS journal*, vol. 283, no. 12, pp. 2245-2258.

Calon, A., Lonardo, E., Berenguer-Llergo, A., Espinet, E., Hernando-Momblona, X., Iglesias, M., Sevillano, M., Palomo-Ponce, S., Tauriello, D.V., Byrom, D., Cortina, C., Morral, C., Barcelo, C., Tosi, S., Riera, A., Attolini, C.S., Rossell, D., Sancho, E. & Batlle, E. 2015, "Stromal gene expression defines poor-prognosis subtypes in colorectal cancer", *Nature genetics*, vol. 47, no. 4, pp. 320-329.

Cappell, K.M., Sinnott, R., Taus, P., Maxfield, K., Scarbrough, M. & Whitehurst, A.W. 2012, "Multiple cancer testis antigens function to support tumor cell mitotic fidelity", *Molecular and cellular biology,* vol. 32, no. 20, pp. 4131-4140.

Cargnello, M., Tcherkezian, J. & Roux, P.P. 2015, "The expanding role of mTOR in cancer cell growth and proliferation", *Mutagenesis,* vol. 30, no. 2, pp. 169-176.

Carreno, B.M., Magrini, V., Becker-Hapak, M., Kaabinejadian, S., Hundal, J., Petti, A.A., Ly, A., Lie, W.R., Hildebrand, W.H., Mardis, E.R. & Linette, G.P. 2015, "A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells", *Science (New York, N.Y.),* .

Carroll, K., Gomez, C. & Shapiro, L. 2004, "Tubby proteins: the plot thickens", *Nature reviews.Molecular cell biology,* vol. 5, no. 1, pp. 55-63.

Celerin, M., Merino, S.T., Stone, J.E., Menzie, A.M. & Zolan, M.E. 2000, "Multiple roles of Spo11 in meiotic chromosome behavior", *The EMBO journal,* vol. 19, no. 11, pp. 2739-2750.

Chalmel, F., Rolland, A.D., Niederhauser-Wiederkehr, C., Chung, S.S., Demougin, P., Gattiker, A., Moore, J., Patard, J.J., Wolgemuth, D.J., Jegou, B. & Primig, M. 2007, "The conserved transcriptome in human and rodent male gametogenesis", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 104, no. 20, pp. 8346-8351.

Chapman, C.J., Healey, G.F., Murray, A., Boyle, P., Robertson, C., Peek, L.J., Allen, J., Thorpe, A.J., Hamilton-Fairley, G., Parsy-Kowalska, C.B., MacDonald, I.K., Jewell, W., Maddison, P. & Robertson, J.F. 2012, "EarlyCDT(R)-Lung test: improved clinical utility through additional autoantibody assays", *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine,* vol. 33, no. 5, pp. 1319-1326.

Chen, H., Rossier, C., Morris, M.A., Scott, H.S., Gos, A., Bairoch, A. & Antonarakis, S.E. 1999, "A testis-specific gene, TPTE, encodes a putative transmembrane tyrosine phosphatase and maps to the pericentromeric region of human chromosomes 21 and 13, and to chromosomes 15, 22, and Y", *Human genetics,* vol. 105, no. 5, pp. 399-409.

Chen, H., Yue, J.X., Yang, S.H., Ding, H., Zhao, R.W. & Zhang, S. 2009, "Overexpression of transketolase-like gene 1 is associated with cell proliferation in uterine cervix cancer", *Journal of experimental & clinical cancer research : CR,* vol. 28, pp. 43-9966-28-43.

Chen, L., Zhou, W.B., Zhao, Y., Liu, X.A., Ding, Q., Zha, X.M. & Wang, S. 2012, "Cancer/testis antigen SSX2 enhances invasiveness in MCF-7 cells by repressing ERalpha signaling", *International journal of oncology,* vol. 40, no. 6, pp. 1986-1994.

Chen, S.R. & Liu, Y.X. 2015, "Regulation of spermatogonial stem cell self-renewal and spermatocyte meiosis by Sertoli cell signaling", *Reproduction (Cambridge, England),* vol. 149, no. 4, pp. R159-R167.

Chen, Y., Schlessinger, D. & Nagaraja, R. 2013, "T antigen transformation reveals Tp53/RB-dependent route to PLAC1 transcription activation in primary fibroblasts", *Oncogenesis*, vol. 2, pp. e67.

Chen, Y.T., Scanlan, M.J., Sahin, U., Tureci, O., Gure, A.O., Tsang, S., Williamson, B., Stockert, E., Pfreundschuh, M. & Old, L.J. 1997, "A testicular antigen aberrantly expressed in human cancers detected by autologous antibody screening", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 5, pp. 1914-1918.

Chen, Y.T., Scanlan, M.J., Venditti, C.A., Chua, R., Theiler, G., Stevenson, B.J., Iseli, C., Gure, A.O., Vasicek, T., Strausberg, R.L., Jongeneel, C.V., Old, L.J. & Simpson, A.J. 2005, "Identification of cancer/testis-antigen genes by massively parallel signature sequencing", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 22, pp. 7940-7945.

Chen, Y.T., Venditti, C.A., Theiler, G., Stevenson, B.J., Iseli, C., Gure, A.O., Jongeneel, C.V., Old, L.J. & Simpson, A.J. 2005, "Identification of CT46/HORMAD1, an immunogenic cancer/testis antigen encoding a putative meiosis-related protein", *Cancer immunity*, vol. 5, pp. 9.

Chen, Z., Li, M., Yuan, Y., Wang, Q., Yan, L. & Gu, J. 2010, "Cancer/testis antigens and clinical risk factors for liver metastasis of colorectal cancer: a predictive panel", *Diseases of the colon and rectum*, vol. 53, no. 1, pp. 31-38.

Cheng, Y.H., Wong, E.W. & Cheng, C.Y. 2011, "Cancer/testis (CT) antigens, carcinogenesis and spermatogenesis", *Spermatogenesis*, vol. 1, no. 3, pp. 209-220.

Cho C. 2012, "Testticular and epididymal ADAMs; expression and function during fertilization", *Nature reviews.Urology*, vol. 9, no. 10, pp. 550-560.

Cho, H., Noh, K.H., Chung, J.Y., Takikita, M., Chung, E.J., Kim, B.W., Hewitt, S.M., Kim, T.W. & Kim, J.H. 2014, "Synaptonemal complex protein 3 is a prognostic marker in cervical cancer", *PloS one*, vol. 9, no. 6, pp. e98712.

Cho, N.W., Dilley, R.L., Lampson, M.A. & Greenberg, R.A. 2014, "Interchromosomal homology searches drive directional ALT telomere movement and synapsis", *Cell*, vol. 159, no. 1, pp. 108-121.

Choi, J. & Chang, H. 2012, "The expression of MAGE and SSX, and correlation of COX2, VEGF, and survivin in colorectal cancer", *Anticancer Research*, vol. 32, no. 2, pp. 559-564.

Choi, H., Jin, S., Kwon, JT., Kim, J., Jeong, J., Kim, J., Jeon, S., Park, ZY., Jung, KJ., Park, K., Cho, C. 2016, "Characterization of mammalian ADAM2 and its absence from human sperm", *PloS one*, vol. 11, no. 6, e0158321.

Chomez, P., De Backer, O., Bertrand, M., De Plaen, E., Boon, T. & Lucas, S. 2001, "An overview of the MAGE gene family with the identification of all human members of the family", *Cancer research*, vol. 61, no. 14, pp. 5544-5551.

Chung, J.Y., Kitano, H., Takikita, M., Cho, H., Noh, K.H., Kim, T.W., Ylaya, K., Hanaoka, J., Fukuoka, J. & Hewitt, S.M. 2013, "Synaptonemal complex protein 3 as a novel prognostic marker in early stage non-small cell lung cancer", *Human pathology*, vol. 44, no. 4, pp. 472-479.

Cicinnati, VR., Shen, Q., Sotiropoulos, GC., Radtke, A., Gerken, G. & Beckebaum S. 2008, "Validation of putative reference genes for gene expression studies in human hepatocellular carcinoma using real-time quantitative RT-PCR", *BMC Cancer*, vol. 8, 350.

Civelekoglu-Scholey, G. & Cimini, D. 2014, "Modelling chromosome dynamics in mitosis: a historical perspective on models of metaphase and anaphase in eukaryotic cells", *Interface focus*, vol. 4, no. 3, pp. 20130073.

Cohen, J.I., Mocarski, E.S., Raab-Traub, N., Corey, L. & Nabel, G.J. 2013, "The need and challenges for development of an Epstein-Barr virus vaccine", *Vaccine*, vol. 31 Suppl 2, pp. B194-6.

Compton, C.C. & Greene, F.L. 2004, "The staging of colorectal cancer: 2004 and beyond", *CA: a cancer journal for clinicians*, vol. 54, no. 6, pp. 295-308.

Constantinidou, A. & Smith, I. 2011, "Is there a case for anti-HER2 therapy without chemotherapy in early breast cancer?", *Breast (Edinburgh, Scotland)*, vol. 20 Suppl 3, pp. S158-61.

Costa, F.F., Le Blanc, K. & Brodin, B. 2007, "Concise review: cancer/testis antigens, stem cells, and cancer", *Stem cells (Dayton, Ohio)*, vol. 25, no. 3, pp. 707-711.

Coulie, P.G., Van den Eynde, B.J., van der Bruggen, P. & Boon, T. 2014, "Tumour antigens recognized by T lymphocytes: at the core of cancer immunotherapy", *Nature reviews.Cancer*, vol. 14, no. 2, pp. 135-146.

Coward, J. & Harding, A. 2014, "Size Does Matter: Why Polyploid Tumor Cells are Critical Drug Targets in the War on Cancer", *Frontiers in oncology*, vol. 4, pp. 123.

Crichton, J.H., Dunican, D.S., Maclennan, M., Meehan, R.R. & Adams, I.R. 2014, "Defending the genome from the enemy within: mechanisms of retrotransposon suppression in the mouse germline", *Cellular and molecular life sciences : CMLS*, vol. 71, no. 9, pp. 1581-1605.

Dakshinamurthy, A.G., Ramesar, R., Goldberg, P. & Blackburn, J.M. 2008, "Infrequent and low expression of cancer-testis antigens located on the X chromosome in colorectal cancer: implications for immunotherapy in South African populations", *Biotechnology journal*, vol. 3, no. 11, pp. 1417-1423.

Dalerba, P., Maccalli, C., Casati, C., Castelli, C. & Parmiani, G. 2003, "Immunology and immunotherapy of colorectal cancer", *Critical reviews in oncology/hematology*, vol. 46, no. 1, pp. 33-57.

Daniel, K., Lange, J., Hached, K., Fu, J., Anastassiadis, K., Roig, I., Cooke, H.J., Stewart, A.F., Wassmann, K., Jasin, M., Keeney, S. & Toth, A. 2011, "Meiotic homologue alignment and its quality surveillance are controlled by mouse HORMAD1", *Nature cell biology*, vol. 13, no. 5, pp. 599-610.

Daniel, K., Trankner, D., Wojtasz, L., Shibuya, H., Watanabe, Y., Alsheimer, M. & Toth, A. 2014, "Mouse CCDC79 (TERB1) is a meiosis-specific telomere associated protein", *BMC cell biology*, vol. 15, pp. 17-2121-15-17.

Daudi, S., Eng, K.H., Mhawech-Fauceglia, P., Morrison, C., Miliotto, A., Beck, A., Matsuzaki, J., Tsuji, T., Groman, A., Gnjatic, S., Spagnoli, G., Lele, S. & Odunsi, K. 2014, "Expression and immune responses to MAGE antigens predict survival in epithelial ovarian cancer", *PloS one*, vol. 9, no. 8, pp. e104099.

Davidoff, M.S., Middendorff, R., Kofuncu, E., Muller, D., Jezek, D. & Holstein, A.F. 2002, "Leydig cells of the human testis possess astrocyte and oligodendrocyte marker molecules", *Acta Histochemica*, vol. 104, no. 1, pp. 39-49.

Davis, D.M. 2007, "Intercellular transfer of cell-surface proteins is common and can affect many stages of an immune response", *Nature reviews.Immunology*, vol. 7, no. 3, pp. 238-243.

Dawson, P.M., Habib, N.A., Rees, H.C. & Wood, C.B. 1987, "Mucosal field change in colorectal cancer", *American Journal of Surgery*, vol. 153, no. 3, pp. 281-284.

de Carvalho, F., Vettore, A.L. & Colleoni, G.W. 2012, "Cancer/Testis Antigen MAGE-C1/CT7: new target for multiple myeloma therapy", *Clinical & developmental immunology*, vol. 2012, pp. 257695.

de Martel, C., Ferlay, J., Franceschi, S., Vignat, J., Bray, F., Forman, D. & Plummer, M. 2012, "Global burden of cancers attributable to infections in 2008: a review and synthetic analysis", *The Lancet.Oncology*, vol. 13, no. 6, pp. 607-615.

de Massy, B. 2013, "Initiation of meiotic recombination: how and where? Conservation and specificities among eukaryotes", *Annual Review of Genetics*, vol. 47, pp. 563-599.

Decock, A., Ongenaert, M., Hoebeeck, J., De Preter, K., Van Peer, G., Van Criekinge, W., Ladenstein, R., Schulte, J.H., Noguera, R., Stallings, R.L., Van Damme, A., Laureys, G., Vermeulen, J., Van Maerken, T., Speleman, F. & Vandesompele, J. 2012, "Genome-wide promoter methylation analysis in neuroblastoma identifies prognostic methylation biomarkers", *Genome biology*, vol. 13, no. 10, pp. R95-2012-13-10-r95.

Dekker, SJ., Kampinga, HH. & Bergink, S. 2015, "DNAJs: more than substrate delivery to HSPA", *Frontiers in molecular biosciences*, vol. 2, 35.

Devor, E.J., Reyes, H.D., Santillan, D.A., Santillan, M.K., Onukwugha, C., Goodheart, M.J. & Leslie, K.K. 2014, "Placenta-specific protein 1: a potential key to many oncofetal-placental OB/GYN research questions", *Obstetrics and gynecology international*, vol. 2014, pp. 678984.

Diamandis, E.P. 2014, "Towards identification of true cancer biomarkers", *BMC medicine,* vol. 12, no. 1, pp. 156-014-0156-8.

Dianatpour, M., Mehdipour, P., Nayernia, K., Mobasheri, M.B., Ghafouri-Fard, S., Savad, S. & Modarressi, M.H. 2012, "Expression of Testis Specific Genes TSGA10, TEX101 and ODF3 in Breast Cancer", *Iranian Red Crescent medical journal,* vol. 14, no. 11, pp. 722-726.

Diaz-Moralli, S., Tarrado-Castellarnau, M., Alenda, C., Castells, A. & Cascante, M. 2011, "Transketolase-like 1 expression is modulated during colorectal cancer progression and metastasis formation", *PloS one,* vol. 6, no. 9, pp. e25323.

Dong, X.Y., Su, Y.R., Qian, X.P., Yang, X.A., Pang, X.W., Wu, H.Y. & Chen, W.F. 2003, "Identification of two novel CT antigens and their capacity to elicit antibody response in hepatocellular carcinoma patients", *British journal of cancer,* vol. 89, no. 2, pp. 291-297.

Drake, C.G., Lipson, E.J. & Brahmer, J.R. 2014, "Breathing new life into immunotherapy: review of melanoma, lung and kidney cancer", *Nature reviews.Clinical oncology,* vol. 11, no. 1, pp. 24-37.

Dray, E., Etchin, J., Wiese, C., Saro, D., Williams, G.J., Hammel, M., Yu, X., Galkin, V.E., Liu, D., Tsai, M.S., Sy, S.M., Schild, D., Egelman, E., Chen, J. & Sung, P. 2010, "Enhancement of RAD51 recombinase activity by the tumor suppressor PALB2", *Nature structural & molecular biology,* vol. 17, no. 10, pp. 1255-1259.

Duffy, D.J. 2015, "Problems, challenges and promises: perspectives on precision medicine", *Briefings in bioinformatics, .*

Dunlop, M.H., Dray, E., Zhao, W., Tsai, M.S., Wiese, C., Schild, D. & Sung, P. 2011, "RAD51-associated protein 1 (RAD51AP1) interacts with the meiotic recombinase DMC1 through a conserved motif", *The Journal of biological chemistry,* vol. 286, no. 43, pp. 37328-37334.

Duro, E. & Marston, A.L. 2015, "From equator to pole: splitting chromosomes in mitosis and meiosis", *Genes & development,* vol. 29, no. 2, pp. 109-122.

Eisenberg E. & Levanon EY. 2013, "Human housekeeping genes, revisited", *Trends in Genetics*, vol. 29, no. 10, pp. 569-574.

Eldai, H., Periyasamy, S., Al Qarni, S., Al Rodayyan, M., Muhammed Mustafa, S., Deeb, A., Al Sheikh, E., Afzal, M., Johani, M., Yousef, Z. & Aziz, M.A. 2013, "Novel genes associated with colorectal cancer are revealed by high resolution cytogenetic analysis in a patient specific manner", *PloS one,* vol. 8, no. 10, pp. e76251.

Eng, K.H. & Tsuji, T. 2014, "Differential antigen expression profile predicts immunoreactive subset of advanced ovarian cancers", *PloS one,* vol. 9, no. 11, pp. e111586.

Erenpreisa, J., Salmina, K., Huna, A., Jackson, T.R., Vazquez-Martin, A. & Cragg, M.S. 2014, "The "virgin birth", polyploidy, and the origin of cancer", *Oncoscience,* vol. 2, no. 1, pp. 3-14.

Erickson, B.K., Conner, M.G. & Landen, C.N.,Jr 2013, "The role of the fallopian tube in the origin of ovarian cancer", *American Journal of Obstetrics and Gynecology,* vol. 209, no. 5, pp. 409-414.

Eskenazi, B., Wyrobek, A.J., Sloter, E., Kidd, S.A., Moore, L., Young, S. & Moore, D. 2003, "The association of age and semen quality in healthy men", *Human reproduction (Oxford, England),* vol. 18, no. 2, pp. 447-454.

Essand, M. & Loskog, A.S. 2013, "Genetically engineered T cells for the treatment of cancer", *Journal of internal medicine,* vol. 273, no. 2, pp. 166-181.

Evans, J.P. 2001, "Fertilin beta and other ADAMs as integrin ligands: insights into cell adhesion and fertilization", *BioEssays : news and reviews in molecular, cellular and developmental biology,* vol. 23, no. 7, pp. 628-639.

Fang, P., Xu, W., Li, D., Zhao, X., Dai, J., Wang, Z., Yan, X., Qin, M., Zhang, Y., Xu, C., Wang, L. & Qiao, Z. 2015, "A novel acrosomal protein, IQCF1, involved in sperm capacitation and the acrosome reaction", *Andrology,*vol. 3, no. 2, pp. 332-344.

Fant, M., Farina, A., Nagaraja, R. & Schlessinger, D. 2010, "PLAC1 (Placenta-specific 1): a novel, X-linked gene with roles in reproductive and cancer biology", *Prenatal diagnosis,* vol. 30, no. 6, pp. 497-502.

Fearon, E.R. & Vogelstein, B. 1990, "A genetic model for colorectal tumorigenesis", *Cell,* vol. 61, no. 5, pp. 759-767.

Feichtinger, J., Aldeailej, I., Anderson, R., Almutairi, M., Almatrafi, A., Alsiwiehri, N., Griffiths, K., Stuart, N., Wakeman, J.A., Larcombe, L. & McFarlane, R.J. 2012b, "Meta-analysis of clinical data using human meiotic genes identifies a novel cohort of highly restricted cancer-specific marker genes", *Oncotarget,* vol. 3, no. 8, pp. 843-853.

Feichtinger, J., Larcombe, L. & McFarlane, R.J. 2014, "Meta-analysis of expression of l(3)mbt tumor-associated germline genes supports the model that a soma-to-germline transition is a hallmark of human cancers",*International journal of cancer.Journal international du cancer,* vol. 134, no. 10, pp. 2359-2365.

Feichtinger, J., McFarlane, R.J. & Larcombe, L.D. 2014, "CancerEST: a web-based tool for automatic meta-analysis of public EST data", *Database : the journal of biological databases and curation,* vol. 2014, no. 0, pp. bau024.

Feichtinger, J., McFarlane, R.J. & Larcombe, L.D. 2012a, "CancerMA: a web-based tool for automatic meta-analysis of public cancer microarray data", *Database : the journal of biological databases and curation,* vol. 2012, pp. bas055.

Feinberg, A.P. & Vogelstein, B. 1983, "Hypomethylation distinguishes genes of some human cancers from their normal counterparts", *Nature,* vol. 301, no. 5895, pp. 89-92.

Ferlin, A., Selice, R., Carraro, U. & Foresta, C. 2013, "Testicular function and bone metabolism--beyond testosterone", *Nature reviews.Endocrinology,* vol. 9, no. 9, pp. 548-554.

Ferns, JE., Theisen, CS., Fibuch, EE. & Seidler, NW. 2012, “Protection against protein aggregation by alpha-crystallin as a mechanism of preconditioning”, *Neurochemical research*, vol. 37, no. 2, pp. 244-252.

Firat-Karalar, EN., Sante, J., Elliott, S. & Stearns T. 2014, “Proteomic analysis of mammalian sperm cells identifies new components of the centrosome”, *Journal of cell science*, vol. 127, no. 19, pp. 4128-4133.

Flemming, A. 2015, "Cancer stem cells: Targeting the root of cancer relapse", *Nature reviews.Drug discovery*, vol. 14, no. 3, pp. 165.

Fratta, E., Coral, S., Covre, A., Parisi, G., Colizzi, F., Danielli, R., Nicolay, H.J., Sigalotti, L. & Maio, M. 2011, "The biology of cancer testis antigens: putative function, regulation and therapeutic potential", *Molecular oncology*,vol. 5, no. 2, pp. 164-182.

Fraune, J., Schramm, S., Alsheimer, M. & Benavente, R. 2012, "The mammalian synaptonemal complex: protein components, assembly and role in meiotic recombination", *Experimental cell research*, vol. 318, no. 12, pp. 1340-1346.

Frazer, I.H. 2004, "Prevention of cervical cancer through papillomavirus vaccination", *Nature reviews.Immunology*, vol. 4, no. 1, pp. 46-54.

Frede, J., Adams, D.J. & Jones, P.H. 2014, "Mutation, clonal fitness and field change in epithelial carcinogenesis", *The Journal of pathology*, vol. 234, no. 3, pp. 296-301.

French, C.A. 2012, "Pathogenesis of NUT midline carcinoma", *Annual review of pathology*, vol. 7, pp. 247-265.

Friedmann-Morvinski, D. & Verma, I.M. 2014, "Dedifferentiation and reprogramming: origins of cancer stem cells", *EMBO reports*, vol. 15, no. 3, pp. 244-253.

Fromm, M.F. 2004, "Importance of P-glycoprotein at blood-tissue barriers", *Trends in pharmacological sciences*, vol. 25, no. 8, pp. 423-429.

Fukuyama, T., Yamazaki, T., Fujita, T., Uematsu, T., Ichiki, Y., Kaneko, H., Suzuki, T. & Kobayashi, N. 2012, "Helicobacter pylori, a carcinogen, induces the expression of melanoma antigen-encoding gene (Mage)-A3, a cancer/testis antigen", *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*, vol. 33, no. 6, pp. 1881-1887.

Gala, M. & Chung, D.C. 2011, "Hereditary colon cancer syndromes", *Seminars in oncology*, vol. 38, no. 4, pp. 490-499.

Gao S, Zhu L, Feng H, Hu Z, Jin S, Song Z, Liu D, Liu J, Hao Y, Li X & Lin B. 2016, “Gene expression profile analysis in response to α1,2-fucosyl transferase (FUT1) gene transfection in epithelial ovarian carcinoma cells”, *Tumour Biology*, [Epub ahead of print].

Gattinoni, L., Powell, D.J.,Jr, Rosenberg, S.A. & Restifo, N.P. 2006, "Adoptive immunotherapy for cancer: building on success", *Nature reviews.Immunology,* vol. 6, no. 5, pp. 383-393.

Gaugler, B., Van den Eynde, B., van der Bruggen, P., Romero, P., Gaforio, J.J., De Plaen, E., Lethe, B., Brasseur, F. & Boon, T. 1994, "Human gene MAGE-3 codes for an antigen recognized on a melanoma by autologous cytolytic T lymphocytes", *The Journal of experimental medicine,* vol. 179, no. 3, pp. 921-930.

Ghafouri-Fard, S., Modarressi, M.H. & Yazarloo, F. 2012, "Expression of testis-specific genes, TEX101 and ODF4, in chronic myeloid leukemia and evaluation of TEX101 immunogenicity", *Annals of Saudi medicine,* vol. 32, no. 3, pp. 256-261.

Ghafouri-Fard, S., Ousati Ashtiani, Z., Sabah Golian, B., Hasheminasab, S.M. & Modarressi, M.H. 2010, "Expression of two testis-specific genes, SPATA19 and LEMD1, in prostate cancer", *Archives of Medical Research,* vol. 41, no. 3, pp. 195-200.

Giacinti, C. & Giordano, A. 2006, "RB and cell cycle progression", *Oncogene,* vol. 25, no. 38, pp. 5220-5227.

Gonzalez de Castro, D., Clarke, P.A., Al-Lazikani, B. & Workman, P. 2013, "Personalized cancer medicine: molecular diagnostics, predictive biomarkers, and drug resistance", *Clinical pharmacology and therapeutics,* vol. 93, no. 3, pp. 252-259.

Gray, A., Raff, A.B., Chiriva-Internati, M., Chen, S.Y. & Kast, W.M. 2008, "A paradigm shift in therapeutic vaccination of cancer patients: the need to apply therapeutic vaccination strategies in the preventive setting",*Immunological reviews,* vol. 222, pp. 316-327.

Guertin, D.A. & Sabatini, D.M. 2007, "Defining the role of mTOR in cancer", *Cancer cell,* vol. 12, no. 1, pp. 9-22.

Guillaudeux, T., Gomez, E., Onno, M., Drenou, B., Segretain, D., Alberti, S., Lejeune, H., Fauchet, R., Jegou, B. & Le Bouteiller, P. 1996, "Expression of HLA class I genes in meiotic and post-meiotic human spermatogenic cells", *Biology of reproduction,* vol. 55, no. 1, pp. 99-110.

Gure, A.O., Tureci, O., Sahin, U., Tsang, S., Scanlan, M.J., Jager, E., Knuth, A., Pfreundschuh, M., Old, L.J. & Chen, Y.T. 1997, "SSX: a multigene family with several members transcribed in normal testis and human cancer",*International journal of cancer.Journal international du cancer,* vol. 72, no. 6, pp. 965-971.

Haffner, A.C., Tassis, A., Zepter, K., Storz, M., Tureci, O., Burg, G. & Nestle, F.O. 2002, "Expression of cancer/testis antigens in cutaneous T cell lymphomas", *International journal of cancer.Journal international du cancer,*vol. 97, no. 5, pp. 668-670.

Hanahan, D. & Weinberg, R.A. 2011, "Hallmarks of cancer: the next generation", *Cell,* vol. 144, no. 5, pp. 646-674.

Hanahan, D. & Weinberg, R.A. 2000, "The hallmarks of cancer", *Cell,* vol. 100, no. 1, pp. 57-70.

Handel, M.A. & Schimenti, J.C. 2010, "Genetics of mammalian meiosis: regulation, dynamics and impact on fertility", *Nature reviews.Genetics,* vol. 11, no. 2, pp. 124-136.

Hanna, M.G.,Jr 2012, "Immunotherapy with autologous tumor cell vaccines for treatment of occult disease in early stage colon cancer", *Human vaccines & immunotherapeutics,* vol. 8, no. 8, pp. 1156-1160.

Hansen, M.A., Nielsen, J.E., Retelska, D., Larsen, N. & Leffers, H. 2008, "A shared promoter region suggests a common ancestor for the human VCX/Y, SPANX, and CSAG gene families and the murine CYPT family", *Molecular reproduction and development,* vol. 75, no. 2, pp. 219-229.

Harper, L., Golubovskaya, I. & Cande, W.Z. 2004, "A bouquet of chromosomes", *Journal of cell science,* vol. 117, no. Pt 18, pp. 4025-4032.

Hasegawa, H., Mori, M., Haraguchi, M., Ueo, H., Sugimachi, K. & Akiyoshi, T. 1998, "Expression spectrum of melanoma antigen-encoding gene family members in colorectal carcinoma", *Archives of Pathology & Laboratory Medicine,* vol. 122, no. 6, pp. 551-554.

Hasegawa, K., Ono, T., Matsushita, H., Shimono, M., Noguchi, Y., Mizutani, Y., Kodama, J., Kudo, T. & Nakayama, E. 2004, "A-kinase anchoring protein 3 messenger RNA expression in ovarian cancer and its implication on prognosis", *International journal of cancer.Journal international du cancer,* vol. 108, no. 1, pp. 86-90.

Hawthorn, L., Lan, L. & Mojica, W. 2014, "Evidence for field effect cancerization in colorectal cancer", *Genomics,* vol. 103, no. 2-3, pp. 211-221.

Hensing, T.A. & Salgia, R. 2013, "Molecular biomarkers for future screening of lung cancer", *Journal of surgical oncology,* vol. 108, no. 5, pp. 327-333.

Herreros-Villanueva, M., Chen, C.C., Yuan, S.S., Liu, T.C. & Er, T.K. 2014, "KRAS mutations: analytical considerations", *Clinica chimica acta; international journal of clinical chemistry,* vol. 431, pp. 211-220.

Hewish, M., Lord, C.J., Martin, S.A., Cunningham, D. & Ashworth, A. 2010, "Mismatch repair deficient colorectal cancer in the era of personalized treatment", *Nature reviews.Clinical oncology,* vol. 7, no. 4, pp. 197-208.

Hirota, T., Takahashi, A., Kubo, M., Tsunoda, T., Tomita, K., Sakashita, M., *et al*. 2012, "Genome-wide association study identifies eight new susceptibility loci for atopic dermatitis in the Japanese population", *Nature Genetics*, vol. 44, no. 11, pp. 1222-12226.

Hobbs, R.M., La, H.M., Makela, J.A., Kobayashi, T., Noda, T. & Pandolfi, P.P. 2015, "Distinct germline progenitor subsets defined through Tsc2-mTORC1 signaling", *EMBO reports,* vol. 16, no. 4, pp. 467-480.

Hodi, F.S., O'Day, S.J., McDermott, D.F., Weber, R.W., Sosman, J.A., Haanen, J.B., Gonzalez, R., Robert, C., Schadendorf, D., Hassel, J.C., Akerley, W., van den Eertwegh, A.J., Lutzky, J., Lorigan, P., Vaubel, J.M., Linette, G.P., Hogg, D., Ottensmeier, C.H., Lebbe, C., Peschel, C., Quirt, I., Clark, J.I., Wolchok, J.D., Weber, J.S., Tian, J., Yellin,

M.J., Nichol, G.M., Hoos, A. & Urba, W.J. 2010, "Improved survival with ipilimumab in patients with metastatic melanoma", *The New England journal of medicine,* vol. 363, no. 8, pp. 711-723.

Hofmann, M. & Ruschenburg, I. 2002, "mRNA detection of tumor-rejection genes BAGE, GAGE, and MAGE in peritoneal fluid from patients with ovarian carcinoma as a potential diagnostic tool", *Cancer,* vol. 96, no. 3, pp. 187-193.

Hofmann, O., Caballero, O.L., Stevenson, B.J., Chen, Y.T., Cohen, T., Chua, R., Maher, C.A., Panji, S., Schaefer, U., Kruger, A., Lehvaslaiho, M., Carninci, P., Hayashizaki, Y., Jongeneel, C.V., Simpson, A.J., Old, L.J. & Hide, W. 2008, "Genome-wide analysis of cancer/testis gene expression", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 105, no. 51, pp. 20422-20427.

Hohenstein, P., Pritchard-Jones, K. & Charlton, J. 2015, "The yin and yang of kidney development and Wilms' tumors", *Genes & development,* vol. 29, no. 5, pp. 467-482.

Hopkins Colon Cancer Center. 2015, "From Polyp to Cancer", John Hopkins Medicine. www.hopkinscoloncancercenter.org [accessed 19th Apr 2015].

Horton, Z., Schlatter, M. & Schultz, S. 2007, "Pediatric germ cell tumors", *Surgical oncology,* vol. 16, no. 3, pp. 205-213.

Hossack, P. 2012, "Cancer Incidence and Mortality in the United Kingdom, 2008-10". Downloaded from the Office for National Statistics website: www.ons.gov.uk [accessed 15th Apr 2015]

Hosoya, N., Okajima, M., Kinomura, A., Fujii, Y., Hiyama, T., Sun, J., Tashiro, S. & Miyagawa, K. 2011, "Synaptonemal complex protein SYCP3 impairs mitotic recombination by interfering with BRCA2", *EMBO reports,* vol. 13, no. 1, pp. 44-51.

Hsiao, Y.H., Su, Y.A., Tsai, H.D., Mason, J.T., Chou, M.C. & Man, Y.G. 2010, "Increased invasiveness and aggressiveness in breast epithelia with cytoplasmic p63 expression", *International journal of biological sciences,* vol. 6, no. 5, pp. 428-442.

Hsieh, P. & Yamane, K. 2008, "DNA mismatch repair: molecular mechanism, cancer, and ageing", *Mechanisms of ageing and development,* vol. 129, no. 7-8, pp. 391-407.

Hsu, S.H., Hsieh-Li, H.M., Huang, H.Y., Huang, P.H. & Li, H. 2005, "bHLH-zip transcription factor Spz1 mediates mitogen-activated protein kinase cell proliferation, transformation, and tumorigenesis", *Cancer research,* vol. 65, no. 10, pp. 4041-4050.

Hsu, SH., Hsieh-Li, H.M. & Li, H. 2004, "Dysfunctional spermatogenesis in transgenic mice overexpressing bHLH-Zip transcription factor, Spz1", *Experimental cell research,* vol. 294, no. 1, pp. 185-198.

Hsu, S.H., Shyu, H.W., Hsieh-Li, H.M. & Li, H. 2001, "Spz1, a novel bHLH-Zip protein, is specifically expressed in testis", *Mechanisms of development,* vol. 100, no. 2, pp. 177-187.

Hu, H., Krasinskas, A. & Willis, J. 2011, "Perspectives on current tumor-node-metastasis (TNM) staging of cancers of the colon and rectum", *Seminars in oncology,* vol. 38, no. 4, pp. 500-510.

Hunder, N.N., Wallen, H., Cao, J., Hendricks, D.W., Reilly, J.Z., Rodmyre, R., Jungbluth, A., Gnjatic, S., Thompson, J.A. & Yee, C. 2008, "Treatment of metastatic melanoma with autologous CD4+ T cells against NY-ESO-1",*The New England journal of medicine,* vol. 358, no. 25, pp. 2698-2703.

Imoto, I., Pimkhaokham, A., Watanabe, T., Saito-Ohara, F., Soeda, E. & Inazawa, J. 2000, "Amplification and overexpression of TGIF2, a novel homeobox gene of the TALE superclass, in ovarian cancer cell lines",*Biochemical and biophysical research communications,* vol. 276, no. 1, pp. 264-270.

Isella, C., Terrasi, A., Bellomo, S.E., Petti, C., Galatola, G., Muratore, A., Mellano, A., Senetta, R., Cassenti, A., Sonetto, C., Inghirami, G., Trusolino, L., Fekete, Z., De Ridder, M., Cassoni, P., Storme, G., Bertotti, A. & Medico, E. 2015, "Stromal contribution to the colorectal cancer transcriptome", *Nature genetics,* vol. 47, no. 4, pp. 312-319.

Iwamori, T., Lin, YN., Ma, L., Iwamori, N. & Matzuk, MM. 2011, "Identification and characterisation of RBM44 as a novel intercellular bridge protein", *PloS one,* vol. 6, no. 2, e17066.

Jaenisch, R. & Young, R. 2008, "Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming", *Cell,* vol. 132, no. 4, pp. 567-582.

Jamal-Hanjani, M., Quezada, S.A., Larkin, J. & Swanton, C. 2015, "Translational Implications of Tumor Heterogeneity", *Clinical cancer research : an official journal of the American Association for Cancer Research,* vol. 21, no. 6, pp. 1258-1266.

Janic, A., Mendizabal, L., Llamazares, S., Rossell, D. & Gonzalez, C. 2010, "Ectopic expression of germline genes drives malignant brain tumor growth in Drosophila", *Science (New York, N.Y.),* vol. 330, no. 6012, pp. 1824-1827.

Jansen, A., Gemayel, R. & Verstrepen, K.J. 2012, "Unstable microsatellite repeats facilitate rapid evolution of coding and regulatory sequences", *Genome dynamics,* vol. 7, pp. 108-125.

Jeffery, J., Sinha, D., Srihari, S., Kalimutho, M. & Khanna, K.K. 2015, "Beyond cytokinesis: the emerging roles of CEP55 in tumorigenesis", *Oncogene,* .

Jiang, C.L., Jin, S.G. & Pfeifer, G.P. 2004, "MBD3L1 is a transcriptional repressor that interacts with methyl-CpG-binding protein 2 (MBD2) and components of the NuRD complex", *The Journal of biological chemistry,* vol. 279, no. 50, pp. 52456-52464.

Jilg, C.A., Ketscher, A., Metzger, E., Hummel, B., Willmann, D., Russeler, V., Drendel, V., Imhof, A., Jung, M., Franz, H., Holz, S., Kronig, M., Muller, J.M. & Schule, R. 2014, "PRK1/PKN1 controls migration and metastasis of androgen-independent prostate cancer cells", *Oncotarget,* vol. 5, no. 24, pp. 12646-12664.

John, T., Starmans, M.H., Chen, Y.T., Russell, P.A., Barnett, S.A., White, S.C., Mitchell, P.L., Walkiewicz, M., Azad, A., Lambin, P., Tsao, M.S., Deb, S., Altorki, N., Wright, G., Knight, S., Boutros, P.C. & Cebon, J.S. 2013, "The role of Cancer-Testis antigens as predictive and prognostic markers in non-small cell lung cancer", *PloS one,* vol. 8, no. 7, pp. e67876.

Jungbluth, A.A., Silva, W.A.,Jr, Iversen, K., Frosina, D., Zaidi, B., Coplan, K., Eastlake-Wade, S.K., Castelli, S.B., Spagnoli, G.C., Old, L.J. & Vogel, M. 2007, "Expression of cancer-testis (CT) antigens in placenta", *Cancer immunity*, vol. 7, pp. 15.

Kalashnikova, E.V., Revenko, A.S., Gemo, A.T., Andrews, N.P., Tepper, C.G., Zou, J.X., Cardiff, R.D., Borowsky, A.D. & Chen, H.W. 2010, "ANCCA/ATAD2 overexpression identifies breast cancer patients with poor prognosis, acting to drive proliferation and survival of triple-negative cells through control of B-Myb and EZH2", *Cancer research*, vol. 70, no. 22, pp. 9402-9412.

Kamata, Y., Kuhara, A., Iwamoto, T., Hayashi, K., Koido, S., Kimura, T., Egawa, S. & Homma, S. 2013, "Identification of HLA class I-binding peptides derived from unique cancer-associated proteins by mass spectrometric analysis", *Anticancer Research,* vol. 33, no. 5, pp. 1853-1859.

Kang, G.H. 2011, "Four molecular subtypes of colorectal cancer and their precursor lesions", *Archives of Pathology & Laboratory Medicine*, vol. 135, no. 6, pp. 698-703.

Kang, KJ., Pyo, JH., Ryu, KJ., Kim, SJ., Ha, JM., Choi, K., Hong, SN., Min, BH., Chang, DK., Son, HJ., Rhee, PL., Kim, JJ., Kim, YH. 2015, "Oncogenic role of BOLL in colorectal cancer", *Digestive disease and sciences*, vol. 60, no. 6, pp. 1663-1673.

Kang, H.A., Shin, H.C., Kalantzi, A.S., Toseland, C.P., Kim, H.M., Gruber, S., Peraro, M.D. & Oh, B.H. 2015, "Crystal structure of Hop2-Mnd1 and mechanistic insights into its role in meiotic recombination", *Nucleic acids research*, vol. 43, no. 7, pp. 3841-3856.

Kapse-Mistry, S., Govender, T., Srivastava, R. & Yergeri, M. 2014, "Nanodrug delivery in reversing multidrug resistance in cancer cells", *Frontiers in pharmacology*, vol. 5, pp. 159.

Kawamura, M., Taki, T., Kaku, H., Ohki, K. & Hayashi, Y. 2015, "Identification of SPAG9 as a novel JAK2 fusion partner gene in pediatric acute lymphoblastic leukemia with t(9;17)(p24;q21)", *Genes, chromosomes & cancer*,vol. 54, no. 7, pp. 401-408.

Kayser, G., Sienel, W., Kubitz, B., Mattern, D., Stickeler, E., Passlick, B., Werner, M. & Zur Hausen, A. 2011, "Poor outcome in primary non-small cell lung cancers is predicted by transketolase TKTL1 expression", *Pathology*,vol. 43, no. 7, pp. 719-724.

Kim, K.H., Yoo, H.Y., Joo, K.M., Jung, Y., Jin, J., Kim, Y., Yoon, S.J., Choi, S.H., Seol, H.J., Park, W.Y. & Nam do, H. 2011, "Time-course analysis of DNA damage response-related genes after in vitro radiation in H460 and H1229 lung cancer cell lines", *Experimental & molecular medicine,* vol. 43, no. 7, pp. 419-426.

Kim, YH., Lee, HC., Kim, SY., Yeom, YI., Ryu, KJ., Min, BH., *et al.* 2011b, "Epigenomic analysis of aberrantly methylated genes in colorectal cancer identifies genes commonly affected by epigenetic alterations", *Annals of surgical oncology*, vol. 18, no. 8, pp. 2338-2347.

Kim, Y., Park, D., Kim, H., Choi, M., Lee, H., Lee, Y.S., Choe, J., Kim, Y.M. & Jeoung, D. 2013, "miR-200b and cancer/testis antigen CAGE form a feedback loop to regulate the invasion and tumorigenic and angiogenic responses of a cancer cell line to microtubule-targeting drugs", *The Journal of biological chemistry*, vol. 288, no. 51, pp. 36502-36518.

Kim, Y., Park, H. & Jeoung, D. 2009, "CAGE, a cancer/testis antigen, induces c-FLIP(L) and Snail to enhance cell motility and increase resistance to an anti-cancer drug", *Biotechnology Letters*, vol. 31, no. 7, pp. 945-952.

Kim, Y., Park, H., Park, D., Lee, Y.S., Choe, J., Hahn, J.H., Lee, H., Kim, Y.M. & Jeoung, D. 2010, "Cancer/testis antigen CAGE exerts negative regulation on p53 expression through HDAC2 and confers resistance to anti-cancer drugs", *The Journal of biological chemistry*, vol. 285, no. 34, pp. 25957-25968.

Kim, Y., Rosenberg, S.C., Kugel, C.L., Kostow, N., Rog, O., Davydov, V., Su, T.Y., Dernburg, A.F. & Corbett, K.D. 2014, "The chromosome axis controls meiotic events through a hierarchical assembly of HORMA domain proteins", *Developmental cell*, vol. 31, no. 4, pp. 487-502.

Klutstein, M. & Cooper, J.P. 2014, "The Chromosomal Courtship Dance-homolog pairing in early meiosis", *Current opinion in cell biology*, vol. 26, pp. 123-131.

Klutstein, M., Fennell, A., Fernandez-Alvarez, A. & Cooper, J.P. 2015, "The telomere bouquet regulates meiotic centromere assembly", *Nature cell biology*, vol. 17, no. 4, pp. 458-469.

Knudson, A.G.,Jr 1971, "Mutation and cancer: statistical study of retinoblastoma", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 68, no. 4, pp. 820-823.

Kohn, M.J., Sztein, J., Yagi, R., DePamphilis, M.L. & Kaneko, K.J. 2010, "The acrosomal protein Dickkopf-like 1 (DKKL1) facilitates sperm penetration of the zona pellucida", *Fertility and sterility*, vol. 93, no. 5, pp. 1533-1537.

Kono, K., Iinuma, H., Akutsu, Y., Tanaka, H., Hayashi, N., Uchikado, Y., Noguchi, T., Fujii, H., Okinaka, K., Fukushima, R., Matsubara, H., Ohira, M., Baba, H., Natsugoe, S., Kitano, S., Takeda, K., Yoshida, K., Tsunoda, T. & Nakamura, Y. 2012, "Multicenter, phase II clinical trial of cancer vaccination for advanced esophageal cancer with three peptides derived from novel cancer-testis antigens", *Journal of translational medicine*, vol. 10, pp. 141-5876-10-141.

Koslowski, M., Tureci, O., Bell, C., Krause, P., Lehr, H.A., Brunner, J., Seitz, G., Nestle, F.O., Huber, C. & Sahin, U. 2002, "Multiple splice variants of lactate dehydrogenase C selectively expressed in human cancer", *Cancer research*, vol. 62, no. 22, pp. 6750-6755.

Kraus, S., Nabiochtchikov, I., Shapira, S. & Arber, N. 2014, "Recent advances in personalized colorectal cancer research", *Cancer letters,* vol. 347, no. 1, pp. 15-21.

Krausz, C. & Chianese, C. 2014, "Genetic testing and counselling for male infertility", *Current opinion in endocrinology, diabetes, and obesity,* vol. 21, no. 3, pp. 244-250.

Krishnadas, D.K., Shapiro, T. & Lucas, K. 2013, "Complete remission following decitabine/dendritic cell vaccine for relapsed neuroblastoma", *Pediatrics,* vol. 131, no. 1, pp. e336-41.

Kumar, R., Ghyselinck, N., Ishiguro, K.I., Watanabe, Y., Kouznetsova, A., Hoog, C., Strong, E., Schimenti, J., Daniel, K., Toth, A. & de Massy, B. 2015, "MEI4: a central player in the regulation of meiotic DNA double strand break formation in the mouse", *Journal of cell science,* vol. 128, no. 9, pp. 1800-1811.

Kumar, V., Kota, V. & Shivaji, S. 2008, "Hamster sperm capacitation: role of pyruvate dehydrogenase A and dihydrolipoamide dehydrogenase", *Biology of reproduction,* vol. 79, no. 2, pp. 190-199.

Kuntz, S., Kieffer, E., Bianchetti, L., Lamoureux, N., Fuhrmann, G. & Viville, S. 2008, "Tex19, a mammalian-specific protein with a restricted expression in pluripotent stem cells and germ line", *Stem cells (Dayton, Ohio),*vol. 26, no. 3, pp. 734-744.

Kuo, P.L., Huang, Y.L., Hsieh, C.C., Lee, J.C., Lin, B.W. & Hung, L.Y. 2014, "STK31 is a cell-cycle regulated protein that contributes to the tumorigenicity of epithelial cancer cells", *PloS one,* vol. 9, no. 3, pp. e93303.

Ladelfa, M.F., Peche, L.Y., Toledo, M.F., Laiseca, J.E., Schneider, C. & Monte, M. 2012, "Tumor-specific MAGE proteins as regulators of p53 function", *Cancer letters,* vol. 325, no. 1, pp. 11-17.

Lam, I. & Keeney, S. 2014, "Mechanism and regulation of meiotic recombination initiation", *Cold Spring Harbor perspectives in biology,* vol. 7, no. 1, pp. a016634.

Larkin, S.E., Holmes, S., Cree, I.A., Walker, T., Basketter, V., Bickers, B., Harris, S., Garbis, S.D., Townsend, P.A. & Aukim-Hastie, C. 2012, "Identification of markers of prostate cancer progression using candidate gene expression", *British journal of cancer,* vol. 106, no. 1, pp. 157-165.

Leal, M., Sapra, P., Hurvitz, S.A., Senter, P., Wahl, A., Schutten, M., Shah, D.K., Haddish-Berhane, N. & Kabbarah, O. 2014, "Antibody-drug conjugates: an emerging modality for the treatment of cancer", *Annals of the New York Academy of Sciences,* vol. 1321, pp. 41-54.

Lee, D.W., Han, S.W., Lee, H.J., Rhee, Y.Y., Bae, J.M., Cho, N.Y., Lee, K.H., Kim, T.Y., Oh, D.Y., Im, S.A., Bang, Y.J., Jeong, S.Y., Park, K.J., Park, J.G., Kang, G.H. & Kim, T.Y. 2013, "Prognostic implication of mucinous histology in colorectal cancer patients treated with adjuvant FOLFOX chemotherapy", *British journal of cancer,* vol. 108, no. 10, pp. 1978-1984.

Leffers, N., Daemen, T., Helfrich, W., Boezen, H.M., Cohlen, B.J., Melief, C.J. & Nijman, H.W. 2014, "Antigen-specific active immunotherapy for ovarian cancer", *The Cochrane database of systematic reviews,* vol. 9, pp. CD007287.

Leslie, N.R., Yang, X., Downes, C.P. & Weijer, C.J. 2007, "PtdIns(3,4,5)P(3)-dependent and -independent roles for PTEN in the control of cell migration", *Current biology : CB,* vol. 17, no. 2, pp. 115-125.

Li, F.Q., Han, Y.L., Liu, Q., Wu, B., Huang, W.B. & Zeng, S.Y. 2009, "Overexpression of human sperm protein 17 increases migration and decreases the chemosensitivity of human epithelial ovarian cancer cells", *BMC cancer,*vol. 9, pp. 323-2407-9-323.

Li, L., Yu, C., Gao, H. & Li, Y. 2010, "Argonaute proteins: potential biomarkers for human colon cancer", *BMC cancer,* vol. 10, pp. 38-2407-10-38.

Li, M., Yuan, Y.H., Han, Y., Liu, Y.X., Yan, L., Wang, Y. & Gu, J. 2005, "Expression profile of cancer-testis genes in 121 human colorectal cancer tissue and adjacent normal tissue", *Clinical cancer research : an official journal of the American Association for Cancer Research,* vol. 11, no. 5, pp. 1809-1814.

Li, M.W., Mruk, D.D. & Cheng, C.Y. 2012, "Gap junctions and blood-tissue barriers", *Advances in Experimental Medicine and Biology,* vol. 763, pp. 260-280.

Li, X., Zhang, T.W., Tang, J.L., Fa, P.P., Lu, J.X., Qi, F.M., Cai, Z.M., Liu, C.X. & Sun, X.J. 2015, "Loss of STAG2 causes aneuploidy in normal human bladder cells", *Genetics and molecular research : GMR,* vol. 14, no. 1, pp. 2638-2646.

Lie, P.P., Cheng, C.Y. & Mruk, D.D. 2013, "Signalling pathways regulating the blood-testis barrier", *The international journal of biochemistry & cell biology,* vol. 45, no. 3, pp. 621-625.

Liggins, A.P., Lim, S.H., Soilleux, E.J., Pulford, K. & Banham, A.H. 2010, "A panel of cancer-testis genes exhibiting broad-spectrum expression in haematological malignancies", *Cancer immunity,* vol. 10, pp. 8.

Lim, S.H., Austin, S., Owen-Jones, E. & Robinson, L. 1999, "Expression of testicular genes in haematological malignancies", *British journal of cancer,* vol. 81, no. 7, pp. 1162-1164.

Lips, E.H., van Eijk, R., de Graaf, E.J., Oosting, J., de Miranda, N.F., Karsten, T., van de Velde, C.J., Eilers, P.H., Tollenaar, R.A., van Wezel, T. & Morreau, H. 2008, "Integrating chromosomal aberrations and gene expression profiles to dissect rectal tumorigenesis", *BMC cancer,* vol. 8, pp. 314-2407-8-314.

Litvinov, I.V., Cordeiro, B., Huang, Y., Zargham, H., Pehr, K., Dore, M.A., Gilbert, M., Zhou, Y., Kupper, T.S. & Sasseville, D. 2014, "Ectopic expression of cancer-testis antigens in cutaneous T-cell lymphoma patients", *Clinical cancer research : an official journal of the American Association for Cancer Research,* vol. 20, no. 14, pp. 3799-3808.

Litvinov, I.V., Netchiporouk, E., Cordeiro, B., Dore, M.A., Moreau, L., Pehr, K., Gilbert, M., Zhou, Y., Sasseville, D. & Kupper, T.S. 2015, "The Use of Transcriptional Profiling to Improve Personalized Diagnosis and Management of Cutaneous T-cell Lymphoma (CTCL)", *Clinical cancer research : an official journal of the American Association for Cancer Research,* .

Litwin, M., Dubis, J., Arczynska, K., Piotrowska, A., Frydlewicz, A., Karczewski, M., Dziegiel, P. & Witkiewicz, W. 2015, "Correlation of HIWI and HILI Expression with Cancer Stem Cell Markers in Colorectal Cancer", *Anticancer Research,* vol. 35, no. 6, pp. 3317-3324.

Liu, F., Zhang, H., Shen, D., Wang, S., Ye, Y., Chen, H., Pang, X., Song, Q. & He, P. 2014, "Identification of two new HLA-A*0201-restricted cytotoxic T lymphocyte epitopes from colorectal carcinoma-associated antigen PLAC1/CP1", *Journal of gastroenterology,* vol. 49, no. 3, pp. 419-426.

Liu, L., Zhang, J., Yuan, J., Dang, Y., Yang, C., Chen, X., Xu, J. & Yu, L. 2005, "Characterization of a human regulatory subunit of protein phosphatase 3 gene (PPP3RL) expressed specifically in testis", *Molecular biology reports,* vol. 32, no. 1, pp. 41-45.

Liu, L., Zhang, J., Yuan, J., Dang, Y., Yang, C., Chen, X., Xu, J. & Yu, L. 2005, “Characterisation of a human regulatory subunit of protein phosphatase 3 gene (PPP3RL) expressed specifically in testis”, *Molecular biology reports*, vol. 32, no. 1, pp. 41-45.

Liu, M., Chen, J., Hu, L., Shi, X., Zhou, Z., Hu, Z. & Sha, J. 2012, "HORMAD2/CT46.2, a novel cancer/testis gene, is ectopically expressed in lung cancer tissues", *Molecular human reproduction,* vol. 18, no. 12, pp. 599-604.

Liu, M., Shi, X., Bi, Y., Qi, L., Guo, X., Wang, L., Zhou, Z. & Sha, J. 2014, "SHCBP1L, a conserved protein in mammals, is predominantly expressed in male germ cells and maintains spindle stability during meiosis in testis",*Molecular human reproduction,* vol. 20, no. 6, pp. 463-475.

Liu, W., Zhai, M., Wu, Z., Qi, Y., Wu, Y., Dai, C., Sun, M., Li, L. & Gao, Y. 2012, "Identification of a novel HLA-A2-restricted cytotoxic T lymphocyte epitope from cancer-testis antigen PLAC1 in breast cancer", *Amino acids,* vol. 42, no. 6, pp. 2257-2265.

Lochhead, P., Chan, A.T., Nishihara, R., Fuchs, C.S., Beck, A.H., Giovannucci, E. & Ogino, S. 2015, "Etiologic field effect: reappraisal of the field effect concept in cancer predisposition and progression", *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc,* vol. 28, no. 1, pp. 14-29.

Longerich, S., Li, J., Xiong, Y., Sung, P. & Kupfer, G.M. 2014, "Stress and DNA repair biology of the Fanconi anemia pathway", *Blood,* vol. 124, no. 18, pp. 2812-2819.

Longhese, M.P., Bonetti, D., Guerini, I., Manfrini, N. & Clerici, M. 2009, "DNA double-strand breaks in meiosis: checking their formation, processing and repair", *DNA repair,* vol. 8, no. 9, pp. 1127-1138.

Lord, C.J. & Ashworth, A. 2012, "The DNA damage response and cancer therapy", *Nature,* vol. 481, no. 7381, pp. 287-294.

Ludwig, J.A. & Weinstein, J.N. 2005, "Biomarkers in cancer staging, prognosis and treatment selection", *Nature reviews.Cancer,* vol. 5, no. 11, pp. 845-856.

Lui, W.Y. & Cheng, C.Y. 2012, "Transcriptional regulation of cell adhesion at the blood-testis barrier and spermatogenesis in the testis", *Advances in Experimental Medicine and Biology,* vol. 763, pp. 281-294.

Lujambio, A., Calin, G.A., Villanueva, A., Ropero, S., Sanchez-Cespedes, M., Blanco, D., Montuenga, L.M., Rossi, S., Nicoloso, M.S., Faller, W.J., Gallagher, W.M., Eccles, S.A., Croce, C.M. & Esteller, M. 2008, "A microRNA DNA methylation signature for human cancer metastasis", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 105, no. 36, pp. 13556-13561.

Luke, J.J. & Hodi, F.S. 2013, "Ipilimumab, vemurafenib, dabrafenib, and trametinib: synergistic competitors in the clinical management of BRAF mutant malignant melanoma", *The oncologist,* vol. 18, no. 6, pp. 717-725.

Luo, B., Yun, X., Fan, R., Lin, Y.D., He, S.J., Zhang, Q.M., Mo, F.R., Chen, F., Xiao, S.W. & Xie, X.X. 2013, "Cancer testis antigen OY-TES-1 expression and serum immunogenicity in colorectal cancer: its relationship to clinicopathological parameters", *International journal of clinical and experimental pathology,* vol. 6, no. 12, pp. 2835-2845.

Lurquin, C., De Smet, C., Brasseur, F., Muscatelli, F., Martelange, V., De Plaen, E., Brasseur, R., Monaco, A.P. & Boon, T. 1997, "Two members of the human MAGEB gene family located in Xp21.3 are expressed in tumors of various histological origins", *Genomics,* vol. 46, no. 3, pp. 397-408.

Mahanta, S., Paul, S., Srivastava, A., Pastor, A., Kundu, B. & Chaudhuri, T.K. 2015, "Stable self-assembled nanostructured hen egg white lysozyme exhibits strong anti-proliferative activity against breast cancer cells",*Colloids and surfaces.B, Biointerfaces,* .

Malumbres, M. & Barbacid, M. 2009, "Cell cycle, CDKs and cancer: a changing paradigm", *Nature reviews.Cancer,* vol. 9, no. 3, pp. 153-166.

Malumbres, M. & Barbacid, M. 2003, "RAS oncogenes: the first 30 years", *Nature reviews.Cancer,* vol. 3, no. 6, pp. 459-465.

Mantovani, A., Allavena, P., Sica, A. & Balkwill, F. 2008, "Cancer-related inflammation", *Nature,* vol. 454, no. 7203, pp. 436-444.

Mao, Z.P., Zhao, L.J., Zhou, S.H., Liu, M.Q., Tan, W.F. & Yao, H.T. 2015, "Expression and significance of glucose transporter-1, P-glycoprotein, multidrug resistance-associated protein and glutathione S-transferase-pi in laryngeal carcinoma", *Oncology letters,* vol. 9, no. 2, pp. 806-810.

Marisa, L., de Reynies, A., Duval, A., Selves, J., Gaub, M.P., Vescovo, L., Etienne-Grimaldi, M.C., Schiappa, R., Guenot, D., Ayadi, M., Kirzin, S., Chazal, M., Flejou, J.F.,

Benchimol, D., Berger, A., Lagarde, A., Pencreach, E., Piard, F., Elias, D., Parc, Y., Olschwang, S., Milano, G., Laurent-Puig, P. & Boige, V. 2013, "Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value", *PLoS medicine*, vol. 10, no. 5, pp. e1001453.

Markman, J.L. & Shiao, S.L. 2015, "Impact of the immune system and immunotherapy in colorectal cancer", *Journal of gastrointestinal oncology*, vol. 6, no. 2, pp. 208-223.

Martin, L.J. & Tremblay, J.J. 2010, "Nuclear receptors in Leydig cell gene expression and function", *Biology of reproduction*, vol. 83, no. 1, pp. 3-14.

Mashino, K., Sadanaga, N., Tanaka, F., Yamaguchi, H., Nagaschima, H., Inoue, H., Sugimachi, K. & Mori, M. 2001, "Expression of multiple cancer-testis antigen genes in gastrointestinal and breast carcinomas", *British journal of cancer*, vol. 85, no. 5, pp. 713-720.

Mathers, C.D., Boerma, T. & Ma Fat, D. 2009, "Global and regional causes of death", *British medical bulletin*, vol. 92, pp. 7-32.

Mathers, C.D. & Loncar, D. 2006, "Projections of global mortality and burden of disease from 2002 to 2030", *PLoS medicine*, vol. 3, no. 11, pp. e442.

Matsa, E., Burridge, P.W. & Wu, J.C. 2014, "Human stem cells for modeling heart disease and for drug discovery", *Science translational medicine*, vol. 6, no. 239, pp. 239ps6.

McFarlane, R.J., Feichtinger, J. & Larcombe, L. 2015, "Germline/meiotic genes in cancer: new dimensions", *Cell cycle (Georgetown, Tex.)*, vol. 14, no. 6, pp. 791-792.

McFarlane, R.J., Feichtinger, J. & Larcombe, L. 2014, "Cancer germline gene activation: friend or foe?", *Cell cycle (Georgetown, Tex.)*, vol. 13, no. 14, pp. 2151-2152.

McKim, K.S. & Hayashi-Hagihara, A. 1998, "mei-W68 in Drosophila melanogaster encodes a Spo11 homolog: evidence that the mechanism for initiating meiotic recombination is conserved", *Genes & development*, vol. 12, no. 18, pp. 2932-2942.

Meccariello, R., Chianese, R., Chioccarelli, T., Ciaramella, V., Fasano, S., Pierantoni, R. & Cobellis, G. 2014, "Intra-testicular signals regulate germ cell progression and production of qualitatively mature spermatozoa in vertebrates", *Frontiers in endocrinology*, vol. 5, pp. 69.

Meek, D.W. & Marcar, L. 2012, "MAGE-A antigens as targets in tumour therapy", *Cancer letters*, vol. 324, no. 2, pp. 126-132.

Mehta, A., Dobersch, S., Dammann, R.H., Bellusci, S., Ilinskaya, O.N., Braun, T. & Barreto, G. 2015, "Validation of Tuba1a as appropriate internal control for normalization of gene expression analysis during mouse lung development", *International journal of molecular sciences*, vol. 16, no. 3, pp. 4492-4511.

Mellman, I., Coukos, G. & Dranoff, G. 2011, "Cancer immunotherapy comes of age", *Nature*, vol. 480, no. 7378, pp. 480-489.

Mendelsohn, J. 2013, "Personalizing oncology: perspectives and prospects", *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 31, no. 15, pp. 1904-1911.

Menon, U., Griffin, M. & Gentry-Maharaj, A. 2014, "Ovarian cancer screening--current status, future directions", *Gynecologic oncology*, vol. 132, no. 2, pp. 490-495.

Micci, F., Haugom, L., Abeler, V.M., Davidson, B., Trope, C.G. & Heim, S. 2014, "Genomic profile of ovarian carcinomas", *BMC cancer*, vol. 14, pp. 315-2407-14-315.

Mikelsaar, R., Nelis, M., Kurg, A., Zilina, O., Korrovits, P., Ratsep, R. & Vali, M. 2012, "Balanced reciprocal translocation t(5;13)(q33;q12) and 9q31.1 microduplication in a man suffering from infertility and pollinosis", *Journal of Applied Genetics*, vol. 53, no. 1, pp. 93-97.

Milde-Langosch, K. 2005, "The Fos family of transcription factors and their role in tumourigenesis", *European journal of cancer (Oxford, England : 1990)*, vol. 41, no. 16, pp. 2449-2461.

Milella, M., Falcone, I., Conciatori, F., Cesta Incani, U., Del Curatolo, A., Inzerilli, N., Nuzzo, C.M., Vaccaro, V., Vari, S., Cognetti, F. & Ciuffreda, L. 2015, "PTEN: Multiple Functions in Human Malignant Tumors", *Frontiers in oncology*, vol. 5, pp. 24.

Minges, J.T., Su, S., Grossman, G., Blackwelder, A.J., Pop, E.A., Mohler, J.L. & Wilson, E.M. 2013, "Melanoma antigen-A11 (MAGE-A11) enhances transcriptional activity by linking androgen receptor dimers", *The Journal of biological chemistry*, vol. 288, no. 3, pp. 1939-1952.

Mirandola, L., J Cannon, M., Cobos, E., Bernardini, G., Jenkins, M.R., Kast, W.M. & Chiriva-Internati, M. 2011, "Cancer testis antigens: novel biomarkers and targetable proteins for ovarian cancer", *International reviews of immunology*, vol. 30, no. 2-3, pp. 127-137.

Mital, P., Hinton, B.T. & Dufour, J.M. 2011, "The blood-testis and blood-epididymis barriers are more than just their tight junctions", *Biology of reproduction*, vol. 84, no. 5, pp. 851-858.

Miyata, H., Satouh, Y., Mashiko, D., Muto, M., Nozawa, K., Shiba, K., Fujihara, Y., Isotani, A., Inaba, K. & Ikawa, M. 2015, "Sperm calcineurin inhibition prevents mouse fertility with implications for male contraceptive", *Science*, vol. 350, no. 6259, pp. 442-445.

Mizukami, Y., Kono, K., Daigo, Y., Takano, A., Tsunoda, T., Kawaguchi, Y., Nakamura, Y. & Fujii, H. 2008, "Detection of novel cancer-testis antigen-specific T-cell responses in TIL, regional lymph nodes, and PBL in patients with esophageal squamous cell carcinoma", *Cancer science*, vol. 99, no. 7, pp. 1448-1454.

Moasser, M.M. 2007, "The oncogene HER2: its signaling and transforming functions and its role in human cancer pathogenesis", *Oncogene*, vol. 26, no. 45, pp. 6469-6487.

Moro, N., Mauch, C. & Zigrino, P. 2014, "Metalloproteinases in melanoma", *European journal of cell biology,* vol. 93, no. 1-2, pp. 23-29.

Mruk, D.D., Su, L. & Cheng, C.Y. 2011, "Emerging role for drug transporters at the blood-testis barrier", *Trends in pharmacological sciences,* vol. 32, no. 2, pp. 99-106.

Murakami, H. & Keeney, S. 2008, "Regulating the formation of DNA double-strand breaks in meiosis", *Genes & development,* vol. 22, no. 3, pp. 286-292.

Nagaraj, S.H., Gasser, R.B. & Ranganathan, S. 2007, "A hitchhiker's guide to expressed sequence tag (EST) analysis", *Briefings in bioinformatics,* vol. 8, no. 1, pp. 6-21.

Negrini, S., Gorgoulis, V.G. & Halazonetis, T.D. 2010, "Genomic instability--an evolving hallmark of cancer", *Nature reviews.Molecular cell biology,* vol. 11, no. 3, pp. 220-228.

Neller, M.A., Lopez, J.A. & Schmidt, C.W. 2008, "Antigens for cancer immunotherapy", *Seminars in immunology,* vol. 20, no. 5, pp. 286-295.

Nelson, B.H. 2015, "New insights into tumor immunity revealed by the unique genetic and genomic aspects of ovarian cancer", *Current opinion in immunology,* vol. 33C, pp. 93-100.

Ng, J.M. & Curran, T. 2011, "The Hedgehog's tale: developing strategies for targeting cancer", *Nature reviews.Cancer,* vol. 11, no. 7, pp. 493-501.

Nirschl, C.J. & Drake, C.G. 2013, "Molecular pathways: coexpression of immune checkpoint molecules: signaling pathways and implications for cancer immunotherapy", *Clinical cancer research : an official journal of the American Association for Cancer Research,* vol. 19, no. 18, pp. 4917-4924.

Nishito, Y., Hasegawa, M., Inohara, N. & Nunez, G. 2006, "MEX is a testis-specific E3 ubiquitin ligase that promotes death receptor-induced apoptosis", *The biochemical journal,* vol. 396, no. 3, pp. 411-477.

Noguchi, T., Ritter, G. & Nishikawa, H. 2013, "Antibody-based therapy in colorectal cancer", *Immunotherapy,* vol. 5, no. 5, pp. 533-545.

Norsett, K.G., Steele, I., Duval, C., Sammut, S.J., Murugesan, S.V., Kenny, S., Rainbow, L., Dimaline, R., Dockray, G.J., Pritchard, D.M. & Varro, A. 2011, "Gastrin stimulates expression of plasminogen activator inhibitor-1 in gastric epithelial cells", *American journal of physiology.Gastrointestinal and liver physiology,* vol. 301, no. 3, pp. G446-53.

Nunes, C.T., Miners, K.L., Dolton, G., Pepper, C., Fegan, C., Mason, M.D. & Man, S. 2011, "A novel tumor antigen derived from enhanced degradation of bax protein in human cancers", *Cancer research,* vol. 71, no. 16, pp. 5435-5444.

Ollinger, R., Childs, A.J., Burgess, H.M., Speed, R.M., Lundegaard, P.R., Reynolds, N., Gray, N.K., Cooke, H.J. & Adams, I.R. 2008, "Deletion of the pluripotency-associated Tex19.1 gene causes activation of endogenous retroviruses and defective spermatogenesis in mice", *PLoS genetics,* vol. 4, no. 9, pp. e1000199.

O'Sullivan, R.J. & Karlseder, J. 2010, "Telomeres: protecting chromosomes against genome instability", *Nature reviews.Molecular cell biology,* vol. 11, no. 3, pp. 171-181.

Ott, P.A., Hamid, O., Pavlick, A.C., Kluger, H., Kim, K.B., Boasberg, P.D., Simantov, R., Crowley, E., Green, J.A., Hawthorne, T., Davis, T.A., Sznol, M. & Hwu, P. 2014, "Phase I/II study of the antibody-drug conjugate glembatumumab vedotin in patients with advanced melanoma", *Journal of clinical oncology : official journal of the American Society of Clinical Oncology,* vol. 32, no. 32, pp. 3659-3666.

Ousati Ashtiani, Z., Ayati, M., Modarresi, M.H., Raoofian, R., Sabah Goulian, B., Greene, W.K. & Heidari, M. 2009, "Association of TGIFLX/Y mRNA expression with prostate cancer", *Medical oncology (Northwood, London, England),* vol. 26, no. 1, pp. 73-77.

Park, S.J., Huh, J.W., Kim, Y.H., Lee, S.R., Kim, S.H., Kim, S.U., Kim, H.S., Kim, M.K. & Chang, K.T. 2013, "Selection of internal reference genes for normalization of quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis in the canine brain and other organs", *Molecular biotechnology,* vol. 54, no. 1, pp. 47-57.

Park, S. & Lippard, SJ. 2012, "Binding interaction of HMGB4 with cisplatin-modified DNA", *Biochemistry*, vol. 51, no. 34, pp. 6728-6737.

Pasquale, R., Fenizia, F., Abate, R.E., Sacco, A., Esposito, C., Forgione, L., Rachiglio, A.M., Bevilacqua, S., Montanino, A., Franco, R., Rocco, G., Botti, G., Denis, M.G., Morabito, A., Luca, A. & Normanno, N. 2015, "Assessment of high-sensitive methods for the detection of EGFR mutations in circulating free tumor DNA from NSCLC patients", *Pharmacogenomics,* , pp. 1-13.

Pattabiraman, D.R. & Weinberg, R.A. 2014, "Tackling the cancer stem cells—what challenges do they pose?", *Nature Reviews Drug Discovery,* vol. 13, no. 7, pp. 497-512.

Pavlou, M.P., Diamandis, E.P. & Blasutig, I.M. 2013, "The long journey of cancer biomarkers from the bench to the clinic", *Clinical chemistry,* vol. 59, no. 1, pp. 147-157.

Perez, D., Hauswirth, F., Jager, D., Metzger, U., Samartzis, E.P., Went, P. & Jungbluth, A. 2011, "Protein expression of cancer testis antigens predicts tumor recurrence and treatment response to imatinib in gastrointestinal stromal tumors", *International journal of cancer.Journal international du cancer,* vol. 128, no. 12, pp. 2947-2952.

Peters, B., Tong, W., Sidney, J., Sette, A. & Weng, Z. 2003, "Examining the independent binding assumption for binding of peptide epitopes to MHC-I molecules", *Bioinformatics (Oxford, England),* vol. 19, no. 14, pp. 1765-1772.

Pinheiro, A., Nunes, MJ., Milagre, I., Rodrigues, E., Silva, MJ., de Almeida, IT. & Rivera, I. 2012, "Demethylation of the coding region triggers the activation of the human testis-specific PDHA2 gene in somatic tissues", *PloS one*, vol. 7, no. 6, e38076.

Plaks, V., Kong, N. & Werb, Z. 2015, "The Cancer Stem Cell Niche: How Essential Is the Niche in Regulating Stemness of Tumor Cells?", *Cell stem cell,* vol. 16, no. 3, pp. 225-238.

Pollheimer, J., Fock, V. & Knofler, M. 2014, "Review: the ADAM metalloproteinases - novel regulators of trophoblast invasion?", *Placenta,* vol. 35 Suppl, pp. S57-63.

Por, E., Byun, H.J., Lee, E.J., Lim, J.H., Jung, S.Y., Park, I., Kim, Y.M., Jeoung, D.I. & Lee, H. 2010, "The cancer/testis antigen CAGE with oncogenic potential stimulates cell proliferation by up-regulating cyclins D1 and E in an AP-1- and E2F-dependent manner", *The Journal of biological chemistry,* vol. 285, no. 19, pp. 14475-14485.

Poulogiannis, G., Frayling, I.M. & Arends, M.J. 2010, "DNA mismatch repair deficiency in sporadic colorectal cancer and Lynch syndrome", *Histopathology,* vol. 56, no. 2, pp. 167-179.

Przemyslaw, L., Boguslaw, H.A., Elzbieta, S. & Malgorzata, S.M. 2013, "ADAM and ADAMTS family proteins and their role in the colorectal cancer etiopathogenesis", *BMB reports,* vol. 46, no. 3, pp. 139-150.

Pui, C.H., Carroll, W.L., Meshinchi, S. & Arceci, R.J. 2011, "Biology, risk stratification, and therapy of pediatric acute leukemias: an update", *Journal of clinical oncology : official journal of the American Society of Clinical Oncology,* vol. 29, no. 5, pp. 551-565.

Radtke, F. & Raj, K. 2003, "The role of Notch in tumorigenesis: oncogene or tumour suppressor?", *Nature reviews.Cancer,* vol. 3, no. 10, pp. 756-767.

Rammensee, H., Bachmann, J., Emmerich, N.P., Bachor, O.A. & Stevanovic, S. 1999, "SYFPEITHI: database for MHC ligands and peptide motifs", *Immunogenetics,* vol. 50, no. 3-4, pp. 213-219.

Rao, C.V. & Yamada, H.Y. 2013, "Genomic instability and colon carcinogenesis: from the perspective of genes", *Frontiers in oncology,* vol. 3, pp. 130.

Rauch, T., Li, H., Wu, X. & Pfeifer, G.P. 2006, "MIRA-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells", *Cancer research,* vol. 66, no. 16, pp. 7939-7947.

Romanuik, T.L., Ueda, T., Le, N., Haile, S., Yong, T.M., Thomson, T., Vessella, R.L. & Sadar, M.D. 2009, "Novel biomarkers for prostate cancer including noncoding transcripts", *The American journal of pathology,* vol. 175, no. 6, pp. 2264-2276.

Rosenberg, S.A. & Restifo, N.P. 2015, "Adoptive cell transfer as personalized immunotherapy for human cancer", *Science (New York, N.Y.),* vol. 348, no. 6230, pp. 62-68.

Rossi, P. & Dolci, S. 2013, "Paracrine mechanisms involved in the control of early stages of Mammalian spermatogenesis", *Frontiers in endocrinology,* vol. 4, pp. 181.

Rousseaux, S., Debernardi, A., Jacquiau, B., Vitte, A.L., Vesin, A., Nagy-Mignotte, H., Moro-Sibilot, D., Brichon, P.Y., Lantuejoul, S., Hainaut, P., Laffaire, J., de Reynies, A., Beer, D.G., Timsit, J.F., Brambilla, C., Brambilla, E. & Khochbin, S. 2013b, "Ectopic

activation of germline and placental genes identifies aggressive metastasis-prone lung cancers", *Science translational medicine,* vol. 5, no. 186, pp. 186ra66.

Rousseaux, S., Wang, J. & Khochbin, S. 2013a, "Cancer hallmarks sustained by ectopic activations of placenta/male germline genes", *Cell cycle (Georgetown, Tex.),* vol. 12, no. 15, pp. 2331-2332.

Ryan, K.M. 2011, "P53 and Autophagy in Cancer: Guardian of the Genome Meets Guardian of the Proteome", *European journal of cancer (Oxford, England : 1990),* vol. 47, no. 1, pp. 44-50.

Sadanandam, A., Lyssiotis, C.A., Homicsko, K., Collisson, E.A., Gibb, W.J., Wullschleger, S., Ostos, L.C., Lannon, W.A., Grotzinger, C., Del Rio, M., Lhermitte, B., Olshen, A.B., Wiedenmann, B., Cantley, L.C., Gray, J.W. & Hanahan, D. 2013, "A colorectal cancer classification system that associates cellular phenotype and responses to therapy", *Nature medicine,* vol. 19, no. 5, pp. 619-625.

Sadanandam, A., Wang, X., de Sousa E Melo, F., Gray, J.W., Vermeulen, L., Hanahan, D. & Medema, J.P. 2014, "Reconciliation of classification systems defining molecular subtypes of colorectal cancer: interrelationships and clinical implications", *Cell cycle (Georgetown, Tex.),* vol. 13, no. 3, pp. 353-357.

Sakai, K., Tsurutani, J., Yamanaka, T., Yoneshige, A., Ito, A., Togashi, Y., De Velasco, M.A., Terashima, M., Fujita, Y., Tomida, S., Tamura, T., Nakagawa, K. & Nishio, K. 2015, "Extended RAS and BRAF Mutation Analysis Using Next-Generation Sequencing", *PloS one,* vol. 10, no. 5, pp. e0121891.

Saltsman, K. 2005, "Cellular Reproduction: Multiplication by Division", *Inside the Cell*, National Insitute of General Medical Sciences: www.nigms.nih.gov [website/publication accessed 19th August 2015].

Sammut, J., Wakeman, J., Stuart, N. & McFarlane, R. 2013, "Cancer/Testis Antigens and Colorectal Cancer", *J Genet Syndr Gene Ther,* vol. 4, no. 149, pp. 2.

Sammut, S.J., Feichtinger, J., Stuart, N., Wakeman, J.A., Larcombe, L. & McFarlane, R.J. 2014, "A novel cohort of cancer-testis biomarker genes revealed through meta-analysis of clinical data sets", *Oncoscience,* vol. 1, no. 5, pp. 349-359.

San Filippo, J., Sung, P. & Klein, H. 2008, "Mechanism of eukaryotic homologous recombination", *Annual Review of Biochemistry,* vol. 77, pp. 229-257.

Sartor, I.T., Zeidan-Chulia, F., Albanus, R.D., Dalmolin, R.J. & Moreira, J.C. 2014, "Computational analyses reveal a prognostic impact of TULP3 as a transcriptional master regulator in pancreatic ductal adenocarcinoma",*Molecular bioSystems,* vol. 10, no. 6, pp. 1461-1468.

Sato, E., Olson, S.H., Ahn, J., Bundy, B., Nishikawa, H., Qian, F., Jungbluth, A.A., Frosina, D., Gnjatic, S., Ambrosone, C., Kepner, J., Odunsi, T., Ritter, G., Lele, S., Chen, Y.T., Ohtani, H., Old, L.J. & Odunsi, K. 2005, "Intraepithelial CD8+ tumor-infiltrating lymphocytes and a high CD8+/regulatory T cell ratio are associated with favorable

prognosis in ovarian cancer", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 51, pp. 18538-18543.

Scanlan, M.J., Gure, A.O., Jungbluth, A.A., Old, L.J. & Chen, Y.T. 2002, "Cancer/testis antigens: an expanding family of targets for cancer immunotherapy", *Immunological reviews*, vol. 188, pp. 22-32.

Scanlan, M.J., Simpson, A.J. & Old, L.J. 2004, "The cancer/testis genes: review, standardization, and commentary", *Cancer immunity*, vol. 4, pp. 1.

Schmitz, J., Schwab, J., Schwenck, J., Chen, Q., Quintanilla-Martinez, L., Hahn, M., Schwenzer, N., Staebler, A., Kohlhofer, U., Aina, O.H., Hubbard, N.E., Reischl, G., Borowsky, A.D., Brucker, S., Nikolaou, K., la Fougere, C., Cardiff, R.D., Pichler, B.J. & Schmid, A.M. 2016, "Decoding intratumoral heterogeneity of breast cancer by multiparametric in vivo imaging: A translational study. *Cancer Research*, vol. 76, no. 18, pp. 5512-5522.

Schumacher, T.N. & Schreiber, R.D. 2015, "Neoantigens in cancer immunotherapy", *Science (New York, N.Y.)*, vol. 348, no. 6230, pp. 69-74.

Schvarzstein, M., Wignall, S.M. & Villeneuve, A.M. 2010, "Coordinating cohesion, co-orientation, and congression during meiosis: lessons from holocentric chromosomes", *Genes & development*, vol. 24, no. 3, pp. 219-228.

Scott, S. & de la Motte, C. 2015. glycoforum.gr.jp/science/hyaluronan/HA12/HA12E.html [accessed 2nd July 2015].

Seoane, J. & De Mattos-Arruda, L. 2014, "The challenge of intratumour heterogeneity in precision medicine", *Journal of internal medicine*, vol. 276, no. 1, pp. 41-51.

Serra, C., Vizoso, F., Alonso, L., Rodriguez, J.C., Gonzalez, L.O., Fernandez, M., Lamelas, M.L., Sanchez, L.M., Garcia-Muniz, J.L., Baltasar, A. & Medrano, J. 2002, "Expression and prognostic significance of lysozyme in male breast cancer", *Breast cancer research : BCR*, vol. 4, no. 6, pp. R16.

Sha, J.H., Zhou, Z.M., Li, J.M., Lin, M., Zhu, H., Zhu, H., Zhou, Y.D., Wang, L.L., Wang, Y.Q. & Zhou, K.Y. 2003, "Expression of a novel bHLH-Zip gene in human testis", *Asian Journal of Andrology*, vol. 5, no. 2, pp. 83-88.

Shah, S.N., Hile, S.E. & Eckert, K.A. 2010, "Defective mismatch repair, microsatellite mutation bias, and variability in clinical cancer phenotypes", *Cancer research*, vol. 70, no. 2, pp. 431-435.

Shantha Kumara, H.M., Grieco, M.J., Caballero, O.L., Su, T., Ahmed, A., Ritter, E., Gnjatic, S., Cekic, V., Old, L.J., Simpson, A.J., Cordon-Cardo, C. & Whelan, R.L. 2012, "MAGE-A3 is highly expressed in a subset of colorectal cancer patients", *Cancer immunity*, vol. 12, pp. 16.

Sharma, S., Qian, F., Keitz, B., Driscoll, D., Scanlan, M.J., Skipper, J., Rodabaugh, K., Lele, S., Old, L.J. & Odunsi, K. 2005, "A-kinase anchoring protein 3 messenger RNA

expression correlates with poor prognosis in epithelial ovarian cancer", *Gynecologic oncology,* vol. 99, no. 1, pp. 183-188.

Sharom, F.J. 2011, "The P-glycoprotein multidrug transporter", *Essays in biochemistry,* vol. 50, no. 1, pp. 161-178.

Shen, L., Kondo, Y., Guo, Y., Zhang, J., Zhang, L., Ahmed, S., Shu, J., Chen, X., Waterland, R.A. & Issa, J.P. 2007, "Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters",*PLoS genetics,* vol. 3, no. 10, pp. 2023-2036.

Shin, O.H. 2014, "Exocytosis and synaptic vesicle function", *Comprehensive Physiology,* vol. 4, no. 1, pp. 149-175.

Shin, Y.H., Choi, Y., Erdin, S.U., Yatsenko, S.A., Kloc, M., Yang, F., Wang, P.J., Meistrich, M.L. & Rajkovic, A. 2010, "Hormad1 mutation disrupts synaptonemal complex formation, recombination, and chromosome segregation in mammalian meiosis", *PLoS genetics,* vol. 6, no. 11, pp. e1001190.

Shiraishi, T., Terada, N., Zeng, Y., Suyama, T., Luo, J., Trock, B., Kulkarni, P. & Getzenberg, R.H. 2011, "Cancer/Testis Antigens as potential predictors of biochemical recurrence of prostate cancer following radical prostatectomy", *Journal of translational medicine,* vol. 9, pp. 153-5876-9-153.

Sibbe, M. & Jarowyj, J. 2013, "Region-specific expression of Dickkopf-like1 in the adult brain. Abbreviated title: Dkkl1 in the adult brain", *Neuroscience letters,* vol. 535, pp. 84-89.

Siddiqi, S. & Matushansky, I. 2012, "Piwis and piwi-interacting RNAs in the epigenetics of cancer", *Journal of cellular biochemistry,* vol. 113, no. 2, pp. 373-380.

Siegel, R.L., Miller, K.D. & Jemal, A. 2015, "Cancer statistics, 2015", *CA: a cancer journal for clinicians,* vol. 65, no. 1, pp. 5-29.

Silver, D.P. & Livingston, D.M. 2012, "Mechanisms of BRCA1 tumor suppression", *Cancer discovery,* vol. 2, no. 8, pp. 679-684.

Simmons, A.R., Baggerly, K. & Bast, R.C.,Jr 2013, "The emerging role of HE4 in the evaluation of epithelial ovarian and endometrial carcinomas", *Oncology (Williston Park, N.Y.),* vol. 27, no. 6, pp. 548-556.

Simo-Riudalbas, L., Diaz-Lagares, A., Gatto, S., Gagliardi, M., Crujeiras, AB., Matarazzo, MR., Esteller, M. & Sandoval, J. 2015, "Genome-wide DNA methylation analysis identifies novel hypomethylated non-pericentromeric genes with potential clinical implications in ICF syndrome", *PloS one,* vol. 10, no. 7, e0132517.

Simon, P., Omokoko, T.A., Breitkreuz, A., Hebich, L., Kreiter, S., Attig, S., Konur, A., Britten, C.M., Paret, C., Dhaene, K., Tureci, O. & Sahin, U. 2014, "Functional TCR retrieval from single antigen-specific human T cells reveals multiple novel epitopes", *Cancer immunology research,* vol. 2, no. 12, pp. 1230-1244.

Simpson, A.J., Caballero, O.L., Jungbluth, A., Chen, Y.T. & Old, L.J. 2005, "Cancer/testis antigens, gametogenesis and cancer", *Nature reviews.Cancer*, vol. 5, no. 8, pp. 615-625.

Siomi, M.C., Sato, K., Pezic, D. & Aravin, A.A. 2011, "PIWI-interacting small RNAs: the vanguard of genome defence", *Nature reviews.Molecular cell biology*, vol. 12, no. 4, pp. 246-258.

Slaughter, D.P., Southwick, H.W. & Smejkal, W. 1953, "Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin", *Cancer*, vol. 6, no. 5, pp. 963-968.

Smith, H.W. & Marshall, C.J. 2010, "Regulation of cell signalling by uPAR", *Nature reviews.Molecular cell biology*, vol. 11, no. 1, pp. 23-36.

Smith, L.B. & Walker, W.H. 2014, "The regulation of spermatogenesis by androgens", *Seminars in cell & developmental biology*, vol. 30, pp. 2-13.

Smith, R.A., Manassaram-Baptiste, D., Brooks, D., Doroshenk, M., Fedewa, S., Saslow, D., Brawley, O.W. & Wender, R. 2015, "Cancer screening in the United States, 2015: a review of current American cancer society guidelines and current issues in cancer screening", *CA: a cancer journal for clinicians*, vol. 65, no. 1, pp. 30-54.

Snippert, H.J., Schepers, A.G., van Es, J.H., Simons, B.D. & Clevers, H. 2014, "Biased competition between Lgr5 intestinal stem cells driven by oncogenic mutation induces clonal expansion", *EMBO reports*, vol. 15, no. 1, pp. 62-69.

Sondak, V.K. & McArthur, G.A. 2015, "Adjuvant immunotherapy for cancer: the next step", *The Lancet.Oncology*, .

Song, M.H., Ha, J.M., Shin, D.H., Lee, C.H., Old, L. & Lee, S.Y. 2012, "KP-CoT-23 (CCDC83) is a novel immunogenic cancer/testis antigen in colon cancer", *International journal of oncology*, vol. 41, no. 5, pp. 1820-1826.

Sood, R., Bonner, T.I., Makalowska, I., Stephan, D.A., Robbins, C.M., Connors, T.D., Morgenbesser, S.D., Su, K., Faruque, M.U., Pinkett, H., Graham, C., Baxevanis, A.D., Klinger, K.W., Landes, G.M., Trent, J.M. & Carpten, J.D. 2001, "Cloning and characterization of 13 novel transcripts and the human RGS8 gene from the 1q25 region encompassing the hereditary prostate cancer (HPC1) locus", *Genomics*, vol. 73, no. 2, pp. 211-222.

Soreide, K., Janssen, E.A., Soiland, H., Korner, H. & Baak, J.P. 2006, "Microsatellite instability in colorectal cancer", *The British journal of surgery*, vol. 93, no. 4, pp. 395-406.

Sorensen, R.D., Schnack, T.H., Karlsen, M.A. & Hogdall, C.K. 2015, "Serous ovarian, fallopian tube and primary peritoneal cancers: A common disease or separate entities - A systematic review", *Gynecologic oncology*, vol. 136, no. 3, pp. 571-581.

Srinivas, P.R., Kramer, B.S. & Srivastava, S. 2001, "Trends in biomarker research for cancer detection", *The Lancet.Oncology*, vol. 2, no. 11, pp. 698-704.

Srinivasan, R. & Wolchok, J.D. 2004, "Tumor antigens for cancer immunotherapy: therapeutic potential of xenogeneic DNA vaccines", *Journal of translational medicine,* vol. 2, no. 1, pp. 12.

Srivastava, P., Paluch, B.E., Matsuzaki, J., James, S.R., Collamat-Lai, G., Taverna, P., Karpf, A.R. & Griffiths, E.A. 2015, "Immunomodulatory action of the DNA methyltransferase inhibitor SGI-110 in epithelial ovarian cancer cells and xenografts", *Epigenetics : official journal of the DNA Methylation Society,* vol. 10, no. 3, pp. 237-246.

Stephens, P.J., Greenman, C.D., Fu, B., Yang, F., Bignell, G.R., Mudie, L.J., Pleasance, E.D., Lau, K.W., Beare, D., Stebbings, L.A., McLaren, S., Lin, M.L., McBride, D.J., Varela, I., Nik-Zainal, S., Leroy, C., Jia, M., Menzies, A., Butler, A.P., Teague, J.W., Quail, M.A., Burton, J., Swerdlow, H., Carter, N.P., Morsberger, L.A., Iacobuzio-Donahue, C., Follows, G.A., Green, A.R., Flanagan, A.M., Stratton, M.R., Futreal, P.A. & Campbell, P.J. 2011, "Massive genomic rearrangement acquired in a single catastrophic event during cancer development", *Cell,* vol. 144, no. 1, pp. 27-40.

Stevanovic, S., Draper, L.M., Langhan, M.M., Campbell, T.E., Kwong, M.L., Wunderlich, J.R., Dudley, M.E., Yang, J.C., Sherry, R.M., Kammula, U.S., Restifo, N.P., Rosenberg, S.A. & Hinrichs, C.S. 2015, "Complete Regression of Metastatic Cervical Cancer After Treatment With Human Papillomavirus-Targeted Tumor-Infiltrating T Cells", *Journal of clinical oncology : official journal of the American Society of Clinical Oncology,* .

Stevens, D., Oegema, K. & Desai, A. 2013, "Meiotic double-strand breaks uncover and protect against mitotic errors in the C. elegans germline", *Current biology : CB,* vol. 23, no. 23, pp. 2400-2406.

Stracci, F., Zorzi, M. & Grazzini, G. 2014, "Colorectal cancer screening: tests, strategies, and perspectives", *Frontiers in public health,* vol. 2, pp. 210.

Stranding S. 2015. Gray's Anatomy, 41st Edition, Elsevier.

Stratton, M.R., Campbell, P.J. & Futreal, P.A. 2009, "The cancer genome", *Nature,* vol. 458, no. 7239, pp. 719-724.

Strunnikov, A. 2013, "Cohesin complexes with a potential to link mammalian meiosis to cancer", *Cell regeneration (London, England),* vol. 2, no. 1, pp. 4-9769-2-4. eCollection 2013.

Sung, P.L., Chang, Y.H., Chao, K.C., Chuang, C.M. & Task Force on Systematic Review and Meta-analysis of Ovarian Cancer 2014, "Global distribution pattern of histological subtypes of epithelial ovarian cancer: a database analysis and systematic review", *Gynecologic oncology,* vol. 133, no. 2, pp. 147-154.

Suzuki, H., Fukuhara, M., Yamaura, T., Mutoh, S., Okabe, N., Yaginuma, H., Hasegawa, T., Yonechi, A., Osugi, J., Hoshino, M., Kimura, T., Higuchi, M., Shio, Y., Ise, K., Takeda, K. & Gotoh, M. 2013, "Multiple therapeutic peptide vaccines consisting of combined novel cancer testis antigens and anti-angiogenic peptides for patients with non-small cell lung cancer", *Journal of translational medicine,* vol. 11, pp. 97-5876-11-97.

Suzuki, M. & Yoshino, I. 2008, "Identification of microRNAs caused by DNA methylation that induce metastasis", *Future oncology (London, England)*, vol. 4, no. 6, pp. 775-777.

Swanton, C., Larkin, J.M., Gerlinger, M., Eklund, A.C., Howell, M., Stamp, G., Downward, J., Gore, M., Futreal, P.A., Escudier, B., Andre, F., Albiges, L., Beuselinck, B., Oudard, S., Hoffmann, J., Gyorffy, B., Torrance, C.J., Boehme, K.A., Volkmer, H., Toschi, L., Nicke, B., Beck, M. & Szallasi, Z. 2010, "Predictive biomarker discovery through the parallel integration of clinical trial and functional genomics datasets", *Genome medicine*, vol. 2, no. 8, pp. 53.

Sylvester BE & Vakiani E. 2015, "Tumor evolution and intratumor heterogeneity in colorectal carcinoma: insights from comparative genomic profiling of primary tumors and matched metastases", *Journal of gastrointestinal oncology*, vol. 6, no. 6, pp. 668-675.

Taguchi, A., Taylor, A.D., Rodriguez, J., Celiktas, M., Liu, H., Ma, X., Zhang, Q., Wong, C.H., Chin, A., Girard, L., Behrens, C., Lam, W.L., Lam, S., Minna, J.D., Wistuba, I.I., Gazdar, A.F. & Hanash, S.M. 2014, "A search for novel cancer/testis antigens in lung cancer identifies VCX/Y genes, expanding the repertoire of potential immunotherapeutic targets", *Cancer research*, vol. 74, no. 17, pp. 4694-4705.

Tajima, K., Obata, Y., Tamaki, H., Yoshida, M., Chen, Y.T., Scanlan, M.J., Old, L.J., Kuwano, H., Takahashi, T., Takahashi, T. & Mitsudomi, T. 2003, "Expression of cancer/testis (CT) antigens in lung cancer", *Lung cancer (Amsterdam, Netherlands)*, vol. 42, no. 1, pp. 23-33.

Takahashi, N., Ohkuri, T., Homma, S., Ohtake, J., Wakita, D., Togashi, Y., Kitamura, H., Todo, S. & Nishimura, T. 2012, "First clinical trial of cancer vaccine therapy with artificially synthesized helper/ killer-hybrid epitope long peptide of MAGE-A4 cancer antigen", *Cancer science*, vol. 103, no. 1, pp. 150-153.

Tamas, K., Walenkamp, A.M., de Vries, E.G., van Vugt, M.A., Beets-Tan, R.G., van Etten, B., de Groot, D.J. & Hospers, G.A. 2015, "Rectal and colon cancer: Not just a different anatomic site", *Cancer treatment reviews*, .

Tammela, J., Jungbluth, A.A., Qian, F., Santiago, D., Scanlan, M.J., Keitz, B., Driscoll, D., Rodabaugh, K., Lele, S., Old, L.J. & Odunsi, K. 2004, "SCP-1 cancer/testis antigen is a prognostic indicator and a candidate target for immunotherapy in epithelial ovarian cancer", *Cancer immunity*, vol. 4, pp. 10.

Tan, T.Z., Miow, Q.H., Huang, R.Y., Wong, M.K., Ye, J., Lau, J.A., Wu, M.C., Bin Abdul Hadi, L.H., Soong, R., Choolani, M., Davidson, B., Nesland, J.M., Wang, L.Z., Matsumura, N., Mandai, M., Konishi, I., Goh, B.C., Chang, J.T., Thiery, J.P. & Mori, S. 2013, "Functional genomics identifies five distinct molecular subtypes with clinical relevance and pathways for growth control in epithelial ovarian cancer", *EMBO molecular medicine*, vol. 5, no. 7, pp. 983-998.

Tan, Y., Liu, L., Liao, M., Zhang, C., Hu, S., Zou, M., Gu, M. & Li, X. 2015, "Emerging roles for PIWI proteins in cancer", *Acta biochimica et biophysica Sinica*, vol. 47, no. 5, pp. 315-324.

Tao, S., Wang, Z., Feng, J., Hsu, FC., Jin, G., Kim, ST., Zhang, Z., Gronberg, H., Zheng, LS., Isaacs, WB., Xu, J. & Sun, J. 2012, "A genome-wide search for loci interacting with known prostate cancer risk-associated genetic variants", *Carcinogenesis*, vol. 33, no. 3, pp. 598-603.

Tarabay, Y., Kieffer, E., Teletin, M., Celebi, C., Van Montfoort, A., Zamudio, N., Achour, M., El Ramy, R., Gazdag, E., Tropel, P., Mark, M., Bourc'his, D. & Viville, S. 2013, "The mammalian-specific Tex19.1 gene plays an essential role in spermatogenesis and placenta-supported development", *Human reproduction (Oxford, England)*, vol. 28, no. 8, pp. 2201-2214.

Tchabo, N.E., Mhawech-Fauceglia, P., Caballero, O.L., Villella, J., Beck, A.F., Miliotto, A.J., Liao, J., Andrews, C., Lele, S., Old, L.J. & Odunsi, K. 2009, "Expression and serum immunoreactivity of developmentally restricted differentiation antigens in epithelial ovarian cancer", *Cancer immunity*, vol. 9, pp. 6.

Tessema, M., Yu, YY., Stidley, CA., Machida, EO., Schuebel, KE., Baylin, SB. & Belinsky, SA. 2009, "Concomitant promoter methylation of multiple genes in lung adenocarcinomas from current, former and never smokers", *Carcinogenesis*, vol. 30, no. 7, pp. 1132-1138.

Thorslund, T., McIlwraith, M.J., Compton, S.A., Lekomtsev, S., Petronczki, M., Griffith, J.D. & West, S.C. 2010, "The breast cancer tumor suppressor BRCA2 promotes the specific targeting of RAD51 to single-stranded DNA",*Nature structural & molecular biology*, vol. 17, no. 10, pp. 1263-1265.

Thorslund, T. & West, S.C. 2007, "BRCA2: a universal recombinase regulator", *Oncogene*, vol. 26, no. 56, pp. 7720-7730.

Tian, L., Wu, X., Lin, Y., Liu, Z., Xiong, F., Han, Z., Zhou, Y., Zeng, Q., Wang, Y., Deng, J. & Chen, H. 2009, "Characterization and potential function of a novel pre-implantation embryo-specific RING finger protein: TRIML1",*Molecular reproduction and development*, vol. 76, no. 7, pp. 656-664.

Tiffen, J.C., Bailey, C.G., Marshall, A.D., Metierre, C., Feng, Y., Wang, Q., Watson, S.L., Holst, J. & Rasko, J.E. 2013, "The cancer-testis antigen BORIS phenocopies the tumor suppressor CTCF in normal and neoplastic cells",*International journal of cancer.Journal international du cancer*, vol. 133, no. 7, pp. 1603-1613.

Topalian, S.L., Drake, C.G. & Pardoll, D.M. 2015, "Immune checkpoint blockade: a common denominator approach to cancer therapy", *Cancer cell*, vol. 27, no. 4, pp. 450-461.

Topalian, S.L., Weiner, G.J. & Pardoll, D.M. 2011, "Cancer immunotherapy comes of age", *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 29, no. 36, pp. 4828-4836.

Trojan, A., Witzens, M., Schultze, J.L., Vonderheide, R.H., Harig, S., Krackhardt, A.M., Stahel, R.A. & Gribben, J.G. 2001, "Generation of cytotoxic T lymphocytes against native and altered peptides of human leukocyte antigen-A*0201 restricted epitopes from

the human epithelial cell adhesion molecule", *Cancer research,* vol. 61, no. 12, pp. 4761-4765.

Tse, B.W., Collins, A., Oehler, M.K., Zippelius, A. & Heinzelmann-Schwarz, V.A. 2014, "Antibody-based immunotherapy for ovarian cancer: where are we at?", *Annals of Oncology : Official Journal of the European Society for Medical Oncology / ESMO,* vol. 25, no. 2, pp. 322-331.

Tsolakis, AV., Grimelius, L. & Islam, MS. 2012, "Expression of the coiled coil domain containing protein 116 in the pancreatic islets and endocrine pancreatic tumors", *Islets,* vol. 4, no. 5, pp. 349-353.

Tureci, O., Chen, Y.T., Sahin, U., Gure, A.O., Zwick, C., Villena, C., Tsang, S., Seitz, G., Old, L.J. & Pfreundschuh, M. 1998, "Expression of SSX genes in human tumors", *International journal of cancer.Journal international du cancer,* vol. 77, no. 1, pp. 19-23.

Tureci, O., Sahin, U., Koslowski, M., Buss, B., Bell, C., Ballweber, P., Zwick, C., Eberle, T., Zuber, M., Villena-Heinsen, C., Seitz, G. & Pfreundschuh, M. 2002, "A novel tumour associated leucine zipper protein targeting to sites of gene transcription and splicing", *Oncogene,* vol. 21, no. 24, pp. 3879-3888.

Tureci, O., Sahin, U., Zwick, C., Koslowski, M., Seitz, G. & Pfreundschuh, M. 1998, "Identification of a meiosis-specific protein as a member of the class of cancer/testis antigens", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 95, no. 9, pp. 5211-5216.

Turner, J.M. 2007, "Meiotic sex chromosome inactivation", *Development (Cambridge, England),* vol. 134, no. 10, pp. 1823-1831.

Uhlen, M., Bjorling, E., Agaton, C., Szigyarto, CA., Amini, B., Andersen, E. *et al.* 2005, "A human protein atlas for normal and cancer tissues based on antibody proteomics", *Molecular and cellular proteomics,* vol. 4, no. 12, pp. 1920-1932.

Ullah, M.F. & Aatif, M. 2009, "The footprints of cancer development: Cancer biomarkers", *Cancer treatment reviews,* vol. 35, no. 3, pp. 193-200.

Usener, D., Schadendorf, D., Koch, J., Dubel, S. & Eichmuller, S. 2003, "cTAGE: a cutaneous T cell lymphoma associated antigen family with tumor-specific splicing", *The Journal of investigative dermatology,* vol. 121, no. 1, pp. 198-206.

Vale, C.L., Tierney, J.F., Fisher, D., Adams, R.A., Kaplan, R., Maughan, T.S., Parmar, M.K. & Meade, A.M. 2012, "Does anti-EGFR therapy improve outcome in advanced colorectal cancer? A systematic review and meta-analysis", *Cancer treatment reviews,* vol. 38, no. 6, pp. 618-625.

Valmori, D., Souleimanian, N.E., Tosello, V., Bhardwaj, N., Adams, S., O'Neill, D., Pavlick, A., Escalon, J.B., Cruz, C.M., Angiulli, A., Angiulli, F., Mears, G., Vogel, S.M., Pan, L., Jungbluth, A.A., Hoffmann, E.W., Venhaus, R., Ritter, G., Old, L.J. & Ayyoub, M. 2007, "Vaccination with NY-ESO-1 protein and CpG in Montanide induces integrated

antibody/Th1 responses and CD8 T cells through cross-priming", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 104, no. 21, pp. 8947-8952.

Van den Eynde, B., Peeters, O., De Backer, O., Gaugler, B., Lucas, S. & Boon, T. 1995, "A new family of genes coding for an antigen recognized by autologous cytolytic T lymphocytes on a human melanoma", *The Journal of experimental medicine,* vol. 182, no. 3, pp. 689-698.

van der Bruggen, P., Traversari, C., Chomez, P., Lurquin, C., De Plaen, E., Van den Eynde, B., Knuth, A. & Boon, T. 1991, "A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma", *Science (New York, N.Y.),* vol. 254, no. 5038, pp. 1643-1647.

van Wezel, T., Middeldorp, A., Wijnen, J.T. & Morreau, H. 2012, "A review of the genetic background and tumour profiling in familial colorectal cancer", *Mutagenesis,* vol. 27, no. 2, pp. 239-245.

Vaughan, S., Coward, J.I., Bast, R.C.,Jr, Berchuck, A., Berek, J.S., Brenton, J.D., Coukos, G., Crum, C.C., Drapkin, R., Etemadmoghadam, D., Friedlander, M., Gabra, H., Kaye, S.B., Lord, C.J., Lengyel, E., Levine, D.A., McNeish, I.A., Menon, U., Mills, G.B., Nephew, K.P., Oza, A.M., Sood, A.K., Stronach, E.A., Walczak, H., Bowtell, D.D. & Balkwill, F.R. 2011, "Rethinking ovarian cancer: recommendations for improving outcomes", *Nature reviews.Cancer,* vol. 11, no. 10, pp. 719-725.

Vicente-Duenas, C., Romero-Camarero, I., Cobaleda, C. & Sanchez-Garcia, I. 2013, "Function of oncogenes in cancer development: a changing paradigm", *The EMBO journal,* vol. 32, no. 11, pp. 1502-1513.

Vilar, E. & Gruber, S.B. 2010, "Microsatellite instability in colorectal cancer-the stable evidence", *Nature reviews.Clinical oncology,* vol. 7, no. 3, pp. 153-162.

Viphakone, N., Cumberbatch, M.G., Livingstone, M.J., Heath, P.R., Dickman, M.J., Catto, J.W. & Wilson, S.A. 2015, "Luzp4 defines a new mRNA export pathway in cancer cells", *Nucleic acids research,* vol. 43, no. 4, pp. 2353-2366.

Viphakone, N., Cumberbatch, M.G., Livingstone, M.J., Heath, P.R., Dickman, M.J., Catto, J.W. & Wilson, S.A. 2015, "Luzp4 defines a new mRNA export pathway in cancer cells", *Nucleic acids research,* vol. 43, no. 4, pp. 2353-2366.

Visvader JE. 2011, “Cells of origin in cancer”, *Nature,* vol. 469, no. 7330, pp. 314-322.

Visvader JE & Clevers H. 2016, “Tissue-specific designs of stem cell hierarchies”, *Nature Cell Biology,* vol. 18, no. 4, pp. 349-355.

Visvader JE & Lindeman GL. 2012, “Cancer stem cells: current status and evolving complexities”, *Cell stem cell,* vol. 10, no. 6, pp. 717-728.

Vizoso, F., Plaza, E., Vazquez, J., Serra, C., Lamelas, M.L., Gonzalez, L.O., Merino, A.M. & Mendez, J. 2001, "Lysozyme expression by breast carcinomas, correlation with

clinicopathologic parameters, and prognostic significance", *Annals of surgical oncology*, vol. 8, no. 8, pp. 667-674.

Vockley, J.G. & Niederhuber, J.E. 2015, "Diagnosis and treatment of cancer using genomics", *BMJ (Clinical research ed.)*, vol. 350, pp. h1832.

Vogelstein, B. & Kinzler, K.W. 2004, "Cancer genes and the pathways they control", *Nature medicine*, vol. 10, no. 8, pp. 789-799.

Wang, J., Emadali, A., Le Bescont, A., Callanan, M., Rousseaux, S. & Khochbin, S. 2011, "Induced malignant genome reprogramming in somatic cells by testis-specific factors", *Biochimica et biophysica acta*, vol. 1809, no. 4-6, pp. 221-225.

Wang, J., Rousseaux, S. & Khochbin, S. 2014, "Sustaining cancer through addictive ectopic gene activation", *Current opinion in oncology*, vol. 26, no. 1, pp. 73-77.

Wang, K., Sturt-Gillespie, B., Hittle, J.C., Macdonald, D., Chan, G.K., Yen, T.J. & Liu, S.T. 2014, "Thyroid hormone receptor interacting protein 13 (TRIP13) AAA-ATPase is a novel mitotic checkpoint-silencing protein", *The Journal of biological chemistry*, vol. 289, no. 34, pp. 23928-23937.

Wang, W., Guan, S., Sun, S., Jin, Y., Lee, K.H., Chen, Y. & Wei, J. 2014, "Detection of circulating antibodies to linear peptide antigens derived from ANXA1 and DDX53 in lung cancer", *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*, vol. 35, no. 5, pp. 4901-4905.

Wang, X., Baddoo, M.C. & Yin, Q. 2014, "The placental specific gene, PLAC1, is induced by the Epstein-Barr virus and is expressed in human tumor cells", *Virology journal*, vol. 11, pp. 107-422X-11-107.

Wang, X.L., Wei, Y.H., Fu, G.L. & Yu, L. 2015, "Testis specific serine/threonine protein kinase 4 (TSSK4) leads to cell apoptosis relying on its kinase activity", *Journal of Huazhong University of Science and Technology.Medical sciences = Hua zhong ke ji da xue xue bao.Yi xue Ying De wen ban = Huazhong keji daxue xuebao.Yixue Yingdewen ban*, vol. 35, no. 2, pp. 235-240.

Wang, Y., Han, K.J., Pang, X.W., Vaughan, H.A., Qu, W., Dong, X.Y., Peng, J.R., Zhao, H.T., Rui, J.A., Leng, X.S., Cebon, J., Burgess, A.W. & Chen, W.F. 2002, "Large scale identification of human hepatocellular carcinoma-associated antigens by autoantibodies", *Journal of immunology (Baltimore, Md.: 1950)*, vol. 169, no. 2, pp. 1102-1109.

Wang, Y., Xu MD., Weng WW., Wei P., Yang YS. & Du X. 2015b, "BCL6 is a negative prognostic factor and exhibits pro-oncogenic activity in ovarian cancer", *American journal of cancer research*, vol. 5, no. 1, pp. 255-266.

Wang, Y., Wang, L. & Wang, Z. 2008, "Transgenic analyses of TGIF family proteins in Drosophila imply their role in cell growth", *Journal of genetics and genomics = Yi chuan xue bao*, vol. 35, no. 8, pp. 457-465.

Watanabe, T. & Lin, H. 2014, "Posttranscriptional regulation of gene expression by Piwi proteins and piRNAs", *Molecular cell*, vol. 56, no. 1, pp. 18-27.

Watkins, J., Weekes, D., Shah, V., Gazinska, P., Joshi, S., Sidhu, B., Gillett, C., Pinder, S., Vanoli, F., Jasin, M., Mayrhofer, M., Isaksson, A., Cheang, M.C., Mirza, H., Frankum, J., Lord, C.J., Ashworth, A., Vinayak, S., Ford, J.M., Telli, M.L., Grigoriadis, A. & Tutt, A.N. 2015, "Genomic complexity profiling reveals that HORMAD1 overexpression contributes to homologous recombination deficiency in triple-negative breast cancers", *Cancer discovery*, .

Wefers, C., Lambert, L.J., Torensma, R. & Hato, S.V. 2015, "Cellular immunotherapy in ovarian cancer: Targeting the stem of recurrence", *Gynecologic oncology*, vol. 137, no. 2, pp. 335-342.

Wei, J., Li, SJ., Shi, H., Wang, HY., Rong, CT., Zhu, P., Jin, SH., Liu, J. & Li, JY. 2013, "Characterisation of Lyzls in mice and antibacterial properties of LYZL6", *Asian journal of andrology*, vol. 15, no. 6, pp. 824-830.

Weissman, S.M., Bellcross, C., Bittner, C.C., Freivogel, M.E., Haidle, J.L., Kaurah, P., Leininger, A., Palaniappan, S., Steenblock, K., Vu, T.M. & Daniels, M.S. 2011, "Genetic counseling considerations in the evaluation of families for Lynch syndrome--a review", *Journal of genetic counseling*, vol. 20, no. 1, pp. 5-19.

Wen, J., Li, H., Tao, W., Savoldo, B., Foglesong, J.A., King, L.C., Zu, Y. & Chang, C.C. 2014, "High throughput quantitative reverse transcription PCR assays revealing over-expression of cancer testis antigen genes in multiple myeloma stem cell-like side population cells", *British journal of haematology*, vol. 166, no. 5, pp. 711-719.

Westbrook, V.A., Schoppee, P.D., Diekman, A.B., Klotz, K.L., Allietta, M., Hogan, K.T., Slingluff, C.L., Patterson, J.W., Frierson, H.F., Irvin, W.P.,Jr, Flickinger, C.J., Coppola, M.A. & Herr, J.C. 2004, "Genomic organization, incidence, and localization of the SPAN-x family of cancer-testis antigens in melanoma tumors and cell lines", *Clinical cancer research : an official journal of the American Association for Cancer Research*, vol. 10, no. 1 Pt 1, pp. 101-112.

White C, Thomas R, Long T, Slyne C, Howe J, Crowther H & Huws DW. 2015, "Cancer in Wales", *Wales Cancer Intelligence and Surveillance Unit* official statistics 2013 data, Health Intelligence Division, Public Health Wales

Whitehurst, A.W. 2014, "Cause and consequence of cancer/testis antigen activation in cancer", *Annual Review of Pharmacology and Toxicology*, vol. 54, pp. 251-272.

Whitehurst, A.W., Bodemann, B.O., Cardenas, J., Ferguson, D., Girard, L., Peyton, M., Minna, J.D., Michnoff, C., Hao, W., Roth, M.G., Xie, X.J. & White, M.A. 2007, "Synthetic lethal screen identification of chemosensitizer loci in cancer cells", *Nature*, vol. 446, no. 7137, pp. 815-819.

Wojtasz, L., Cloutier, J.M., Baumann, M., Daniel, K., Varga, J., Fu, J., Anastassiadis, K., Stewart, A.F., Remenyi, A., Turner, J.M. & Toth, A. 2012, "Meiotic DNA double-strand breaks and chromosome asynapsis in mice are monitored by distinct HORMAD2-

independent and -dependent mechanisms", *Genes & development*, vol. 26, no. 9, pp. 958-973.

Woods, N.T., Mesquita, R.D., Sweet, M., Carvalho, M.A., Li, X., Liu, Y., Nguyen, H., Thomas, C.E., Iversen, E.S.,Jr, Marsillac, S., Karchin, R., Koomen, J. & Monteiro, A.N. 2012, "Charting the landscape of tandem BRCT domain-mediated protein interactions", *Science signaling*, vol. 5, no. 242, pp. rs6.

Wu, Y., Dowbenko, D., Pisabarro, M.T., Dillard-Telm, L., Koeppen, H. & Lasky, L.A. 2001, "PTEN 2, a Golgi-associated testis-specific homologue of the PTEN tumor suppressor lipid phosphatase", *The Journal of biological chemistry*, vol. 276, no. 24, pp. 21745-21753.

Xia, B., Sheng, Q., Nakanishi, K., Ohashi, A., Wu, J., Christ, N., Liu, X., Jasin, M., Couch, F.J. & Livingston, D.M. 2006, "Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2", *Molecular cell*, vol. 22, no. 6, pp. 719-729.

Xia, J., Jia, P., Hutchinson, K.E., Dahlman, K.B., Johnson, D., Sosman, J., Pao, W. & Zhao, Z. 2014, "A meta-analysis of somatic mutations from next generation sequencing of 241 melanomas: a road map for the study of genes with potential clinical relevance", *Molecular cancer therapeutics*, vol. 13, no. 7, pp. 1918-1928.

Xu, H., Gu, N., Liu, Z.B., Zheng, M., Xiong, F., Wang, S.Y., Li, N. & Lu, J. 2012, "NY-ESO-1 expression in hepatocellular carcinoma: A potential new marker for early recurrence after surgery", *Oncology letters*, vol. 3, no. 1, pp. 39-44.

Xu, K., Liu, P. & Wei, W. 2014, "mTOR signaling in tumorigenesis", *Biochimica et biophysica acta*, vol. 1846, no. 2, pp. 638-654.

Xu, X., Tang, X., Lu, M., Tang, Q., Zhang, H., Zhu, H., Xu, N., Zhang, D., Xiong, L., Mao, Y. & Zhu, J. 2014, "Overexpression of MAGE-A9 predicts unfavorable outcome in breast cancer", *Experimental and molecular pathology*, vol. 97, no. 3, pp. 579-584.

Yakirevich, E., Sabo, E., Lavie, O., Mazareb, S., Spagnoli, G.C. & Resnick, M.B. 2003, "Expression of the MAGE-A4 and NY-ESO-1 cancer-testis antigens in serous ovarian neoplasms", *Clinical cancer research : an official journal of the American Association for Cancer Research*, vol. 9, no. 17, pp. 6453-6460.

Yamada, R., Takahashi, A., Torigoe, T., Morita, R., Tamura, Y., Tsukahara, T., Kanaseki, T., Kubo, T., Watarai, K., Kondo, T., Hirohashi, Y. & Sato, N. 2013, "Preferential expression of cancer/testis genes in cancer stem-like cells: proposal of a novel sub-category, cancer/testis/stem gene", *Tissue antigens*, vol. 81, no. 6, pp. 428-434.

Yan, H.H., Mruk, D.D., Lee, W.M. & Cheng, C.Y. 2008, "Blood-testis barrier dynamics are regulated by testosterone and cytokines via their differential effects on the kinetics of protein endocytosis and recycling in Sertoli cells", *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, vol. 22, no. 6, pp. 1945-1959.

Yang, F., Eckardt, S., Leu, N.A., McLaughlin, K.J. & Wang, P.J. 2008, "Mouse TEX15 is essential for DNA double-strand break repair and chromosomal synapsis during male meiosis", *The Journal of cell biology*, vol. 180, no. 4, pp. 673-679.

Yang, P., Huo, Z., Liao, H. & Zhou, Q. 2015, "Cancer/testis antigens trigger epithelial-mesenchymal transition and genesis of cancer stem-like cells", *Current pharmaceutical design*, vol. 21, no. 10, pp. 1292-1300.

Yeh, C.H., Kuo, P.L., Wang, Y.Y., Wu, Y.Y., Chen, M.F., Lin, D.Y., Lai, T.H., Chiang, H.S. & Lin, Y.H. 2015, "SEPT12/SPAG4/LAMINB1 complexes are required for maintaining the integrity of the nuclear envelope in postmeiotic male germ cells", *PloS one*, vol. 10, no. 3, pp. e0120722.

Yi, F., Ni, W., Liu, W., Pan, X., Han, X., Yang, L., Kong, X., Ma, R. & Chang, R. 2013, "SPAG9 is overexpressed in human astrocytoma and promotes cell proliferation and invasion", *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*, vol. 34, no. 5, pp. 2849-2855.

Yilmaz-Ozcan, S., Sade, A., Kucukkaraduman, B., Kaygusuz, Y., Senses, K.M., Banerjee, S. & Gure, A.O. 2014, "Epigenetic mechanisms underlying the dynamic expression of cancer-testis genes, PAGE2, -2B and SPANX-B, during mesenchymal-to-epithelial transition", *PloS one*, vol. 9, no. 9, pp. e107905.

Yin, L.L., Chung, C.M., Chen, J., Fok, K.L., Ng, C.P., Jia, R.R., Ren, X., Zhou, J., Zhang, T., Zhao, X.H., Lin, M., Zhu, H., Zhang, X.H., Tsang, L.L., Bi, Y., Zhou, Z., Mo, F., Wong, N., Chung, Y.W., Sha, J. & Chan, H.C. 2009, "A suppressor of multiple extracellular matrix-degrading proteases and cancer metastasis", *Journal of Cellular and Molecular Medicine*, vol. 13, no. 9B, pp. 4034-4041.

Yokoe, T., Tanaka, F., Mimori, K., Inoue, H., Ohmachi, T., Kusunoki, M. & Mori, M. 2008, "Efficient identification of a novel cancer/testis antigen for immunotherapy using three-step microarray analysis", *Cancer research*,vol. 68, no. 4, pp. 1074-1082.

Yoon, H.J., Kozminsky, M. & Nagrath, S. 2014, "Emerging role of nanomaterials in circulating tumor cell isolation and analysis", *ACS nano*, vol. 8, no. 3, pp. 1995-2017.

Yoshitake, H., Yokoi, H., Ishikawa, H., Maruyama, M., Endo, S., Nojima, M., Yoshida, K., Yoshikawa, H., Suzuki, F., Takamori, K., Fujiwara, H. & Araki, Y. 2012, "Overexpression of TEX101, a potential novel cancer marker, in head and neck squamous cell carcinoma", *Cancer biomarkers : section A of Disease markers*, vol. 12, no. 3, pp. 141-148.

Younes, A., Gopal, A.K., Smith, S.E., Ansell, S.M., Rosenblatt, J.D., Savage, K.J., Ramchandren, R., Bartlett, N.L., Cheson, B.D., de Vos, S., Forero-Torres, A., Moskowitz, C.H., Connors, J.M., Engert, A., Larsen, E.K., Kennedy, D.A., Sievers, E.L. & Chen, R. 2012, "Results of a pivotal phase II study of brentuximab vedotin for patients with relapsed or refractory Hodgkin's lymphoma", *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 30, no. 18, pp. 2183-2189.

Yuan, J., Gnjatic, S., Li, H., Powel, S., Gallardo, H.F., Ritter, E., Ku, G.Y., Jungbluth, A.A., Segal, N.H., Rasalan, T.S., Manukian, G., Xu, Y., Roman, R.A., Terzulli, S.L., Heywood, M., Pogoriler, E., Ritter, G., Old, L.J., Allison, J.P. & Wolchok, J.D. 2008, "CTLA-4 blockade enhances polyfunctional NY-ESO-1 specific T cell responses in metastatic melanoma patients with clinical benefit", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 105, no. 51, pp. 20410-20415.

Zampieri, M., Ciccarone, F., Palermo, R., Cialfi, S., Passananti, C., Chiaretti, S., Nocchia, D., Talora, C., Screpanti, I. & Caiafa, P. 2014, "The epigenetic factor BORIS/CTCFL regulates the NOTCH3 gene expression in cancer cells", *Biochimica et biophysica acta,* vol. 1839, no. 9, pp. 813-825.

Zhang, K., Tang, S., Cao, S., Hu, L., Pan, Y., Ma, H., Guo, X., Wu, S., Shen, H. & Hu, Z. 2014, "Association of polymorphisms at HORMAD2 and prognosis in advanced non-small-cell lung cancer patients", *Cancer epidemiology,* vol. 38, no. 4, pp. 414-418.

Zhang, S., Yue, J.X., Yang, J.H., Cai, P.C. & Kong, W.J. 2008, "Overexpression of transketolase protein TKTL1 is associated with occurrence and progression in nasopharyngeal carcinoma: a potential therapeutic target in nasopharyngeal carcinoma", *Cancer biology & therapy,* vol. 7, no. 4, pp. 517-522.

Zhang, X., Yang, H., Lee, J.J., Kim, E., Lippman, S.M., Khuri, F.R., Spitz, M.R., Lotan, R., Hong, W.K. & Wu, X. 2010, "MicroRNA-related genetic variations as predictors for risk of second primary tumor and/or recurrence in patients with early-stage head and neck cancer", *Carcinogenesis,* vol. 31, no. 12, pp. 2118-2123.

Zhou, X., Yu, Y., Tao, J. & Yu, L. 2014, "Production of LYZL6, a novel human c-type lysozyme, in recombinant Pichia pastoris employing high cell density fed-batch fermentation", *Journal of bioscience and bioengineering,* vol. 118, no. 4, pp. 420-425.

Zimmermann, A.K., Imig, J., Klar, A., Renner, C., Korol, D., Fink, D., Stadlmann, S., Singer, G., Knuth, A., Moch, H. & Caduff, R. 2013, "Expression of MAGE-C1/CT7 and selected cancer/testis antigens in ovarian borderline tumours and primary and recurrent ovarian carcinomas", *Virchows Archiv : an international journal of pathology,* vol. 462, no. 5, pp. 565-574.

Zou, S.W., Zhang, J.C., Zhang, X.D., Miao, S.Y., Zong, S.D., Sheng, Q. & Wang, L.F. 2003, "Expression and localization of VCX/Y proteins and their possible involvement in regulation of ribosome assembly during spermatogenesis", *Cell research,* vol. 13, no. 3, pp. 171-177.

Zuo, X., Chen, L., Liu, L., Zhang, Z., Zhang, X., Yu, Q., Feng, L., Zhao, X. & Qin, T. 2016, “Identification of a panel of complex autoantigens (LGALS3, PHB2, MUC1 and GK2) in combination with CA15-3 for the diagnosis of early-stage breast cancer. *Tumor biology*, vol. 37, no. 1, pp. 1309-1317.

Appendix - Table of Contents

SECTION 3 – ADDITIONAL FIGURES 258

Appendix

Section 1 – Further discussion

Discussion of additional germline genes assessed using TLDA cards

MeiCT genes

PFN3 and *PPP3R2* were both identified by our group as meiCT genes but have not otherwise been described as CTAs. The normal physiological role of PFN3 remains unclear but it interacts with proteins close to the acrosome of developing sperm and its functions are believed to be distinct from other somatic members of this protein group (Behnen *et al*., 2009). Low-moderate levels of expression were seen in three separate cancer types on the TLDA arrays, though not clearly expressed in any of the colon cancer samples (see Figure 5.16, group AC). *PPP3R2* has previously been shown to be a testis-specific gene, which may have a role in calcium-mediated signal transduction pathways and like many CT genes is activated by demethylation of its promoter region within the testis (Liu *et al*., 2005). The TLDA array analysis revealed this gene to have a testis/CNS restricted expression pattern and in addition be expressed in one ovarian cancer sample. The TLDA result therefore concurs with our previous validation of these genes as CT genes that may be of clinical relevance in a subset of cancers.

We characterised *TRIM42* as a novel meiCT gene, which was not detectable in any cancer tissues or cell lines but had a significant upregulation in prostate cancer when microarray datasets were meta-analysed (Sammut *et al*., 2014). No other studies linking this gene to cancer have been identified in the literature and the TLDA analysis confirmed our previous findings of the gene expression being restricted to testis only; there was no expression in any of the cancer types we tested but those did not include prostate cancer for which an upregulation had been suggested by the microarray datasets (see Figure 5.16, group AC). *C20orf201* is another poorly characterised gene, which had been identified previously as a meiCT gene that was testis/CNS restricted in normal tissue and significantly upregulated in ovarian cancer in the clinical datasets (Feichtinger *et al*., 2012b). The TLDA gene expression results revealed testis/CNS-selective expression, with additional expression in at least one sample of each of the four cancer types tested (see Figure 5.16, group AC).

ODF3 and *ODF4* are testis-specific genes that have not been extensively investigated in cancer though both are considered CT genes (Almeida *et al.*, 2009). We had also characterised these genes as candidate meiCT genes, with both genes appearing to be upregulated in ovarian cancer though only *ODF4* displayed a statistically significant upregulation in gene expression when combined clinical datasets were meta-analysed (Feichtinger *et al*., 2012b). *ODF4* has been found to be expressed in chronic myeloid leukaemia (Ghafouri-Fard *et al*., 2012). Neither gene was found to be expressed in prostate cancer samples but *SPATA19* was, in 40% of the samples tested, in a separate study (Ghafouri-Fard *et al*., 2010); this gene was present on the extended germline gene lists compiled by the McFarlane and Knochbin groups (e.g., see Figure 5.17, group CEF). *ODF3* was not expressed in a small cohort of breast cancer samples tested in another study (Dianatpour *et al*., 2012). We found *ODF3* to be expressed in testis but none of the other samples tested. *ODF4* on the other hand was testis-restricted in the normal tissues but was strongly expressed in one colon cancer and one lung cancer sample (e.g., see Figure 5.16, group ACE). It has not been previously identified as a candidate CTA in either of these cancer types, which makes this a novel finding of this research; its possible role in tumorigenesis remains obscure.

SMC1B and *RAD21L* are meiosis specific genes, which encode subunits of the cohesin complex (Strunnikov, 2013). We identified both of these genes as potential CTAs previously and a possible role for these genes in contributing to chromosomal instability through reactivation of their meiotic function in cancer has been postulated (Feichtinger *et al*., 2012; Strunnikov, 2013). *SMC1B* was revealed as having a micro-RNA binding site single nucleotide polymorphism associated with recurrence in head and neck cancers (Zhang *et al*., 2010). SMC1B has been shown to be present in pancreatic cancer and to interact with p53 (Ansari *et al*., 2015). Increased expression levels have also been observed in a radiotherapy-resistant non-small cell lung cancer cell line compared to a radiosensitive lung cancer cell line (Kim *et al*., 2011). Expression of *SMC1B* has also been linked to more aggressive forms of breast cancer (Hsiao *et al*., 2010). The TLDA analysis revealed testis-selective expression in normal tissue. Expression was seen in foetal liver and also a NAT colonic sample; this latter finding was made more surprising by the fact that the corresponding cancer sample did not express the gene (see Figure 5.16, group A). Two melanoma samples and one ovarian cancer sample also expressed the gene, so it may be contributing to oncogenesis/chromosomal instability in these cancer types.

SPZ1 has previously been shown to have an expression profile tightly restricted to the testis in normal tissues (Sha *et al*., 2003). We independently found it to have a testis-restricted expression profile and classified it as a meiCT gene (Sammut *et al*., 2014). Furthermore, meta-analysis of microarray datasets indicated this gene was significantly upregulated in ovarian cancer (Sammut *et al*., 2014). The TLDA analysis once again revealed a testis-restricted expression pattern in normal tissues. In the cancer samples it was expressed in one colon cancer sample and two ovarian cancer samples (e.g., see Figure 5.16, group ACF). This increases the evidence or likelihood that this gene is clinically relevant in ovarian cancer and this is certainly worthy of further exploration. It has been shown using a mouse model that *SPZ1* regulates *PCNA* expression and promotes cellular proliferation (Hsu *et al*., 2005). It was further suggested that *SPZ1* may act as an oncogene through interaction with the Ras signalling pathway. *KRAS* mutations are common in colorectal cancers and given that *SPZ1* was found to be expressed in one of the colon cancer samples, it would be interesting to investigate if there is an interaction with Ras in a subset of patients with CRC.

We characterised *TRIML1* as a novel meiCT gene that has not otherwise been associated with oncogenesis (Feichtinger *et al*., 2012b). It has been identified as a possible regulator of embryo development in its early preimplantation stage (Tian *et al*., 2009), which is consistent with its categorisation and inclusion here as a germline-associated gene. The TLDA analysis revealed a testis-selective expression profile with weak-moderate expression in normal colon and ovarian samples (see Figure 5.16, group AC). There was strong expression in one of the colon cancer samples only. A possible oncogenic role for this gene remains obscure.

STRA8 is involved in the early stages of meiosis and believed to be an inducer of meiosis in both the testis and ovary; retinoic acid (RA) promotes the expression of this gene in spermatogonia (Rossi and Dolci, 2013). *STRA8* was identified as a novel meiCT gene but has not otherwise been described as a CTA (Feichtinger *et al*., 2012b). The TLDA analysis in this study has indicated, as expected, testis-restricted expression in normal tissues. *STRA8* was expressed in all ovarian cancer samples tested (2 out of the 4 samples were not tested for this gene) and one of the lung cancer samples (see Figure 5.16, group AC). This is a novel and potentially important finding, though an oncogenic or predictive role for this gene in these cancer types remains to be established. It has been suggested that RA may, through *STRA8* expression, drive mTORC1 activation in spermatogonial progenitor cells to influence cellular differentiation within the testis (Hobbs *et al*., 2015). Given the importance of mTOR

signalling in cancer as well as stem cell self-renewal (Cargnello *et al*., 2015; Guertin and Sabatini, 2007; Xu *et al*., 2014), this raises the possibility of *STRA8* impacting on such signalling pathways if ectopically activated in cancer. This is somewhat speculative and would need to be investigated by further research but the confirmation of *STRA8* expression in lung and ovarian cancer is important and interesting.

TULP2 belongs to a family of genes that were originally thought to encode transcription factors and have been implicated as having roles in various biological and signal transduction processes primarily within the CNS (Boggon *et al*., 1999; Carroll *et al*., 2004). *TULP2* has high levels of expression in the testis and low levels of expression in the normal retina, which is another immunologically privileged site; its function remains poorly characterised (Carroll *et al*., 2004). The gene was included in the spermatocyte-associated and germline-specific genes compiled by the McFarlane and Knochbin groups, respectively (Feichtinger *et al*., 2012b; Rousseaux *et al*., 2013). It is also listed as a CT gene but I am not aware of other researchers proposing this gene as a CTA or linking it to cancer (Almeida *et al*., 2009). The TLDA analysis revealed that the gene was expressed in normal testis, foetal liver, ovary and CNS tissues, consistent with it being a testis/CNS-selective CT gene. Within the cancer samples, one colon cancer, one melanoma and two ovarian cancers expressed this gene (e.g., see Figure 5.17, group CEF). Its relevance as a biomarker or therapeutic target in these cancer types remains to be established but another member of this gene family *TULP3*, has recently been shown to be a predictive biomarker in pancreatic cancer (Sartor *et al*., 2014).

LYZL6 is expressed in the male testis and belongs to a family of lysozymes that can cause bacterial cell wall disruption. It has been confirmed, using a yeast expression system, that the protein has antibacterial properties (Zhou *et al*., 2014). I found the gene to be expressed in a melanoma sample and testis only, suggesting that it may be a novel CT gene (see Figure 5.16, group ACF). It was identified as a testis-specific germline gene but not associated with a worse outcome in lung cancer in the Knochbin group study (Rousseaux *et al*., 2013b). It has not been previously described as a CTA and has no known oncogenic effects. Different lysozymes have been shown to suppress proliferation of breast cancer cells (Mahanta *et al*., 2015) but conversely the presence of the protein has been shown to be of differing prognostic significance in breast cancers in men and women (Serra *et al*., 2002; Vizoso *et al*., 2001). It should be noted that these studies were not specifically investigating *LYZL6* or the gene product but some degree of functional overlap between these proteins of the innate immune

system will be expected. The possible relevance of *LYZL6* in cancer requires further clarification but the expression of this gene in cancer is a novel finding of this research.

Our group identified and classified *SEPT12* as a meiCT gene that had a significant upregulation in ovarian cancer microarray datasets (Feichtinger *et al*., 2012b). *SPAG4* is considered a known CTA (Almeida *et al*., 2009), although we found this gene to be widely expressed in the normal tissues tested. There was, however, more consistent and higher expression in the cancer samples. Thus, it may be another example of a CT gene that is expressed in normal tissues but at low levels; it should be remembered that the fold change in gene expression in normal tissues compared to testis and cancer, is an alternative way of classifying CT genes as discussed in the introduction. Consistent with our findings, it has separately been shown that normal heart, lung, kidney, spleen, uterus and placenta expressed *SPAG4* but the strength of the bands on the agarose gel was much weaker in these tissues compared to testis (Yeh *et al*., 2015). This confirms that the readout from the TLDA card is most likely accurate, even though *SPAG4* is considered a CTA. *SEPT12* displayed a testis/CNS-selective expression pattern in the normal tissues we tested; in the cancer samples it was not expressed in the colon cancer but was found in a melanoma, ovarian and lung cancer sample. There was moderately strong expression in the ovarian cancer and melanoma sample that expressed *SEPT12* and *SPAG4* (see Figure 5.16, group ACF and Figure 5.18, group E). It has been shown that the products of these two genes co-localise in developing sperm cells and functionally interact (Yeh *et al*., 2015). It is possible that there is a functional interaction when they are co-expressed in cancer, though this finding could also relate to a common mechanism of transcriptional upregulation, such as demethylation, in cancer.

Known CTA-encoding genes

Although *SPANXC/E* was expressed in three of the NAT colonic samples; this gene was also strongly expressed in the recurrent malignant melanoma sample and the metastatic squamous carcinoma of the lung (e.g., see Figure 5.21, group BEF). *SPANX* was one of three CT genes previously shown to correlate with the presence of liver metastases in CRC (Chen *et al*., 2010). There are numerous members of this gene family and the promotor region is shared with other CT genes, including *VCY*, suggesting a common ancestry (Hansen *et al*., 2008). The specificity of primers to detect individual members of such paralogous gene families creates a difficulty in comparing expression profiles between studies (the assay selected for

the TLDA cards did not differentiate between *SPANXC* and *SPANXE*). *SPANX* genes have previously been shown to be expressed in ovarian cancer cell lines, as well as melanoma cell lines and tumour specimens (Westbrook *et al*., 2004). This is consistent with the expression seen in the melanoma and ovarian cancer samples in this research (see Figure 5.21). More recently the presence of CT genes, including a member of the *SPANX* family, have been shown to negatively correlate with EMT markers and be associated with the epithelial rather than mesenchymal phenotype in a CRC cell line model (Yilmaz-Ozcan *et al*., 2014). As the products of these genes have been shown to elicit immune responses *in vivo* they might be attractive targets for immunotherapy (Almanzar *et al*., 2009).

CTAGE1 is a CTA gene that has been shown to be of potential importance in patients with lymphomas, glioma and melanoma (Akiyama *et al*., 2014; Liggins *et al*., 2010; Litvinov *et al*., 2014; Usener *et al*., 2003). In this study *CTAGE1* was confirmed as a gene with restricted tissue expression in the testis in normal tissues and was expressed in an ovarian cancer sample (e.g., see Figure 5.21, group BEF). This raises the possibility of this CT gene being clinically relevant in ovarian cancer, which has not been described previously.

DDX53 also belongs to the category of known CT genes and was also included in the extended list of germline genes compiled by the Knochbin group (Almeida *et al*., 2009; Rousseaux *et al*., 2013). It was not present on the list of human orthologues from which the meiCT genes were derived. From the TLDA analysis conducted here it would have been categorised as testis/CNS-selective, so would not have been discarded at this validation stage. Within the cancer samples there was moderately high expression in one of the colon cancer samples tested as well as an ovarian cancer sample (e.g., see Figure 5.21, group BEF). The gene was not expressed in either of the two lung cancer samples tested, although it has been proposed and investigated as a possible diagnostic marker in lung cancer patients (Chapman *et al*., 2012; Wang *et al*., 2014). Functionally *DDX53*, which is also known as *CAGE* and *CT26*, encodes the CAGE protein that has been shown to increase cellular proliferation by enhancing cyclin-dependant kinase activity (Por *et al*., 2010) and also contributes to drug resistance via a negative impact on p53 levels in hepatocellular and melanoma cell lines (Kim *et al*., 2010). Thus, whenever the gene is expressed, it is potentially clinically important and may be a useful predictive biomarker in combination with other cancer-associated genes.

TTK has previously been described as a CT gene that is able to elicit a T-cell response in patients with squamous cell carcinoma of the oesophagus and was subsequently used as part of a cancer vaccine targeting multiple CTAs (Kono *et al*., 2012; Mizukami *et al*., 2008). Higher levels of *TTK* expression have also been linked with recurrent prostate cancer (Shiraishi *et al*., 2011). The TTK antigen has been one of multiple targets of novel cancer vaccine therapies, evaluated in early phase clinical trials in lung cancer and cholangiocarcinoma (Aruga *et al*., 2013; Suzuki *et al*., 2013). The gene was expressed in bone marrow, stomach and normal ovarian tissue as well as being widely expressed in NAT colonic tissue in this study. Using our stringent screening approach, we would have excluded this gene from further analysis on this basis. Higher levels of expression for this gene were seen in the colon cancer samples, as well as in melanoma, lung and ovarian cancer (see Figure 5.18, group E). Given that this CTA has been used as a target for immunotherapy without serious adverse side effects, implies that low levels of expression of this gene produce such low amounts of protein in these tissues that it is not clinically relevant. It also highlights the complexity of selecting targets for disease stratification and/or treatment in cancer.

SPAG9 is considered a CT gene (Almeida *et al*., 2009), though neither the McFarlane group nor the Knochbin group identified this gene as spermatocyte- or germline-associated in the list compiled independently (Feichtinger *et al*., 2012b; Rousseaux *et al*., 2013b). It has been highlighted as a possible cancer biomarker and/or therapeutic target in many cancer types and implicated in promotion of tumour proliferation and invasion (Jilg *et al*., 2014; Mirandola *et al*., 2011; Yi *et al*., 2013). It has been shown here that *SPAG9* was widely expressed in both normal and cancerous tissues, although there was a more consistently high expression in the cancer samples (see Figure 5.18, group E). However, the high levels of expression in the normal tissues is at odds with what is known about this gene and it could suggest a lack of specificity with the primers for this target. The assay I chose for this gene was exon-spanning so this should obviate the possibility of genomic contamination but it would be important to repeat this using a different assay/primer set. The alternative conclusion is that *SPAG9* would not be a suitable candidate to take forward using a more focused list of genes using TLDA cards (i.e. looking for an "on-off" signal). It has been shown recently that a fusion transcript occurs between *SPAG9* and *JAK2* in an aggressive form of acute lymphoblastic leukaemia (Kawamura *et al*., 2015). It is possible that previous studies have detected *SPAG9* as a fusion gene, which is why it was absent from our and the Knochbin groups expanded gene lists (see Appendix B, on CD for full lists).

TPTE is a well characterised CT gene that has been implicated as playing a role in several cancer types (Almeida *et al*., 2009; Dong *et al*., 2003; Simon *et al*., 2014). It was first described as a CT gene not long after its initial discovery where it was shown to share significant homology to the tumour suppressor gene *PTEN* and gained its alias name as *PTEN2* (Chen *et al*., 1999; Dong *et al*., 2003; Wu *et al*., 2001). More recently several epitopes for this gene have been identified that may be important in targeting the TAA that is encoded by this gene (Simon *et al*., 2014). *TPTE* was not identified as a candidate meiCT gene but Rousseaux and colleagues found that the expression of this gene along with 25 other germline genes predicted a worse outcome in lung cancer (Rousseaux *et al*., 2013b). We found the gene to be expressed in a melanoma and ovarian cancer sample and as expected the gene expression was restricted to testis in normal tissue (e.g., see Figure 5.18, group EF). The gene was not expressed in any of the colon cancer samples tested. The expression of this gene in ovarian cancer and melanoma appears to be a novel finding as no reference in the English language linking the gene to these cancer types has been identified. Mutations in *TPTE* have been shown to permit phosphatase activity and impact on cell migration (Leslie *et al*., 2007). Clarification of a possible role for this gene as an oncogenic driver or predictive biomarker should be investigated further across a range of cancer types.

Other tissue-specific genes identified by the Knochbin group

We chose to include *STK31* as it was present on both our extended list of spermatocyte-associated genes and the list of germline tissue-restricted genes that Rousseaux and colleagues compiled (Feichtinger *et al*., 2012; Rousseaux *et al*., 2013b). In normal tissues the gene was shown to have testis/CNS-selective expression, being expressed in ovarian and normal adjacent colonic cancer tissue types only. In the cancer samples, the gene appeared to have higher expression in four of the colon cancer samples and in addition had moderate expression in both lung cancer samples and an ovarian cancer sample. Although not included on CTdatabase (Almeida *et al*., 2009) it was proposed as a novel CT gene implicated in gastrointestinal tract cancer (Yokoe *et al*., 2008). It was subsequently shown that overexpression of *STK31* in a duodenal cancer cell line promoted invasive abilities but not proliferation and conversely knockdown led to apoptosis (Kuo *et al*., 2014). STK31 localised with the centrosomes of this cell line and it was concluded that the protein may be involved in cell cycle regulation through an interaction with the spindle assembly checkpoint mechanism

(Kuo *et al*., 2014). Through analysis of driver mutations in melanoma, *STK31* has also been identified as a possible target for novel therapies in this cancer type (Xia *et al*., 2014). We found the gene to be expressed in melanoma also. The expression of the gene widely in normal adjacent colon cancer tissues may limit the use of the protein as a target for immunotherapeutic approaches. Expression was not demonstrated in the commercially obtained normal colonic sample used, so it is possible that the gene activation results from field cancerisation within the colonic mucosa. Further analysis is required to establish to what extent there is expression in normal tissues as my results indicate there may be higher expression than other groups have reported (Yokoe *et al*., 2008).

We identified *TKTL2* as an orthologue of a mouse spermatocyte-associated gene but did not validate it as a candidate meiCT gene (Feichtinger *et al*., 2012b). Rousseaux and colleagues also identified this gene as being testis-specific and moreover, it was included in their final list of 26 genes that predicted an unfavourable outcome in lung cancer patients (Rousseaux *et al*., 2013b). On the TLDA analysis the gene had a testis/CNS-restricted expression pattern and in addition one of the colon cancer samples expressed this gene (see Figure 5.17, group CF). The Knochbin group study was the first to link this gene to cancer (Rousseaux *et al*., 2013b). A related gene, *TKTL1*, has been linked to various cancer types, including cervical and nasopharyngeal cancer where it promoted proliferation of cells *in vitro* and it was additionally correlated with disease progression in nasopharyngeal cancer (Chen *et al*., 2009; Zhang *et al*., 2008). These studies did not find *TKTL2* to be upregulated. *TKTL1*, which is not testis-specific, has also been linked to adverse clinical outcomes in lung and CRC (Diaz-Moralli *et al*., 2011; Kayser *et al*., 2011). A plausible mechanism for the gene's association with a worse outcome in solid tumours is through its effect on glucose metabolism and energy supply to tumour cells within a hypoxic tumour microenvironment (Bentz *et al*., 2013). Given that *TKTL2* is a member of a group of genes that can independently predict a poor outcome in lung cancer and the gene appears to be expressed occasionally in colon cancer, further work should be conducted to both establish a function for this gene and assess whether it could form part of a predictive panel of genes across a wider range of cancer types.

TUBA3C has been implicated as playing a pivotal role in sperm motility as well as autosomal dominant genetic syndromes (Mikelsaar *et al*., 2012). There had been no report of a link to cancer in the medical literature for this gene until Knochbin and colleagues identified this gene as a member of a predictive cohort of genes in lung cancer (Rousseaux *et al*., 2013b).

Other members of the TUBA family were identified as orthologues of mouse spermatocyte-associated genes but these were not validated as meiCT genes (Feichtinger *et al*., 2012b). In this study we found the gene to be expressed in one of the lung cancer samples, which is consistent with the Rousseaux paper (see Figure 5.22, group F). There was also very weak/borderline expression in all the melanoma samples tested (2 out of the 4 samples were not tested for this gene) providing the first evidence that this gene may be a biomarker for this cancer type also but this requires clarification through further research to confirm definite expression and clinical relevance. None of the colon cancer samples expressed this gene and it was restricted to the testis in the normal tissues that were tested.

VCY has been implicated in ribosomal assembly during spermatogenesis and was another gene identified as a predictor of poor outcome in lung cancer (Rousseaux *et al*., 2013b; Zou *et al*., 2003). It has since been confirmed as a CTA with varying frequency in different lung cancer types (Taguchi *et al*., 2014). It has not otherwise been described as a CT gene. It had a testis-restricted expression profile in normal tissues and appeared to be expressed strongly in melanoma only within the cancer samples we tested (see Figure 5.22, group F). Thus, it would be interesting to investigate the expression and function of this gene further in melanoma.

Section 2 – Primers used for RT-PCR

Table 1. List of primers used for RT-PCR, annealing temperatures and PCR cycling conditions

Gene/primer name	Sequence (5' – 3')	Length	CG %	Tm	Size	PCR cycle
ACTB - F	CAAACATGATCTGGGTCATCT	**21**	**42.9**	**55.6**	**353**	**E**
ACTB – R	GCTCGTCGTCGACAACGGCTC	**21**	**66.7**	**66.0**		
ACTL9 - F	CAGTCGGTGCTGTCTGTCTA	**20**	**55**	**59.1**	**488**	**E**
ACTL9 - R	CCGCAGAGAAGCACGTTTTG	**20**	**55**	**60.4**		*****
ADAM2 – F	CTGTGCTGAAGGACCATGCT	**20**	**55**	**60.3**	**610**	**B**
ADAM2 – R	GCCACTGTCAATACTCCCAC	**20**	**55**	**58.3**		
ASB17 – F	GGGTCAGTGGGGATATCACT	**20**	**55**	**58.2**	**327**	**B**
ASB17 – R	GGGCTTGGACAGTATACTGG	**20**	**55**	**57.4**		
BOLL – F	ATGACAGAGCTGGAGTATCC	**20**	**50**	**56.1**	**657**	**A**
BOLL – R	TGGCACTTGGAGCATAAACC	**20**	**50**	**58.5**		
BTG4 – F	GCCTCATCAGACGTTTCCTC	**20**	**55**	**58.4**	**125**	**F**
BTG4 – R	AAGGGCTGTTTCAAGTTTTCAA	**22**	**36.4**	**57.2**		
C1orf141 – F1	CCCGTTGTGCTTTGAGGATG	**20**	**55**	**59.5**	**565**	**I**
C1orf141 – R1	TGTGGGTATGCTCGTTGGTT	**20**	**50**	**59.6**		
C1orf141 – F2	ACATCCGCGTCTAAGGCAAT	**20**	**50**	**59.8**	**367**	**B**
C1orf141 – R2	CAAAGCACAACGGGAGCAAG	**20**	**55**	**60.3**		
C10orf67 – F1	GCTTTTCCTCCTCCTTGAGG	**20**	**55**	**57.9**	**293**	**E**
C10orf67 – R1	CCCAAAATCTACCTGGAGGG	**20**	**55**	**57.3**		
C10orf82 – F	CTGCCAAGGAATGTCCAAGGA	**21**	**52.4**	**60.3**	**134**	**H**
C10orf82 – R	TGACAGGTTTCAGTTTCGGGG	**21**	**52.4**	**60.5**		
C12orf37 – F	ATGGGATGCTTTGATTGCTC	**20**	**45**	**56.4**	**143**	**F**
C12orf37 – R	AGAGATGTGAGGCTGGCAGT	**20**	**55**	**60.9**		
C12orf50 – F	ATGCTTCTGGGAAACTCAGC	**20**	**50**	**58.2**	**539**	**A**
C12orf50 – R	CAGTCTTTGGTTTCCCATGC	**20**	**50**	**57.3**		
C16orf78 – F	CAGGGGAAGAAGAAACAAGC	**20**	**50**	**56.6**	**405**	**A**
C16orf78 – R	GTCTCTTATGAAGGTTGCCC	**20**	**50**	**55.8**		
C2orf61 – F2	GGGCAATACAACGTGCTTCC	**20**	**55**	**59.8**	**204**	**C**
C2orf61 – R2	ACAGCTGGGCAAGAATCGAG	**20**	**55**	**60.4**		
C3orf22 - F	CAGTGCCTGCAAGAAGTCTC	**20**	**55**	**58.8**	**294**	**E**

C3orf22 – R	CAAAGGTTGCACAGTGGAGG	**20**	**55**	**59.3**		
C3orf30 – F	GACCAGACTGACCACTTAGC	**20**	**55**	**57.6**	**563**	**B**
C3orf30 – R	AAATTCAGGGGGTCCTCTGG	**20**	**55**	**59**		
C6orf81 – F	GCAAACAGGATGAGTATGCC	**20**	**50**	**56.9**	**603**	**A**
C6orf81 – R	GCAGGGACTGTTCGTTGTAA	**20**	**50**	**58.1**		
C8orf74 - F	CCAGGTGGTCAAGTTCACAG	**20**	**55**	**58.4**	**551**	**E**
C8orf74 - R	GGTCCAAGATGGCGAATGTG	**20**	**55**	**59.3**		
C9orf153 – F1	GCCACCCTTCCTCAATGTTCA	**21**	**52.4**	**60.6**	**711**	**C**
C9orf153 – R1	TGGAGATGGGGTCTTTCTATGTTTT	**25**	**40**	**60**		
C9orf153 – F2	ATGTTCCTCACTGGAGACACC	**21**	**52.4**	**59.4**	**273**	**C**
C9orf153 – R2	ATGTTTTCCAGGCTTGTTCATT	**22**	**36.4**	**57.0**		
CAPZA3 - F	GTACCACTCTGCATCGATGG	**20**	**55**	**58.1**	**645**	**E**
CAPZA3 - R	GTCAGAGAGTATCCTGTGCC	**20**	**55**	**57.1**		*****
CCDC116 – F1	TGTGAGGGACAAACTCCTGC	**20**	**55**	**59.9**	**721**	**B**
CCDC116 – R1	GTGAGACATGCTGGACTTCG	**20**	**55**	**58.7**		
CCDC116 – F2	GCTTTCAGTGGACACAGGAG	**20**	**55**	**58.5**	**555**	**B**
CCDC116 – R2	GCGGCTTCTTCTTCGTGAAC	**20**	**55**	**59.8**		
CCDC63 – F	GGACACCTCTCAGTACAACC	**20**	**55**	**57.3**	**614**	**E**
CCDC63 – R	GCGGAAGGTTGATGTCAGTG	**20**	**55**	**59.2**		
CCDC73 – F	TGAGCTGCAAAGGGAGAAGG	**20**	**55**	**60**	**881**	**C**
CCDC73 – R	GTTCTCGAGTCCCTGGAAGC	**20**	**60**	**60.1**		
CCDC83 – F	AATCCTCGTCATCTGCTGCT	**20**	**50**	**59.2**	**90**	**F**
CCDC83 – R	TCCAGCAGCTTGGGTAAGATA	**21**	**47.6**	**58.5**		
CCIN – F	CAGCGCACTCATCAATTGGG	**20**	**55**	**59.9**	**369**	**C**
CCIN – R	CTAAGTGCTGCTGCCCGATA	**20**	**55**	**59.9**		*****
CPA5 – F	CAGCACCAACAGCTTCAGTT	**20**	**50**	**59**	**76**	**G**
CPA5 – R	CATTACAAAGTTGTCAATCCAGCTA	**25**	**36**	**57.9**		
DDI1 – F	GAGCAGCAAAGGGAAAAGGC	**20**	**55**	**60**	**591**	**A**
DDI1 – R	GTCCGAAGACTCATCTTGCC	**20**	**55**	**58.4**		*****
DNAJC5G – F	GATGAGCCTCTATGCAGTGC	**20**	**55**	**58.5**	**464**	**B**
DNAJC5G – R	GCTGACTCTGGACATTCTGC	**20**	**55**	**58.6**		
DPEP3 – F	CCTGGACAGGCTTAGAGACG	**20**	**60**	**59.5**	**72**	**H**
DPEP3 – R	TCCTGGGACTGGCATGAG	**18**	**61.1**	**58.3**		
DYDC1 - F	CCTTGGGGCCTGTTTAACTC	**20**	**55**	**58.5**	**444**	**E**
DYDC1 - R	CTGCTCAAGTTAGGTGCTCC	**20**	**55**	**58.3**		

EB13 – F	CCTGCAGTGGAAGGAAAGG	**19**	**57.9**	**58.1**	**100**	**G**
EB13 – R	AGGGTCCAGGAGCAATCC	**18**	**61.1**	**58.3**		
FAM170A – F1	CGTTTCTTCTTCACGCAAGC	**20**	**50**	**58.3**	**305**	**A**
FAM170A – R1	TCACTTCTTCCATCCTTGGC	**20**	**50**	**57.5**		
FAM170A – F2	TGAAGAGTCCCAGGAAACCG	**20**	**55**	**59.3**	**305**	**B**
FAM170A – R2	CTTCTTCCATCCTTGGCTGC	**20**	**55**	**58.9**	**+ 446**	
FAM194A - F	GCTGAACCTGAATGTCTGGC	**20**	**55**	**59.2**	**413**	**B**
FAM194A - R	GCTGCATGAGGGTCAATAGC	**20**	**55**	**59.1**		
FAM71B - F	GCAGTTGTGTCCCTCTTCTG	**20**	**55**	**58.5**	**537**	**E**
FAM71B - R	GCCATGGACTTGCTGGATTC	**20**	**55**	**59.3**		
FIGLA – F	CTCAACCGTGGTTTTGCCAG	**20**	**55**	**60**	**371**	**C**
FIGLA – R	TCCAGACTTCTGGTTGGGGA	**20**	**55**	**60.1**		
FLJ43944 – F	ACTGTAATGCTCGGCAAGGT	**20**	**50**	**59.7**	**X**	**G**
FLJ43944 – R	CTGAGTTTTCTGGGGAGGTGG	**21**	**57.1**	**60.3**		
GK2 – F	TATTGGGAGCCCAGTGCAAG	**20**	**55**	**60**	**309**	**C**
GK2 – R	GCTCCAAACGCTTACTCCCT	**20**	**55**	**60**		*****
H2AFB1 – F	CGAGCGGAGCTTTCGTTTTC	**20**	**55**	**60.2**	**264**	**D**
H2AFB1 – R	CAGGGCCACTTGAGAGATGG	**20**	**60**	**60.1**		*****
HEATR7B1 - F	CAGCAGCCAGAATCTGATGG	**20**	**55**	**58.7**	**633**	**E**
HEATR7B1 - R	CAGGAACTTTGCCCACCTTG	**20**	**55**	**59.3**		
HIST1H3C – F	GTCTAGTAATGAACCAATCAGTCTGG	**26**	**42.3**	**59.2**	**X**	**F**
HIST1H3C – R	AGTCCCGCGCGATAAAAT	**18**	**50**	**57.2**		*****
HMGB4 – F	CCAGCTAAAGCCTAAGGCAA	**20**	**50**	**57.9**	**449**	**A**
HMGB4 – R	CGAAGTACTTAGCTCTCAGG	**20**	**50**	**54.8**		*****
IQCF1 - F	GAGCAGAGTCAAAGGCAGAG	**20**	**55**	**58.3**	**481**	**E**
IQCF1 – R	GAGTCCAGCAAGATCTCCAG	**20**	**55**	**57.4**		
ISM2 – F	ACTCGGCCCTGTGGCTAT	**18**	**61.1**	**60.4**	**87**	**H**
ISM2 – R	GGTGTCCTTGTCCTCAGTGC	**20**	**60**	**60.6**		
KIAA1257 – F	GACGATTCTTGGACGATTCAA	**21**	**42.9**	**56.4**	**86**	**F**
KIAA1257 – R	AATTCCTTGATCTCCATCATGC	**22**	**40.9**	**56.5**		
KLF17 – F1	CTCGACAGTACCTTCTGACG	**20**	**55**	**57.5**	**461**	**E**
KLF17 – R1	CCCGCATATGTCGTCTAAGC	**20**	**55**	**58.6**		
KLF17 – F2	AGTAGAGAAGAACTCCAGGC	**20**	**50**	**56**	**314**	**B**

KLF17 – R2	GATGAGTCTTCTGGTGTTGC	**20**	**50**	**56.5**		
LGALS14 – F	TCACTTTTGTCAAGGACCCAC	**21**	**47.6**	**58.4**	**211**	**H**
LGALS14 – R	TTGTGACGCACATAGATGCAC	**21**	**47.6**	**59.3**		
LOC441601 – F	GCCAAAGCAGACACACTCAC	**20**	**55**	**59.7**	**X**	**I**
LOC441601 – R	ACAATGTTCACCATTTTCAACTGT	**24**	**33.3**	**58**		
LYZL6 – F	GGCGCTACTCATCTATTTGG	**20**	**50**	**56.1**	**348**	**A**
LYZL6 – R	CCGGACACAATCCTTTTTGC	**20**	**50**	**57.9**		
MAGEB6 – F	TGAGCCTGAAAAGTGCTGTC	**20**	**50**	**58.4**	**93**	**H**
MAGEB6 – R	TGTGGGTAGGAAGACTAGGAAGG	**23**	**52.2**	**60.6**		
NBPF4 – F	CATGCTGAGAGCGCAGAG	**18**	**61.1**	**58.3**	**62**	**F**
NBPF4 – R	TTTCATCCTGCCATCCTTTG	**20**	**45**	**56**		
NY-ESO-1 – F	CTGAATGGATGCTGCAGATGC	**21**	**52.4**	**60.0**	**279**	**E**
NY-ESO-1 – R	CTGCGTGATCCACATCAACAG	**21**	**52.4**	**59.6**		
OR7E156P – F	TGAGAATTCTCTTGCCTGCC	**20**	**50**	**57.9**	**X**	**G**
OR7E156P – R	GAGAAAAGGGTCCCAGAGAT	**20**	**50**	**56.2**		
PDHA2 – F1	CGAGTTGCCCAGAAATCAGC	**20**	**55**	**59.6**	**374**	**B**
PDHA2 – R1	AGCTCTGCGAGAATGGATCG	**20**	**55**	**60**		*****
PDHA2 – F2	CTCACCTCAGCCCCAATTTC	**20**	**55**	**57.2**	**679**	**B**
PDHA2 – R2	CCTATAGGGCTCATGGTGTG	**20**	**55**	**58.5**		*****
PDILT – F	GACGAACCCAGAAATGGACG	**20**	**55**	**58.9**	**404**	**B**
PDILT – R	GCTGTGACATCGATCTTGGC	**20**	**55**	**59.4**		
PIWIL1 – F	AGAGGTTACCAGACCAGAATGG	**22**	**50**	**59.2**	**77**	**J**
PIWIL1 – R	GTGTGGGAGAAACACTACCACTT	**23**	**47.8**	**60.4**		
PPP3R2 – F1	GGGCAGGAGGTTTAAGAAGT	**20**	**50**	**57.1**	**401**	**A**
PPP3R2 – R1	CCACAGCACTGAATTCCTCA	**20**	**50**	**57.8**		*****
PPP3R2 – F2	GTCCACAATGGGAAACGAGG	**20**	**55**	**58.8**	**509**	**B**
PPP3R2 – R2	GGACCAGCTTCTTGTGGATC	**20**	**55**	**58.3**		*****
PRPS1L1 – F	GACGGAGTGCCTCTGGTAG	**19**	**63.2**	**59.2**	**728**	**C**
PRPS1L1 – R	CAGCTGCGAGGCAGATTGTA	**20**	**55**	**60.5**		*****
RBM44 – F	CAGATGTGTCGTCGCCATTG	**20**	**55**	**59.6**	**548**	**E**
RBM44 – R	GAGACGTCAACTCCTGTCAG	**20**	**55**	**57.7**		
RBM46 – F	GGAAATTTGGCGGTCCTC	**18**	**55.6**	**56.1**	**70**	**G**
RBM46 – R	CCTACAAAAACTTCACAGCCTCT	**23**	**43.5**	**58.9**		
RFX4 – F	CATCACCAAGCAAACCCTTT	**20**	**45**	**56.8**	**117**	**G**

RFX4 – R	GACTCGATGGGAGACTGCTC	**20**	**60**	**59.6**		
RNF133 – F	GGTGGGGAGAAAGCACATCA	**20**	**55**	**60**	**347**	**C**
RNF133 – R	GGCATGTCCCATGGGGTAAA	**20**	**55**	**60**		*****
ROPN1 – F	CCAAAGTGGATGGGGAGAT	**19**	**52.6**	**56.4**	**89**	**I**
ROPN1 – R	GATTATACCATCAGGGCCAATTA	**23**	**39.1**	**55.9**		
SATL1 – F	GCTGGAGACTGCCCAGAAAT	**20**	**55**	**60**	**228**	**C**
SATL1 – R	CTTGCCAGTCCATGAGTCGT	**20**	**55**	**60**		
SHCBP1L – F	CGTGTCGAGCTCATTGAGTA	**20**	**50**	**57.5**	**471**	**A**
SHCBP1L – R	GTCAGCTTTGGACACCACAA	**20**	**50**	**58.6**		
SLC25A31 – F	GGTACAAAGGCATGGTGGAC	**20**	**55**	**58.8**	**422**	**B**
SLC25A31 – R	CCCTGTACCGCGAAGAACAT	**20**	**55**	**60.1**	**+ 695**	
SPDYA – F	AAGTTAAGGGACCAGCTCTG	**20**	**50**	**56.8**	**293**	**A**
SPDYA – R	TGTTTTTCTGTCACCCCTGC	**20**	**50**	**58.6**		
SPO11 – F1	TGGTAACCAGACTGTCGTCG	**20**	**55**	**59.4**	**331**	**E**
SPO11 – R1	ATCAGGAACTCCCTTTCCCG	**20**	**55**	**59.1**		
SPO11 – F2	ACGGGAAAGGGAGTTCCTGA	**20**	**55**	**60.5**	**324**	**J**
SPO11 – R2	TCTCCAAAATGGTTGGCAGGT	**21**	**47.6**	**60.1**		
SPZ1 – F	CTGCTAAGTCAGCTGAGATG	**20**	**50**	**55.7**	**403**	**A**
SPZ1 – R	TTGTCTTCTTTCTCTGGGGC	**20**	**50**	**57.4**		*****
TBC1D21 – F	CCCTCTCTCCTGAAAACAGC	**20**	**55**	**57.9**	**419**	**B**
TBC1D21 – R	CTTGTCGATGAGGACGTTGC	**20**	**55**	**59.3**		
TGIF2LX – F1	ACCAGAGCACAAGAAGAAGC	**20**	**50**	**58.1**	**578**	**A**
TGIF2LX – R1	GCTCTTGCTTCTTCTCTAGC	**20**	**50**	**56**		*****
TGIF2LX – F2	GAGTCCGTTAAGATCCTCCG	**20**	**55**	**57.2**	**532**	**B**
TGIF2LX – R2	GCTTCTTCTCTAGCTCCAGC	**20**	**55**	**57.8**		*****
TKTL2 – F	ACGACCGGTTCATCCTCTC	**19**	**57.9**	**58.8**	**61**	**F**
TKTL2 – R	TCCACCCAAGCAGCATAGA	**19**	**52.6**	**58.3**		
TMC7 – F1	GCAAACCATCTGCTGGATCG	**20**	**55**	**59.6**	**331**	**B**
TMC7 – R1	ATGGATGAAGGACTGCAGCG	**20**	**55**	**60.5**		
TMC7 – F2	GCGGTATAGCAGCAAGTCTT	**20**	**50**	**58.1**	**971**	**A**
TMC7 – R2	GGATTGTCAGACGGATCTCA	**20**	**50**	**56.7**		
TMC7 – F3	TCCTCCAAGAATTGCCAAGC	**20**	**50**	**58.5**	**396**	**A**
TMC7 – R3	CCAGGAATCTCAAGAAGGAG	**20**	**50**	**54.9**		
TMEM146 - F	GGAAATACTGACCCCACTGC	**20**	**55**	**58.3**	**485**	**E**
TMEM146 – R	GGTCGCAGCCAACTGAAATC	**20**	**55**	**59.8**		

TPTE – F	AGGCACCTGCGAAAGAAAG	**19**	**52.6**	**58.4**	**85**	**G**
TPTE – R	CGTGCTAACACACTTTCACTGAT	**23**	**43.5**	**59.5**		
TRIM42 – F	GAAGCTCCTCCATGTTGTCC	**20**	**55**	**58.3**	**524**	**B**
TRIM42 – R	CCACTTAGCTCGGTTCTTCC	**20**	**55**	**58**		
TSGA13 – F	CCTGCAGACTAGCAACCCAA	**20**	**55**	**60**	**379**	**B**
TSGA13 – R	TGCCCGACTGCATCAGAAAT	**20**	**50**	**60**		
TSSK2 – F	AAGAACCTGACCTGCGAGTG	**20**	**55**	**60**	**363**	**C**
TSSK2 – R	CTAGGTGCTTGCTTTCCCCA	**20**	**55**	**60**		*****
TUBA3C – F	CGGAGGAGCTCAACATGC	**18**	**61.1**	**58.2**	**82**	**J**
TUBA3C – R	AGTTCCCAGCAGGCATTG	**18**	**55.6**	**57.6**		
UBL4B – F	CACGCTGAAGAGACTGGTGT	**20**	**55**	**60**	**352**	**D**
UBL4B – R	CTCCTCTGCCAGGAGGTACT	**20**	**60**	**60**		*****
UBQLN3 - F	GGAGATGATACGTAGCCAGG	**20**	**55**	**57**	**573**	**E**
UBQLN3 – R	CCAGTACTGTTCTCTGTGGG	**20**	**55**	**57.3**		*****
UMODL1 - F	GTGAAAGGAAGGAGGACGAC	**20**	**55**	**57.9**	**563**	**E**
UMODL1 – R	CAGCAGGTCATTCTGGAAGG	**20**	**55**	**58.3**		
VCY – F	GGCCAAGGAGACAGGAAAG	**19**	**57.9**	**57.7**	**80**	**H**
VCY – R	CGGCCACCTTGGTAGTCTT	**19**	**57.9**	**59.3**		
YI72E07 – F	TGCTACTTGCTGTGTGACCTG	**21**	**52.4**	**60.5**	**X**	**H**
YI72E07 – R	ATGGTGTGACTTCCTCCTCAG	**21**	**52.4**	**59.1**		
ZSWIM2 - F	AAACACCTTGGGATTCCCTG	**20**	**50**	**57.7**	**466**	**A**
ZSWIM2 -R	GGCATGAATTGCACTTGTGG	**20**	**50**	**58.3**		

The letters in the right hand column refer to the PCR cycling conditions used – refer to Table 2.2.
X denotes that expected product size uncertain.
*In PCR cycle column indicates that primer set did not span an intron.

Section 3 – Additional Figures

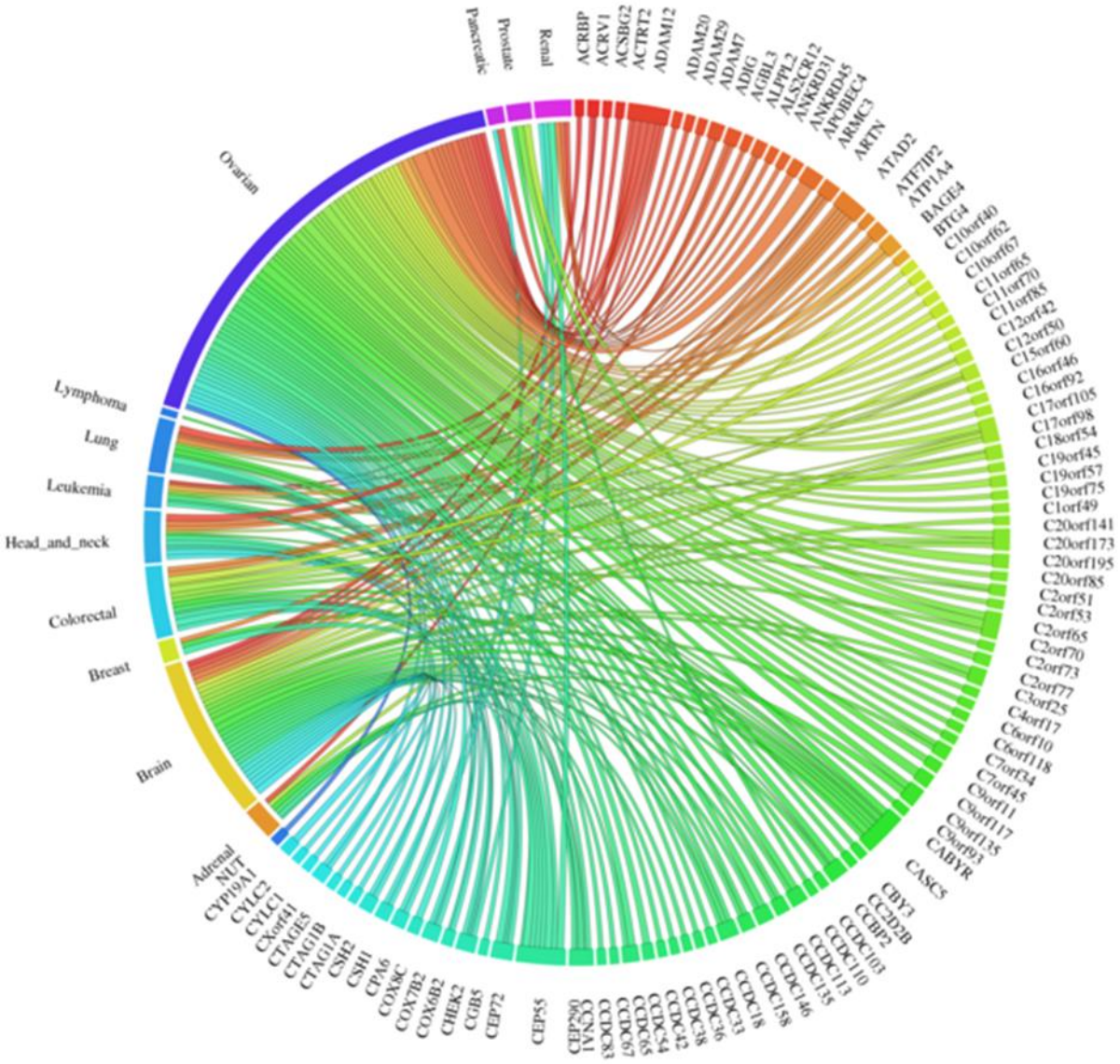


Figure 1. Germline genes displaying meta-change upregulation in cancer.

The definitive list of germline genes (comprising over 1000 genes) were assessed using CancerMA (Feichtinger *et al*., 2012a). The website is limited to ~300 targets on a single run, so the results for only a selection of the germline genes were provided in Figure 4.14. Here, genes beginning with the letter A through to C are included. Brain cancer and ovarian cancer feature most commonly as cancer types displaying a statistically significant upregulation in gene expression when the microarray datasets were meta-analysed, with CRC and lung cancer being the next most common.

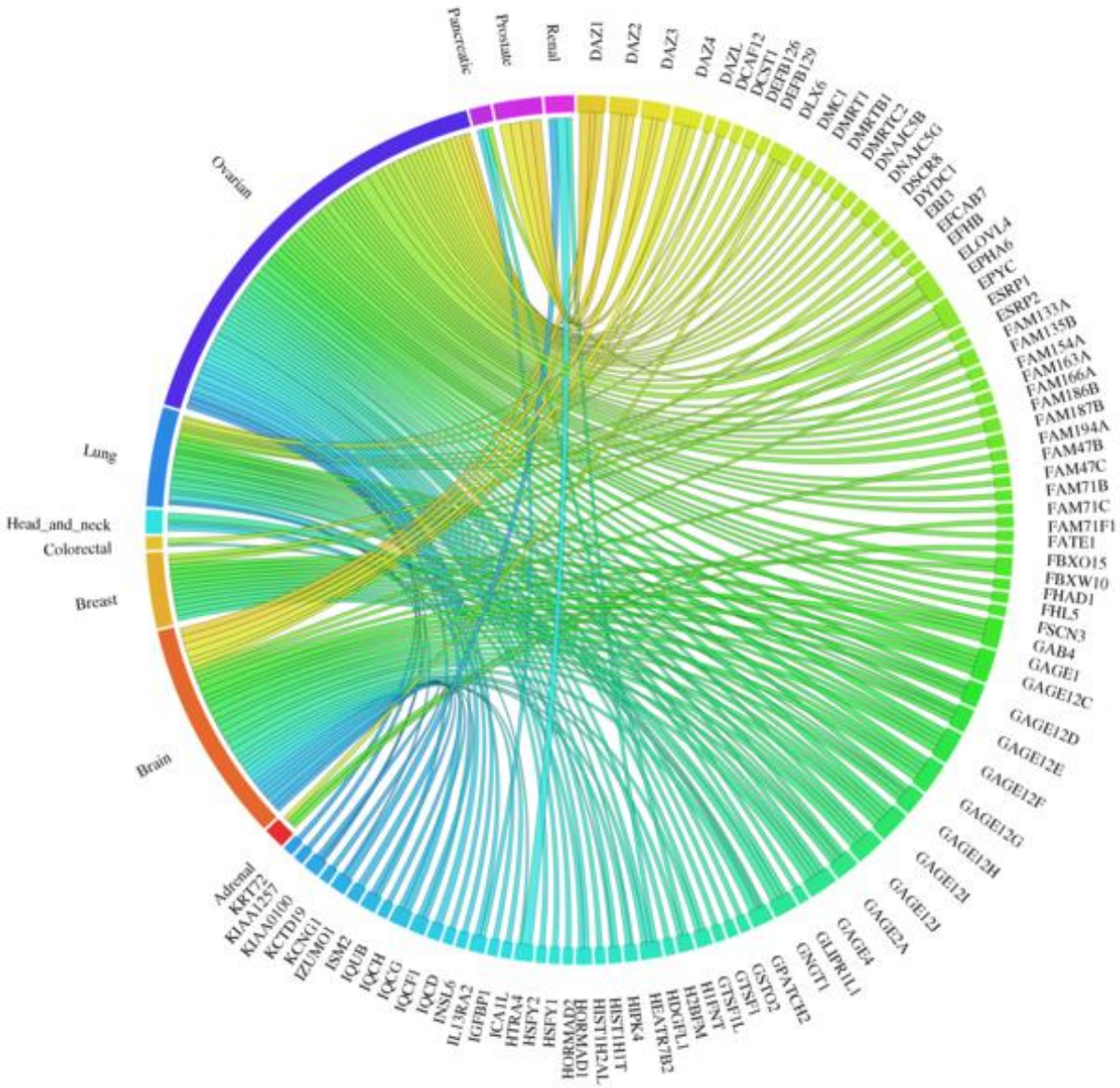


Figure 2. Germline genes displaying meta-change upregulation in cancer.

The definitive list of germline genes (comprising over 1000 genes) were assessed using CancerMA (Feichtinger *et al*., 2012a). The website is limited to ~300 targets on a single run, so the results for only a selection of the germline genes were provided in Figure 4.14. Here, genes beginning with the letter D through to K are included. Brain cancer and ovarian cancer were once again the most common cancer types displaying a significant upregulation in gene expression when the microarray datasets were meta-analysed, with breast cancer featuring more prominently in this cohort of genes.

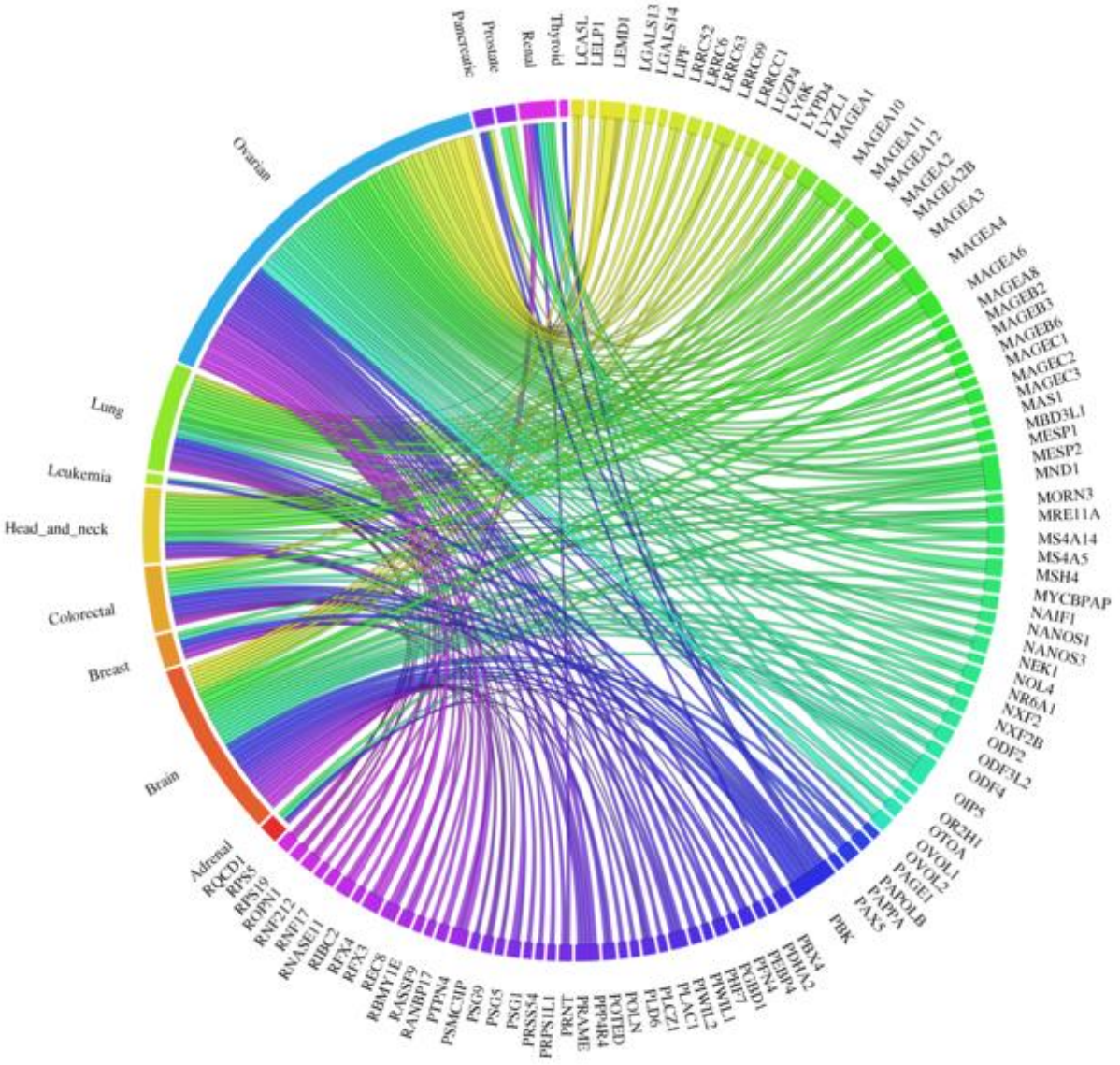


Figure 3. Germline genes displaying meta-change upregulation in cancer.

The final cohort of genes (not shown in Figures 1 and 2 above or Figure 5.12) is shown here, with names starting with the letter L through to R. Brain cancer and ovarian cancer were once again the most common cancer types displaying a significant upregulation in gene expression when the microarray datasets were meta-analysed but for this cohort head and neck cancer featured more commonly than in the other examples. The combined list of meta-upregulated genes provides one avenue on which to narrow down on a smaller cohort of genes for individual cancer types.

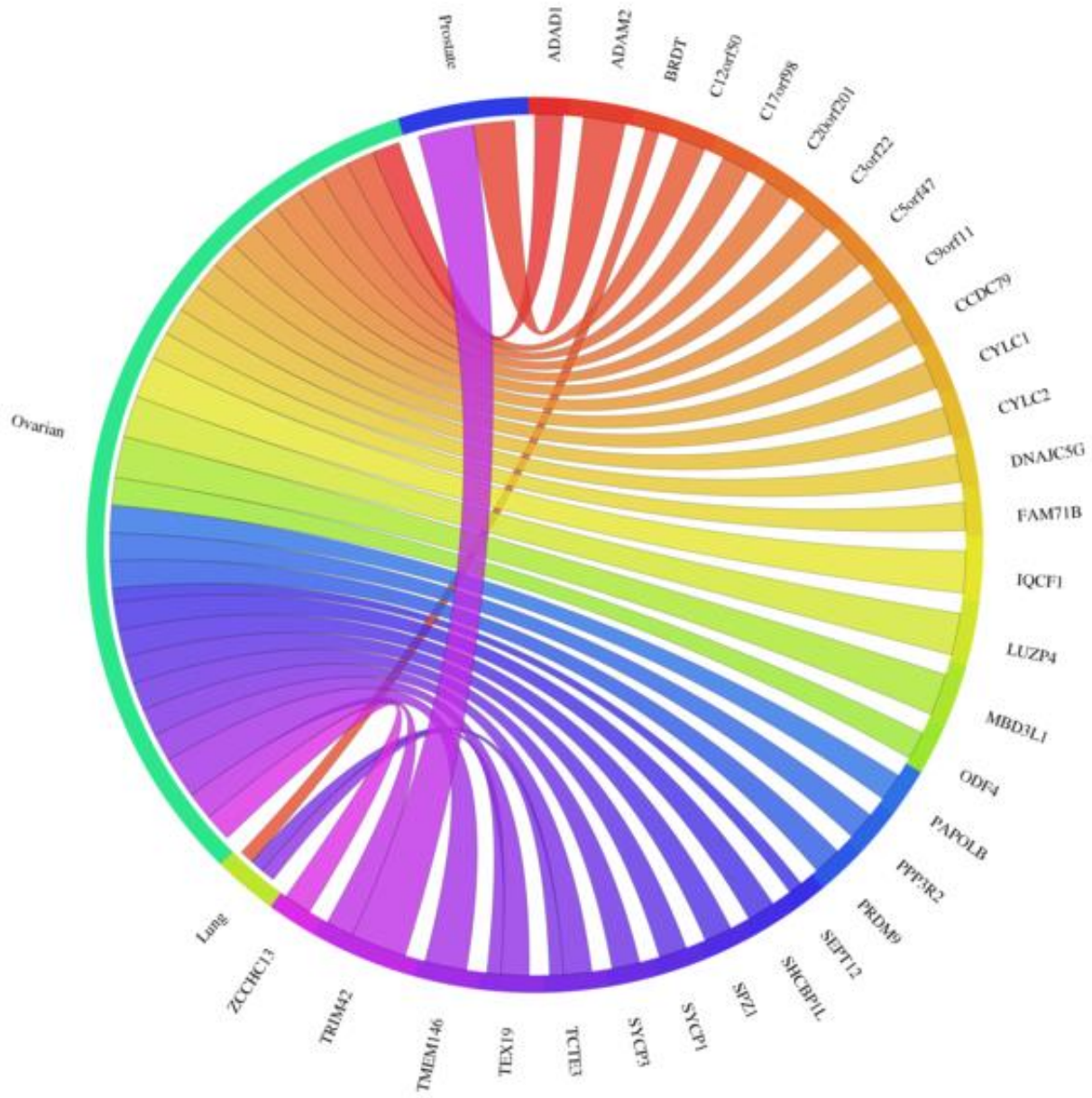


Figure 4. Circos plot displaying meta-upregulated meiCT genes in cancer.

When challenging the entire list of meiCT genes against CancerMA (Feichtinger *et al*., 2012a), ovarian cancer stood out as the predominant cancer type displaying meta-upregulations. The thickness of the line connecting individual genes to the associated cancer type in which they are upregulated corresponds to the degree of upregulation or strength of association. The reason for the prominence of ovarian cancer is unclear, as many ovarian tumours are believed to originate outside of the ovaries, although dormant meiosis-associated genes could conceivably be reactivated in cancer.

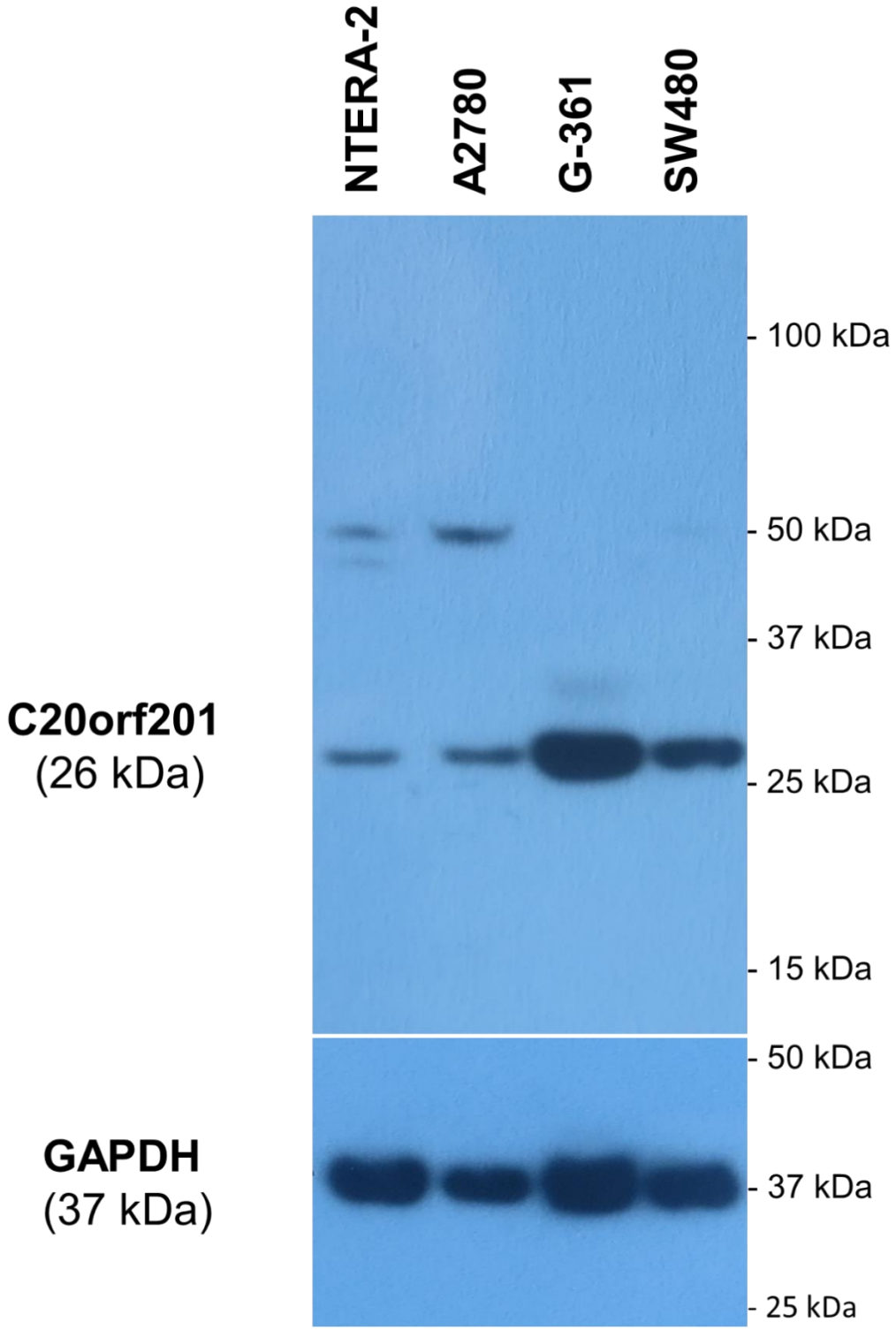


Figure 5. Western Blot indicating the presence of C20orf201 in cancer cell lines.

Western blot was performed on whole cell extracts from NTERA-2, A2780, G-361 and SW480 cells. Detection of C20orf201 was with the ab108142 antibody (Abcam; Lot. GR125175-1, 1:1000). All the cells line lysates produced a band of the expected size, and some in addition a fainter band at ~50 kDa. Secondary antibody: anti-rabbit IgG, HRP-linked (Cell Signaling, #7074; 1:3000 dilution). GAPDH is shown as a loading control.

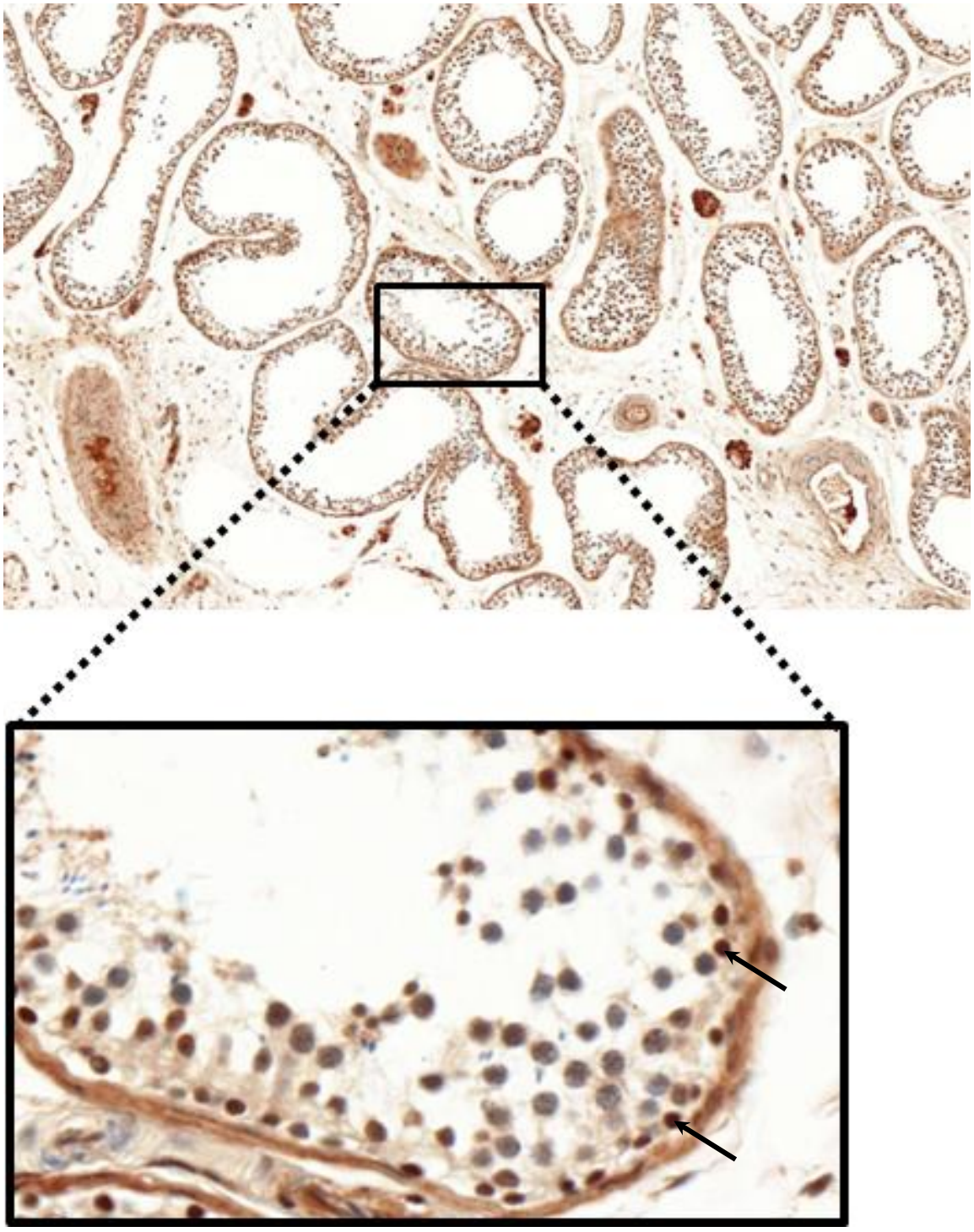


Figure 6. IHC staining of C20orf201 in normal testis.

IHC was performed on the semi-automated Ventana machine (Roche) using the rabbit polyclonal antibody against C20orf201 (Abcam, ab108142). Positive staining was seen in germ cells of various stages of development/differentiation with strongest positive staining seen in the spermatogonial cells (arrows). There was also positive staining seen in some cells surrounding the seminiferous tubules, indicating that the protein (or antigen detected by the antibody) was not restricted to germ cells of the testis. A few post-meiotic spermatid/sperm cells are visible with small nuclei that displayed negative staining. Images acquired on an AMG EVOS® xl Core microscope.

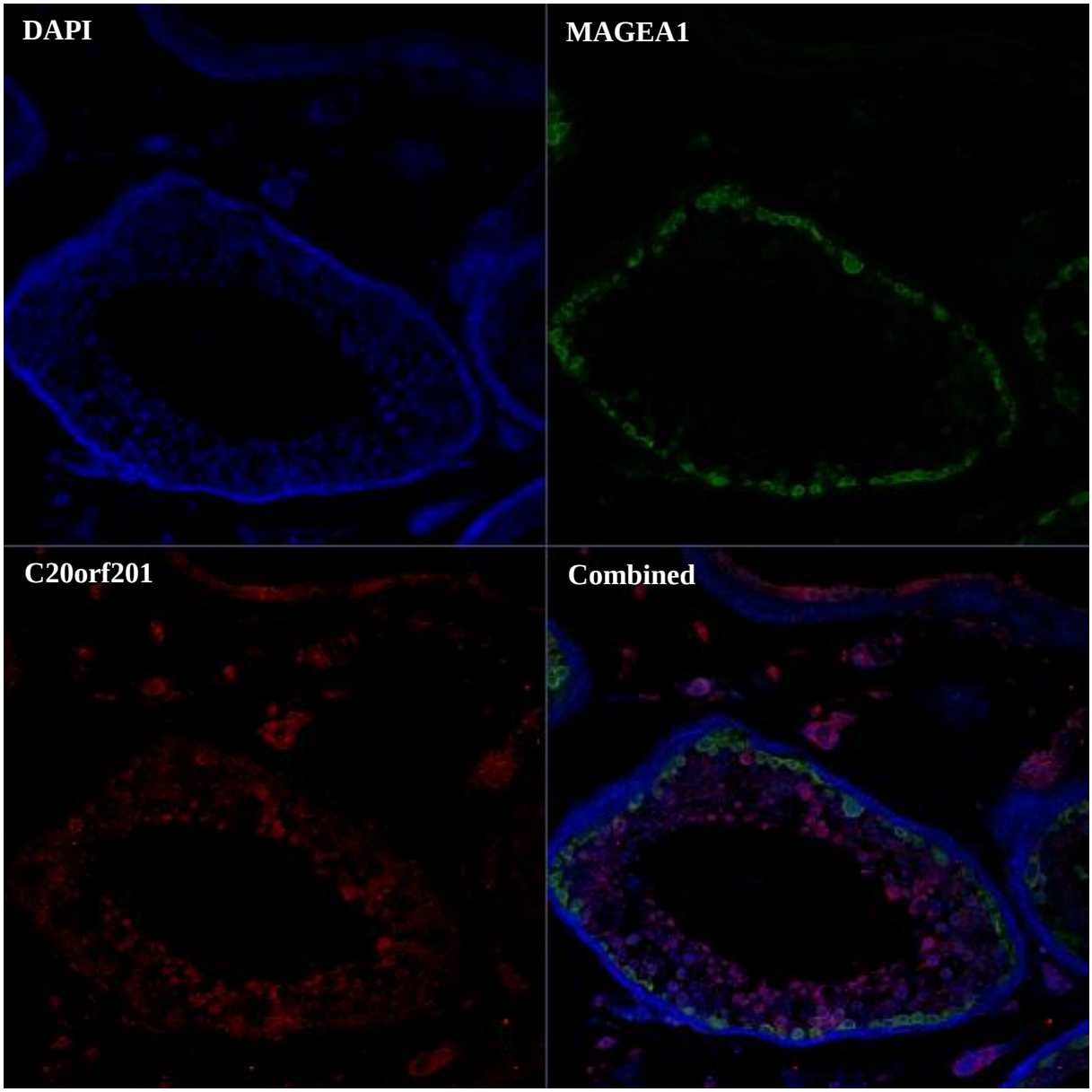


Figure 7. Immunofluorescence (IF) staining for C20orf201 and MAGEA1 in normal testis (x40).

IF was carried out on FFPE tissue blocks of normal testis showing MAGEA1 (green) staining detected by the mouse monoclonal antibody (LSBio, LS-C87868) localised to the spermatogonial layer of the seminiferous tubule. The staining for C20orf201 (red) was with the rabbit polyclonal antibody (Abcam; ab170783). Similar but less distinct staining pattern was obtained using the rabbit polyclonal antibody to C20orf201 (Abcam; ab108142) and the stronger staining seen on IHC in the spermatogonial layer could not be reproduced. C20orf201 stained developing spermatocytes more strongly and there appeared to be no clear co-localisation with MAGEA1. Images acquired on Zeiss LSM 710 confocal microscope.

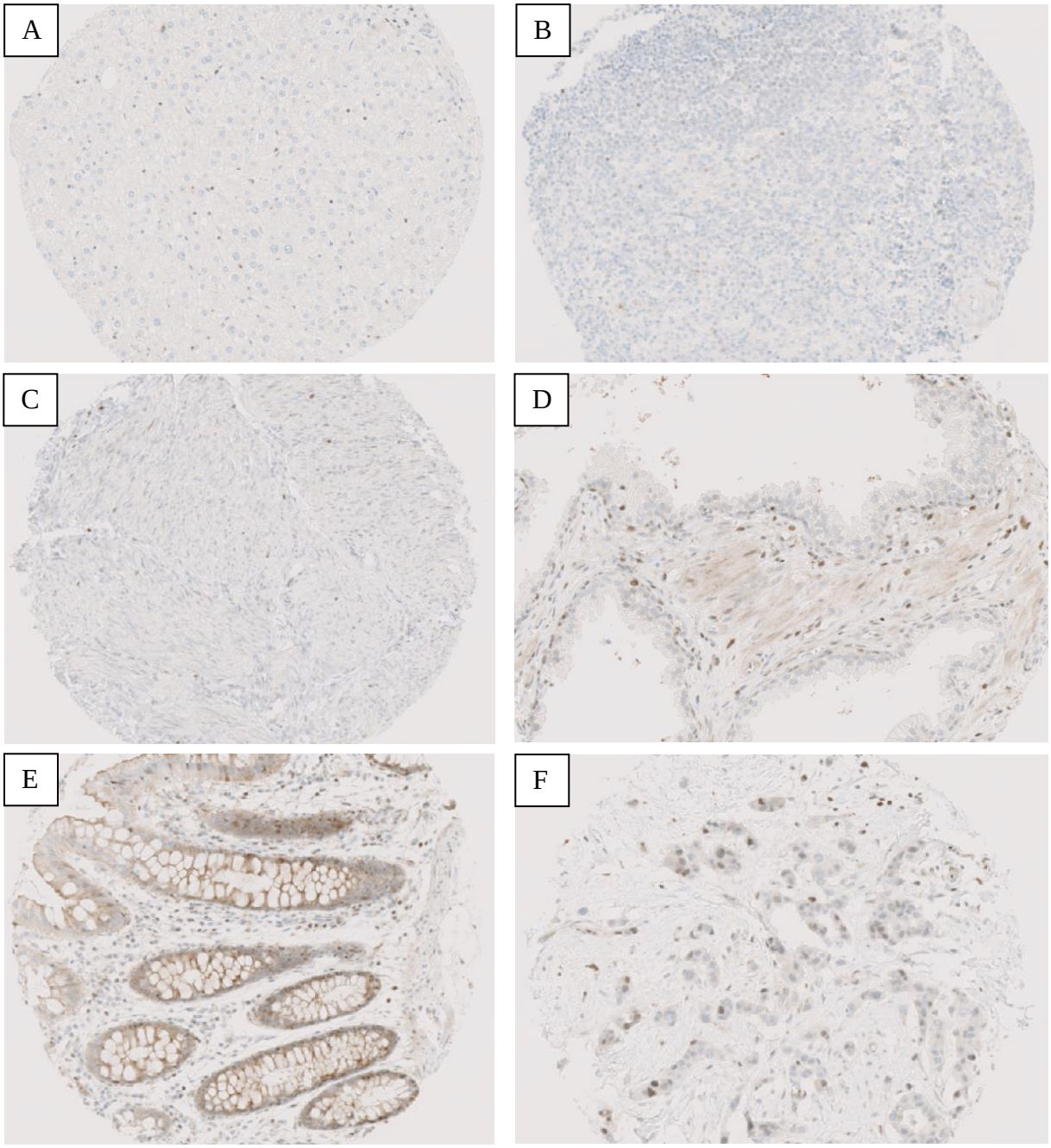


Figure 8. IHC staining of C20orf201 in human tissues.

Selected images taken from IHC performed on a TMA (CHTN_TEST2). IHC was performed on the semi-automated Ventana® machine (Roche) using the rabbit polyclonal antibody against C20orf201 (Abcam; ab108142, Lot. – GR125175-1). Negative staining was seen in normal liver, smooth muscle and spleen: images (A), (B) and (C), respectively. Positive nuclear staining was seen in a proportion of the stromal cells in normal prostate (D) but the prostate epithelium displayed negative staining. There was weak positive staining in normal colon (E) that was largely cytoplasmic (the example shown here displayed stronger staining than what was generally seen). There was also patchy positive nuclear staining seen in breast cancer samples, as shown in (F). Images obtained on a Zeiss Axio Scan.Z1 digital slide scanner.

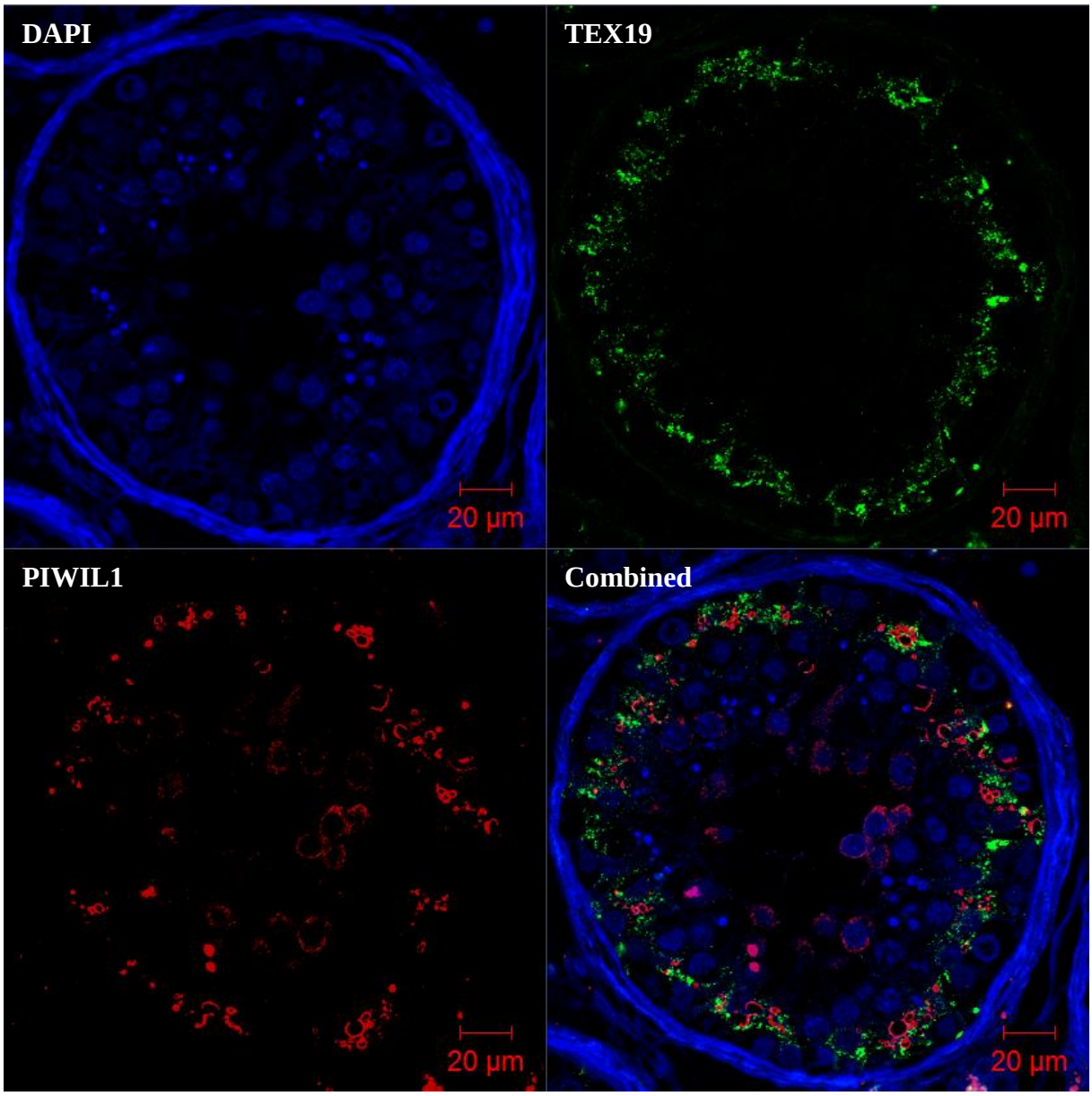


Figure 9. TEX19 and PIWIL1 staining in normal testis (x40).

IF was carried out on a 4 µm section of FFPE tissue blocks of normal testis showing TEX19 (green) staining detected by the rabbit polyclonal antibody (Abcam, ab185547), which again localised between developing spermatocytes but towards the basal compartment of the seminiferous tubule. The co-staining for PIWIL1 was with the mouse monoclonal antibody (Sigma, SAB4200365). The staining for PIWIL1 appeared stronger in the adluminal compartment where a higher proportion of cells would be undergoing meiosis. Although there was no clear co-localisation between these two proteins, there was a region on the interface between the spermatogonial and adluminal compartments where cytoplasmic staining in the germ cells for PIWIL1 was quite strong; this is the same region of the seminiferous tubule that TEX19 was consistently found (although not in the precise same cellular region). Images acquired on Zeiss LSM 710 confocal microscope.

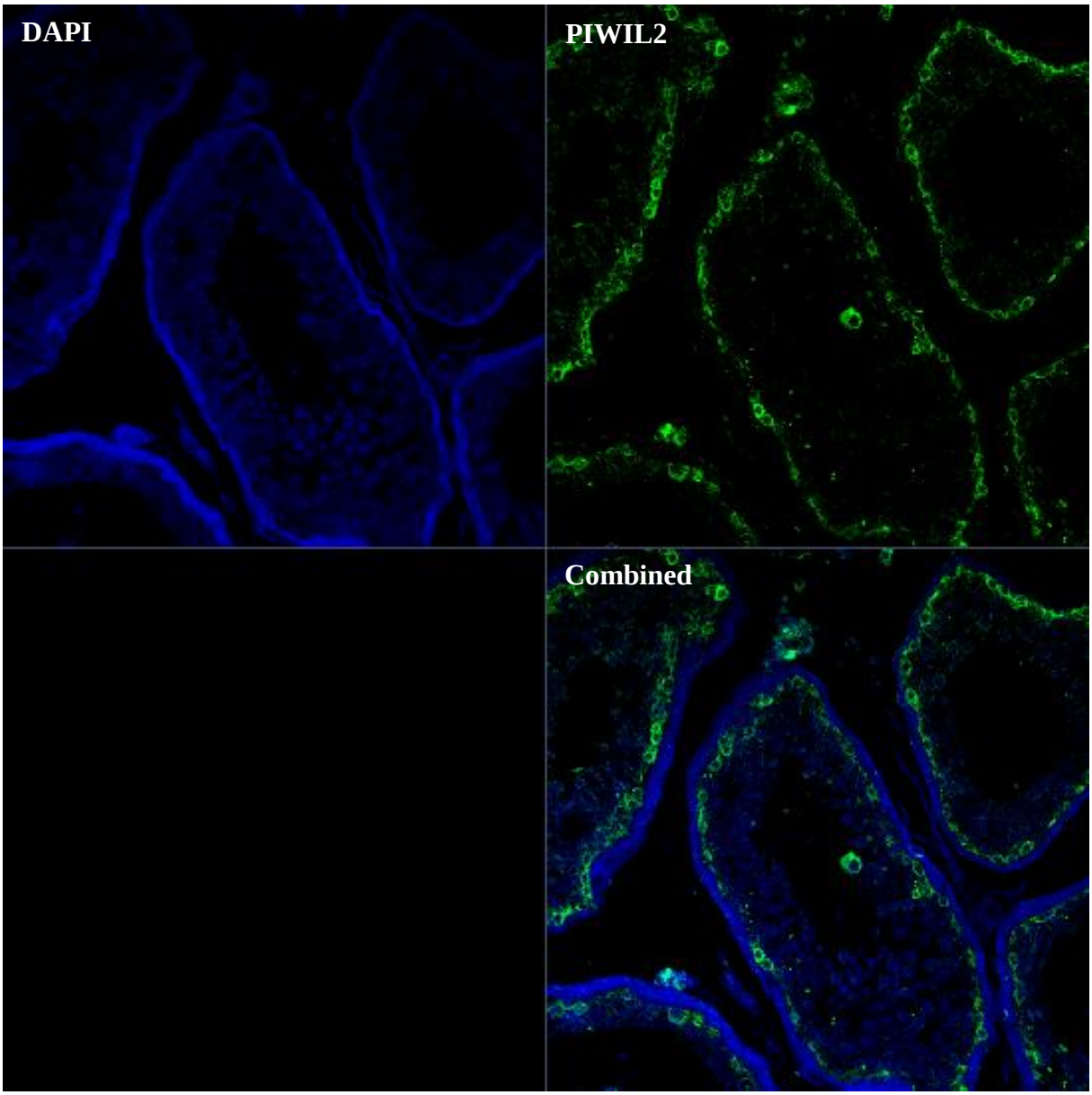


Figure 10. PIWIL2 staining in normal testis (x40).

IF was carried out on FFPE tissue blocks of normal testis showing PIWIL2 (green) staining detected by the mouse monoclonal antibody (Abnova, MAB0843). In this example, more distinct staining was observed in the cytoplasm of germ cells in the spermatogonial compartment. Comparison can be made to Figure 4.8 in Chapter 4, where co-staining for TEX19 was performed. Images acquired on Zeiss LSM 710 confocal microscope.

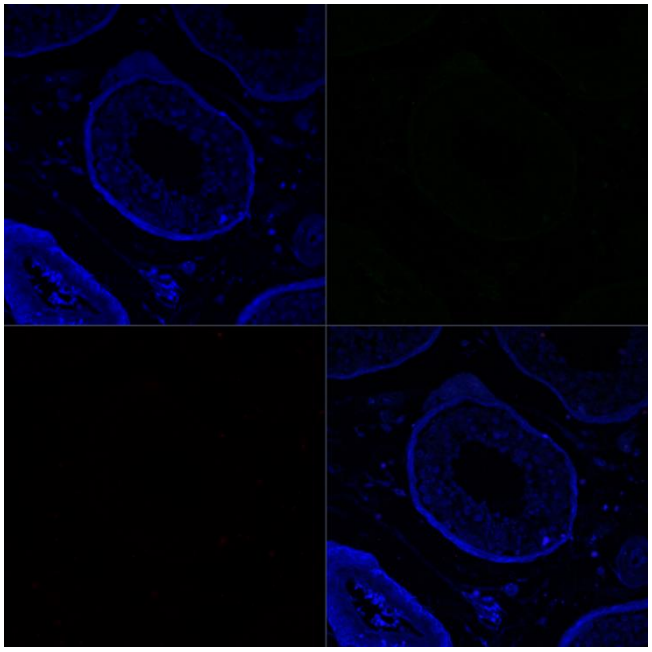


Figure 11. Negative control for IF analysis on normal testis (40x)

Negative control for Figures 4.6-4.8, in Chapter 4. Top left – DAPI staining. Top right – Alexa Fluor 488 (green) rabbit secondary antibody only. Bottom left – Alexa Fluor 568 (red) mouse secondary antibody only. Bottom right – maximum intensity projection of combined image. Images acquired on Zeiss LSM 710 confocal microscope.

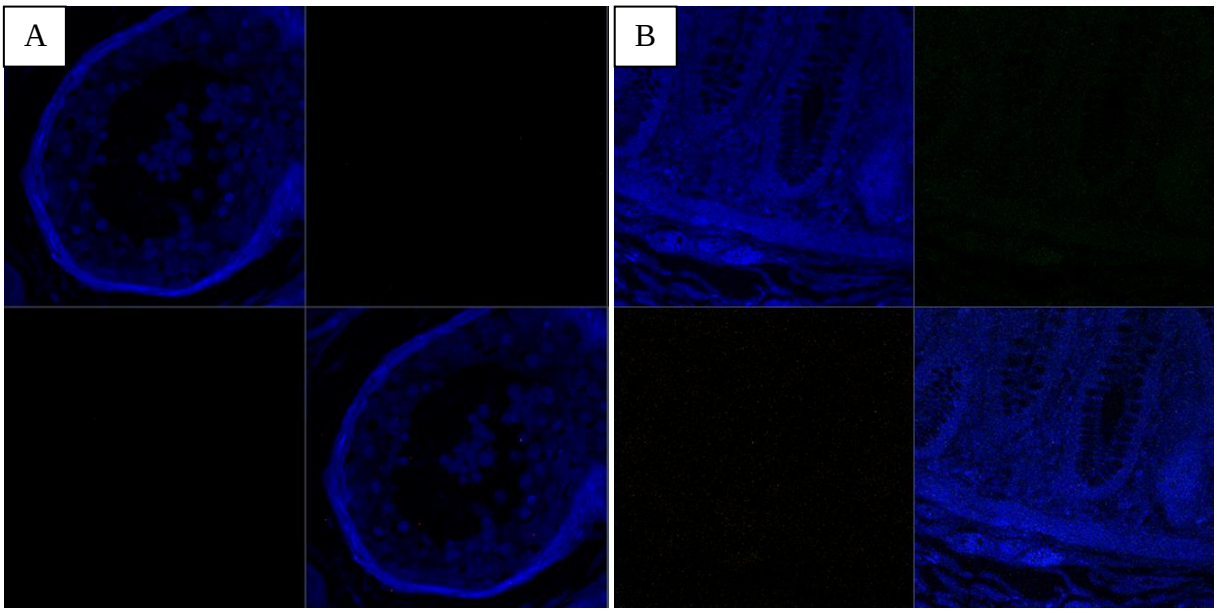


Figure 12. Negative control for IF analysis on normal testis and colon (40x)

Negative controls for Figures 4.12-4.16. Negative control performed in normal testis shown in (A) and negative control in normal colon in (B). Top left of each image – DAPI staining. Top right of each image – Alexa Fluor 488 (green) rabbit secondary antibody only. Bottom left of each image – Alexa Fluor 647 (red) mouse secondary antibody only. Bottom right of each image – maximum intensity projection of combined image. Images acquired on Zeiss LSM 710 confocal microscope.

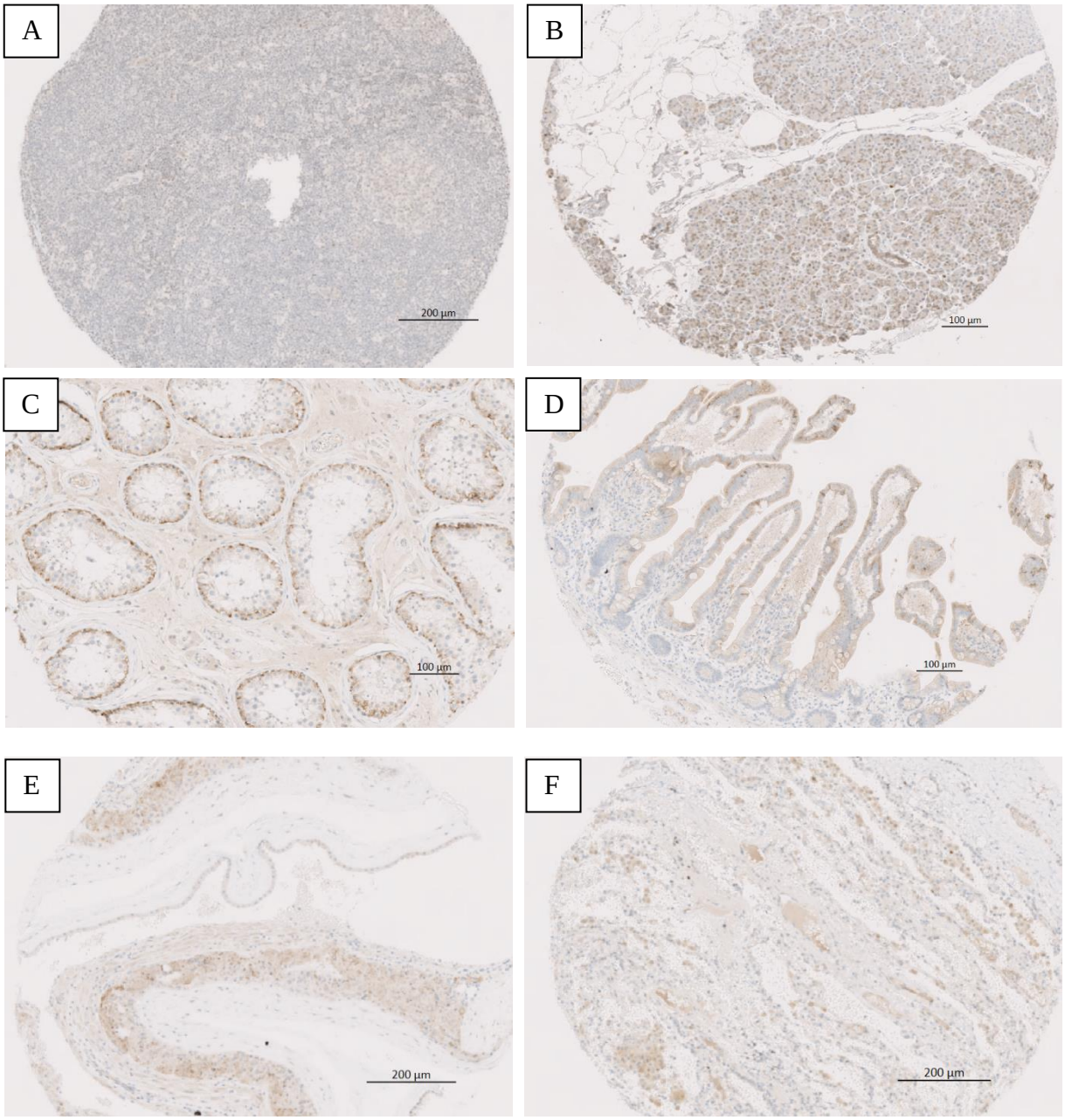
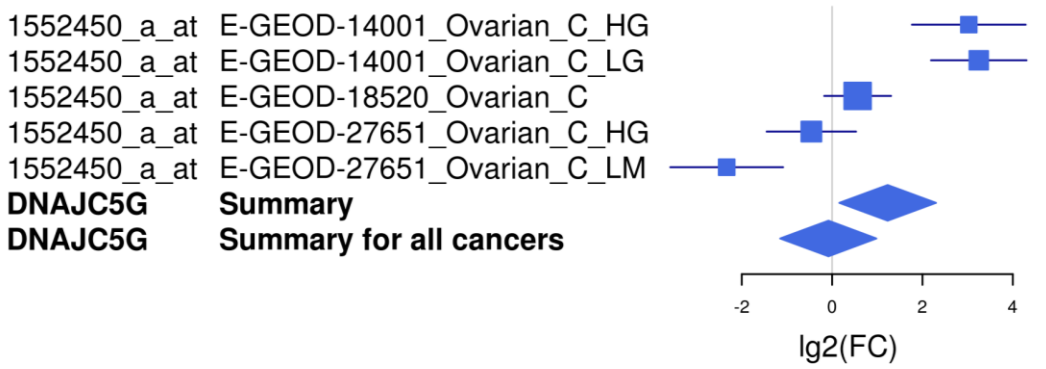
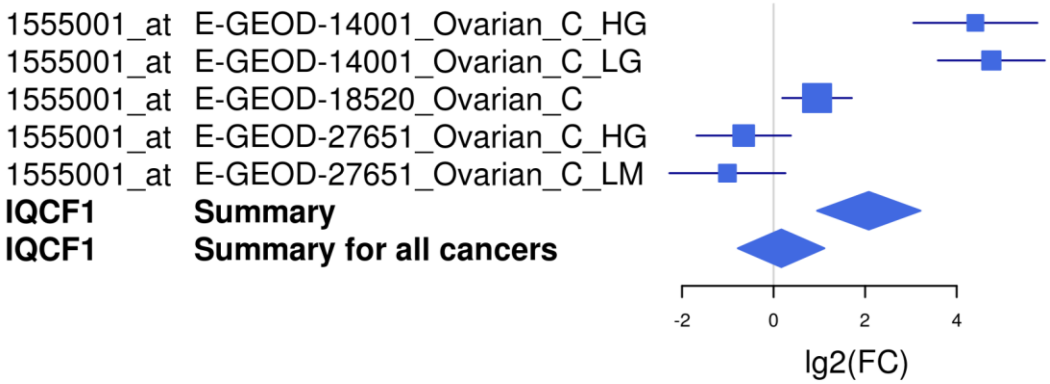


Figure 13. IHC staining for TEX19 in human tissues.

IHC was performed on the semi-automated Ventana® machine (Roche) using the rabbit polyclonal antibody against TEX19 (Abcam; ab185547) on a normal human TMA (CHTN_NORM2). The majority of the tissues included on the array, which included 150 individual tissue cores, stained negative for TEX19. Moreover, the majority of the remaining tissues stained only weakly positive – this was predominantly cytoplasmic staining. Here an example of some of the staining patterns seen is provided. (A) Shows an example of what was considered negative staining in a normal section of lymph node. (B) Is a section of pancreas and this was considered to display weak focally positive nuclear and cytoplasmic staining. Normal testis (C) was included on the array and produced a similar staining pattern to that seen previously with moderate staining seen towards the basal compartment of the seminiferous tubules; acting as a positive control. Weak but consistently positive staining was seen in intestinal epithelial cells but predominantly adjacent to the lumen (i.e. adluminal); a sample of small intestine with stronger than typical staining is shown in (D). Weak cytoplasmic staining also seen in amniotic membrane shown in (E) and in a sparse fashion within normal bronchial epithelium shown in (F). Images obtained on a Zeiss Axio Scan.Z1 digital slide scanner.



(A)



(B)

Figure 14. Forest plots for two of the identified meiCT genes, *DNAJC5G* and *IQCF1*.

These Forrest plots display a statistically significant meta-upregulation in gene expression for both genes in ovarian cancer when combined microarray datasets are analysed. There is an approximate 1.6 log-fold change over matched normal *vs.* cancer array sets for *DNAJC5G* (A) and an approximate 2.1 log-fold change for *IQCF1* (B). The upper five squares illustrate the individual microarray studies, with the confidence intervals for the individual studies represented by the horizontal lines. The size of the squares is proportional to the weight assigned to the individual study. The upper diamond is a summary of all five data-sets showing significant upregulation for ovarian cancer. When all the cancer-types on the arrays were meta-analysed there was no significant change in expression over normal tissues (lower diamond).

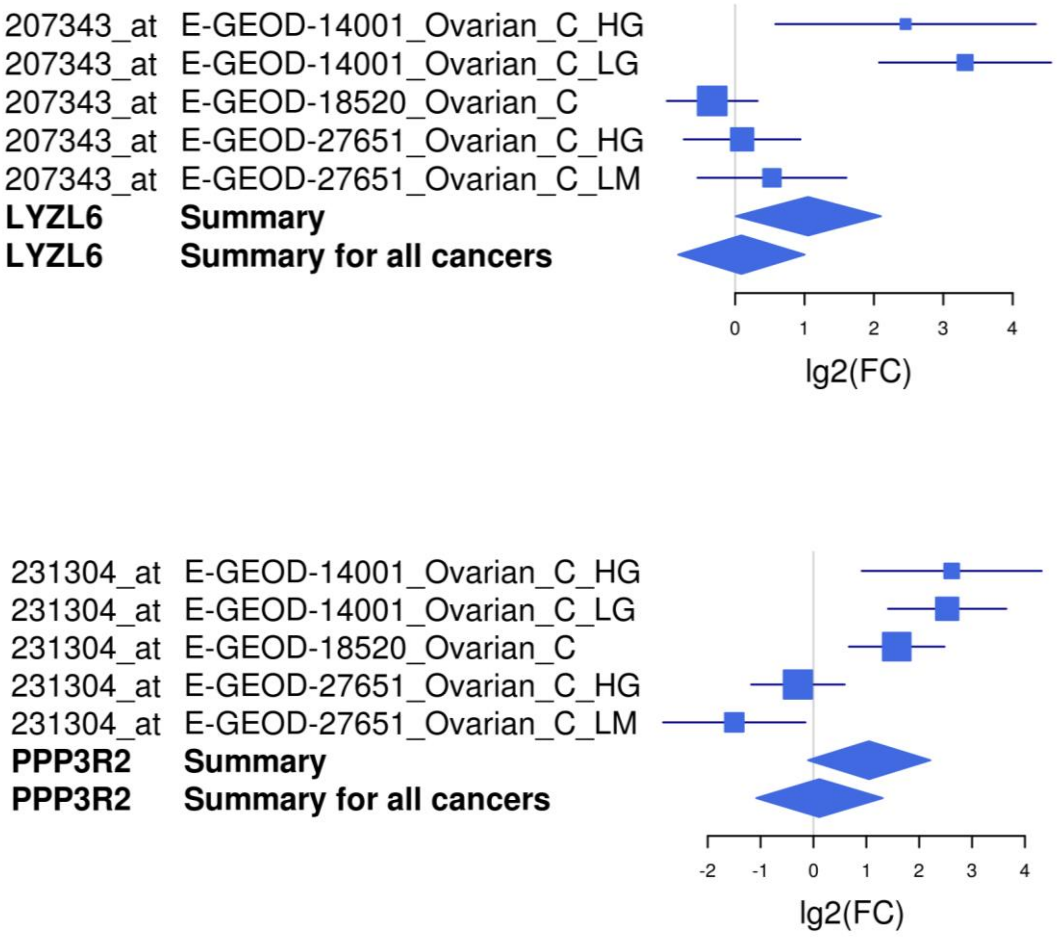


Figure 15. Forest plot for two of the identified meiCT genes, *LYZL6* and *PPP3R2*.

These Forrest plots display a statistically significant meta-upregulation in gene expression for both genes in ovarian cancer when combined microarray datasets are analysed. There is an approximate 1.2 log-fold change over matched normal *vs.* cancer array sets for both genes. The upper five squares illustrate the individual microarray studies, with the confidence intervals for the individual studies represented by the horizontal lines. The size of the squares is proportional to the weight assigned to the individual study. The upper diamond is a summary of all five data-sets showing significant upregulation for ovarian cancer. When all the cancer-types on the arrays were meta-analysed there was no significant change in expression over normal tissues (lower diamond). The Forrest plot for *SPZ1*, which also displayed a meta-upregulation in ovarian cancer, is shown in Figure 3.7.

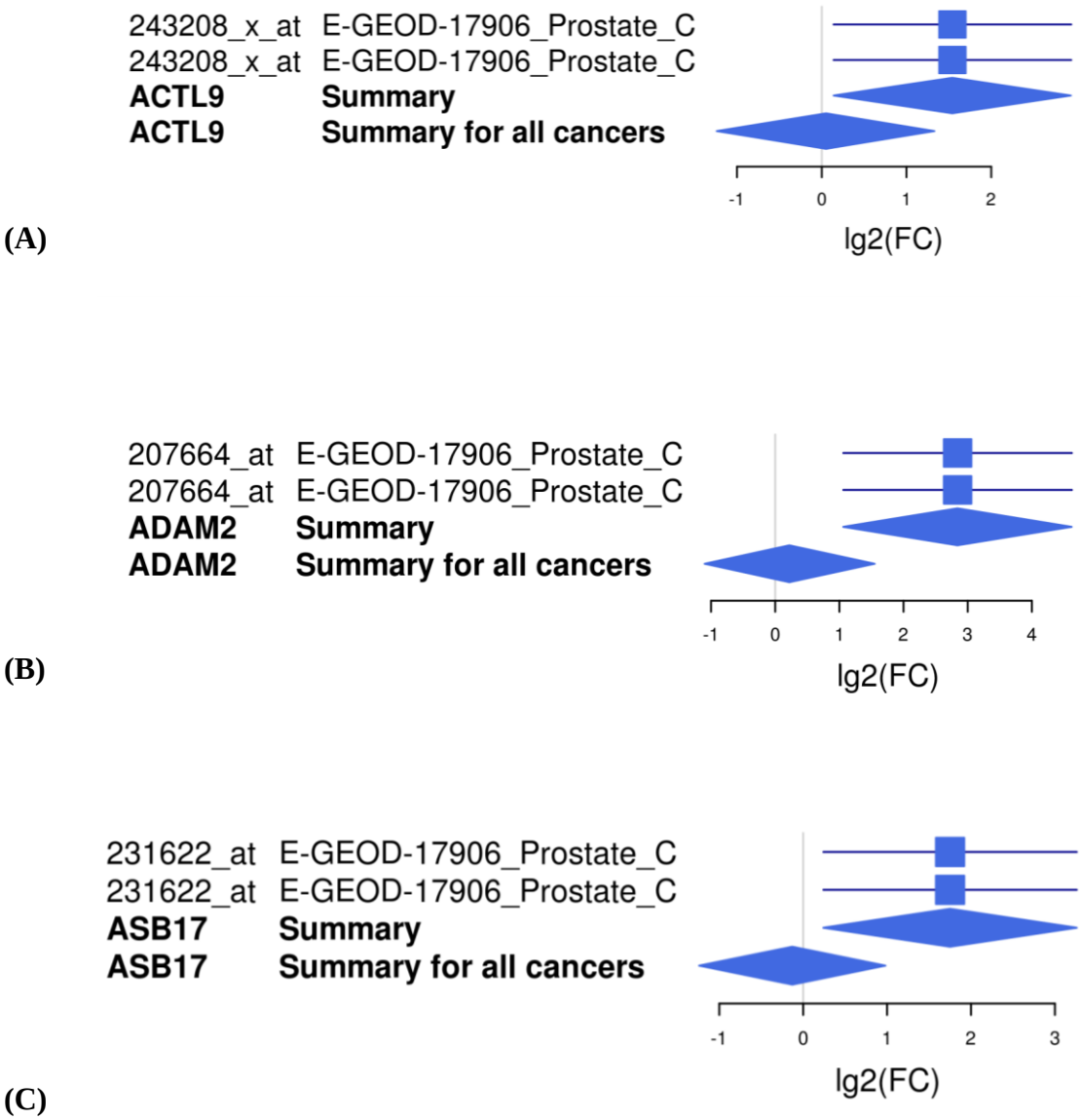


Figure 16. Forest plots for three of the identified meiCT genes: *ACTL9, ADAM2* and *ASB17*.

These Forrest plots display a statistically significant meta-upregulation in gene expression in prostate cancer when combined microarray datasets are analysed. There is an approximate 1.7 log-fold change over matched normal *vs.* cancer array sets for *ACTL9* (A), an approximate 2.9 log-fold change for *ADAM2* (B), and a 1.8 log-fold change for *ASB17* (C). The upper two squares illustrate the individual microarray studies, with the confidence intervals for the individual studies represented by the horizontal lines. The size of the squares is proportional to the weight assigned to the individual study. The upper diamond is a summary of the data-sets showing significant upregulation for prostate cancer. When all the cancer-types on the arrays were meta-analysed there was no significant change in expression over normal tissues (lower diamond).

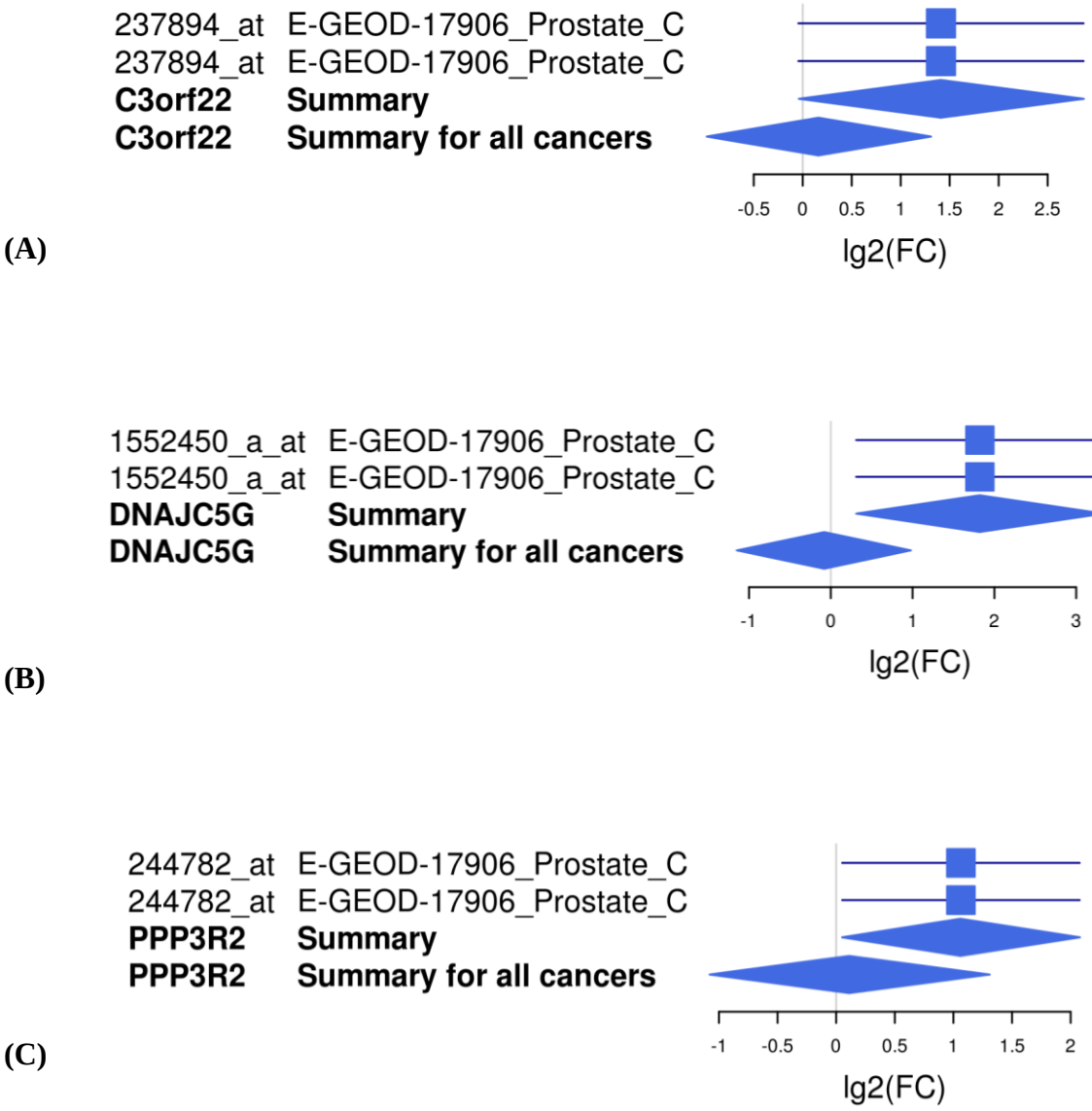


Figure 17. Forest plots for three of the identified meiCT genes: *C3orf22, DNAJC5G* and *PPP3R2*.

These Forrest plots display a statistically significant meta-upregulation in gene expression in prostate cancer when combined microarray datasets are analysed. There is an approximate 1.5 log-fold change over matched normal *vs.* cancer array sets for *C3orf22* (A), an approximate 1.9 log-fold change for *DNAJC5G* (B), and a 1.1 log-fold change for *PPP3R2* (C). The upper two squares illustrate the individual microarray studies, with the confidence intervals for the individual studies represented by the horizontal lines. The size of the squares is proportional to the weight assigned to the individual study. The upper diamond is a summary of the data-sets showing significant upregulation for prostate cancer. When all the cancer-types on the arrays were meta-analysed there was no significant change in expression over normal tissues (lower diamond).

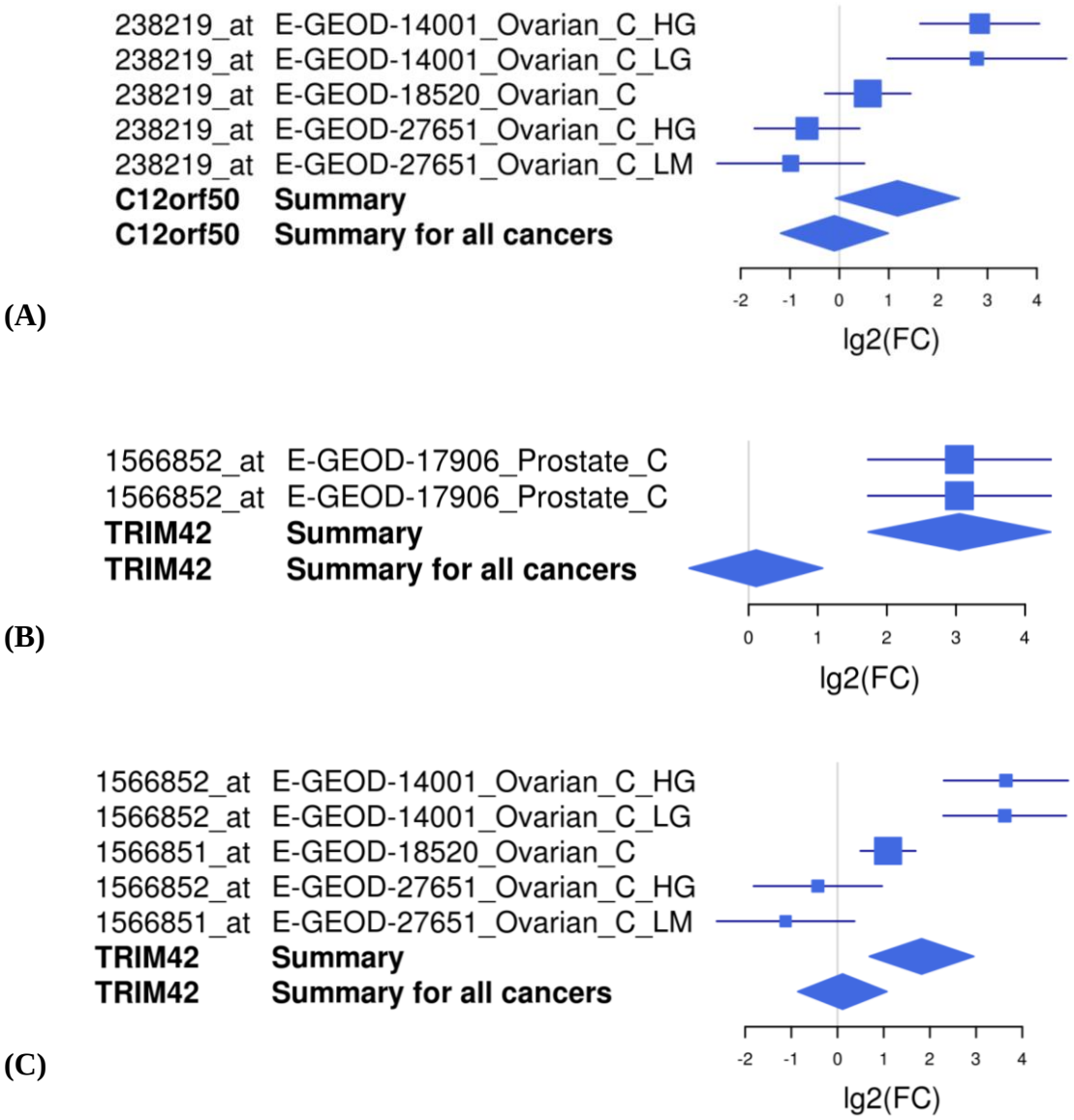


Figure 18. Forest plots for two of the identified meiCT genes: *C12orf50* and *TRIM42*.

Forrest plots displaying a statistically significant meta-upregulation in gene expression in prostate and/or ovarian cancer when combined microarray datasets are analysed. These genes did not appear to be expressed in the limited cancer samples we tested through PCR. There was however an approximate 1.2 log-fold change over matched normal *vs.* cancer array sets for *C12orf50* (A), an approximate 3.1 log-fold change in prostate cancer (B) and 1.9 log-fold change in ovarian cancer (C) for *TRIM42*. The upper squares illustrate the individual microarray studies, with the confidence intervals for the studies represented by the horizontal lines. The size of the squares is proportional to the weight assigned to the individual study. The upper diamond is a summary of the data-sets showing significant upregulation for relevant cancer type. When all the cancer-types on the arrays were meta-analysed there was no significant change in expression over normal tissues (lower diamond).

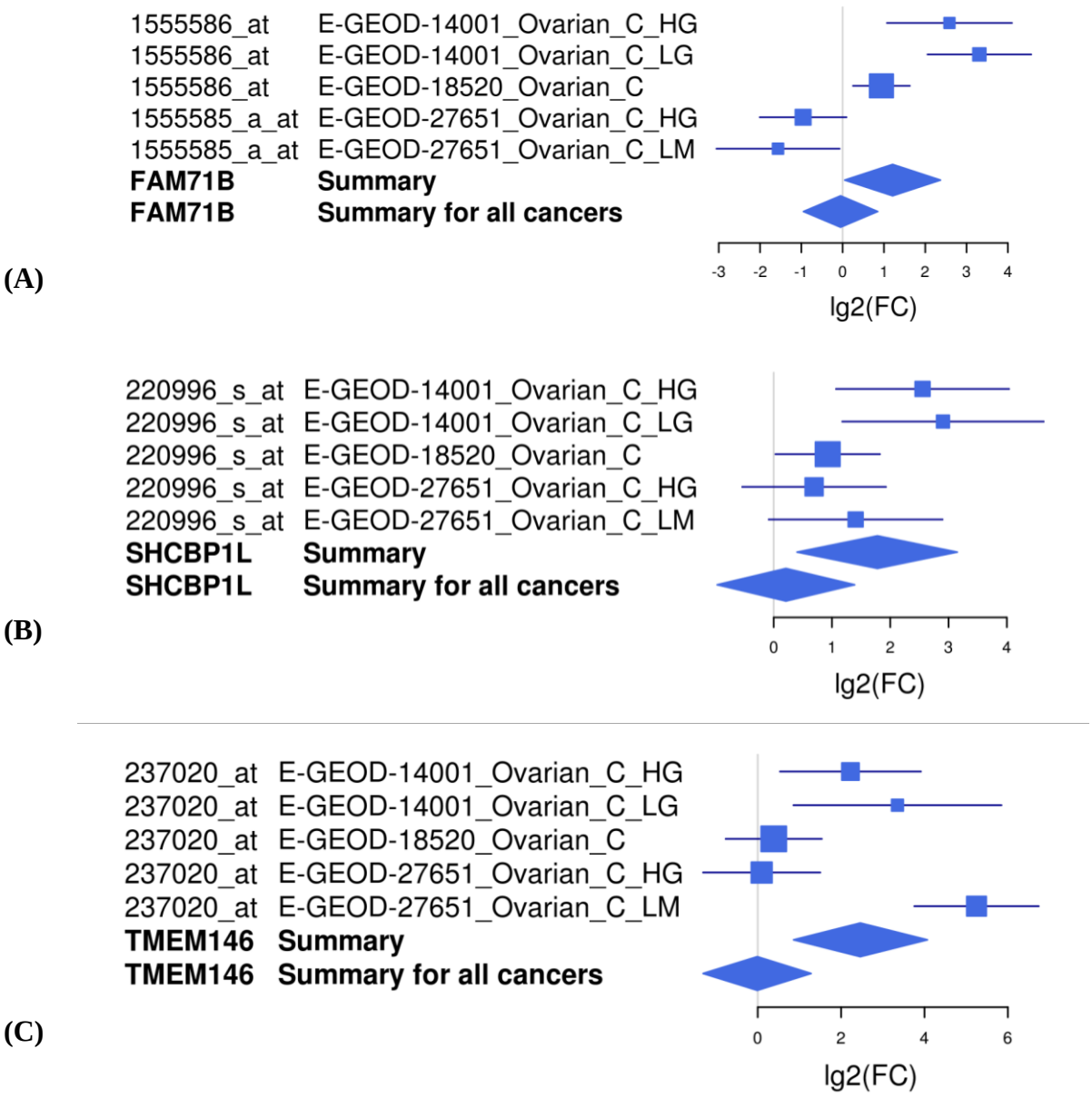


Figure 19. Forest plots for three of the identified meiCT genes: *FAM71B, SHCBP1L* and *TMEM146*.

These genes did not appear to be expressed in the cancer samples we tested through PCR. However, there was a meta-upregulation in gene expression in ovarian cancer for all three genes when combined microarray data are analysed. There was an approximate 1.3 log-fold change over matched normal *vs.* cancer array sets for *FAM71B* (A), an approximate 1.9 log-fold change for *SHCBP1L* (B) and a 2.4 log-fold change for *TMEM146 (C)*. The upper squares illustrate the individual microarray studies, with the confidence intervals for the studies represented by the horizontal lines. The size of the squares is proportional to the weight assigned to the individual study. The upper diamond is a summary of the data-sets showing significant upregulation for relevant cancer type. When all the cancer-types on the arrays were meta-analysed there was no significant change in expression over normal tissues (lower diamond).

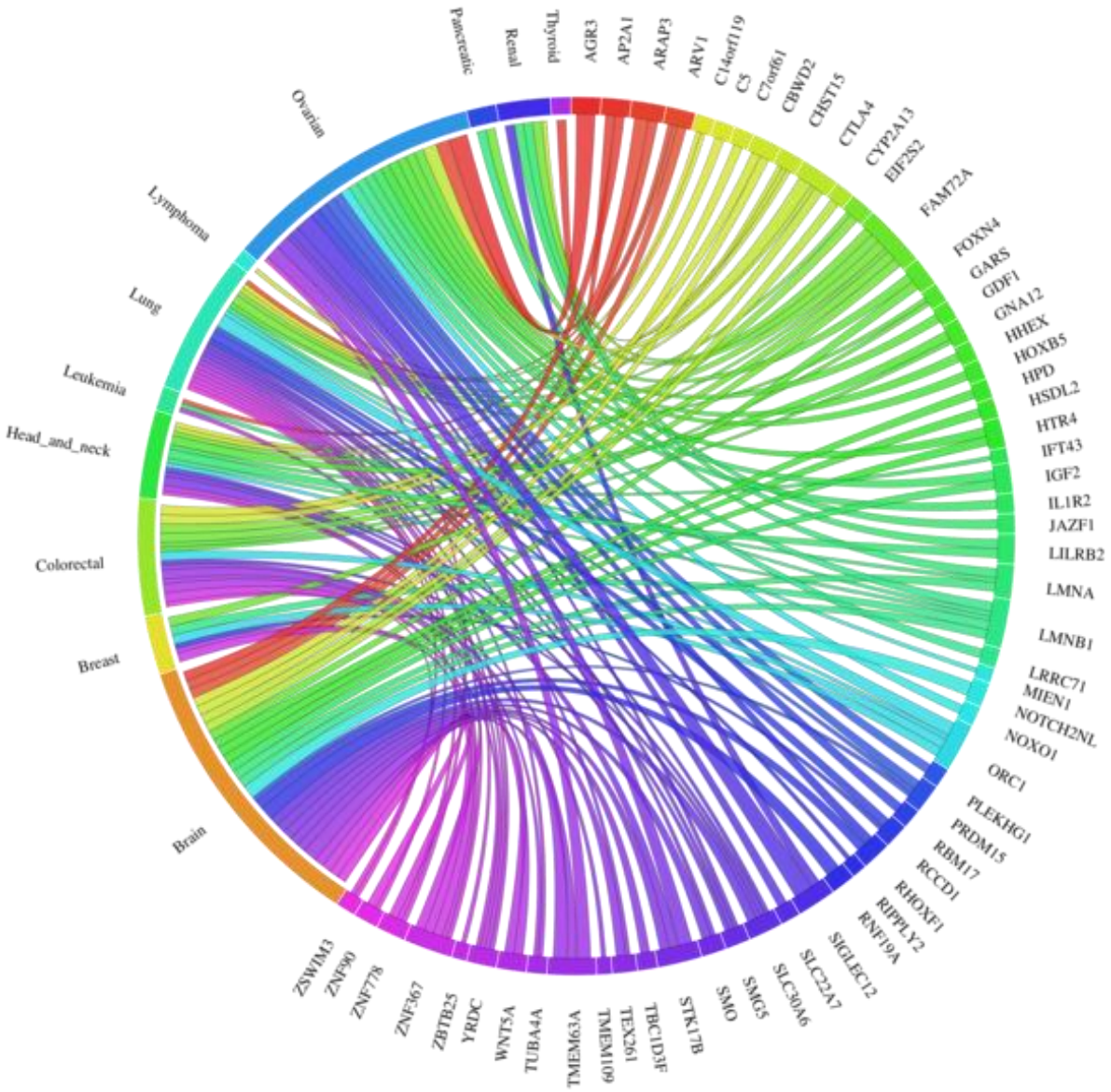


Figure 20. Circos plot displaying meta-upregulated randomly selected genes.

100 genes were randomly selected from a list of protein-encoding genes provided by the HUGO gene nomenclature committee (HGNC) – www.genenames.org.uk (website accessed 18th May 2016). 59 of these randomly selected genes displayed a meta-upregulation in at least one cancer type. 23 of the genes were not included on the arrays, so overall 76.6% displayed at least one upregulation in at least one cancer type. Brain, ovarian and lung cancer displayed the most upregulations but in comparison to Figure 4 above, there was not a striking preponderance for ovarian cancer as was seen for the meiCT gene cohort. The thickness of the line connecting individual genes to the associated cancer type in which they are upregulated corresponds to the degree of upregulation or strength of association.

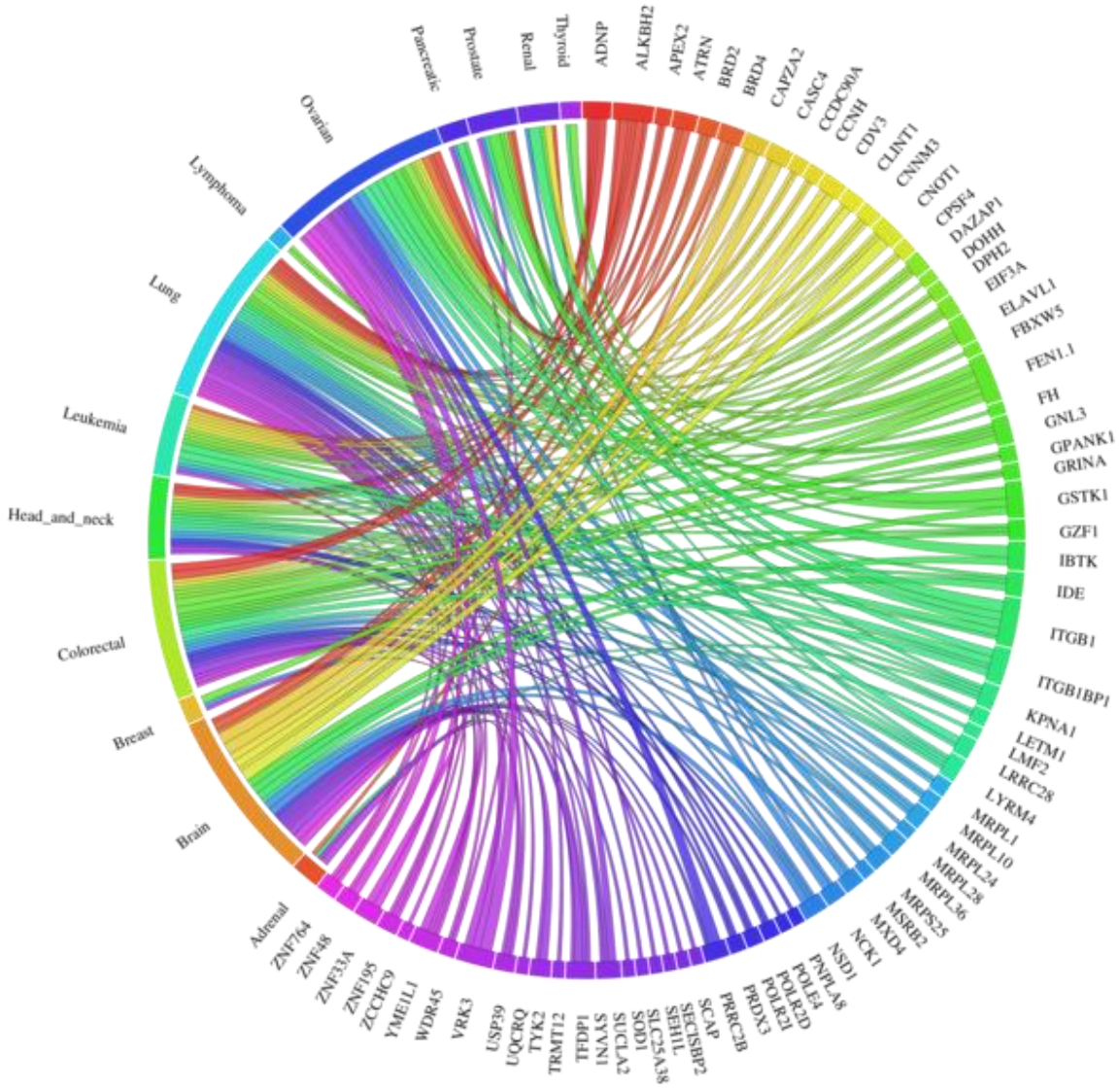


Figure 21. Circos plot displaying meta-upregulated randomly selected 'housekeeping' genes.

Housekeeping genes are expressed in both normal and pathological tissues and some display fairly constant expression in both normal and pathological states (Eisenberg and Levanon, 2013). They are thus useful for quantitative PCR experiments, acting as a normalisation factor, though should ideally be optimised for each disease state or sample cohort. 100 genes were randomly selected from a list of housekeeping genes, freely available at www.tau.ac.il (website accessed 10th June 2016). Of the 100 randomly selected housekeeping genes, 73 were shown to have a meta-upregulation in at least one cancer type. However, only 6 of these genes were not included on the arrays. So, overall 77.6% displayed at least one up-regulation in at least one cancer type. The thickness of the line connecting individual genes to the associated cancer type in which they are upregulated corresponds to the degree of upregulation or strength of association.

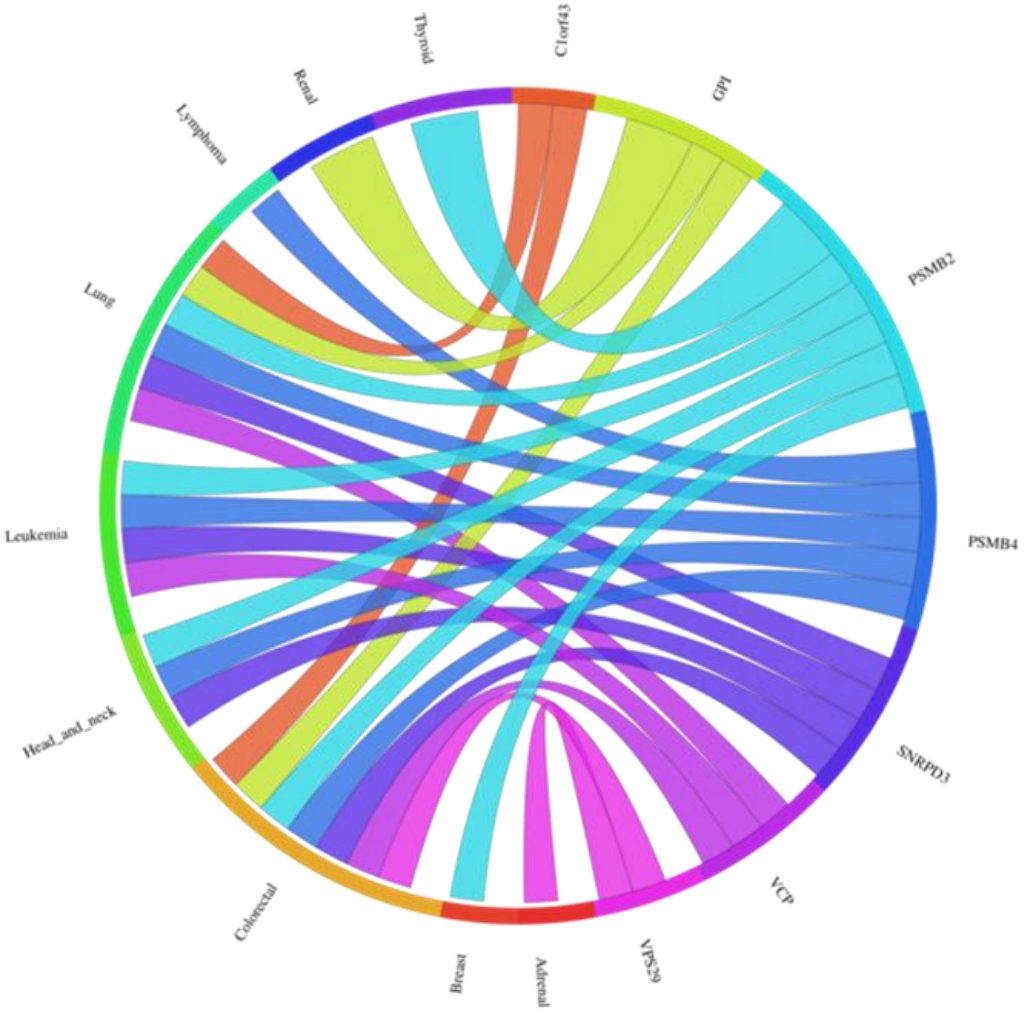


Figure 22. Circos plot displaying meta-upregulated selected 'housekeeping' genes.

Eleven housekeeping genes were put forward as the most consistent genes displaying constant expression in different tissue states (Eisenberg and Levanon, 2013). When subject to analysis through the cancerMA pipeline (Feichtinger *et al.*, 2012a), seven of the genes displayed a meta-upregulation in at least one cancer type. One of the genes was not present on the arrays, so overall 70% displayed at least one upregulation in at least one cancer type: only slightly lower than the randomly selected cohort of genes. Although these genes would be expected to show constant expression, the fact they do in selected cancers is still not unsurprising given the heterogenous nature of the disease. The proportion of genes displaying upregulations is slightly lower when compared to other gene cohorts. As acknowledged in the text, there are limitations of cancerMA as there are with other existing pipelines and the results produced should not be over-interpreted. The thickness of the line connecting individual genes to the associated cancer type in which they are upregulated corresponds to the degree of upregulation or strength of association.

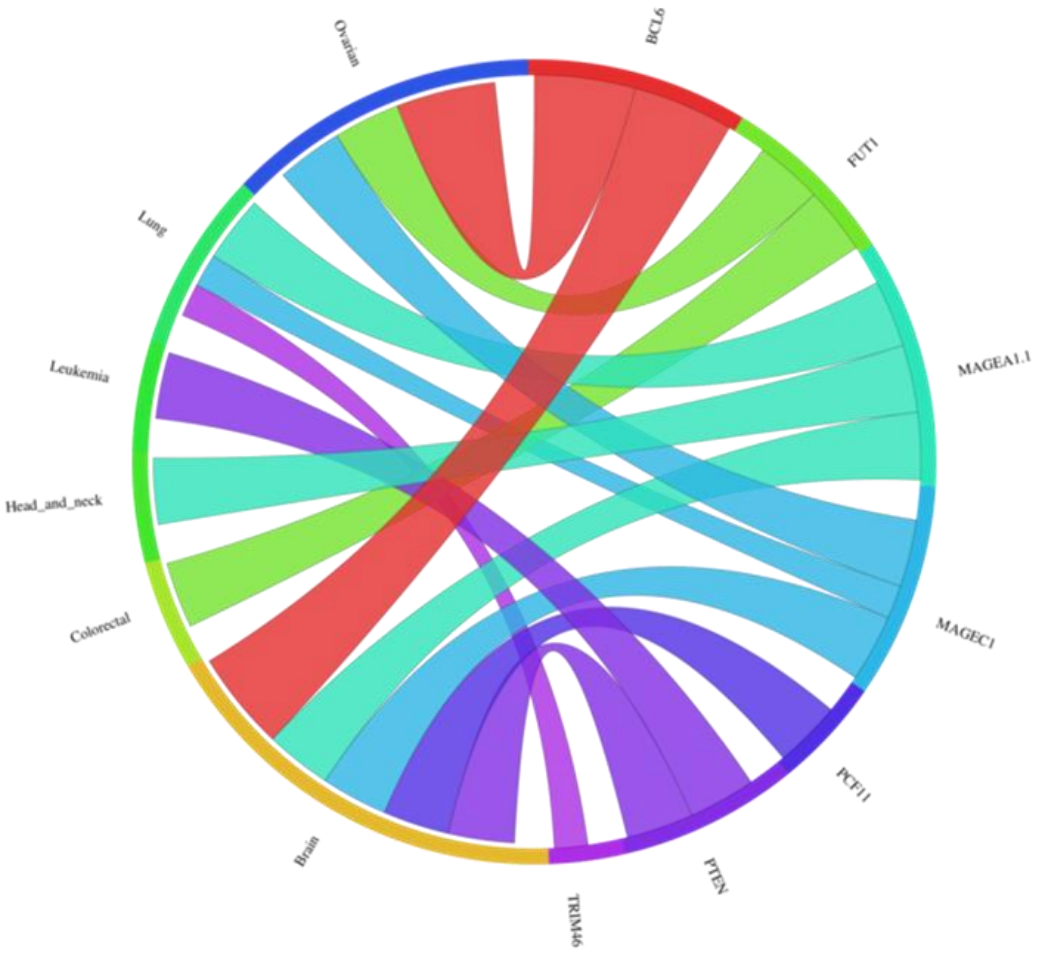


Figure 23. Circos plot displaying meta-upregulated genes previously shown to be upregulated in ovarian cancer.

Seven genes have recently been shown in one study to be upregulated in ovarian cancer samples (Gao *et al.*, 2016). When subject to analysis through cancerMA, 3 of these 7 genes displayed a meta-upregulation in ovarian cancer. As highlighted in the text cancerMA has limitations, one being the restricted number of arrays included and the inherent weaknesses of the raw microarray data on which it is based. It is not surprising that there is not an exact correlation to other published work using different methodologies in different samples. *BCL6* was shown in a separate study to be upregulated in different subtypes of ovarian cancer (Wang *et al.*, 2015b), so it is reassuring to see that cancerMA displays a meta-upregulation for this gene. The thickness of the line connecting individual genes to the associated cancer type in which they are upregulated corresponds to the degree of upregulation or strength of association.