Bangor University

DOCTOR OF PHILOSOPHY

Identification and functional characterisation of potential novel cancer-testis genes in human cancer cells

Almatrafi, Ahmed

Award date:
2014

Awarding institution:
Bangor University

[Link to publication](#)

Identification and functional characterisation of potential novel cancer-testis genes in human cancer cells

Ahmed Mubrik Almatrafi

Bangor University
College of Natural Science
School of Biological School
Ph.D. Thesis 2014

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: Ahmed Mubrik Almatrafi

Title: Identification and functional characterisation of potential novel cancer-testis genes in human cancer cells

Supervisor/Department: Dr. Ramsay James McFarlane/ School of Biological Science

Funding body (if any): University of Taibah, Medinah, Kingdom of Saudi Arabia

Qualification/Degree obtained: Ph.D.

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). Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

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 repositories, 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 repositories after expiry of a bar on access.

Signed ………………………………………… (candidate)
Date …………………………………………

<u>Statement 4</u>:
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 : ………………………………………….

This thesis is dedicated to my parents & family

For their endless love, support & encouragement

Summary

The identification and characterisation of novel cancer-specific biomarkers and/or targets is a key challenge for establishment of successful diagnostic, prognostic and immunotherapeutic strategies. Cancer/testis (CT) genes are an attractive group of genes that encode proteins that are restricted to the human testis and malignant tumours but are not expressed in healthy somatic cells. The testes are an immunologically privileged site; therefore, CT genes represent important targets/biomarkers in the diagnosis and treatment of cancers.

Here, targeted genes identified from a literature search for meiosis-specific genes as well as from bioinformatics pipelines were analysed using RT-PCR in a range of normal and cancerous cell types. Ten out of 24 genes were identified as promising CT genes as they were expressed in cancer cells but not in normal tissues. A subsequent meta-analysis of seven testis-restricted genes in a microarray data set also identified five genes that were up-regulated in clinical samples derived from patients. The use of hypomethylation (5-aza-2˙-deoxycytidine) and/or histone deacetylation (trichostatin; TSA) drugs further identified a new class of CT genes with refractory transcriptional silencing. In contrast, *STRA8* and *TDRD12* were transcriptionally regulated by DNA methylation and histone acetylation.

Western blot analysis revealed the presence of PRDM9 protein in nearly all cancer cell lines; PRDM9 was localised in the nucleus in three cancer cell lines and in the cytoplasm in NT2 cells. Knockdown of PRDM9 protein in HCT116 and SW480 cells reduced the survival of cancer cells, with a major effect seen with siRNA-7. This may indicate an oncogenic role for PRDM9 in cancer cells, this was explored further using an over expression system.

Finally, PRDM7 and PRDM9 were overexpressed in *E. coli* cells using a Glutathione S-transferase (GST) Gene Fusion System, but the proteins were found mostly in the insoluble fraction and were difficult to purify.

Acknowledgments

The research work presented in this thesis was carried out at the Department of Biological Science, North West Cancer Research Institute, Bangor University.

First and foremost, I would like to express my sincere gratitude to my excellent supervisor, Dr. Ramsay James McFarlane, for the patient guidance, encouragement, motivation, enthusiasm and immense knowledge; his guidance helped me throughout my research and during my writing of this thesis. He is more than a supervisor to me and I have learned many things from him. My sincere thanks also go to Dr. David Warren Pryce for his advice and teaching since I was a Master's student at Bangor University. Thank you to all members of the McFarlane & Wakeman lab in the past and present, especially Dr. Jane Wakeman, Dr. Natalia Gomez Escobar and Dr. Ellen Vernon. You provide an excellent research atmosphere and friendly people who were helpful and provided the required answers to my questions. I really enjoyed doing my PhD with you.

I would like to thank the government of Saudi Arabia, especially the University of Taibah, for their sponsorships and support during my master's and PhD study.

Words cannot explain what I feel toward my mother and father. Thank you very much for everything you have done for me. Thank you for raising me, supporting me and standing with me since I was born; without you, I could not finish my PhD. Thanks to my lovely wife and little daughter Ghazal, for standing with me, helping me and giving me the motivation to make this dream come true, especially your daily crying, Ghazal.

My brothers and sisters Abdullah, Yasser, Ibrahim, Mohammad and my younger sister Ream: without your support, standing with me, asking about me and calling me every day, your older brother could not have finished his PhD. Lastly, my special and deep thanks go to my grandmother and grandfather in heaven, who passed away during thesis writing.

Thank you all

Ahmed Almatrafi

Abbreviation

5-AZA-CdR	5-aza-2'-deoxycytidine
AEs	The axial elements
amp	ampicillin
ATM	ataxia telangiectasia protein
B-ALL	B- cell precursor acute lymphoblastic leukaemia
BLAST	Basic Local Alignment Search Tool
BLM	Bloom syndrome protein
bp	Base pair(s)
BRCA1	Breast cancer susceptibility 1
BSA	Bovine serum albumin
BSP	bisulphite sequencing PCR
BBB	Blood-brain barrier
BTB	Blood-testis barrier
CancerMA	meta-analysis microarray web tool
cDNA	Complementary DNA
CDS	Coding DNA sequence
CE	The central element
CG	Cancer germline genes
CIN	Chromosome instability
CNS	Central nervous system
COs	Crossover events
CpG	- cytosine-phosphate-guanine-
Cq	Quantification cycle
CRC	Colorectal cancer
CTAs	Cancer testis antigens
C-terminal	Carboxy-terminal domain
DEPC	Diethylpyrocarbonate
dH$_2$O	Distilled water
dHJs	Double Holliday junctions
D-loop	Disassociation loop
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DNMTi	DNA methyltransferase inhibitor

DP	Protein dimerise
DSBs	Double strand breaks
EB	Elution buffer
EBV	Epstein-Barr virus antigen
ECACC	European collection of cell cultures
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
ELDA	Extreme limiting dilution analysis
EST	Expressed sequence tag
FBS	Foetal bovine serum
G_0	Resting/quiescent phase
G_1	Gap-1 phase
G_2	Gap-2 phase
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GST	Glutathione S-transferase tag
H3K36	Histone 3 lysine 36
H3K36me3	Trimethylation of lysine 36 of histone H3
H3K4	Histone 3 lysine 4
H3K4me3	Trimethylation of lysine 4 of histone H3
HATs	Histone acetylransferases
HDACs	Histone deacetylases
HDMs	Histone demethylases
HMTs	Histone methyltransferases
HNPCC	Hereditary non-polyposis carcinoma coli
HPV	Human papillomavirus antigen
HR	Homologous recombination
HRP	Horseradish peroxidase
IPTG	isopropyl β-D-1-thiogalactopyranoside
kb	Kilobase
KDa	Kilo Dalton
KRAB	Kruppel-association box
LB	Luria bertani
LEs	The lateral elements
LOH	Loss of heterozygosity
meiCT	Meiosis-specific CT antigen genes
MHC	Major histocompatibility complex
MRN complex	MRE11-RAD50-NBS1
MRX complex	Mre11-Rad50-Xrs2

MSCI	Meiotic sex chromosome inactivation
MSP	methylation-specific PCR
N	Nuclear lysate fractions
NCBI	National Centre for Biotechnology Information
NCO	Non-crossover
NHEJ	Nonhomologous end joining
nM	Nanomolar
non X-CT	Non chromosome cancer testis genes
NRT	No reverse transcriptase
NTC	No template control
N-terminal	Amino-terminal domain
ORF	Open reading frame
PBS	Phosphate buffered saline
PBST	Phosphate buffered saline-tween-20
PCR	Polymerase chain reaction
pH	Power of hydrogen
PS	Placenta-specific
PVDF	Polyvinylidene difluoride
qRT-PCR	Quantitative, real time PCR
r.p.m.	Rotation per minute
RB	Retinoblastoma-associated gene
RNA	Ribonucleic acid
RPA	protein A
RT	Room temperature
RT-PCR	Reverse transcriptase PCR
SAM	S-Adenosylmethionine
SC	Synaptonemal complex
SCC	Sister chromatid cohesion
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium Dodecyl Sulphate- Polyacrylamide Gel Electrophoresis
SEM	Standard error of the mean
SEREX	Serological expression cloning
SNPs	Single nucleotide polymorphisms
SOC	Super Optimal broth with Catabolite repression
ssDNA	3' single-stranded DNA
SSXRD	Synovial sarcoma X breakpoint domain
TAA	Tumour-associated antigen
TAE	Tris-acetate-EDTA

TBE	Tris-borate-EDTA
TFs	Transverse filament
TS	Testis-specific
TS/PS	Testis-specific/placenta-specific genes
TSA	Trichostatin A
TSGs	Tumour suppressor genes
TSS	Transcription start site
UCSC	Genome Browser bioinformatics website
UV	Ultraviolet light
V-onc	viral oncogenes
W	Whole cell extract
WB	Western blot
X-CT	X- chromosome cancer testis genes

Contents

List of Figures

List of Tables

Chapter 5.0

Chapter 6.0

Chapter 1.0 Introduction

1.1 Cell cycle

Two types of cell division occur in eukaryotic organisms: mitosis and meiosis. During mitotic cell division, a single diploid parental cell divides to produce two identical daughter cells that have an equal number of chromosomes. Mitotic cell division normally occurs in somatic cells as part of the normal cell cycle and maintains the chromosome complement. Mitosis is important for tissue homeostasis and replacement of dead cells and damaged tissues (reviewed in Walczak *et al.*, 2010; Silkworth and Cimini, 2012).

Meiosis is a special form of cell division required for sexual reproduction in higher eukaryotes. In meiosis, diploid cells in the gonads divide to generate haploid gametes, the sperm and egg cells (reviewed in Marston and Amon, 2004). This process is a key source of genetic variation and information reshuffling between paternal and maternal chromosomes and generates new combinations of alleles in the population (Segurel, 2013). Meiosis and mitosis are highly regulated, so that errors in these two processes are associated with genetic diseases (Miller *et al.*, 2013).

During the normal mitotic cell cycle, each cell duplicates the genomic content and divides to generate two daughter cells. The cell cycle has four stages: Gap-1 (G_1), Synthesis (S), Gap-2 (G_2; also known as interphase) and mitosis (M). During interphase, cells spend most of their lives growing in size and replicating the chromosomal DNA (Figure 1.1). During the M phase, the replicated sister chromatids are segregated into the two daughter cells (reviewed in Kronja and Orr-Weaver, 2011).

The cell cycle in eukaryotic cells is highly conserved and controlled by a series of restricted checkpoints. These checkpoints are required to ensure genome integrity and to provide the opportunity for the cells either to complete cell division or to exit into a resting/quiescent phase (the G_0 phase). Central to this process are set of proteins termed cyclins and their associated cyclin-dependent kinase (Cdk) proteins. For instance, if the chromosome is not attached correctly to the mitotic spindle during the M phase, checkpoints will prevent the

cell from completing division. Consequently, dysregulation of cell cycle regulators is a hallmark of cancer and tumourigenesis (Aarts *et al.*, 2013).
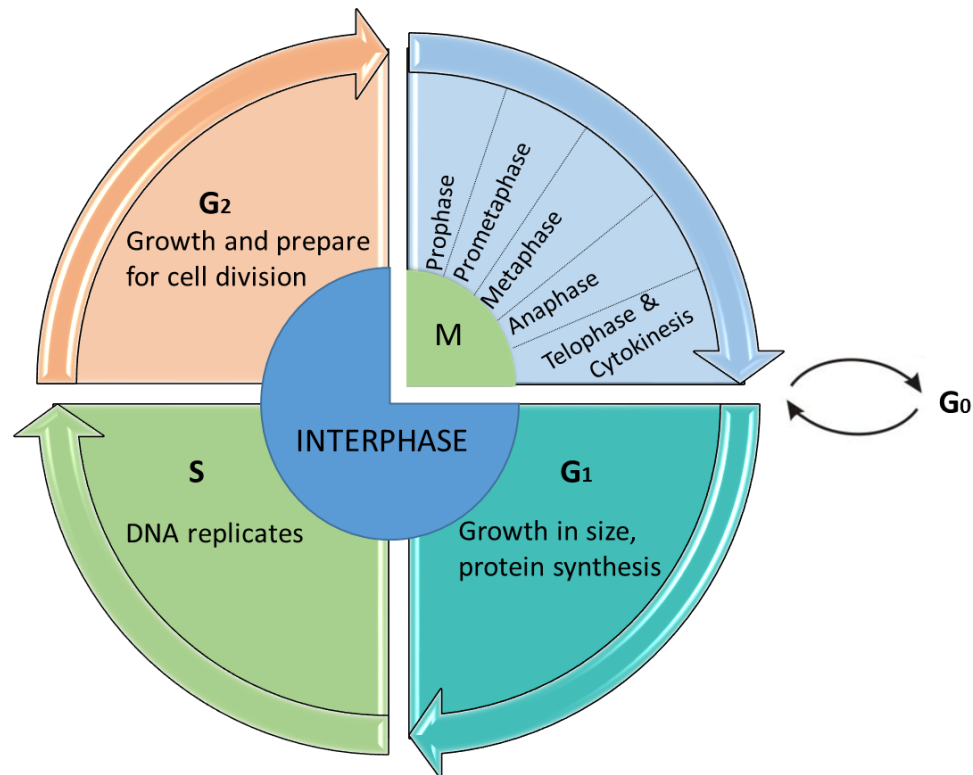
Figure 1.1. **A schematic representation of the events in the eukaryotic cell cycle.** The cell cycle has two main stages: interphase and mitosis (M). Interphase consists of three phases: Gap-1 (G_1), Synthesis (S) and Gap-2 (G_2). During the G_1 phase, the cell grows in size and synthesise mRNA and structural proteins in preparation for DNA synthesis. During the S phase, DNA is replicated so that each chromosome is duplicated. During the G_2 phase, the cell grows in size, protein is synthesised and the cell prepares to enter into mitotic cell division. The M-phase is composed of two parts: nuclear segregation (mitosis) and cytoplasmic cell division (cytokinesis). Mitosis consists of five sub stages: prophase, prometaphase, metaphase, anaphase and telophase. G_0 is the resting phase or quiescent state occurring between the cell cycle.

1.1.1 Mitotic cell division

During the S-phase of mitosis, DNA replicates once, followed by one round of chromosome segregation, resulting in two identical daughter cells. Mitotic cell division is divided into different distinct phases in terms of the chromosome organisation and behaviour: prophase, prometaphase, metaphase, anaphase, telophase and cytokinesis (Figure 1.2). During the prophase stage, the replicated chromosomes, which contain two sister chromatids held by cohesin complex proteins (see Section 1.2.3 for details), begin to condense, gradually becoming shorter, thicker and easily visualised. The two centrosomes migrate to opposite poles of the cell and the mitotic spindle begins to assemble out of the nucleus (reviewed in Walczak *et al*., 2010; Silkworth and Cimini, 2012). During the prometaphase stage, the nuclear envelope breaks down in higher eukaryotes and the chromosomes attach to the mitotic spindle microtubules via a structure known as the kinetochore. In the metaphase stage, the chromosomes line up at the equator of the cell on the metaphase plate, halfway between the two spindle poles. The spindle microtubules are attached to paired kinetochores, before chromatid separation is signalled. During the anaphase stage, the two sister paired chromatids are separated and move in synchrony toward the opposite poles, resulting in the disruption of the cohesin between the sister chromatids. In the telophase stage, the chromosomes arrive at the opposite poles and decondense, and a new nuclear envelope reforms around each daughter nucleus. During cytokinesis, the cytoplasm of the cell is segregated into two distinct daughter cells (reviewed in Petronczki *et al*., 2003; Walczak *et al*., 2010).

1.1.1.1 Meiotic cell division

In contrast to mitosis, meiosis is a specialised form of cell division required for sexual reproduction and produces four different daughter cells, each with one set of chromosomes (reviewed in Marston and Amon, 2004). The meiotic programme begins with a single phase of DNA replication, followed by two successive rounds of chromosome segregation (meiosis I and meiosis II). In meiosis I (reductional division), the chromosome numbers are halved as the homologous chromosomes segregate into two individual nuclei (Figure 1.2). During meiosis II (equational division), the sister chromatids segregate from each other into four

individual gametes without an intervening S-phase (reviewed in Petronczki *et al.*, 2003; Marston and Amon, 2004; Lichten and de Massy, 2011).

The stages of meiosis I and II have some similarities to mitosis but also have some fundamental differences. In the first meiotic division, homologous chromosome pair and become physically connected, followed by their subsequent segregation from each other. In the second meiotic division, the chromosomes segregate in a mitotic-like fashion without a further DNA replication step. There are different distinct stages of meiosis I and II: prophase, metaphase, anaphase and telophase (reviewed in Handel and Schimenti, 2010).

During prophase I, homologous chromosomes pairing is initiated to allow close association by synapsis. This process allows the exchange of paternal and maternal DNA to form homologous recombination that generated crossover (CO) events. During metaphase I, the joined homologous chromosome pairs (bivalents) align along the metaphase plate, to direct bipolar tension on the bivalent prior to anaphase I. At anaphase I, the recombination mediated connections are resolved and chromosomes reductionally segregate. At telophase I, the nuclear membrane generates the two daughter cells, ending meiosis I (reviewed in Handel and Schimenti, 2010).

Cells then undergo meiosis II without an intervening S-phase or DNA replication. During meiosis II, the sister chromatids, which remain connected via centromere-associated cohesion, align on the metaphase plate halfway between the two spindle pores in metaphase II, followed by pulling of the sister chromatids to opposite poles in anaphase II. During telophase II, cells divide to generate four genetically different haploid gametes cells, known as spermatids in the male and ova in the female (Handel and Schimenti, 2010; Miller *et al.*, 2013).

1.1.1.2   The sub-stages of prophase I

In eukaryotic cells, prophase I is one of the longest and most complex phases of meiotic cell division. Several biological events occurring in prophase I can differentiate meiosis from mitosis. Prophase I is divided into five sub-stages based on the cytological landmarks of chromosome structure and movements to leptotene, zygotene, pachytene, diplotene and diakinesis (Zickler and Kleckner, 1998; Zickler, 2006) (summarised in Table 1.1).
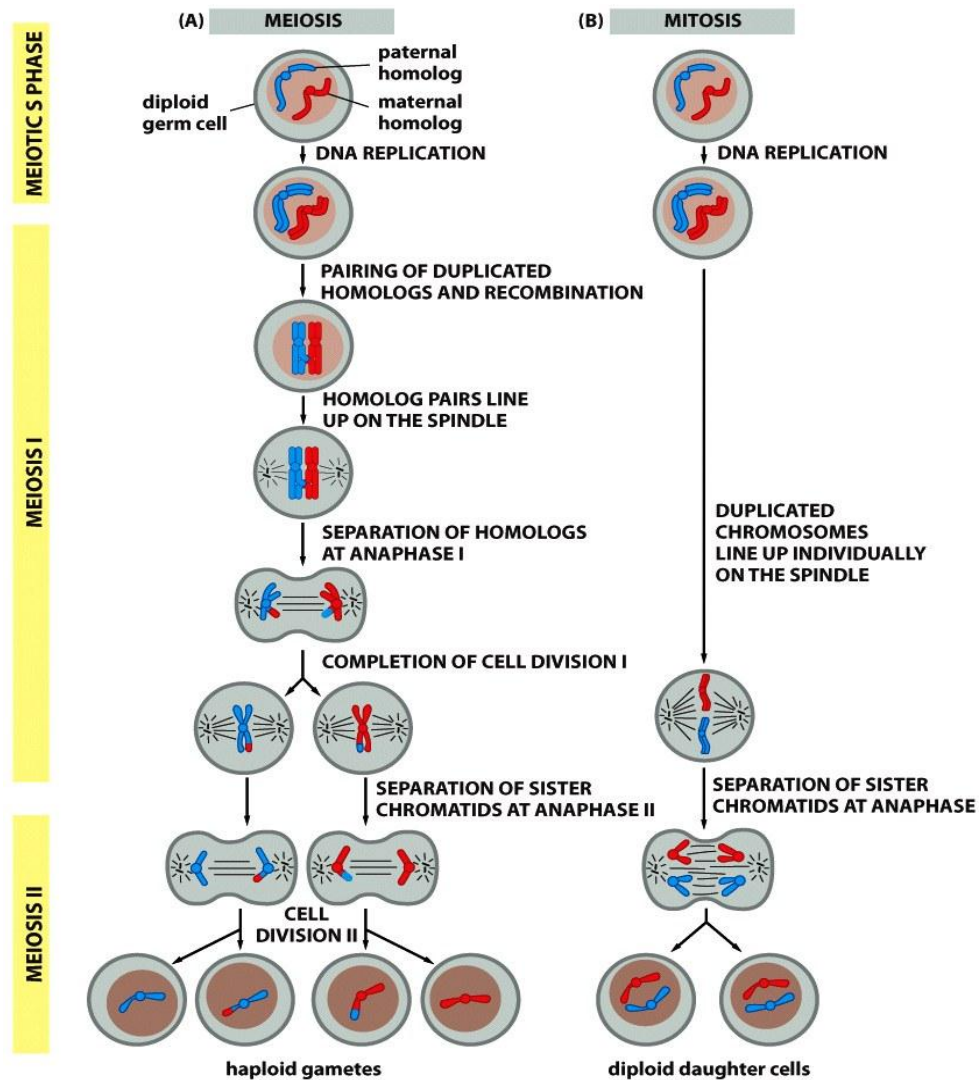
Figure 1.2. Comparison between meiotic and mitotic cell division. Four nonidentical haploid cells are generated during meiotic cell division, whereas two identical diploid cells are produce during mitosis. (A) Meiosis involves two rounds of chromosome segregation: meiosis I (reductional division), and meiosis II (equational division), which are followed by a single phase of DNA replication. (B) Mitosis in contrast, involves one round of chromosome segregation after DNA replication to generate two identical daughter cells (Adapted from Alberts *et al.*, 2013).

Table 1.1. **Summary of the key events occurring in the sub-stages of prophase I.**

Sub- stage of prophase I	Discretion of meiosis events
Leptotene	Chromosomes start to condense and become thin. Telomeres begin to cluster as late leptotene as homologue pairing initiates.
Zygotene	Telomere clustering continues, leading to formation of the "bouquet structure". Homologue pairing is complete and synapsis occurs. Synapsis initiates as a result of the formation of a highly proteinaceous axis component known as the synaptonemal complex (SC).
Pachytene	Synapsis is complete. Reciprocal recombination between homologues forms chiasmata (resolution of which can form crossovers). The resulting cross-linked structure is term a bivalent.
Diplotene	The bivalents begin to separate but remain attached by the chiasmata.
Diakinesis	Chromatids condense further. Centrioles/spindle pole bodies begin to migrate to opposite poles of the nucleus. Bivalents congregate on the metaphase plate.

1.1.2 The main different between mammalian male and female meiosis

Meiosis is a germline-specific division in which haploid gametes sperm and egg are produced from a diploid cell. Although, the meiosis process is similar, crucial differences exist between female and male meiosis. In mammalian females, meiosis initiates before birth during foetal ovarian development. During oocyte development, the homologous recombination pairs remain associated and arrest in prophase I for a long period until ovulation, depending on the species (for example, months in mice or years in humans). Recombination is re-activated in the oocyte at puberty as a result of hormone stimulation (Figure 1.3). The second arrest occurs in metaphase II and requires fertilisation to complete meiosis. The result of this process is only one mature egg and two polar bodies, which afterward degenerate. In the male germline, meiosis occurs at puberty and proceeds continuously; the spermatocytes need about one week to complete meiosis II (reviewed in Von Stetina and Orr-Weaver, 2011; Baillet and Mandon-Pepin, 2012).
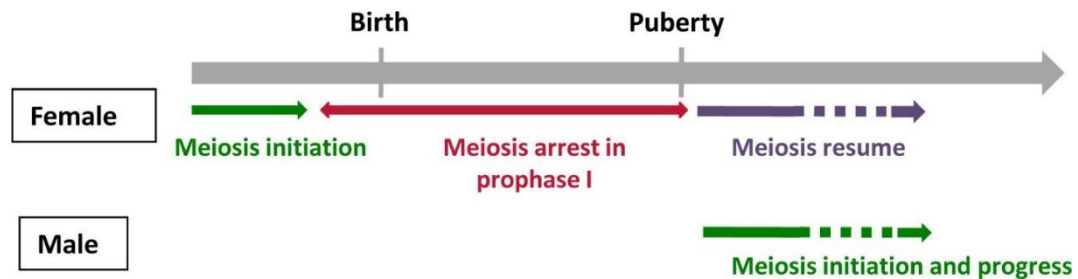
Figure 1.3 Time course comparison between meiosis in both male and female mammals. Meiosis is initiated during foetal development in females with two arrests: one in prophase I and released after puberty as a result of hormonal stimulation; and the second one in metaphase II, which requires fertilization to complete meiosis. In males, meiosis is a continuous process after puberty, occurring without any arrest. (Adapted from Baillet and Mandon-Pepin, 2012).

1.2   Conserved features in meiosis

Many model organisms, such as yeasts (*Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*) and mice, share common molecular mechanism of meiosis many of which appear to be conserved in humans, and they can therefore be used to study the complexity of meiosis. Studying the meiosis mechanism directly in humans is difficult; consequently, many processes remain incompletely understood. However, three main features are conserved in meiosis and are required for successful meiotic division. First, homologous chromosome pairing is essential for initiation of recombination between homologous chromosomes, which leads to genetic information exchange. The physical connections between homologous chromosomes are termed chiasmata and the outcome of this recombination is a reciprocal exchange crossover (CO) between homologous chromosomes (Zickler and Kleckner, 1999). Second, during mitosis and meiosis II, the microtubule spindle of sister kinetochores must attach to the opposite poles, a process known as bi-orientation or bipolar orientation. In contrast, during meiosis I, homologous chromosomes attach to the same spindle pole, known as mono-orientation or monopolar orientation. Third, during meiosis, cohesion between the sister chromatids dissolves in two steps. During anaphase I, cohesin in the chromosome arm is eliminated but centromeric cohesin is protected to hold the sister chromatid for completion of meiosis I. Cohesin complex is completely destroyed in anaphase II (Bishop and Zickler 2004). These events are driven by coordination of a number of meiosis-specific proteins. Some of the key events in meiosis I and aberrant of expression of these genes in germline cells in normal somatic cells could be involved in aneuploidy and cancer development (Bishop and Zickler, 2004; Lindsey *et al.*, 2013; Feichtinger *et al.*, 2014b; McFarlane *et al.*, 2014).

1.2.1   Homologous recombination

Programmed homologous recombination (HR) is an important feature of meiotic prophase I and a key source of genetic information exchange for generation of new combinations of alleles in the population. In addition, HR is source of genomic instability associated with various human diseases (Segurel, 2013). Two distinct outcomes are recognised for the

recombination products: crossover (CO) and non-crossover (NCO), in which at least one obligated crossover event per bivalent is required during meiosis (Baudat *et al*., 2013).

1.2.1.1 Double strand breaks (DSBs)

DNA DSBs are among the most hazardous types of genomic DNA damage, where both strands of the DNA helix break. Several factors are known to induce DSBs, including endogenous factors such as DNA replication problems or DNA damage by agents such as exposure to ultraviolet (UV) light, ionising radiation and chemical modification of the DNA agents (reviewed in Abbotts *et al*., 2014). In addition, DSBs are deliberately initiated in meiosis after DNA replication to maintaining genome integrity and ensure accurate homologous segregation (Longhese *et al*., 2009). Unpaired DSBs are the major source of the genomic instability and chromosome rearrangement. For instance, mismatch repair of the DNA in a germline mutation is associated with one of the most common inherited cancer-predisposition syndromes known as hereditary non-polyposis carcinoma coli (HNPCC) (reviewed in Abbotts *et al*., 2014).

Two major pathways are available for repair of DSBs: nonhomologous end joining (NHEJ) and homology–direct recombination (HR). In mammalian cells, HR mediated DNA repair occurs during the S and G_2 phases of the cell cycle and it tends to use the sister chromatids/homologous chromosomes as a template for the synthesis new DNA at the DSB site. In contrast, the NHEJ occurs in the G_1 phase, during which the DNA ends are recognised and rejoined at the two broken ends, rather than using a homologous sequence to guide DNA repair (reviewed in Lord and Ashworth, 2012).

1.2.1.2 Meiotic recombination

Meiotic recombination is initiated by induction of meiotic DSBs, which are generated by a conserved topoisomerase II-like enzyme protein SPO-11 (Bergerat *et al*., 1997; Keeney *et al*., 1997). SPO11 orthologues are involved in meiotic recombination in all sexually reproducing species including yeasts (known as Spo11 in *S. cerevisiae and* Rec12 in *S. pombe*) and mammals.

Mammals have two main variants of SPO11 proteins: SPO11-α and SPO11-β (reviewed in Metzler-Guillemain and de Massy, 2000; Romanienko and Camerini-Otero, 2000; Bellani *et*

al., 2010). SPO11-β is the larger variant and contains amino acids encoded by an additional exon. It is proposed to be responsible for the generation of the majority of DSBs because it is expressed in early prophase I when the majority of the DSBs occur. SPO11-α is produced in late prophase I at the beginning of the early pachytene (Bellani *et al*., 2010). The specific function of SPO11-α remains elusive; however, the absence of SPO11-α is associated with a reduction in the number of DSBs in the sex chromosome in male mice (Kauppi *et al*., 2011).

SPO11 acts as a dimer and catalyses breakage of both strands of the DNA molecule and remains covalently linked at the 5' end through a tyrosine that is conserved in SPO11 orthologues. The SPO11 is removed through the association of several proteins such as the MRN complex (MRE11-RAD50-NBS1) in mammals and fission yeast, or the MRX complex (Mre11-Rad50-Xrs2) in budding yeast. The outcome is a DSB with a 3' single-stranded DNA (ssDNA) overhang (Figure 1.4) (reviewed in de Massy, 2013). An involvement of the MRN/MRX complex proteins has been identified in the activation of the DNA damage checkpoint and telomere maintenance in both meiotic and mitotic cells (Longhese *et al*., 2010). The nuclease ExoI and the BLM (Bloom syndrome protein) are thought to mediate DSB resection from the 5'-3' direction, away from the DSB, whereas Mre11 endonucleases resect from the 3'-5' direction towards the DSB end (Garcia *et al*., 2011; Shibata *et al*., 2014). The ssDNA-binding protein, or protein A (RPA) plays an important role in the affinity and binding of Rad51 to ssDNA. Strand invasion into a homologous form is initiated by loading of Rad51 and DMC1 (in meiosis) recombinases to create nucleoprotein filaments (reviewed in Holthausen *et al*., 2010). The meiosis-specific RAD51 paralogue DMC1, which is a member of the recombinase Rad51/RecA family, also associates with Rad51 to search for a homologous pair template to form the dissociation loop (D-loop) (Kagawa and Kurumizaka, 2010; Dray *et al*., 2011). A number of other factors including the breast and ovarian cancer-related proteins (BRCA1 and BRCA2), as well as the PALB2 protein, are involved in the formation of Rad51 nucleoprotein filaments (Yu *et al*., 2001; O'Donovan and Livingston, 2010).

Double Holliday junctions (dHJs) are then formed from the result of ssDNA invasion and ligation to the other complementary strand, with association of different proteins such as Rad52 (reviewed in Lok and Powell, 2012) and Rad54 (reviewed in Mazin *et al*., 2010). The

cleavage of dHJs generates either crossovers (COs) or non-crossovers (NCOs), depending on the enzymatic cleavage orientation (Schwacha and Kleckner, 1995). In *vitro*, three endonuclease enzymes can resolve the dHJs including Slx1-4(BTBD12), Mus81-Mms4 (EME1) complex and GEN1. However, *in vivo,* the enzymes responsible for creating dHJs remain poorly understood. For example, in *S. cerevisiae*, the majority of dHJs are resolved by EXO1, Mlh1-Mlh4 and Sgs1 (Zakharyevich *et al*., 2012).

1.2.1.3   Initiation of DSB Hotspots

Recombination events are not randomly distributed among the genome; in fact, they appear to be concentrated only in specific segments, termed hotspots. In humans, several studies have shown that the initiation of recombination occurs in narrow sites approximately 1 to 2 kb long (Jeffreys *et al*., 2001). These specific sites are marked by trimethylation of lysine 4 of Histone H3 (H3K4) (Borde *et al*., 2009; Buard *et al*., 2009; Smagulova *et al*., 2011). In mice and humans, the PR-domain containing protein 9 (PRDM9; known as Meisetz in mice) has been identified to have a critical role in regulating recombination by binding to these hotspots (Baudat *et al*., 2010; Myers *et al*., 2010; Parvanov *et al*., 2010).

In mammals, PRDM9 is a zinc finger protein belonging to a large family of PRDM proteins (this family contains 16 members in mice and 17 in humans). It consists of three main domains: a PR/SET domain (which confers the activity of histone methyltransferase); a C2H2 zinc finger repeat domain (reviewed in Fumasoni *et al*., 2007; Baudat *et al*., 2013) and a KRAB (Kruppel-association box) protein-protein binding domain, which is normally found in transcriptional repressors (Figure 1.5 A) (Birtle and Ponting, 2006; Oliver *et al*., 2009). It is the only member of this family with this domain, other than the paralogous protein PRDM7, which is encoded by a duplicate of PRDM9 (Fumasoni *et al*., 2007).

The production of the zinc finger protein PRDM9 is restricted to germline cells and encodes a histone methyltransferase activity at the PR/SET domain for trimethylation of lysine 4 of Histone H3 (H3K4) in mice (Hayashi *et al*., 2005). Recent studies found this protein also has high mono-, di- and trimethylation activity of lysine 4 of Histone H3 (H3K4) (Wu *et al*., 2013; Eram *et al*., 2014) as well as trimethylation activity of lysine 36 of Histone H3 (H3K36) activity (Eram *et al*., 2014). Furthermore, PRDM9 was found to methylate the other core histones,

H2A, H2B and H4, although to a lesser degree than H3. A single C321P mutation in the PR/SET domain led to weak activity of PRDM9 through inhibition of interaction with S-Adenosylmethionine (SAM) (Wu *et al.*, 2013; Xiaoying *et al.*, 2014). In addition, the zinc finger protein has a degenerate 13-bp motif, CCNCCNTNNCCNC, which associated with DSB locations in human hotspots (Myers *et al.*, 2010). A mouse study showed that *Prdm9/Meisetz* is expressed in the early stages of meiosis and that disruption of this gene leads to hybrid sterility in both sexes resulting from a severe deficiency in the DSB repair pathway, deficient pairing of homologues and failure of sex body formation (Hayashi *et al.*, 2005). The zinc finger domain contains tandem repeats; gene sequencing of this domain in 20 mouse strains identified five different zinc finger repeats responsible for binding different DNA specific sites (Parvanov *et al.*, 2010).

The molecular mechanism of hotspot activation remains poorly understood. However, the zinc finger protein is proposed to bind to the specific DNA motif via the C2H2 zinc finger domain (Figure 1.5 B). Subsequently, the PR/SET domain activates H3K4me3 on adjacent nucleosomes. The KRAB may play a critical role in interactions with other proteins for inducing DSBs by the SPO11 protein (reviewed in Parvanov *et al.*, 2010; Baudat *et al.*, 2013).
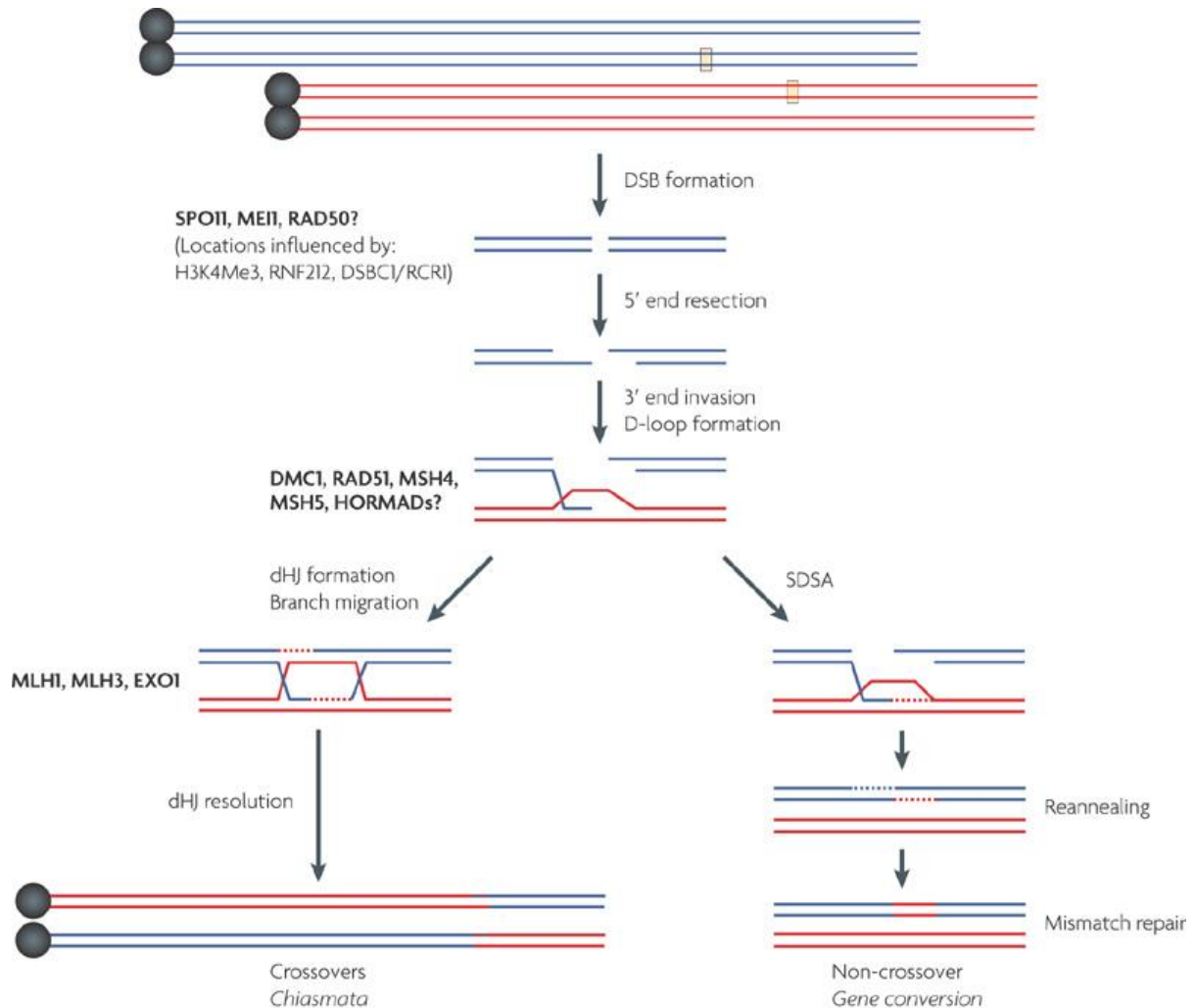
Figure 1.4. **Schematic diagram of meiotic recombination through the DSB pathway.** Two main recombination pathways are presented: the non-crossover (NCO) pathway on the right and the crossover (CO) pathway in the left. The DSB is initiated by cleavage of both strands of the DNA by SPO11 at the 5' ends of the DNA, followed by removal of SPO11 by MRN/MRX protein complex to generate 3' single strand DNA (ssDNA) overhangs. Then, a single 3' end overhang invades the other homologue to initiate a D-loop; DNA synthesis then occurs, using the other homologues as a template. This formation is unstable and the decision is made in this step to generate crossover (CO) or non-crossover (NCO). In the CO pathway, double holiday junctions (dHJs) can be resolved to generate crossover or dissolved to produce non-crossover depending on the enzymatic cleavage orientation. Otherwise, non-crossover can occur as a result of the D-loop dissociation and invasion of the opposite end of the original strand. This pathway is termed synthesis dependent strand annealing (SDSA), which leads to gene conversion. (Adapted from Handel and Schimenti, 2010).

(A)

(B)
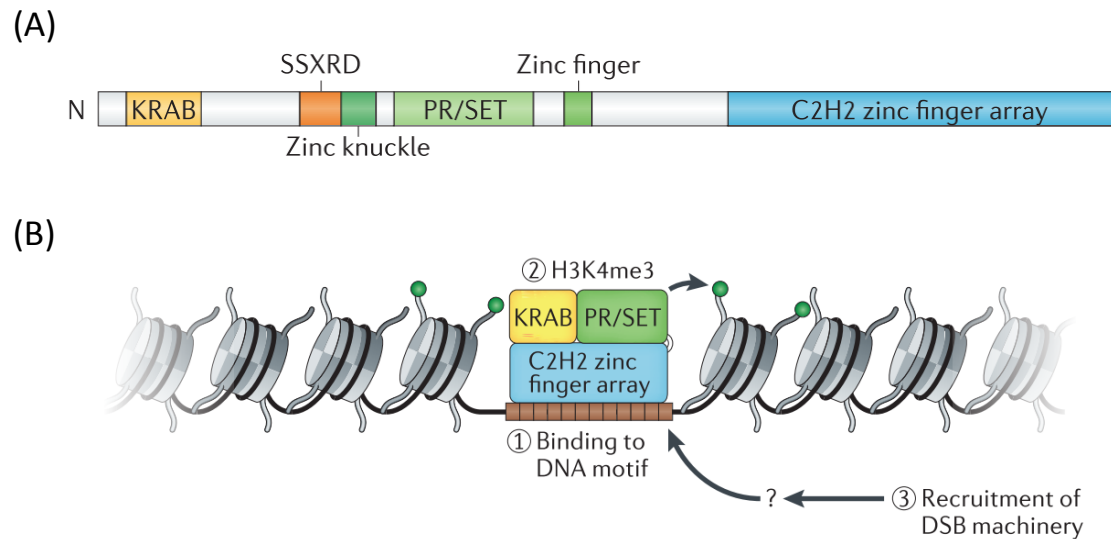
Figure 1.5. **Schematic showing the PRDM9 structure and the mechanism of hotspot activation. (A)** PR domain-containing protein 9 (PRDM9) is a histone methyltransferase composed of three domains: KRAB (Kruppel-associaton box) containing SSXRD (synovial sarcoma X breakpoint domain); PR/SET domains encoding a histone methyltransferase activity and a C2H2 zinc finger domain. **(B)** The proposed mechanism of PRDM9 binding to the specific DNA motif via the C2H2 zinc finger domain. Subsequently, the PR/SET domain activates H3K4me3 on adjacent nucleosomes. The KRAB (Kruppel-association box) may play a critical role in interaction with other proteins for recruitment of the DSBs by the SPO11 protein (Adapted from Baudat and Frédéric, 2013).

1.2.2 The synaptonemal complex

The synaptonemal complex (SC) is a large, proteinaceous, zipper-like structure. It is involved in the mediation of the stable connection between the two homologous chromosomes (synapsis) during prophase I of meiosis. The SC is known to have a critical function in the formation the crossover during meiosis I. It has been proposed to play an important roles in CO interference in a number of orgasims but its actual role is not well understood. However, failure of SC assembly can lead to infertility and production of aneuploid germ cells in mammals (Handel and Schimenti, 2010; Fraune *et al*., 2014). In females, oocyte aneuploidy can be caused by a defect in synapsis and this can lead to pregnancy loss (Fraune *et al*., 2012; 2014). Three highly conserved domains of the SC are known: (1) the lateral elements (LEs), (2) the transverse filaments (TFs), and (3) the central element (CE)(Figure 1.6) (Page and Hawley, 2004).

During the early stage of the leptotene, the homologous chromosomes pair to form the axial elements (AEs) along each chromosome. The two AEs of the homologues are then known as lateral elements (LEs) at zygotene, when the homologues are fully synapsed (Fraune *et al*., 2012). The main known structural proteins specific to the LEs are SYCP2 (Offenberg *et al*. 1998) and SYCP3 (Lammers *et al*., 1994). An involvement of cohesion complex proteins (cohesin cores), including REC8, STAG3 and SMC1β proteins, is proposed to facilitate the formation and localisation of AEs/LEs (Page and Hawley, 2004). Mutations in the proteins involved in AE formation lead to disruption of sister chromatid cohesin, recombination and chromosome segregation (Shin *et al*., 2010). HORMA-domain proteins (HORMAD1 and HORMAD2) are critical component of AEs and preferentially localise to the unsynapsed chromosome axis to facilitate SC formation. In the mouse, HORMAD1 is thought to ensure that sufficient unpaired DSBs are available for successful homology searches, to establish a local alignment between homologous chromosomes and to facilitate CO formation (Daniel *et al*., 2011).

The second structure of the SC is the transverse filament (TF) which consists of large coiled-coil homodimers. The TFs are thought to serve as bridges between the two LEs and the central element (Figure 1.7). SYCP1 is the only TF protein known; it consists of two globular heads associated head-to-head with CE via the N-terminal; the C-terminal interacts with the LEs (Liu *et al.*, 1996) .

The third structure of the SC is the central element (CE), which is composed of four known proteins: SYCE1 and SYCE2 (Costa *et al.*, 2005), SYCE3 (Schramm *et al.*, 2011) and TEX12 (Hamer *et al.*, 2006). In mammals, SYCE1 and SYCE3 co-localise in the CE and interact with SYCP1. This distribution of SYCE1 and SYCE3 leads to failure of SC initiation (Costa *et al.*, 2005; Hamer *et al.*, 2006; Schramm *et al.*, 2011). In contrast, SYCE2 associates with TEX12 to form a complex that indirectly interacts with SYCE1 (Hamer *et al.*, 2008; Davies *et al.*, 2012). These finding suggest that SYCE1 and SYEC3 have a role in the initiation of synapsis (Schramm et al., 2011), whereas, SYCE2 and TEX12 have a role in SC extension (Davies *et al.*, 2012).
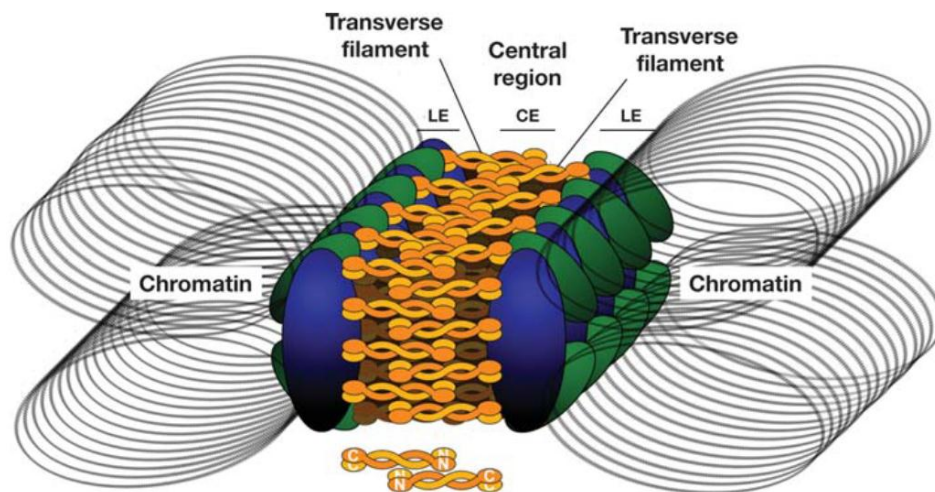
Figure 1.6. **Schematic showing the structure of the synaptonemal complex (SC).** SC is composed of two lateral elements (LEs), which are associated with the central element (CE) via transverse filaments (TFs) (Adapted from Page and Hawley, 2004).
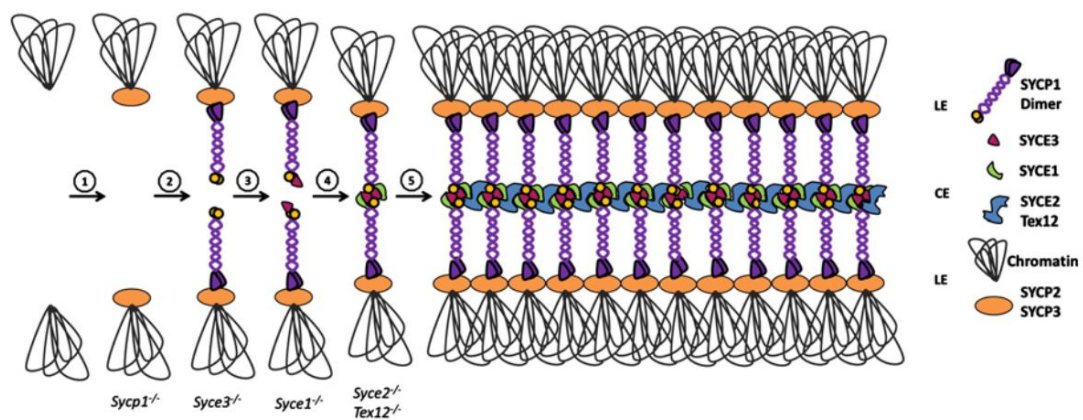
Figure 1.7. A model of the synaptonemal complex (SC) assembly. The two lateral elements (LEs) are initiated from the axial elements (AEs) composed of SYCP2, SYCP3, cohesin cores and HORMA-domain proteins. Transverse filaments consist of SYCP1 and form a bridge between the two LEs and the central element (CE). CE is made up from SYCE1, SYCE2, SYCE3 and TEX12. (Adapted from Fraune *et al*., 2012)

1.2.3 The cohesin complex

Cohesin is a multiprotein structure that holds two sister chromatids together in its tripartite ring structure from the S-phase until the metaphase-to-anaphase transition. It is essential for faithful chromosome segregation in meiotic and mitotic cell division. Cohesin plays a fundamental role in the life of the cell, including generation sister chromatid cohesion (SCC) (Mehta *et al.*, 2012) and formation and repair of DNA double-strand breaks (DSBs) in both meiosis (Kim *et al.*, 2010) and mitosis (Sjögren and Nasmyth, 2001). It is also involved in the assembly of the AEs of the synaptonemal complex during meiosis (Kim *et al.*, 2010; Klein *et al.*, 1999) and acts as a transcriptional regulator in yeast (Lin *et al.*, 2011). Recently, mutation of cohesin subunit *SA2* (*STAG2*) in mice has confirmed an association with aneuploidy and tumourigenesis (Remeseiro *et al.*, 2012). Furthermore, mutations of STAG2, RAD21, SMC1 and SMC3 of cohesin genes have been identified in acute myeloid leukaemia (Welch *et al.*, 2012). They also linked to other genetic diseases referred to as cohesinopathies (Skibbens *et al.*, 2013).

The cohesin complex is made up of two main structures: (1) two members of the structural maintenance of chromosome (SMC) family, SMC1 and SMC3, and (2) klesin proteins (Figure 1.8). SMC1 and SMC3 are conserved from yeast to humans and are rod-shape proteins with a coiled coil flanked globular domain with an ATPase activity at one end and a dimerisation domain at the other end (Gruber *et al.*, 2003). The dimerisation domain "hinges" of SMC1 and SMC3 are tightly associated with each other and form a V-shaped structure. The ATPase head of these proteins is associated with the SCC1/RAD21 α-klesin subunits. The fourth domain of cohesin is SCC3/SA, which interacts with SCC1/RAD21 as well as Pds5 and WapL subunits (reviewed in Mehta *et al.*, 2012).

In mammals, two types of SMC1 are identified: SMC1α that is expressed during mitosis and SMC1β that is expressed in meiosis. The α-klesin subunits are also composed of three subunits: SSC1/RAD21 in mitosis or RAD21L and REC8 in meiosis. The α-klesin binding subunit SCC3/SA has two homologs: stromalin antigens termed SA1/STAG1 and SA2/STAG2 in mitosis, whereas in meiosis these are defined as SA3/STAG3. The Pds5 protein is

composed of two subunits: PDS5A and PDS5B, so these create approximately 18 possible combinations for the formation of the cohesin complex (Nasmyth, 2011).

The mechanism of creating and triggering cohesin during meiosis and mitosis is not fully understood; however, four different models have been proposed for the cohesin ring: (1) a one ring (embrace) model (Figure 1.9 A); (2) a two ring model (Figure 1.9 B); (3) a multimeric bracelet (Figure 1.9 C); and (4) a multimeric rod shaped model (Figure 1.9 D). The simplest model and the most popular is the one ring (embrace), which explains how cohesin holds two sister chromatids together with a single ring in terms of known protein-protein interactions (Mehta *et al.*, 2012; Nasmyth, 2011).
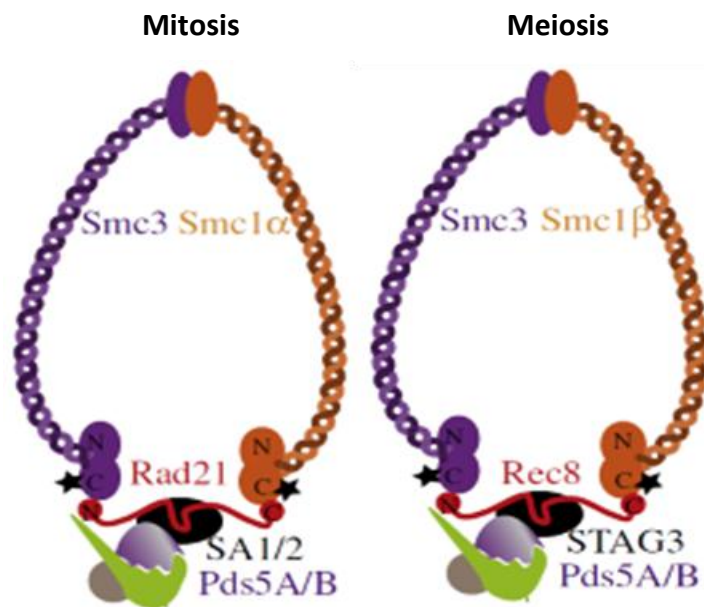
Figure 1.8. **Schematic composition of the cohesin complex structure in meiosis and mitosis**. Two members of the SMC protein family—SMC1α in mitosis and SMC1β in meiosis (brown)—as well as SMC3 (purple) are involved. The α-klesin subunits bind to SMC1 and SMC3 heterodimers to form a V-shaped structure of cohesin. (Adapted from Mehta *et al.*, 2012).
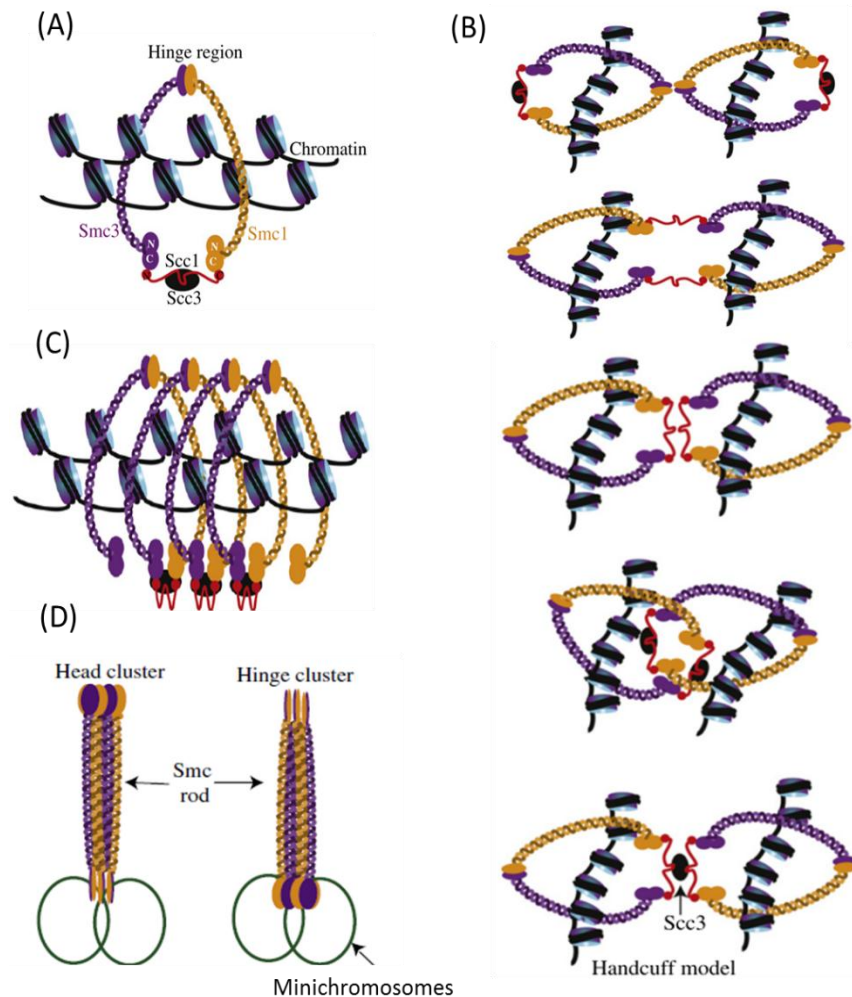
Figure 1.9. **Schematic composition of four ring models proposed for the sister chromatids holding by cohesin**. **(A)** Embrace model where the cohesin ring holds the sister chromatids by one ring. **(B)** The two ring model entraps the sister chromatids with five possibilities and the handcuff model is more popular due to experimental evidence supporting this model.**(C)** The bracelet model suggests that the sister chromatids are linked with cohesin rings by oligomerisation in hand to hand fashion. **(D)** The Rod model suggests that the cohesin rings are linked with each other by their SMC rod region to form a rod thick structure. (Adapted from Mehta *et al*., 2012).

1.3   Errors in meiosis and mitosis

Genome integrity is based on the faithful segregation of duplicated chromosomes to generate daughter cells during cell division. During meiosis, failures in chromosome pairing and /or mis-segregation lead to a non-disjunction state that is associated with aneuploid gametes. Aneuploidy in gametes is a condition in which the chromosome number is abnormal: gametes carry an extra copy of the chromosome (termed trisomy) or one copy (known as monosomy). In humans, aneuploidy is the most commonly type of chromosome abnormality found to be associated with birth defects, infertility and miscarriage. Nearly one-third of miscarriages are aneuploid and approximately 10-30% of fertilised human eggs are aneuploid. Down's Syndrome, for instance, is a result of trisomy of chromosome 21. In contrast, monosomies are lethal: no monosomies during foetal development are viable and therefore lead to spontaneous abortions (reviewed in Hassold and Hunt 2001; 2007).

In mitosis, faithful segregation of sister chromatids rely on the accurate attachment of the microtubules from opposite poles to the sister kinetochores. Defects and/or mis-attachment of the microtubules to sister kinetochores lead to mis-segregation of the chromosomes, which are hallmarks of aneuploidy (Gregan *et al.*, 2011). A high level of chromosome mis-segregation is known as chromosome instability (CIN), which is a common feature of cancer. Most solid tumours are aneuploid, with a very high level of mis-segregation (reviewed in Thompson *et al.*, 2010). The expression of meiosis-specific-genes is restricted to germ cell lines and aberrant expression of these genes is found in cancer cells for different genes, including the meiotic hotspot activator gene *PRDM9*, the meiotic regulator gene *STRA8* and cohesion complex genes *SMC1β* and *RAD21L.* It has been postulated that aberrant meiotic gene expression in somatic cells might derive aneuploidies (Simpson *et al.*, 2005; Feichtinger *et al.*, 2012a).

1.4 Cancer

Cancer is one of the main health problems and is the leading cause of death after cardiovascular diseases in the developed world (Jemal *et al.*, 2011). More than 100 forms of cancer are known and are primarily classified according to the original tissue (Stratton *et al.*, 2009). Based on the statistics of GLOBOCAN in 2008, approximately 7.6 million deaths from cancer and 12.7 million cancer cases occurred worldwide. The annual number of deaths caused by cancer is projected to increase to more than 11 million by 2030.

Several factors are associated with the appearance and development of cancer, including exogenous and endogenous factors. Most (90-95%) cancers are attributed to exogenous factors, such as the environment and lifestyle. The lifestyle factors include exposure to UV radiation and chemical carcinogens, smoking, obesity, diet, alcohol consumption and infectious organisms (Anand *et al.*, 2008). A recent study found that gene mutations detected in lung cancers were 10-fold higher in smokers than in non-smokers (Govindan *et al*. 2012). In contrast, only 5-10% of cancer cases are associated with genetically inherited defects (Anand *et al.*, 2008).

Cancers can be defined as a complex of genetic diseases, leading to uncontrolled cell division, decreased cell death, invasion and destruction of adjacent tissues and spread (metastasis) to other organs of the body. Tumours are initiated in a multistep process resulting from the genetic mutation of DNA sequencing and/or epigenetic alteration (Michor *et al.*, 2005; Stratton *et a*l., 2009; Tomasetti *et al.*, 2013).

1.4.1 Tumourigenesis development and hallmarks of cancer

Normal cells transform into a carcinogenic condition in a multistep process, where epigenetic and genetic alterations are accumulate. Carcinogenesis exhibits complex phenotypic features that are easy to distinguish from those of normal tissues. One of the most dangerous transformations of cancer is metastasis. In this process, cancer cells from a primary tumour become detached and initiate a local invasion of stromal tissue. The cells then spread via the blood and lymphatic vessels (intravasation) to translocate and survive within the bloodstream. Cancer cells then escape from bloodstream to penetrate the parenchyma (extravasation), where they then form small nodules term micrometastases (Figure. 1.10).

The growth and proliferation of micrometastases in a suitable tumour microenvironment creates "colonies" where the cells divide to initiate the new tumours (reviewed in Bacac and Stamenkovic, 2008; Chaffer and Weinberg, 2011; Hanahan and Weinberg, 2011).

Hanahan and Weinberg (2011) have proposed several features of cancer cells that differentiate them from normal cells, and they term these hallmarks of cancer. They proposed that cancer cells have the ability to achieve immortality through replication, evasion of growth suppressors, resistance to apoptosis, enhancement of proliferative signalling, sustained angiogenesis, and the ability for tissue invasion and metastasis (Figure 1.11). In addition, these cells can evade immune destruction and can reprogram energy metabolism including genomic instability and mutation, as well as promote inflammation (Hanahan and Weinberg, 2001; 2011).

1.5   The role of oncogene and tumour suppressor genes in tumorigenesis

In normal cells, the process of cell growth and differentiation are highly regulated and tightly controlled. Otherwise, loss of cell proliferation is one of the cancer cell hallmarks. Mutations in three classes of genes have been identified to be involved in tumourigenesis: (1) tumour-suppresser genes; (2) oncogenes; and (3) genomic stability genes including genes involved in DNA damage and repair, chromosomal segregation and recombination (reviewed in Vogelstein and Kinzler, 2004; Negrini *et al*., 2010; Hanahan and Weinberg, 2011).

1.5.1 Tumour suppressor genes

Tumour suppressor genes (TSGs), or anti-oncogenes, are genes that encode the proteins required to inhibit cell growth or induce cell death. Several types of tumour suppressors have been identified through their characteristics in different types of cancers, such as *p53* (TP53), which is considered the guardian of the genome, and *RB* (retinoblastoma-associated) genes. These genes lose or inactivate their function via genetic mutation, DNA methylation and loss of heterozygosity (LOH) (reviewed in Thoma *et al*., 2011). One of the essential hypotheses of tumour suppressor function loss is the "Two-hit hypothesis" (Knudson 1971), which suggests that two alleles have to lose their function in order to cause cancer (Figure 1.12). One

mutation is inherited via the germinal cells and the other through somatic cells. For example, the mutation of the *RB1* gene is presented in various types of cancer, such as lung, breast and eye. The RB protein acts as a checkpoint in the cell cycle, controlling entry of the cell into the S-phase. This protein mainly binds to the transcription factors family E2F. A mutation can affect the binding of transcriptional factors (E2F1-E2F2) with the protein dimerise (DP) complex in the RB pathway (reviewed in Chen *et al*., 2009; Thoma *et al*., 2011).

In addition, the guardian of the genome, TP53, tumour suppressor protein is important for regulating the cell cycle checkpoint, and for activating the senescence and apoptosis programmes. Loss or mutation of the *p53* gene has been found in about half of human cancers. It is involved in metastatic processes such as cell migration and invasion (Muller *et al*., 2011).

1.5.2 Oncogenes

The term "oncogene" refers to a group of genes capable of causing cancer by stimulating cell proliferation. Proto-oncogenes are normal cellular genes that encode protein enhanced cell growth by stimulating cell division, preventing cell differentiation and impairing cell death. Mutations, chromosomal translocations and over-expression can lead to these genes becoming oncogenes. However, on their own, they do not have the ability to cause cancer but instead act in collaboration with other oncogenes or with the loss of cellular tumour suppressor genes. Oncogenes encode specific cellular proteins that have important roles in apoptosis, chromatin remodelling, signal transduction, transcriptional factor induction, growth regulation, growth factor receptor responses or hormone receptor responses (reviewed in Croce, 2008). For instance, RAS subfamilies are a large group of small GTP-binding proteins that are involved in cell signalling, growth regulation and survival. They are able to activate and inhibit the growth factor signalling via tyrosine kinase receptor pathways. A single mutation in the RAS oncogenes (KRAS, HRAS and NRAS) occurs in 30% of human cancers (Croce, 2008; Pylayeva-Gupta *et al*., 2011).

In addition, viral oncogenes (V-onc) can initiate and promote the development of several types of cancer resulting from persistent virus infection. Viral oncogenes can mutate pro-oncogenes by inserting their promoters into the host's chromosomal DNA (e.g., in the

enhancer or promoter), which then leads to activation of the transcription factor genes. Over-expression of these genes commonly alters expression of tumour suppressor genes (reviewed in Ranzani *et al.*, 2013). For example, mutations of the sarcoma or *src* oncogenes are driven by retroviral Rous Sarcoma Virus RSV (reviewed in Vogt, 2012).

1.5.3   Genome stability genes

Genome stability genes (caretaker genes) are a group of genes involved in repairing of mistakes in the DNA initiated either during DNA replication or after induction by exposure to mutagens (reviewed in Vogelstein and Kinzler 2004; Negrini *et al.*, 2010). In addition, this group of genes is associated with meiotic recombination and chromosomal segregation, and includes genes such as ataxia telangiectasia protein (*ATM*), breast cancer susceptibility 1 (*BRCA1* and *BRCA2*), and Nijmegen breakage syndrome 1 (*NBS1*) which are mutated in various types of cancer including breast cancer, leukaemias, lymphomas and ovarian cancer (reviewed in Negrini *et al.*, 2010).

 This group of genes keeps the rate of genetic alterations to a minimum level; consequently, abnormality of expression of these genes is found to increase the rate of alteration of other genes (reviewed in Vogelstein and Kinzler 2004; Negrini *et al.*, 2010).
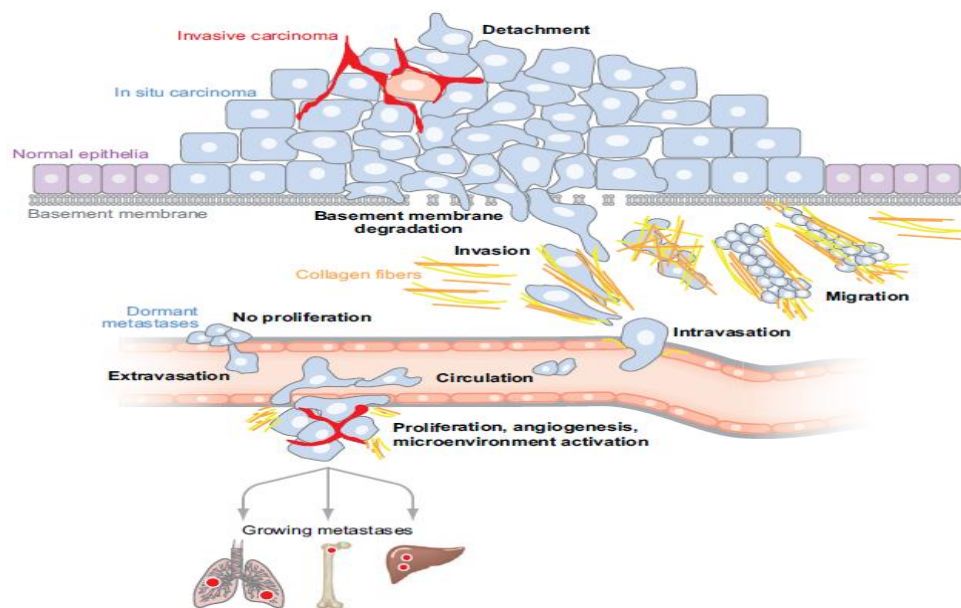
Figure 1.10. Steps of metastasis. Starting with primary tumour, breaks in the basement membrane in epithelial tissues results in loss adherens junctions. Cells then travel via the bloodstream to penetrate the parenchyma to form a cancer microenvironment (Adapted from Bacac and Stamenkovic, 2008).
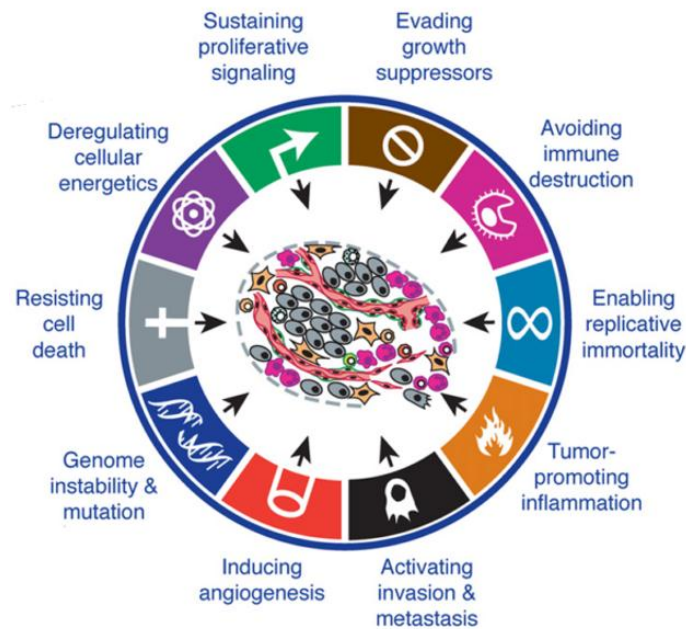
Figure 1.11. **A model of cancer hallmarks proposed by Hanahan and Weinberg.** They suggested that cancer cells have the ability to achieve immortality of replication, evasion of growth suppressors, resistance to apoptosis, promotion of proliferative signaling, sustained angiogenesis, the ability for tissue invasion and metastasis. In addition, the cells can evade immune destruction and reprogram energy metabolism including genomic instability and mutation, and promote inflammation. (Adapted and modified from Hanahan and Weinberg, 2011).
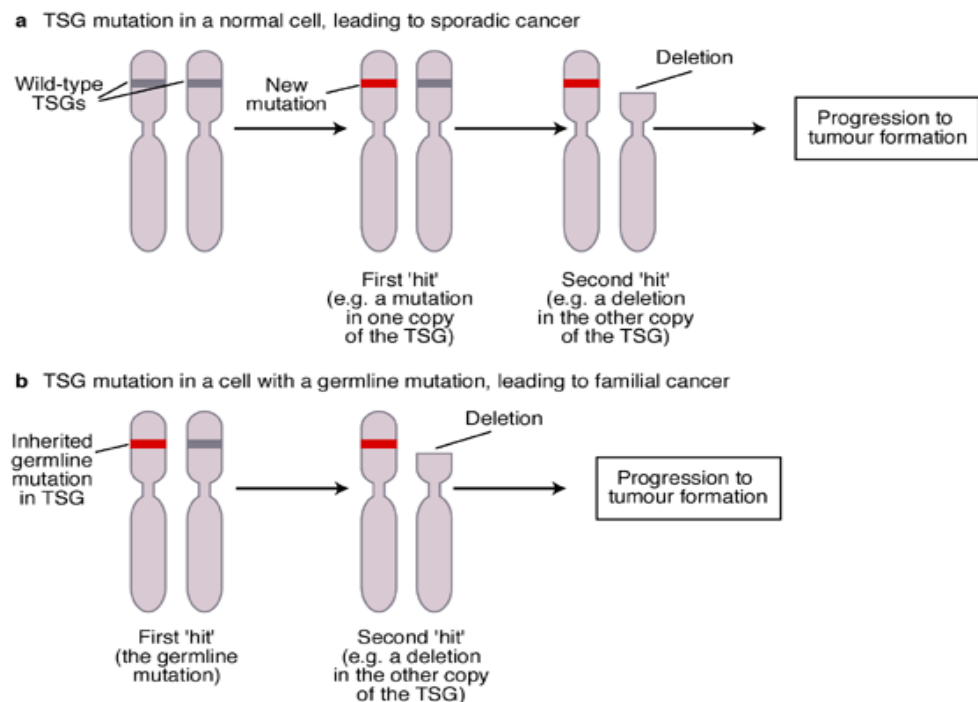
Figure 1.12. **A model of the two-hit hypothesis proposed by Knudson for involvement of tumour suppressor genes in tumourigenesis.** (a) A normal cell has two copies of TSG, so to develop cancer, two individual "hits" of mutation are required. (b) The normal cell has already one "hit" in TSG resulting in a germline mutation, so only one hit is required to initiate cancer. (Adapted from expert reviews in molecular medicine website 2001; http://www-ermm.cbcu.cam.ac.uk).

1.6 Classification of Tumour-associated Antigens (TAAs)

Alterations in gene expression and sequence, such as mutation of proto-oncogenes, tumour suppressor genes and instability genes, and epigenetic changes can all lead to the abnormal production of proteins. These proteins are known as tumour-associated antigens (TAAs). TAAs have the ability to induce the immune system by production of a single epitope (peptide) recognised by the immune system. To be useful for cancer immunotherapy, TAAs must be: (A) restricted to the tumour but not expressed in healthy cells; (B) be present stably and homogenously in most detected tumours; (C) targeted by cytotoxic T lymphocytes (Krishnadas *et al.*, 2013). The identification and characterisation of TAAs has become a large area of research in cancer immunotherapy. Human TAAs can be classified into different types according to their expression patterns:

1. **Viral antigens:** Several types of tumours are caused by virus infections that promote cells to produce viral proteins that can be recognised by the immune system. For example, human papillomavirus (HPV) antigen is associated with development of cervical cancer and the Epstein-Barr virus (EBV) antigen is expressed with Hodgkin's and Burkitt's lymphoma (Vogt, 2012).

2. **Differentiation antigens:** These antigens have been identified as being expressed in malignant and normal cells of similar lineages. For example, tyrosine, MART1/MelanA and gp100 glycoprotein are expressed normally in melanocytes, but are also found in melanoma cancer. This protein can be a target for the immune system, which can eradicate the tumour (Barrio *et al.*, 2012; Kozłowska *et al.*, 2013).

3. **Overexpressed antigens:** Compared to normal tissues, several genes are identified as being either down-regulated or up-regulated in cancer tissues. Peptides derived from over-expressed proteins 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 proteins. For example, overexpressed antigens, such as transmembrane mucin MUC1, is associated with many carcinomas (Singh *et al.*, 2006; Kaur *et al.*, 2014).

4. **Cancer testis antigens (CTAs):** This group of genes encode proteins with expression restricted to human germ lines and malignant tumours. Because of their unique

expression pattern, CTAs are becoming useful targets for cancer vaccines, biomarkers and immunotherapy (Krishnadas *et al*., 2013; Whitehurst, 2014).

1.7   Modalities cancer treatment

Cancer is diagnosed and treated in the late stages when the cancer cells invade the body and turn into metastatic cells. Treatment can usually include surgery, chemotherapy, radiotherapy or a combination of these treatments (Suri, 2006; Aly, 2012). These treatments offer some advantages to the patients and may be associated with negative side effects as well as failure to completely target all cancer cells. Surgery is the best method only for solid tumours (Aly, 2012). Unfortunately, use of radiotherapy can cause many side effects including anorexia, fatigue, nausea, changes in taste and increased risk of microbial infection as well as skin irritation. Furthermore, long-term radiotherapy can lead to loss of memory, endocrine dysfunction, incontinence and gait dysfunction (Eichler and Plotkin, 2008; Aly 2012).

Chemotherapy treatment can also cause many side effects such as anorexia, emesis, fatigue, nausea, loss of hair, nosebleeds, diarrhoea, stomach ache and swelling, as well as loss or gain of weight (Carelle *et al*., 2002). Since cancer cells multiply more rapidly than normal cells, they are more susceptible to chemotherapy drugs than normal cells. However, chemotherapy does not specifically target cancer cells but instead targets all dividing cells, including normal cells that are multiplying at some point. In addition, chemotherapy resistance can arise from different mechanisms including constant mutation of cancer cells, which leads to failure of patient treatment (Rivera and Gomez, 2010).

The immune system of cancer patients normally fails to fight the rapid growth of tumours and therefore development of new immunotherapeutic strategies for early identification, diagnosis and treatment of cancer is a fundamental goal of cancer immunotherapies. Cancer immunotherapy has an edge over other conventional treatments because it provokes the immune system of the patient to fight cancer cells. In addition, it has fewer side effects and specifically targets (e.g., TAAs) (Aly, 2012; Pardoll, 2003).

The immune system is able to recognise and distinguish between self and non-self antigens. In fact, the main goal of immune therapy is to enhance the patient's immune system to

attack and destroy cancer cells with the help of their T-cell, B-cell and natural killer cells (Pardoll, 2003; Mellman *et al*., 2011; Aly, 2012; Harris and Drake, 2013). Based on this concept, one important aim is to identify cancer specific biomarkers/targets for clinical and immunotherapy approaches. The CTA genes represent an excellent immunological target for clinical approaches such as adaptive cell therapy (see Section 1.9).

1.8 Cancer testis antigens

The CTA genes are a large family of genes that encode proteins normally found in normal male germ cells, the placenta and trophoblasts, but not in normal somatic cells (reviewed in Whitehurst, 2014). These genes are expressed in various tumour types at high frequency in ovarian cancer, bladder cancer, lung cancer and melanomas, whereas they are expressed at low levels in colorectal cancer, renal cancer and lymphoma/leukaemia (reviewed in Simpson *et al*., 2005; Fratta *et al*., 2011). The testes are an immunologically privileged site because of the tight junctions occurring between adjacent Sertoli cells that form the blood–testis barrier (BTB). This barrier prevents autoimmune reactions by diffusing the large molecules that carry in blood into the lumen of the seminiferous tubule (Mirandola *et al*., 2011). In addition to the BTB, testicular germ cells also lack the major histocompatibility complex MHC class I antigens, implying that their unique expression pattern could make them good targets for cancer therapy (Caballero and Chen, 2009).

1.8.1 Identification of CT antigens

In the early 1960s, the first alpha-fetoprotein was described as a serum marker for germ cell and hepatoma cancers (Abelev *et al*., 1963). Subsequently, a serum marker for epithelial and colon cancer (carcinoembryonic antigens, CEA) was discovered (Gold and Freedman, 1965). However, in 1991, van der Bruggen and his colleagues discovered the first CTA, melanoma antigen-1 (MAGE-A1), using T-cell epitope cloning (van der Bruggen *et al*., 1991). Based on this technique, several CTA genes were identified, including MAGE-A2, MAGE-A3, BAGE and GAGE-1 (Gaugler *et al*., 1994; Boël *et al*., 1995; De Backer *et al*., 1999). Later, in 1995, the SEREX (serological expression cloning) technique was developed by Pfreundschuh and his colleagues to determine tumour antigens. This method was based on the immunological

screening of tumour cDNA expression libraries using antibodies from cancer patients, instead of using T-cells (Sahin *et al.*, 1997). Since then, several CT antigens have been identified using the SEREX technique, such as the immunogenic tumour antigens, New York oesophageal squamous cell carcinoma 1 (NY-ESO-1) protein (Chen *et al.*, 1997), synovial sarcoma/X breakpoint (SSX) family (Tureci *et al.*, 1998a) and meiosis-specific-protein synaptonemal complex protein 1 (SCP1)(Tureci *et al.*, 1998b). The term 'cancer testis antigen' was coined by Chen and his colleagues, who described the similarity of these genes (Chen *et al.*, 1997). To date, at least 255 CTA proteins have been classed as CTAs and organised into 138 gene families (http://www.cta.lncc.br/index.php), although not all of these meet the strictest CTA classification as some of these were found to be represent with very low level in normal tissues (see Section 1.8.3).

1.8.2 Classification of CT antigens

The CTA genes are normally expressed in the male germ cells but not in normal somatic cells. Expression of some CTA genes has been identified in some immunologically privileged areas such as central nervous system (CNS); genes in this class are termed cancer testis/CNS genes. However, CTA genes have been sub-classified according to their expression in normal and cancer tissues into:

(i)   **Testis and/or CNS restricted:** Gene expression in this group is restricted only to the testis or testis and some of central nervous tissues as well as cancer tissues.

(ii)   **Testis and/or CNS selective:** Gene expression in this group is found in the testis and/or CNS as well as in at least one or two normal tissues, as well as cancer tissues (Feichtinger *et al.*, 2012a).

In addition, CTA genes are divided into the genes encoded on the X chromosome and are termed X-CT genes, while non X-CT genes are located on autosomes (Simpson *et al.*, 2005). Approximately 10% of genes on chromosome X are reported as CT genes and are make up more than half of the discovered CTA genes. The X-CT genes predominantly constitute a multigene group organised along the X chromosome in well-defined clusters (Figure 1.13). For instance, the *GAGE*, *PAGE* and *XAGE* super-families and all members of the *SSX* family are located on the Xp11 region. The Xq 24 –Xq 28 region is proposed to have the highest density

of CTA genes and contains the *MAGE-A*, *MAGE-C* and *NY-ESO-1* multigene families (Simpson *et al*., 2005; Caballero and Chen, 2009). Three or more X-CTA genes are also expressed in 40% of breast cancers and 65% of melanomas (Sahin *et al*., 1998). However, X-chromosome in mammalian spermatogenesis is reported to transcriptionally silence during the meiotic zygotene to pachytene transition in a unique mechanism called meiotic sex chromosome inactivation (MSCI), which is based upon DSBs formation in unsynapsed chromatin (Turner,2007). This inactivation mechanism proposed that the vast majority of CTA genes are silenced during meiosis, as most of CTA genes are X-encoded and therefore might have non-meiotic roles in the testis (Feichtinger *et al*., 2012a).

Non X-CTA genes are distributed around the genome and they exist predominantly as single-copy genes. This group of genes is generally expressed in the late stages of germ cell differentiation (Simpson *et al*., 2005). For example, the synaptonemal complex genes *SCP1* (Tureci *et al*., 1998b), *HORMD1* (Chen *et al*., 2005a) and meiosis-specific endonuclease *SPO11* (Koslowski *et al*., 2002) are identified non X-CTA genes.

A recent study based on bioinformatics analysis followed by validation of novel CTA genes described a new class of clinical CT genes called meiosis-specific CT antigen genes (meiCT) (Feichtinger *et al*., 2012a; Sammut *et al*., 2014). This study went one step further by incorporating meta-analyses of cancer gene expression data sets in different types of tumours in order to identify a highly specific cancer biomarker. Most of these genes are autosomally encoded and some serve to drive meiotic dynamics, including the meiotic regulator gene *STRA8*, meiotic hotspot regulator gene *PRDM9,* and cohesion subunits regulator genes *SMC1β* and *RAD21L* (Feichtinger *et al*., 2012a).

1.8.3 Expression of CTA antigens

In addition to the expression in testis, certain CTA genes are expressed in some non- germ line tissues such as the placenta; this includes the *MAGE* family, *XAGE2* and *XAGE3* genes (Silva *et al*., 2007). Quantitative RT-PCR analysis reveals that the expression of CTA genes in non-testis tissues is less than < 1% of their expression in testis. However, the expression of CTA genes has been evaluated in different types of cancers. It appears to be associated with high expression in ovarian cancer, lung cancer and melanoma. Also, it showed low levels of

expression in colon cancer, renal cancer, pancreatic cancer, lymphomas and leukaemia (Hofmann *et al*., 2008). Furthermore, epithelial cancers, including breast, bladder and prostate cancer, show moderate expression of CTA genes. Several studies have revealed that CTA genes tend to be co-expressed in the same positive tumours and that CTA genes are often expressed at high levels in the later clinical stages, depending on tumour grade (Table 1.2) (reviewed in Caballero and Chen, 2009). For instance, in bladder cancer, *NY-ESO-1* is often down regulated, with no expression in grade 1 tumours, but expression in 23% and 40 % of grade 2 and 3 tumours, respectively (Kurashige *et al*., 2001). On the other hand, CTA genes have been investigated at the protein level in different normal and cancer tissues. For example, NY-ESO-1, MAGE-A, and SCP-1 proteins have been evaluated in cancer tissues and antibodies against these genes were produced (reviewed in Caballero and Chen, 2009).

Recently, a large-scale study has demonstrate a new sub-class of testis-specific and placenta-specific (TS/PS) genes that have applications in prognostics and patient stratification. These genes were found to be epigenetically silenced in normal tissues and associated with lung cancers delineated patients with aggressive, metastasis prone tumours with poor prognosis (Rousseaux *et al.*, 2013a).

Moreover, it has been demonstrated that germline genes in *Drosophila melanogaster* are required for the oncogenic process and that the human orthologues of these *Drosophila* genes have up-regulated expression in a range of human cancers, although the functional implications for oncogenesis of this up-regulation remains unclear (Feichtinger *et al*., 2014 b).

1.8.4   Functional roles of CT antigens

The biological functions of CTA in both germ cells and tumour development are poorly understood, but some evidence indicates that CTA have a critical role in human tumourigenesis (Fratta *et al*., 2011). Indeed, it has also been proposed CTAs would be involved in drug resistance (Whitehurst *et al*., 2007). CTAs can also be classified according to their known functions:

1. **Transcriptional regulation:** CTAs serve as transcriptional regulators of genes such as MAGE-A1 (van der Bruggen *et al*., 1991), SSX2 (de Bruijn *et al*., 2002), BORIS (Loukinov *et al*., 2002) and BRDT (Scanlan *et al*., 2000).

2. **Control of cellular signalling:** such as MAGEA1, SGY1, LIP1 (reviewed in Ghafouri-Fard and Modarressi, 2009).

3. **Meiotic recombination:** such as cohesion subunits *SMC1β* and *RAD21L;* the meiotic regulator STRA8 and meiotic hot spot regulator PRDM9 (Feichtinger *et al.,* 2012a), as well as SPO11 (Koslowski *et al.,* 2002) and SCP-1 (Tureci *et al.,* 1998b).

4. **Involvement in spermatozoa components:** such as TSGA10 (reviewed in Modarressi *et al.*, 2004; Yuan *et al.*, 2013).

5. **Helicase-like feature**: such as HAGE, CAGE (reviewed in Modarressi and Ghafouri-Fard, 2011)

6. **Evasion of apoptosis:** such as CAGE (Modarressi and Ghafouri-Fard, 2011).

7. **Enzymatic actions**: such as TPTE, LDHC, TSP50 and ADAM2 (reviewed in Modarressi and Ghafouri-Fard, 2011).

8. **Cell to cell binding**: such as SPA17, TPX1, ADAM2 (reviewed in Modarressi and Ghafouri-Fard, 2011).
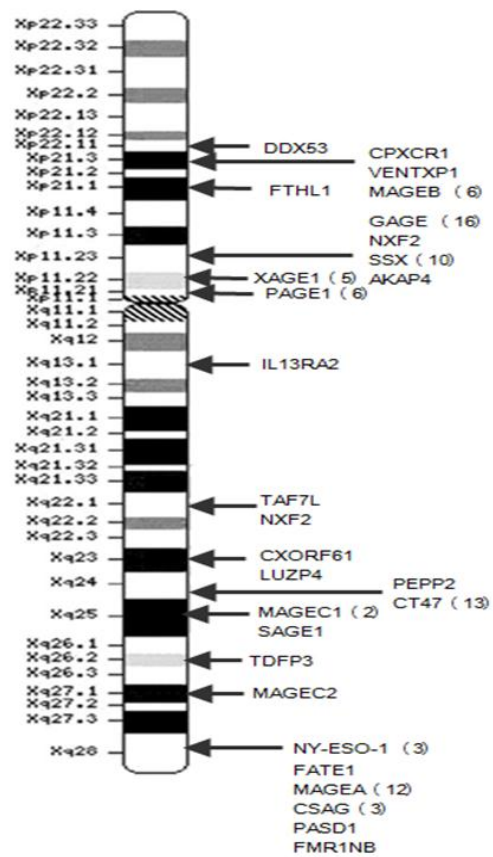
CTAs families on the X- Chromosome

Figure 1.13. **Distribution of CTA genes on the X- Chromosome.** X-CT antigens are organised in multigene families (Caballero and Chen, 2009).

Table 1.2. **Association between CTA protein expression and clinicopathologic features and prognosis** (reviewed in Caballero and Chen, 2009)**.**

Tumour Type	Antigen	Association
Melanoma	MAGE-A1, MAGE-A2, MAGE-A3, MAGE-A4	Tumour thickness and metastasis
Lung cancer	MAGE-A4, MAGE-A10, MAGE-C1	High grade tumours, pathologic stage and nodal as well as pleural invasion
Pancreatic cancer	MAGE-A3	Poor survival
Hepatocellular	MAGE-C1	Reduced overall survival
Melanoma	NY-ESO-1	Thicker primary lesions and a higher frequency of metastatic disease

1.8.5 Epigenetic involvement in CTA regulation

The term "epigenetic" refers to inheritable alteration in gene expression without changes in the DNA nucleotide sequencing. Epigenetics is an important process for regulation of gene expression and normal mammalian cell development. Epimutations can lead to increased genomic instability by silencing of tumour suppressor genes and/or activation of oncogenes—either independently or in association with genetic mutations or deletion. However, the biological mechanism that leads to the abnormality of epigenetics remains poorly understood (Sharma *et al*., 2010). Several CTA genes are usually found to co-express in positive tumours, which suggests that their activation in cancer is due to a global gene activation mechanism rather than stochastic individual events (De Smet and Loriot, 2013). The involvement of two major epigenetics processes has been described in CTA gene regulation: methylation of cytosine bases in DNA and post-translational histone modifications (Fratta *et al*., 2011).

1.8.5.1 DNA methylation

DNA methylation is a major modification of genomic DNA; it leads to the silencing of gene expression and plays an essential role in the association of histone modification in mammalian gene regulation. It involves the covalent addition of a methyl group on the 5' position of cytosine bases in CpG dinucleotides, catalysed by DNA methyltransferases (DNMTs) (reviewed in Ballestar, 2011). CpG islands, which include about 76% of human gene promoters, are located at the 5' end of the genes and commonly remain unmethylated during cell development and differentiation (Davuluri *et al*., 2001; Ramírez *et al*., 2004). Hypomethylation and/or hypermethylation of CpG island promoter regions are associated with initiation and progression of cancer (reviewed in Ballestar, 2011).

Most of the studied CTA genes are methylated in normal somatic cells but are activated by hypomethylation (demethylation) during spermatogenesis (Fratta *et al*., 2011). The first evidence to reveal that CTAs can be activated by DNA hypomethylation was presented by Weber *et al.,*(1994) who found that treatment of melanoma cells lines with DNA methyltransferase inhibitors (DNMTis) such as 5-aza-2'-deoxycytidine (5-AZA-CdR) activated the expression of the *MAGE-A1* gene (Weber *et al*., 1994). The 5-aza-2'-deoxycytidine agent

entraps the DNA methyltransferase (DNMT) enzyme in a covalent complex with the DNA, which leads to a progressive loss of DNA methylation during cell division (Zendman *et al.*, 2003a). Several CTA genes are activated by promoter DNA hypomethylation, including *NY-ESO-1*, *XAGE1* (Lim *et al.*, 2005), *LAGE1* and *MAGE-A2*, *MAGE-A3* and *MAGE-A4* (De Smet *et al.*, 1999; Sigalotti *et al.*, 2008). However, some CTA genes that lack CpG dinucleotides are activated with DNMT inhibitors, whereas several CTA genes are not activated in the presence of CpG dinucleotides. For example, the *SPANX* gene is activated by 5-aza-2'-deoxycytidine treatment even though this gene does not have a canonical CpG island in its promoter (Zendman *et al.*, 2003b; Menendez *et al.*, 2007).

The CCCTC-binding factor (CTCF) is a highly conserved, ubiquitous zinc finger factor that plays various roles in gene regulation including gene activation, methylation sensitivity and X chromosome inactivation (Ohlsson *et al.*, 2001). A study by Loukinov et al. (2002) was identified as a paralogous gene of CTCF, known as the brother of regulator of imprinted sites (*BORIS/CTCFL*). BORIS affects various aspects of the epigenetic regulation program during the development of germ cells (Loukinov *et al.*, 2002). *BORIS* was identified as being expressed only in the testis and in several types of cancers and can promote the expression of several CTA genes, whereas CTCF is normally expressed in somatic cells (Hong *et al.*, 2005; Vatolin *et al.*, 2005). The binding site of BORIS protein was identified in several *MAGE-A* genes expressed in head and neck cancers (Smith *et al.*, 2009). Overexpression of *BORIS* in normal primary fibroblasts cells by 5-aza-2'-deoxycytidine demethylation treatment leads to activation of 12 CTA genes. Down-regulation of *BORIS* using RNA interference prior to 5-aza-2'-deoxycytidine treatment reduces the ability of *MAGE-A1* expression (Vatolin *et al.*, 2005). However, some evidence suggests that presentation of BORIS protein at the promoter of some CTA genes may involve the recruitment of the histone methyltransferase SET1A as part of the transcriptional complex, but the process remains poorly understood and needs further investigation (Nguyen *et al.*, 2008; Fratta *et al.*, 2011).

1.8.5.2 Histone modifications

Histone proteins, which comprise the nucleosome core, include a globular C-terminal domain and a flexible N-terminal tail. Several post translational modifications are associated with the

N-terminal tail, such as ubiquitination, acetylation, methylation and phosphorylation. These modifications play essential roles in cellular proliferation, including DNA repair, replication and transcription. Histone modifications are usually controlled by enzymes that balance the action of covalent modification by adding and removing acetyl and methyl groups. For example, histone acetylransferases (HATs) add acetyl groups, and histone methyltransferases (HMTs) add methyl groups to the N-terminal. In contrast, histone deacetylases (HDACs) and histone demethylases (HDMs) play opposite roles. Different cancer types overexpress HDACs, resulting in a loss of histone acetylation (reviewed in Fratta *et al*., 2011). For example, treatment of cancer cells with histone deacetylase inhibitors (HDACI) such as trichostatin A has minor effects on expression of the CTA genes, *MAGE-A1*, *MAGE-A2*, *MAGE-A3* and *MAGE-A12*, whereas using this treatment in combination with 5-aza-CdR upregulates the transcription of these genes (Wischnewski *et al*., 2006).

1.9 Clinical application of CT antigen genes

As previously discussed, the testes are an immunologically-privileged site because of the blood–testis barrier (BTB) and because testicular germ cells lack the major histocompatibility complex MHC class I antigens, which prevents the immune system from recognising and attacking the CTA proteins as a "nonself" proteins. For these reasons, CTA genes can be promising targets for cancer diagnostic and immunotherapy treatment (reviewed in Ghafouri-Fard and Modarressi, 2009; Fratta *et al*., 2011; Modarressi and GhafouriFard, 2011). Two major strategies of immunotherapy in cancer are used: (1) passive immunotherapy, which includes adaptive cell therapy or using monoclonal antibodies (2) active immunotherapy, which relies on the use of vaccines.

1.9.1 Adoptive cell therapy

Adoptive immunotherapy is based on the isolation of the lymphocyte T-cells from a patient and culturing and enriching these cells *in vitro* to express the specific tumour antigens, followed by re-infusion into the patient again (Restifo *et al*., 2012). One interesting example was provided by Hunder et al., (2008) who treated a patient with melanoma metastasis using autologous CD4+ T cells against the CTA NY-ESO-1. They isolated and expanded CD4+ T cells

from the patient *in vitro* against NY-ESO-1 and infused the cells back into the patient. Two months later, after the infusion of the cells into the patient, no nodal and pulmonary tumours occurred and after two years, the patient remained free of cancer (Hunder *et al.*, 2008).

1.9.2 Monoclonal antibodies

Monoclonal antibodies are now used for cancer treatment. Several commercial monoclonal antibodies are produced to target cancer cells; one example is Cetuximab, which used for colorectal cancer treatment (Van Cutsem *et al.*, 2009). They are specific for targeting tumour association antigens (TAAs), and they have shown some therapeutic efficacy. For example, antibodies conjugated either chemotherapeutic drugs or radioactive isotopes are used for treatment of haematological malignancies and have shown some promising results. In contrast, non-leukaemic cancers are commonly treated with unconjugated antibodies that normally target the growth factor receptor, such as human epidermal growth factor receptor 1 and 2 (ERBB1 and ERBB2), which are expressed in 10-40% of breast cancers and other carcinomas (Weiner *et al.*, 2010).

1.9.3 Active immunotherapy

Several studies have found humoral immune responses to CTAs in different types of tumours. For example, different antibodies were produced against SCP-1, NY-ESO-1, SSX-2 and CTSP-1 in breast cancer. Antibodies against MAGE-A3, SSX2 and NY-ESO-1 in multiple myelomas have been developed. In addition, cancer vaccine trials have been carried out in melanoma patients using MAGE-A3 or NY-ESO-1 and have reported a regression in the tumour nodules. For NY-ESO-1, 34 trials have been performed, including NY-ESO-1 peptide, protein and pox-NY-ESO-1. One of the most encouraging vaccines to positively affect cancer is NY-ESO-1 protein/ISCOMATRIX®. Clearly, CTAs could be a potential target for reducing the risk of melanomas, as well as being useful for immunosystem adaptation (reviewed in Caballero and Chen, 2009).

1.10 Project aims

The main aim of this project was to identify new potential CT gene targets/biomarkers by manually searching the literature for meiosis-specific genes or by searching the large bioinformatics pipelines developed by the Mcfarlane group (Feichtinger *et al.,* 2012a).

The overarching goals of this project are:

1. Identify novel potential CTA genes through validation these genes in RNAs isolated from normal human tissues and cancerous cell types using RT-PCR analysis.

2. Identify the effect of epigenetic inhibitors on the expression of the CTA genes using RT-PCR analysis.

3. Function analysis of novel CTA PRDM9 in human cancer cell lines.

4. Study the histone methyltransferase activity of PRDM7 and PRDM9.

Chapter 2.0   Materials and Methods

2.1   Human cancer cell line sources

The embryonal carcinoma NTERA-2 (clone D1) cell line was kindly provided by Prof. P.W. Andrews (University of Sheffield). The ovarian cancer cell line A2780 was a generous gift from Prof. P. Workman (Cancer Research UK Centre for Cancer Therapeutics, Surrey, UK). The 1321N1, COLO800, COLO857, G-361, HCT116, HT29, LoVo, MCF7, MM127, SW480 and T84 cell lines were obtained from the European Collection of Cell Cultures (ECACC). The two ovarian adenocarcinoma cell lines PEO14 and TO14 were purchased from Cancer Research Technology Ltd. The H460 and MDA-MB-453 lines were obtained from the American Type Culture Collection (ATCC). The stably transfected cell line HeLa Tet-On® 3G with pCMV-Tet3G was purchased from Clontech. Cancer cell lines are tested for authenticity once per annum by LGC StandardsTM (authentication tracking number 710236782).

2.2   Cell culture growth maintenance

Cancer cell lines were cultured in media supplemented with foetal bovine serum (FBS) (Invitrogen; GIBCO 10270). HeLa Tet-On® 3G cells were grown in Tet System Approved (FBS) medium (Clontech; 631106) in the presence of 100 µg/ml Geneticin G418 (Invitrogen; GIBCO 10131035). All cancer cell lines were maintained in humidified incubators at 37°C in a 5% CO_2 atmosphere ; 10% CO_2 for the NTERA-2 (clone D1) cells. Cancer cells were regularly tested for mycoplasma infection using a LookOut® Mycoplasma PCR Detection kit (Sigma Aldrich; MP0035) according to the manufacturer's protocol. All media and cell line growth conditions are listed in Table 2.1.

Table 2.1. **Description and growth conditions for the human cancer cell lines used in this study**

Cell Line	Description	Media	CO_2
G-361	Human Caucasian malignant melanoma	McCoy's 5A medium + GLUTAMAX^{TM} (Invitrogen, GIBCO 36600) supplemented with 10% FBS	5%
HCT116	Human colon carcinoma		
HT29	Human Caucasian colon adenocarcinoma		
MCF7	Human Caucasian breast adenocarcinoma	DMEM + GLATAMAX^{TM} + supplemented with 10% FBS and 1xNEAA (non-essential amino acids)	5%
NTERA-2 (clone D1)	Human Caucasian pluripotent embryonal carcinoma	Dulbecco's modified Eagle's medium (DMEM) + GLUTAMAX^{TM} (Invitrogen, GIBCO 61965) supplemented with 10% FBS	10%
1321N1	Human brain astrocytoma		5%
SW480	Human colon adenocarcinoma		
A2780	Human ovarian carcinoma		
HeLa Tet-On® 3G	Stably transfected Human cervical cancer with pCMV-Tet3G vector	Dulbecco's modified Eagle's medium (DMEM) + GLUTAMAX^{TM} (Invitrogen, GIBCO 61965) + 10% Tet System Approved FBS	
COLO800	Human melanoma	Roswell Park Memorial Institute 1640 medium (RPMI 1640) + GLUTAMAX^{TM} (Invitrogen, GIBCO 61870) supplemented with 10% FBS	
COLO857	Human melanoma		
H460	Large cell lung carcinoma		
PEO14	Ovarian Adenocarcinoma, peritoneal ascites	RPMI 1640 + GLUTAMAX^{TM} + supplemented with 10% FBS and 2 mM sodium pyruvate	5%
TO14	Ovarian Adenocarcinoma, solid metastasis		
MM127	Human malignant melanoma	RPMI 1640 + GLUTAMAX^{TM} supplemented with 10% FBS and 25 mM HEPES	5%
T84	Human colon carcinoma	Ham's F12 + DMEM (1:1) + GLUTAMAX^{TM} (Invitrogen, GIBCO 31331) + 10% FBS	5%
MDA-MB-453	Human breast carcinoma	Leibovitz's (L-15) medium + GLUTAMAX^{TM} (Invitrogen, GIBCO 31415) + 10% FB	0%

2.3   Preparation of cancer cell line stocks

Confluent cells were washed twice with 1x PBS and then trypsinised with 1x trypsin – EDTA (Invitrogen, GIBCO 1370163). The cells were counted using either a haemocytometer slide or an automated cell counter (Bio-Rad) and then centrifuged at 100 x g for 5 minutes. Cell pellets were resuspended in a freezing medium (1:9 DMSO: FBS) and transferred into a labelled cryotube. For short term storage, the cells placed at -80°C for 24 hours. For long-term storage, cells were stored in liquid nitrogen.

2.4   Thawing of stored cancer cell lines

The cell vials were rapidly thawed in a 37°C water bath with gentle agitation. The cell suspension was immediately diluted into 5 ml of prewarmed medium and mixed gently. After centrifugation at 100 x g for 5 minutes, the medium was aspirated and the cells resuspended in 10 ml growth medium. The cells were split into two T25 flasks and grown for 24 hours at 37°C in a humidified incubator with the required CO_2 concentration.

2.5   RNA extraction

Total RNA was isolated using Trizol reagent (Invitrogen; 15596-026). Confluent cells were homogenised in Trizol (1 ml Trizol/5 x 10^6 cells) and held at room temperature (RT) for 5 minutes. Chloroform (200 μl per 1 ml of Trizol) was added to each sample and the homogenate was vigorously shaken for 15 seconds, followed by incubation for 5 minutes at RT. Samples were then centrifuged at 12,000 x g for 15 minutes at 4°C. The aqueous layer was then removed to a new Eppendorf tube and 500 μl of isopropanol was added. After incubation at RT for 10 minutes, the samples were centrifuged again at 12,000 x g for 20 minutes. The supernatant was removed and the pellet was washed with 70% ethanol and re-centrifuged at 7,500 x g for 5 minutes at 4°C. The supernatant was discarded again and the cell pellet was left to dry at RT for 5-10 minutes, and then 100 μl RNase free water containing 2 μl DNase I (Sigma; D5319) was added to each RNA preparation sample. The samples were incubated at 37°C for 10 minutes and then at 75°C for 10 minutes. RNA concentration was measured with a NanoDrop (ND_1000) spectrophotometer.

2.6 cDNA synthesis (Reverse Transcriptase)

Total RNA was prepared from the cancer cell lines generated in our lab and listed in Table 2.1. RNA from normal human tissues was supplied by Clontech (Catalogue number; 636643).

In addition, panels of RNA from cancer cell lines and tumours were purchased from both Clontech and Ambion. RNA was used to synthesise cDNA using a SuperScript III First Strand Synthesis Kit (Invitrogen; 18080-051). Samples (1-2 μg) of total RNA were used according to the manufacturer's protocol. Oligo-dT primer was used to transcribe RNA into single-strand cDNA which binds to the poly-A tail of mRNAs. The cDNA was diluted eight time and PCR using *β- Actin* primers was used as endogenous control for the cDNA.

2.7   Polymerase chain reaction (PCR)

Gene sequences were obtained from the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov). Primers were designed to span more than one intron where possible. Two methods of primer design were used: a manual method for picking primers or by using Primer 3 software (available at http://primer3.ut.ee/).

For PCR amplification, 2 μl of diluted cDNA was supplemented with 25 μl of BioMix™ Red (Bioline; BIO-25006) and 1 μl each of the forward and reverse primer, and the final volume was adjusted with ddH$_2$O to 50 μl.

RT-PCR for samples was initiated with a pre-cycling melting step at 96°C for 5 minutes, followed by 40 cycles of denaturing at 96°C for 30 seconds. For all of the primer settings, an annealing step was carried out between 58-62°C for 30 seconds. The extension temperature was 72°C for 30 seconds and the final extension temperature was 72°C for 5 minutes.

As positive controls, RT-PCR of *β-Actin* and previously known CT antigen genes *GAGE1*, *MAGEC1* and *SSX2* were evaluated in all normal and cancer cDNA samples and run on 1% agarose gels stained with ethidium bromide.

2.8 Demethylation and/or histone deacetylase analysis using RT-PCR

(Adapted from Mossman *et al.* 2010)

Colorectal cancer cell lines HCT116 and SW480 were treated with 5'-aza-2' deoxycytidine (a demethylating agent; Sigma, A3656) and/or Trichostatin A (a histone deacetylase inhibitor;

Sigma, T1952). Demethylation was induced by treating cells with different concentrations (0.1, 0.5, 1.0, 5.0, 10.0 and 15.0 µM) of 5'-aza-2' deoxycytidine for 48 or 72 hours. Deacetylase assays were performed by treatment with 150, 300 or 600 nM TSA for 24 hours. The experiments were also carried out using 0.1, 0.5, 1.0, 5.0, 10.0 and 15.0 µM 5'-aza-2' deoxycytidine and 300 nM TSA for 48 hours. The media were replaced every 24 hours with fresh media containing 5'-aza-2' deoxycytidine and/or TSA. Cells were treated with 10 µl DMSO (Sigma; D8418) to control for effects of DMSO on gene expression since DMSO was used as the solvent for the treatment chemicals. Total RNA was extracted from untreated and treated cells as described in Section 2.5. RT-PCR was performed to investigate the effect of treatment with 5'-aza-2' deoxycytidine and/or Trichostatin A on gene expression. Primers and genes used in this experiment are listed in Tables 2.2, 2.3 and 2.4. Only genes that showed no expression in both the HCT116 and the SW480 cell lines were used in this experiment.

2.8.1 DNA methylation and histone deacetylase analysis using RT-PCR

(Adapted from Mossman *et al.* 2010)

HCT116 and SW480 cell lines were treated with 300 nM TSA for 24 hours. The HCT116 cells were also treated with 1 µM 5'-aza-2' deoxycytidine for 48 or 72 hours, as well as with a combination of 1 µM 5'-aza-2' deoxycytidine and 300 nM TSA for 48 hours. Total RNA was extracted as described in Section 2.5. The remaining cells were washed twice with 1x PBS and allowed to grow under normal conditions with no drugs. RNA was extracted from these cells after 3, 6 and 9 days. The cells media were changed at least every two days, and cDNA was synthesised for RT-PCR.

2.9 DNA purification methods

2.9.1 Direct RT-PCR product purification

PCR reactions (50 µl) were directly purified using the High Pure PCR Product Purification Kit (Roche Applied Science; 11732676001), according to the manufacturer's protocol.

2.9.2 Purification of DNA from agarose gel

PCR products were run in 1% of agarose gels stained with ethidium bromide in 1XTBE or TAE buffer. Specific fragments were excised from the gel using a sterile scalpel. The purification process was carried out as per the manufacturer's instructions included with the PCR purification Kit (Roche) or Geneclean (MP; 111102400).

2.9.3 DNA sequencing

DNA was sequenced by the Eurofins MWG Company (Germany). For DNA purified from PCR reaction, the amount sent (together with corresponding forward and/or reverse primers) was:

5 ng/µl of (300-1000 bp)

10 ng/µl (>1000 bp)

For plasmid DNA, this amount was 50-100 ng/µl (up to >3000 bp)

The gene sequencing results were blasted and aligned against corresponding genes using the Basic Local Alignment Search Tool (BLAST) website http://blast.ncbi.nlm.nih.gov/Blast.cgi and EMBL European Bioinformatics Institute Website http://www.ebi.ac.uk/.

Table 2.2. **Primer sequencing of positive control and previously known CT antigens and their expected sizes (bp)**

Gene	Chr	Primer	Primer Sequence	Annealing Temp °C	Product Size (bp)
β-Actin*	7	F	5'-TGCTATCCCTGTACGCCTCT-3'	58	675
		R	5'-CGTCATACTCCTGCTTGCTG-3'		
GAGE1*	X	F	5'- TAGACCAAGGCGCTATGTAC -3'	58.4	245
		R	5'-CATCAGGACCATCTTCACAC-3'		
MAGEC1*	X	F	5'-CCTGTGAGCTCCTCTTTCTC-3'	60.5	650
		R	5'-GACTATGGAGAGGAGACTGGA-3'		
SSX2*	X	F	5'-CAGAGATCCAAAAGGCC -3'	58.4	407
		R	5'- CTCGTGAATCTTCTCAGAGG-3'		

* This genes and primers were used in methylation experiment

Table 2.3. **Primer sequencing of identified CT genes and their expected sizes (bp)**

Gene	Chr	Primer	Primer Sequence	Annealing Temp °C	Product Size (bp)
ALX1	12	F	5'-CGAATGTCTCCCGTGAAAGG-3'	60.5	622
		R	5'-CTGCGATACTGGAAGACCTC-3'		
APOBEC4	1	F	5'-CATGCTAGCAGTTGCACTGG-3'	60.5	667
		R	5'-GCATTGGTGGTAAGTCCCTG-3'		
ARRDC5	19	F	5'-CAACAAGGCAGACTACGTGC-3'	60.5	628
		R	5'-GCGAGTGTGCATGATCTCAC-3'		
C4orf17	4	F	5'-CCTCATCCCAGAAGAGTCTG-3'	60.5	628
		R	5'-CTGCTGCTGGTTCCATTGAG-3'		
C9orf11	9	F	5'-CCAAATGGCACTGAGTCTGA -3'	58.4	652
		R	5'-CCGACTCATCGTTTTCATGC-3'		
C9orf12	9	F	5'-GGTCTGTCGATACTGCATGC-3'	60.5	597
		R	5'-CTCTCCTCGGGAAACTCTTC-3'		
C9orf43	9	F	5'-CAGCACTGCGATATCCTGAA-3'	58.4	546
		R	5'-CATCCTCATCATCAGTGTCC-3'		
C9orf128	9	F	5'-CATCAGGAGCCTTCCATCTC-3'	60.5	731
		R	5'-CGTCTCTTGAGCCAGAACTC-3'		
C12orf12	12	F	5'-CAGCGTACAATAGACCGCAC-3'	60.5	748
		R	5'-CACACCTCCTGGTCATACTC-3'		
*C16orf92**	16	F	5'-TCACAGAGTTTGCTTGGGTC-3'	58.4	279
		R	5'-GGAAGTTTATGTGGGTGCAG-3'		
*DDX4**	5	F	5'-GTGCTACTCCTGGAAGACTG-3'	60.5	756
		R	5'-CCAACCATGCAGGAACATCC-3'		
*IQCF3**	3	F	5'-GCAGTAGAAAGACAGAGGCG-3'	60.5	356
		R	5'-CCATCAGTCTGGAAGGCAAG-3'		
KDM4D	11	F	5'-GCATACAGAGGACATGGACC-3'	60.5	733
		R	5'-GGATGATGGAGTTGAGCTGG-3'		
MAGE-B5	X	F	5'-CCTCCACTGAGAGTTCATGC-3'	60.5	655
		R	5'-CTTGGGCTCTCTTCCTCATC-3'		
*NT5C1B**	2	F	5'-CGGCAGGAAAATCTACGAGC-3'	60.5	647
		R	5'-CTGTAACCAGGTAGGTCCTG-3'		
*ODF4**	17	F	5'-GACAAGATGGGAGACTGCTG-3'	60.5	602
		R	5'-GGTGTCTGTGATCGTCTGTG-3'		
*PRDM1***	6	F	5'-CAGTGCCTTCTCCTTTACCG-3'	60.5	768
		R	5'-ATGTCATCCTCCACGTCCTC-3'		
*PRDM4***	12	F	5'-GGGGACAGGTCATGTAGATG-3'	60.5	714
		R	5'-TGTCCCTGGGTAGGAAGATG-3'		
*PRDM6***	5	F	5'-GCACCTGGATTGGACCTTTC-3'	60.5	384
		R	5'-CTTGTCTGCACATGGCTTCC-3'		

Gene	Chr	Primer	Sequence	Tm	Size
*PRDM7***	16	F1	5'-CTTCATTGACAGCTGTGCTG-3'	58.4	278
		R	5'-ACTCATAGCAGTTTCTCCCC-3'		
*PRDM7****	16	F2	5'-GATGGCCTTCAGAGGAGAAC -3'	58.4	755
		R	5'-GACCAGACCAGCAGTTCACA -3'		
*PRDM7****	16	F6	5'-CTTCATTGACAGCTGTGCTG-3'	60.5	607
		R	5'-AGTTCCTGGCCATACTCATC -3'		
*PRDM9****	5	F1	5'-GCAGCCAACAATGGATACTC -3'	60.5	734
		R	5'-CCACACTCTCCATACTTGAC -3'		
*PRDM9****	5	F4	5'-CAGGCTCAGAAACCAGTGTC-3'	60	655
		R	5'-GTTCCTGGCCGTATTCATCC-3'		
PRDM11	11	F	5'-AAAGCTTCCAGCAAGTGGAC -3'	60.5	680
		R	5'-TACATCCCCCTCATCAAAGC -3'		
*PRSS54***	16	F	5'-CTCACACAGAGTATCCAGTC-3'	58.4	646
		R	5'-GTGATGGTCCTTCTTTGTCC-3'		
*SEPT12***	16	F	5'-CTGCAGCTGCATTCACTGAC-3'	60.5	608
		R	5'-CGGATAAGCAGGTCTCTCAG-3'		
*TDRD12***	19	F	5'-GAGCTAAAGTGCTGGTGCAG-3'	60.5	641
		R	5'-CTGAGGTCACCGACAATACC-3'		
*TMEM202***	15	F	5'-CCAAGTGCCTCGATGTCATG-3'	60.5	629
		R	5'-GATACTGTAGTCACCGGACC-3'		
TNRC18	7	F	5'-CTCTTCTTCCTCCACCACAG-3'	60.5	725
		R	5'-CATGCGAGGACTGGTATAGC-3'		
TSSK1B	5	F	5'-CGAAGCTCGCAAGAAGTTCC-3'	60.5	651
		R	5'-GGTTTTGTCTCAGGCTGTGC-3'		
WDR20	14	F	5'-GCCACAGCAGAAAGTGTCTC-3'	60.5	601-735
		R	5'-CTACAACACTGACCCAGGAC-3'		

*These genes and primers were used in methylation experiments

 ** These genes and primers were used in overexpression system Tet-on 3G experiment

Table 2.4. RT-PCR primers of validated genes from HCT116 and SW480 cells treated with 5'-aza- 2' deoxycytidine and TSA, singly or in combination

Gene	Chr	Primer	Primer Sequence	Annealing Temp °C	Product Size (bp)
C17orf98[‡]	17	F	5'-ATGGCGTACCTGAGCGAGT-3'	58	394
		R	5'-GGTAGCCAAACCTTCCATTG-3'		
C20orf201[‡]	20	F	5'-ATCTGCTCTTCGGCGACCTG-3'	62.5	505
		R	5'-ACACTCTCAGTCGCCGTCAC-3'		
SMC1β[‡]	22	F	5'-TCAAGAAATCGAGGCCCACC-3'	58	368
		R	5'-CTGGGGCCACACAGTTATAG-3'		
STRA8[‡]	7	F	5'-TGGCAGGTTCTGAATAAGGC-3'	58.4	723
		R	5'-GAAGCTTGCCACATCAAAGG-3'		
SYCP1[‡]	1	F	5'- GGTCAGCAGAAAGCAAGCAAC -3'	61	645
		R	5'-GGCAGATGTCCACAGATAGTC-3'		

‡ Genes were evaluated as CT antigen genes (Feichtinger *et al.* 2012)

2.10 Real time quantitative qRT-PCR

Total RNA was prepared from cell cultures at the confluent stage using an RNeasy Plus Mini kit (Qiagen; 74134) according to the manufacturer's protocols. The concentration and quality of the RNA was measured using a NanoDrop (ND_1000) spectrophotometer. A QuantiTect Reverse Transcription Kit (Qiagen; 205310) was used to synthesise cDNA from 1 µg of total RNA according to the manufacturer's protocol; the resulting cDNA then was diluted 10-fold.

The qRT-PCR primers were commercial primers ordered from Qiagen or personally designed using Primer 3 software. These were used to perform SYBR® Green-based quantitative real time qRT-PCR. The quantitative real time PCR results for target genes were normalised against results obtained for reference genes (Tables 2.5 and 2.6).

The Go Taq qPCR Master Mix (Promega; A6001) was used to set up the quantitative RT-PCR reaction, following the manufacturer's protocol. Each well of a Hard-Shell® 96 well plate (BioRad; 9655) contained 1.5 µl cDNA in a final reaction volume of 25 µl, and triplicate repeats were carried out for each reaction.

The target sequences were amplified with a pre-cycling hold at 95°C for 3 minutes, then followed by 40 cycles of 95°C for 10 seconds. All primers underwent an annealing temperature at 60°C for 30 seconds, followed by 95°C for 10 seconds. A quantitative Bio-Rad

CFX machine was used to perform the qPCR experiment and the data were analysed with Bio-Rad CFX Manager Software (version 2).

Table 2.5. Commercial quantitative real time PCR primers used in this study and their sources

Tested Gene	Assay Name	Source
β-Actin	Hs_ACTB_1_SG	Qiagen; QT00095431
GAPDH	Hs_GAPDH_2_SG	Qiagen; QT0192646
LAMIN A	Hs_LMNA_1_SG	Qiagen; QT00083349
PRDM9	Hs_PRDM9_1_SG	Qiagen; QT01023631

Table 2.6. Personally designed quantitative real time PCR primers used in this study and their details and sources

Gene	Chr	Primer	Primer Sequence	Annealing Temp °C	Product Size (bp)
PRDM9	5	F1	5'-GCCTTAAGAGTGGAACAGCG-3'	60	140
		R1	5'-GACACTGGTTTCTGAGCCTG-3'		
		F2	5'-GACACAGAGAGAACAGAGCG-3'		87
		R2	5'-CTCTGCCCATTCTTCCTTGG-3'		
PRDM7	16	F	5'-CTGCTATGAGTATGTGGATGGA-3'	60.5	118
		R	5'-GATCTGCCTGTGGTACTGGA-3'		

2.11 Western blotting protocol

2.11.1 Whole protein extractions (WPEs) from cancer cell lines and quantification

Protein was extracted for western blot assays by washing confluent cell cultures twice with cold 1x PBS. Cell pellets were then treated with RIPA buffer (Sigma; R0278) containing one complete mini, EDTA-free, protease inhibitor cocktail tablet per 10 ml (Roche; 11836170001) and mixed by pipetting. The solution was then mixed by vortexing, and incubated on ice for 30 minutes with gentle vortexing every 10 minutes. After centrifugation at high speed for 20 minutes at 4°C, a Pierce ® BCA Protein Assay Kit (Thermo Scientific; 23227) was used to assess the protein concentration in the supernatant according to the manufacturer's protocol.

2.11.2   SDS page and western blotting

Approximately 30 μg of protein lysates were mixed with 3.7 μL of NuPAGE® LDS Sample buffer (4x) (Invitrogen; NP0007) and 1.5 μL NuPAGE® Sample Reducing Agent (10x) (Invitrogen; NP0004) and the samples were then boiled at 95˚C for 5-10 minutes.

Protein samples were run on a pre-cast NuPAGE® 4-12% Bis-Tris gel (Invitrogen; NP0322) alongside Precision Plus Protein Dual Color Standards (Bio-Rad; 161-0374). Samples were run for 2 hours at 100 Volt in (1x) MOPS running buffer (Invitrogen; NP0001), and then transferred to an activated Immobilon-P (PVDF) membrane (Millipore; IPVH00010) at 400 mA for 2-4 hours using (1x) transfer buffer (380 mM glycine, 50 mM Tris). Membranes were incubated for at least one hour in 5% dry skimmed milk in blocking buffer (1x) PBS/Tween (0.3%), followed by an overnight incubation with a required concentration of the primary antibody at 4˚C (Table 2.7 A). Membranes were washed three times with a blocking buffer (1x) PBS/Tween (containing 5% of milk) for 10 minutes, followed by incubation with the required concentration of the secondary antibody for one hour at room temperature (Table 2.7 B). Membranes were washed again three times, 10 minutes each time, with the blocking buffer. Enhanced chemiluminescence (ECL; Thermo Scientific; 34087) was used to detect the required protein, which visualised using X-Ray films (Thermo Scientific; 34091).

2.11.3 Crude subcellular fractionation

Subcellular fractionation of proteins was performed by separating, cytoplasmic fractions after resuspending cell pellets in lysis buffer C (1% Triton-X-100, 10 mM $MgCl_2$, 1 mM AEBSF, Roche complete protease inhibitor cocktail) and an equal volume of hypotonic buffer (50 mM Tris-HCl pH 7.4, 0.1 M sucrose, 1 mM AEBSF, Roche complete protease inhibitor cocktail). The lysed cell fractions were placed on ice for 30 minutes and spun down at 6,000 x g for 2 minutes. The supernatants, which contained the cytoplasmic fractions, were collected and transferred to fresh tubes. The cell pellets, which included the nuclear fractions, were re-suspended in lysis buffer N (50 mM Tris-HCl pH 7.4, 100 mM KAc, 1 mM AEBSF, Roche complete protease inhibitor cocktail). Both cytoplasmic and nuclear fractions were mixed

with an equal volume of Laemmli buffer. Samples were heated at 100°C for 5 minutes with briefly vortexed every two minutes; samples were then run on SDS-page gels.

Table 2.7 A. List of primary antibodies and their optimum dilutions

Antibody	Source	Application	Dilution
Monoclonal mouse anti α-Tubulin	Sigma; T6074	WB	1/8000
Polyclonal rabbit anti-*PRDM9*	Abcam; Ab85654	WB	1/1000
Monoclonal mouse anti-Lamin A/C (636)	Santa Cruz; sc-7292	WB	1/100

Table 2.7 B. List of secondary antibodies and their optimum dilutions

Antibody	Source	Application	Dilution
Donkey anti-rabbit IgG(H+L)	Jackson Immunoresearch; 711-035-152	WB	1/40,000
Donkey anti-mouse IgG(H+L)	Jackson Immunoresearch; 711-035-150	WB	1/40,000
Rabbit anti-goat	Sigma; A5420	WB	1/40,000

2.12 Gene knockdown using small interfering RNA (siRNA)

Cells were seeded into 6-well plates at 1.5 x 10⁵ or 2.0 x 10⁵ cells per well and incubated to reach approximately 40% confluent growth. The transfection complex was made by adding 10 nM siRNA (Qiagen) and 6 µl HiPerfect reagent (Qiagen; 301705) to 100 µl serum free medium and incubating for 25 minutes at room temperature (Table 2.8). A negative non-interference control (Qiagen; 1022076) was used to assess the siRNA knockdown and was prepared in the same way. The transfection mixture was added dropwise to each well and the plates were shaken gently. After 48 hours, the medium was replaced and a second hit of siRNA was applied. The treated and untreated cells were harvested after the first and third hits to perform western blot analysis as described in Section 2.11.

Table 2.8. List of siRNA used to knock down the *PRDM9* gene.

Gene	siRNA Name	Source	Target sequencing
PRDM9	Hs_PRDM9_ 5	Qiagen; SI04299890	5'-TTCCCTTATCACTGAAGGCAA -3'
	Hs_PRDM9_ 6	Qiagen; SI04299890	5'-CACGGGAGACTGTGAAGAGCA -3'
	Hs_PRDM9_ 7	Qiagen; SI04299890	5'-CCACACAGCCGTAATGACAAA-3'
	Hs_PRDM9_ 8	Qiagen; SI04299890	5'-GTGGACAAGGTTTCAGTGTTA -3'
NI (Non-Interference siRNA	AllStars Negative Control siRNA	Qiagen; 1022076	--

2.13 Extreme limiting dilution analysis (ELDA)

(Adapted from Hu and Smyth, 2009)

The effect of *PRDM9* gene knockdown was examined in parental HCT116 and SW480 cells by seeding cells in 96-well plates (Costar; Corning 3474) at concentrations of 1000, 100, 10 and 1 cell in 100 µl of medium. A further 6 well repeats were used for untreated cells, negative control siRNA, Hiperfect treated cells and *PRDM9* siRNA treated cells. The transfection complex was made by adding 0.1 nM siRNA (Qiagen) containing 0.3 µl HiPerfect reagent (Qiagen; 301705) to 4.7 µl serum free medium and incubating for 25 minutes at room temperature. A negative control non-interference siRNA (Qiagen; 1022076) was prepared in

the same way. The *PRDM9* siRNA transfection mixture, negative control siRNA or Hiperfect only were added dropwise to the medium (6 well repeats for each condition) alongside the untreated cells. The cells were incubated for 10 days in a humidified incubator at 37°C with 5% CO_2. Cells were supplemented with 50 µl serum free medium and the transfection mixture re-applied after 2 and 6 days of incubation.

After 10 days of incubation, the numbers of wells showing positive cells growth were counted by light microscopy. The frequency of cell proliferation was determined using the ELDA web tool (http://bioinf.wehi.edu.au/software/elda/).

2.14 Gene cloning

2.14.1 PCR for amplification of *PRDM7* and *PRDM9*

The full open reading frame (ORF) sequence of *PRDM9* and both the long (*I*) and short (*II*) isoforms of *PRDM7* were obtained from the National Center for Biotechnology information (NCBI; http://www.ncbi.nlm.nih.gov). The forward and reverse primers were both designed with a *Bam*HI restriction site, and the full length of *PRDM7* long isoform *(I)* clone was purchased from (Genscript). Cloning into the cancer cell lines using the Tet-on 3G system was facilitated by adding a Kozak sequence (ACC) to the forward primers and a 6X His tag to the reverse primers. No Kozak or 6 x His tag sequences were added to the primers for the protein expression experiment (Table 2.9. and 2.10).

The PCR amplification was carried out using 2 µl of diluted testis cDNA (containing ≈150 ng/µl cDNA), 25 µl of Phusion® High-Fidelity PCR Master Mix with GC Buffer (NEB; M0532s), 1 µl of each forward and reverse primer, and sufficient ddH_2O to make a final volume of 50 µl.

The PCR conditions were a pre-cycling melting step at 98°C for 30 seconds, followed by 35 cycles of denaturing at 98°C for 10 seconds and annealing temperatures between 58-62°C for 30 seconds. The extension was carried out at 72°C for 30 seconds per 1 kb, followed by a final extension step at 72°C for 10 minutes. A 5 µl sample of the PCR reaction was run on a 1% agarose gel as a check.

Table 2.9. List of primer genes used for cloning into pNEB193 or pGEM 3ZF (+) vectors.

Gene	Primer	Primer Sequence	Annealing Temp °C	Product Size (bp)
PRDM9	F	5'-GGAGGATCCATGAGCCCTGAAAAGTCCC-3'	62	2685
	R	5'-GGAGGATCCTTACTCATCCTCCCTGGAG-3'		
PRDM7 (II)	F	5'- GGAGGATCCATGTGTCAGAACTTCTTCATTGA-3'	58.4	516
	R	5'- GGAGGATCCTCAGAATATTGCCGCTCCTG - 3'		

Table 2.10. List of primer genes used for cloning into pTRE 3G vectors.

Gene	Primer	Primer Sequence	Annealing Temp °C	Product Size (bp)
PRDM9	F	5'-GGAGGATCCACCATGAGCCCTGAAAAGTCCC-3'	62	2685
	R	5'-GGAGGATCCTTAGTGGTGGTGGTGGTGGTGCTCATCCTCCCTGGAG-3'		
PRDM7 (I)	F	5'- GGAGGATCCACCATGAGCCCTGAAAGGTC -3'	60	1479
	R	5'-GGAGGATCCTCAGTGGTGGTGGTGGTGGTGAGAGTTTGG ACCTTTC -3'		
PRDM7 (II)	F	5'-GGAGGATCCACCATGTGTCAGAACTTCTTCATTGA - 3'	58.4	516
	R	5'-GGAGGATCCTCAGTGGTGGTGGTGGTGGTGGAATATTGCCGCTCCTG - 3'		

2.14.2 Purification and DNA digestion

PCR reactions were mixed with blue loading dye (6x) (NEB; B7021S), run on 1% agarose gels, and then specific bands were extracted. The purification was performed using a Roche kit as per the manufacturer's protocol. The purified PCR product was digested with 1 µl of *Bam*HI restriction enzyme (≈20 units; Promega; R6021), with 6 µl of *Bam*HI buffer E, and 4 µl of ddH$_2$O in a total volume of 60 µl. Samples were incubated at 37°C for 2 hours, then run on a 1% agarose gel. The appropriate bands were selected after digestion and purified with a Roche kit as per the manufacturer's protocol.

2.14.3 Preparation and plasmid digestion

All plasmid sources and universal primer details are listed in Table 2.11. A 2.5 µg sample of the plasmid was digested with 1 µl of *Bam*HI restriction enzyme supplemented with 10 µl of *Bam*HI buffer E, and 86.5 µl of ddH$_2$O in a total volume of 100 µl. The mixture was incubated at 37°C for 2 hours, followed by addition of 11 µl Antarctic phosphatase reaction buffer (NEB; B0289S) and 1 µl Antarctic phosphatase (NEB; M0289S). The mixture was incubated at 37°C for 15 minutes, followed by a heat inactivation step for 5 minutes at 65°C. The reaction was run on a 1% agarose gel and purified using the Roche kit.

2.14.4 Gene ligation and transformation

The *PRDM9* gene was ligated into the pGEM 3Zf (+) plasmid and *PRDM7 (II)* into pNEB193 plasmid. A 50 ng sample of digested plasmid was mixed with a 1–9 fold molar excess of the insert and made up with sterile water (Sigma; W4502) to 10 µl. A 10 µl volume of 2 x Quick ligase reaction buffer and 1 µl of Quick T4 DNA ligase (NEB; M2200S) were added and mixed thoroughly. The mixture was briefly centrifuged, incubated at room temperature (25°C) for 5 minutes and then chilled on ice.

Samples (5 µl) of recombinant plasmid *PRDM9*:pGEM 3ZF (+); *PRDM7*(I):pUC57; and *PRDM7* (II):pNEB193 and 1 µl of control (cut) and uncut vectors were transformed in a 50 µl volume of 5-alpha Competent *E. coli* (NEB; C2987I). Samples were chilled on ice for 30 minutes, followed by heat shock at 42°C for 30 seconds, and then chilled on ice again for 5 minutes. A 950 µl volume of Super Optimal broth with Catabolite repression (SOC) was pipetted into the mixture and incubated at 37°C with vigorous shaking (250 rpm) for an hour.

Samples were plated onto LB agar plates containing 100 mg/ml ampicillin (Sigma; A9518), 20 mg/ml X-gal (Thermo; R0941) and 40 µl 100 mM IPTG (Thermo; R1171) for blue/white selection. Genes cloned into plasmids without the lacZ selection gene were plated onto LB agar plates with 100 mg/ml of ampicillin only and the plates were incubated at 37°C overnight. The media used for *E. coli* and bacterial strains are listed in Tables 2.12 and 2.13.

Table 2.11. List of plasmids and their universal primers used in this study

Vector	Name	Universal primers	Size (bp)	Source
pNEB193	M 13 uni	5'- AGG GTT TTC CCA GTC ACG ACG TT- 3'	2713	NEB
	M13 rev	5'- GAG CGG ATA ACA ATT TCA CAC AGG- 3'		
pGEM 3ZF(+)	M 13 uni	5'- AGG GTT TTC CCA GTC ACG ACG TT- 3'	3197	Promega
	M13 rev	5'- GAG CGG ATA ACA ATT TCA CAC AGG- 3'		
PUC57	F	5'- GTAAAACGACGGCCAGTG- 3'	2710	Genscrip
	R	5'- GGAAACAGCTATGACCATG- 3'		
pGEX-2T	pGex F	5'- ATA GCA TGG CCT TTG CAG G- 3'	4948	GE Healthcare
	pGex R	5'- GAG CTG CAT GTG TCA GAG G- 3'		
pTRE3G	F	5'- ATTCCACAACACTTTTGTCT- 3'	3431	Clontech
	R	5'- GGTTCCTTCACAAAGATCCTC- 3'		

2.14.5 Colony screening

The PCR reaction was performed using internal primers for *PRDM9* and *PRDM7* isoforms *(I)* and *(II)* to confirm that the correct colonies were selected. White colonies were picked up into 10 µl of LB liquid medium and 2 µl was used for the PCR reaction. The PCR conditions are as described in detail in Section 2.7, and the primer details are listed in Tables 2.3 and 2.4. The remaining medium, which contained the positive PCR colonies, was inoculated into 10 ml of LB liquid medium with 100 mg/ml ampicillin, and incubated at 37°C overnight with vigorous shaking at 250 rpm.

2.14.6 Plasmid extraction from *E. coli*

A mini preparation was carried out using 5 ml of *E. coli* overnight culture containing 100 mg/ml of ampicillin. Plasmids were isolated from bacterial culture using a QIAprep Spin Miniprep Kit (250) (Qiagen, 27106) according to the manufacturer's protocol. Bacterial cells were harvested by centrifugation at 4000 xg for 10 minutes at 4 °C. The pellet was resuspended in P1 buffer, transferred to a fresh Eppendorf tube, 250 µl of Buffer P2 was added, and then 350 µl of N3 buffer was added. The tubes were gently inverted and then centrifuged at 13,000 xg for 10 minutes. The supernatant was transferred to a spin column and centrifuged at 13,000 xg for 1 minute. A 750 µl volume of PE washing buffer was added, the columns were centrifuged again at 13,000 xg for 1 minute and the supernatant wash was discarded. The columns were then centrifuged again to remove residual washing buffer. Samples were then eluted from the columns into 50 µl of elution buffer (EB).

Cell culture transfection plasmids were extracted using a HiSpeed plasmid Midi Kit (Qiagen; 12643) as per the manufacturer's instructions.

A 500 ng sample of each purified plasmid was digested with *Bam*HI enzyme to confirm correct cloning, as described in Section 2.14.3. Samples with correctly cloned genes were sent for sequencing to confirm the correct orientation and to detect any gene mutations. The universal primers and internal primers used for sequencing are listed in Tables 2.3and 2.11.

Table 2.12. Medium recipe for *E. coli* growth

Media	Amount
Luria Broth (LB)	
Tryptone	10 g
Yeast extract	5 g
NaCl	10 g
For LB agar plates add:	
Agar	14 g
Water	up to 1litre
Minimal Medium M9 salt (1X)	
Na_2HPO_4	6 g
KH_2PO_4	3 g
NaCl	0.5 g
NH_4Cl	1 g
Water	up to 800 ml
Add the sterile component after medium autoclaving	
100 mM $CaCl_2$	
1 M $MgSO_4$	1 ml
Glycerol	1 ml
Autoclaved Water	0.3% final
	up to 1L final

Table 2.13. The *E. coli* strains used in this study

Bacteria	Source
5-alpha Competent *E. coli*	NEB; C2987I
BL21 (DE3) Competent *E. coli*	NEB; C2527I
SoluBL21 Chemically Competent *E. coli*	Genlantis; C700200

2.14.7 Subcloning of *PRDM9* and *PRDM7* genes into a protein production vector pGEX-2T

The *PRDM9* and both *PRDM7* isoforms transcript sequences were subcloned into a pGEX-2T vector containing a glutathione S-transferase (GST) tag at the upstream for protein function analysis. A 1 µg sample of pGEX-2T vector (GE Healthcare; 28-546-53) was digested with 1 µl of *Bam*HI restriction enzyme (Promega; R6021) and ligated with the full length of *PRDM9* and/or *PRDM7* (Isoforms *I* or *II*) in the manner described in Section 2.14.4. Recombinant plasmids were transformed into BL21 (DE3) and/or SoluBL21 competent cells and then plated on LB agar medium containing 100 mg/ml of ampicillin and incubated overnight at 37°C. Colony screening and plasmid purification were performed as described in Sections 2.14.5 and 2.14.6. Samples with correctly cloned genes were sent for sequencing to confirm the correct orientation and to check for unwanted mutations. Universal forward F-pGEX, reverse R-pGEX and internal primers used for sequencing are listed in Tables 2.3 and 2.11.

2.14.8 Expression and purification of soluble and insoluble GST protein

A fresh and healthy single colony was picked from colonies previously transformed *PRDM9* and *PRDM7* (both isoform *I* and *II*) with pGEX-2T vector and resuspended in 10 ml of LB medium containing 100 mg/ml ampicillin. The flasks were then incubated overnight at 37°C at a shaking speed of 180-200 rpm. For protein expression, 2 ml of the (BL21) bacterial growth medium, after overnight incubation, was transferred into a Falcon tube containing 18 ml of fresh LB medium containing 100 mg/ml ampicillin and shaken at 37°C and 180-200 rpm. The optical density (OD) was measured periodically using a spectrophotometer until the OD_{600} had reached 0.6–0.8. Production of recombinant protein was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 100 µM and protein synthesis was allowed to progress for 5 hours at 37°C.

2.14.8.1 Expression and purification of insoluble GST protein using Triton-100X

Insoluble protein samples were incubated in a range of temperatures from 15°C to 37°C and different concentrations of IPTG from 40 to 400 µM to solubilise the protein as much as possible. The cell pellets from samples taken before and after induction were collected by centrifugation for 20 minutes at 4000 xg at 4°C. The pellets were resuspended in 500 µl ice

cold 1X PBS or STE buffer (Sigma; 85810) containing one complete mini EDTA-free protease inhibitor cocktail tablet per 10 ml (Roche; 11836170001). The suspension was sonicated on ice for 5 minutes (30 seconds on and 30 seconds off) at high level. A 20 μl sample of sonicated cell lysate (whole cell lysis) was taken to check the induction of soluble and insoluble protein. 1% of Triton-100X was added to the remaining sample and centrifuged at 12,000x g, 4°C for 30 minutes to precipitate insoluble cell parts. The supernatant (soluble fraction) was collected in new Eppendorf tubes.

The insoluble fraction (pellet) was resuspended in 300 μl of ice cold 1X PBS. Protein solubility was checked in 10 μl samples of both soluble and insoluble protein as well as in whole bacterial lysate and uninduced samples. Protein samples were mixed with an equal volume of 2x Laemmli buffer (Sigma-Aldrich; S3401) and heated at 100°C for 10 minutes with a quick vortexing every two minutes; samples were then run on 4-12% SDS-page gels for Coomassie blue staining.

2.14.8.2 Production and purification of insoluble GST protein using sodium lauroyl sarcosinate (sarkosyl)

(Adapted from Park *et al.* 2011)

The insoluble fraction GST-fusion protein was purified by resuspending the cell pellet in 300 μl of ice cold STE buffer containing 2% sarkosyl, 7 mM dithiothreitol (DTT) and 100 μg/ml lysozyme. Samples were then incubated on ice for 15 minutes and the suspension was sonicated on ice for 5 minutes (30 seconds on and 30 second off) at high level. Samples were centrifuged at 18,000 xg for 30 minutes and the supernatants were transferred to new Eppendorf tubes. A volume of 2% Triton X-100 was added to the cell lysate and incubated for 30 minutes on ice.

2.14.8.3 Expression and purification of insoluble GST protein using formic acid

(Adapted from Lee *et al.* 2008)

The GST-fusion proteins from genes of interest were purified from the other fraction of bacterial protein using a Pierce GST spin purification kit (Thermo; 16106) according to the

manufacturer's instructions. The protein yields were enhanced by adding 2% formic acid buffer (Sigma; F0507) to increase the amount of protein obtained from the columns.

2.14.9 Coomassie Brilliant Blue staining

The SDS-page gels containing induced proteins were immersed in 10 to 15 ml of Coomassie Instant Blue (Expedeon; ISB1L) for 1-2 hours with gentle shaking. Gels were washed several times with ultrapure water to remove residual stain. The gels were visualised using a Bio-Rad image reader.

2.15 Establishment of a double Tet-On 3G stable cell line.

2.15.1 Titration of a puromycin selection marker (kill dose curve)

A HeLa Tet-On 3G cell line was seeded into 6-well plates in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% Tet System Approved FBS (Clontech; 631106) and 100 µg/ml G418. The cells were incubated in a humidified incubator at 37°C and 5% CO_2 until the cell density reached the confluent stage. Seven doses of puromycin, from 0.25 to 2 µg, together with untreated cells, were tested to optimise the minimum dose of puromycin required to kill all cells after 3-5 days. The lowest optimum dose chosen for single colony selection was 1 µg/ml.

2.15.2 Generation of a double stable HeLa Tet-On 3G stable cell line

A double stable HeLa Tet-On 3G cell line with *PRDM9* or *PRDM7* isoform *(I)* or *(II)* was created by culturing a HeLa Tet-On 3G cell line in 6-well plates and growing the cells to a density nearly at the confluent stage. A 2 µg sample of recombinant vector pTRE3G containing *PRDM9*, or pTRE3G containing either isoform I or II of *PRDM7,* was then co-transfected in each well along with 100 ng of a linear puromycin selection marker (20:1 ratio) using Xfect transfection reagent (Clontech; PT5003-2) according to the manufacturer's protocols. Control cells were co-transfected with only the pTRE3G vector without any cloning. All plasmid cloning steps and purifications are described in Section 2.14. After 48 hours, cells were split into 4 x 10 cm dishes and grown for a further 48 hours. Fresh media containing 1 µg/ml puromycin and 100 µg/ml G418 were then added to each dish; the media, puromycin and G418 were refreshed every 4 days. A large, healthy single colony was

transferred into a 24-well plate using a cloning cylinder (Sigma; C2059-1EA). The cells were cultured and maintained with appropriate concentrations of selective antibiotic until they reached the confluent stage. Each well was then split into 3 wells of 6-well plates and transferred into T75 flasks. Some cells were frozen in liquid nitrogen and the others were used for screening to determine the best fold of gene induction.

2.15.3 Screening of a double stable HeLa Tet-On 3G stable cell line

Each colony obtained from Section 2.15.2 was tested in the presence and absence of doxycycline (Sigma; D9891-5G) to check the gene induction. For each individual clone to be tested, 1/4 of the total amount of cells was seeded into 4 x 10 cm dishes. Doxycycline was added to plates at a concentration of 1000 ng/ml; a negative control was grown in the absence of doxycycline. Cells were incubated at 37°C in 5% CO_2 for 48 hours. The cells were harvested and used for RT-PCR and qRT-PCR to compare induced to uninduced expression of *PRDM9* and *PRDM7 (I)* and *(II)*.

Chapter 3.0 RT-PCR analysis of selected potential novel cancer-testis antigen genes

3.1 Introduction

Sexual reproduction relies on a special form of cell division to reduce the chromosome number by half through the process known as meiosis. During meiosis, diploid cells in the testes and ovaries divide to produce four distinct haploid cells (sperm and egg) (reviewed in Petronczki *et al.*, 2003; Marston and Amon, 2004). Meiosis is a key source of genetic variation and information exchange between paternal and maternal chromosomes for the creation of new combinations of alleles in the population (Marston and Amon, 2004; Silkworth and Cimini, 2012). In contrast, mitotic cell division occurs in somatic cells as a part of tissue homeostasis and for replacement of dead cells and damaged tissues (Walczak *et al.*, 2010; Silkworth and Cimini, 2012). Meiosis and mitosis are tightly controlled and regulated; however, errors in these two processes may lead to defects in chromosome segregation and can cause aneuploidy. Genetic instability is one of the cancer hallmarks and results in alterations of tumour suppressor genes, proto-oncogenes and instability genes (Vogelstein and Kinzler, 2004; Negrini *et al.*, 2010).

Several meiosis-specific genes are expressed in different types of cancer; these genes include *SPO11* (Koslowski *et al.*, 2002), *SYCP1* (Tureci *et al.*, 1998), *SYCP3 (*Modarressi *et al.*, 2004) and *HORMAD1* (Chen *et al.*, 2005a). Each of these genes has a unique function in meiosis, involving chromosome alignment and meiotic recombination. Activation of the expression of these genes in cancer cells suggests that this may be a consequence of induction of a gametogenic programme during tumourigenesis (Old, 2001; Simpson *et al.*, 2005). Thus, the presence of meiotic proteins in cancer cells represent an excellent source of potential cancer diagnostic, vaccine and immunotherapy tools.

Meiosis-specific genes are a class of genes which considered to be a rich source of CTA genes (for example see introduction Section 1.8). In this project, two approaches were taken for identification of novel cancer testis CTA genes: (i) A manual literature search for meiosis-specific genes; (ii) a systematic approach using a bioinformatics pipeline. We initially hypothesised that meiosis-specific genes might be a rich source of novel CTA genes. Several of the known meiosis-specific genes have been identified as excellent candidate for CTA genes, which supported our hypothesis. For this reason, we select from the literature a group of meiosis-specific genes to initiate screening to test our hypothesis. In the present case, only one meiosis-specific gene was manually selected to start this project: the testis-specific serine/threonine kinase 1 (*TSSK1*) gene.

Manually searching in the literature for highly restricted CTA genes is not only time consuming but it can identify only a limited number of meiosis-specific genes. Thus, a bioinformatics pipeline was developed to identify novel CTA candidate genes based on previously published data from microarray and expressed sequence tag (EST) studies (Feichtinger *et al*., 2012a; 2012b; 2014a). In brief, this pipeline took advantage of a previously microarray study which identified 744 genes with highly restricted expression to meiotic spermatocytes in mice (Chalmel *et al.*, 2007). This group of genes was then mapped to the human orthologues, which resulted in the identification of 408 human genes (Figure 3.1). MitoCheck analysis (http://www.mitocheck.org) was applied to eliminate non-meiosis specific genes, yielding a final number of 375 genes. These genes were then checked through a previous bioinformatics pipeline using two methods: either microarray analysis from ArrayExpress and GEO (developed by Julia Feichtinger http://www.cancerma.org.uk/) or EST analysis based on the Unigene database (developed by Julia Feichtinger http://www.cancerest.org.uk/) (Feichtinger *et al.*, 2012a; 2012b; 2014a).

The 375 identified human orthologous genes were analysed using the microarray analysis tool against an 80 cancer datasets to determine gene up-regulation. Only 40 candidate genes were found to be predicted highly restricted CTA genes using this method. Due to the

limitations in identifying new candidate genes via this route a second bioinformatics approach was developed using EST data analysis (Feichtinger *et al.*, 2014a).

ESTs are unedited, short sub-sequences (200-500 nucleotides in length), generated by single pass sequencing of a 5′ and/or 3′ complementary DNA (cDNA) sequence that are afterward clustered and counted (Adams *et al.*, 1992; Feichtinger *et al.*, 2014a). Several studies have used this technique to identified potential CTA genes; for example, Hofmann et al. (2008) and Kim et al. (2007). The 375 human orthologous genes were analysed again using the EST pipeline (developed by J. Feichtinger), which identified 177 candidate genes. These genes were classified according to their expression in the EST pipeline to different sub-classes. Class 1 contains 75 genes that are testis-restricted; class 2 consists of 9 genes that are cancer/testis restricted; class 3 contains 21 genes that are cancer-testis/CNS restricted; and class 4 contains 72 genes that are testis/CNS restricted (Feichtinger *et al.*, 2012a; 2014a).

In this study, 2 genes were chosen from 40 candidate genes identified via microarray analysis and 21 genes were also selected from 177 candidate genes from EST analysis results. The aim of the work reports in this chapter was to validate these genes as potential CT genes/cancer biomarkers.
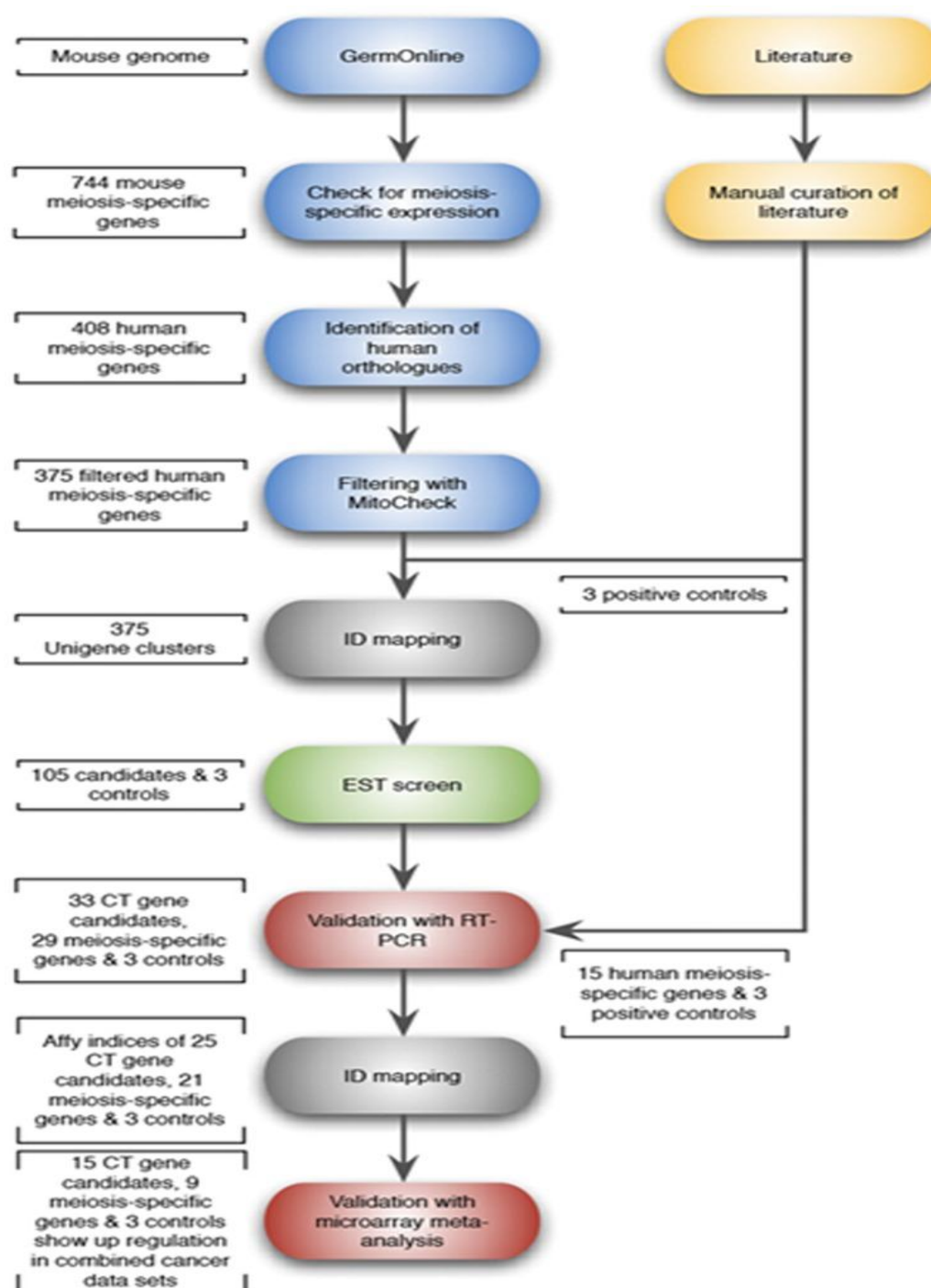
Figure 3.1. **Schematic flow diagram showing the pipeline for the identification of new potential cancer testis (CT) genes using bioinformatics tools.** Based on a large scale microarray study, 744 mouse meiosis-specific genes were selected as a starting point: 408 human orthologues could be identified and 375 human meiosis-specific genes remained after filtering to eliminate non-testis specific genes. All 375 candidates as well as 3 controls (*MAGE-C1, GAGE1* and *SSX2*) were fed into the EST analysis pipeline, which returned 105 candidate genes which were subjected to RT-PCR validation/microarray meta-analysis.

3.2   Results

3.2.1   Identification and validation of the putative meiosis-specific *TSSK1B* gene as a potential CT antigen

Based on manually searching of meiosis-specific genes in the literature, the *TSS1KB* gene was selected for initial screening (Table 3.1). Reverse transcriptase PCR (RT-PCR) was used to analyse the *TSSK1B* gene expression in 21 human normal tissues. A panel of total RNA from 21 human normal tissues was purchased from Clontech and Ambion, which included the positive control testis and other 20 other normal adult tissues. cDNA was synthesised from the total RNA preparations and the cDNA quality was assessed by RT-PCR of the *β-actin* gene. RT-PCR primers were designed to span more than one intron, where possible to avoid genomic DNA contamination. Previously identified CT antigens genes *GAGE1*, *MAGEC1* and *SSX2* were used as positive control genes.

The RT-PCR profile of *TSSK1B* shows a strong band in the positive control testis and no expression was observed in other normal tissues with the exception of a faint band in the foetal brain (Figure 3.2). To confirm that the correct target gene was amplified, the two bands derived from the testis and foetal brain were sent for DNA sequencing, and the result showed identical matching of the gene (see Table 3.4 for a sequencing summary).

Based on the RT-PCR analysis result for the normal adult tissues, the *TSSK1B* was classified as a testis/CNS restricted candidate gene, although the CNS expression is only in foetal tissue. To address whether this gene might be a potential CTA gene, *TSSK1B* expression was analysed in 33 cancer cell lines and solid tumours (Figure 3.3). It was expressed in different cancer cell lines and solid tumours including a moderate RT-PCR band in lung (MRC-5) and ovarian cancer cells (TO14); or faint expression in prostate cancer cells (PC-3), breast (MCF-7 and MDA-MB-453) and epidermal carcinoma cells (A-431). Some of the bands produced by RT-PCR were purified and DNA sequenced to ensure the correct gene sequence had been amplified (see Table 3.4 for the sequencing results).

Table 3.1. Gene function roles of the meiosis-specific gene *TSSK1B*, which was selected for initial screening for CT genes.

Gene Name	Known function	Classification after validation	Reference
TSSK1B	Meiotic serine/threonine kinase	Restricted CT/CNS antigen gene	(Li *et al.*, 2011)
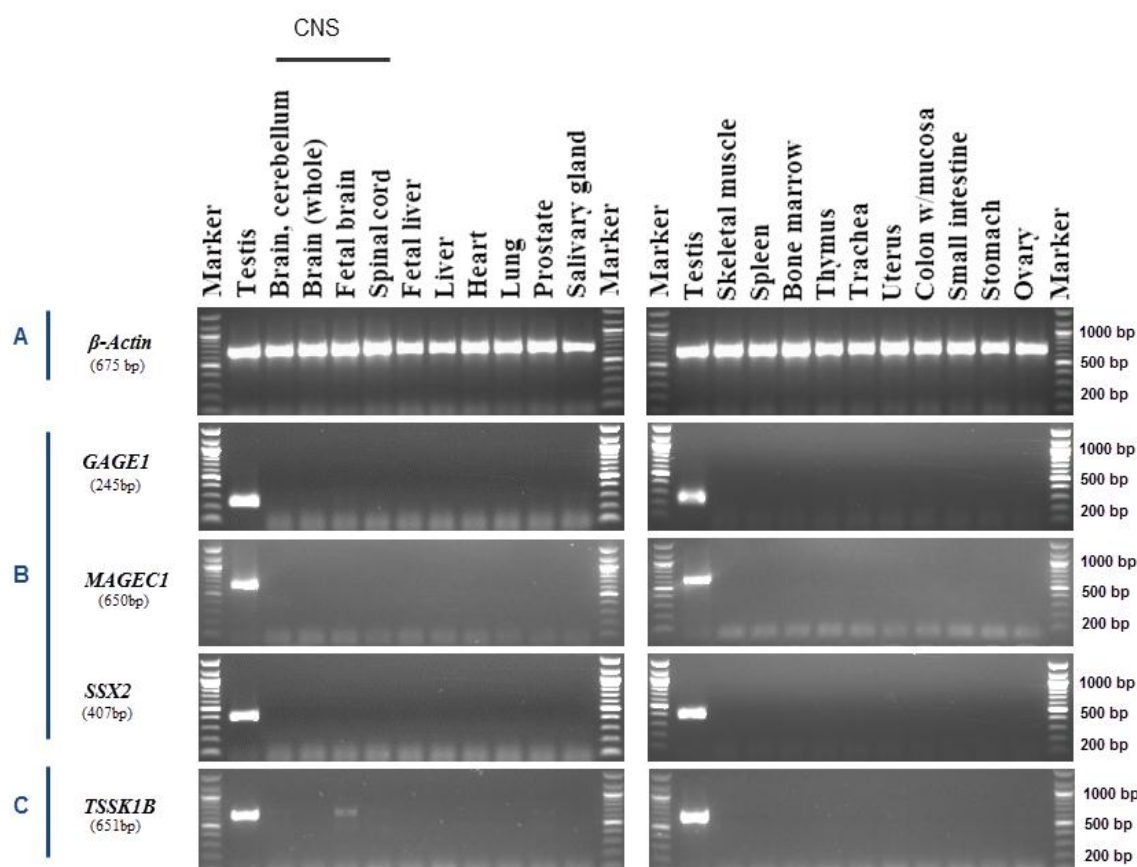
Figure 3.2. **RT-PCR analysis of *TSSK1B* gene in normal human tissues.** Agarose gels showing the *TSSK1B* gene expression profile in different normal human tissues. The cDNA was generated from 21 normal human tissues RNA (collected *post mortem*). The *TSSK1B* gene appears to be a testis/CNS restricted candidate gene, which is expressed in testis and foetal brain. Gene sequencing of two bands appearing in the testis and foetal brain show significant similarity to *TSSK1B* gene. (A) Expression of the *β-Actin* gene is a positive control for the cDNA samples. (B) The expression profile of *GAGE1*, *MAGEC1* and *SSX2* genes, which were previously identified as CTA genes, evaluated in normal human tissues. (C) *TSSK1B* profile
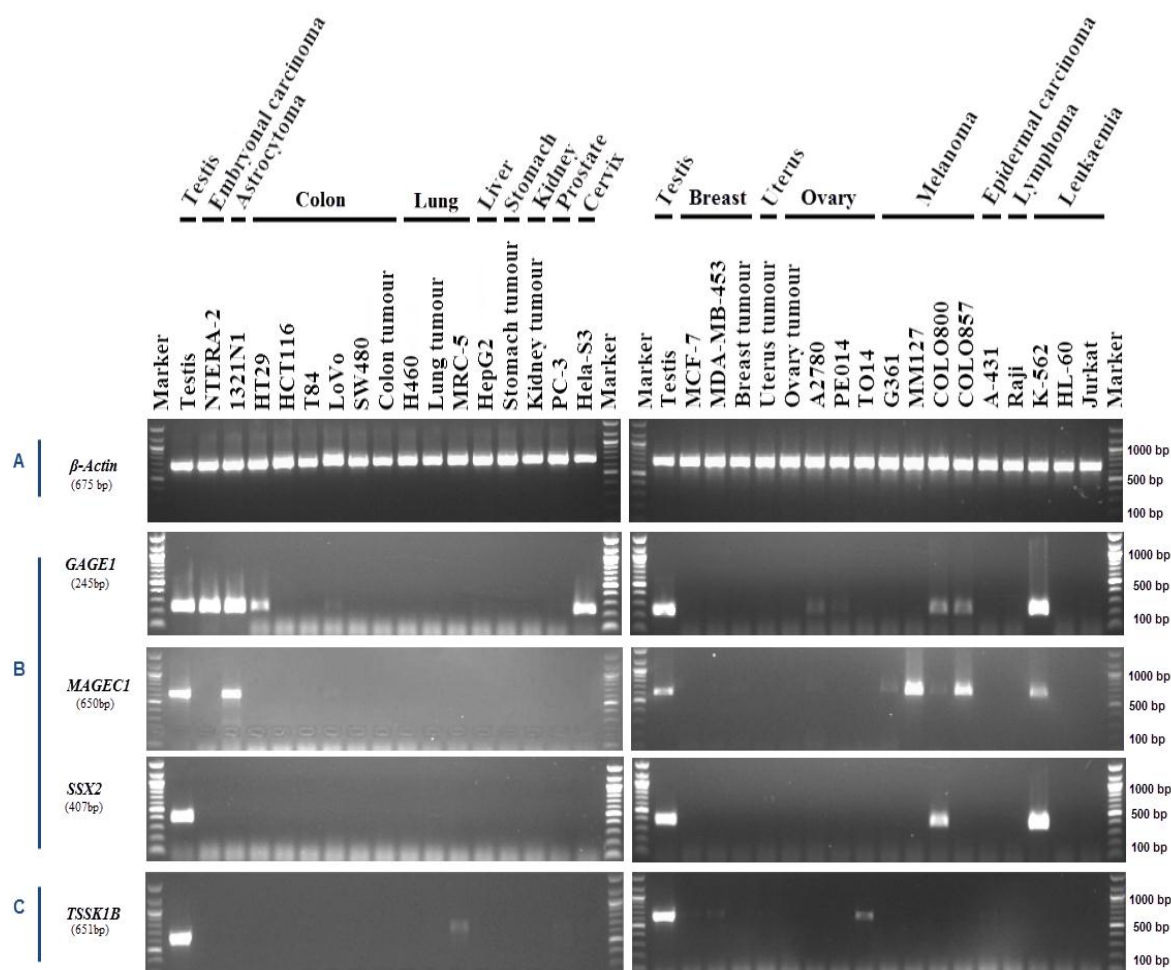
Figure 3.3. RT-PCR expression analysis of the *TSSK1B* gene in cancer cell lines and solid tumours. Agarose gels showing the *TSSK1B* gene RT-PCR analysis in some cancer cell lines and tissues. cDNA was generated from 33 cancer cell lines and solid tumours. The *TSSK1B* gene appears to be expressed in a range of cancer cell lines including lung, breast and ovarian cancer cell lines. (A) Analysis of the *β-Actin* gene is a positive control for the cDNA samples. (B) The expression profile of *GAGE1*, *MAGEC1* and *SSX2* genes, which were previously identified as CT antigen genes, evaluated in cancer cell lines and tissues. (C) Analysis of *TSSK1B* expression in cancer cells.

3.2.2 Identification and validation of potential novel CT antigen genes, predicted through bioinformatics analysis.

The literature search of highly restricted CTA genes identified only a limited number of genes, only one gene *TSSK1β* which was analysed here (Table 3.1). The meiosis-specific genes searched in the literature appear to be an excellent source of potential CTA genes. However, manual curation in the literature of new CTA genes is limited due to the few meiosis-specific-genes characterised to date, and it is also time-consuming. This study took advantage of the bioinformatics approach by using previously published microarray data and EST analysis data (Figure 3.1) (Feichtinger *et al*., 2012a; 2012b; 2014a).

3.2.3 Identification and validation of potential novel CT antigen genes predicted through microarray analysis.

Two genes, *ALX1* and *WDR20,* were selected from 40 candidate genes identified via a microarray analysis study developed by J. Feichtinger (www.cancerma.org.uk/). In order to assess their meiosis specificity (Table 3.2), the expression of these genes was validated using 21 normal human tissues RNA subjected to 40 cycles of RT-PCR. cDNA was generated from the total RNA preparations obtained from Clontech and Ambion. The cDNA quality was assessed using the expression of the *β-actin* gene. Again, *GAGE1*, *MAGEC1* and *SSX2,* were used as positive control CTA genes and intron-spanning primers were designed for each gene to avoid genomic DNA contamination.

The RT-PCR analysis of the *ALX1* and *WDR20* genes gave bands in multiple normal human tissues (Figure 3.4). *ALX1* was expressed in 5 normal tissues as well as in normal testis tissues, whereas *WDR20* was expressed in all normal tissues. Based on these results, the *ALX1* and *WDR20* genes were dismissed and not validated in cancer tissues. The PCR products of these genes were sequenced to ensure the correct gene had been amplified and the results of the sequencing are shown in Table 3.4.

Table 3.2. **Gene function roles of predicted meiosis-specific genes through microarray analysis.**

Gene Name	Functional known	Classification after validation	Referencing
ALX1	May play a role in cervix development and chondrocyte differentiation	Dismissed	(Zhao *et al.*, 1993; Cai, 1998)
WDR20	It is regulate the activity of the USP12-UAF1 deubiquitinating enzyme complex.	Dismissed	(Kee *et al.*, 2010)
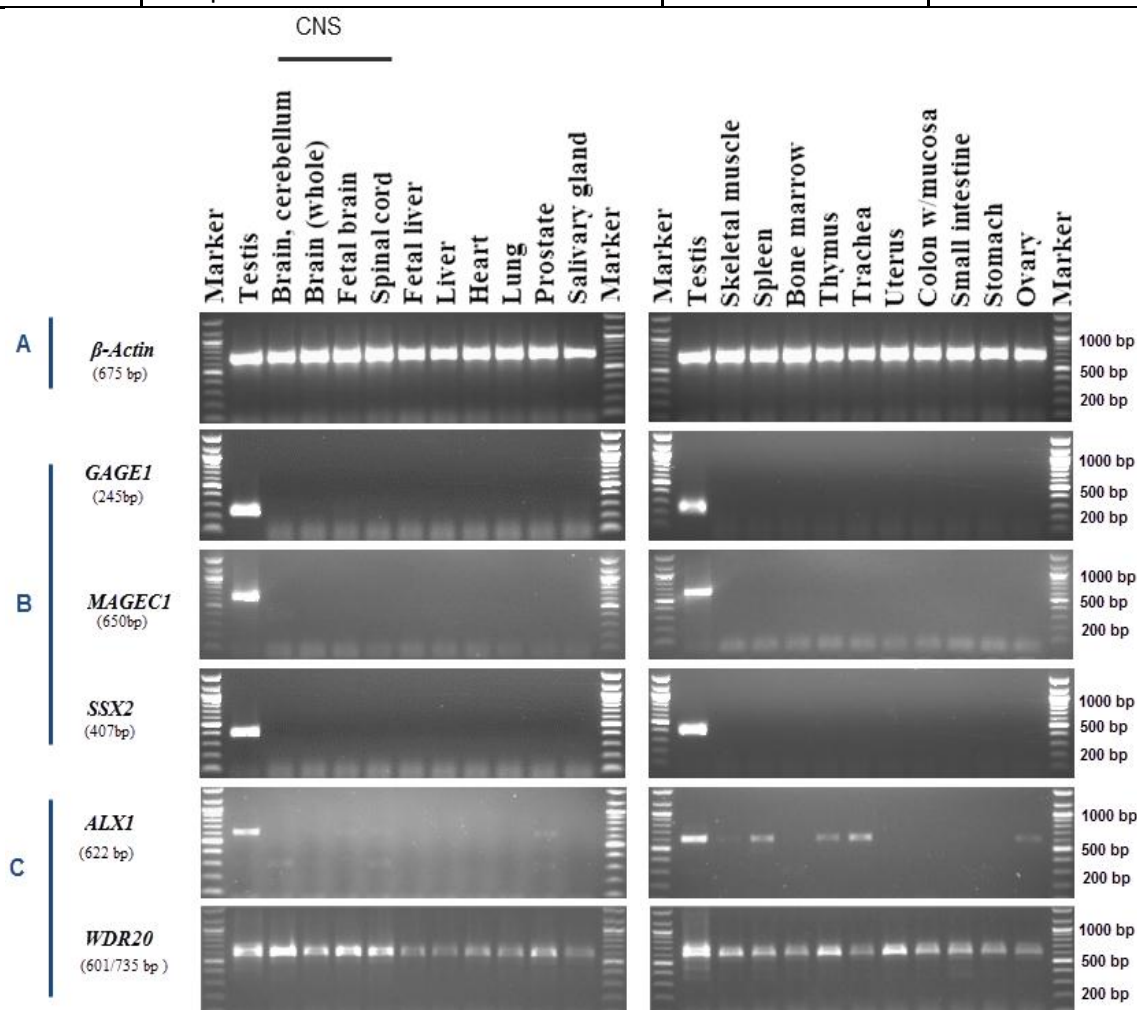
Figure 3.4. **RT-PCR validation for *ALX1* and *WDR20* genes in normal human tissues.** Agarose gels showing the expression profile of predicted candidate genes in 21 normal human tissues. cDNA was generated from the total RNA obtained from a range of normal human tissues (collected *post mortem*). The *ALX1* gene was expressed in multiple normal tissues whereas the *WDR20* gene was expressed in all tissues (C). (A) Expression of the *β-Actin* gene as a positive control for the cDNA samples. (B) The expression profile of the *GAGE1*, *MAGEC1* and *SSX2* genes, which were previously identified as CT antigen genes, evaluated in normal human tissues.

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

The second bioinformatics approach used to identify novel cancer-testis-antigen genes was an EST pipeline (Feichtinger *et al.*, 2012a; 2014a). A selection of 21 genes was made from the 177 candidate genes predicted to be meiosis-specific genes (www.cancerest.org.uk/). These genes were classified according to their expression profile in the EST pipeline into three sub-classes (Table 3.3). Class 1 contained 13 predicted testis-restricted genes; class 2 consists of 2 predicted cancer/testis restricted genes; class 3 contained 6 predicted cancer-testis/CNS restricted genes (Feichtinger *et al.*, 2014).

The meiosis specificity was assessed by validating expression of these genes in 21 normal human tissues using 40 cycles of RT-PCR. The cDNA was generated from the total RNA preparations from normal human tissues obtained from Clontech and Ambion. The cDNA quality was assessed using the *β-actin* gene. Previously identified CTA genes, *GAGE1*, *MAGEC1* and *SSX2,* were used as positive control genes and intron-spanning primers were designed for each gene to avoid genomic DNA contamination.

Based on the RT-PCR profile from the normal somatic tissues, these genes where classified into different groups: (i) dismissed genes, if the expression appeared in more than two normal human tissues (non-testis); these genes were not analysed further in the cancer cell lines and tissues; (ii) testis-restricted genes that showed restricted expression to the testis and no expression in normal somatic (non-testis) tissues; (iii) testis/CNS-restricted genes that showed expression in the testis and central nervous system (CNS) only; (iv) testis-selective genes, which were expressed in the testis and in no more than two normal somatic tissues. (v) testis/CNS selective genes, which were expressed in the testis, CNS and no more than two normal human tissues. Further analyses of these genes were carried out (with the exception of the dismissed group) in 33 cancer cell lines and tissues.

Table 3.3. **Functional roles of predicted novel CTA genes in an EST analysis pipeline.**

Gene Name	Known function	Referencing
	Class 1: Testis-restricted	
C4orf17	Protein function uncharacterised, unknown function.	
C9orf11	Cell membrane protein and Involve in acrosome formation in murine testis.	(Li *et al*., 2006)
C9orf43	Protein function uncharacterised, unknown function.	
C12orf12	Protein function uncharacterised, unknown function.	
C16orf92	Protein function uncharacterised, unknown function.	
DDX4	May play a role in sperm motility and in germ cell development.	(Castrillon *et al*., 2000)
IQCF3	Protein function uncharacterised, unknown function.	
MAGE-B5	Protein function uncharacterised, unknown function but it was found to express in testis and a range of tumours.	(Lucas *et al*., 2000)
NT5C1B	Plays a role in producing the adenosine important in biochemical processes in the body.	(Wade *et al*., 2013)
PRDM7	Probably involved in transcriptional regulation and it might participate in cell differentiation and tumourigenesis.	(Fumasoni *et al*., 2007)
TDRD12	Plays a role in RNA biogenesis by interaction with Piwi protein in mice.	(Pandey *et al*., 2013)
TMEM202	Protein function uncharacterised, unknown function.	
TNRC18	trinucleotide repeat containing 18.	
	Class 2: Cancer/testis-restricted	
ARRDC5	Protein function uncharacterised, unknown function.	
ODF4	Also known OPPO1 and it may be involved in the sperm tail structure.	(Kitamura *et al*., 2003)
	Class 3: Cancer/testis-CNS restricted	
APOBEC4	May play a role in RNA/DNA editing cytidine deaminase to uridine.	(Rogozin *et al*., 2005)
C9orf12	Also known IPPK; Inositol-pentakisphosphate 2-kinase and it may play a role to protect cells from TNF-alpha-induced apoptosis.	(Verbsky *et al*., 2002)
C9orf128	Protein function uncharacterised, unknown function.	
KDM4D	Also known as JMJD2D, lysine (K)-specific demethylase 4D and plays a role in the histone core. It interacts with the tumour suppressor Rb, histone deacetylases (HDACs) and has diverse genomic functions.	(Gray *et al*., 2005)
PRSS54	Protein function uncharacterised, unknown function.	
SEPT12	Involved in formation of the sperm head and tail in the spermatogenesis process.	(Kuo *et al*., 2013)

3.2.5 RT-PCR analysis of predicted CT genes in normal human tissues

The RT-PCR profiles of 5 out of 21 genes showed expression in multiple normal human tissues: these were *APOBEC4*, *C9orf12*, *C9orf43*, *KDM4D* and *TNRC18*. *C9orf12*, *C9orf43* and *TNRC18* were expressed in all normal tissues including the testis, whereas *APOBEC4* and *KDM4D* were expressed in 12 normal tissues (Figure 3.5). The RT-PCR products of these genes were purified and sequenced to confirm that the correct DNA sequence was amplified (summary of sequencing in Table 3.4). Based on these results, these genes were dismissed and the expression was not analysed in cancer cell lines and tissues. Three of the dismissed genes (*APOBEC4*, *C9orf12* and *KDM4D*) were classified as class 3 genes (cancer-testis/CNS restricted) after the EST pipeline analysis. In contrast, *C9orf43* and *TNRC18* were classified as class 1 (testis-restricted). The RT-PCR validations identified twelve genes that were apparently testis restricted. No expression was determined in other somatic (non-testis) tissues, confirming the classification as testis-restricted genes. These genes were *ARRDC5*, *C4orf17*, *C16orf92*, *DDX4*, *IQCF3*, *NT5C1B*, *TMEM202*, *C12orf12*, *MAGE-B5*, *ODF4*, *PRDM7* and *TDRD12* (Figure 3.6 A and 3.6 B). Only *MAGE-B5* is located on chromosome X; the other genes have an autosomal location. DNA sequencing of the testis bands of these genes confirmed the correct sequence of the gene being amplified (Table 3.4 for sequencing summary).

C16orf92 showed a faint band in the testis with an expected size of 279 bp. Two bands with unexpected size (about 350 bp) appeared in the skeletal muscle and bone marrow; these were sequenced and showed no relation with the *C16orf92* gene indicating they are the product of mispriming. The *ODF4* gene was previously reported to be a CTA gene, with no expression in normal tissues and high expression in leukaemia and prostate cancer (Ghafouri-Fard *et al*., 2012; Ghafouri-Fard and Modarressi, 2012). The RT-PCR expression of the *ODF4* gene shows several unexpected bands appearing higher and lower than the expected size (602 bp). DNA sequencing of the RT-PCR product of these bands (approximately 800 bp in brain cerebellum and less than 400 bp in other tissues) did not

show any significant relationship to the *ODF4* indicating they are the products of cross reactivity (summary of sequencing, Table 3.4).

ODF4 and *ARRDC5* were classified as class 2 (cancer/testis-restricted) based on the original EST pipeline data, whereas the *C4orf17*, *C16orf92*, *DDX4*, *IQCF3*, *NT5C1B*, *TMEM202*, *C12orf12*, *MAGE-B5*, *PRDM7* and *TDRD12* genes were sub-classified after the EST analysis as a class 1 (testis-restricted). Most of these genes are predicted to encode proteins with uncharacterised functions and they might be novel potential CTA genes. To address whether these genes might be potential CTA genes, expression of these genes was analysed in 33 cancer cell lines and solid tumours.

The RT-PCR expression profile of the *SEPT12* gene indicated expression in the normal testis tissue with an expected size of 608 bp (Figure 3.7). Two faint bands appeared in both the brain cerebellum and the foetal brain. DNA sequencing of the PCR products of these bands and the normal testis tissue showed high similarity to the *SEPT12* gene (for sequencing results, see Table 3.4). Based on this result, *SEPT12* was classified as testis/CNS restricted, which fit with the EST pipeline classification of this gene as a class 3 cancer-testis/CNS restricted gene. To address whether this gene might be a potential CTA gene, *SEPT12* expression was analysed in 33 cancer cell lines and solid tumours.

C9orf128 was originally classified as a cancer/testis-CNS restricted gene based on the expression profile obtained from the EST pipeline. However, the RT-PCR profile of *C9orf128* shows a faint band in the prostate and trachea in addition to a strong band in the normal testis with the expected size of 731 bp (Figure 3.8). Therefore, *C9orf128* was classified according to the RT-PCR profile analysis as testis-selective. The PCR products of the testis, prostate and trachea were sequenced and these results showed significant identity to this gene. At this stage, *C9orf128* was not dismissed but its expression was analysed in the 33 cancer cell lines and tissues (see summary of sequencing, Table 3.4).

Based on the original EST pipeline result, *C9orf11* was classified as testis-restricted while *PRSS54* was cancer/testis-CNS restricted. The RT-PCR validations of *C9orf11* and *PRSS54*

genes shows apparent expression in the testis, CNS and no more than two normal somatic tissues (Figure 3.9). *C9orf11* gave faint bands in the brain cerebellum, foetal brain, spinal cord, lung and uterus, with the expected size of 652 bp, whereas *PRSS54* was expressed in brain cerebellum, spinal cord and prostate with the expected RT-PCR band size of 646 bp. DNA sequencing of the RT-PCR products from the testis and two somatic tissues showed significant identity with these genes (summary of gene sequencing results in Table 3.4). According to their RT-PCR profile, these genes were classified as testis/CNS selective.
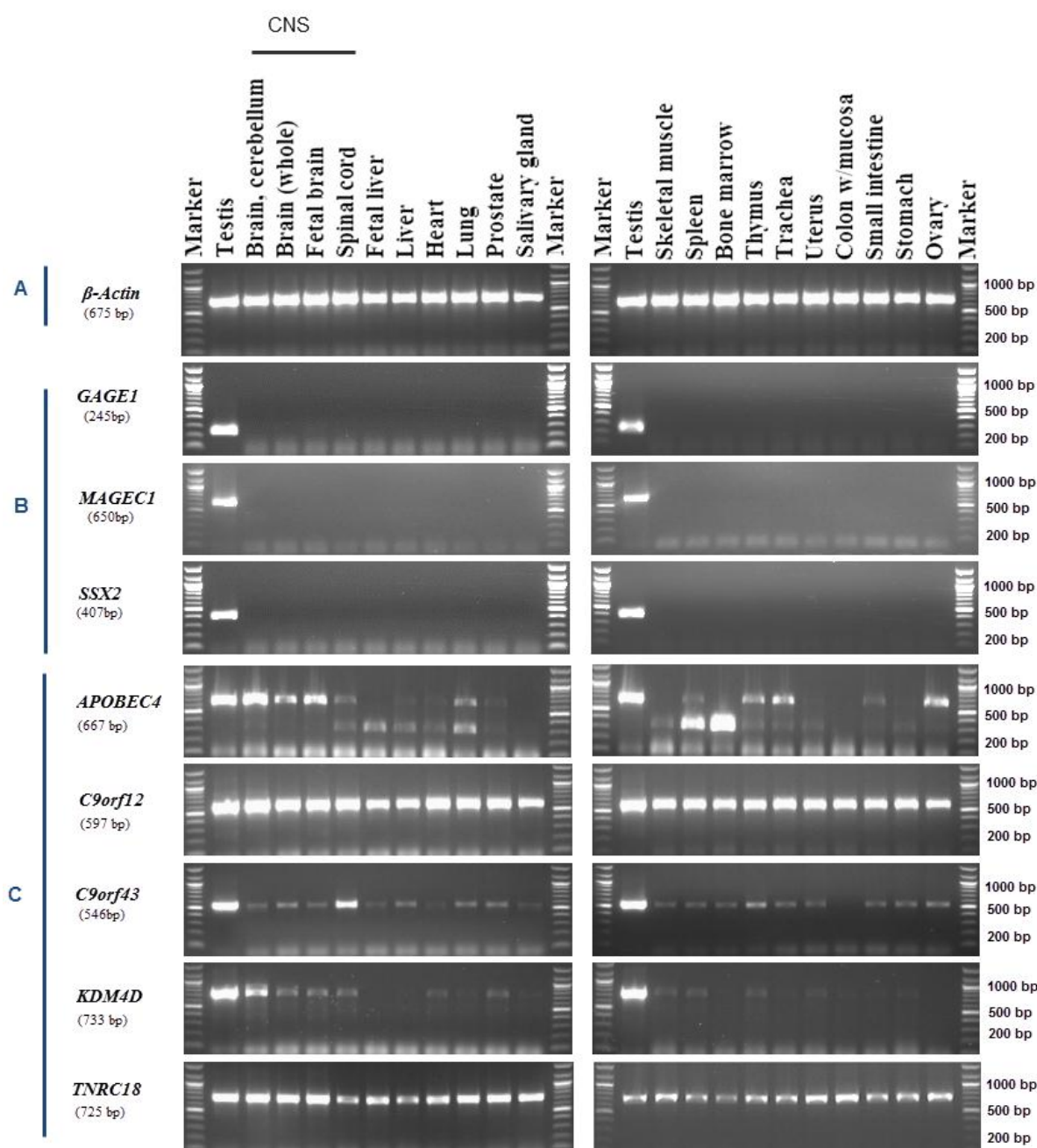
Figure 3.5. **RT-PCR validation for dismissed genes, which were derived from the EST pipeline, in normal human tissues.** Agarose gels showing the expression profiles of *APOBEC4*, *C9orf12*, *C9orf43*, *KDM4D* and *TNRC18* genes in 21 normal human tissues. cDNA was generated from the total RNA prepared from different normal human tissues (collected *post mortem*). (C) All these genes were expressed in most of the normal tissues. (A) Expression of the *β-Actin* gene is a positive control for the cDNA samples. (B) The expression profile of *GAGE1*, *MAGEC1* and *SSX2* genes, which were previously known as CT antigen genes, evaluated in normal human tissues.
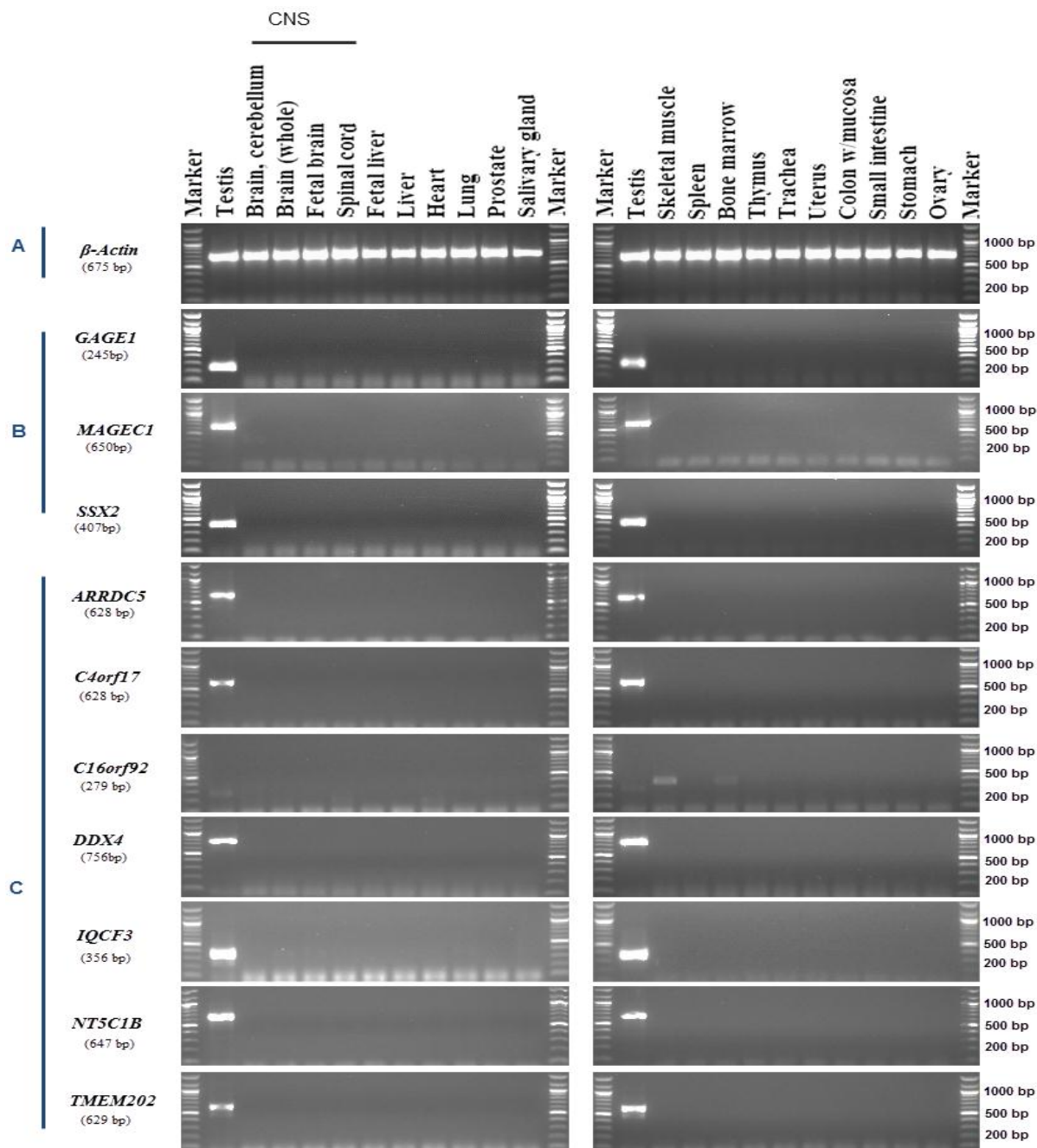
Figure 3.6 A. RT-PCR analysis of the testis-restricted CT genes obtained from an EST pipeline in normal human tissues. Agarose gels showing the expression profile of seven testis-restricted candidate genes in 21 normal human tissues. cDNA was generated from the total RNA prepared from a range of normal human tissues (collected *post mortem*). (C) All these genes showed no expression in normal somatic (non-testis) tissues. For *C16orf92,* DNA sequencing of two bands in skeletal muscle and bone marrow showed no relationship with the *C16orf92* gene. (A) Expression of *β-Actin* gene is a positive control for the cDNA samples. (B) The expression profile of *GAGE1*, *MAGEC1* and *SSX2* genes, which were previously identified as CT antigen genes, evaluated in normal human tissues.
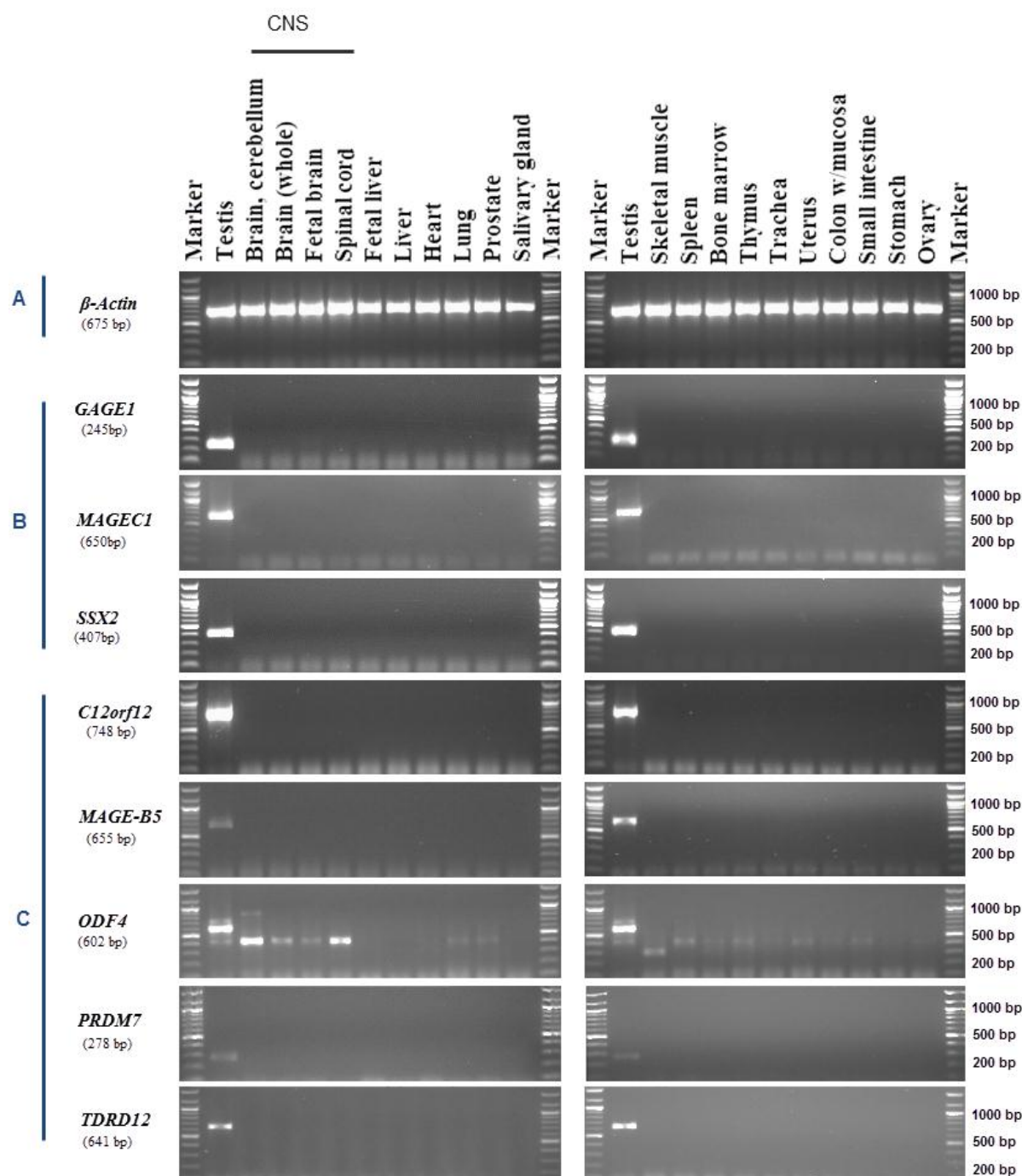
Figure 3.6 B. **RT-PCR expression profiles for the testis-restricted CT genes, obtained from an EST pipeline, in normal human tissues.** Agarose gels showing the RT-PCR analysis of five testis-restricted candidate genes in 21 normal human tissues. cDNA was generated from the total RNA prepared from many normal human tissues (collected *post mortem*). (C) All these candidate genes appear to be expressed only in the testis. For the *ODF4* gene, the higher band and the lower bands were DNA sequenced in different normal tissues and showed no match with the *ODF4* gene indicating they are the products of cross reactivity. (A) Expression of the *β-Actin* gene is a positive control for the cDNA samples. (B) The expression profile of *GAGE1*, *MAGEC1* and *SSX2* genes, which were previously identified as CT antigen genes, evaluated in normal human tissues.
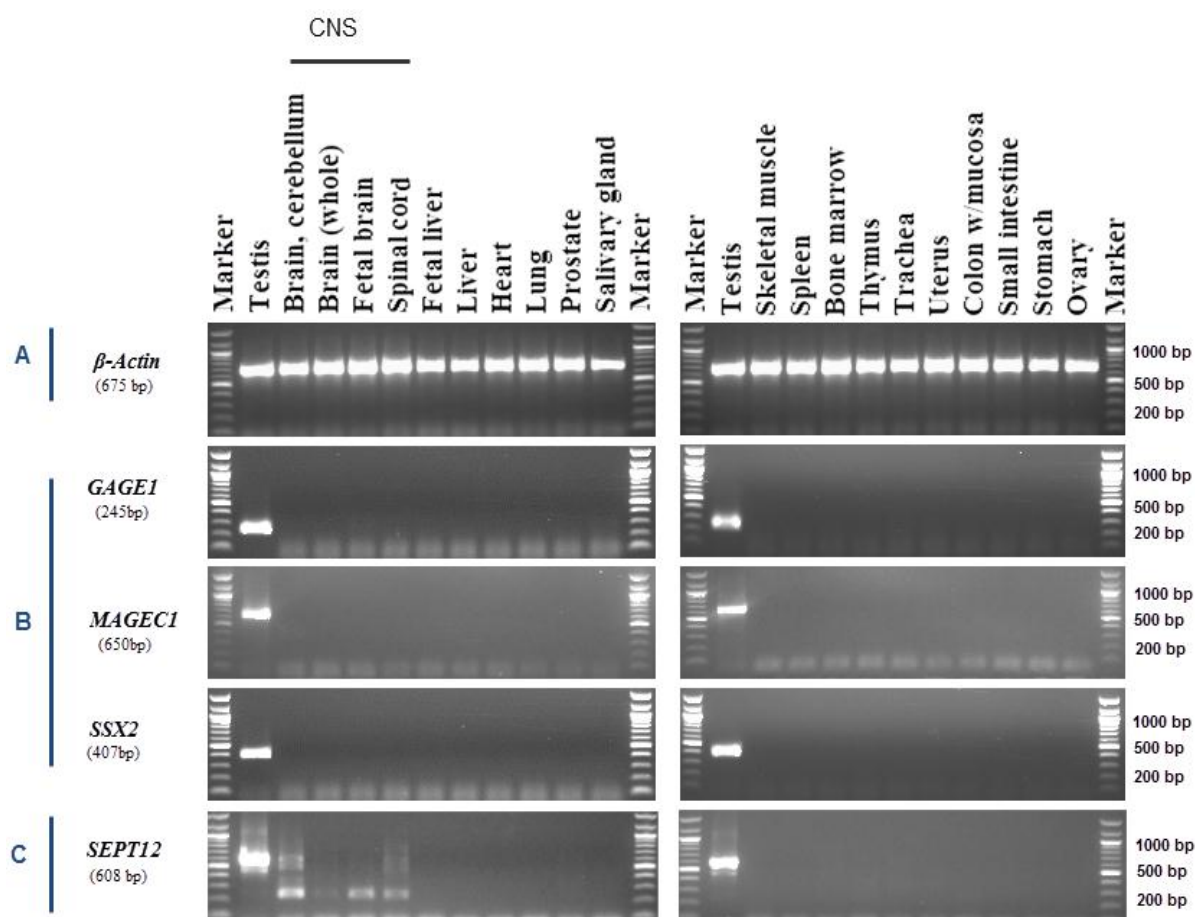
Figure 3.7. **RT-PCR expression profile for the *SEPT12* gene in normal human tissues.** Agarose gels showing the RT-PCR analysis of the testis/CNS-restricted *SEPT12* candidate gene in 21 normal human tissues. cDNA was generated from a range of normal human tissues (collected *post mortem*). (C) *SEPT12* is expressed in testis and CNS normal tissues (Brain cerebellum and Spinal cord). The lower bands were sequenced in CNS normal tissues and found to be unrelated to the *SEPT12* gene. (A) Expression of the *β-Actin* gene is the positive control for the cDNA samples. (B) The expression profile of *GAGE1*, *MAGEC1* and *SSX2* genes, which were previously identified as CT antigen genes, evaluated in normal human tissues.
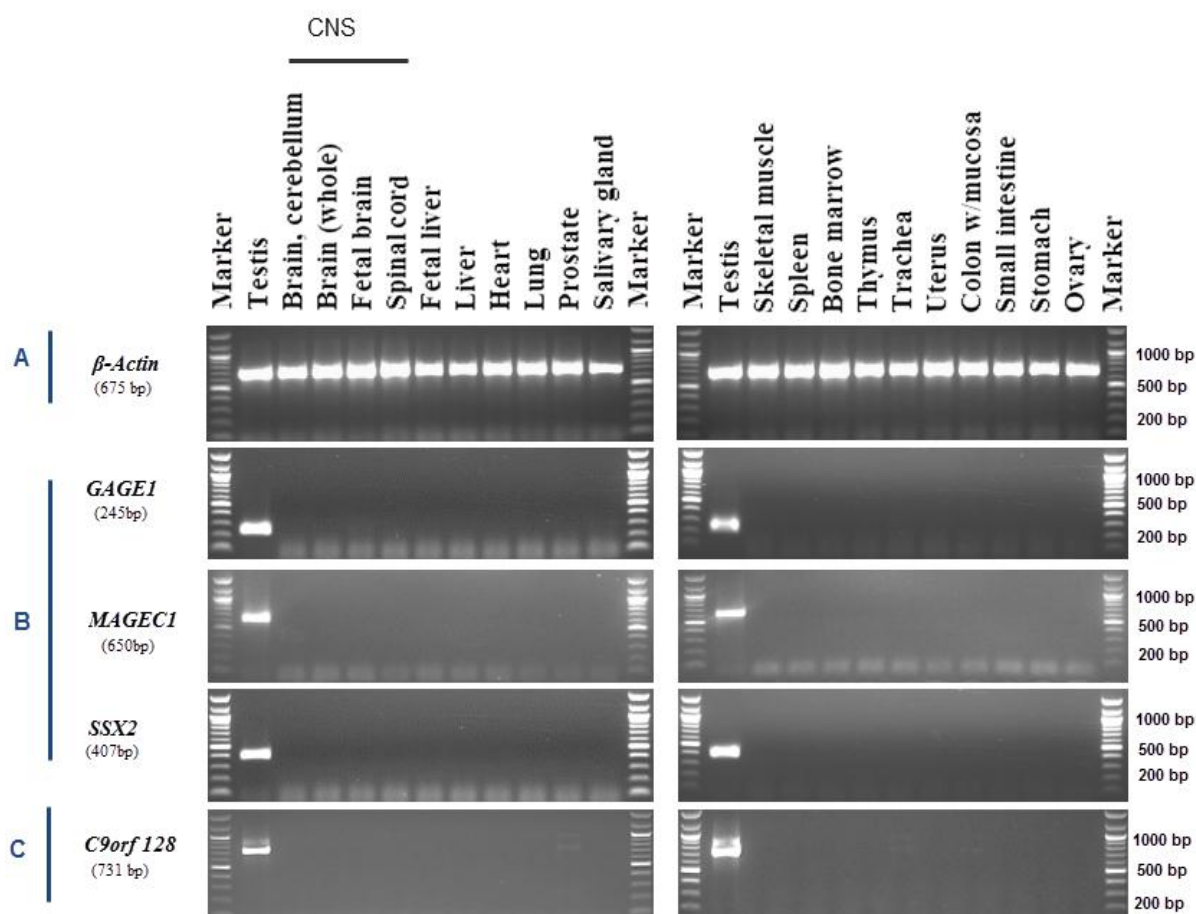
Figure 3.8. **RT-PCR expression profiles for the *C9orf128* gene in normal human tissues.** Agarose gels showing the RT-PCR analysis of testis selective *C9orf128* candidate gene in 21 normal human tissues. cDNA was generated from a range of normal human tissues (collected *post mortem*). (C) *C9orf128* was expressed in the testis and faint bands appeared in prostate and trachea tissues. Gene sequencing of these bands confirmed a relationship with the *C9orf128* gene. (A) Expression of *β-Actin* gene is the positive control for the cDNA samples. (B) The expression profile of *GAGE1*, *MAGEC1* and *SSX2* genes, which were previously identified as CT antigen genes, evaluated in normal human tissues.
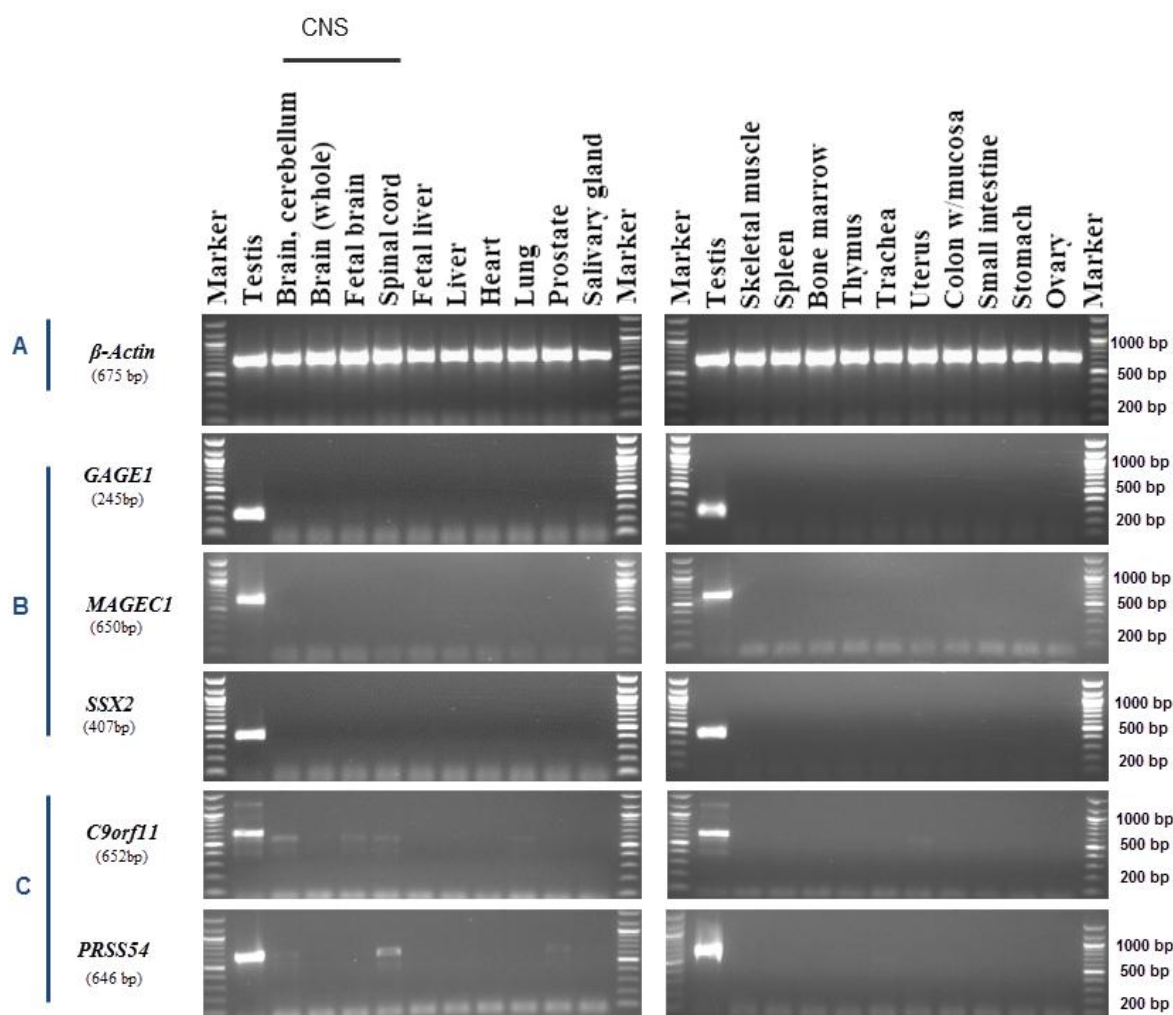
Figure 3.9. **RT-PCR expression profile for the testis/CNS-selective CT antigen *C9orf11* and *PRSS54* genes in normal human tissues.** Agarose gels showing the RT-PCR analysis of *C9orf11* and *PRSS54* in 21 normal human tissues. cDNA was generated from a range of normal human tissues (collected *post mortem*). (C) Expression of *C9orf11* and *PRSS54* in the testis and a number of normal tissues including CNS normal tissues. (A) Expression of the *β-Actin* gene is a positive control for the cDNA samples. (B) The expression profile of *GAGE1*, *MAGEC1* and *SSX2* genes, which were previously identified as CT antigen genes, evaluated in normal human tissues.

3.2.6 RT-PCR analysis of predicted CT genes in cancer cell lines and solid tumours

Based on the RT-PCR analysis result of the CT predicted genes in the normal human tissues, 16 out 21 genes were found to be candidate genes, whereas the other 5 genes were dismissed due to their expression in multiple normal human tissues. The 16 genes were classified based on their expression in the normal somatic tissues as 12 genes that were testis-restricted, one gene that was testis/CNS-restricted, one gene that was testis-selective and two genes that were testis/CNS-selective. To address whether these genes could be a potential CTA genes, their expression was assessed in 33 cancer cell lines and solid tumours from 16 different types of cancers. Most of the genes identified are located on autosomes, with exception of *MAGE-B5,* which is X encoded. According to the RT-PCR analysis result, 14 of the 21 genes fit the classification of the original EST pipeline, whereas 7 genes showed different results in the RT-PCR analysis.

The RT-PCR validations of 16 genes in cancer cell lines showed 7 genes expressed in the testis only. No expression of these genes was observed in the 33 cancer cells and tissues. These genes were *ARRDC5*, *C4orf17*, *C16orf92*, *DDX4*, *IQCF3*, *NT5C1B* and *TMEM202* genes (Figure 3.10). *ARRDC5* was classified based on the original EST data analysis as cancer/testis-restricted but no expression was observed in the cancer cells. *ARRDC5* was analysed in the normal human tissues and appeared as a testis-restricted candidate. The other 6 genes were classified based on the original EST analysis as testis-restricted and the RT-PCR validation in the normal somatic tissues and cancer cells showed similar results to the EST data analysis profile. To sum up, these genes were classified according to the PCR profile results as belonging to the testis-restricted class.

RT-PCR products for these genes from the testis were sent for DNA sequencing and the results showed high identity with each gene (summary of sequencing, Table 3.4). The *IQCF3* gene showed a high molecular weight band with an unexpected size of approximately 1000 bp in an ovarian cancer cell line (TO14) and the DNA sequence indicated no relation with

IQCF3 gene. In addition, the *C16orf92* gene showed some faint bands with unexpected size of about 400 bp in breast cancer cell lines (MCF-7 and MDA-MB-453); again DNA sequencing confirmed this band to be unrelated to the *C16orf92* gene.

The second group of genes was classified based on the RT-PCR analysis in normal and cancer cells as cancer/testis restricted. Among the 16 positive candidate genes showed expression in less than two normal tissues, only five genes were expressed in the testis (no expression in other somatic tissues) and expressed in several of the cancer cell lines and solid tumours. These genes were *C12orf12*, *MAGE-B5*, *ODF4*, *PRDM7* and *TDRD12* (Figure 3.11). Based on the original EST pipeline data, these genes were classified as testis-restricted, with exception of *ODF4,* which was classified as cancer/testis-restricted.

C12orf12 gene was expressed in the testis with the expected size of 748 bp, with a strong band present in the ovarian TO14 cell line and faint bands in colorectal cancer cell line (SW480), ovarian tumour, and the ovarian PE014 cell line. The *MAGE-B5* gene is the only member of this class located on the X chromosome and it was expressed in the testis and ovarian cancer cell line (TO14) (summary of gene sequencing, Table 3.4).

ODF4 was previously published as a prostate and leukaemia CTA gene (Ghafouri-Fard *et al*., 2012; Ghafouri-Fard and Modarressi, 2012). Further analysis of this gene was performed in 16 other types of cancer. *ODF4* was expressed in a range of cancer cell lines and tissues including a faint band in embryonic carcinoma, brain cancer, liver cancer, prostate cancer and uterine cancer. This gene also showed moderate bands in breast cancers, ovarian cancers, melanoma and leukaemia. Sequencing of unexpected bands (less than 400 bp) in the colon cancer cell line (Lovo), lung cancer (MRC-5) and melanoma (Colo800) revealed no significant similarity to *ODF4* (Table 3.4).

The *PRDM7* gene contains two isoforms and intron-spanning primers were designed to cover both isoforms with an expected size of 278 bp. This gene showed restricted expression to the

testis in the normal tissues and showed a faint band in a melanoma cancer cell line (G361) and leukaemia cell line (K562). Gene sequencing of these bands revealed high similarity to the *PRDM9* gene. Unexpected sizes bands appeared in the ovarian cancer cell line (TO14) and melanoma cell line (colo857) but were unrelated to the *PRDM7* gene. The *TDRD12* gene was expressed in the normal testis and no gene expression was observed in other somatic tissues. The expression of this gene was also detected in the embryonic carcinoma cancer cell line (NT2). Gene sequencing performed on the purified PCR product showed high similarity to *TDRD12* (summary of sequencing, Table 3.4).

SEPT12 was classified as a class 3 cancer-testis/CNS restricted gene based on the original EST pipeline classification. The RT-PCR profile of *SEPT12* in normal somatic tissues showed it is expressed in the testis as well as in the CNS and the gene was classified as testis/CNS restricted. Further analysis of *SEPT12* in cancer cell lines showed expression in a colorectal cancer cell line (SW480), breast cancer cell lines (MCR-7 and MDA-MB-453) and ovarian cancer cell line (PE014) (Figure 3.12). PCR products of *SEPT12* were sent for DNA sequencing and the results appeared significantly similar to the gene, (summary of sequencing results, Table 3.4).

C9orf128 was originally classified as a class 3 cancer-testis/CNS restricted gene according to the expression profile obtained from the EST pipeline analysis. However, the RT-PCR profile of *C9orf128* expression in normal human tissues showed it to be testis-selective, because a faint band was observed in the prostate and trachea. The RT-PCR profile of *C9orf128* expression in cancer cell lines showed faint bands in the prostate cancer (PC-3) and uterus tumour as well as the normal testis, with the expected size of 731 bp (Figure 3.13). Therefore, *C9orf128* was classified according to the RT-PCR profile analysis as a cancer/testis-selective gene. The PCR products from the testis, prostate cancer and uterus tumour were DNA sequenced and these results showed significant identity to this gene (summary of sequencing, Table 3.4).

Based on the original EST pipeline result, *C9orf11* was classified as a class 1 testis-restricted gene, while *PRSS54* was a class 3 cancer-testis/CNS restricted gene. The *C9orf11* and *PRSS54* genes were classified as testis/CNS selective based on the results from the RT-PCR analysis in normal human tissues. Expression of *C9orf11* and *PRSS54* was analysed in the cancer cell lines and these genes were expressed in some of the cancer cells (Figure 3.14). *C9orf11* was expressed in colorectal cancer cells (T84), uterus tumour, ovarian cancer cells (PE014) and leukaemia cells (K-562) as well as in the normal testis. In contrast, *PRSS54* was expressed only in ovarian cancer cells (TO14) in addition to the normal testis. According to their expression by RT-PCR in normal and cancer tissues, these genes were classified as cancer-testis/CNS selective. DNA sequencing of the purified RT-PCR products showed significant identity to the actual genes (Table 3.4).
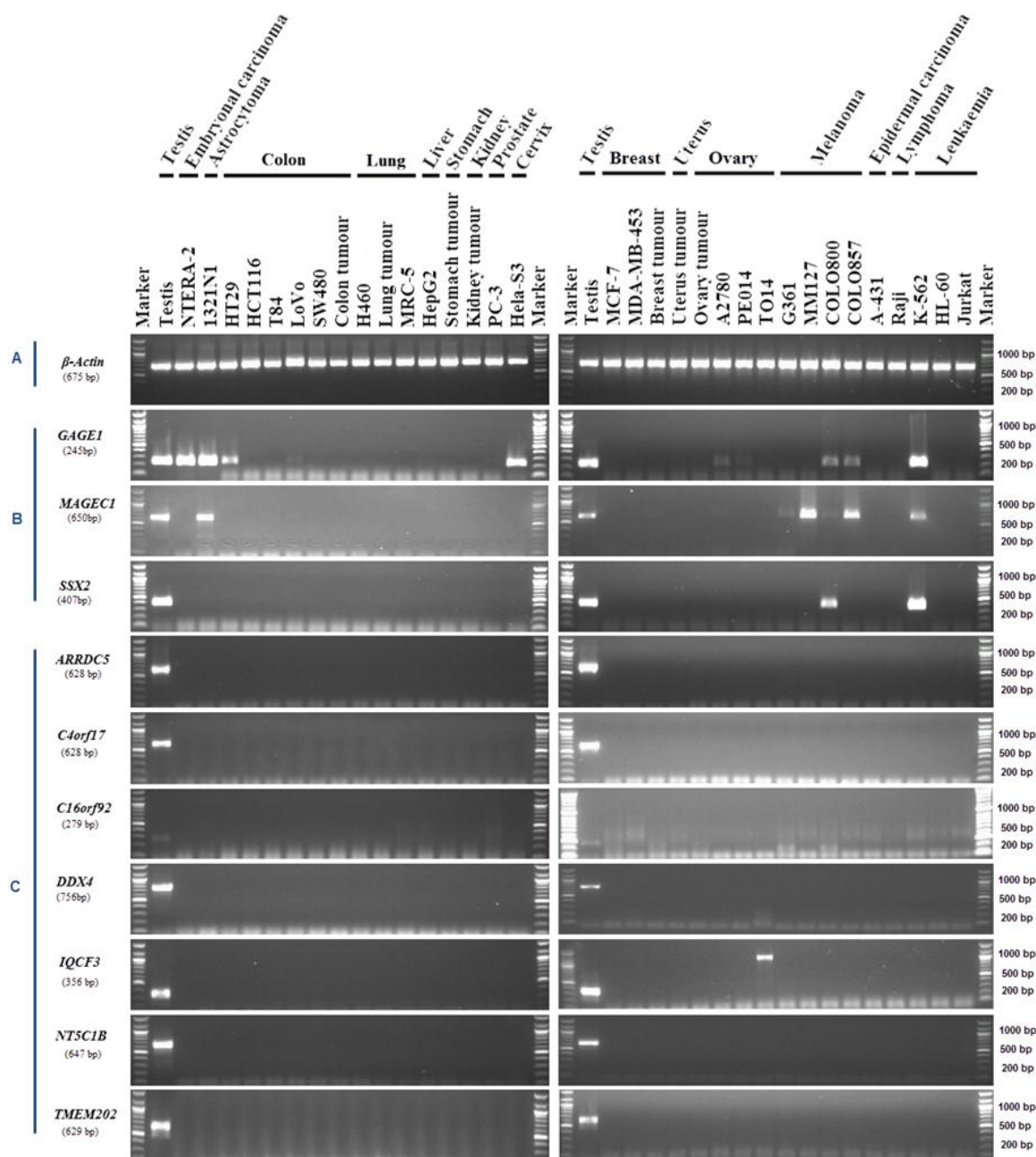
Figure 3.10. **RT-PCR Screening for the predicted testis-restricted CT genes, obtained from an EST pipeline, in cancer cell lines and solid tumours.** Agarose gels showing the RT-PCR analysis of 7 testis-restricted genes in different cancer cell lines and tissues. cDNA was generated from the total RNA from 33 cancer cell lines and tissues. (C) All candidate genes show restricted expression to the testis and no expression in cancer tissues. For *IQCF3,* DNA sequencing of the higher band, which appeared in the TO14 ovary cell line, confirmed no relation with the *IQCF3* gene. (A) Expression of the *β-Actin* gene is the positive control for the cDNA samples. (B) The expression profile of *GAGE1*, *MAGEC1* and *SSX2* genes, which were previously identified as CT antigen genes, evaluated in cancer tissues.
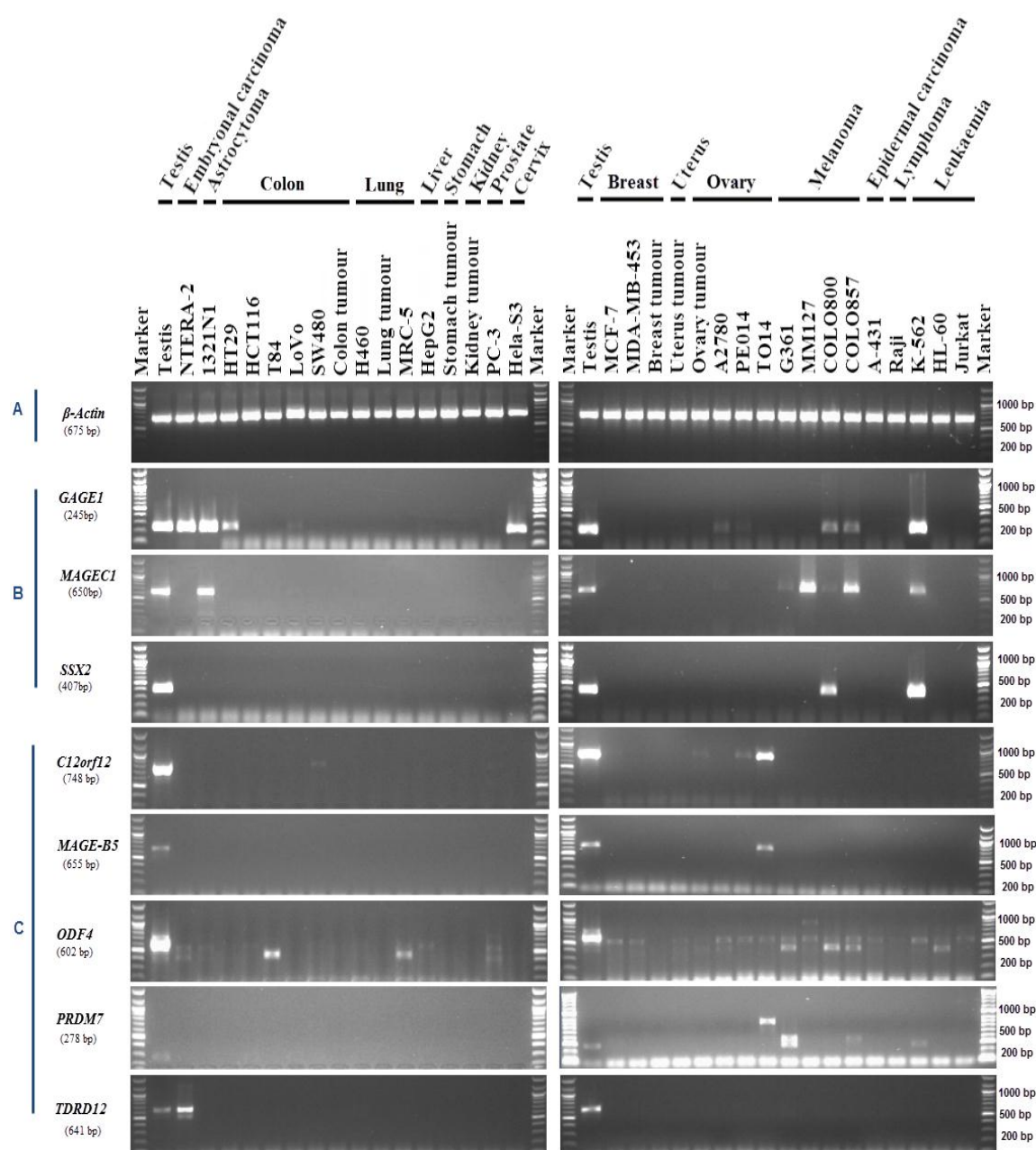
Figure 3.11. **RT-PCR analysis for the cancer/testis-restricted predicted CT genes, obtained from an EST pipeline, in cancer cell lines and solid tumours.** Agarose gels showing the RT-PCR analysis of five cancer/testis- restricted candidate genes in a range of cancer cell lines and tissues. cDNA was generated from the total RNA from 33 cancer cell lines and tissues. (C) All these candidate genes were expressed in the testis and in one or more cancer tissues. (A) Expression of *β-Actin* gene was the positive control for the cDNA samples. (B) The expression profile of *GAGE1*, *MAGEC1* and *SSX2* genes, which were previously known as CT antigen genes, evaluated in cancer tissues.
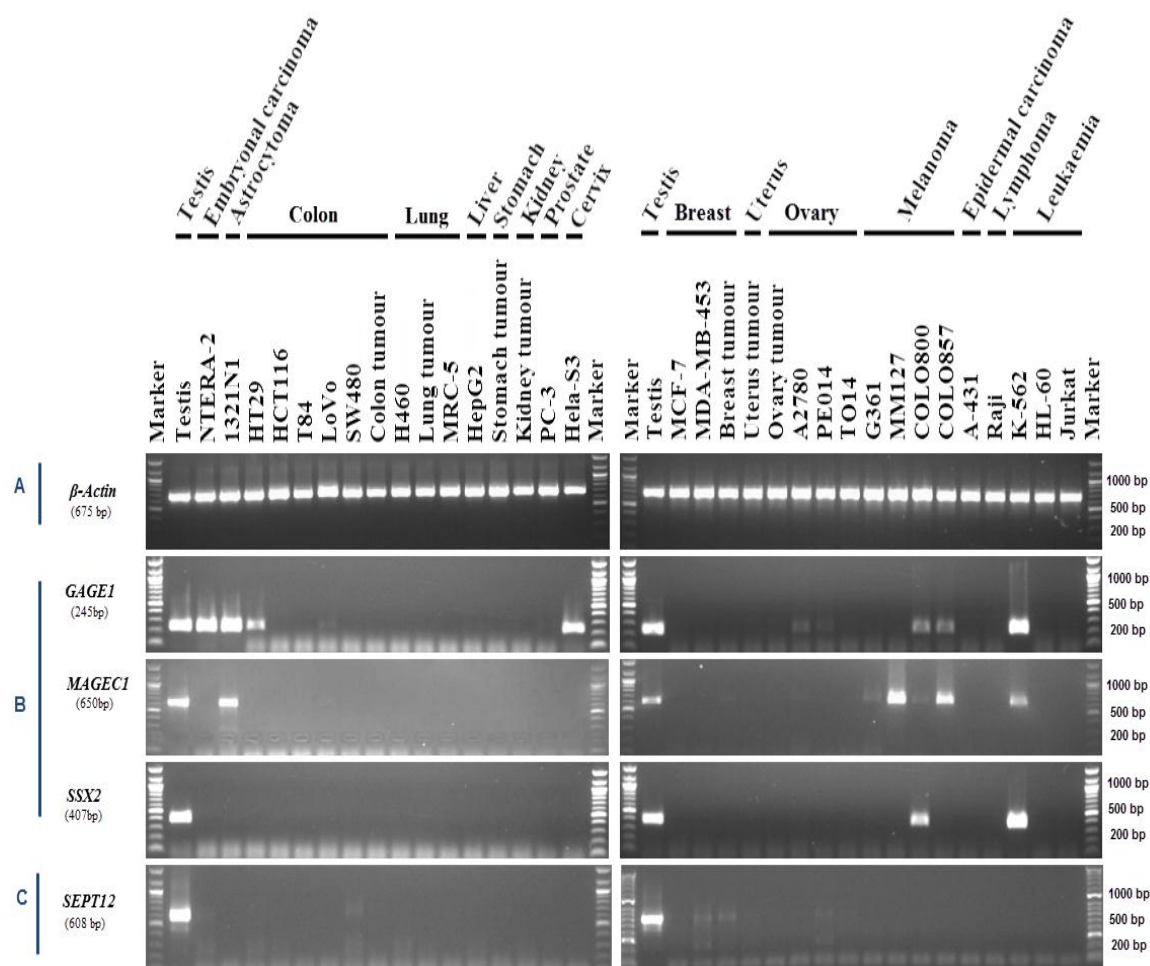
Figure 3.12.   **RT-PCR validation for the _SEPT12_ gene in cancer cell lines and solid tumours.**
Agarose gels showing the RT-PCR analysis of _SEPT12_ in a range of cancer cell lines and tissues. cDNA
was generated from the total RNA from 33 cancer cell lines and tissues. (C) _SEPT12_ was expressed in
SW480, MCR-7, MDA-MB-453 and PE014 cancer cell lines as well as in the normal testis. (A) Expression
of _β-Actin_ gene was the positive control for the cDNA samples. (B) The expression profile of _GAGE1_,
MAGEC1 and _SSX2_ genes, which were previously identified as CT antigen genes, evaluated in cancer
tissues.
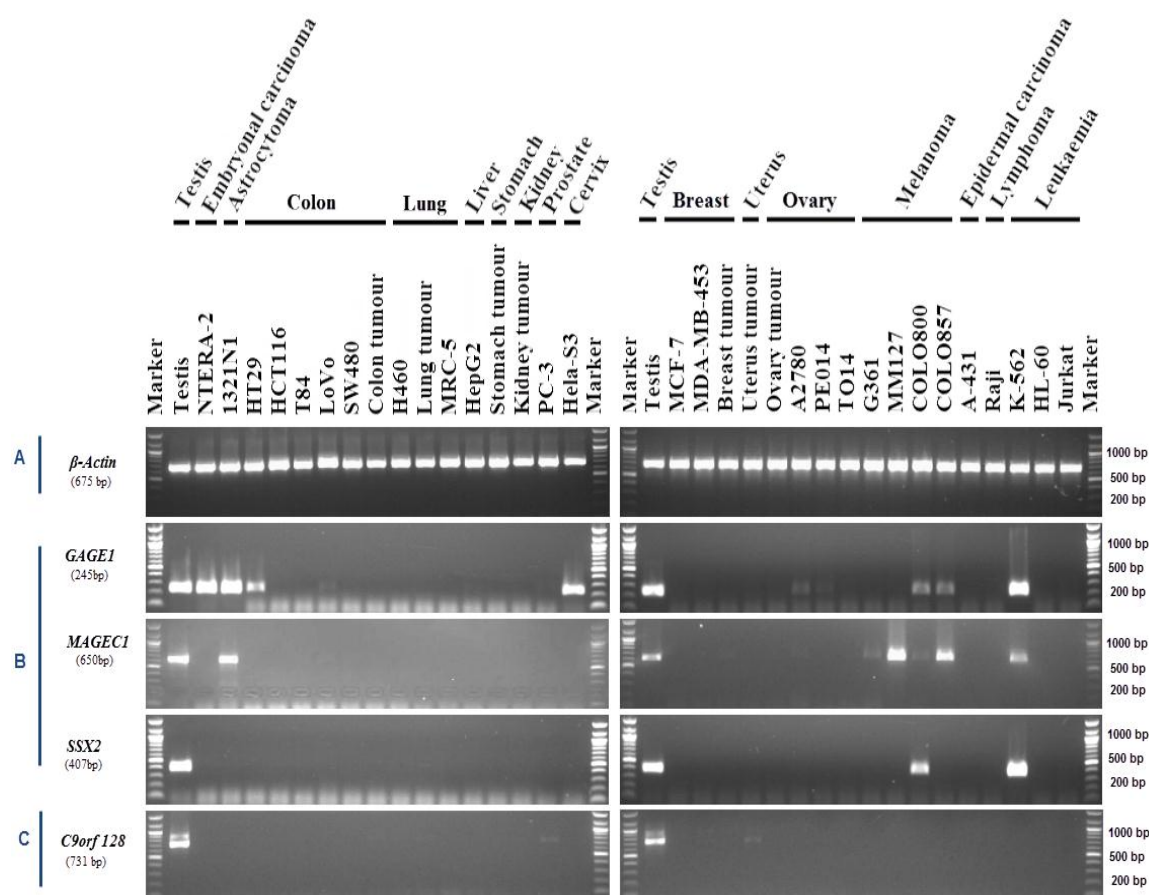
Figure 3.13. RT-PCR Screening for the *C9orf128* gene in cancer cell lines and solid tumours. Agarose gels showing the RT-PCR analysis of *C9orf128* in multiple of cancer cell lines and tumours. cDNA was generated from the total RNA from 33 cancer cell lines and tissues. (C) *C9orf128* was expressed in prostate cell line PC-3 and a uterus tumour as well as in the normal testis. (A) Expression of *β-Actin* gene was the positive control for the cDNA samples. (B) The expression profile of *GAGE1*, *MAGEC1* and *SSX2* genes, which were previously identified as CT antigen genes, evaluated in cancer tissues.
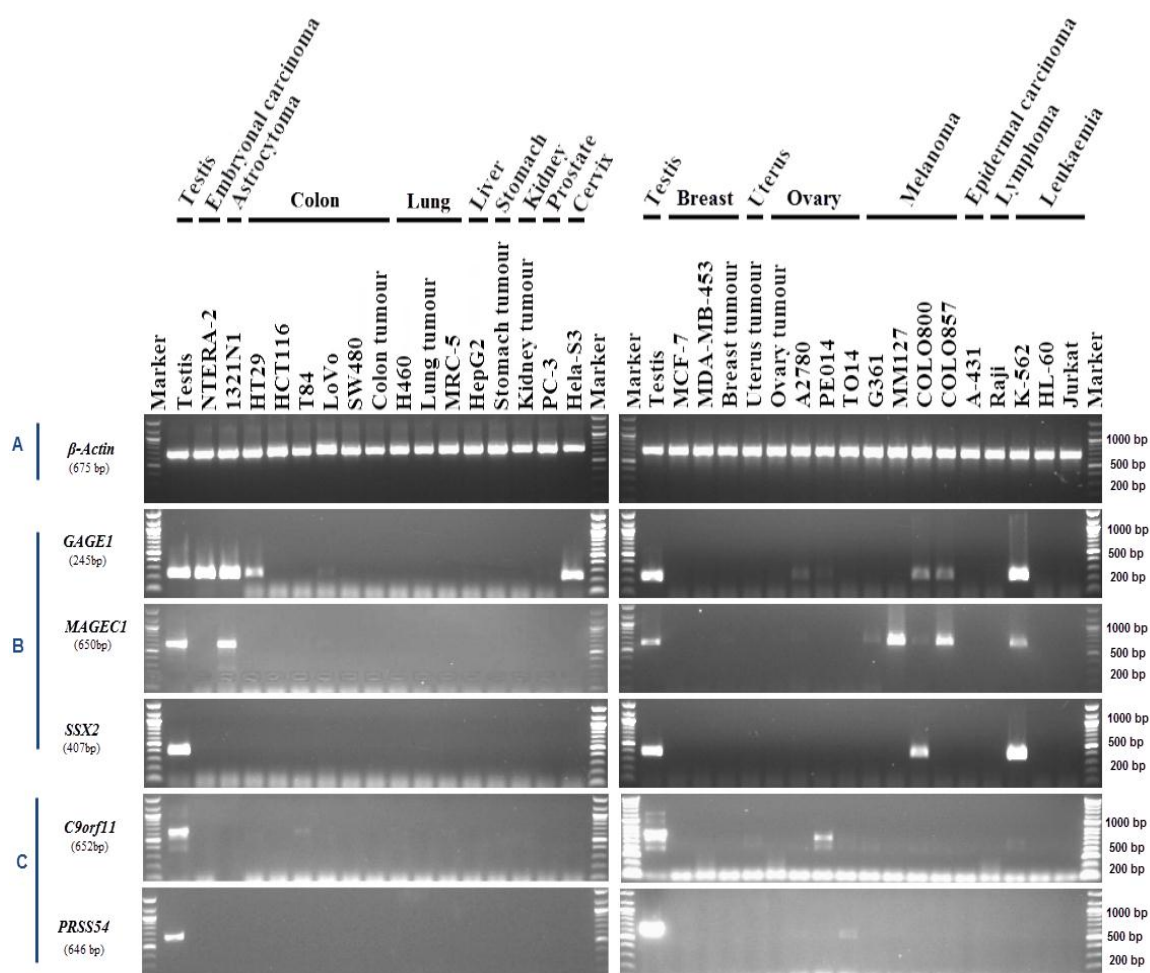
Figure 3.14.   RT-PCR expression profiles for the *C9orf11* and *PRSS54* genes in cancer cell lines and solid tumours. Agarose gels showing the RT-PCR analysis of *C9orf11* and *PRSS54* genes in multiple cancer cell lines and tumours. cDNA was generated from the total RNA from 33 cancer cell lines and tissues. (C) *C9orf11* was expressed inT84, uterus tumour, ovary PE014 and K-562 cell lines, while *PRSS54* was expressed only in TO14 cells as well as in the normal testis. (A) Expression of *β-Actin* gene was the positive control for the cDNA samples. (B) The expression profile of *GAGE1*, *MAGEC1* and *SSX2* genes, which were previously identified as CT antigen genes, evaluated in cancer tissues.

3.2.7 DNA sequencing and classification after RT-PCR validation in normal and cancer tissues.

Table 3.4. Summary of DNA sequencing results and the gene classification after RT-PCR validation in normal and/or cancer tissues.

Gene name	Primer	Sequenced in normal tissues	Sequenced in cancer samples	Sequence identity obtained from NCBI blast (%)	Classification after RT-PCR validation
β-Actin	F	Testis	-	100%	-
ALX1	F	Testis	Trachea	99%	Dismissed
APOBEC4	F	Brain, cerebellum	-	99%	Dismissed
ARRDC5	F	Testis	-	99%	Testis-restricted
C12orf12	F	Testis	Ovarian tumour	99%	Cancer/ testis-restricted
C16orf92	F	Testis Skeletal muscle	-	100% 98% TTN	Testis-restricted
C4orf17	F	Testis	-	99%	Testis-restricted
C9orf11	F	Testis Brain cerebellum	- PE014	100% 98%	Testis/CNS-selective
C9orf12	F	Testis	-	100%	Dismissed
C9orf128	F	Testis Brain cerebellum	- Uterus tumour	100% 99%	Cancer/testis-selective
C9orf43	F	Testis	-	98%	Dismissed
DDX4	F	Testis	-	100%	Testis-restricted
GAGE1	F	Testis	-	99%	Control
IQCF3	F	Testis -	- TO14	100% 99%	Testis-restricted
KDM4D	F	Testis	-	100%	Dismissed
MAGE-B5	F	Testis	TO14	100%	Cancer/ testis-restricted
MAGEC1	F	Testis	-	99%	Control
NT5C1B	F	Testis	-	100%	Testis-restricted

ODF4	F	Testis Foetal brain	A2780 & K562 - T84	99% 99% EPB41L3 99% KAP12	Cancer/ testis-restricted
PRDM7	F	Testis -	K562 G361	99% RDM7 98% RDM9 99% PRDM7 97% PRDM9	Cancer/ testis-restricted
PRSS54	F	Testis Spinal cord-	 TO14	99% 100%	Testis/CNS-selective
SEPT12	F	Testis Brain, cerebellum-	- - MRC-7	100% 98% 99%	Cancer/Testis-CNS-restricted
SSX2	F	Testis	-	99%	Control
TDRD12	F	Testis	NT2	100%	Cancer/ testis-restricted
TMEM202	F	Testis	-	100%	Testis-restricted
TNRC18	F	Prostate	-	99%	Dismissed
TSSK1B	F	Foetal brain -	MRC-5 TO14	99% 100%	Cancer/Testis-CNS-restricted
WDR20	F	Testis	-	100%	Dismissed

3.3 Discussion

Cancers are complex genetic diseases that lead to uncontrolled cell division, decrease in cell death, invasion, destruction of adjacent tissues and spread to other organs of the body (Michor *et al.*, 2005; Stratton *et al.*, 2009; Tomasetti *et al.*, 2013). Cancer is normally diagnosed and treated in the late stages when the cancer cells invade the body and turn into metastatic cells (Suri, 2006; Aly, 2012). Alteration of gene expression and mutations can lead to the production of abnormal proteins termed tumour-associated antigens (TAAs). The identification and characterisation of novel TAAs as potential cancer biomarkers and/or targets is a key challenge in the successful immunotherapeutic strategies aimed at tumour rejection (Krishnadas *et al.*, 2013). CTA genes are attractive targets for cancer therapies due to their restricted expression pattern to the testis and malignant tumours.

In this chapter, two methods were used for the identification of novel CTA genes. The first was a manual search of the literature for meiosis-specific genes, which can be a rich source of CTA genes. Due to the limitations of manually searching for restricted CT genes, a second, systematic approach was developed to identify potential CTA genes based on microarray and EST databases.

3.3.1 Gene expression of the *TSSK1B* gene

Several known meiosis-specific genes have been found to be expressed in different types of cancer including DSB initiator gene *SPO11* (Koslowski *et al.*, 2002) and the synaptonemal complex protein genes *SYCP1* (Tureci *et al.*, 1998), *SYCP3* (Modarressi *et al.*, 2004) and *HORMAD1* (Chen *et al.*, 2005a). The abnormal expression of meiosis-specific genes in cancer cells makes them a potential source of biomarkers for cancer diagnostic, vaccine and immunotherapy treatments (Old, 2001; Simpson *et al.*, 2005; Caballero and Chen, 2009; Feichtinger *et al.*, 2012a). *TSSK1B* is a meiosis-specific gene that was selected based on the manual search in the literature and displayed a promising CT gene characteristics. It was expressed only in the testis and foetal brain and in different cancer cell lines (see grid summary Figure 3.15 and Figure 3.16).

3.3.2 Gene expression of microarray analysis predicted genes

ALX1 and *WDR20* genes were selected from 40 genes derived from microarray analysis. *ALX1* is a transcription factor gene which may be involved in cervix development and chondrocyte differentiation (Zhao *et al*., 1993; Cai, 1998). In contrast, *WDR20* may encode a protein that regulates the activity of the USP12-UAF1 deubiquitinating enzyme complex (Kee *et al*., 2010). Due to their expression in the normal tissues, no further investigation of these genes was performed in cancer cells (see grid summary Figure 3.15). Different intron-spanning primers were designed for these genes, but gave the same results. A high number of RT-PCR cycles were used and this might be the reason for detection of the low level expression of *ALX1* in the other normal tissues.

Microarray technology is a powerful tool for studying the expression of thousands of genes from different populations of cells at the same time in a single experiment (Russo *et al*., 2003). Microarray analysis can be affected by the noise in the background and poor sensitivity as well as the RNA quality of the tested samples, which can lead to inaccurate results (Russo *et al*., 2003). Only two genes *ALX1* and *WDR20* derived from the microarray analysis in this study and were dismissed due to their expression in normal tissues. Due to the limited number of genes obtained from the microarray analysis, a second bioinformatic approach (EST analysis) was used to identify more candidate genes.

3.3.3 Gene expression of EST analysis predicted genes

21 out of 177 genes were chosen from the EST pipeline for further validation in normal and cancer tissues. Based on the RT-PCR profile in the normal somatic tissues, 16 of the 21 genes were identified as promising CTA genes and thus were further screened in the cancer cell lines and solid tumours. Whereas, 5 genes were dismissed due to their expression in more than two non-testis normal tissues. Candidate genes were categorised to five categories depending on their expression profile in normal somatic and/or cancer tissues using RT-PCR analysis. These categories are (i) testis-restricted; (ii) cancer/testis restricted; (iii) cancer/testis CNS-restricted; (iv) cancer/testis-selective; and (v) cancer/testis-CNS selective genes.

3.3.4 Gene expression of dismissed genes from the EST analysis

The RT-PCR profiles of *APOBEC4*, *C9orf12*, *C9orf43*, *KDM4D* and *TNRC18* were dismissed because their expression in more than two normal tissues (summary grid Figure 3.15). Based on the original EST pipeline category, *C9orf43* and *TNRC18* were classified as class 1 testis-restricted genes. In contrast, the other three dismissed genes, *APOBEC4*, *C9orf12* and *KDM4D,* were classified as class 3 cancer-testis/restricted genes.

3.3.5 Gene expression of testis-restricted genes

RT-PCR analysis of *ARRDC5*, *C4orf17*, *C16orf92*, *DDX4*, *IQCF3*, *NT5C1B* and *TMEM202* showed no expression was observed for these genes in both normal and cancer tissues (see grid summary Figure 3.15 and 3.16). Based on the initial EST analysis, most of these genes were categorised as a class 1 testis-restricted with the exception of *ARRDC5,* which was classified as class 2 cancer-testis restricted. Most of these genes encoded proteins of uncharacterised function and they might be novel potential CTA genes.

3.3.5.1 Meta-analysis of testis-restricted genes

The seven testis-restricted genes were investigated in samples derived from patients using a meta-analysis microarray web tool (CancerMA, http://www.cancerma.org.uk/; (Feichtinger *et al*., 2012b). Basically, the meta-analysis pipeline was derived from microarray data sets from 13 untreated types of cancer and compared to normal tissues with a total of 80 cancer data sets. These genes were analysed as well as 3 positive X-CTAs (*GAGE1*, *MAGEC1* and *SSX2*) to identify the meta-changes in these genes in relation to corresponding cancer types based on the microarray dataset. Two genes were not present in the array: *ARRDC5* and *TMEM202*. In addition, the meta-analysis result showed statistically significant up-regulation of 5 genes (Figure 3.17). *C4orf17* was up-regulated in brain cancer whereas *IQCF3* was up-regulated in adrenal cancer. The meta-analysis result also showed a significant up-regulation of the *NT5C1B* gene in leukaemia. The *C16orf92* and *DDX4* genes were identified to be up-regulated in ovarian cancer.

 The three positive control genes, *GAGE1*, *MAGEC1* and *SSX2,* were validated with the meta-analysis to assess our results. The control genes showed a significant up-regulation in ovarian cancer and lung cancer. Also, *GAGE1* and *SSX2* were significant up-regulated in brain cancer.
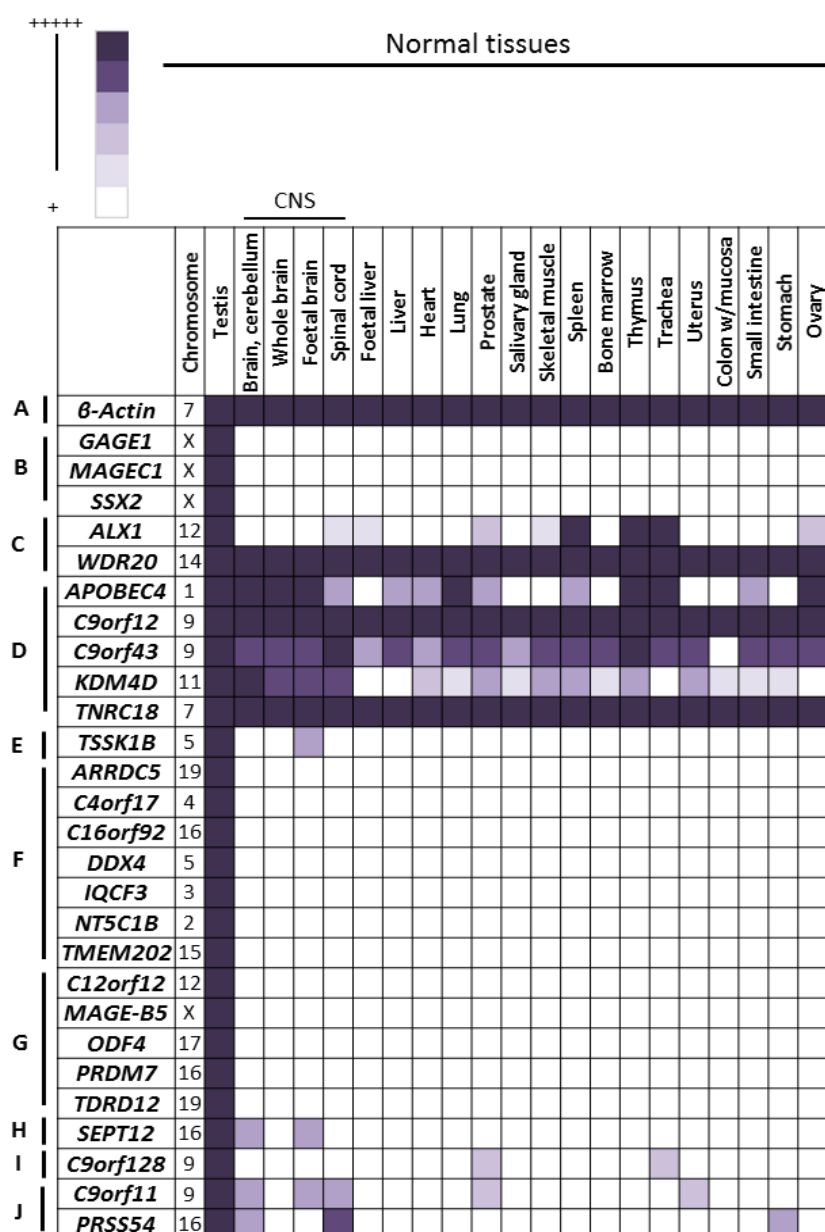
Figure 3.15. **Grid representation of the gene expression profiles of 24 genes validated in normal human tissues.** Gene expressions are represented in the grid in the rows, while each column represents a range of normal human tissues. RT-PCR expression profiles of each gene are visualised in 1% agarose gels stained with ethidium bromide. The intensity of the bands is graded with different shades of purple (see the key at the top left) (+) is weak expression band and (+++++) is strong expression band. In the grid, (A) expression of the β-*Actin* gene shows the positive control for the cDNA samples. (B) The expression profile of *GAGE1*, *MAGEC1* and *SSX2* genes, which were previously identified as CT antigen genes; (C) dismissed genes derived from microarray analysis; (D) dismissed genes derived from EST pipeline. (E) Expression of meiotic-specific *TSSK1B* gene; (F) testis-restricted genes; (G) cancer/testis restricted genes; (H) testis/CNS restricted *SETP12* gene; (I) testis-selective *C9orf128*; (J) Testis/CNS-selective genes; *C9orf11* and *PRSS54*.
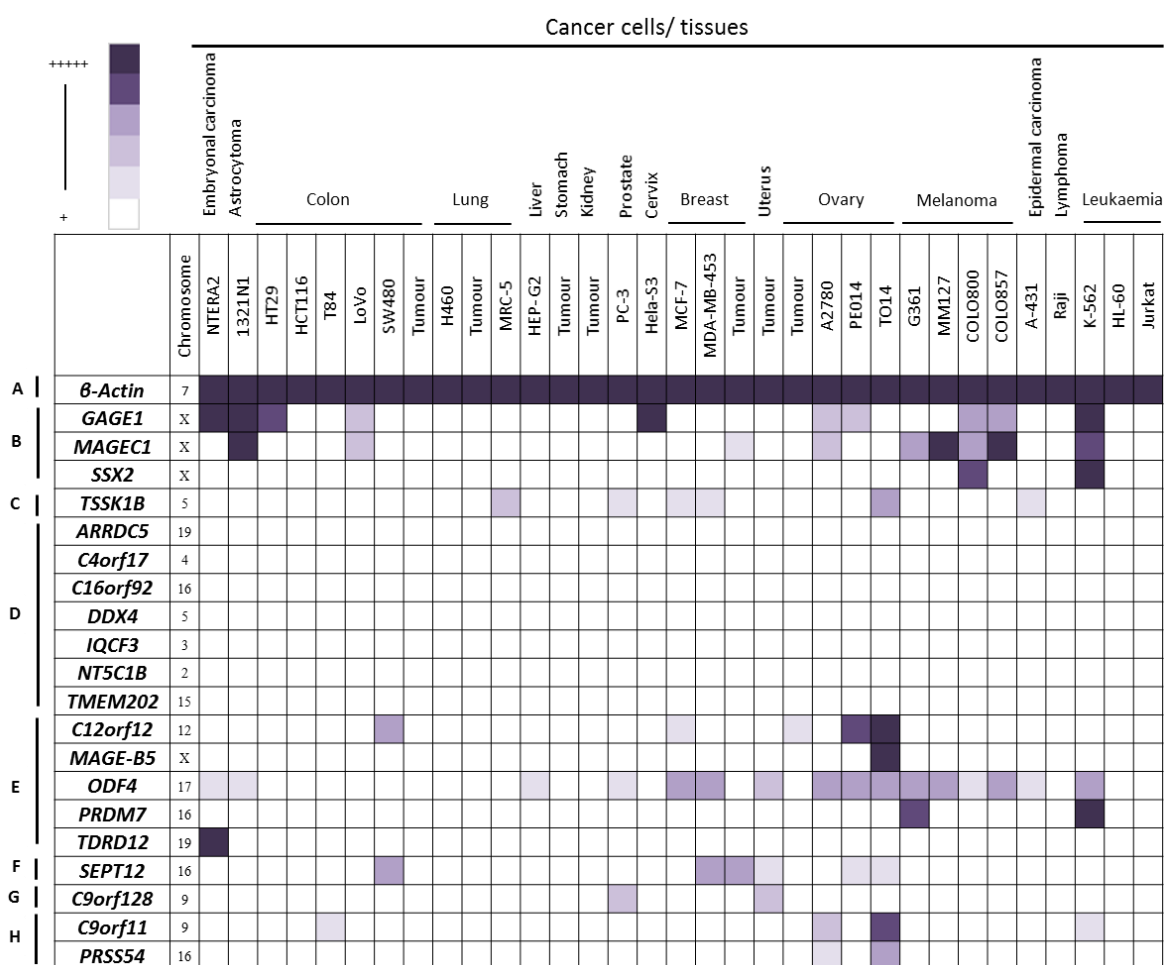
Figure 3.16. **Grid representation of the gene expression profiles of 17 candidate genes in cancer cell lines and tissues.** Gene expressions are represented in the grid in the rows, while each column represents a range of normal human tissues. RT-PCR expression profiles of each gene are visualised in 1% agarose gels stained with ethidium bromide. The intensity of the bands is graded with different shades of purple (see the key at the top left) (+) is weak expression band and (+++++) is strong expression band. In the grid, (A) expression of *β-Actin* gene shows the positive control of the cDNA samples. (B) The expression profile of *GAGE1, MAGEC1* and *SSX2* genes, which were previously identified as CT antigen genes, (C) Expression of meiotic-specific *TSSK1B* gene; (D) testis-restricted genes; (E) cancer/testis restricted genes; (F) testis/CNS restricted *SEPT12* gene; (G) testis-selective *C9orf128*; (H) testis/CNS-selective genes; *C9orf11* and *PRSS54*.
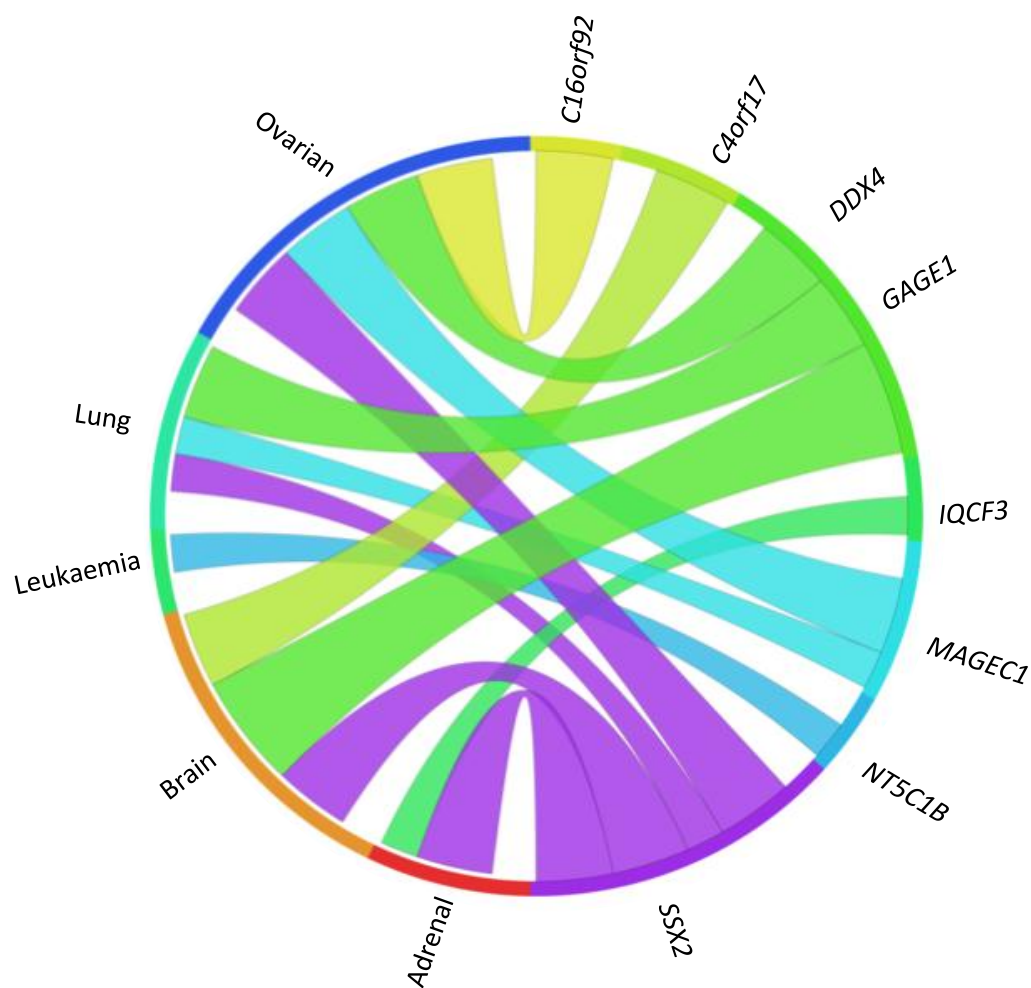
Figure 3.17. **The circus plot for microarray analysis of testis-restricted genes showing the meta-change in gene expressions in relation to corresponding cancer types.** Seven testis-restricted candidate genes and three positive control genes were analysed in meta-analysis using the online web tool CancerMA (http://www.cancerma.org.uk/; Feichtinger *et al.*, 2012b). Each connection between the gene and a corresponding type of cancer indicates a statistically significant up-regulation in that type of cancer; these gene are *C4orf17*, *C16orf92*, *DDX4*, *IQCF3*, *NT5C1B* and three positive controls, *GAGE1*, *MAGEC1* as well as *SSX2*. Two genes not appearing in the array are *ARRDC5* and *TMEM202*. The cancerMA tool analyses the expression of these genes against 13 types of cancer *vs* normal tissue derived from different array studies.

3.3.6 Gene expression of novel cancer/testis-restricted genes

The RT-PCR expression profiles of five genes indicated that they were cancer/testis-restricted genes. These genes were expressed in normal testis and no expression was determined in non-testis normal tissues (summary Figure 3.15). These genes were validated in the 33 cancer cell lines and solid tumours and appeared to be expressed in more than one cancer cell line/tissue (summary Figure 3.16). These genes included *C12orf12*, *MAGE-B5*, *ODF4*, *PRDM7* and *TDRD12*. Some of these genes are encoded with uncharacterised protein functions and the others have known protein functions. All these genes have autosomal locations with the exception of *MAGE-B5* which is X-chromosome encoded. This group of genes might be considered as promising biomarkers and /or targeted genes for immunotherapeutic strategies based on their mRNA expression.

C12orf12 (also termed *CCER1*) is an uncharacterised protein found to be expressed in the colorectal cancer cell line (SW480) and 3/4 of ovarian cancer cell lines/tumours. The expression of *C12orf12* was observed in the normal testis but not in other somatic tissues. The coding sequence of *C12orf12* did not consist of any introns, so an intron-spanning primer could not be designed.

MAGE-B5 belongs to a large group of known CTA genes, the *MAGE-A*, *B* and *C* families, which are encoded on the X- chromosome; this gene was the only X-CT candidate gene in this study (Lucas *et al*., 2000). The expression of *MAGE-B5* was identified in the ovarian cancer cell line (TO14) as well as in the normal testis. A previous study by Lucas et al*. (2000) on 18 different types of cancer determined that *MAGE-B5* was expressed in the seminoma but no expression was detected in other tumours. However, in that previous study did not analyse *MAGE-B5* in ovarian cancer. In our case, this gene was expressed in 16 different types of cancer including ovarian cancer. The clinical relevance of *MAGE-B5* in meta-analysis microarray was not apparent. However, some of the *MAGE* family members have been found to have oncogenic activity, such as *MAGE-A2* which was shown to be involved with other genes in down-regulation of the tumour suppressor gene *p53* (Monte *et al*., 2006). *C12orf12* and *MAGE-B5* were expressed mostly in ovarian cancer, where these genes might play a normal role in meiosis, such as in foetal ovary development and they may preferentially reactivate in

ovarian cancer. These genes might be a potential genes used as diagnostic tool and/or treatment of ovarian cancers.

ODF4 was expressed in a range of cancer cell lines and solid tumours as well as in the normal testis. This gene was a previously identified as a CTA gene in leukaemia and prostate cancers (Ghafouri-Fard *et al*., 2012; Ghafouri-Fard and Modarressi, 2012). Our results showed that this gene was expressed in a range of cancer cells, not only in leukaemia and prostate cancer. The meta-analysis result for *ODF4* showed a significant up-regulation in adrenal and ovarian cancers (Figure 3.18).

The RT-PCR profile of *PRDM7* showed expression in the leukaemia cell line (K562), melanoma cell line (G361) and in the normal testis tissues. The meta-analysis result of *PRDM7* revealed statistically significant up-regulation in ovarian cancer (Figure 3.18). *PRDM7* is a member of *PRDM* gene family which consists of 17 human genes. This family is characterised by the presence of the PR-domain containing protein and the SET domain of histone methytransferases. Some of the PRDM genes may be involved in transcriptional regulation, cell differentiation and consequently neoplastic transformation (Fog *et al*., 2012). *PRDM7* is the only member of PRDM family that was generated by a duplication with the hot spot activator *PRDM9* (Fumasoni *et al*., 2007).

The *TDRD12* (also termed *ECAT8*) gene was expressed only in the embryonic carcinoma cell line (NT2) as well as in the normal testis tissue. The RT-PCR expression profiles for the 21 genes in the normal and cancer tissues suggested that *TDRD12* is the only potential biomarker that is expressed in germ line tumours, with the exception of low expression of the *ODF4* gene. *TDRD12* has a unique function in RNA biogenesis through its interaction with the piwi protein in mice (Pandey *et al*., 2013). Normally, germ cell lines emerging through embryogenesis differentiate to give mature gametes (spermatocytes and oocytes). Alteration of the germ cell development may cause testicular tumours. One interesting hypothesis proposed that testicular tumours may emerge from the germ cells that fail to be mitotically inactive during the resting phase of G_0 arrest (reviewed in Heaney *et al*., 2012). The expression of *TDRD12* in both testis and a germ cell line (NT2) might be a result of

mitotic inactivation and could represent DNA hypomethylation of this gene in NT2 as most of

the CTA can be regulated through epigenetic events (reviewed in Fratta *et al.*, 2011).
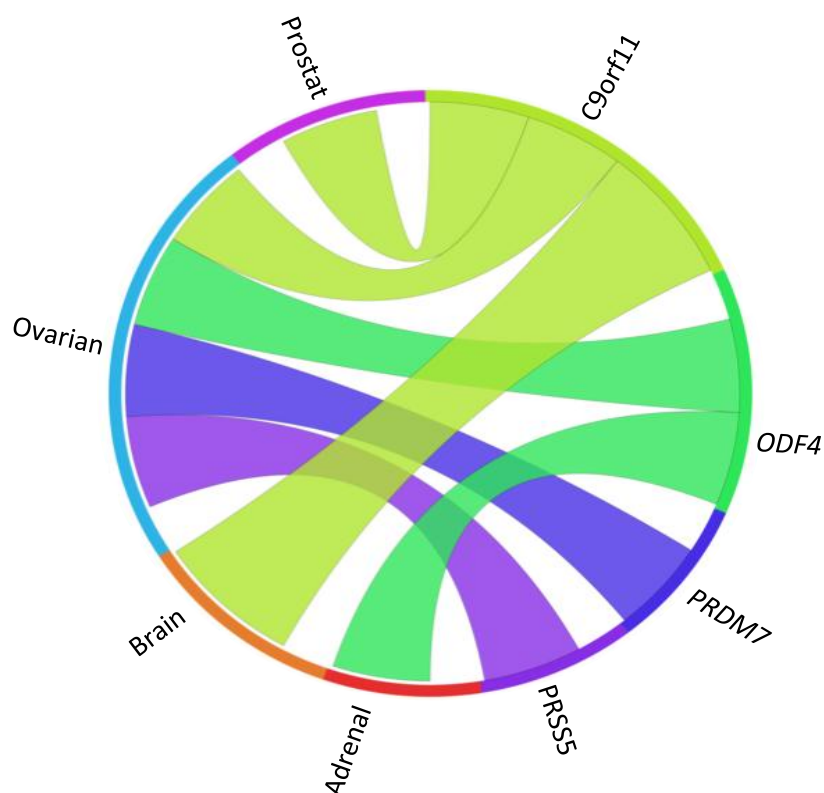
Figure 3.18. **The circus plot for microarray analysis of genes expressed in more than one cancer cell line/tissue using RT-PCR.** The meta-change expression of candidate genes in relation to corresponding of cancer types. Ten candidate genes showing expression in more than one cancer cell line/tissue were analysed in clinical meta-analysis using the online web tool CancerMA (http://www.cancerma.org.uk/; Feichtinger *et al.*, 2012b). Each connetion between the gene and a corresponding type of cancer indicates a statistically significant up-regulation in that type of cancer. Only 4 genes out 10 showed statistically significant up-regulation of these genes: *C9orf11*, *ODF4*, *PRDM7* and *PRSS54.* The CancerMA tool analysed the expression of these genes against 13 types of cancer *vs* normal tissue derived from different array studies.

3.3.7 Gene expression of cancer/testis-CNS restricted genes

The RT-PCR profile of *SEPT12* in normal somatic tissues revealed expression in the normal testis as well as in the CNS, including the brain cerebellum and foetal brain tissues. This gene also was expressed, with a faint band produced from some of the colorectal cancer, breast cancer and ovarian cancer cell lines (see summary grid Figure 3.15 and 3.16).

The meta-analysis result of *SEPT12* shown no significant up-regulation. *SEPT12* can be used as a potential cancer testis/CNS-restricted gene because the CNS (brain cerebellum and foetal brain tissues) as well as the testis are immune privileged sites. The blood-brain barrier (BBB) isolates the CNS from the immune system and does not allow leukocyte entry (Bechmann and Woodroofe, 2014). However, the expression of *SEPT12* in CNS tissues might reflect the nature of the normal somatic tissues used in this study.

Most of extracted RNA was obtained from "normal tissues" *post mortem* and were collected from different individual patients of different ages and with different health problem. These tissues were often pooled together and some of these tissues might have harboured undiagnosed cancer, which would explain the aberrant expression of some CTA genes (Feichtinger *et al.*, 2012a).

3.3.8 Gene expression of cancer/testis selective genes

Three of the 21 genes identified by the EST pipeline analysis were CT-selective genes. *C9orf128* was expressed in two normal somatic tissues, the prostate and trachea, as well as in the normal testis. This gene was also expressed in a prostate cancer cell line (PC-3) and uterus tumour (see summary grid Figure 3.15 and 3.16).

C9orf128 (also called *FAM221B*) is a protein of uncharacterised function and the expression of this gene indicates that it may play a role in prostate development and tumourigenesis. However, evidence from a clinical study indicated that a large number of people can carry microscopic cancer cells that do not reach a level of becoming clinically detectable disease (Naumov *et al.*, 2009). For this reason, the "normal" somatic prostate and trachea RNA used in the present study to validate *C9orf128* could conceivably have carried undetected microscopic cancer.

In addition to *C9orf128*, two genes were found to be CT/CNS-selective genes. *C9orf11* expression was found in normal CNS, testis, lung and uterus tissues, whereas *PRSS54* was expressed in normal CNS, testis and prostate tissues. The expression of *C9orf11* was also detected in cancer cell lines (T84, PE014 and K-562) as well as uterus tumour. In contrast, *PRSS54* was only expressed in ovarian cancer cells (TO14) (see summary grid Figure 3.15 and 3.16).

The meta-analysis result of these genes showed a significant up-regulation of *C9orf11* in ovarian, brain and prostate cancers, while *PRSS54* showed a statistically significant up-regulation in ovarian cancer (Figure 3.18). The clinical meta-analysis results are similar to those observed with RT-PCR.

The aberrant expression of these genes may be a result of the nature of "normal" RNA purchased from Clontech and Ambion. Typically, normal tissues from a number of individuals obtained *post mortem* and are pooled together. It is might that some of these "normal" tissues had undiagnosed microscopy neoplastic cells and possibly have been aberrantly expressing these genes.

For example, Chen and colleagues (2005) was noticed some different CT genes expression in the similar normal tissue types but from different source. They found some genes expressed in normal tissues but their expression could not be detected in similar type of the normal tissues but obtained from different source (Chen et al., 2005b; Feichtinger *et al*., 2012a). Although these companies follow a high standard of quality, these genes need to be investigated in a new panel of normal RNA from other sources to confirm their classification as CT-selective genes.

3.4 Conclusions

This chapter focused on identification and validation of potential novel genes. Due to their unique expression and immunogenicity, CTA antigens can be used as potential cancer biomarkers and in targeting of cancer immunotherapy. In this project, two approaches were taken for the identification of novel cancer testis CTAs genes: (i) searching manually in the literature for meiosis-specific genes and (ii) a systematic approach using bioinformatics

technology. Only one meiosis-specific gene (*TSSK1B*) was selected from the literature and it appeared to be a potential CTA gene.

In this study, a systematic approach was developed to identify new CTA genes using publicly available microarrays and EST databases (developed by J.Feichtinger; Feichtinger *et al.*, 2012a; 2012b; 2014a). 100% of the genes selected from the microarray analysis data were dismissed, whereas 76% of genes derived from the EST analysis pipeline were found to be potential testis-specific genes. The analysis of EST pipelines was more specific than microarray pipeline analysis and was used as an additional filter to exclude non-testis-specific genes.

RT-PCR analyses were used to validate the expression of the predicted genes in 21 normal somatic tissues. If the gene expression was determined in more than two normal somatic (non-testis) tissues, that gene was dismissed [with the exception of the CNS tissues which are immune privileged due to the blood-brain-barrier (BBB)].

In total, 17 genes were identified as promising testis-specific genes and were validated in 33 cancer cell lines and solid tumours. This identified 5 potential cancer/testis-restricted antigen genes that were expressed in the normal testis as well as in several cancer cell lines/tissues. These genes were *C12orf12*, *MAGE-B5*, *ODF4*, *PRDM7* and *TDRD12*. Three genes were identified as either cancer/testis-selective such as (*C9orf128*) or cancer/testis-CNS selective (*C9orf11* and *PRSS54*); whereas one gene was testis/CNS-restricted (*SEPT12*).

Based on the RT-PCR profiles and the meta-analysis of these genes, six genes were expressed in ovarian cancers and might be used as biomarkers/targets for that type of cancer. These genes may play a critical role in normal meiosis, such as in foetal ovary development, and may also preferentially reactivate in ovarian cancer.

Another class of genes was expressed only in the testis but the expressions of these genes were not detected in the other cancers or normal tissues that were investigated. These genes were *ARRDC5*, *C4orf17*, *C16orf92*, *DDX4*, *IQCF3*, *NT5C1B* and *TMEM202*.

The meta-analysis showed that *C4orf17* was up-regulated in brain cancer whereas *IQCF3* was up-regulated in adrenal cancer. *NT5C1B* was also up-regulated in leukaemia and *C16orf92* was up-regulated in ovarian cancer as well as being down-regulated in head and neck cancer.

Thus, these genes require more validation in different cancer cell lines and solid tumours to address whether they represent potential tissue restricted CTA genes.

These genes were sub-classified as meiCT genes, which are extensively expressed in ovarian cancers. All of predicted CTA genes in this study were non-X-coded and only one gene was located on the X-chromosome. These genes now need to be studied at the protein level to identify their functions, which may include oncogenic activity. The CTA genes are also known to be regulated by epigenetic changes; and in the next chapter, several identified CTA genes were studied for epigenetic effects.

Chapter 4.0 RT-PCR analysis of epigenetic regulation of potential novel cancer testis antigen genes in colon cancer cell lines

4.1 Introduction

Epigenetic mechanisms play an important role in the regulation of gene expression and normal mammalian cell development. Differentiating male germ cells undergo a distinct of series epigenetics transition that are characterised by global chromatin changes; these factors include DNA and histone modifications as well as chromatin remodelling (Nottke *et al*., 2009; Meikar *et al*., 2013). Epimutations can lead to increased genomic instability by the silencing of tumour suppressor genes and/or activation of oncogenes either independently or in association with genetic mutations or deletions. However, the biological mechanism which leads to the abnormality of epigenetics remains poorly understood (Sharma *et al*., 2010).

Cancer germline (CG) or cancer-testis (CT) antigen genes are normally expressed in the germ cells and aberrantly expressed in different types of cancers, but these genes are not expressed in normal somatic tissues (Simpson *et al*., 2005; Link *et al*., 2009). Several CTA genes are usually found co-expressed in positive tumours, which suggests that their activation in cancers is due to a global gene activation mechanism rather than stochastic individual events. Regulation of the CTA genes is proposed to occur at the transcriptional level by epigenetic signals that affect not only repeated sequences but also single copy genes, which may lead to alteration of gene expression in cancers (De Smet *et al*., 1999; De Smet and Loriot, 2013).

4.1.1 The role of DNA methylation in CTA gene regulation

DNA methylation is an enzymatic reversible reaction characterised by addition of a methyl group (CH_3) onto the 5' position of cytosine bases in CpG dinucleotides to form 5-methylcytosine. The reaction is catalysed by a key family of enzymes known as DNA

methyltransferases (DNMTs) which transfer a methyl group from S-adenosylmethionine (SAM) to cytosine. Hypermethylation (an increased level of methylation) of the CpG island of a gene promoter is often associated with gene silencing, as it prevents transcription factors (TFs) from binding to gene promoters. For example, tumour suppressor genes (TSGs) can be silenced through promoter hypermethylation. By contrast, promoter hypomethylation (a decreased level of methylation) is associated with expression of silenced genes in cancer cells (For example, see Coppedè, 2014).

Different definitions have been proposed for the CpG island; however, Takai and Jones (2002) defined the CpG islands as a genomic region of DNA greater than 500 bp in size, containing greater than 55% of G:C content and an observed/expected ratio of CpG greater than 0.65 (Takai and Jones, 2002).

CTA genes are highly methylated and transcriptionally repressed in normal somatic cells but their promoter regions are associated with DNA hypomethylation (demethylation) in the testis and sperm as well as in human tumours (Marchal *et al.*, 2004; Ehrlich and Lacey, 2013). Bisulphite DNA sequencing of the promoter regions of several CTA genes confirmed that CTA gene promoters are hypomethylated among different types of tumours when compared with the normal matched samples (Glazer *et al.*, 2009; Smith *et al.*, 2009). Previous studies showed that treatment of cancer cell lines that do not express CTA genes with epigenetic modulating inhibitors, including drugs targeting DNA methyltransferase (DNMT) and/or histone deacetylase (HDAC), activate the expression of several of CTA genes (reviewed in Link *et al.*, 2009; Fratta *et al.*, 2011).

In addition, CTA genes (*SSX*, *NY-ESO-1* and *N-RAGE*) are expressed in human adult mesenchymal and foetal stem cells, which show low levels of DNA methylation (Cronwright *et al.*, 2005). This sensitivity of CTA genes to epigenetic modulation inhibitors, particularly methylation drugs, might be a result of the CpG-rich islands that are frequently found in their promoter regions (De Smet *et al.*, 1999; De Smet and Loriot, 2013). By contrast, genes with a low density of CpG islands in their promoter regions are affected very little by DNA methylation and frequently display an inconstant relationship between transcriptional silencing and promoter methylation (Boyes and Bird, 1992; De Smet and Loriot, 2013).

However, some of the CTA genes that lack rich CpG sites are activated by DNMT inhibitors. For example, the *SPANX* gene was activated by 5-aza-2'-deoxycytidine treatment even though this gene lacks a CpG island in its promoter region (Zendman *et al.*, 2003b; Menendez *et al.*, 2007).

The possible mechanisms underlying DNA demethylation in cancer cells remain poorly understood. However, two processes of DNA demethylation in cancer are known: passive and active demethylation. Passive demethylation is based on the inhibition of DNA methyltransferase enzymes such as DNMT1, DNMT3A and DNMT3B, which commonly preserve DNA methylation marks during mammalian development. Active demethylation, on the other hand, refers to involvement of demethylation enzyme activity such as Ten-eleven translocation (TET) proteins that modifies the methyl group from 5-methylcytosine (5mC) upon base excision DNA repair pathway (BER) (reviewed in Chen and Riggs, 2011; De Smet and Loriot 2013).

Active CTA gene promoters in cancer cells commonly rely on a hypomethylation of the gene promoter. Previous study of the *MAGEA1* promoter shows that protection of the promoter region against DNA remethylation was dependent on the transcriptional activation levels. This indicates that maintenance of the promoter hypomethylation of CTA genes in cancer cells depends on the existence of appropriate transcription factors and their activation (De Smet *et al.*, 2004). For instance, the SP1 transcription factor was found to regulate the activity of several CTA genes including the *MAGE* family as well as the *NY-ESO-1* gene by binding to their promoters in an unmethylated state (De Smet *et al.*, 1995; Kang *et al.*, 2007). SP1- binding elements were shown to preserve the methylation-free events of *NY-ESO-1* and to play a role in protecting the promoter regions against DNA remethylation (Kang *et al.*, 2007). SP1 transcription factors have been proposed to cooperate with the BORIS transcription factor in the demethylation of CTA genes (Vatolin *et al.*, 2005; Kang *et al.*, 2007).

4.1.2 The role of histone modification in CTA genes

The basic unit of chromatin is the nucleosome, which consists of approximately 146 bp of the DNA wrapped around an octamer of four histone proteins H4, H3, H2A and H2B (Sharma *et*

al., 2010). Histone modifications and DNA methylation can influence each other during mammalian development (Cedar and Bergman, 2009). Specific histone H3 modification patterns are associated with CTA gene expression status in malignant tumours (James *et al.*, 2006; Woloszynska-Read *et al.*, 2007; Link *et al.*, 2009). Treatment of cancer cells with a histone deacetylase inhibitor (HDACI) alone is insufficient to activate CTA genes but when supplied in combination with an inhibitor of DNA methylation, this upregulates the transcription level of these genes (Wischnewski *et al.*, 2006).

Previous studies have indicated that CTA genes are transcriptionally silenced in colorectal cancer due to DNA hypermethylation and are induced by genetic targeting of DNMT enzymes in colon cancer cells (Karpf *et al.*, 2004; James *et al.*, 2006; Link *et al.*, 2009). Colorectal cancer (CRC) is the third leading cause of death among the most common malignancies in men and the second in women in developing countries. In the USA an estimated 71,830 males and 65,000 females are diagnosed annually as new cases of colorectal cancer; and approximately 24,040 women and 26,270 men may die from colon cancer (Siegel *et al.*, 2013). HCT116 and SW480 colon adenocarcinoma cell lines are therefore proposed to be useful models for the study of the mechanism of regulation of CTA genes in human cancers.

In this chapter, two epigenetic inhibitors were used to examine the effects of hypomethylation and/or histone deacetylase treatments on the expression of potential novel CTA genes (meiCT genes) which were identified in Chapter 3.0 as potentially encoding immunotherapeutic targets. The DNA hypomethylation agent 5-aza-2'-deoxycytidine (5-AZA-CdR) was shown to induce the expression of silenced CTA genes in different cancer cells depending on the doses and duration of the cell treatments. The histone deacetylase inhibitor Trichostatin A (TSA) has been reported to have a small effect on CTA genes expression when supplied alone but can promote the up-regulation of gene expression when supplied in combination with 5-AZA-CdR (reviewed in De Smet and Loriot, 2013; Whitehurst, 2014). In the work presented here, two colorectal cancer cell lines HCT116 and SW480, were treated with different doses of 5-AZA-CdR and/or TSA for different incubation times in

order to compare the influence of epigenetic treatments on gene expression in both cancer cell types.

Here, 11 potential candidate CTA genes studied in Chapter 3.0 and 5 genes previously published by Feichtinger *et al.,* (2012a) were analysed for activation using RT-PCR in colorectal cancer cells treated with 5-AZA-CdR and/or TSA inhibitors. Two well-known CTA genes, *GAGE1* and *SSX2,* were used in addition to a positive control to assess gene expression. These genes were chosen because their expression cannot be detected in untreated HCT116 and SW480 cancer cells.

The 16 candidate genes were previously categorised into three classes based on their RT-PCR profiles in both normal and cancer cells. Class 1 contains 6 genes that are testis-restricted; class 2 contains 7 genes that are cancer-testis restricted; and class 3 contains 3 genes that are cancer/testis-CNS restricted (Feichtinger *et al.*, 2012a).

4.2   Results

4.2.1   The influence of 5-AZA-CdR and/or TSA inhibitors on the expression of testis-restricted genes in HCT116 cancer cells

The influence of DNA demethylation and histone deacetylation on the expression status of CTA genes was investigated by treating the human colon cancer cells lines HCT116 and SW480 with different doses of 5-AZA-CdR and/or TSA inhibitors. The HCT116 and SW480 cells were treated with a range of concentrations of 5-AZA-CdR (0.1, 0.5, 1.0, 5.0, 10.0 and 15.0 μM) for 48 or 72 hours. Cells were also treated with the histone deacetylation TSA inhibitor at different concentrations (150, 300 and 600 nM) for 24 hours. The combined influence of DNA demethylation and histone deacetylation on CTA gene induction was examined by treating cells with different doses of 5-AZA-CdR (0.1, 0.5, 1.0, 5.0, 10.0 and 15.0 μM) in combination with 300 nM TSA for 48 hours.

The HCT116 and SW480 cells treated with DMSO were used to compare the gene expression since the drugs were dissolved in DMSO. Untreated HCT116 and SW480 cells were also used as a negative control to compare the expression of CTA genes in treated cells; the normal testis was also employed as a positive control.

Two well-known X-CTA genes, *GAGE1* and *SSX2,* previously reported to be actively transcribed via DNA hypomethylation were used to assess gene expression. The 16 CTA genes were previously categorised based on their RT-PCR profile in both normal and cancer cells into three classes (Table 4.1).

RT-PCR was used to validate the expression in both treated and untreated cell lines. The cDNA was synthesised from total RNA preparations isolated from both treated and untreated cells. The cDNA quality was assessed using the *β-actin* gene in this study. Primers were designed to span more than one intron where possible to avoid genomic DNA contamination.

The RT-PCR validation of six testis-restricted genes in HCT116 cells treated with 5-AZA-CdR identified two genes that were apparently activated by DNA hypomethylation, whereas four

genes showed no effect of the inhibitor on gene expression (Figure 4.1). *DDX4* was expressed after treatment of HCT116 cells with 0.5 to 15 µM of 5-AZA-CdR for 48 hours. The *DDX4* gene was also expressed after treatment with 0.1 to 15 µM of 5-AZA-CdR for 72 hours. The *IQCF3* gene displayed faint bands after treatment of HCT116 cells with 5-AZA-CdR for 72 hours. Treatment with different concentration of 5-AZA-CdR resulted in no activation of four genes *ARRDC5*, *C4orf17*, *NT5C1B* and *TMEM202;* these genes all remained silenced.

The expression profile of the control genes *GAGE1* and *SSX2* showed activation after the cells were treated with 5-AZA-CdR. The intensity of the bands for *DDX4*, *IQCF3* and control genes increased throughout the time course and/or in response to the concentration of inhibitor uses as the treatment.

The possibility that histone deacetylation can induce the expression of testis-restricted genes was explored by performing RT-PCR analysis of HCT116 cells treated with the TSA inhibitor. The expression profiles of six testis-restricted genes showed that these genes remained silenced when the cells were treated with different concentrations of TSA (150, 300 and 600 nM) for 24 hours. Only the *GAGE1* control gene showed induction after the cells were treated with 300 nM and 600 nM TSA (Figure 4.2).

The potential for a combined treatment of 5-AZA-CdR and TSA to activate the expression of testis-restricted genes was examined by analysing gene expression in HCT116 cells treated with 5-AZA-CdR at different doses (0.1, 0.5,1.0,5.0,10.0 and 15.0 µM) in combination with 300 nM TSA for 48 hours. The expression profiles of six testis-restricted genes indicated that they remained silenced with the combined treatment, with the exception of the *IQCF3* gene, which showed significant expression.

Surprisingly, the combination of histone deacetylation and DNA demethylation inhibitors resulted in transcriptional silencing of *DDX4* expression and increased the expression of *IQCF3* in HCT116 cells. The band intensity for the expression of the control *SSX2* gene also imply decreased expression following treatment with 300 nM TSA (Figure 4.3; although

quantitative RT-PCR would be need to confirm this). The PCR products of these genes were sequenced to ensure that the correct gene had been amplified. The sequencing results are shown in Table 4.2.

4.2.2 The influence of 5-AZA-CdR and/or TSA inhibitors on the expression of testis-restricted genes in SW480 cancer cells

Treatment of HCT116 cells with different concentrations of 5-AZA-CdR and/or TSA for short and long time courses had no effect on the expression of four genes, which remained silenced. However, CTA genes are expressed in some cancer cells and lack expression in others. This indicated CTA genes can be expressed in different patterns in cancer cells (Simpson *et.al*, 2005; Feichtinger *et al.*, 2012a). We next asked whether treatment of SW480 cells with epigenetic inhibitors might induce the expression of testis-restricted genes.

The RT-PCR analysis of six testis-restricted genes in SW480 cells treated with 5-AZA-CdR revealed the induction of only two genes, *DDX4* and *IQCF3*. No expression of *ARRDC5*, *C4orf17*, *NT5C1B* or *TMEM202* was detected in SW480 cells after treatment (Figure 4.4). The *DDX4* gene showed faint bands after treatment of SW480 cells with 5 to 15 µM of 5-AZA-CdR for 48 hours. The expression of *DDX4* was also up-regulated with moderate intensity bands after treatment with 0.5 to 15 µM of 5-AZA-CdR for 72 hours. This result indicated that *DDX4* was more strongly induced in by 5-AZA-CdR in HCT116 cells than in SW480 cells. In addition, the intensity of the *DDX4* bands was influenced by the incubation time and the treatment doses.

On the other hand, *IQCF3* showed faint expression after treatment of SW480 cells with 5-AZA-CdR but only after 72 hours of treatment. Increasing the treatment concentration and/or the incubation time did not result in any activation of the expression of *ARRDC5*, *C4orf17*, *NT5C1B* or *TMEM202*. Both CTA control genes *GAGE1* and *SSX2* were activated and showed strong band intensity in SW480 cells treated with 5-AZA-CdR (Figure 4.4).

No significant expressions of testis-restricted genes were observed in SW480 cells treated with TSA, with the exception of *IQCF3*. Interestingly, *IQCF3* was induced in SW480 cells at

TSA concentrations of 300 nM and 600 nM, whereas this gene was not expressed in HCT116 cells at these TSA concentrations (Figure 4.5).

In addition, co-treatment of SW480 cells with 5-AZA-CdR and TSA did not induce the expression of *ARRDC5*, *C4orf17*, *DDX4*, *NT5C1B* or *TMEM202*. The only testis-restricted gene to show significant expression was *IQCF3* (Figure 4.6), which displayed moderate bands intensity after treatment of SW480 cells with 0.1 to 15 μM of 5-AZA-CdR in combination with 300 nM TSA for 48 hours. However, the same level of expression observed in the testis could not be attained in most cases for the testis-restricted genes. The PCR products of these genes were purified and sequenced to confirm that the correct DNA sequence was amplified (summary of sequencing, Table 4.2).

Table 4.1. Classification and expression of CTA genes were previously categories based on their RT-PCR profile in both normal and cancer cells.

Gene name	Chromosome location	Expression in normal tissues	Expression in cancer tissues	Expression in colon cancer
Positive control genes				
GAGE1	X	No	Embryonic carcinoma, cervix, colon, ovary, melanoma and leukaemia	HT29 and LoVo cell lines
SSX2	X	No	Melanoma and leukaemia	No
Class 1: testis-restricted genes				
ARRDC5	19	No	No	No
C4orf17	4	No	No	No
DDX4	5	No	No	No
IQCF3	3	No	No	No
NT5C1B	2	No	No	No
TMEM202	15	No	No	No
Class2: cancer/testis-restricted genes				
*C17orf98	17	No	Embryonic carcinoma	No
ODF4	17	No	Embryonic carcinoma, Breast, cervix, ovary, stomach melanoma and leukaemia	No
PRDM7	16	No	melanoma and leukaemia	SW480 cells
*PRDM9	5	No	Embryonic carcinoma, breast, cervix, colon, ovary, melanoma and leukaemia	LoVo cell line
*SMC1β	22	No	Embryonic carcinoma, breast and leukaemia	No
*STRA8	7	No	Breast, lung, liver and uterus	No
TDRD12	19	No	Embryonic carcinoma	No
Class3: cancer/testis-CNS-restricted				
*C20orf201	20	CNS	Breast, ovary, melanoma and leukaemia	No
SEPT12	16	CNS	Breast, colon and ovary	SW480 cells
*SYCP1	1	CNS	Embryonic carcinoma, melanoma and ovary	No

*These genes were published in Feichtinger *et al*., (2012a).
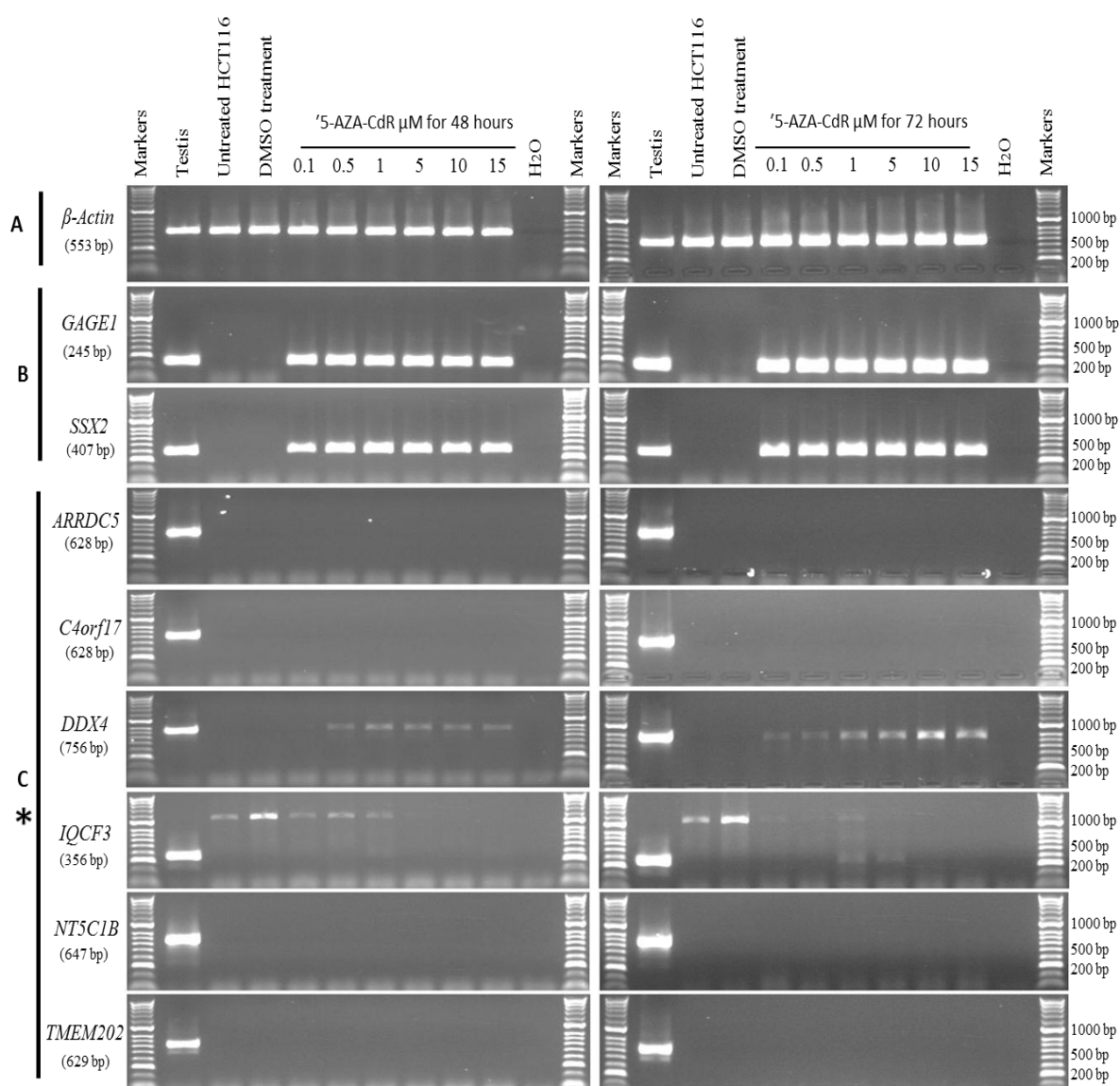
Figure 4.1. The effect of 5-AZA-CdR on testis-restricted genes expression in the HCT116 cancer cell line. Agarose gels showing the expression of six testis-restricted genes after HCT116 cells treatment with a range of concentrations of 5-AZA-CdR (0.1, 0.5, 1.0, 5.0, 10.0 and 15.0 μM) for 48 hours (on the left) or 72 hour (on the right). Untreated HCT116 cells were used to compare with treated cells and a testis sample served as a positive control. Control HCT116 cells were treated with DMSO since DMSO was the solvent used for the 5-AZA-CdR solution. (C) *DDX4* and *IQCF3* were slightly activated after treatment for 72 hours, whereas the four genes *ARRDC5*, *C4orf17*, *NT5C1B* and *TMEM202* remained silenced. (A) Expression of the *β-Actin* gene shows the positive control of the cDNA samples in both cases. (B) The expression profiles of known CT antigen genes, *GAGE1* and *SSX2*, show that these genes are activated after cell treatment with 5-AZA-CdR and the intensity of the bands increases with the time and concentration of the treatment.
*This is mispriming product not belong to *IQCF3* gene
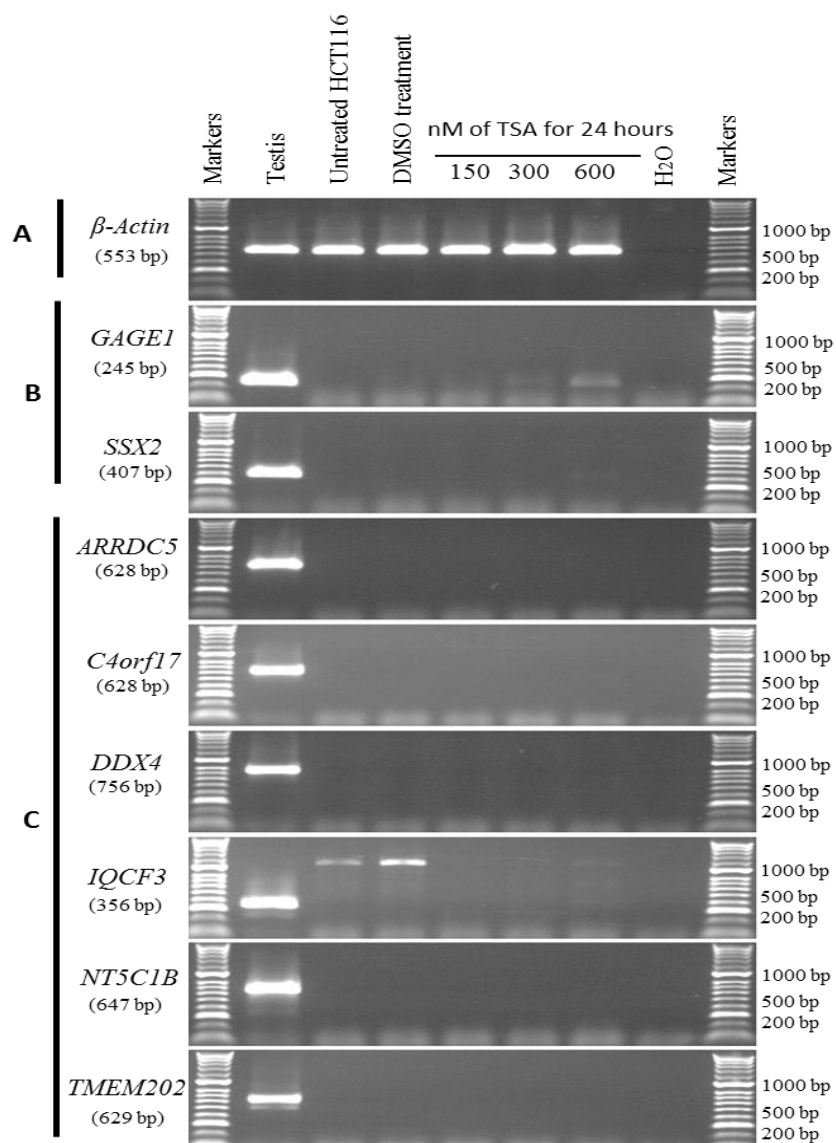
Figure 4.2. **The effect of TSA on testis-restricted genes expression in the HCT116 cancer cell line.** Agarose gels showing the expression of six testis-restricted genes after treatment of HCT116 cells with a range of concentrations of TSA (150 nM, 300 nM and 600 nM) for 24 hours. Untreated HCT116 cells were used to compare with treated cells and a testis sample served as a positive control. Control HCT116 cells were treated with DMSO since this was the solvent used for the TSA solution. (A) *β-Actin* served as a positive control for the cDNA. (B) *GAGE1* and *SSX2* served as control CTA genes. (C) The expression of six testis-restricted CT genes appears to remain silenced even after treatment with different concentrations of TSA.
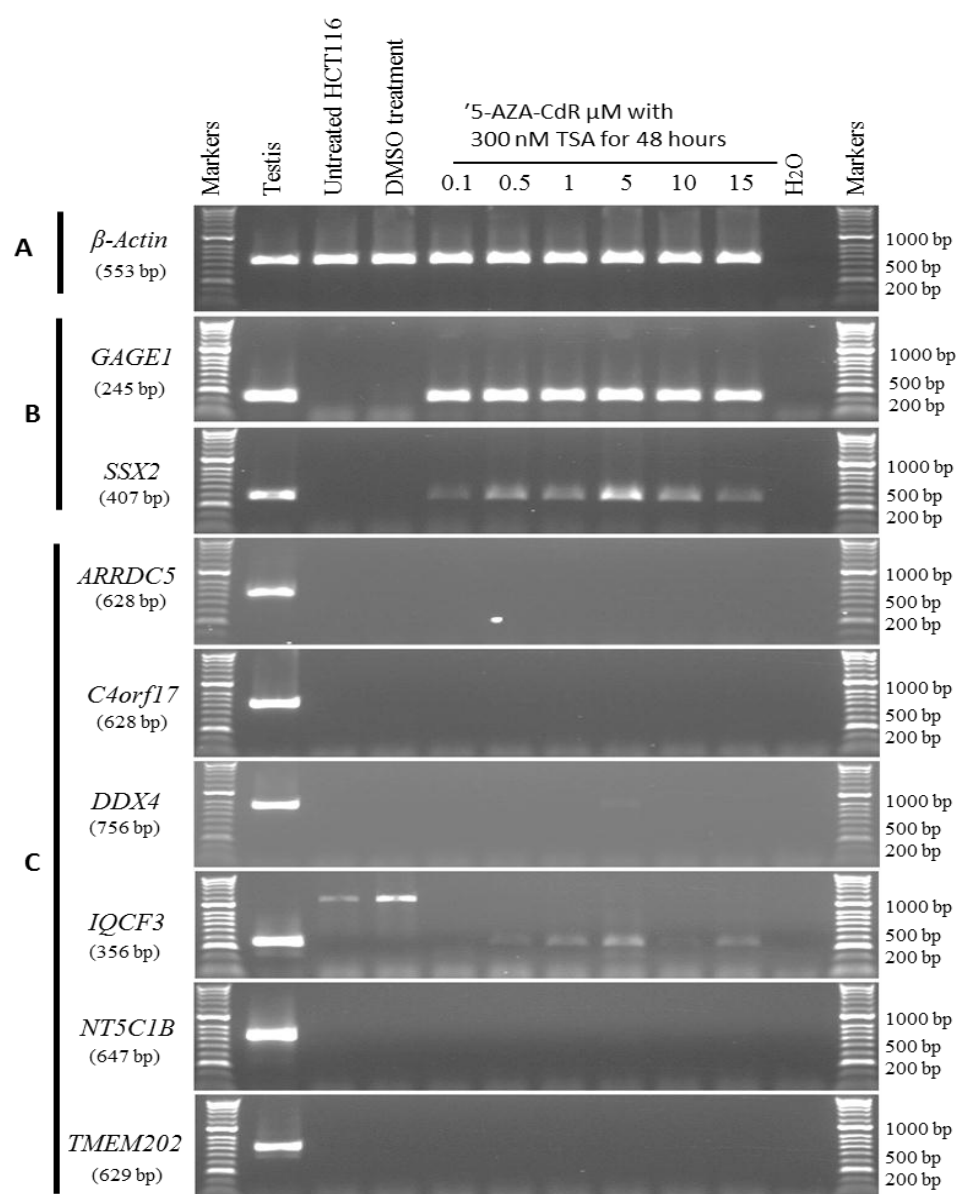
Figure 4.3. **The effect of co-treatment of 5-AZA-CdR with TSA on testis-restricted genes expression in HCT116 cancer cell line.** Agarose gels showing the expression of six testis-restricted genes after treatment of HCT116 cells with a range of concentrations of 5-AZA-CdR (0.1, 0.5, 1.0, 5.0, 10.0 and 15.0 µM) and 300 nM TSA for 48 hours. Untreated HCT116 cells were used to compare with treated cells and a testis sample served as a positive control. (Control HCT116 cells were treated with DMSO since this was the solvent used for the 5-AZA-CdR solution.) (A) *β-Actin* served as a positive control for the cDNA. (B) *GAGE1* and *SSX2* served as control CTA genes. (C) Six testis-restricted CT genes appear to remain silenced with exception of *IQCF3* gene which has significant expression.
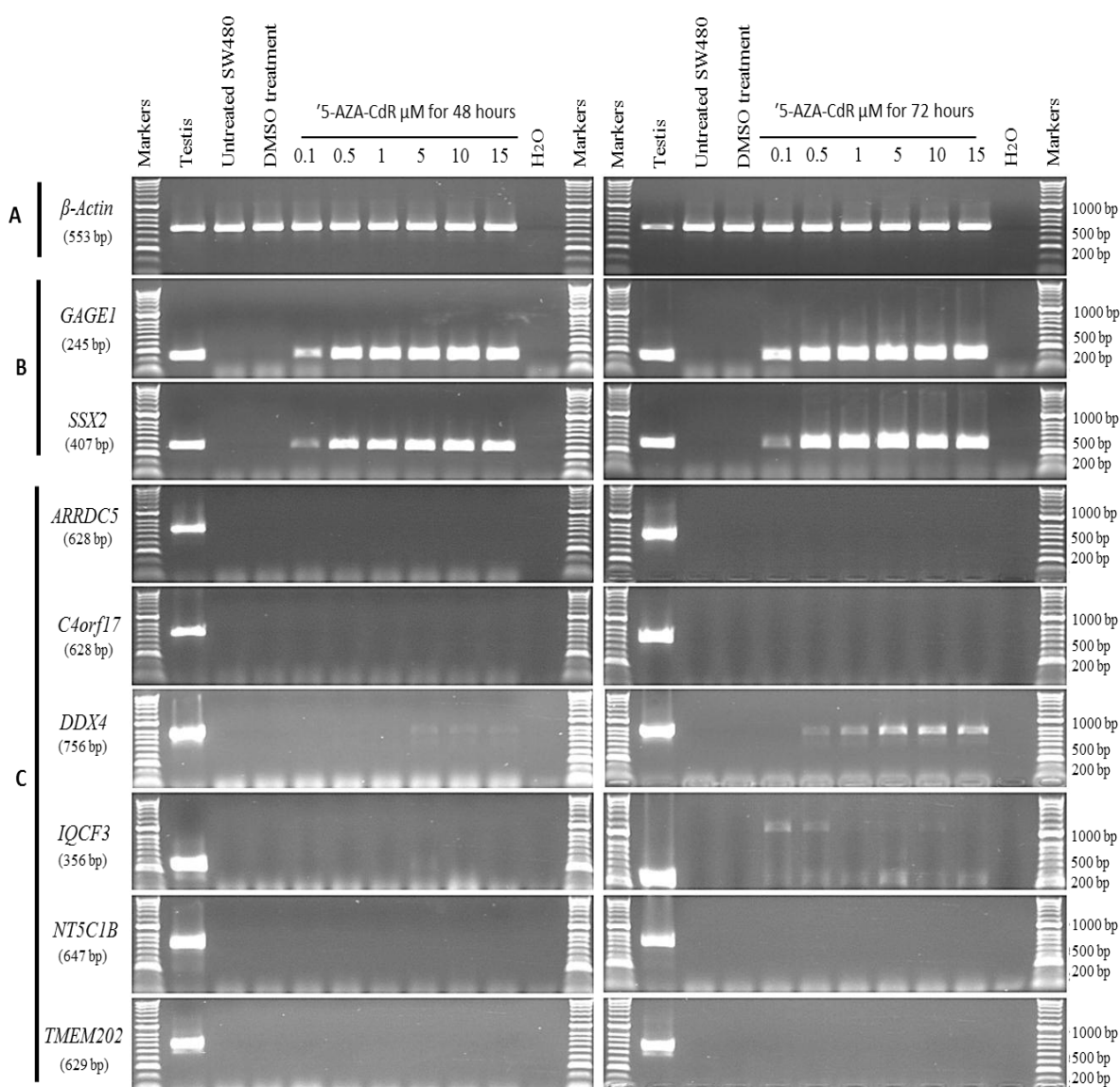
Figure 4.4. **The effect of 5-AZA-CdR on testis-restricted genes expression in SW480 cancer cell line.** Agarose gels showing the expression of six testis-restricted genes after treatment of SW480 cells with a range of concentrations of 5-AZA-CdR (0.1, 0.5,1,5,10 and 15 µM) for 48 hours (on the left) or 72 hour (on the right). Untreated SW480 cells were used to compare with treated cells and a testis sample served as a positive control. Control SW480 cells were treated with DMSO since this was the solvent used for the 5-AZA-CdR solution. (A) *β-Actin* served as a positive control for the cDNA. (B) *GAGE1* and *SSX2* served as control CTA genes. (C) *DDX4* and *IQCF3* were activated after cell treatment for 72 hours whereas four other genes, *ARRDC5*, *C4orf17*, *NT5C1B* and *TMEM202,* remained silenced.
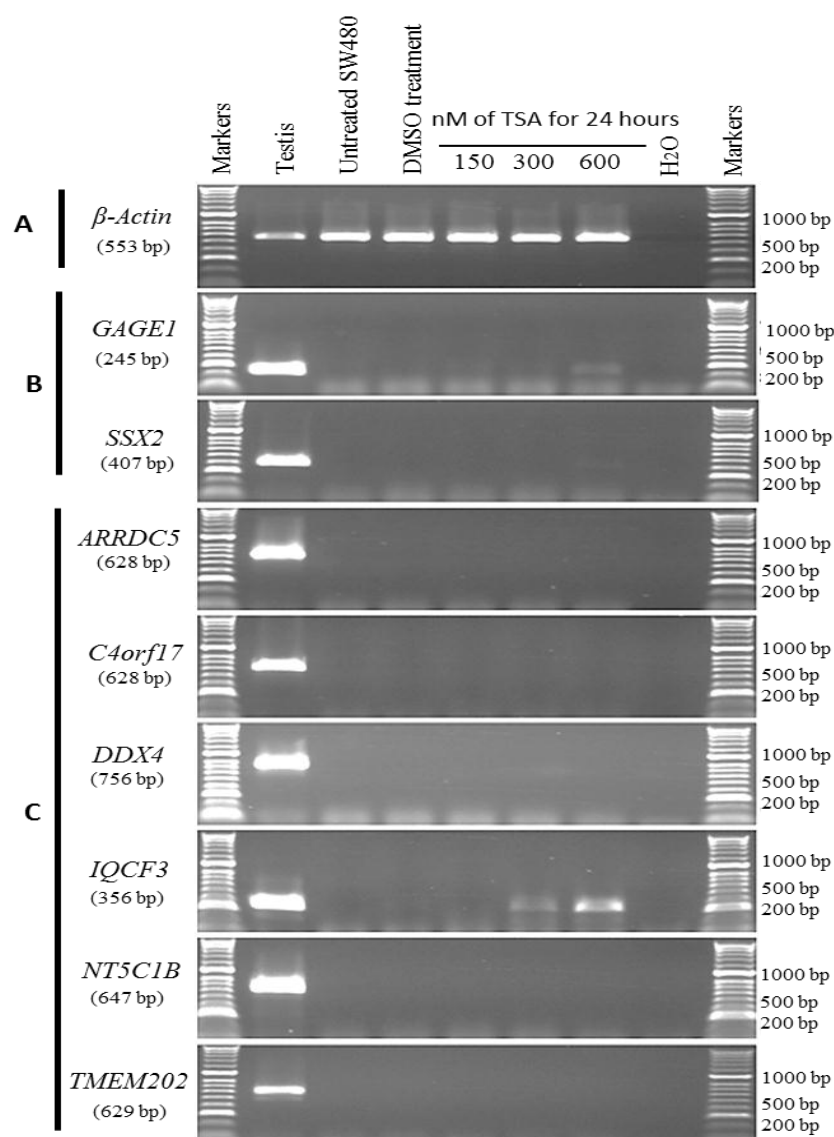
Figure 4.5. **The effect of TSA on testis-restricted genes expression in SW480 cancer cell line.** Agarose gels showing the expression of six testis-restricted genes after treatment of SW480 with a range of concentrations of TSA (150 nM, 300 nM and 600 nM) for 24 hours. Untreated SW480 cells were used to compare with treated cells and a testis sample served as a positive control. Control SW480 cells were treated with DMSO since this was the solvent used for the TSA solution. (A) *β-Actin* served as a positive control for the cDNA. (B) *GAGE1* and *SSX2* served as control CTA genes. (C) Only *IQCF3* showed significant expression after TSA treatment, whereas the five other testis-restricted genes remained silenced.
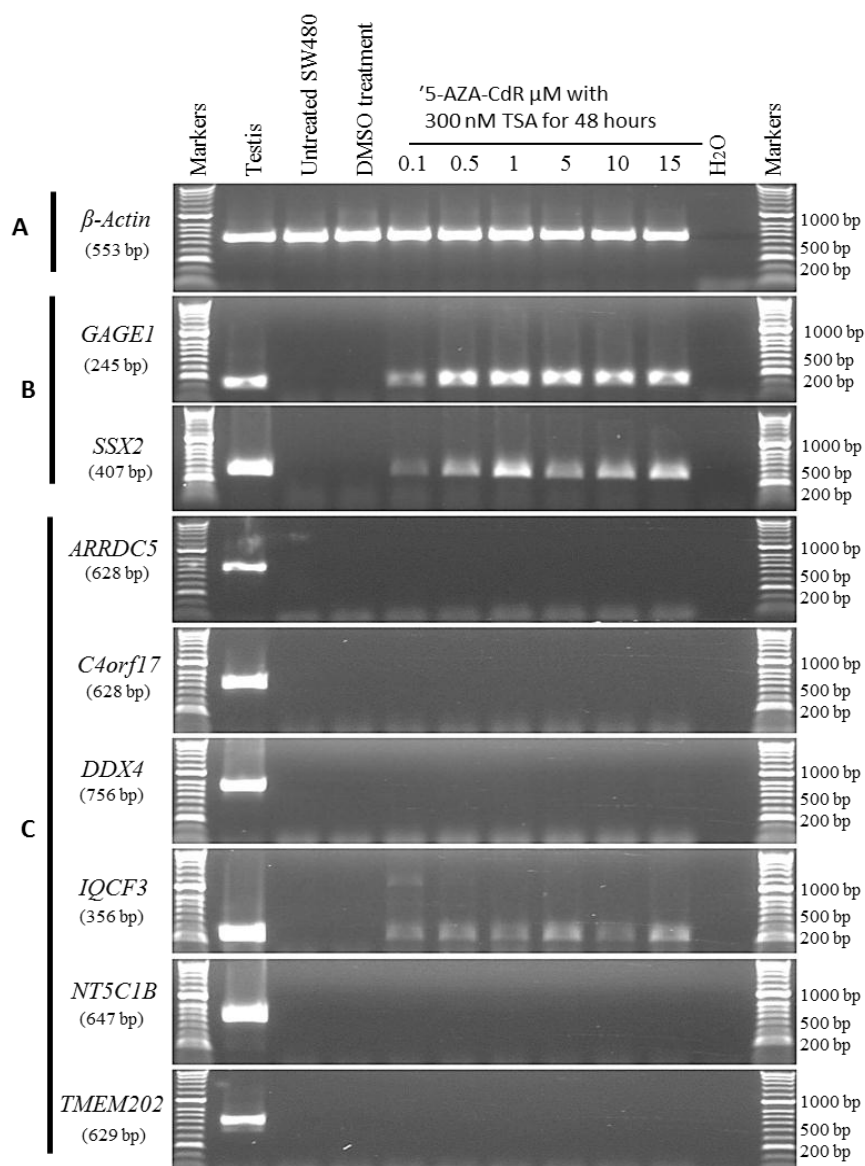
Figure 4.6. **The effect of co-treatment of 5-AZA-CdR with TSA on testis-restricted genes expression in SW480 cancer cell line.** Agarose gels showing the expression of six testis-restricted genes after SW480 cells treatment with range of concentration of 5-AZA-CdR (0.1, 0.5,1.0,5.0,10.0 and 15.0 μM) with 300 nM TSA for 48 hours. Untreated SW480 cells were used to compare with treated cells and a testis sample served as a positive control. Control SW480 cells were treated with DMSO since this was the solvent used for the 5-AZA-CdR and TSA solutions. (A) *β-Actin* served as a positive control for the cDNA. (B) *GAGE1* and *SSX2* served as control CTA genes (C) Five testis-restricted CT genes showed no induction following co-treatment with 5-AZA-CdR with TSA; only the *IQCF3* gene showed significant expression.

4.2.3   The influence of 5-AZA-CdR and/or TSA inhibitors on the expression of potential cancer/testis-restricted genes in HCT116 cancer cells

Here, 5 potential novel cancer/testis-restricted CT genes which were previously published by Feichtinger et al. (2012a) as well as two genes, *ODF4* and *TDRD12,* identified in Chapter 3.0, were analysed using RT-PCR in both untreated colorectal cancer cells and cells treated with 5-AZA-CdR and/or TSA inhibitors. The expression profiles of these genes are restricted to human germ lines and are observed in several types of cancer cells (Table 4.1). The aberrant expression of these genes remains poorly understood; however, epigenetic alteration-particularly DNA hypomethylation and histone modification-were described as regulators of CTA gene expression (reviewed in Fratta *et al.*, 2011).

The RT-PCR analysis of seven CT genes in HCT116 cells treated with 5-AZA-CdR showed four genes that were apparently activated by DNA hypomethylation, whereas the expression of the other 3 genes were not affected (Figure 4.7). *STRA8* and *TDRD12* showed strong bands after cell treatment with 0.1 to 15 µM 5-AZA-CdR for 48 or 72 hours. The level of band intensity of *STRA8* and *TDRD12* were similar to that observed for testis control. The expressions of *C17orf98* and *SMC1β* were influenced by the 5-AZA-CdR concentration as well as the time course. *SMC1β* displayed faint bands after cells were treated with 0.1 to 15 µM 5-AZA-CdR for 72 hours. The expressions of the *ODF4*, *PRDM7* and *PRDM9* genes were not influenced by 5-AZA-CdR, despite the use of different concentrations and short and long courses of the drug.

The possibility that histone deacetylation could influence the expression of CTA genes was addressed by performing RT-PCR validations in HCT116 cells treated with TSA. Only one of the 7 genes showed activation treatment of cells with 150, 300 and 600 nM TSA (Figure 4.8). This result suggests that *STRA8* can be expressed through both DNA demethylation and histone deacetylation in HCT116 cells. Expressions of *ODF4*, *PRDM7* and *PRDM9* were not observed either under the influence of DNA hypomethylation or histone alterations; however, *C17orf98*, *SMC1β* and *TDRD12* showed re-expression following DNA hypomethylation.

The potential for a combination of 5-AZA-CdR and TSA treatments to stimulate the expression of CTA genes was evaluated by analysing these genes using RT-PCR in HCT116 cells co-treated with 5-AZA-CdR (0.1, 0.5, 1.0, 5.0, 10.0 and 15.0 µM) and 300 nM TSA for 48 hours. The *C17orf98*, *ODF4*, *PRDM7*, *PRDM9* and *SMC1β* genes showed no expression in these co-treated cells, whereas *STRA8* and *TDRD12* appeared to be activated, as indicated by strong band intensity for *STRA8* and moderate intensity for *TDRD12* (Figure 4.9). The data indicated that treatment HCT116 cells with TSA in combination with 5-AZA-CdR prevented the expression of *C17orf98* and *SMC1β*. Also, co-treatment of HCT116 cells reduced the expression of *TDRD12* as well as the *SSX2* positive control gene. The PCR products of these genes were sequenced to ensure the correct gene had been amplified; the results of the sequencing are shown in Table 4.2.

4.2.4 The influence of 5-AZA-CdR and/or TSA inhibitors on the expression potential cancer/testis-restricted genes in SW480 cancer cells

ODF4, *PRDM7* and *PRDM9* were expressed in several cancer cells/tissues (Table 4.1); but these genes were not expressed in HCT116 cells treated with epigenetic inhibitors. On the other hand, the genes *C17orf98*, *SMC1β*, *STRA8* and *TDRD12* from a similar category were induced after HCT116 cells were treated with a DNA hypomethylation inhibitor. We asked whether these genes might have different behaviour patterns after epigenetic drug treatment in SW480 cells. The results supported the idea that these genes might have different patterns in the same type of cancer; for example, *PRDM9* was expressed only in one colorectal cell line (LoVo) and remained silent in 5 cancer cell lines/tissues (Feichtinger *et al*., 2012a).

The RT-PCR profile of seven CTA genes in SW480 cells treated with 5-AZA-CdR revealed five genes that were apparently activated by DNA hypomethylation, whereas expression of two genes was unaffected (Figure 4.10). *TDRD12* was expressed after treatment of SW480 cells with 0.5 to 15 µM 5-AZA-CdR for 48 or 72 hours. The level of band intensity of *TDRD12* was similar to that in the testis, particularly after 72 hours at high treatment doses. The *STRA8* gene was activated and showed faint to moderate band intensity after cells were incubated with high treatment doses for 48 or 72 hours. The *C17orf98*, *PRDM9* and *SMC1β* genes

showed weak expression after 72 hours treatment. The data suggest that treatment of SW480 cells with 5-AZA-CdR may restore the expression of *PRDM9* while the same treatment doses and incubation times could not restore expression of this gene in HCT116 cells.

The band intensities of *STRA8* and *SMC1β* in SW480 cells treated with 5-AZA-CdR were lower than in treated HCT116 cells. The *STRA8* and *SMC1β* expression was also not activated in SW480 cells treated with low doses. *PRDM7* showed a faint band in untreated SW480 cells and *ODF4* remained silenced; neither gene showed a significant response to 5-AZA-CdR treatment in either SW480 or HCT116 cells, even when different doses of 5-AZA-CdR and short and long course treatments were used.

Two of the 7 genes were activated after treatment of SW480 cells with TSA inhibitor (Figure 4.11). *C17orf98* and *STRA8* expressions were slightly activated with 600 nM TSA, while the other five genes showed no expression. This result suggests that *STRA8* can be expressed through histone deacetylation in both HCT116 and SW480 cells but the expression in HCT116 cells appeared to be stronger. Also, *C17orf98* was expressed with a faint band in SW480 cells treated with TSA but not in treated HCT116 cells. The data suggest that the expressions of *ODF4*, *PRDM7*, *PRDM9* and *SMC1β* are not induced in either HCT116 or SW480 cells by histone alterations.

In addition, co-treatment of SW480 cells with 5-AZA-CdR and TSA did not promote expression of *ODF4* and *PRDM7* whereas 5 genes were influenced in their expression (Figure 4.12). The expression of *ODF4* and *PRDM7* was not influenced by 5-AZA-CdR and/or TSA. *C17orf98*, *PRDM9* and *SMC1β* were expressed as faint bands particularly at high treatment doses. This result indicates that the expression of *PRDM9* could be activated through co-treatment of SW480 cells but not in HCT116 cells. *STRA8* and *TDRD12* showed activation, with strong band intensity for *TDRD12* and moderate for *STRA8*. The PCR products of these genes were sequenced to ensure the correct gene had been amplified and the results of the sequencing are shown in Table 4.2.
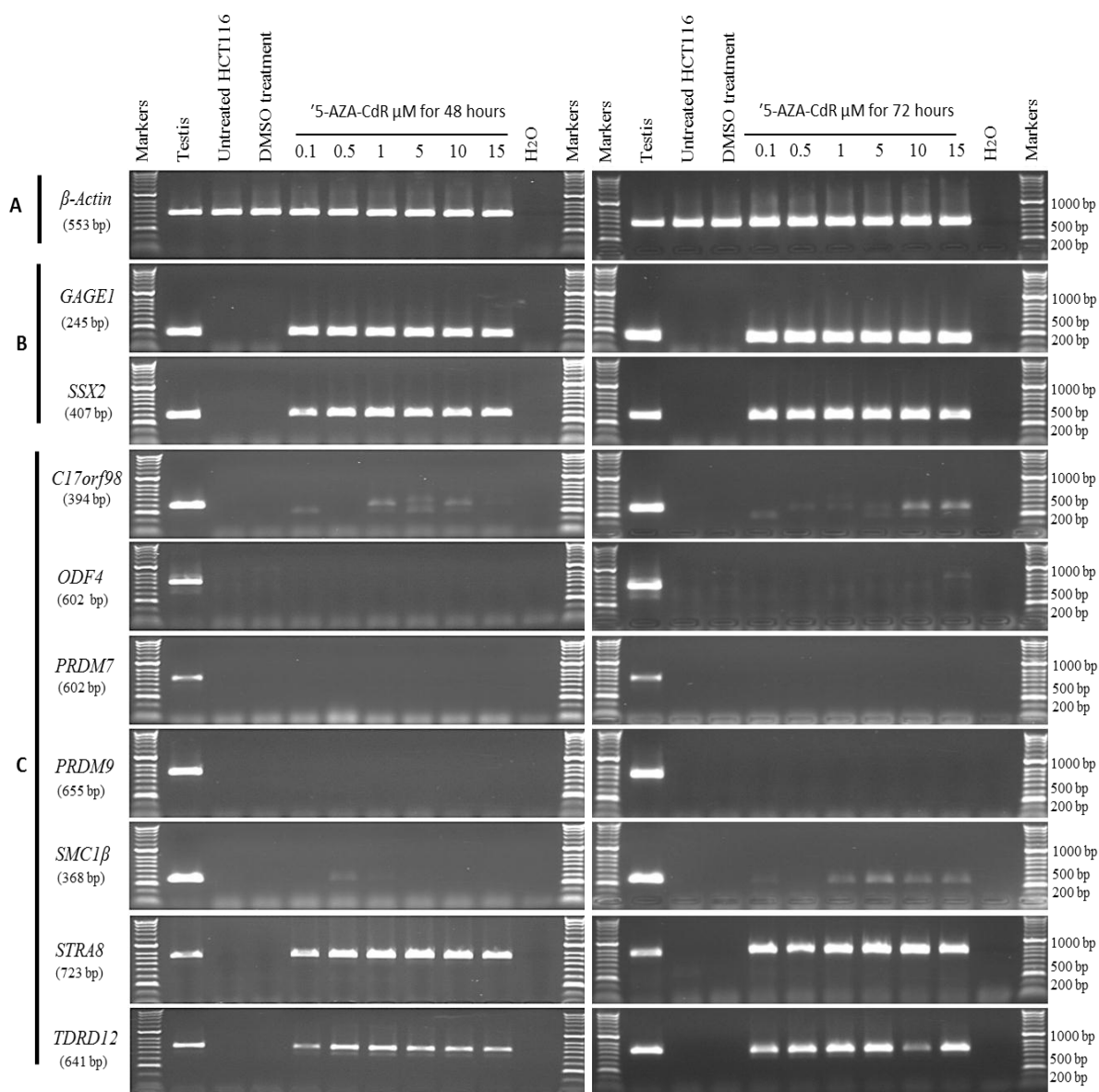
Figure 4.7. The effect of 5-AZA-CdR on the potential cancer/testis-restricted genes expression in HCT116 cancer cell line. (A) *β-Actin* serves as positive control of the cDNA. (B) *GAGE1* and *SSX2* serves as control CTA genes. (C) The *STRA8* and *TDRD12* genes appear to be activated with strong expression, whereas *C17orf98* and *SMC1β* show moderate expression after treatment of HCT116 cells with 5-AZA-CdR. *ODF4*, *PRDM7* and *PRDM9* genes show no expression in response to the methylation treatment.
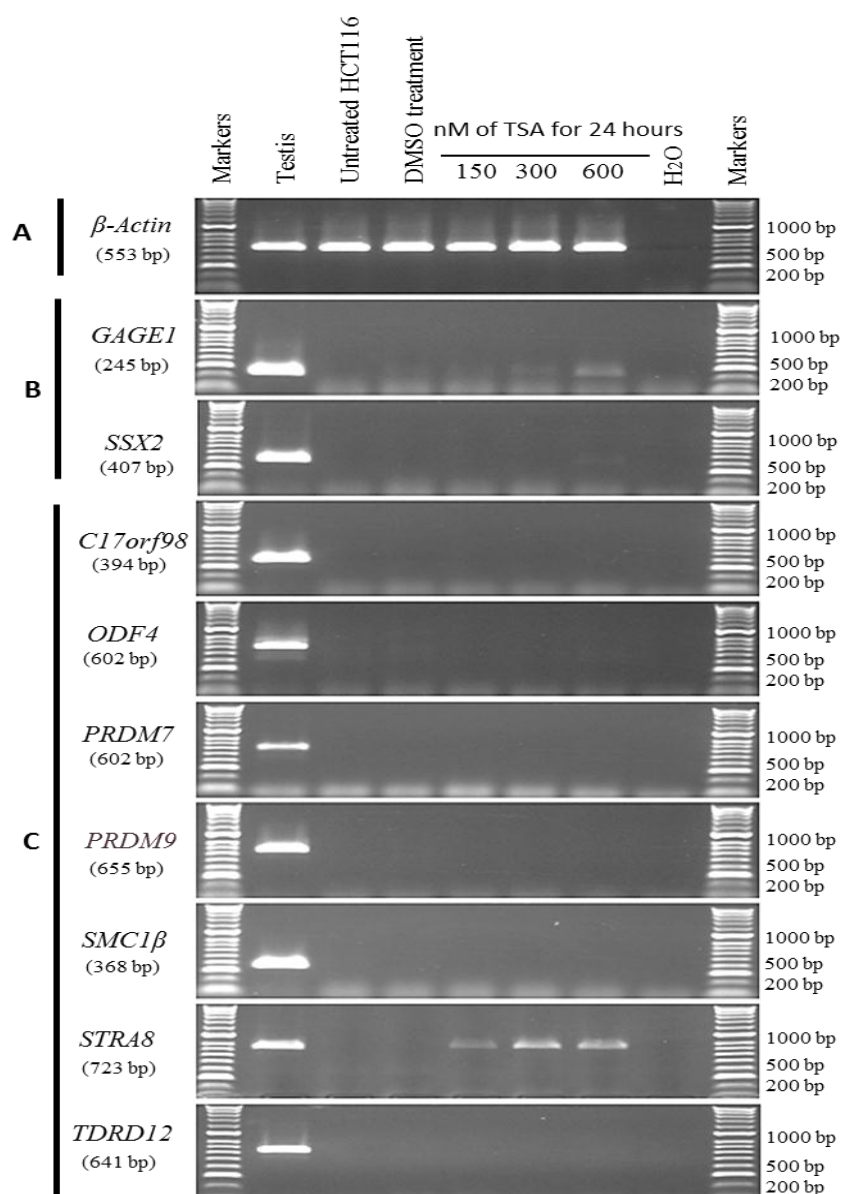
Figure 4.8. The effect of TSA on the potential cancer/testis-restricted genes expression in HCT116 cancer cell line. (A) *β-Actin* serves as positive control of the cDNA. (B) *GAGE1* and *SSX2* serves as control of CTA genes. (C) The expression of seven CTA genes remained silenced after TSA treatment, with exception of the *STRA8* gene which showed significant expression after treatment with 150, 300 and 600 nM TSA.
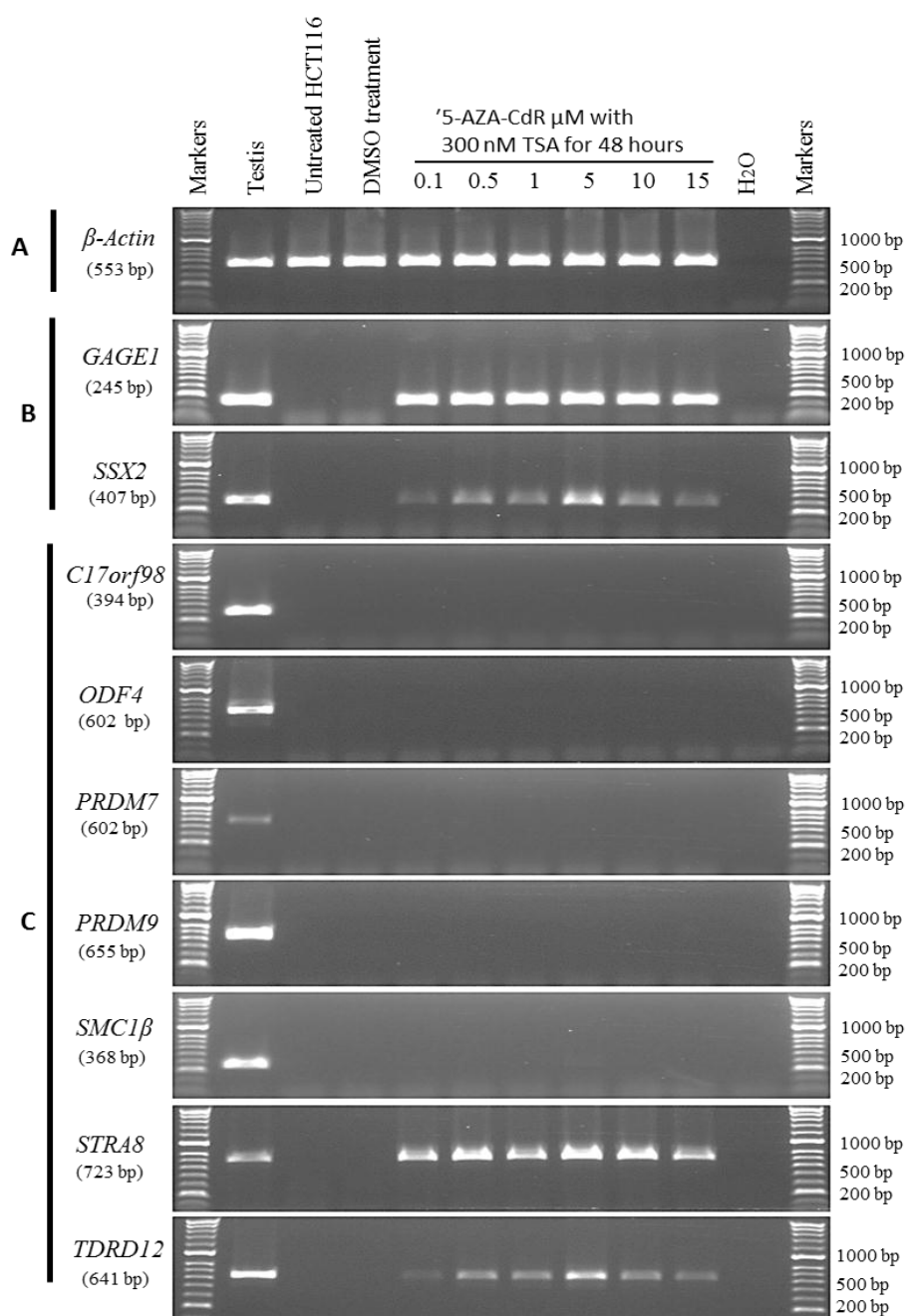
Figure 4.9. The effect of co-treatment with 5-AZA-CdR and TSA on potential cancer/testis-restricted genes expression in HCT116 cancer cell line. (A) *β-Actin* serves as positive control of the cDNA. (B) *GAGE1* and *SSX2* serves as control of CTA genes. (C) *STRA8* and *TDRD12* genes appear to be activated with strong expression whereas *C17orf98*, *ODF4*, *PRDM7*, *PRDM9* and *SMC1β* genes show no expression after co-treatment of HCT116 cells with 5-AZA-CdR and TSA.
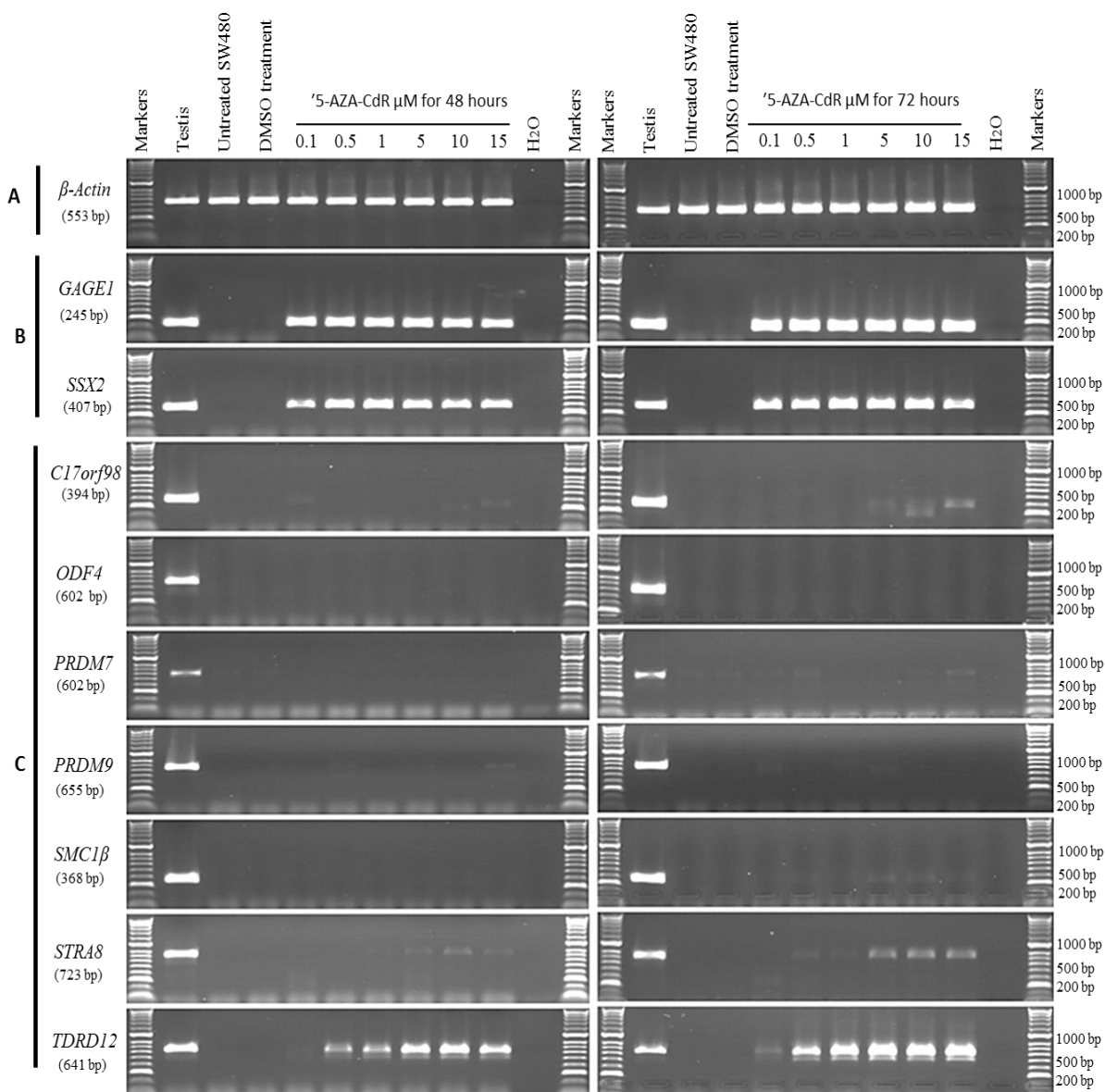
Figure 4.10. **The effect of 5-AZA-CdR on expression of potential cancer/testis-restricted genes in SW480 cancer cell line.** (A) *β-Actin* served as a positive control of the cDNA. (B) *GAGE1* and *SSX2* served as control CTA genes. (C) *TDRD12* expression showed stronger bands than STRA8 gene expression. *C17orf98*, *PRDM9* and *SMC1β* were expressed with weak expression after for a 72 hour treatment. The *ODF4* and *PRDM7* genes showed no significant response to 5-AZA-CdR treatment.
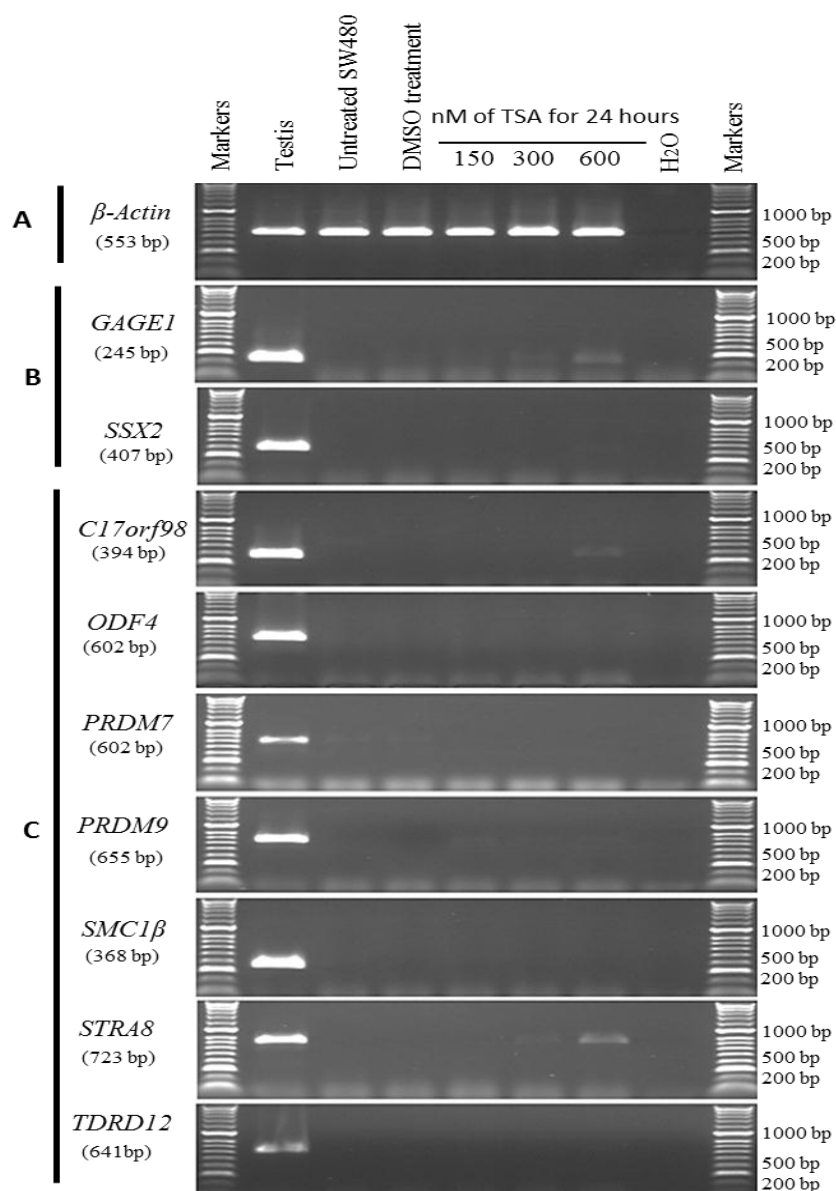
Figure 4.11. **The effect of TSA on the potential cancer/testis-restricted genes expression in SW480 cancer cell line.** (A) *β-Actin* served as a positive control for the cDNA. (B) *GAGE1* and *SSX2* served as control CTA genes (C) *C17orf98* and *STRA8* were slightly activated by a 600 nM TSA treatment. All five other genes showed no expression after TSA treatment.
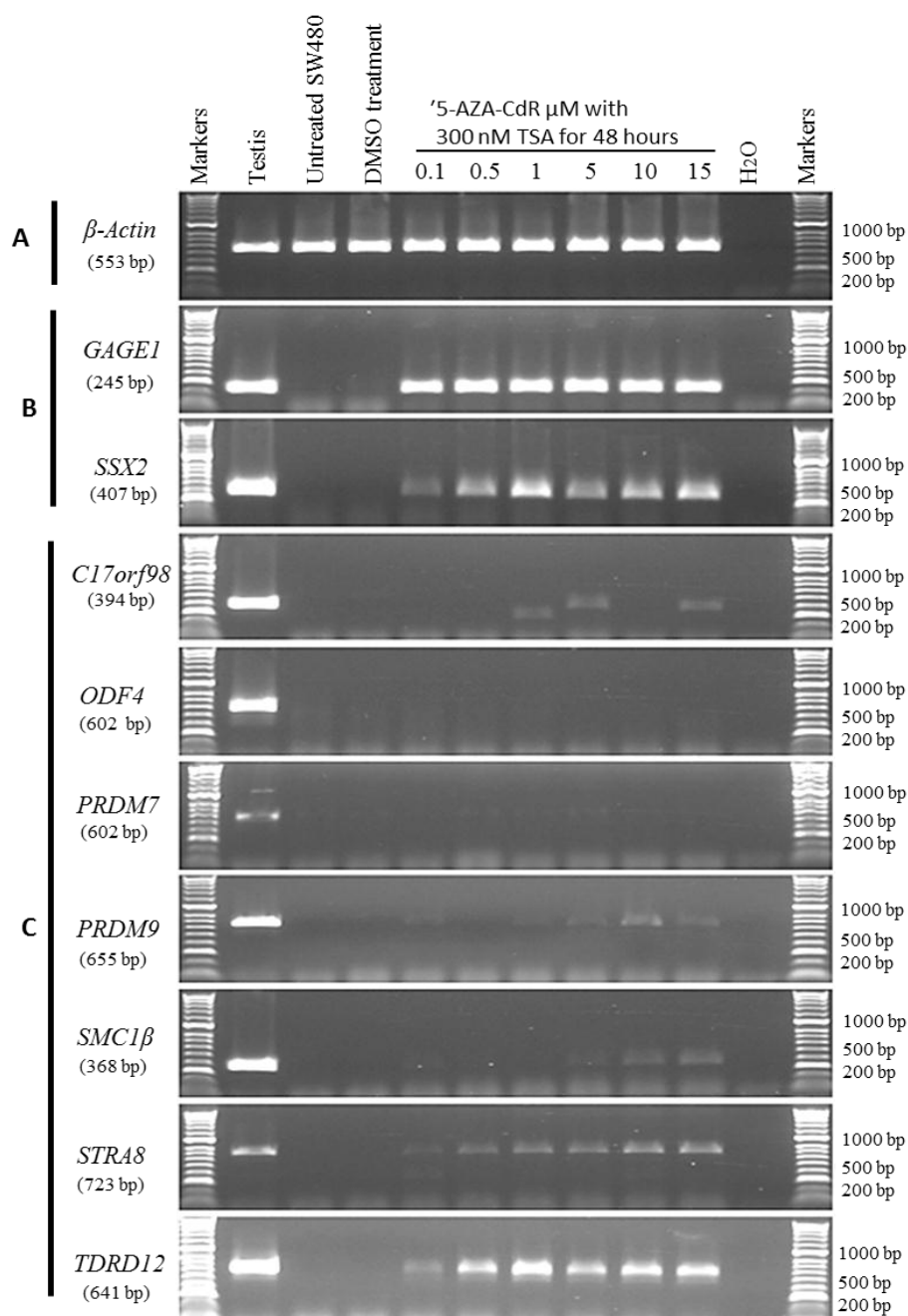
Figure 4.12. **The effect of co-treatment of 5-AZA-CdR with TSA on potential cancer/testis-restricted genes expression in SW480 cancer cell line.** (A) *β-Actin* served as a positive control for the cDNA. (B) *GAGE1* and *SSX2* served as control CTA genes. (C) Co-treatment activated the expression of five genes *C17orf98*, *PRDM9* and *SMC1β, STRA8* and *TDRD12*. The *ODF4* and *PRDM7* genes showed no significant expression after cell treatment.

4.2.5 The influence of 5-AZA-CdR and/or TSA inhibitors on the expression of cancer/testis-CNS-restricted genes in HCT116 cancer cells

Here, two cancer/testis-CNS restricted genes which were previously published by Feichtinger et al. (2012a), as well as the *SEPT12* gene identified in Chapter 3.0, were analysed using RT-PCR in both untreated colorectal cancer cells and in cells treated with 5-AZA-CdR and/or TSA inhibitors. The expression profiles of these genes are restricted to human germ lines, CNS, and several cancer cell types (Table 4.1).

C20orf201, *SEPT12* and *SYCP1* were activated in HCT116 cells treated with the 5-AZA-CdR inhibitor with faint band expression (Figure 4.13). *C20orf201* was expressed with moderate band intensity for most of the treatment doses after 48 hours but gene expression was noted at low doses (0.1, 0.5 and 1.0 µM) after incubation for 72 hours. These data suggest that using high concentrations of 5-AZA-CdR for 72 hours may prevent the expression of *C20orf201*. However, *SEPT12* and *SYCP1* were activated with 0.5 to 15 µM of 5-AZA-CdR at 48 or 72 hours.

The *C20orf201*, *SEPT12* and *SYCP1* genes were analysed in treated HCT116 cells with TSA inhibitor using RT-PCR (Figure 4.14). No significant expression was evident for *C20orf201* and *SYCP1* after treatment but *SEPT12* showed some expression in cells treated with 600 nM TSA for 24 hours.

In addition, *SEPT12* and *SYCP1* showed some activation in the co-treated HCT116 cells whereas, no expression of *C20orf201* was observed at all concentration (Figure 4.15). The data suggest that co-treatment of HCT116 cells with 5-AZA-CdR and TSA prevented the expression of *C20orf201* and *SYCP1* which were activated by 5-AZA-CdR treatment only. *SEPT12* showed faint expression after treatment with 1 and 5 µM 5-AZA-CdR in combination with 300 nM TSA. The PCR products of these genes were sequenced to ensure the correct gene had been amplified and the results of the sequencing are shown in Table 4.2.
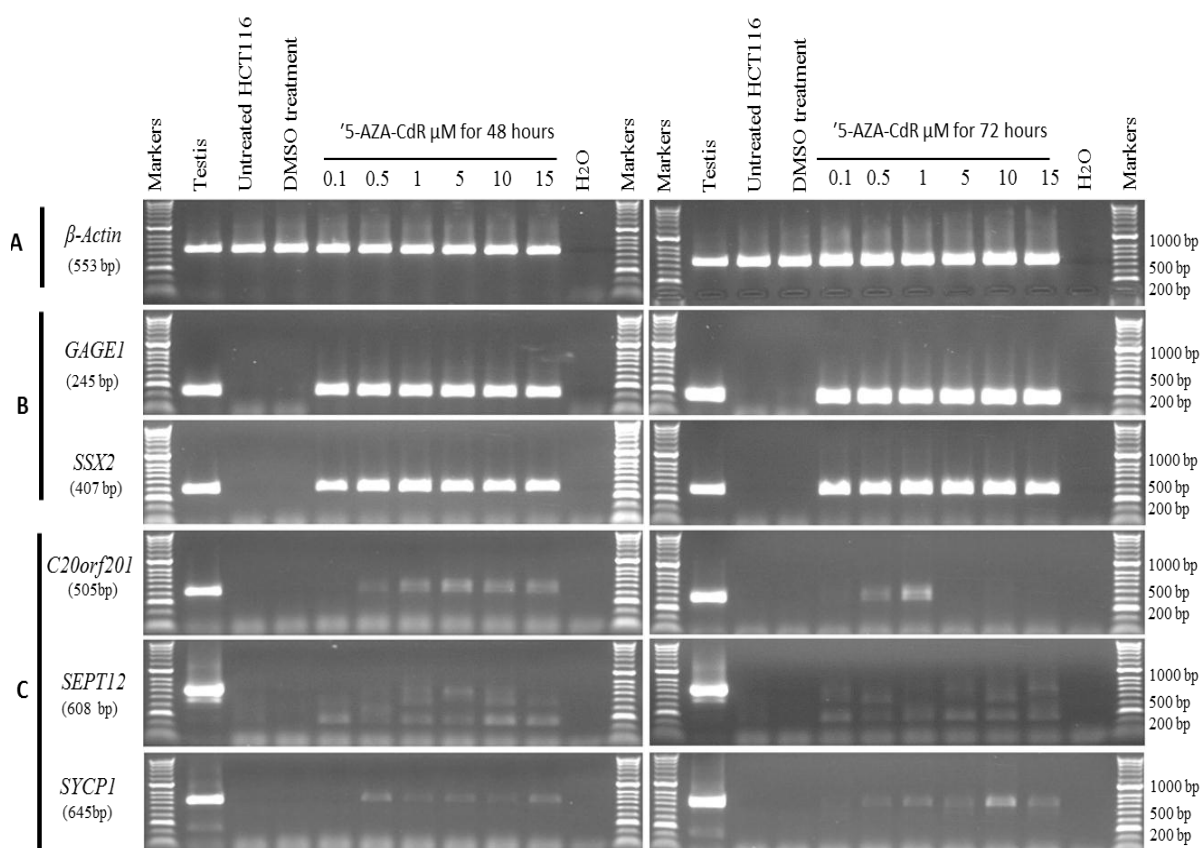
Figure 4.13. **The effect of 5-AZA-CdR on the cancer/testis-CNS restricted genes expression in HCT116 cancer cell line.** (A) *β-Actin* served as positive control for the cDNA. (B) *GAGE1* and *SSX2* served as control CTA genes. (C) The expression of *C20orf201*, *SEPT12* and *SYCP1* after treatment of HCT116 cells with 5-AZA-CdR for 48 hours (on the left) or 72 hours (on the right).
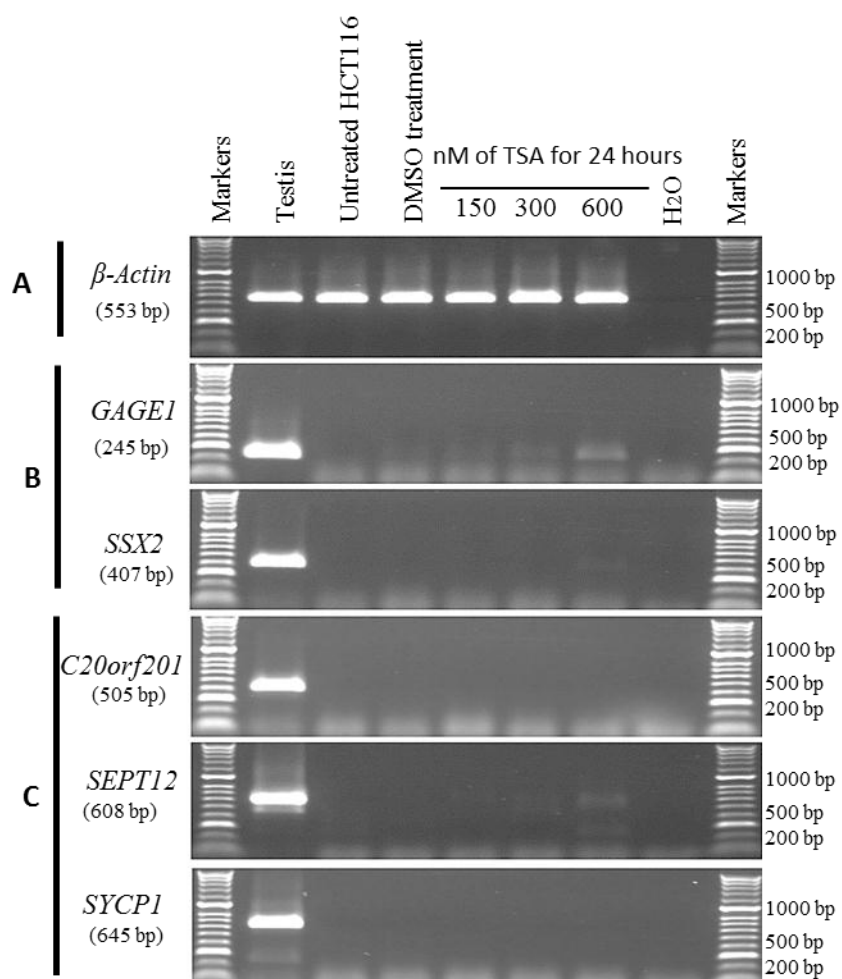
Figure 4.14. The effect of TSA on the cancer/testis-CNS restricted genes expression in HCT116 cancer cell line. (A) *β-Actin* served as a positive control for the cDNA. (B) *GAGE1* and *SSX2* served as control CTA genes (C) Treatment with TSA showed no effect on gene expression of *C20orf201* and *SYCP1* but the *SEPT12* gene showed a slight expression following treatment with 600 nM TSA.
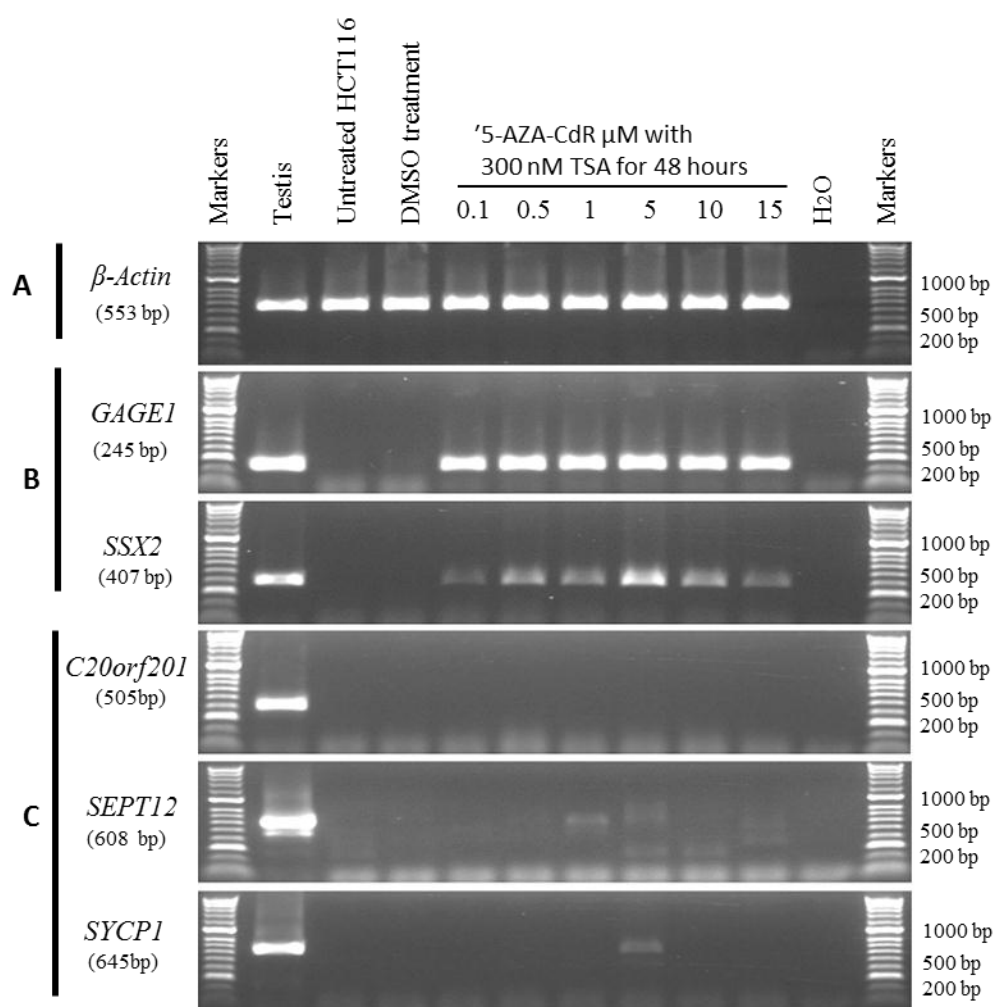
Figure 4.15. **The effect of co-treatment with 5-AZA-CdR and TSA on cancer/testis-restricted genes expression in the HCT116 cancer cell line.** (A) *β-Actin* served as a positive control for the cDNA. (B) *GAGE1* and *SSX2* served as control CTA genes. (C) *C20orf201* and *SYCP1* showed no expression in HCT116 cells co-treated with 5-AZA-CdR and TSA while *SEPT12* was induced with weak expression.

4.2.6 The influence of 5-AZA-CdR and/or TSA inhibitors on the expression of cancer/testis-CNS restricted genes in SW480 cancer cells

SW480 cells were treated with 5-AZA-CdR and/or TSA to investigate if *C20orf201*, *SEPT12* and *SYCP1* genes might have similar behaviour to that seen in HCT116 cells treated with epigenetic inhibitors. *SEPT12* was expressed in SW480 cells and this gene was used to investigate if the epigenetic treatments could increase the expression of this gene.

The RT-PCR profile of *C20orf201*, *SEPT12* and *SYCP1* in SW480 cells treated with 5-AZA-CdR for 48 or 72 hours showed no significant activation of these genes and no increased band intensity for *SEPT12* (Figure 4.16), although these genes had been activated in HCT116 cells treated with 5-AZA-CdR.

No significant expression of *C20orf201* or *SYCP1* was obtained for SW480 cells treated with TSA. *SEPT12* showed a slight increase in band intensity with a 600 nM TSA treatment (Figure 4.17). *SEPT12* was also expressed with faint bands after treatment of HCT116 cells with 600 nM TSA. These data suggest that TSA treatment of HCT116 and SW480 cells could not induce the expression of *C20orf201* or *SYCP1*.

Co-treatment of SW480 cells with 5-AZA-CdR and TSA also had no effect on the expression of *C20orf201*, and *SYCP1*. These genes were also not activated by co-treatment in HCT116 cells. On the other hand, a slight increase in the *SEPT12* band intensity was noted after co-treatment particularly at high concentrations, which was also observed for co-treated HCT116 cells (Figure 4.18). The PCR products of these genes were purified and sequenced to confirm that the correct DNA sequence was amplified (summary of sequencing, Table 4.2).
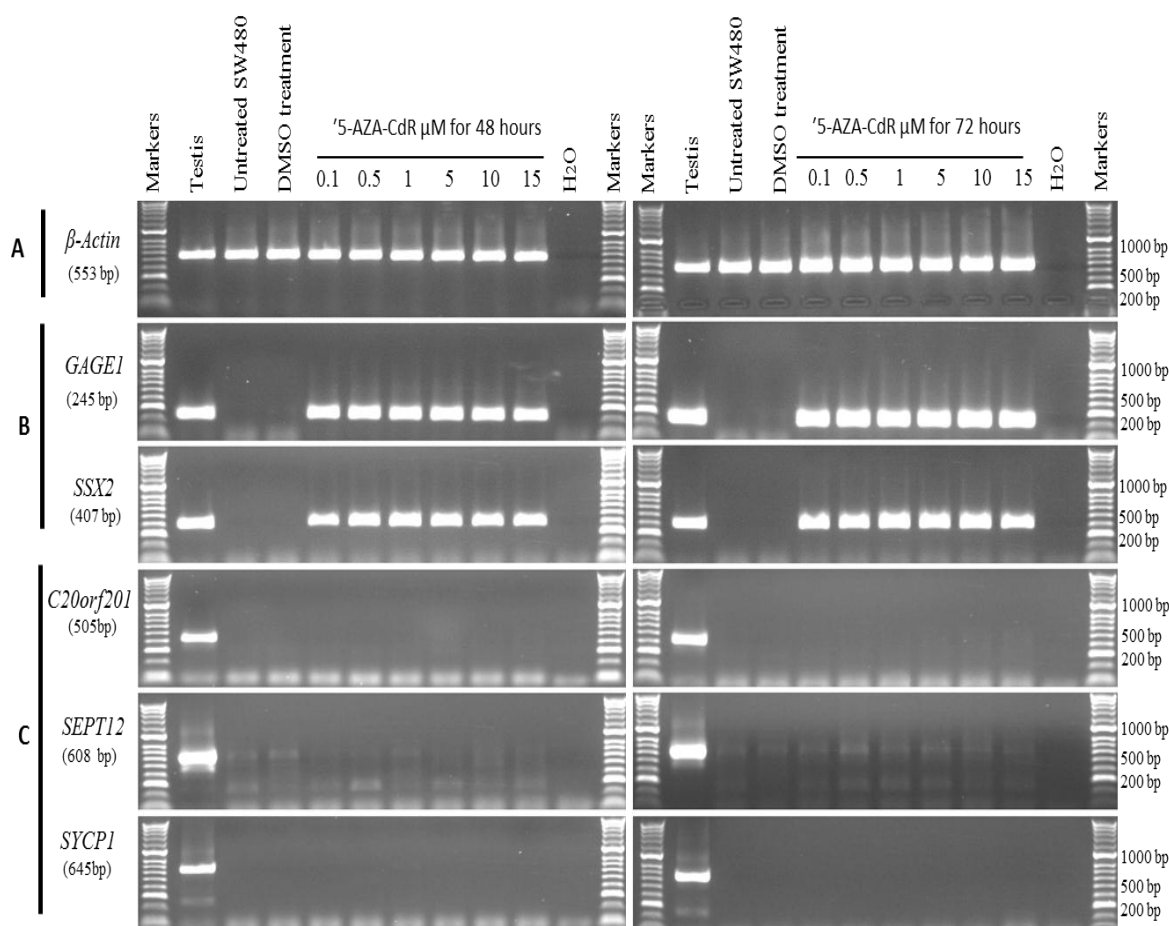
Figure 4.16. **The effect of 5-AZA-CdR on the cancer/testis-CNS restricted genes expression in the SW480 cancer cell line.** (A) *β-Actin* served as a positive control of the cDNA. (B) *GAGE1* and *SSX2* served as control CTA genes. (C) Treatment of SW480 cells with 5-AZA-CdR resulted in no significant activation of *C20orf201, SEPT12* and *SYCP1* gene expression.
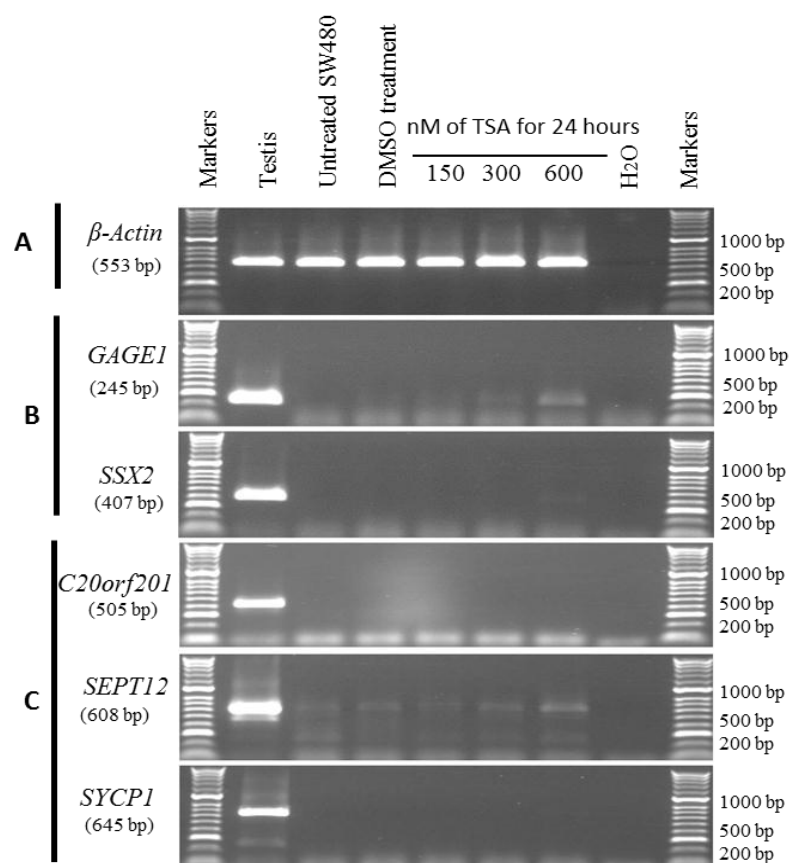
Figure 4.17. **The effect of TSA on the cancer/testis-CNS restricted genes expression in the SW480 cancer cell line.** (A) *β-Actin* served as positive control of the cDNA. (B) *GAGE1* and *SSX2* served as control CTA genes. (C) Treatment with TSA showed no effect on gene expression of *C20orf201* and *SYCP1.* The *SEPT12* gene showed a slight expression after treatment with 600 nM TSA but it is expressed even in untreated cells.
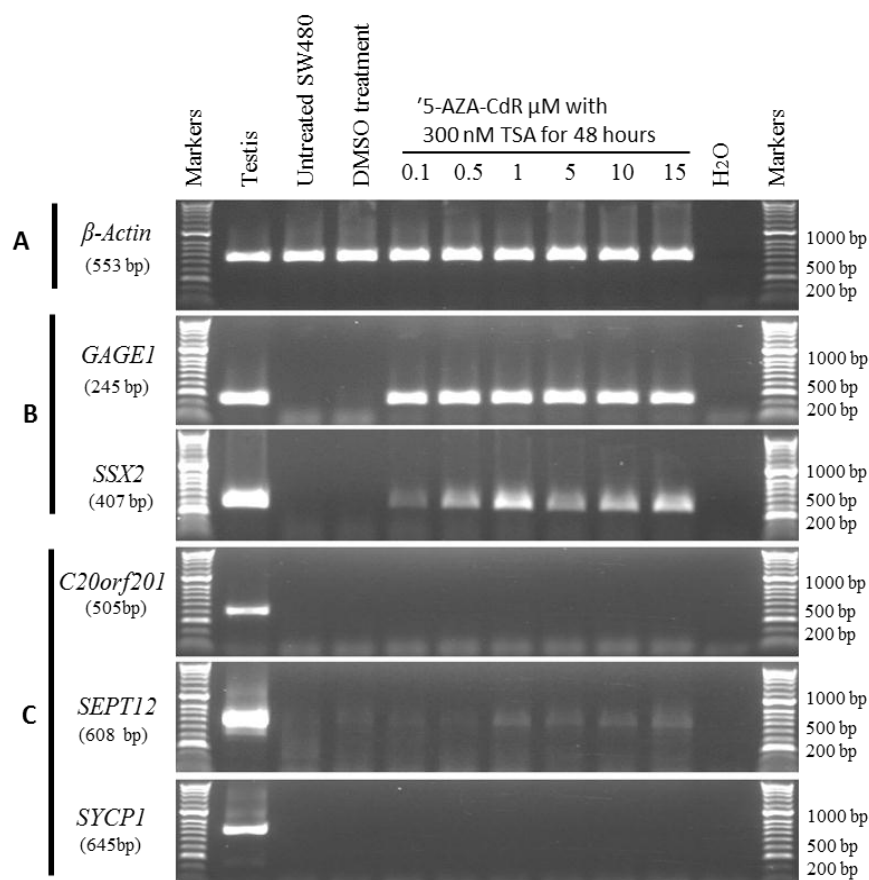
Figure 4.18. **The effect of co-treatment of 5-AZA-CdR with TSA on cancer/testis-CNS-restricted genes expression in SW480 cancer cell line.** (A) *β-Actin* served as a positive control of the cDNA. (B) *GAGE1* and *SSX2* served as control CTA genes. (C) Co-treatment of SW480 with 5-AZA-CdR and TSA induced no significant expression of *C20orf201* and *SYCP1* but a slight increase in *SEPT12* expression was noted.

4.2.7 The expression of CT genes after growth of HCT116 cells in drug free culture.

Two mechanisms were proposed to regulate the expression of CTA genes in cancer cells: DNA methylation and histone modification. The DNA methylation inhibitor 5-AZA-CdR was shown to activate the expression of several silenced CT genes whereas the histone deacetylase inhibitor TSA had little influence on the CT gene expression. Co-treatment of cancer cells with 5-AZA-CdR and TSA up-regulated the expression of CT genes more than TSA alone (reviewed in De Smet and Loriot, 2013; Whitehurst, 2014).

In this chapter, *DDX4* testis-restricted and *C20orf201* CT/CNS-restricted genes were induced after HCT116 cells were treated with 5-AZA-CdR. The *STRA8* and *TDRD12* CTA-restricted genes also showed a significant up-regulation after HCT116 cells were treated with 5-AZA-CdR alone or combined with TSA. Furthermore, *STRA8* was influenced by TSA inhibitor treatment in both HCT116 and SW480 cells. We next asked whether the expression pattern of these genes could still be observed in the HCT116 cell line 9 days after treatment removal. Here, HCT116 and SW480 cells were treated with 300 nM TSA for 24 hours. In addition, HCT116 cells were treated with 1 µM of 5-AZA-CdR for 48 and 72 hours alone or in combination with 300 nM TSA for 48 hours. The effects of 5-AZA-CdR and/or TSA and their persistence of the expression of CTA genes was examined by washing the treated cells twice with PBS and culturing them in drug-free medium for 3, 6 and 9 days. Total RNA was isolated from untreated, DMSO treated, and 5-AZA-CdR and/or TSA treated as well as drug-free treated cells after 3, 6 and 9 days.

The RT-PCR profile of *STRA8* indicated that it was silenced in untreated HCT116 and SW480 cells. The expression of *STRA8* was activated after treatment of HCT116 and SW480 cells with 300 nM TSA for 24 hours. *STRA8* expression disappeared after 3 days of cell culture in drug-free medium (Figure 4.19). These data suggest that the loss of *STRA8* expression in both cancer cell types may be due to histone acetylation.
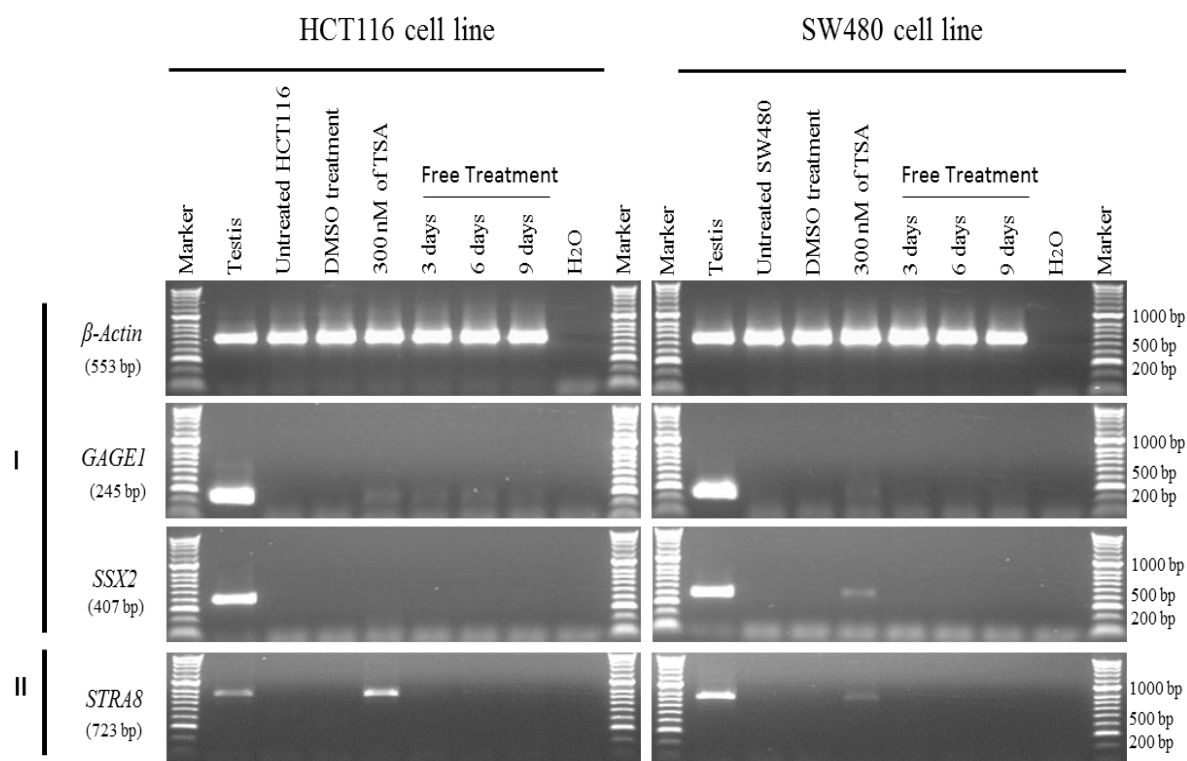
Figure 4.19. **The effect of TSA and the persistence of its effects on expression of _STRA8_ in colorectal cancer cell lines.** Two cancer cell lines, HCT116 (on the left) and SW480 (on the right) were treated with 300 nM trichostatin (TSA) for 24 hours. Cells were then washed twice with PBS and cultured in drug-free medium for 3, 6 and 9 days. RNA was then isolated (from untreated, DMSO treated, 300 nM TSA treated, and drug-free treated cells after 3, 6, and 9 days). RT-PCR expression profiles of STRA8 were visualised in 1% agarose gels stained with ethidium bromide. (I) Expression of _β-Actin_ served as a positive control for the cDNA and _GAGE1_ and _SSX2_ served as control CTA genes. (II) The expression _STRA8_ disappeared after 3 days of drug-free growth.

However, treatment of HCT116 cells with 5-AZA-CdR induced the expression of 8 genes in this study. Only four genes *DDX4*, *TDRD12*, *STRA8*, *C20orf201* and control genes *GAGE1* as well as *SSX2* were validated in HCT116 cells treated with 1 µM 5-AZA-CdR for 48 or 72 hours. These genes were also analysed in HCT116 cells after 3, 6 and 9 days of drug free growth, as well as in cells treated with DMSO as it was the solvent for 5-AZA-CdR (Figure 4.20). *STRA8*, *GAGE1* and *SSX2* were still being expressed after 3, 6 and 9 days of drug-free growth. These data suggested that the transcriptional status of *STRA8*, *GAGE1* and *SSX2* may have been permanently reversed. Support for this idea comes from a study by Mossman et al., (2010) who found that *MAGE-A3* was still expressed after 10 days of drug-free growth (Mossman *et al*., 2010). *TDRD12* expression was observed at days 3, 6 and 9 of drug-free growth but the expression at day 9 was lower than in the treated cells. In addition, the expression of *DDX4* and *C20orf201* were induced after cell treatment with 5-AZA-CdR for 48 or 72 hours. This activation disappeared after 6 and 9 days of drug-free growth. This result suggested that global DNA methylation can lead to gene silencing, which was supported by the findings of Mossman et al., (2010), who reported that treatment of HCT116 and SW480 cells with 15 µM 5-AZA-CdR for 72 hours led to a more than 50% decrease in the global methylation and at day 10 of drug-free growth, the level of global methylation increased to reach the level before treatment (Mossman *et al.*, 2010).

The HCT116 cells were also co-treated with 1 µM 5-AZA-CdR and 300 nM TSA to investigate the influence of the histone deacetylation in combination with DNA demethylation and the persistence of *TDRD12* and *STRA8* expression (Figure 4.20). *TDRD12* expression was displayed at day 3, 6 and 9 of drug-free growth. The expression was decreased after 6 and 9 days of drug-free treatment. By contrast, *STRA8* was still expressed at 3 and 6 days and had disappeared at day 9 of drug-free treatment. Both control genes *GAGE1* and *SSX2* were activated after 3, 6 and 9 days of drug-free growth, with a minor decrease in *SSX2* expression at day 9. The PCR products of these genes were purified and sequenced to confirm that the correct DNA sequence was amplified (summary of sequencing, Table 4.2).
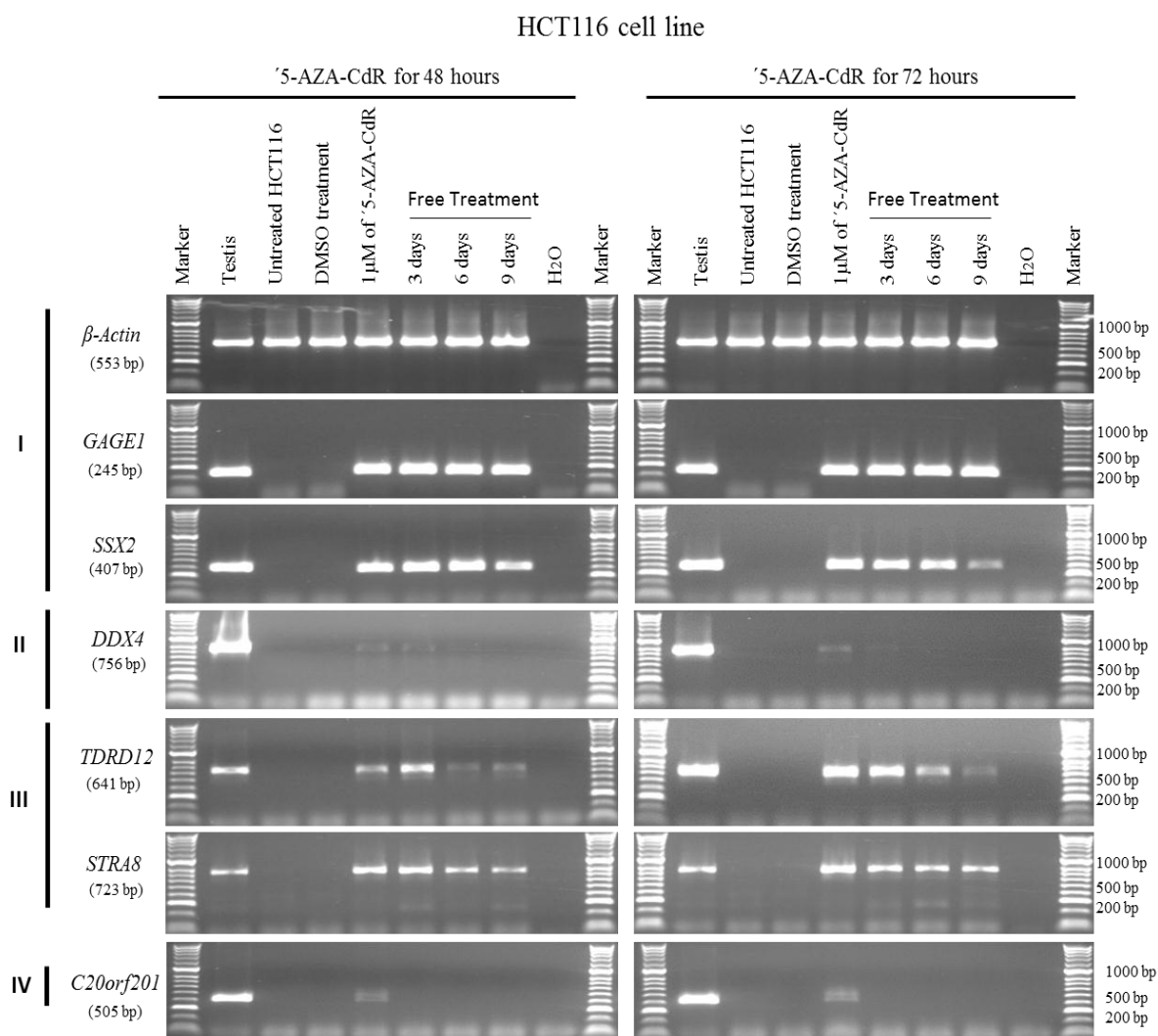
Figure 4.20. **The effect of 5-aza 2' deoxycytidine and the persistence of expression of potential novel CT genes in the HCT116 cancer cell line.** HCT116 cells were treated with 1 µM 5-aza 2' deoxycytidine for 48 hours (on the left) or 72 hour (on the right). Cells were then washed twice with PBS and cultured in drug-free medium for 3, 6 and 9 days. RNA was isolated (from untreated, DMSO treated, 1 µM 5-aza 2' deoxycytidine treated, and drug-free treated cells after 3, 6 and 9 days). RT-PCR expression profiles of each gene are visualised in 1% of agarose gel stained with ethidium bromide. (I) Expression of the *β-Actin* gene served as a positive control for the cDNA ; whereas, *GAGE1* and *SSX2* served as control CTA genes.(II) The expression profile of testis-restricted genes *DDX4* (III) The expression profile of cancer/testis restricted genes; *TDRD12* , *STRA8.* (IV) The expression profile of the cancer/testis-CNS restricted *C20orf201* gene.
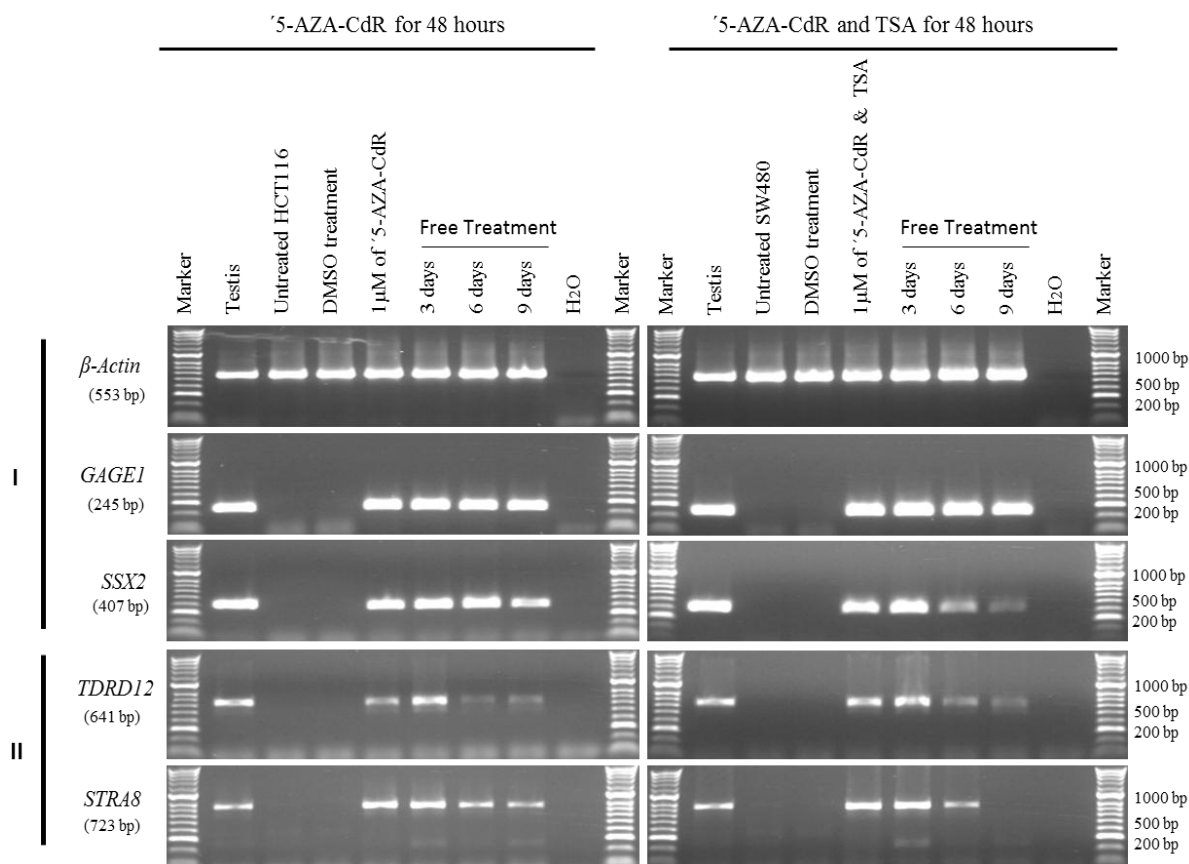
Figure 4.21. **Comparison of the effects of 5-aza 2′ deoxycytidine and/or a combination with TSA and the persistence of *TDRD12* and *STRA8* expression in the HCT116 cancer cell line.** HCT116 cells were treated with 1 µM 5-aza 2′ deoxycytidine (on the left) or in combination with 300 nM TSA (on the right) for 48 hours. Cells were then washed twice with PBS and cultured in drug-free medium for 3, 6 and 9 days. RNA was isolated (from untreated, DMSO treated, drugs treated, and drug-free treated cells after 3, 6, and 9 days). RT-PCR expression profiles of each gene was visualised in 1% agarose gel stained with ethidium bromide. (I) Expression of *β-Actin* served as a positive control for the cDNA, whereas, *GAGE1* and *SSX2* served as control CTA genes.(II) The expression profile of cancer/testis restricted genes; *TDRD12* , *STRA8.*

Table 4.2 **Summary of gene sequencing results after RT-PCR validation in HCT116 and SW480 cells treated with epigenetic inhibitors.**

Gene name	Primer	Expected Size (bp)	Approximately Sequenced product (bp)	Treated colon cancer cells or normal testis	Sequence identity obtained from NCBI blast (%)
Control genes					
β-Actin	F	553	405	Testis	99%
		553	380	HCT116	100%
GAGE1	F	245	208	Testis	99%
		245	122	HCT116 (5-AZA-CdR)	100%
		245	203	SW480 (5-AZA-CdR)	99%
SSX2	F	407	372	Testis	99%
		407	346	HCT116 (5-AZA-CdR)	100%
		407	365	SW480 (5-AZA-CdR)	99%
Class 1: testis-restricted genes					
ARRDC5	F	628	521	Testis	99%
C4orf17	F	628	585	Testis	99%
DDX4	F	756	500	Testis	100%
		756	387	HCT116 (5-AZA-CdR)	100%
IQCF3	F	356	303	Testis	99%
	F	356	314	HCT116 (5-AZA-CdR & TSA)	100%
NT5C1B	F	647	601	Testis	100%
TMEM202	F	629	447	Testis	100%
Class 2: cancer/testis-restricted genes					
C17orf98	F	394	354	Testis	100%
		394	311	HCT116 (5-AZA-CdR)	100%
ODF4	F	602	258	Testis	99%
PRDM7	F	602	201	Testis	100% *PRDM7* 97% *PRDM9*
PRDM9	F	655	604	Testis	99% *PRDM9* 97% *PRDM7*
		655	536	SW480 (5-AZA-CdR & TSA)	100% *PRDM9* 97% *PRDM7*
SMC1β	F	368	332	Testis	99%
		368	316	HCT116 (5-AZA-CdR)	100%
STRA8	F	723	121	HCT116 (5-AZA-CdR)	100%
		723	153	SW480 (5-AZA-CdR)	99%

TDRD12	F	641	308	Testis	100%
		641	371	HCT116 with 5-AZA-CdR	99%
	F	641	141	SW480 with 5-AZA-CdR	100%
Class 3: cancer/testis-CNS restricted genes					
C20orf201	F	505	373	Testis	99%
		505	298	HCT116 (5-AZA-CdR)	100%
SEPT12	F	608	553	Testis	99%
		608	351	HCT116 (5-AZA-CdR & TSA)	99%
SYCP1	F	645	260	Testis	100%
		645	125	HCT116 with 5-AZA-CdR	100%

4.3   Discussion and Conclusions

Epigenetic mechanisms play an important role in the regulation of the expression of CTA genes via DNA methylation and histone modifications to maintain the transcriptional state. These two features, and particularly global DNA hypomethylation, commonly disrupt the expression of CTA genes in cancer cells. (De Smet *et al.*, 1999; De Smet and Loriot, 2013; Whitehurst, 2014). The reversible nature of the epigenetic alteration has led to development of epigenetic drugs which are being tested as therapeutic agents against cancers (Sharma *et al.*, 2010). The use of epigenetic drugs such as the DNA hypomethylation inhibitor 5-AZA-CdR can increase the number of CTA genes expressed in tumours, which can be recognised as non-self by the immune system and thereby have application in cancer immunotherapy (De Smet and Loriot, 2013). In this chapter, we investigated the effect of DNA demethylation and/or histone deacetylation on the expression of 16 candidate CT genes. In addition, the influence of "DNA re-methylation" was examined by the growing treated cells in drug-free medium for up to 9 days.

4.3.1   Predicted CpG islands upstream of CTA genes

Treatment of HCT116 and SW480 cells with 5-AZA-CdR and/or TSA appears to induce the expression of 10 of the 16 candidate CTA genes. Two of the 6 genes from class 1 testis-restricted genes; 5 of 7 genes from class 2 cancer/testis-restricted genes; and 3 of 3 genes from class 3 cancer/testis-CNS restricted genes were activated after cell treatment with epigenetic inhibitors (Figure 4.22).

The sensitivity of CTA genes to epigenetic modulator inhibitors, and particularly methylation drugs, might be due to their CpG-rich islands which are frequently found in their promoter regions (De Smet *et al.*, 1999; De Smet and Loriot, 2013). In contrast, genes with a low density of CpG islands in their promoter regions are little effected by DNA methylation which frequently displays an inconstant relationship between transcriptional silencing and promoter methylation (Boyes and Bird, 1992; De Smet and Loriot, 2013).

CpG sites are genomic regions of DNA at least 200 bp long and containing greater than 50% of G:C content and an observed/expected ratio of CpG greater than 0.60 (Gardiner-Garden and Frommer, 1987). Takai and Jones (2002) defined CpG islands as genomic DNA greater

than 500 bp long and containing at least 55% G:C content and with an observed/expected ratio of more than 0.65 (Takai and Jones, 2002).   Several websites and software programs were designed following these criteria to predict the CpG sites which are usually found near the transcription start site (TSS) (± 1500 bp from TSS). The UCSC Genome Browser bioinformatics website (https://genome.ucsc.edu/; Gardiner-Garden and Frommer, 1987) was used to predict the CpG islands of the 16 CTA genes in this study. Six out of 16 genes showed rich CpG islands, whereas 10 genes did not show any CpG islands based on the Gardiner-Garden and Frommer (1987) criteria or are still not documented (Table 4.3.).

4.3.1.1 CpG island prediction and epigenetic control of testis-restricted gene expression

Different expression patterns were observed for CTA genes from the same categories after the epigenetic inhibitor treatment. *DDX4* was expressed giving moderate RT-PCR band intensity in HCT116 and SW480 cells treated with 5-AZA-CdR (Figure 4.22). The data shown in Table 4.3 suggest that the induction of *DDX4* expression may due to the CpG islands upstream of the *DDX4* promoter region with 299 bp size and containing 30 CpG dinucleotides.

By contrast, *IQCF3* may be slightly affected by treatment of HCT116 and SW480 cells with the 5-AZA-CdR inhibitor as it shows very weak expression band intensity. The intensity of the bands was increased by co-treatment of the cells with 5-AZA-CdR and TSA. *IQCF3* showed different behaviour by its expression in SW480 cells and lack of expression in HCT116 cells after treatment with TSA under the same conditions (Figure 4.22). Searching for CpG islands upstream of *IQCF3* using the Gardiner-Garden and Frommer, (1987) criteria did not reveal any CpG islands. A study by Suzuki et al.(2002) using a microarray approach showed that treatment of colon cancer cells with TSA inhibitor alone induced the expression of a class of genes that lack CpG island methylation. In contrast, densely methylated genes could not be activated with only TSA treatment but required the AZA-CdR pre-treatment (Suzuki *et al.*, 2002).

Expression of 4 genes, *ARRDC5*, *C14orf17*, *NT5C1B* and *TMEM202,* was not induced after the cell treatments. Analysis of the TSS upstream of these genes showed only *ARRDC5* has a rich

CpG island of approximately 234 bp size and 18 CpG dinucleotides (Table 4.3). However, *ARRDC5* did not show re-expression after DNA hypomethylation and/or histone deacetylation, which suggested that perhaps another mechanism is controlling of the expression of these genes. However, a great effect was seen in the gene expression of the positive CT genes *GAGE1* and *SSX2*. *GAGE1* showed strong expression following treatment with AZA-CdR alone or with TSA in both HCT116 and SW480 cells. The *SSX2* gene also showed strong expression after AZA-CdR treatment in both cell types but co-treatment with TSA decreased their expression level. Both of these genes are located on chromosome X, which is known to be inactivated in females through epigenetic regulation (Gartler and Goldman, 2005). The activation of *GAGE1* and *SSX2* is not surprising as several CTA genes, such as *MAGE-A3* located in chromosome X, are activated through the DNA methylation due to their CpG-rich promoter regions (De Smet *et al*., 1999). The CpG island map predicted 34 CpG dinucleotides near the TSS of *GAGE1* of about 485 bp in length. The CpG island of *SSX2* seems not be documented in UCSC Genome Browser and other websites (Smith *et al*., 2011).

4.3.1.2 CpG island prediction and epigenetic control of cancer/testis-restricted gene expression

Five of 7 genes from class 2 cancer/testis-restricted genes showed expression following epigenetic inhibitor treatment. Two genes, *ODF4* and *PRDM7,* were previously found to be expressed in more than one cancer type (Table 4.1). Treatment of HCT116 and SW480 cells with 5-AZA-CdR and/or TSA did not activate the expression of these genes, with the exception of *PRDM7* which is expressed in SW480 cells normally. Based on the USCU CpG islands Browser search results, *ODF4* and *PRDM7* did not appear to have CpG islands upstream of their TSS using Gardiner-Garden and Former (1987) criteria. This may suggest that these genes are poor in CpG islands or that the CpG islands of these genes are still not documented. However, the absence of expression of these genes after treatment of HCT116 and SW480 cells may indicate that these genes lack rich CpG islands and their aberrant expression in cancer cells may due to unknown mechanisms (Figure 4.22).

The expressions of the *C17orf98*, *SMC1β* and *PRDM9* genes were also previously reported in several types of cancers (Table 4.1) (Feichtinger *et al*., 2012a). Treatment of HCT116 and

SW480 cells with 5-AZA-CdR or co-treatment with TSA activated the expression of *C17orf98* and *SMC1β* with moderate band intensity. *PRDM9* was expressed only after SW480 cells were treated with 5-AZA-CdR and displayed very weak band intensity (Figure 4.22). A search for CpG dinucleotides near the TSS of *C17orf98*, *SMC1β* and *PRDM9* did not predict rich CpG islands. These data may suggest that the CpG islands of these genes are not yet documented or the poor expression is due to the low density of CpG islands in their promoter regions. It has been proposed that CTA genes with poor CpG regions are little effected by DNA methylation which frequently displays an inconstant relationship between transcriptional silencing and promoter methylation (Boyes and Bird, 1992; De Smet and Loriot, 2013).

STRA8 and *TDRD12* displayed expression with strong band intensity in HCT116 cells treated with 5-AZA-CdR or with TSA, whereas SW480 cells treated with 5-AZA-CdR or with TSA led to strong induction of *TDRD12* and moderate activation of *STRA8*. *STRA8* expression was also affected by TSA treatment alone in both HCT116 and SW480 cells, while *TDRD12* remained silenced (Figure 4.22).   Searching for CpG islands upstream of the *TDRD12* and *STRA8* promoter regions showed that two CpG islands are predicted in the TSS of both *TDRD12* and *STRA8* Table 4.3. In *TDRD12*, a rich CpG island 586 bp in size containing 56 CpG dinucleotides was predicted and the other was 299 bp long and contained 28 CpG dinucleotides. *STRA8,* in contrast, has two CpG islands 335 and 451 bp long containing 28 and 42 CpG dinucleotides, respectively. This observation may indicate clear evidence that DNA hypomethylation leads to activation of *TDRD12* and *STRA8* because of their CpG-rich islands. However, to address whether these genes are activated due to DNA hypomethylation, methylation-specific PCR (MSP) should be performed to assess the methylation status of CpG islands (Herman *et al*., 1996).

4.3.1.3 CpG island prediction and epigenetic control of cancer/testis-CNS restricted gene expression

The expression of *C20orf201*, *SEPT12* and *SYCP1* was activated with moderate band intensity after treatment of HCT116 or SW480 cells with epigenetic inhibitors (Figure 4.22). *C20orf201* was expressed in HCT116 cells treated with 5-AZA-CdR alone. This gene expression was not observed following co-treatment of HCT116 cells or with TSA treatment alone. No expression

of this gene was seen in SW480 cells treated with 5-AZA-CdR and/or TSA. *C20orf201* has a rich CpG island 998 bp long and containing 123 CpG dinucleotides near to the TSS (Table 4.3). However, this gene consists of rich CpG dinucleotides in the promoter region and was not activated in treated SW480 cells, which suggested that another mechanism controls the expression of the gene. Expression of *SYCP1* was induced in HCT116 cells treated with 5-AZA-CdR only and no expression was observed in treated SW480 cells. The CpG island map predicted 37 CpG dinucleotides at the *SYCP1* TSS with an approximate size of 494 bp. The lack of expression of *SYCP1* in treated SW480 and the re-expression in HCT116 after DNA hypomethylation treatment suggest that perhaps this control occurs via another mechanism in SW480 cells. The *SYCP1* gene seems to have a CpG-rich region in the TSS, more investigation of this gene is needed. *SEPT12* was also expressed in HCT116 cells after treatment with 5-AZA-CdR and the gene was expressed in untreated SW480 cells. Searching for the CpG dinucleotides at the TSS of *SEPT12* did not show any CpG-rich islands.

4.3.2 The influence of epigenetic inhibitors on CTA gene expression after 9 days of drug-free growth

STRA8 was the only gene expressed with moderate band intensity after treatment of HCT116 and SW480 cells with TSA drugs. This gene expression was no longer observed after 3 days of drug-free growth (Figure 4.19). The TSA inhibitor is known as an anti-cancer drug which causes accumulation of acetylated histones through inhibition of histone deacetylase (Yoshida *et al.*, 1990; Mossman *et al.*, 2010). The process of histone acetylation followed by TSA on the CTA genes is not well documented. However, it was reported to have a synergistic effect with 5-AZA-CdR in activation of silenced gene expression in cancer cells (Cameron *et al.*, 1999; Mossman *et al.*, 2010).

Based on the quantitative PCR of HCT116 and SW480 cells treated with 5-AZA-CdR followed by culture in drug-free medium for 10 days, Mossman and Scott, (2011) classified the gene expression to different classes. (i) Long term reactivated (which show no expression in untreated but expression after 4 and 10 days) ; (ii) short term reactivated (which are not expressed after 4 days); and (iii) always-expressed (which are expressed at all-time points) (Mossman and Scott, 2011). Here, *TDRD12*, *STRA8* and the control genes *GAGE1* and *SSX2*

were expressed 9 days after the treatment with 5-AZA-CdR. Based on their expression in HCT116 cells treated with 5-AZA-CdR, these genes can be classified as long term reactivated genes. However, *DDX4* and *C20orf201* showed no expression after 3 days of treatment and can be classified as short term reactivated genes (Figure 4.20). In addition, co-treatment with 5-AZA-CdR and TSA resulted in faint expression of *TDRD12* at day 9 of the drug-free treatment whereas *STRA8* had disappeared by day 6. *GAGE1* was always expressed in co-treatment while the expression level of *SSX2* was reduced after 6 and 9 days of drug-free growth (Figure 4.23).

In summary, our results showed that some CTA genes were activated after 5-AZA-CdR treatment and others were not. Activation does not always rely on DNA hypomethylation, so methylation-specific PCR (MSP) and/or bisulphite sequencing PCR (BSP) are required to confirm this finding. MSP and BSP are designed to distinguish between the methylated and unmethylated CpG sites. In addition, HCT116 cells have been reported to harbour epigenetic alterations and one of the colon cancer lines most affected by 5-AZA-CdR treatment, which may explain the expression of several CTA genes in that cell line (Mossman *et al.*, 2010; Ikehata *et al.*, 2012) . Further investigations of these genes are required in other cancer cell lines to confirm this result and quantitative qPCR should be used to determine the effect of co-treatment on gene expression.
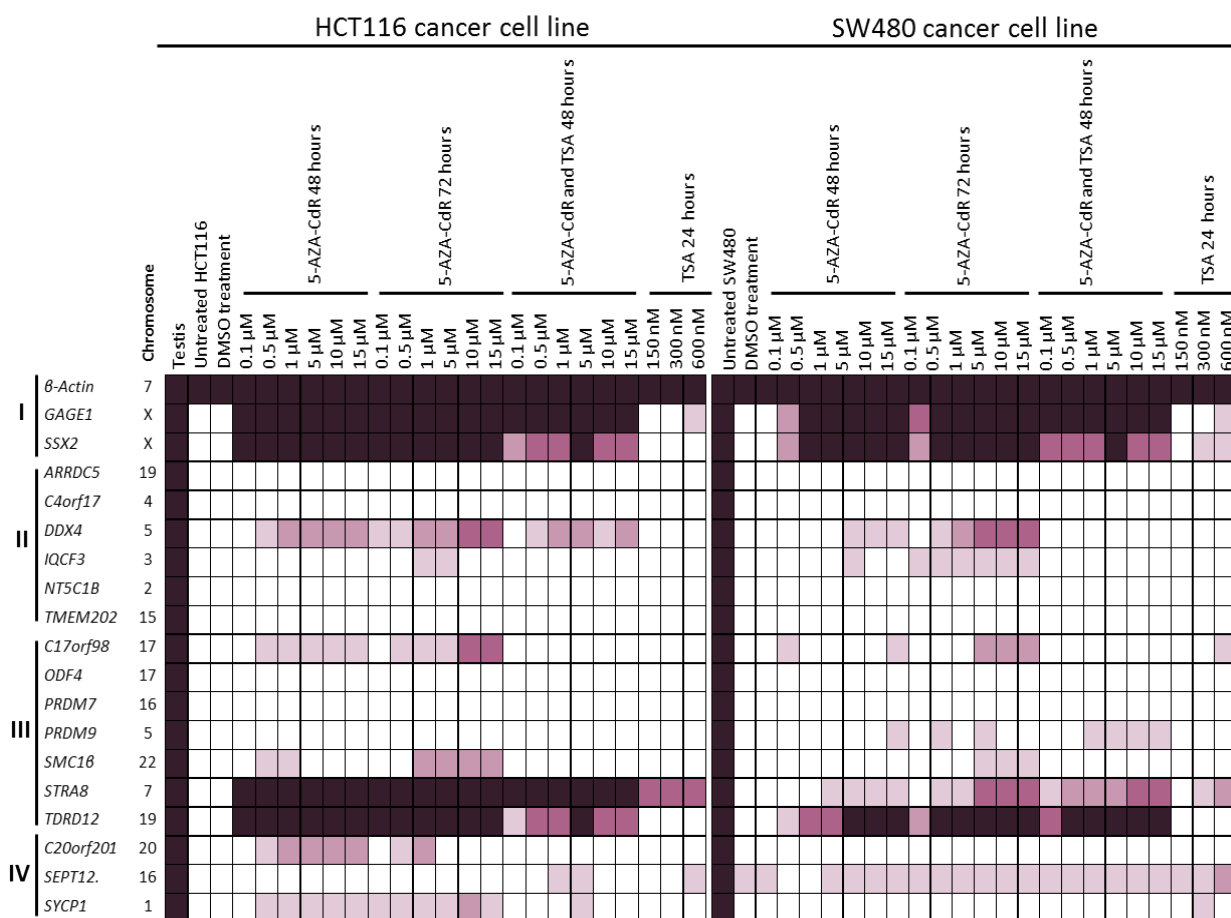
Figure 4.22. Grid representation of the effect of 5-AZA-CdR and/or TSA on potential novel CT genes in colorectal cancer cell lines. Two cancer cell lines, HCT116 (on the left) and SW480 (on the right) were treated with different concentrations of 5-AZA-CdR and/or TSA. Untreated and DMSO treated cells were used to assess the effect of treatment on gene expression. In the grid, each column represents the concentration of treatment and the incubation time, while gene expression is represented in the rows. RT-PCR expression profiles of each gene is visualised in 1% of agarose gel stained with ethidium bromide. The intensity of the bands are graded with different shades of purple colour (see the key at the top left) (+) is weak intensity band and (+++++) is strong intensity band. In the grid, (I) Expression of β-*Actin* gene shows the positive control of the cDNA samples, and the profile expression of previously known CT genes *GAGE1* and *SSX2*.(II) The expression profile of testis-restricted genes; *ARRDC5*, *C4orf17*, *DDX4*, *IQCF3*, *NT5C1B* and *TMEM202*. (III) The expression profile of cancer/testis restricted genes; *C17orf98*, *ODF4*, *PRDM7*, *PRDM9*, *SMC1β*, *STRA8* and *TDRD12*. (IV) The expression of testis/CNS restricted genes; *C20orf201*, *SEPT12* and *SYCP1*.

Table 4.3. Predicted CpG islands upstream of the potential novel CT antigen genes.

Gene	Chromosome location	CpG island length (bp)	CpG count	% CpG	%C or G	Ratio of observed to expected
Positive control gene						
GAGE1	Xp11.23	485	34	14.0	67.6	0.63
Class 1: testis-restricted genes						
ARRDC5	19p13.3	234	18	15.4	62.4	0.80
DDX4	5q11.2	299	30	20.1	63.5	1.01
Class 2 : cancer/testis-restricted genes						
STRA8	7q33	335	28	16.7	65.1	0.81
		451	42	18.6	68.7	0.79
TDRD12	19q13.11	586	56	19.1	72.7	0.73
		299	28	18.7	70.2	0.76
Class 3 : cancer/testis-CNS-restricted genes						
C20orf201	20q13.33	998	123	24.6	72.7	0.94
SYCP1	1p13.2	494	37	15.0	59.3	0.86
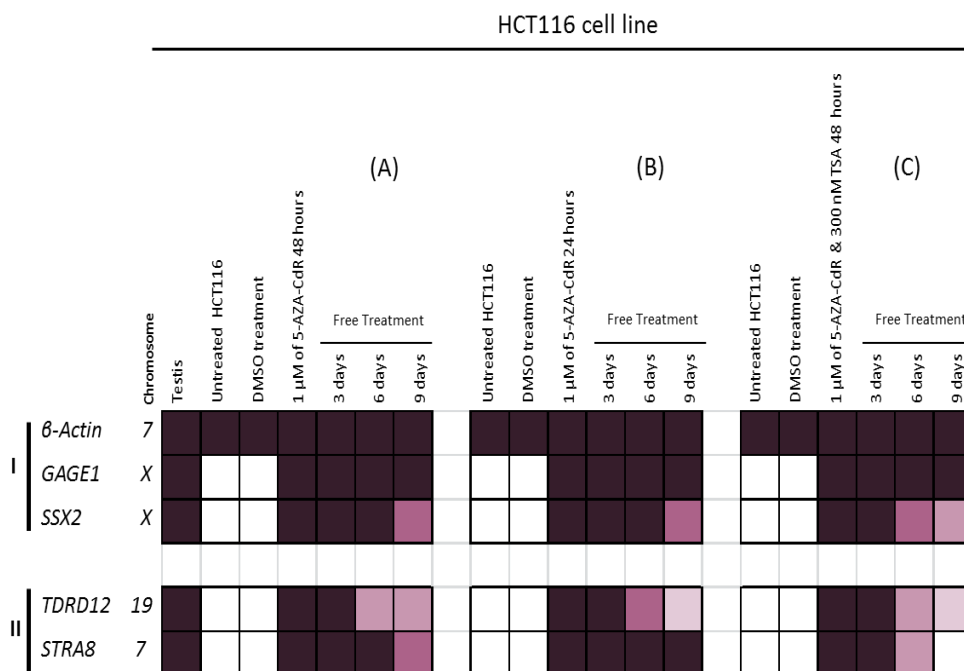
Figure 4.23. **Grid representation of the effect of 5-AZA-CdR and/or combination with TSA and the persistence of *TDRD12* and *STRA8* expression in the HCT116 colorectal cancer cell line.** HCT116 cells treated with 1 µM 5-AZA-CdR for 48 hours (A) or for 72 hours (B) or in combination with 300 nM TSA for 48 hours (C). Each column in the grid represents the positive control (testis), untreated cells, DMSO treated, drug treatment and drug-free treatment after 3, 6 and 9 days, while gene expressions are represented in the rows. RT-PCR expression profiles of each gene are visualised in 1% of agarose gel stained with ethidium bromide. The intensity of the bands are graded with different shades of purple colour (see the key at the top left) (+) is a weak intensity band and (+++++) is a strong intensity band. In the grid, (I) Expression of the β-*Actin* gene is the positive control for the cDNA samples, and the expression profile of previously known CT genes *GAGE1* and *SSX2*. (II) The expression profile of cancer/testis restricted genes; *TDRD12* and *STRA8.*

Chapter 5.0 Identification and analysis the biological role of PRDM9 in cancer cells

5.1 Introduction

The PRDM protein family is characterised by the presence of a conserved N-terminal PR domain followed by variable number of zinc-finger repeats that mediate specific DNA sequencing binding and protein-protein interaction (reviewed in Fumasoni *et al.*, 2007; Fog *et al.*, 2012). The PR domain was initially characterised in PRDM1 (positive regulatory domain binding factor-1; PRDI-BF-1) and PRDM2 (retinoblastoma protein-interacting zinc finger 1; RIZ1). This domain shares high homology with the catalytic SET domains (Suppressor of variegation 3-9, Enhancer of zeste and Trithorax) that characterise of a group of histone methyltransferases (HMTs) (reviewed in Fog *et al.*, 2012; Hohenauer and Moore 2012; Di Zazzo *et al.*, 2013).

The molecular mechanism of some zinc finger proteins was identified to be involved in a range of developmental processes. These proteins play critical roles in transcriptional regulators of cellular differentiation and maturation (Hohenauer and Moore, 2012). Some of the PRDM proteins, such as PRDM1 (Karube *et al.*, 2011), PRDM2 (Steele-Perkins *et al.*, 2001; Kim, Geng and Huang, 2003), PRDM5 (Shu *et al.*, 2011) and PRDM12 (Reid and Nacheva, 2003), act as tumour suppressors. For example, inactivation of PRDM1 can lead to natural killer cell lymphoma (Karube *et al.*, 2011) as well as diffused large B cell lymphoma (Mandelbaum *et al.*, 2010). Other PRDMs such as PRDM3 (Morishita, 2007; Koos *et al*. 2011), PRDM13 (Behrends *et al*., 2003) and PRDM14 (Nishikawa *et al*., 2007) are also oncogenic. For instance, PRDM14 is overexpressed frequently in breast cancers (Nishikawa *et al.*, 2007) and can initiate lymphoid leukaemia (Dettman *et al*., 2011).

PRDM7 and PRDM9 are the only members of this family that have expression restricted to germ cells entering meiotic prophase (Hayashi *et al*., 2005). PRDM9 is involved in regulation of meiotic recombination hotspots (Baudat *et al.*, 2010; Myers *et al.*, 2010; Parvanov *et al.*, 2010). In mice, disruption of *Prdm9* $^{-/-}$ directly regulates the expression of the testis-specific

RIK gene (also known as *Morc2b*). This indicates that *PRDM9* (*Meisetz* in mouse) can also play a critical role as a transcriptional factor that regulates meiotic genes (Hayashi *et al.*, 2005).

RT-PCR analysis demonstrated that overexpression of *PRDM9* is expressed in different cancer cell lines but not in normal human tissues other than the testis (Feichtinger *et al.*, 2012a) . In addition, PRDM9 protein has been detected in the NTERA-2 cancer cell line but not in normal prostate tissues, suggesting this gene as a potential CT antigen biomarker (Feichtinger *et al.*, 2012a). Another recent study found an excess of a rare allele of *PRDM9* linked with development of childhood leukaemogenesis as well as genomic rearrangements (Hussin *et al.*, 2013).

The main aim of this chapter was to investigate the *in vivo* biological roles of PRDM9 in cancer cells. The goals were to identify the relationship between *PRDM9* and the paralogous gene *PRDM7* as well as the human orthologues of *Rik* (*Morc2b*). Cancer cell lines were also investigated for localisation (nuclear or cytoplasmic) of the PRDM9 protein, as well as for the effect of gene knockdown on cell survival. Finally, the role of PRDM9 as a transcription factor was investigated by gene overexpression in a HeLa Tet-on 3G system.

5.2   Results

5.2.1 Comparison of *PRDM9* and *PRDM7* expression in normal and cancer tissues

PRDM7 has been reported to arise from a *PRDM9* duplication segment and undergoes marked structural rearrangements in human. Both genes show high frequency similarity in their DNA sequencing, but a decreased number of encoded zinc fingers occurs in *PRDM7* (Fumasoni *et al.*, 2007). Both genes consist of the PR/SET domain in addition to the KRAB (Kruppel-association box) protein-protein binding domain, which is lacking in the other family members (Figure 5.1) (Birtle and Ponting, 2006; Oliver *et al.*, 2009).

The expression of *PRDM7* and *PRDM9* was compared by designing two sets of internal RT-PCR primers to distinguish the expression patterns of these genes in normal and cancer tissues. Gene alignment of *PRDM7* and *PRDM9* identified an 89-nuclotide duplicon in the *PRDM7* sequencing that does not exist in the *PRDM9* sequence (Figure 5.2). The distinguishing primers were designed based on the 89-nucleotide duplicon region. Both genes showed no expression in normal tissues (Figure 5.3) but were co-expressed in leukaemia K-562 cells (Figure 5.4). *PRDM9* was also expressed in embryonal carcinoma cells, colon cancer cells, prostate cancer cells, ovary cancer cells, breast cancer cells and melanoma cells.

5.2.2 The relation between *PRDM9* and the human orthologue *Morc2b* (*Rik*) genes in normal and cancer tissues

The meiosis-specific *Rik* (*Morc2b*) gene was identified as being directly regulated by the *Prdm9* gene in mice (Hayashi *et al.,* 2005). We asked whether aberrant expression of *PRDM9* in human cancer cells might lead to activation of human *MORC* family genes. Four human orthologous genes (*MORC1*, *MORC2*, *MORC3* and *MORC4*) were identified and validated in normal and cancer cell lines using RT-PCR analysis.

The *PRDM9* and *MORC* family genes showed different expression patterns in normal and cancer tissues, with the exception of the *MORC1* gene (Figure 5.5 and Figure 5.6). *MORC2*, *MORC3* and *MORC4* were expressed in nearly all normal and cancer tissues, which indicated a negative relation between these genes and *PRDM9*. However, *MORC1* showed expression

in normal testis and spinal cord tissues. The expression of this gene was determined in a few

cancer cell lines including T84, SW480 and K-562, which might indicate a relative relationship

between the *MORC1* and *PRDM9* genes.
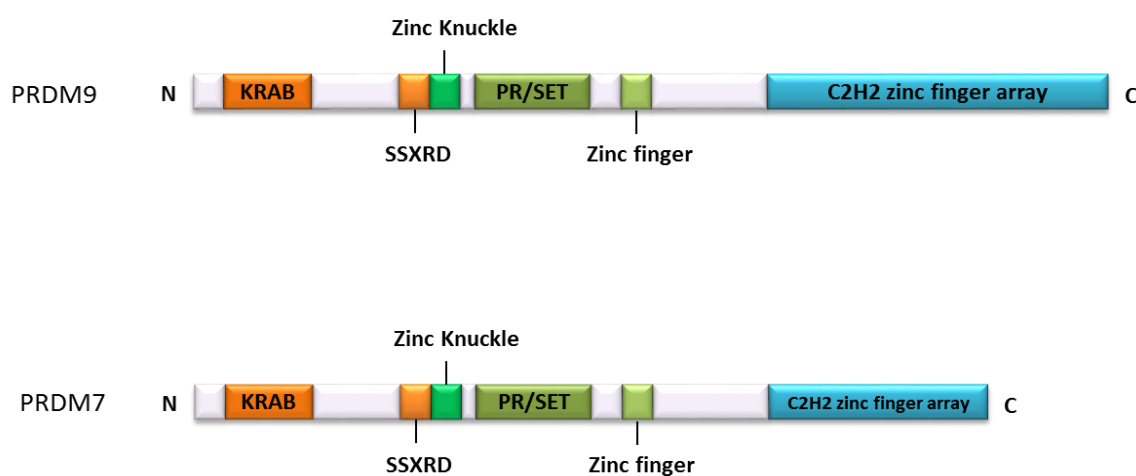
Figure 5.1. **Schematic showing the structure of PRDM9 and PRDM7.**Both genes are composed of three domains: a PR/SET domain encoding a histone methyltransferase activity, a KRAB protein-protein binding box and a C2H2 zinc finger repeated array, which is longer in PRDM9.

```
                10        20        30        40        50        60        70        80
PRDM9  ATGAGCCCTGAAAAGTCCCAAGAGGAGAGCCCAGAAGAAGACACAGAGAGAACAGAGCGGAAGCCCATGGTCAAAGATGC
       :::::::::::::X:::::::::::::::::X:::::::::::::::::::::::::::::::::::::::::::::::::::::
PRDM7  ATGAGCCCTGAAAGGTCCCAAGAGGAGAGCCCAGAAGGAGACACAGAGAGAACAGAGCGGAAGCCCATGGTCAAAGATGC
                10        20        30        40        50        60        70        80

                90       100       110       120       130       140       150       160
PRDM9  CTTCAAAGACATTTCCATATACTTCACCAAGGAAGAATGGGCAGAGATGGGAGACTGGGAGAAAACTCGCTATAGGAATG
       ::::::::::::::::::::::::::::::::::::::::::::X:::::::::::::::::::::::::::::::::::::::
PRDM7  CTTCAAAGACATTTCCATATACTTCACCAAGGAAGAATGGGCAGAAATGGGAGACTGGGAGAAAACTCGCTATAGGAATG
                90       100       110       120       130       140       150       160

               170       180       190       200       210       220       230       240
PRDM9  TGAAAGGAACTATAATGCACTGATTACTATAGGTCTCAGAGCCACTCGACCAGCTTTCATGTGTCACCGAAGGCAGGCC
       ::::::X::::::::::::::::::::::X::::::::::::::::::::::::::::::::::::::::::::::::::::::
PRDM7  TGAAATGAACTATAATGCACTGATTACTGTAGGTCTCAGAGCCACTCGACCAGCTTTCATGTGTCACCGAAGGCAGGCC
               170       180       190       200       210       220       230       240

               250       260       270       280       290       300       310       320
PRDM9  ATCAAACTCCAGGTGGATGACACAGAAGATTCTGATGAAGAATGGACCCCTAGGCAGCAAGTCAAACCTCCTTGGATGGC
       :::::::::::::::::::::::::::::::::::X:::::::::::::X::::::::::::::::::::::::::::::::::
PRDM7  ATCAAACTCCAGGTGGATGACACAGAAGATTCCGATGAAGAATGGACACCTAGGCAGCAAGTCAAACCTCCTTGGATGGC
               250       260       270       280       290       300       310       320

               330       340       350       360       370       380       390       400
PRDM9  CTTAAGAGTGGAACAGCGTAAACACCAGAAGGGAATGCCCAAGGCGTCATTCAGTAATGAATCTAGTTTGAAAGAATTGT
       :::X::::XX::::::X:::::::::::::::::::::::::::::::::::X::::::::::::::::::X:::::::::
PRDM7  CTTCAGAGGAGAACAGAGTAAACACCAGAAGGGAATGCCCAAGGCGTCATTCAATAATGAATCTAGTTTGAGAGAATTGT
               330       340       350       360       370       380       390       400

               410       420       430       440       450       460       470       480
PRDM9  CAAGAACAGCAAATTTACTGAATGCAAGTGGCTCAGAGCAGGCTCAGAAACCAGTGTCCCCTTCTGGAGAAGCAAGTACC
       ::X:::::XX:::::::::::::X::::::X:::::::::::::::::::::::::::::::::::X:::::::::::::::::
PRDM7  CAGGAACGGCAAATTTACTGAATACAAGTGACTCAGAGCAGGCTCAGAAACCAGTGTCCCCTCCTGGAGAAGCAAGTACC
               410       420       430       440       450       460       470       480

               490       500       510       520       530       540       550       560
PRDM9  TCTGGACAGCACTCAAGACTAAAACTGGAACTCAGGAAGAAGGAGACTGAAAGAAAGATGTATAGCCTGCGAGAAAGAAA
       :::::::::::::::X::::::::::::::::::::::::X:::::::::::::X:::::::::::::::::::::::::::::
PRDM7  TCTGGACAGCACTCTAGACTAAAACTGGAACTCAGGAGGAAGGAGACTGAAGGAAAGATGTATAGCCTGCGAGAAAGAAA
               490       500       510       520       530       540       550       560

               570       580       590       600       610       620       630       640
PRDM9  GGGTCATGCATACAAAGAGGTCAGCGAGCCGCAGGATGATGATTACCTCTATTGTGAGATGTGTCAGAACTTCTTCATTG
       :::::::::::::::::::::X::::::::::X::::::::::X:::::::::::::::::::::::::::::::::::::::
PRDM7  GGGTCATGCATACAAAGAGATCAGCGAGCCACAGGATGATGACTACCTCTATTGTGAGATGTGTCAGAACTTCTTCATTG
               570       580       590       600       610       620       630       640

               650       660       670       680       690       700       710       720
PRDM9  ACAGCTGTGCTGCCCCATGGGCCCCCTACATTTGTAAAGGACAGTGCAGTGGACAAGGGGCACCCCAACCGTTCAGCCCTC
       :::::::::::::::X:::::::::::::::::::::::::::::::::::::::::::::::X:::::::::::::::::::
PRDM7  ACAGCTGTGCTGCCTCATGGGCCCCCTACATTTGTAAAGGACAGTGCAGTGGACAAGGGGCATCCCAACCGTTCAGCCCTC
               650       660       670       680       690       700       710       720

               730       740       750       760       770       780       790       800
PRDM9  AGTCTGCCCCCAGGGCTGAGAATTGGGCCATCAGGCATCCCTCAGGCTGGGCTTGGAGTATGGAATGAGGCATCTGATCT
       :::::::::::X:::::::::::::::::::::::::::::::::::::::::::::::::::::::X::::::::::::::::
PRDM7  AGTCTGCCCCCGGGGCTGAGAATTGGGCCATCAGGCATCCCTCAGGCTGGGCTTGGAGTATGGAACGAGGCATCTGATCT
               730       740       750       760       770       780       790       800
               810       820       830       840       850       860       870       880
PRDM9  GCCGCTGGGTCTGCACTTTGGCCCTTATGAGGGCCGAATTACAGAAGACGAAGAGGCAGCCAACAATGGATACTCCTGGC
       :::X:::::::::::::::::::::::::X::::::::::::::::::::::::::::::::::::::X:::::X:::::::
PRDM7  GCCACTGGGTCTGCACTTTGGCCCCTATGAGGGCCGAATTACAGAAGACGAAGAGGCAGCCAACAGTGGATATTCCTGGC
               810       820       830       840       850       860       870       880
```

```
              890       900       910       920       930       940       950       960
PRDM9   TGATCACCAAGGGGAGAAACTGCTATGAGTATGTGGATGGAAAAGATAAATCCTGGGCCAACTGGATGAGGTATGTGAAC
        :X:::::::::::::::::::::::::::::::::::::::::::::::::::X::::::::::::::::::::::::::::::::
PRDM7   TAATCACCAAGGGGAGAAACTGCTATGAGTATGTGGATGGAAAAGATAAATCCTCGGCCAACTGGATGAGGTATGTGAAC
              890       900       910       920       930       940       950       960

              970       980       990      1000      1010      1020      1030      1040
PRDM9   TGTGCCCGGGATGATGAAGAGCAGAACCTGGTGGCCTTCCAGTACCACAGGCAGATCTTCTATAGAACCTGCCGAGTCAT
        :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
PRDM7   TGTGCCCGGGATGATGAAGAGCAGAACCTGGTGGCCTTCCAGTACCACAGGCAGATCTTCTATAGAACCTGCCGAGTCAT
              970       980       990      1000      1010      1020      1030      1040

             1050      1060
PRDM9   TAGGCCAGGCTGTGAACTGCTGGTCTGGT-------------------------------------------------
        :::::::::::::::::::::::::::::::
PRDM7   TAGGCCAGGCTGTGAACTGCTGGTCTGGTCTGGGGATGAGTATGGCCAGGAACTGGGCATCAGATCTTCTATAGAACCTG
             1050      1060      1070      1080      1090      1100      1110      1120

                                                  1070      1080      1090      1100      1110
PRDM9   ------------------------------ATGGGGATGAATACGGCCAGGAACTGGGCATCAAGTGGGGCA
                                        ::::::::::X::X:::::::::::::::::::::::::::::::
PRDM7   CCGAGTCATTAGGCCAGGCTGTGAACTGCTGGTCTGGTATGGGGATGAGTATGGCCAGGAACTGGGCATCAAGTGGGGCA
             1130      1140      1150      1160      1170      1180      1190      1200

             1120      1130      1140      1150      1160      1170      1180      1190
PRDM9   GCAAGTGGAAGAAAGAGCTCATGGCAGGGAGAGAACCAAAGCCAGAGATCCATCCATGTCCCTCATGCTGTCTGGCCTTT
        :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
PRDM7   GCAAGTGGAAGAAAGAGCTCATGGCAGGGAGAGAACCAAAGCCAGAGATCCATCCATGTCCCTCATGCTGTCTGGCCTTT
             1210      1220      1230      1240      1250      1260      1270      1280

             1200      1210      1220      1230      1240      1250      1260      1270
PRDM9   TCAAGTCAGAAATTTCTCAGTCAACATGTAGAACGCAATCACTCCTCTCAGAACTTCCCAGGACCATCTGCAAGAAAACT
        ::::::::X::::::::::::::::::X:::::::::::::::::::::::::::::::::::::::::::::::::::::::::
PRDM7   TCAAGTCAAAAATTTCTCAGTCAACATGTGGAACGCAATCACTCCTCTCAGAACTTCCCAGGACCATCTGCAAGAAAACT
             1290      1300      1310      1320      1330      1340      1350      1360

             1280      1290      1300      1310      1320      1330      1340      1350
PRDM9   CCTCCAACCAGAGAATCCCTGCCCAGGGGATCAGAATCAGGAGCAGCAATATCCAGATCCACACAGCCGTAATGACAAAA
        X:::::::::::::::::::::::::::::::::::::::::::X:::::::X:X:::::::X:X::X:::::::::::::
PRDM7   TCTCCAACCAGAGAATCCCTGCCCAGGGGATCAGAATCAGGAGCGGCAATATTCTGATCCACGCTGCTGTAATGACAAAA
             1370      1380      1390      1400      1410      1420      1430      1440

             1360      1370      1380      1390      1400      1410      1420      1430
PRDM9   CCAAAGGTCAAGAGATCAAAGAAAGGTCCAAACTCTTGAATAAAAGGACATGGCAGAGGGAGATTTCAAGGGCCTTTTCT
        :::::::::::::::::::::::::::::::::::::::::::
PRDM7   CCAAAGGTCAAGAGATCAAAGAAAGGTCCAAACTCTTGA----------------------------------------
             1450      1460      1470

             1440      1450      1460      1470      1480      1490      1500      1510
PRDM9   AGCCCACCCAAAGGACAAATGGGGAGCTGTAGAGTGGGAAAAAGAATAATGGAAGAAGAGTCCAGAACAGGCCAGAAAGT
PRDM7   --------------------------------------------------------------------------------

             1520      1530      1540      1550      1560      1570      1580      1590
PRDM9   GAATCCAGGGAACACAGGCAAATTATTTGTGGGGGGTAGGAATCTCAAGAATTGCAAAAGTCAAGTATGGAGAGTGTGGAC
PRDM7   --------------------------------------------------------------------------------

             1600      1610      1620      1630      1640      1650      1660      1670
PRDM9   AAGGTTTCAGTGTTAAATCAGATGTTATTACACACCAAAGGACACATACAGGGGAGAAGCTCTACGTCTGCAGGGAGTGT
PRDM7   --------------------------------------------------------------------------------
```

165

```
              1680      1690      1700      1710      1720      1730      1740      1750
PRDM9  GGGCGGGGCTTTAGCTGGAAGTCACACCTCCTCATTCACCAGAGGATACACACAGGGGAGAAGCCCTATGTCTGCAGGGA

PRDM7  _____


              1760      1770      1780      1790      1800      1810      1820      1830
PRDM9  GTGTGGGCGGGGCTTTAGCTGGCAGTCAGTCCTCCTCACTCACCAGAGGACACACACAGGGGAGAAGCCCTATGTCTGCA

PRDM7  _____


              1840      1850      1860      1870      1880      1890      1900      1910
PRDM9  GGGAGTGTGGGCGGGGCTTTAGCCGGCAGTCAGTCCTCCTCACTCACCAGAGGAGACACACAGGGGAGAAGCCCTATGTC

PRDM7  _____


              1920      1930      1940      1950      1960      1970      1980      1990
PRDM9  TGCAGGGAGTGTGGGCGGGGCTTTAGCCGGCAGTCAGTCCTCCTCACTCACCAGAGGAGACACACAGGGGAGAAGCCCTA

PRDM7  _____


              2000      2010      2020      2030      2040      2050      2060      2070
PRDM9  TGTCTGCAGGGAGTGTGGGCGGGGCTTTAGCTGGCAGTCAGTCCTCCTCACTCACCAGAGGACACACACAGGGGAGAAGC

PRDM7  _____


              2080      2090      2100      2110      2120      2130      2140      2150
PRDM9  CCTATGTCTGCAGGGAGTGTGGGCGGGGCTTTAGCTGGCAGTCAGTCCTCCTCACTCACCAGAGGACACACACAGGGGAG

PRDM7  _____


              2160      2170      2180      2190      2200      2210      2220      2230
PRDM9  AAGCCCTATGTCTGCAGGGAGTGTGGGCGGGGCTTTAGCAATAAGTCACACCTCCTCAGACACCAGAGGACACACACAGG

PRDM7  _____


              2240      2250      2260      2270      2280      2290      2300      2310
PRDM9  GGAGAAGCCCTATGTCTGCAGGGAGTGTGGGCGGGGCTTTCGCGATAAGTCACACCTCCTCAGACACCAGAGGACACACA

PRDM7  _____


              2320      2330      2340      2350      2360      2370      2380      2390
PRDM9  CAGGGGAGAAGCCCTATGTCTGCAGGGAGTGTGGGCGGGGCTTTAGAGATAAGTCAAACCTCCTCAGTCACCAGAGGACA

PRDM7  _____

              2400      2410      2420      2430      2440      2450      2460      2470
PRDM9  CACACAGGGGAGAAGCCCTATGTCTGCAGGGAGTGTGGGCGGGGCTTTAGCAATAAGTCACACCTCCTCAGACACCAGAG

PRDM7  _____
```

```
            2480      2490      2500      2510      2520      2530      2540      2550
PRDM9    GACACACACAGGGGAGAAGCCCTATGTCTGCAGGGAGTGTGGGCGGGGCTTTCGCAATAAGTCACACCTCCTCAGACACC

PRDM7    ————————————————————————————————————————————————————————————————————————————


            2560      2570      2580      2590      2600      2610      2620      2630
PRDM9    AGAGGACACACACAGGGGAGAAGCCCTACGTCTGCAGGGAGTGTGGGCGGGGCTTTAGCGATAGGTCAAGCCTCTGCTAT

PRDM7    ————————————————————————————————————————————————————————————————————————————


            2640      2650      2660      2670      2680
PRDM9    CACCAGAGGACACACACAGGGGAGAAGCCCTACGTCTGCAGGGAGGATGAGTAA

PRDM7    ——————————————————————————————————————————————————
```

Figure 5.2. Gene alignments of *PRDM9* and *PRDM7* show high similarity between both genes. *PRDM7* is shorter than *PRDM9* and lacks some of the zinc finger repeats (Turquoise), which is the main feature of PRDM9. The 89-nucleotide segment (Red) is present only in *PRDM7* and does not have any repetitive elements. Both genes showed identical sequencing with the exception of some nucleotides (Green). Identical sequencing of the start codon and stop codon cloning primers of *PRDM7* to the *PRDM9* sequences (Yallow).
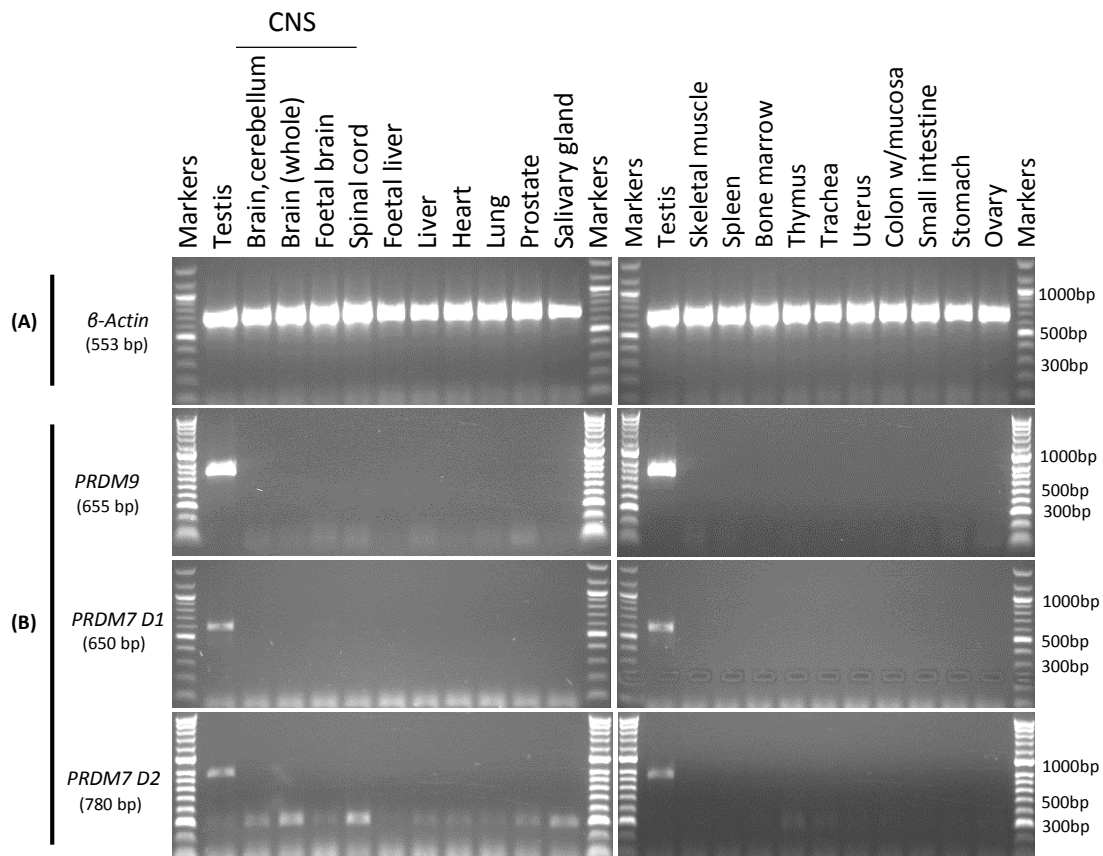
Figure 5.3. **Comparison of the expression of *PRDM9* and *PRDM7* in normal human tissues**. Agarose gels stained with ethidium bromide showing the expression pattern of *PRDM9* and *PRDM7* genes in 21 normal tissues. (A) *β-Actin* serves as a positive control for the cDNA samples. (B) *PRDM9* and *PRDM7* showed no expression in non-testis somatic tissues. *PRDM9* and *PRDM7* share highly similar sequences and distinguishing primers D1 and D2 were used to distinguish these two genes.
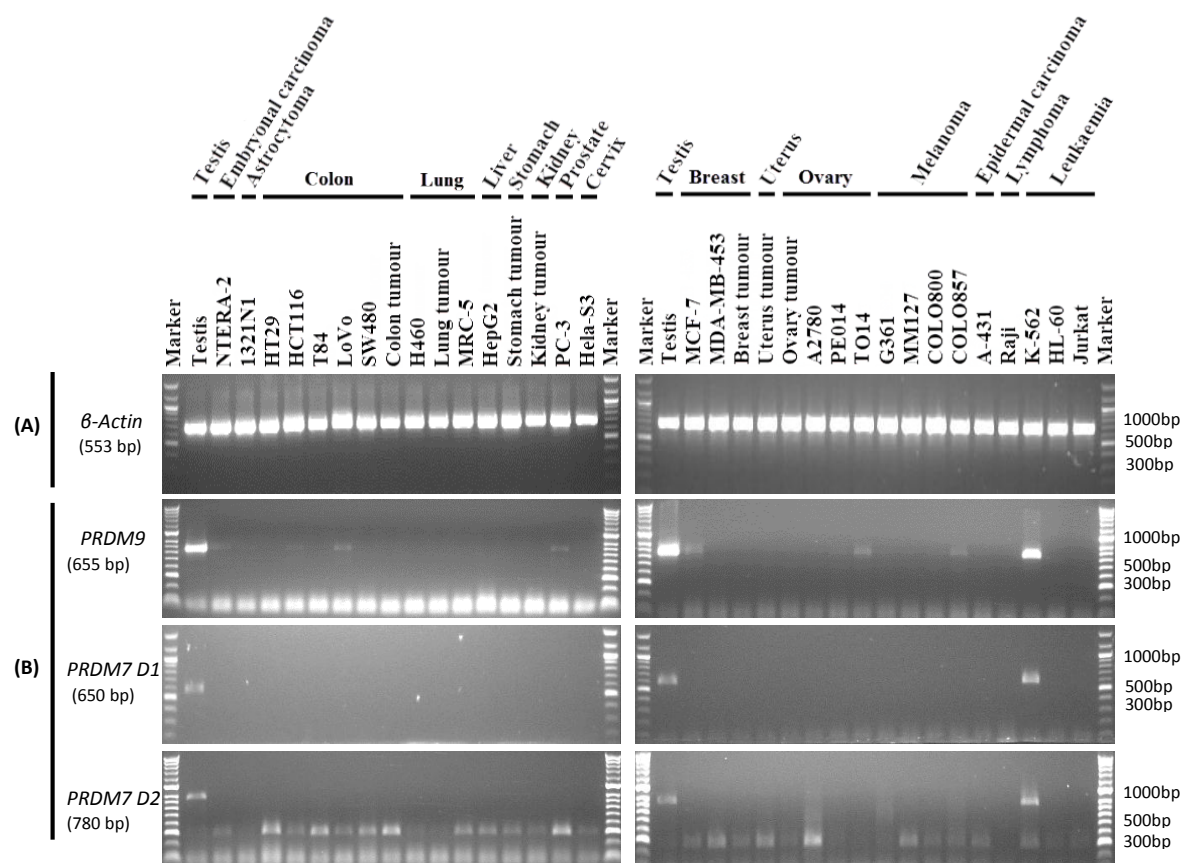
Figure 5.4. Comparison of the expression of *PRDM9* and *PRDM7* in cancer cell lines and tissues. Agarose gels stained with ethidium bromide showing expression pattern of *PRDM9* and *PRDM7* genes in 33 cancer cell lines and tissues. (A) *β-Actin* served as a positive control for cDNA samples. (B) *PRDM9* alone was expressed in different cancer cells while both genes were expressed in leukaemia K-562 cells with high intensity bands. *PRDM9* and *PRDM7* share highly similar sequences and distinguishing primers D1 and D2 were used to distinguish these two genes.
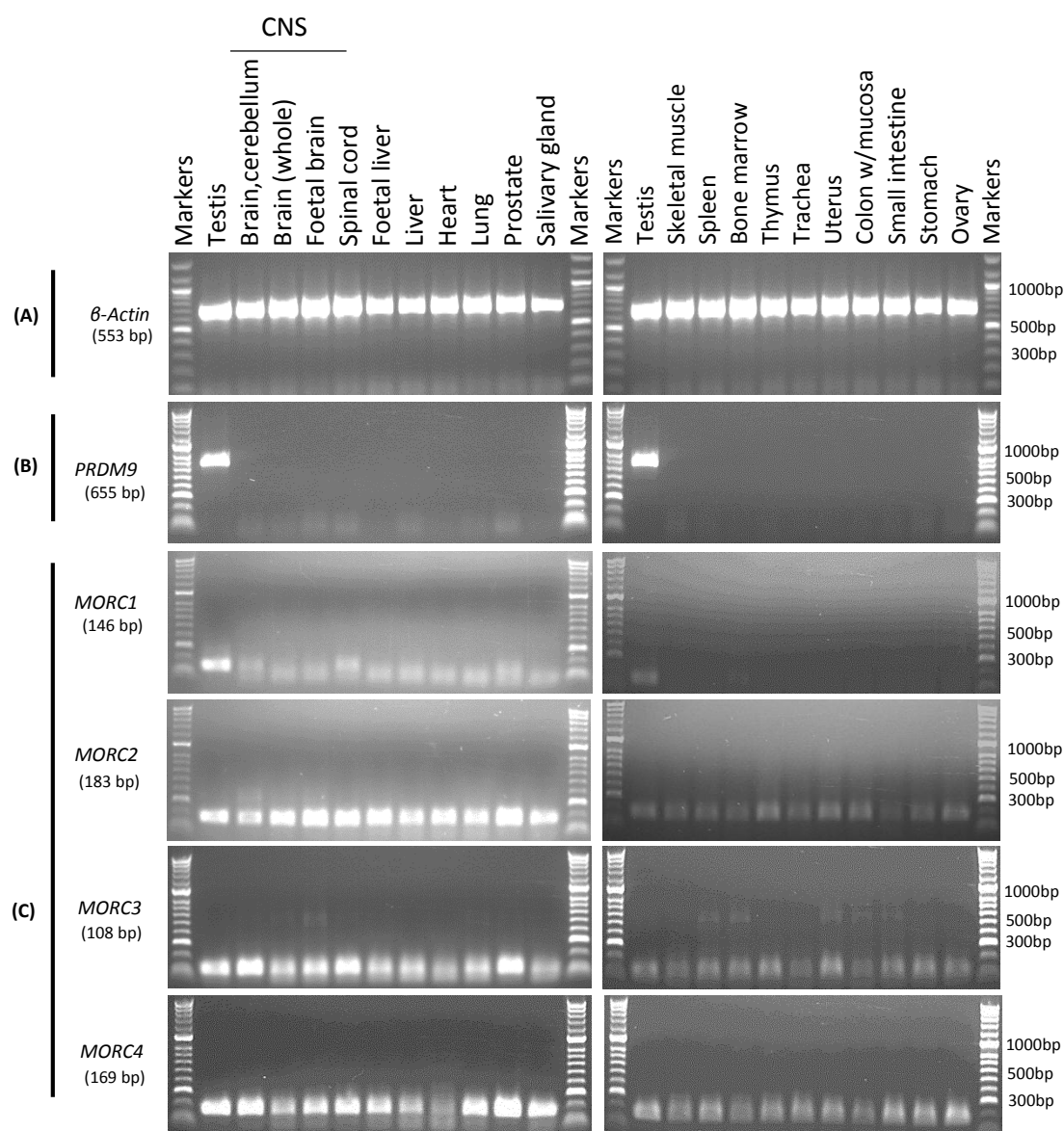
Figure 5.5. RT-PCR screening analysis of *MORC* family genes expression and comparison with *PRDM9* gene expression in a range of human normal tissues. Agarose gels showing different expression patterns between *PRDM9* and *MORC* family genes, with the exception of the *MORC1* gene, in 21 normal human tissues. (A) *β-Actin* gene served as a positive control for cDNA samples. (B) *PRDM9* is expressed only in the testis and not in other normal tissues. (C) *MORC* family gene were expressed in most of the human normal tissues. Two set of primers for *MORC2*, *MORC3* and *MORC4* indicate a negative relation between these genes and *PRDM9*, with the exception of *MORC1* that is expressed in the normal testis and spinal cord only.
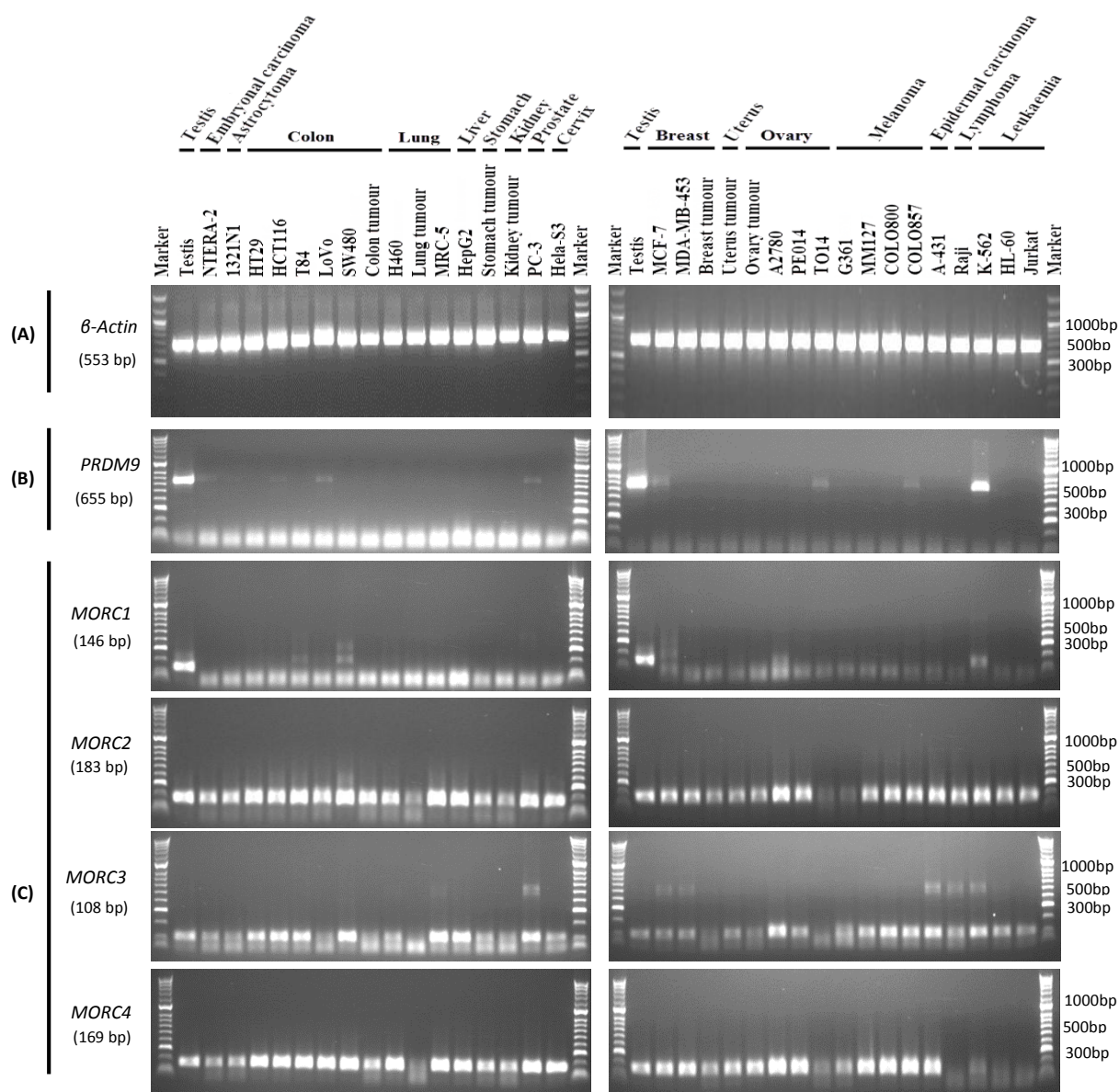
Figure 5.6. **RT-PCR screening analysis of *MORC* family gene expression compared with *PRDM9* gene expression in a range of cancer cell lines and tissues**. Agarose gels stained with ethidium bromide showing expression patterns of *PRDM9* and *MORC1* genes in 33 cancer cell lines and tissues. (A) *β-Actin* served as the positive control for cDNA samples. (B) *PRDM9* is expressed in most of the cancer cell lines and tissues. (C) *MORC2*, *MORC3* and *MORC4* genes are expressed in most of the cancer cell lines tissues. The *MORC1* gene is only expressed in a few cancer tissues, indicating perhaps a relative relationship between this gene and the *PRDM9* gene.

5.2.3 Protein analysis and cellular localisation of PRDM9 in different cancer cell lines

PRDM9 expression was identified in several types of cancer cell lines and tissues using RT-PCR analysis. Based on this result, PRDM9 was analysed at the protein level in 8 different cancer types using polyclonal anti-PRDM9 antibody (Abcam; ab85654). Whole cell lysates were prepared from 10 cancer lines (NT2, 1321N1, HCT116, SW480, H460, MRC5, HeLa, MCF-7, A2780 and K-562). The PRDM9 protein was detected in nearly all cancer cell lines with an expected size of approximately 102 kDa, using anti-α-tubulin antibody as a positive control (Figure 5.7 A). This result indicates that aberrant expression of PRDM9 might encode a clinically important oncogenic antigen (Feichtinger *et al.*, 2012a).

The cellular localisation of PRDM9 was assessed by preparing nuclear (N), cytoplasmic (C) and whole cell (W) proteins from 1321N1, A2780, and HCT116, as well as NT2 cell lines. PRDM9 was present in 1321N1, A2780 and HCT116 cells in the nucleus whereas it was cytoplasmic in the NT2 cell line (Figure 5.6 B). α-tubulin was used as a cytoplasmic and lamin A/C as a nuclear positive control.

5.2.4 siRNA knockdown of PRDM9

In the present study, knockdown of PRDM9 was attempted using four different siRNAs (5, 6, 7 and 8) to determine the effect of PRDM9 knockdown on 1321N1, A2780, HCT116, SW480, K-562, MCf-7 and NT2 cells. Different conditions were examined to optimise successful PRDM9 knockdown, including increasing the concentration of siRNA and transfection reagent as well as the number of siRNA hits. Protein concentration was measured using the BCA protein assay and approximately 30 µg protein was loaded per well.

PRDM9 knockdown was carried out in leukaemia K-562 and ovarian cancer A2780 cell lines using a single hit or three hits. Western blot analysis showed failure of the PRDM9 knockdown using siRNAs 5, 6, 7 and 8 in both K-562 and A2780 cells compared to the negative control non-interference as well as untreated cells (Figure 5.8). Three hits were used for PRDM9 knockdown in K-562 cells. In contrast, one or three hits appeared to have no effect on PRDM9 knockdown in the A2780 cell line, even when using a combination of siRNAs. Anti-α-tubulin antibody was used as positive loading control.

PRDM9 knockdown was performed in embryonal carcinoma NT2 and astrocytoma 1321N1 cell lines. Western blot analysis showed successful knockdown of PRDM9 in NT2 cells after single hits with siRNA5 and siRNA8 (Figure 5.9). However, no knockdown of PRDM9 was observed in 1321N cells after a single hit with any of the four siRNAs. We increased the siRNA hits to three hits as a further attempt at knockdown of PRDM9 in 1321N1 cells. The western blot results showed that increasing the hits to three led to successful knockdown of PRDM9 in 1321N1 cells (Figure 5.9).

PRDM9 knockdown also failed in breast cancer Mcf-7 cells using all four siRNAs, even with three hits, but showed a slight effect when a combination of siRNAs was used (Figure 5.10). In contrast, successful knockdown was observed using three hits of siRNAs in colorectal cancer cell lines HCT116 and SW480. PRDM9 was successfully knocked down using three hits of siRNA7 in both HCT116 and SW480 cells compared to the negative control non-interference as well as untreated cells (Figure 5.10). In addition, siRNAs 6 and 8 showed slight knockdown in both cell lines. Protein loading was assessed using anti-α-tubulin antibody, which served as a positive control.
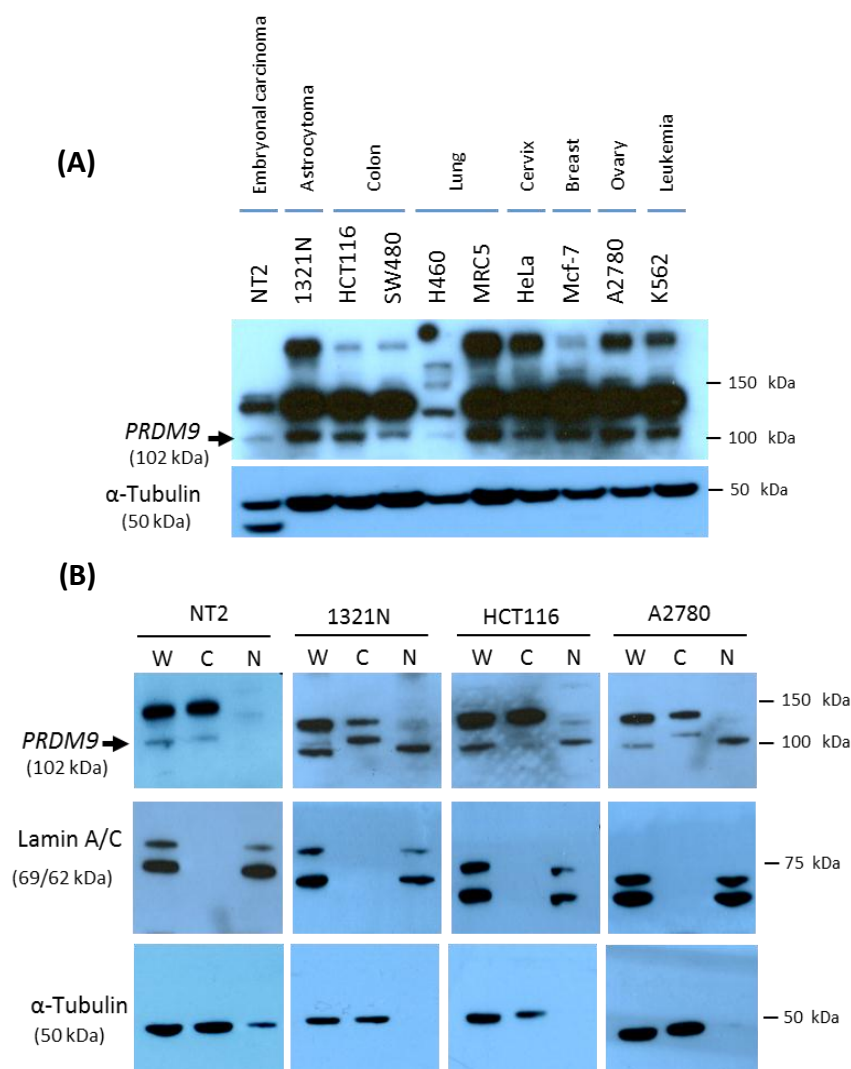
Figure 5.7. **Western blot analysis of PRDM9 in different cancer cell lines. (A)** The expression of PRDM9 protein was determined in 10 cancer cell lines. α-tubulin was used as a positive control. **(B)** The cellular localization of PRDM9 was determined in NT2, 1321N, HCT116 and A2780 cells using whole cell lysates (W), cytoplasmic lysate fractions (C) and nuclear lysate fractions (N). PRDM9 was present in the nucleus in 1321N, HCT116 and A2780 cells, but in the cytoplasm in the NT2 cell line. α-tubulin was used as a cytoplasmic positive control and lamin A/C serves as a nuclear control.
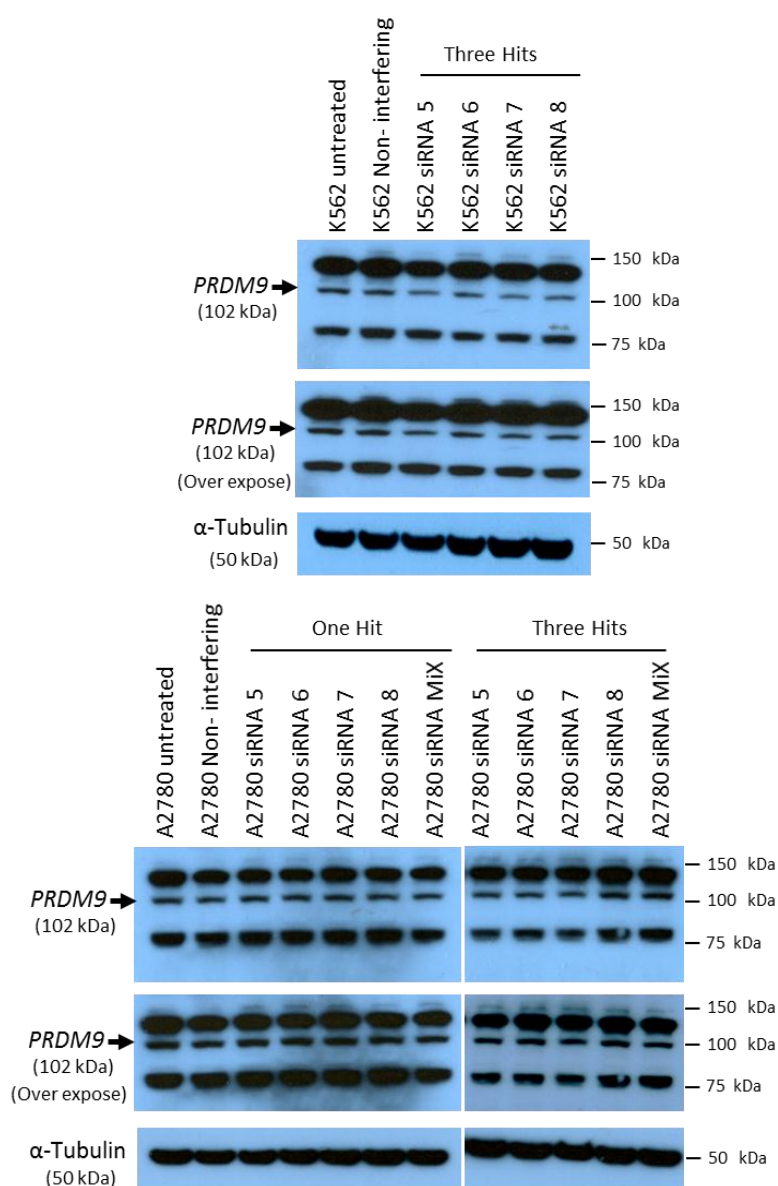
Figure 5.8. **Western blot analysis of PRDM9 knockdown in K-562 and A2780 cells.** Four different siRNA numbers (5, 6, 7 and 8) as well as a combination of these siRNAs were used to knockdown PRDM9. No successful knockdown was observed using three hits in K-562 cells when compared to the negative control non-interference as well as untreated cells. PRDM9 knockdown failed using either one or three hits in A2780 cell lines, even with combinations of siRNAs. α-tubulin served as a positive loading control.
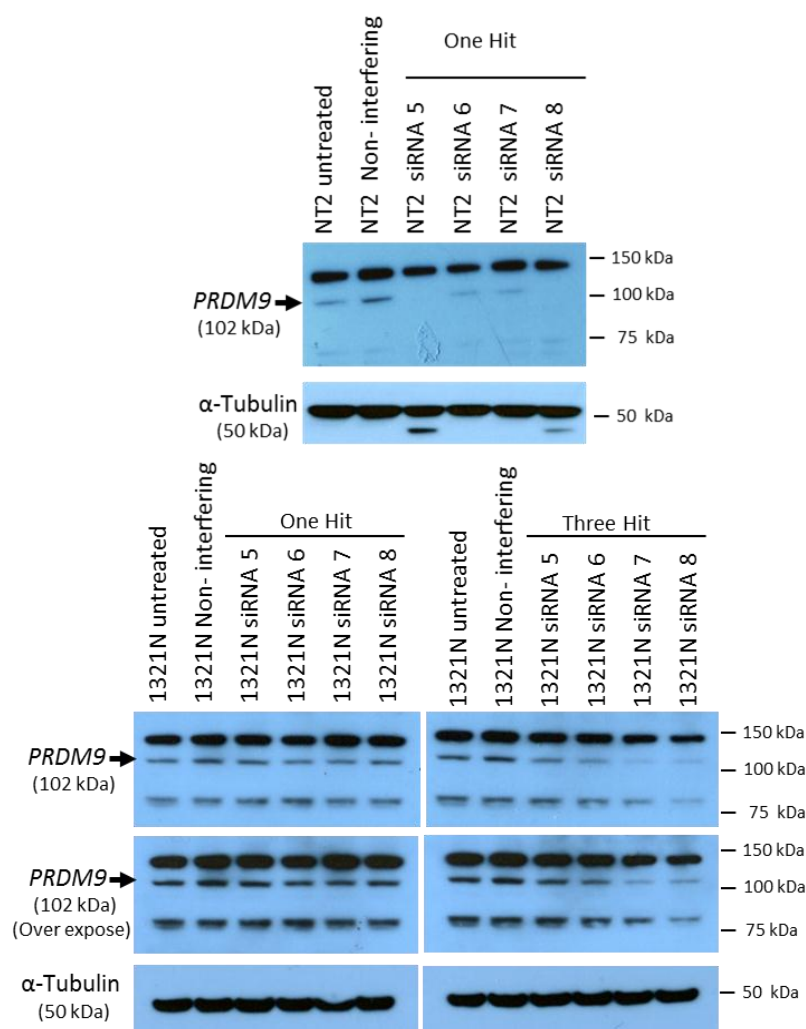
Figure 5.9. Western blot analysis of PRDM9 knockdown in NT2 and 1321N1 cells. Four different siRNA numbers (5, 6, 7 and 8) were used to knockdown PRDM9 in NT2 and 1321N1 cells. Successful knockdown was observed after one hit of the siRNA in NT2 cells and three hits in 1321N1 cells when compared to the negative control non-interference as well as untreated cells. In NT2, PRDM9 knockdown was successful with siRNA5 and siRNA8 after only one hit. In contrast, the knockdown was only slightly successful in 1321N1 cells using siRNA7 and siRNA8 after three hits. α-tubulin served as a positive loading control.
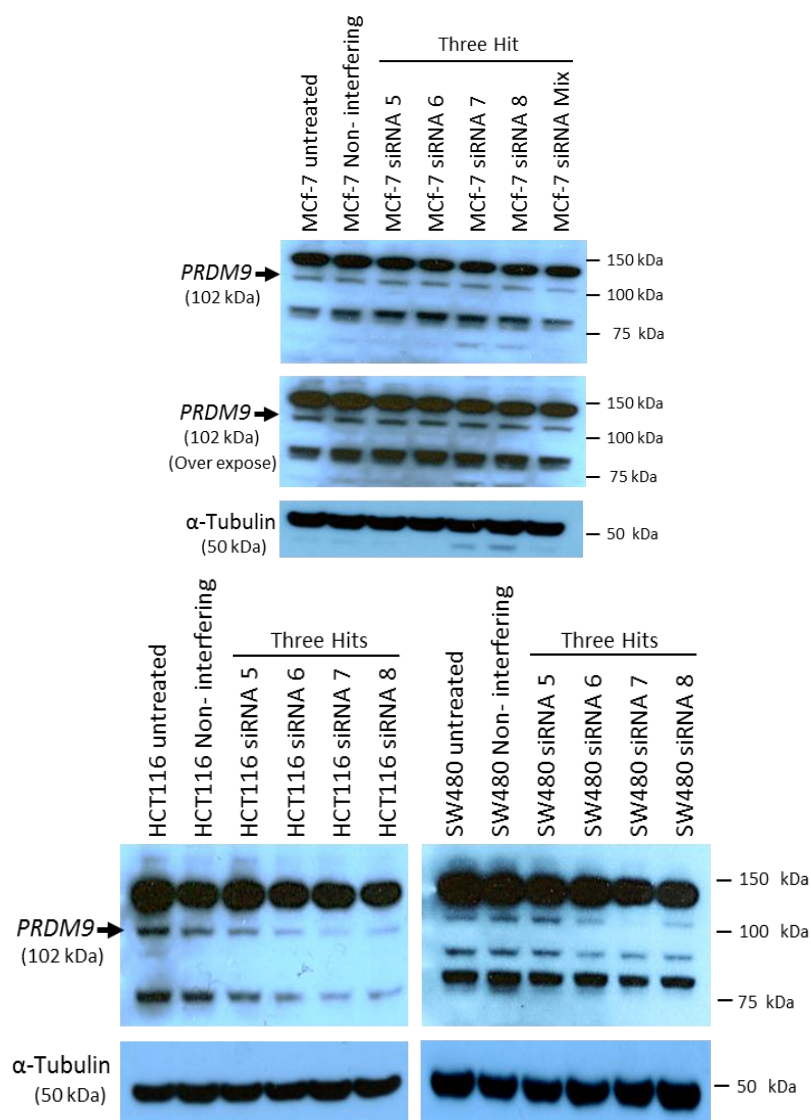
Figure 5.10. **Western blot analysis of PRDM9 knockdown in Mcf-7, HCT116 and SW480 cells.** Four different siRNA numbers (5, 6, 7 and 8) as well as a combination of these siRNAs were used to knockdown PRDM9 in Mcf-7, HCT116 and SW480 cells. PRDM9 knockdown failed using four siRNAs in Mcf-7 cells but a slight effect was seen with a combination of these siRNAs. In contrast, successful knockdown was observed using three hits of the siRNA7 in HCT116 and SW480 cells compared to the negative control non-interference as well as untreated cells. The siRNAs 6 and 8 showed a slight knockdown in HCT116 and SW480 cells. α-tubulin served as positive loading control.

5.2.5 Analysis of proliferation of HCT116 and SW480 cells transfected with PRDM9 siRNAs

Western blot analysis showed successful knockdown of PRDM9 in HCT116 and SW480 cells after three hits of siRNA. PRDM9 has been identified to serve as a meiotic hotspot regulator but its aberrant function in cancer cells remains poorly understood (Baudat *et al.*, 2010; Myers *et al.*, 2010; Parvanov *et al.*, 2010). We asked whether PRDM9 knockdown could influence the proliferation potential of HCT116 and SW480 cells. The effect of PRDM9 knockdown was examined in adherent parental cells by seeding 1, 10, 100 or 1000 cells in 96-well plates in 100 µl of growth medium. Six well repeats were used for untreated cells, negative control siRNA, HiPerFect treated cells and PRDM9 cells treated with siRNA (siRNAs 5, 6, 7 and 8). The transfection complex was made by adding 0.1 nM siRNA containing 0.3 µl HiPerFect reagent to 4.7 µl serum-free medium. A negative control non-interference siRNA was prepared in the same way. The PRDM9 siRNA transfection mixture, negative control siRNA or HiPerFect only were added to the culture upon cell seeding and re-applied after 2 and 6 days to the parental HCT116 and SW480 cells. Untreated cells were used to compare the effect of the siRNA treatments on cell proliferation after 10 days of incubation. Light microscopy was used to count the number of wells showing positive growth greater than that observed for the cells transfected with the negative control siRNA. The extreme limiting dilution analysis (ELDA) webtool (http://bioinf.wehi.edu.au/software/elda/) was used to determine frequencies of the cell proliferation.

PRDM9 knockdown using siRNA-5 and siRNA-7 in parental HCT116 cells resulted in a statistically significant 8-fold reduction in proliferation (P<0.01) when compared to cells transfected with the negative control siRNA. Transfection with siRNA-6 also reduced the cell proliferation by 6-fold, while siRNA-8 transfection caused a 4-fold reduction (Table 5.1 A) (Figures 5.11 and 5.12). Pairwise comparisons of cell proliferation frequencies between untreated cells when compared to either cells transfected with negative control siRNA or treated with HiPerFect only revealed no statistically significant differences (P> 0.01). Comparison of cell proliferation between negative control siRNA transfected cells and cells treated with HiPerFect only also showed no statistically significant differences (P> 0.01)

(Table 5.1 B). In contrast, pairwise comparisons of all untreated cells, negative control siRNA transfected cells and HiPerFect treated cells with the PRDM9 siRNA-5, siRNA-6 and siRNA-7 transfected groups showed highly significant differences (P< 0.01).

In addition, knockdown of PRDM9 with siRNA7 in parental SW480 cells resulted in a 14-fold reduction in cell proliferation (P< 0.01) when compared to cells transfected with the negative control siRNA (Table 5.2 A) (Figures 5.13 and 5.14). Pairwise comparisons revealed no significant differences (P> 0.01) in cell proliferation frequencies between untreated/negative control siRNA transfected cells, untreated/HiPerFect only treated cells and negative control siRNA transfected/HiPerFect only treated cells (Table 5.2 B). In contrast, all pairwise comparisons with the PRDM9 siRNA-7 transfected group revealed highly significant differences (P< 0.01).

<u>Table 5.1 A.</u> Extreme limiting dilution analysis (ELDA) assay showing the influence of PRDM9 siRNA-mediated knockdown on proliferation of HCT116 cells

Number of cells per well	Number of wells plated	Untreated cells	Negative control siRNA	HiPerFect only	*PRDM9* siRNA-5	*PRDM9* siRNA-6	*PRDM9* siRNA-7	*PRDM9* siRNA-8
Number of well showing positive growth of HCT116 cells after 10 days culturing								
1000	6	6	6	6	6	6	6	6
100	6	6	6	6	6	6	6	6
10	6	6	6	6	1	2	1	3
1	6	0	0	0	0	0	0	0
Cell proliferation frequency (95% CI)		1/4 (1/11-1/2)	1/4 (1/11-1/2)	1/4 (1/11-1/2)	1/32 (1/82-1/13)	1/23 (1/64-1/9)	1/32 (1/82-1/13)	1/16 (1/46-1/6)
P value	<< 0.00190 pairwise comparisons presented below							

*The overall test for differences in cells frequencies between any of the groups.

Table 5.1 B. Extreme limiting dilution analysis (ELDA) pairwise tests for differences in HCT116 cell self-renewal frequencies

HCT116 parental cells				
Group 1	Group 2	Chisq	DF	Pr(>Chisq)
PRDM9 siRNA-5	PRDM9 siRNA-6	0.192	1	0.661
PRDM9 siRNA-5	PRDM9 siRNA-7	0	1	1
PRDM9 siRNA-5	PRDM9 siRNA-8	0.86	1	0.354
PRDM9 siRNA-5	HiPerFect-only	7.57	1	0.00593
PRDM9 siRNA-5	Negative control siRNA	7.57	1	0.00593
PRDM9 siRNA-5	Untreated cells	7.57	1	0.00593
PRDM9 siRNA-6	PRDM9 siRNA-7	0.192	1	0.661
PRDM9 siRNA-6	PRDM9 siRNA-8	0.240	1	0.624
PRDM9 siRNA-6	HiPerFect-only	5.36	1	0.0206
PRDM9 siRNA-6	Negative control siRNA	5.36	1	0.0206
PRDM9 siRNA-6	Untreated cells	5.36	1	0.0206
PRDM9 siRNA-7	PRDM9 siRNA-8	0.86	1	0.354
PRDM9 siRNA-7	HiPerFect-only	7.57	1	0.00593
PRDM9 siRNA-7	Negative control siRNA	7.57	1	0.00593
PRDM9 siRNA-7	Untreated cells	7.57	1	0.00593
PRDM9 siRNA-8	HiPerFect-only	3.37	1	0.0663
PRDM9 siRNA-8	Negative control siRNA	3.37	1	0.0663
PRDM9 siRNA-8	Untreated cells	3.37	1	0.0663
HiPerFect-only	Negative control siRNA	0	1	1
HiPerFect-only	Untreated cells	0	1	1
Negative control siRNA	Untreated cells	0	1	1

 * The table shows the Chi-Square statistic, degrees of freedom (DF), and the probability value (Pr > ChiSq) with a value below than 0.05 indicating a statistically significant difference.
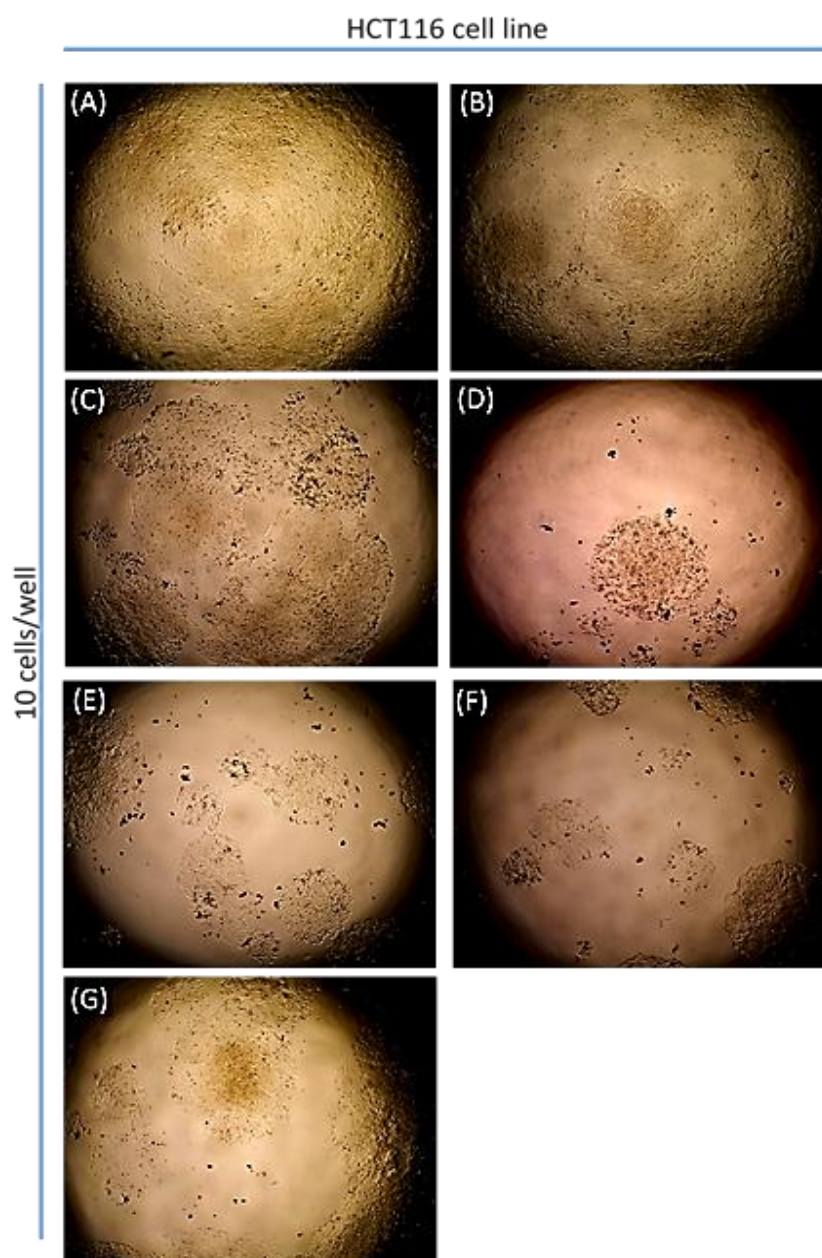
Figure 5.11. The influence of using different types of PRDM9 siRNA on HCT116 cell growth (10 cells seeded per well) after 10 days. Untreated HCT116 cells, cells transfected with non-interference siRNA and/or HiPerfect transfection only (served as a negative control) were used to assess the effect of using PRDM9 siRNA on cell growth. Four types of PRDM9 siRNA were used to test cell survival after PRDM9 knockdown. (A) Untreated HCT116 cells, (B) Cells transfected with non-interference siRNA, (C) HiPerfect transfection only, (D) siRNA-5 transfection, (E) siRNA-6 transfection, (F) siRNA-7 transfection, (G) siRNA-8 transfection. HCT116 cell growth was influenced by siRNA-5, siRNA-6 and siRNA-7 compared to untreated or negative control transfection.
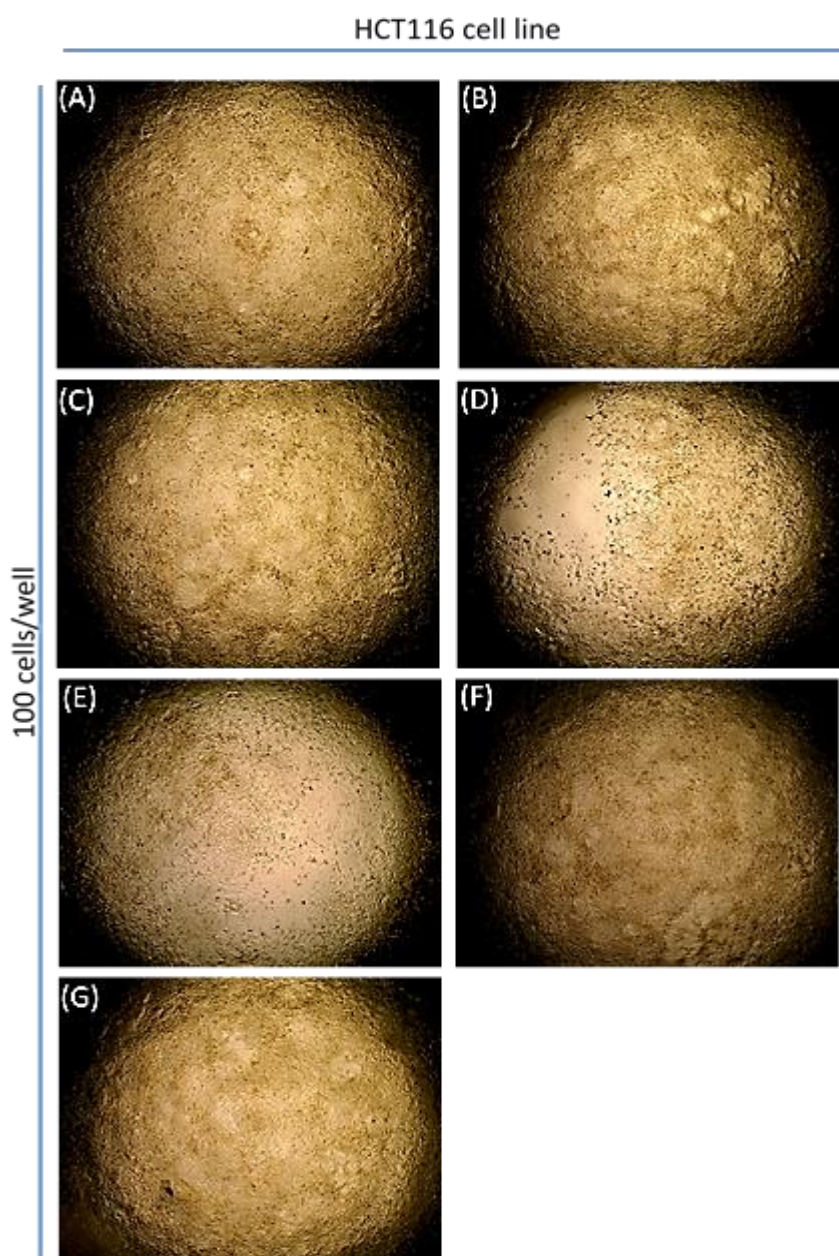
Figure 5.12. **The influence of using different types of PRDM9 siRNA on HCT116 cell growth (100 cells seeded per well) after 10 days.** Untreated HCT116 cells, cells transfected with non-interference siRNA and/or HiPerfect transfection only (served as a negative control) were used to assess the influence of using PRDM9 siRNA on cell growth. Four types of PRDM9 siRNA were used to test cell survival after PRDM9 knockdown. (A) Untreated HCT116 cells, (B) cells transfected with non-interference siRNA, (C) HiPerfect transfection only, (D) siRNA-5 transfection, (E) siRNA-6 transfection, (F) siRNA-7 transfection, (G) siRNA-8 transfection. HCT116 cell growth was slightly influenced by siRNA 5 and 6 compared to untreated or negative control transfection.

Table 5.2 A. Extreme limiting dilution analysis (ELDA) assay showing the influence of PRDM9 siRNA-mediated knockdown on proliferation of SW480 cells

Number of cells per well	Number of wells plated	Untreated cells	Negative control siRNA	HiPerFect only	*PRDM9 siRNA-5*	*PRDM9 siRNA-6*	*PRDM9 siRNA-7*	*PRDM9 siRNA-8*
			Number of well showing positive growth of SW480 cells after 10 days culturing					
1000	6	6	6	6	6	6	6	6
100	6	6	6	6	6	6	5	6
10	6	6	6	6	4	1	1	2
1	6	0	0	0	0	0	0	0
Cell forming frequency (95% CI)		1/4 (1/11-1/2)	1/4 (1/11-1/2)	1/4 (1/11-1/2)	1/11 (1/29-1/4)	1/32 (1/82-1/13)	1/57 (1/138-1/24)	1/23 (1/64-1/9)
P value	<< 0.0000275							

*The overall test for differences in cells frequencies between any of the groups.

Table 5.2 B. Extreme limiting dilution analysis (ELDA) pairwise tests for differences in SW480 cell self-renewal frequencies

SW480 parental cells				
Group 1	Group 2	Chisq	DF	Pr(>Chisq)
PRDM9 siRNA-5	siRNA-6	2.17	1	0.141
PRDM9 siRNA-5	siRNA-7	5.64	1	0.0176
PRDM9 siRNA-5	siRNA-8	1.07	1	0.302
PRDM9 siRNA-5	HiPerFect-only	1.75	1	0.186
PRDM9 siRNA-5	Negative control siRNA	1.75	1	0.186
PRDM9 siRNA-5	Untreated cells	1.75	1	0.186
PRDM9 siRNA-6	siRNA-7	0.701	1	0.402
PRDM9 siRNA-6	siRNA-8	0.192	1	0.661
PRDM9 siRNA-6	HiPerFect-only	7.57	1	0.00593
PRDM9 siRNA-6	Negative control siRNA	7.57	1	0.00593
PRDM9 siRNA-6	Untreated cells	7.57	1	0.00593
PRDM9 siRNA-7	siRNA-8	1.68	1	0.194
PRDM9 siRNA-7	HiPerFect-only	13.9	1	0.000198
PRDM9 siRNA-7	Negative control siRNA	13.9	1	0.000198
PRDM9 siRNA-7	Untreated cells	13.9	1	0.000198
PRDM9 siRNA-8	HiPerFect-only	5.36	1	0.0206
PRDM9 siRNA-8	Negative control siRNA	5.36	1	0.0206
PRDM9 siRNA-8	Untreated cells	5.36	1	0.0206
HiPerFect-only	Negative control siRNA	0	1	1
HiPerFect-only	Untreated cells	0	1	1
Negative control siRNA	Untreated cells	0	1	1

* The table shows the Chi-Square statistic, degrees of freedom (DF), and the probability value (Pr > ChiSq) with a value below than 0.05 indicating a statistically significant difference.
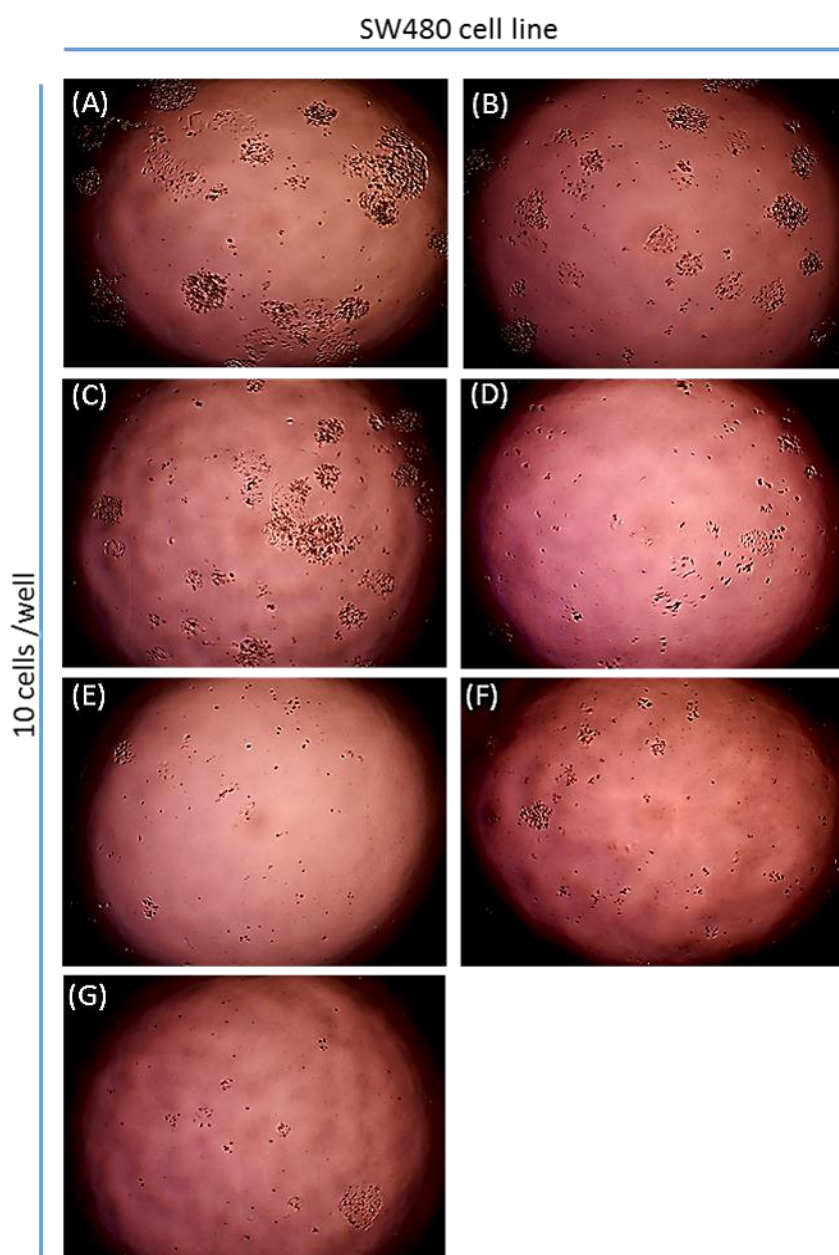
Figure 5.13. **The influence of using different types of PRDM9 siRNA on SW480 cell growth (10 cells seeded per well) after 10 days.** Untreated SW480 cells, cells transfected with non-interference siRNA and/or HiPerfect transfection only (served as a negative control) were used to assess the influence of using PRDM9 siRNA on cell growth. Four types of PRDM9 siRNA were used to examine cell survival after PRDM9 knockdown. (A) Untreated SW480 cells, (B) cells transfected with non-interference siRNA, (C) HiPerfect transfection only, (D) siRNA-5 transfection, (E) siRNA-6 transfection, (F) siRNA-7 transfection, (G) siRNA-8 transfection. SW480 growth was inhibited by siRNAs compared to untreated or negative control transfection.
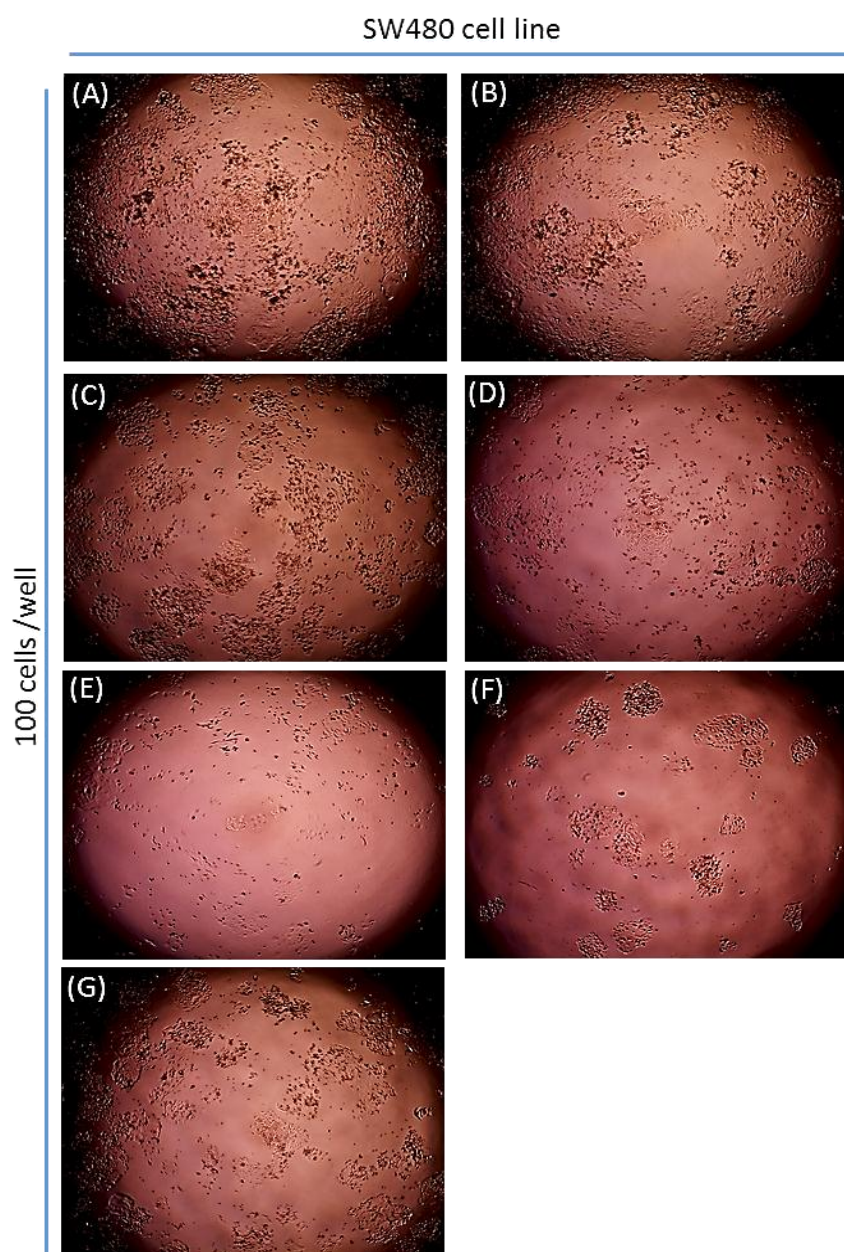
SW480 cell line

Figure 5.14. **The influence of using different types of PRDM9 siRNAs on SW480 cell growth (100 cells seeded per well) after 10 days.** Untreated SW480 cells, cells transfected with non-interference siRNA and/or HiPerfect transfection only (served as a negative control) were used to assess the effect of using PRDM9 siRNA on cell growth. Four types of PRDM9 siRNA were used to test cell survival after PRDM9 knockdown. (A) Untreated SW480 cells, (B) cells transfected with non-interference siRNA, (C) HiPerfect transfection only, (D) siRNA-5 transfection, (E) siRNA-6 transfection, (F) siRNA-7 transfection, (G) siRNA-8 transfection. SW480 growth was inhibited using siRNA-6 and siRNA-7 compared to untreated or negative control transfection.

5.2.6 Cloning and sequencing of full length variants of *PRDM9* and *PRDM7*

The full-length human *PRDM9* open reading frame (NCBI, Gene ID 56979) has a predicted size of 2685 bp and the short isoform of *PRDM7* (*II*) (NCBI, Gene ID 11105) has a predicted size of 516 bp, both were amplified from the cDNA of normal human testis RNA using Phusion high-fidelity PCR Master Mix with GC buffer (Figure 5.15). The correct *PRDM9* and *PRDM7* (II) bands were extracted from the gels, purified, and then digested with *Bam*HI restriction enzyme and purified.

Different sets of primers, phusions and PCR conditions were used to amplify the long isoform of *PRDM7* with estimated size 1479 bp but this gene could not be amplified due to the identical sequencing of the start codon and stop codon primers to the *PRDM9* sequence (see gene alignment Figure 5.2). Therefore, the full-length recombinant PUC57::*PRDM7* (Named pAMO5) long isoform (*I*) with *Bam*HI restriction sites was purchased (Genscript).

A digested (*Bam*HI) and purified pNEB193 plasmid was used to clone the *PRDM7 (II)* short isoform (Figure 5.16 A). Due to the size similarity of the *PRDM9* and pNEB193 plasmid, a digested and purified pGEM-T easy vector pGEM-3ZF(+) (3199 bp) was used to clone the full length of *PRDM9* (Figure 5.16 B).

The ratios of *PRDM9* and *PRDM7* (*II*) genes to the plasmids were used (6:1 and/or 9:1; insert: plasmid). Ampicillin resistance and blue/white screening with lac-Z selection were used for successful detection of ligated plasmids with *PRDM9* or *PRDM7 (I)* and *(II)* genes after transformation into the *E. coli* DH5α strain. Large white colonies were selected and PCR reactions were performed using internal primers for *PRDM9* or *PRDM7* isoforms (*I*) and *(II)* to confirm that the correct colonies were selected (Figure 5.17). Colonies with positive PCR were selected for overnight growing.

The primers of *PRDM9* and *PRDM7* genes had incorporated *Bam*HI sites, so the recombinants plasmids were checked also by restriction digestion of the purified plasmids using *Bam*HI to confirm correct cloning. Analysis of the digestion pGEM-3ZF (+)::*PRDM9* (Named pAMO1) in 0.8% agarose gel confirmed the presence of the insert *PRDM9* with the correct size of approximately 2685 bp (Figure 5.18). Similarly, digestion of pUC57::*PRDM7* (*I*)

(Named pAMO5) with *Bam*HI enzyme showed successful cloning of *PRDM7* (*I*) with 1479 bp
(Figure 5.19 A); and digestion of pNEB193::*PRDM7* (II) (Named pAMO6) indicated correct
cloning of *PRDM7* (*II*) with a single band of 516 bp (Figure 5.19 B).

Samples with the correctly cloned genes were sent for DNA sequencing to confirm the
correct orientation and to check for unwanted mutations. The colonies chosen for further
analysis did not contain any mutations. Universal primers and internal primers used for these
checks are listed in Tables 2.3 and 2.11 (Method and Materials).
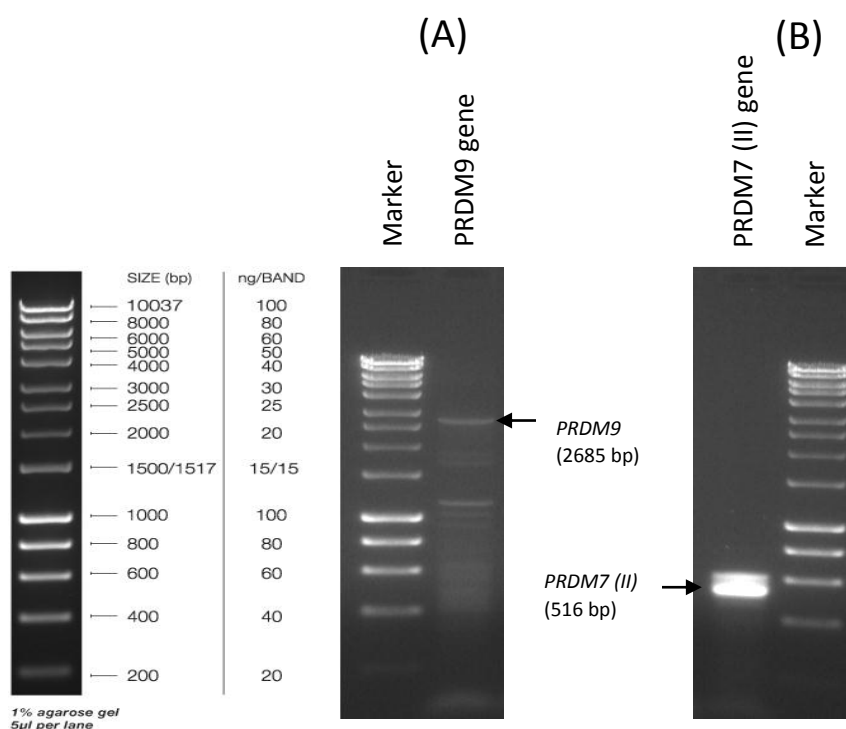
Figure 5.15. **PCR amplification of the full open reading frame of *PRDM9* and *PRDM7* (*II*) short
isoform.** Ethidium bromide stained agarose gels show the product of (A) *PRDM9* with 2685 bp and (B)
PRDM7 (II) short isoform with size 516 bp.

Figure 5.16. **Digestion of pNEB193 and pGEM-3Zf(+) plasmids with *Bam*HI restriction enzyme.** An 0.8% agarose gel stained with ethidium bromide shown in figure (A) pNEB193 plasmid with size 2713 bp, lane 1 digestion of pNEB193 before the purification and lane 2 after purification. Figure (B) showing pGEM-3ZF (+) plasmid. Lane 1: uncut pGEM-3ZF (+) and lane 2: digested and purified plasmid. 5 μl of HyperLadder 1 kb was used as marker.
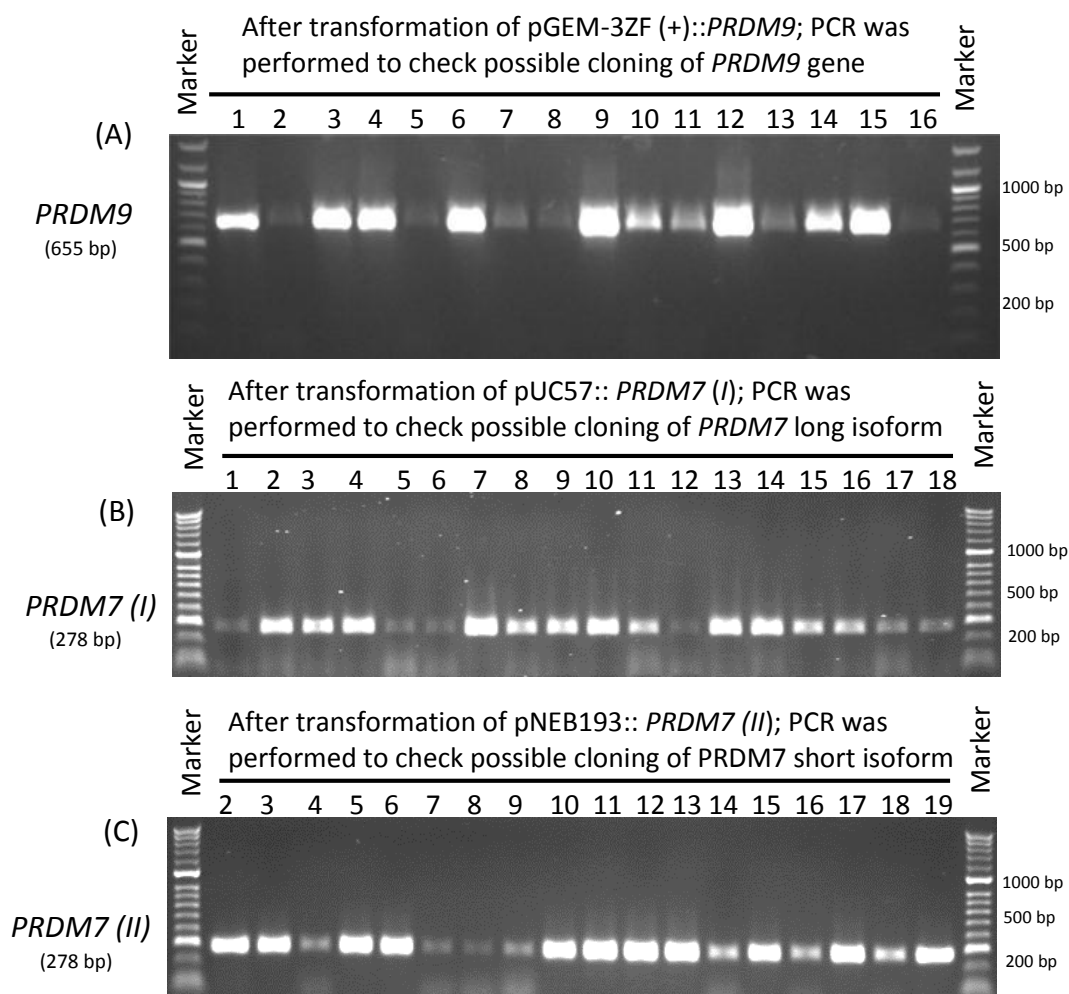
Figure 5.17. **PCR profile analyses both *PRDM9* and *PRDM7* genes after transformation into *E.coli*.** *PRDM9* and *PRDM7* (*II*) were ligated at *Bam*HI site with ratio of (1:6 and/or 1:9 plasmid: insert) as well as purchased *PRDM7* (*I*):PUC57 transformed into DH5α *E.coli*. Random large white colonies were picked using the blue/white screening technique. PCR was performed to check the possibility of transformation using internal primers for *PRDM9* and *PRDM7* genes. PCR products were fractionated on 1% agarose gels stained with ethidium bromide. (A) *PRDM9* gene was cloned into pGEM-3ZF (+) plasmid; Lane 1: positive testis control; Lanes 2-16: colonies containing *PRDM9* gene. (B) *PRDM7* gene long isoform was cloned into pUC57 plasmid; and (C) *PRDM7* gene short isoform was cloned into pNEB193 plasmid; Lane 1: positive testis control; Lanes 2-18 or 19: colonies containing *PRDM7* genes in (B) long isoform and (C) short isoform. 5 μl of HyperLadder II was used as marker.
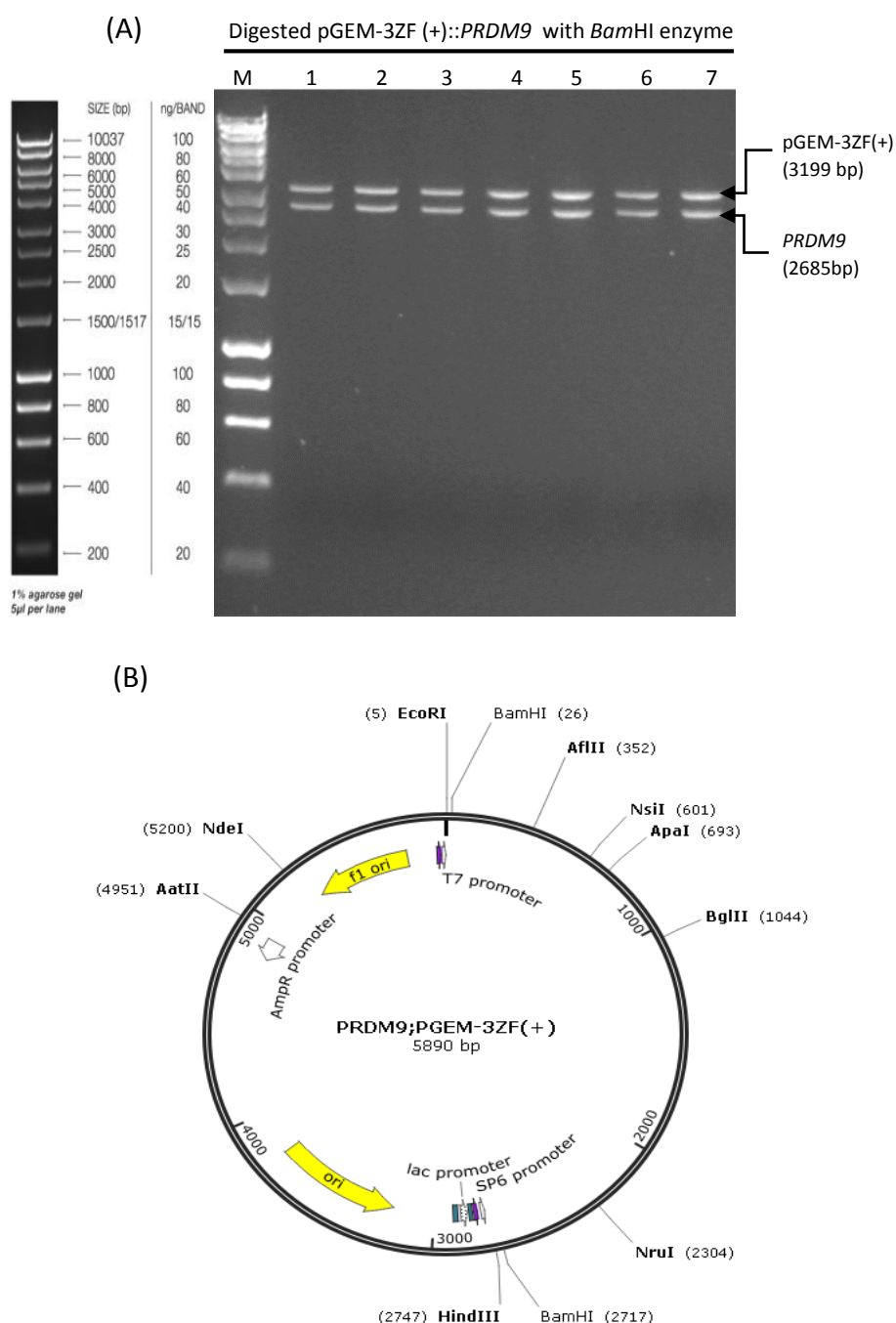
Figure 5.18. Digestion of pGEM-3ZF (+)::*PRDM9* (pAMO1) construct with *Bam*HI restriction enzyme and construct map. (A) 0.8% agarose gel stained with ethidium bromide shows: in Lane 1-7, digestion of pGEM-3ZF (+)::*PRDM9* construct with the *Bam*HI enzyme. Two bands were obtained in all seven colonies, the upper one being the pGEM-3ZF (+) vector with approximately 3199 bp. The lower band is the *PRDM9* gene with approximately 2685 bp. (B) pGEM-3ZF (+)::*PRDM9* recombinant plasmid showing the restriction site of *Bam*HI used to clone the *PRDM9* gene. 5 µl of HyperLadder 1 kb was used as marker.
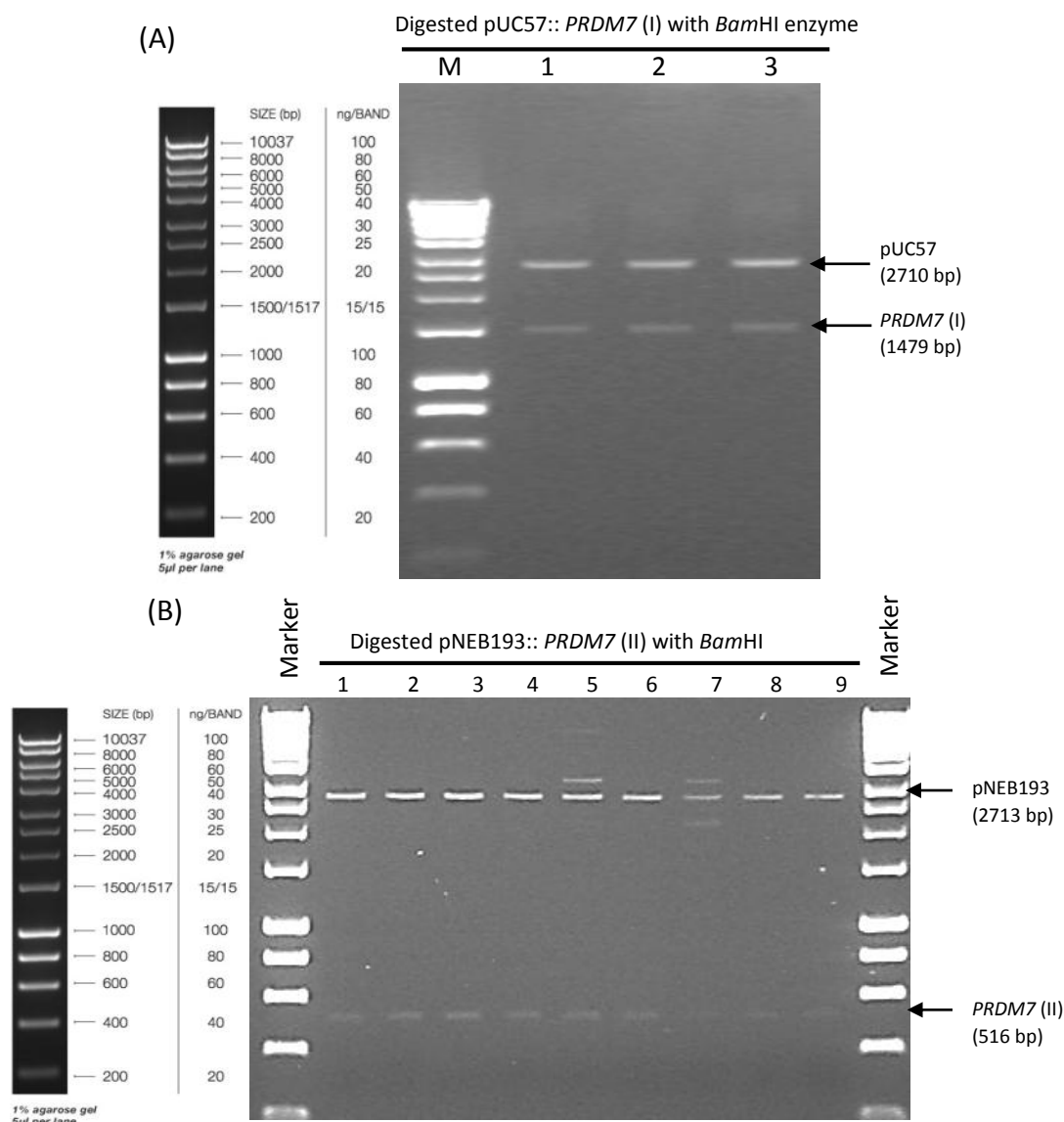
Figure 5.19. **Digestion of constructed plasmids of *PRDM7* long and short isoforms with *BamHI*** **restriction enzyme.** 0.8% agarose gels stained with ethidium bromide show in (A) *PRDM7 (I)* long isoform, Lane 1-3: digestion of pUC57:: *PRDM7 (I)* (pAMO5) construct with *Bam*HI enzyme. Two bands were obtained in the three colonies, the upper one is the pUC57 vector with approximately 2710 bp. The lower band is the *PRDM7 (I)* gene with approximately 1479 bp. (B) Digestion of pNEB193:: *PRDM7 (II)* (pAMO6) with *Bam*HI enzyme gave two bands shown in lanes 1-9, which correspond to the pNEB193 plasmid (upper band, 2713 bp) and the *PRDM7* (II) gene (lower band, 516 bp). 5 µl of HyperLadder 1 kb was used as marker.

5.2.7 Subcloning of *PRDM9* and *PRDM7* genes into mammalian expression system Tet-on 3G plasmid.

One of the central hypotheses in this chapter that was PRDM9 acts as a transcriptional factor that activates the expression of meiosis genes in cancer cells. This hypothesis was examined by overexpressing *PRDM9*, *PRDM7* (*I*) and *PRDM7* (*II*) in the HeLa Tet on-3G system.

The Tetracycline-Inducible Gene Expression Systems (Tet-On 3G) is one of the most common, inducible gene expression systems in mammalian cells. In this system, a high level of transgenes can be expressed in targeted cells under the regulation of a Tet-On 3G transactivator protein. The TRE3G promoter (P_{TRE3G}) controls the level of target gene expression in the presence or absence of doxycycline (DOX) (Gossen and Bujard, 1992; Vigna *et al.*, 2002; Loew *et al.*, 2010).

Here, the full lengths of *PRDM9*, *PRDM7* (*I*) and *PRDM7* (*II*) were amplified from the constructed plasmids generated in Section 5.2.6 using Phusion high-fidelity PCR. The primers contained a Kozak sequence (ACC) in the forward primers and a 6X His tag in the reverse primers, in addition to *Bam*HI sites (Figure 5.20). The digested and purified pTRE3G plasmid with *Bam*HI restriction enzyme was used to clone *PRDM9*, *PRDM7* (*I*) and *PRDM7* (*II*) individually (Figure 5.21).

The ratios of insert to plasmid used were 6:1 and/or 9:1 (inserts: plasmid). Large white colonies were selected after the ligated plasmids were transformed into *E*. *coli* under ampicillin resistance selection growth. PCR reactions were performed to confirm that the correct colonies were selected (Figure 5.22). The pTRE3G::*PRDM9* (Named pAMO7) vector was digested using *Bam*HI restriction enzyme (Figure 5.23), as were pTRE3G::*PRDM7*(*I*) (Named pAMO8) or pTER3G:: *PRDM7* (*II*) (Named pAMO9) (Figure 5.24). Samples with the correctly cloned genes were sent for DNA sequencing to confirm the correct orientation and check for unwanted mutations. The colonies chosen for further analysis did not contain any mutations.
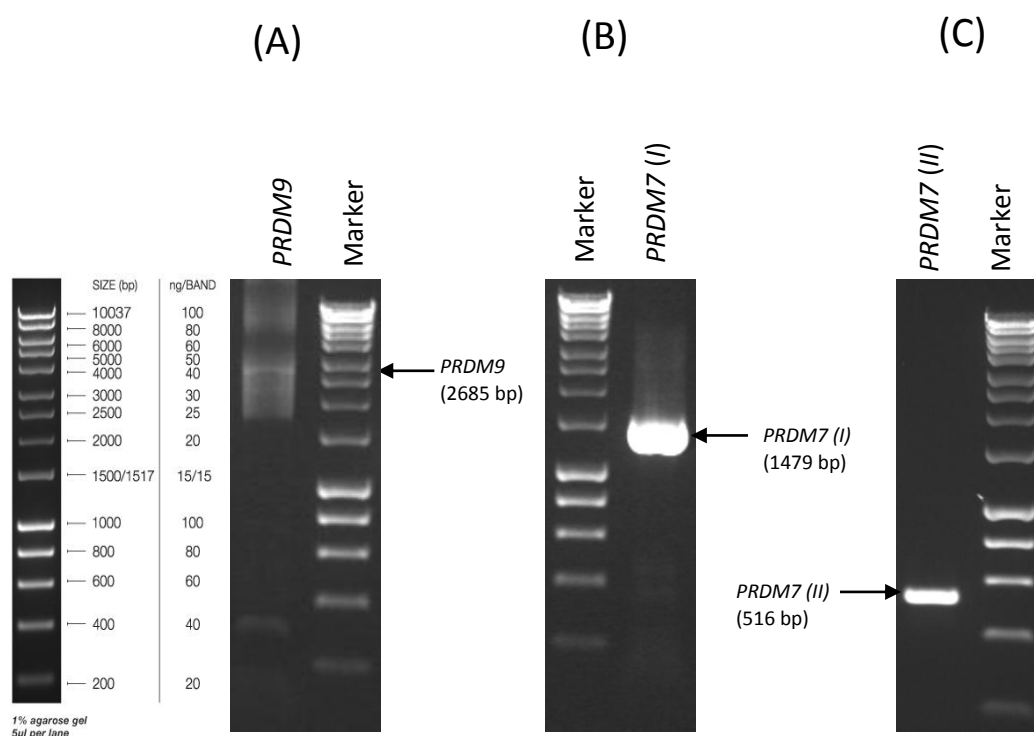
Figure 5.20. **PCR amplification of the full open reading frame of *PRDM9* and *PRDM7* genes using constructed plasmids.** Ethidium bromide stained agarose gels show the product of (A) *PRDM9* with 2685 bp (B) *PRDM7(I)* long isoform with 1479 bp;. and (C) *PRDM7 (II)* short isoform with 516 bp;. Kozak sequencing (ACC) was added to the forward primers before the start codon in the N-terminal and 6X His tag were added to the C-terminal in the reverse primers before the stop codon.
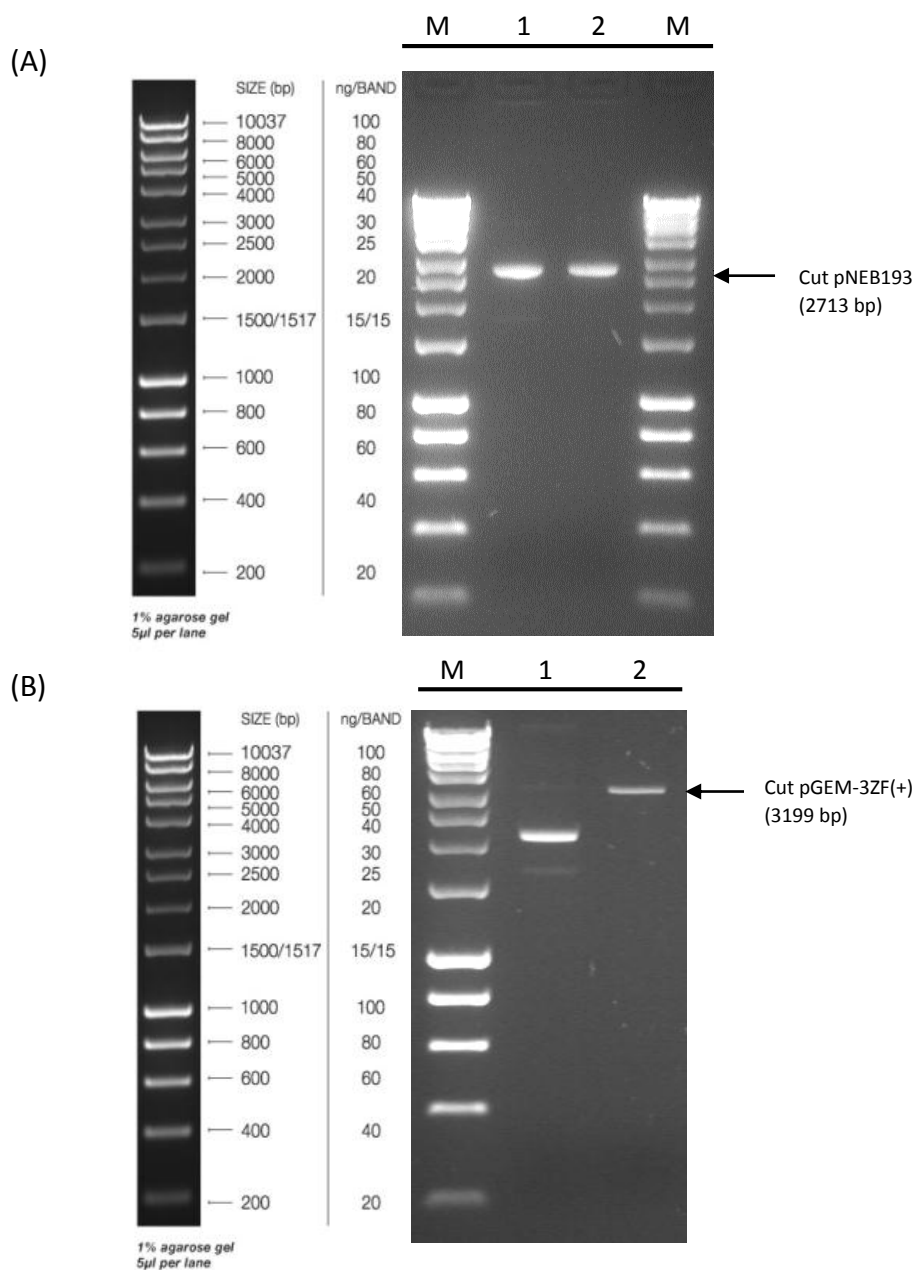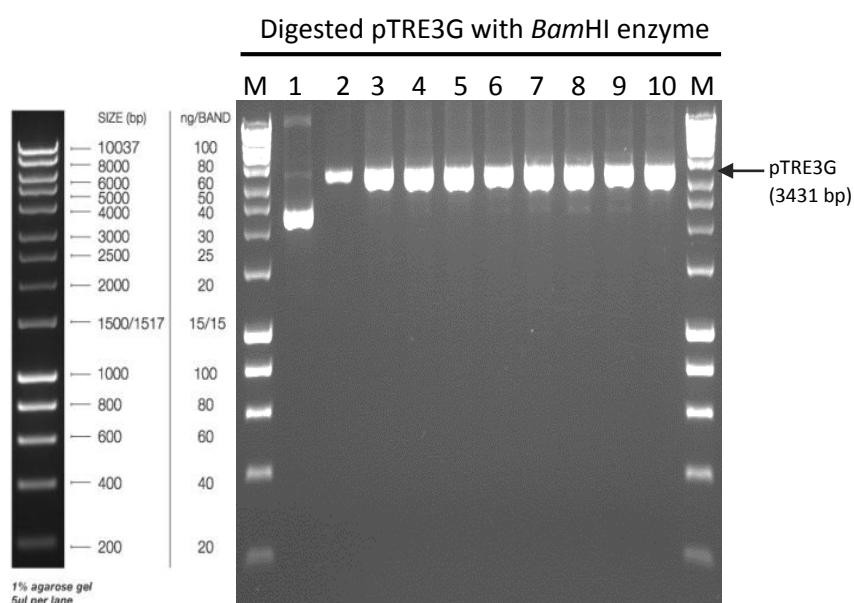
Figure 5.21. **Digestion of pTRE3G plasmid with *BamHI* restriction enzyme.** 0.8% agarose gel stained with ethidium bromide shows in Lane 1: uncut *pTRE3G* vector; Lane 2: cut and purified pTRE3G vector; Lanes 3-10: digested pTRE3G plasmid before purification. Lane M is Hyperladder I.
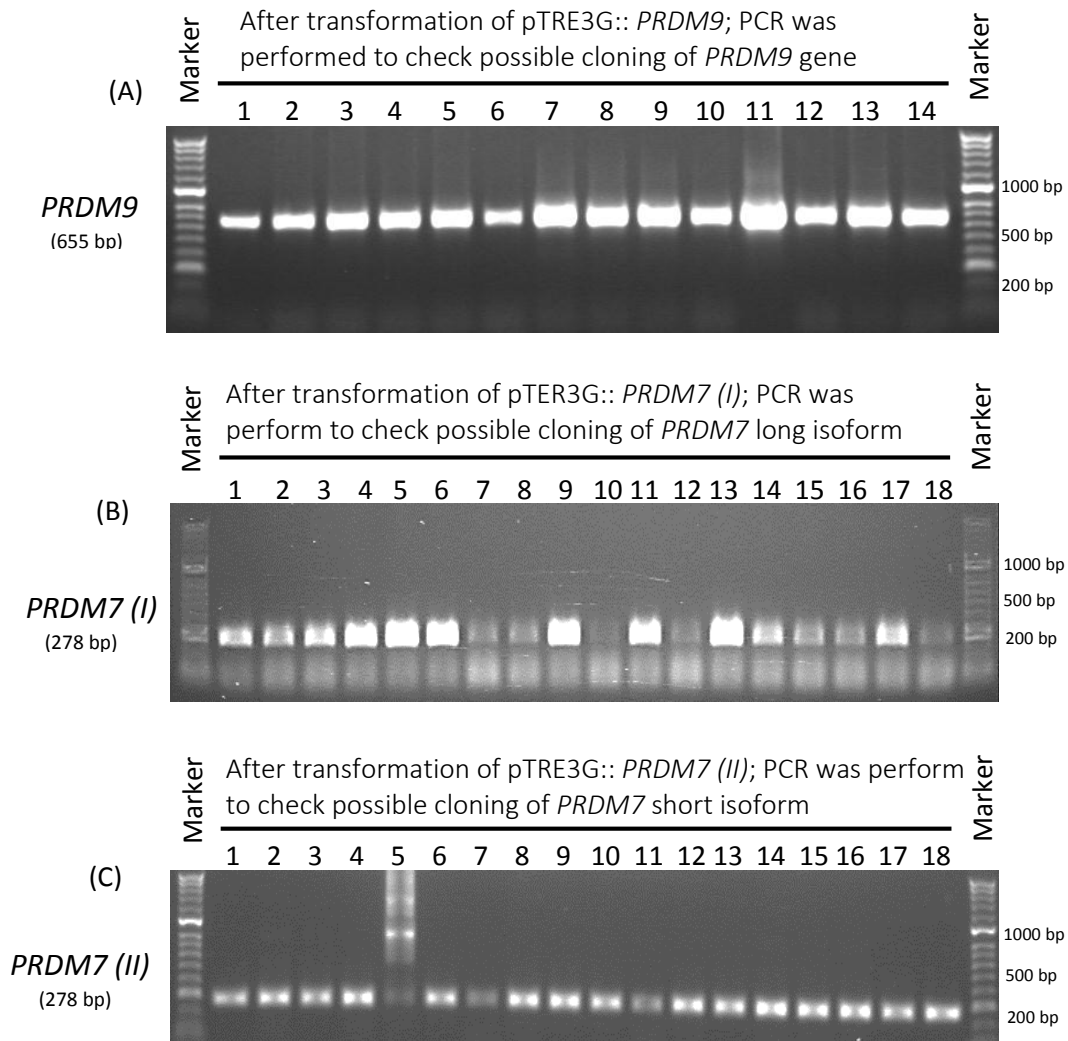
Figure 5.22. **PCR profile analyses of *PRDM9* and *PRDM7* genes after cloning into *pTRE3G* and transformation into *E. coli*.** Genes were ligated with *pTRE3G* plasmid at ratio of 1:6 and/or 1:9 (plasmid: insert) and transformed into *E. coli*. Random large white colonies were picked and PCR was performed to check the possibility of transformation. PCR products were fractionated on 1% agarose gels stained with ethidium bromide. (A) *PRDM9* gene; Lane 1: positive testis control; Lanes 2-14: colonies containing *PRDM9* gene. (B) *PRDM7* gene long isoform; and (C) *PRDM7* gene short isoform; Lane 1: positive testis control; Lanes 2-18: colonies containing *PRDM7* genes in (B) long isoform and (C) short isoform.
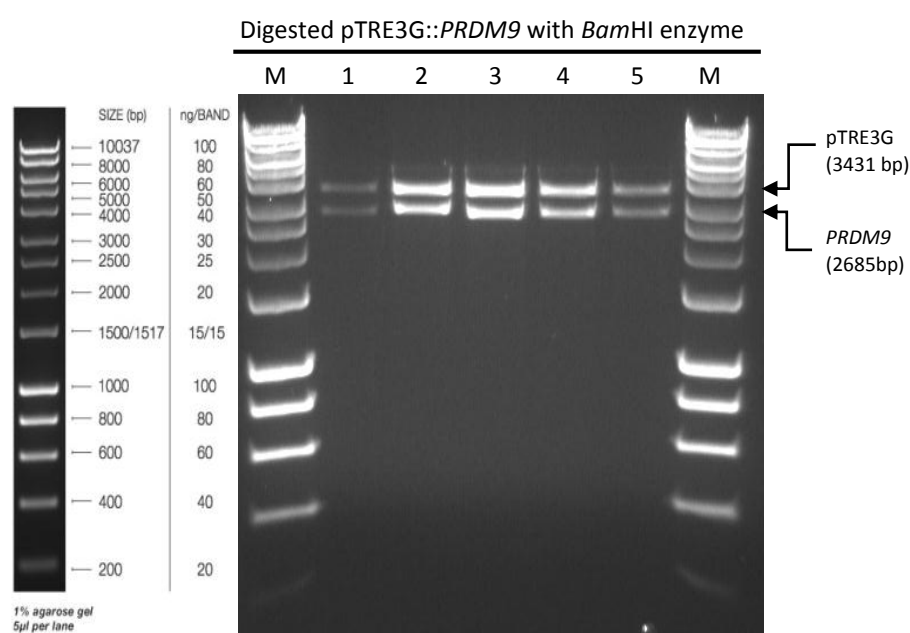
Figure 5.23. Digestion of *pTRE3G::PRDM9* (pAMO7) construct with *Bam*HI restriction enzyme. 0.8% agarose gel stained with ethidium bromide shows in Lane 1-5: digestion of pTRE3G::*PRDM9* construct with *Bam*HI enzyme. Two bands were obtained in all five colonies higher one belongs to pTRE3G vector with approximately 3431 bp. The lower band is the *PRDM9* gene with approximately 2685 bp. A 6X His tag sequence was added to the reverse primers and a Kozak sequence to the forward primer. M is a HyperLadder 1 kb marker.
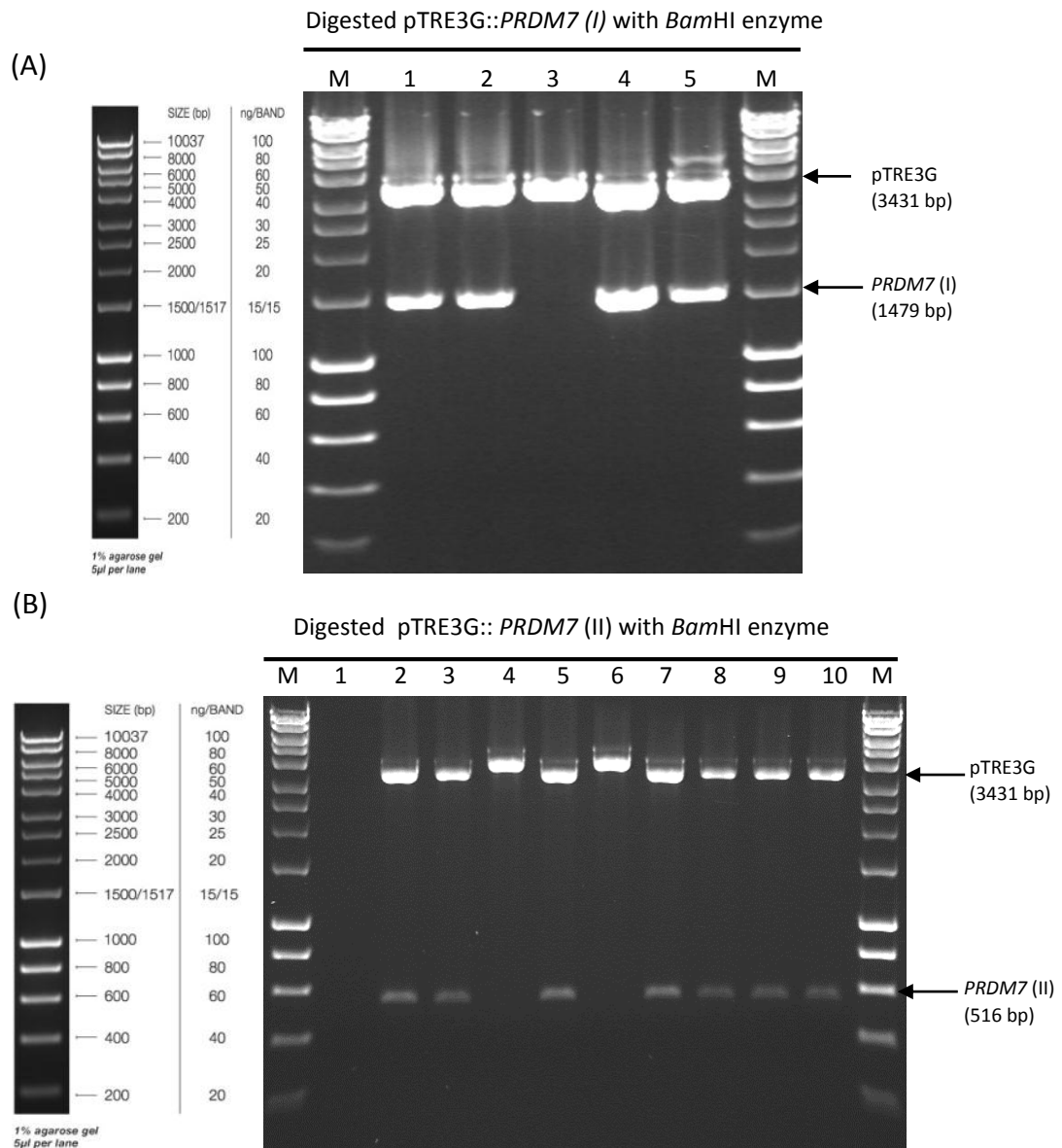
Figure 5.24. Digestion of pTRE3G::*PRDM7* constructed plasmids with *BamHI* restriction enzyme. 0.8% agarose gels stained with ethidium bromide show in (A) *PRDM7 (I)* long isoform, Lane 1-5: digestion of pTRE3G::*PRDM7 (I)* (pAMO8) construct with *Bam*HI enzyme. Two bands were obtained in four colonies; the upper one is the pTRE3G vector with approximately 3431 bp. The lower band is the *PRDM7(I)* gene with approximately 1479 bp.(B) Digestion of pTRE3G::*PRDM7 (II)* (pAMO9) with *Bam*HI enzyme. Two bands were obtained in lanes 2, 3, 5, 7, 8, 9 and 10, which correspond to the pTRE3G plasmid and *PRDM7*(*II*) gene. A His tag sequence was added to the reverse primers and a Kozak sequence to the forward primer. M is a HyperLadder 1 kb marker.

5.2.8 Establishment of a double-stable HeLa Tet-On 3G cell line

The generation of a double stable Tet-on 3G cell line containing *PRDM9*, *PRDM7* (*I*) or *PRDM7* (*II*) individually was one of the major parts of this research project. Creation of such a system for overexpression of *PRDM9* and *PRDM7* in human cancer cells was aimed at studying the transcription factor activities (TFs) of these genes in cancer cells. *PRDM9* was identified to be aberrantly expressed in several cancer cells and was hypothesised to drive the expression of several meiosis genes in cancer cells (Feichtinger *et al.,* 2012a).

In the Tet-on 3G system, two steps are required to generate a double stable cell line expressing *PRDM9* or *PRDM7* genes. In the first transfection, cells are transfected with pCMV-Tet3G plasmid to constitutively express the transactivator protein under the G418 antibiotic selection. Here, HeLa Tet-on 3G and Jurkat Tet-on 3G cells containing the pCMV-Tet-3G vector were purchased from Clonetech and cDNA was generated from each cell type. PCR was carried out to analyse the expression of *PRDM9* and *PRDM7* in these cells. No expression was observed for either *PRDM9* or *PRDM7* in HeLa Tet-on 3G and Jurkat Tet-on 3G cells (Figure 5.25).

The double-stable cell line was created by cotransfecting HeLa Tet-on 3G individually with pTRE3G::*PRDM9*, pTRE3G::*PRDM7*(*I*) or pTRE3G::*PRDM7*(*II*) plasmids along with a linear selection marker for puromycin antibiotic. This was followed by selection for double-stable transfected cells that were resistant to puromycin and G418 antibiotics (Figure 5.26 A and B).

5.2.8.1 Selection of double-stable HeLa Tet-on 3G cells

Before HeLa Tet-on 3G cell cotransfection, a cell kill curve was optimised using different concentrations of puromycin antibiotic. Cells were grown in media without antibiotic for 48 hours and then exposed to 0 -2.5 µg/ml puromycin for 3-5 days (Figure 5.27). The minimum antibiotic dose that killed all cells in 5 days was 1.0 µg/ml.

The HeLa Tet-On 3G cell line was cotransfected with pTRE3G plasmid containing either *PRDM9*, *PRDM7*(*I*) or *PRDM7* (*II*) genes or with empty pTRE3G plasmid to assess the gene inductions, along with the puromycin linear selection marker, at ratios of 20:1 (constructed plasmids:puromycin marker). After 4 days of cells growth in antibiotic free media, 100 µg/ml of G418 and 1 µg/ml puromycin were added. Most of the cells died, with the exception of

single cells that appeared after 4 days of treatment and had possible transfection (Figure 5.28). After two weeks of cell culture in media containing optimised antibiotic concentrations, drug-resistant colonies begin to appear (Figure 5.29). Large and healthy colonies with possible gene cloning were picked up using cloning cylinders and grown in separate 10 cm plates and then in T75 flasks.

Each clone was then tested individually by comparing the expression of *PRDM9* or *PRDM7* in the presence or absence of 1 µg/ml doxycycline. Cells were grown in this experiment in a high quality Tet system approved FBS which was functionally tested for this inducible system and free of tetracycline contamination.
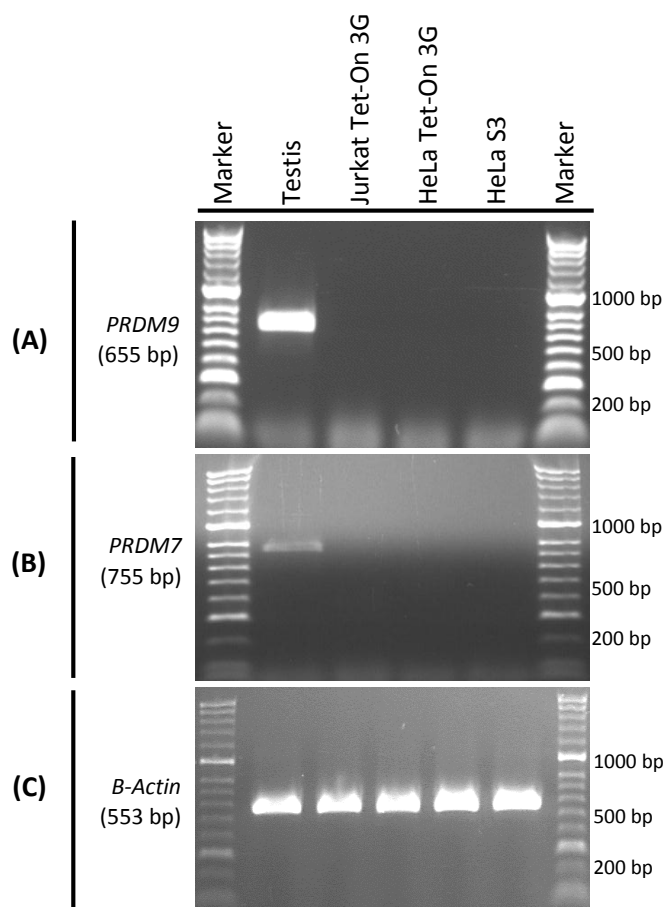
Figure 5.25. **RT-PCR of analysis expression of *PRDM9* and *PRDM7* in Jurkat Tet -on 3G, HeLa Tet - on 3G and HeLa S3 cell lines**. Agarose gels stained with ethidium bromide showed no expression of (A) *PRDM9* and (B) *PRDM7* in either the Jurkat or HeLa cancer cell lines. A testis sample was used as a positive control.(C) *β-Actin* served as a positive control for the cDNA samples.
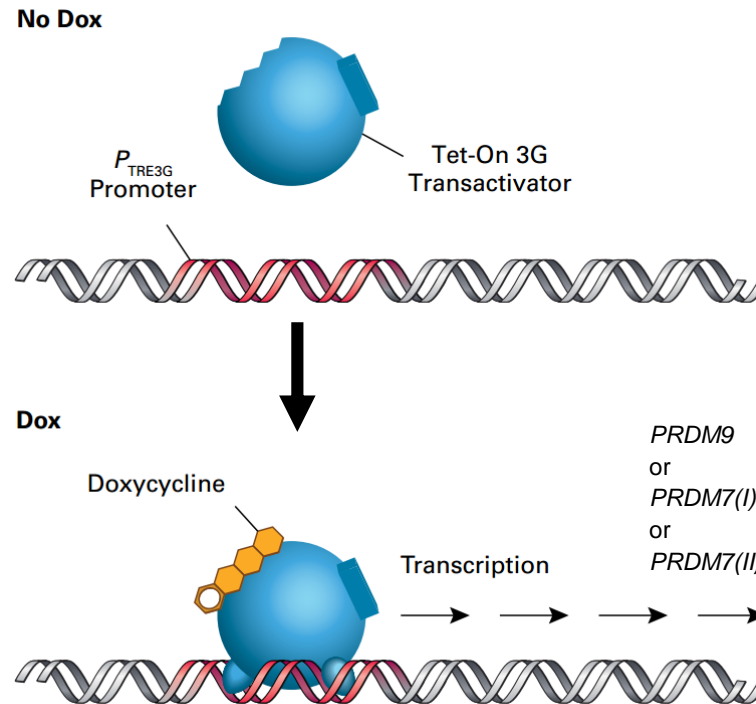
Figure 5.26 A. **The Tet-on 3G inducible system allow to induce the expression of *PRDM9* or *PRDM7* in the presence of doxycycline only.** Tet-on 3G transactivator protein (blue) undergoes a conformational alteration in the presence of doxycyline. This will allow this protien to bind to the tet operator (tetO) element (red) within the promoter P_{TRE3G} of TRE3G plasmid. This in turn induce the expression of PRDM9 or PRDM7.
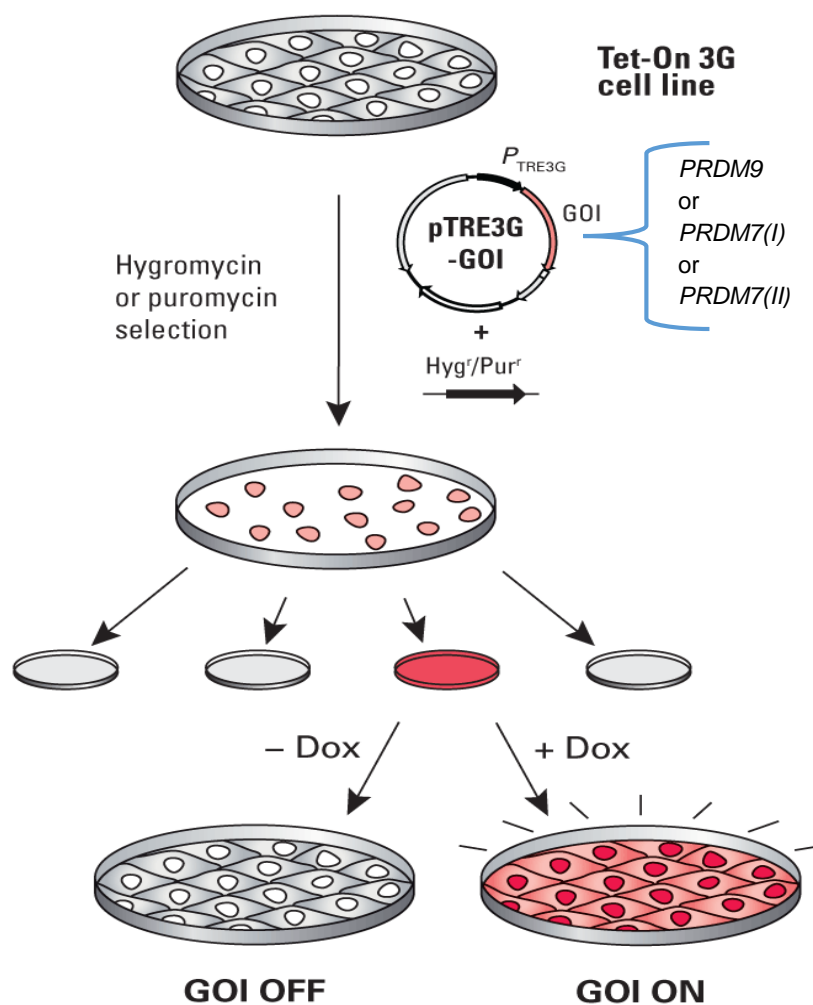
Figure 5.26 B. **Diagram summarising the methods of creating a double Tet on 3G cell line.** A HeLa Tet-on 3G line that expresses the tetracycline (Tet)-regulated transactivator protein was used to induce the expression of *PRDM9* and PRDM7 under the control of a tetracycline-responsive promoter. The constructed pTRE3G plasmids containing genes of interest were transfected to HeLa Tet-on 3G cells along with the puromycin linear marker. Doxycycline (Dox) was used for overexpression of *PRDM9* or *PRDM7* expression in HeLa Tet-on 3G when it was added to the cells.
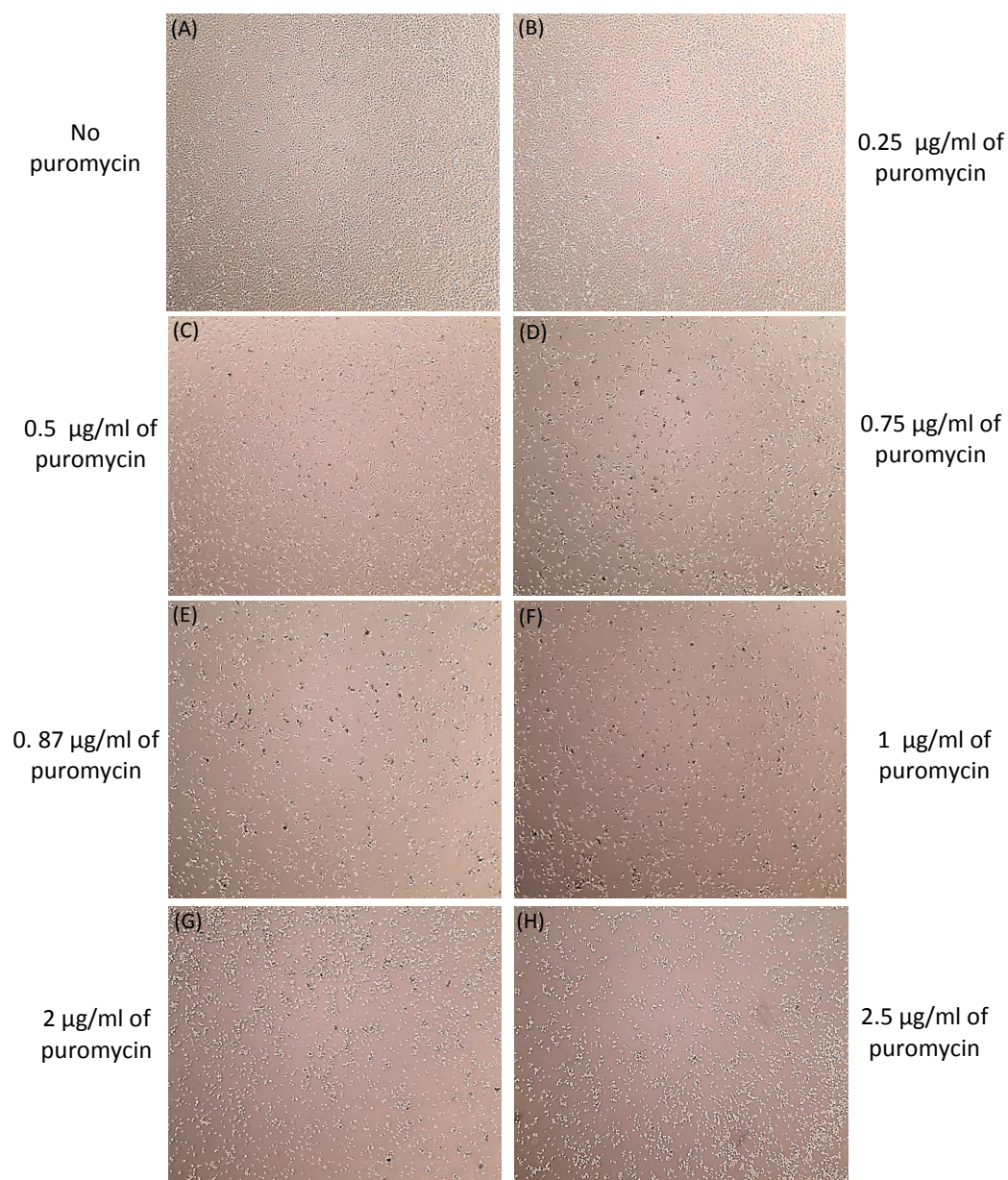
Figure 5.27. **Untransfected HeLa Tet-On 3G cells were exposed to a range of puromycin antibiotic dosages to generate a cell killing curve.** Appropriate doses of puromycin were determined by growing untransfected cells in normal media without puromycin antibiotic for 48 hours. Seven doses of puromycin (0 to 2.5 µg) were used then to optimize the minimum dose that would kill all cells after 3-5 days. (A) Untreated cells were used to compare the effect of the antibiotic on treated cells. From (B) to (H) different concentrations of puromycin antibiotic and the optimum dose based on colony selection was (D) 1 µg/ml.

PRDM9 transfection   PRDM7(I) transfection

PRDM7(II) transfection   pTRE3G vector transfection
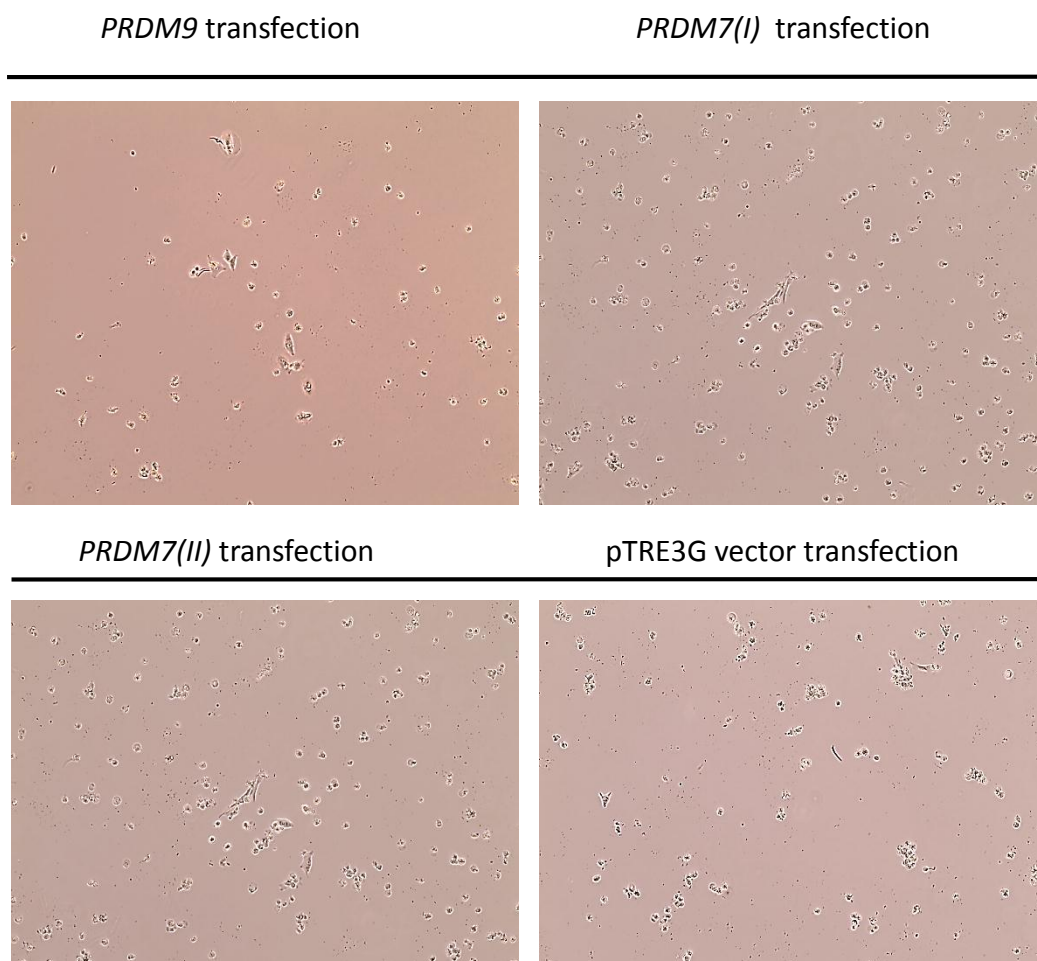
Figure 5.28. HeLa Tet-On 3G cell line transfected with *PRDM9, PRDM7(I)* or *PRDM7 (II)* genes after 4 days of cell growth. Cells were transfected with pTRE3G plasmid containing either *PRDM9*, *PRDM7(I)* or *PRDM7(II)* along with the puromycin linear selection marker at a ratio of 20: 1 (construct: puromycin marker). Empty pTRE3G plasmid was transfected into the cells to assess the gene induction. After 4 days of growth, most of the cells had died except the cells with possible transfection. Single cells started to appear 4 days after adding 1 µg/ml puromycin antibiotic.

PRDM9 transfection   *PRDM7(I)* transfection

PRDM7(II) transfection   pTRE3G vector transfection
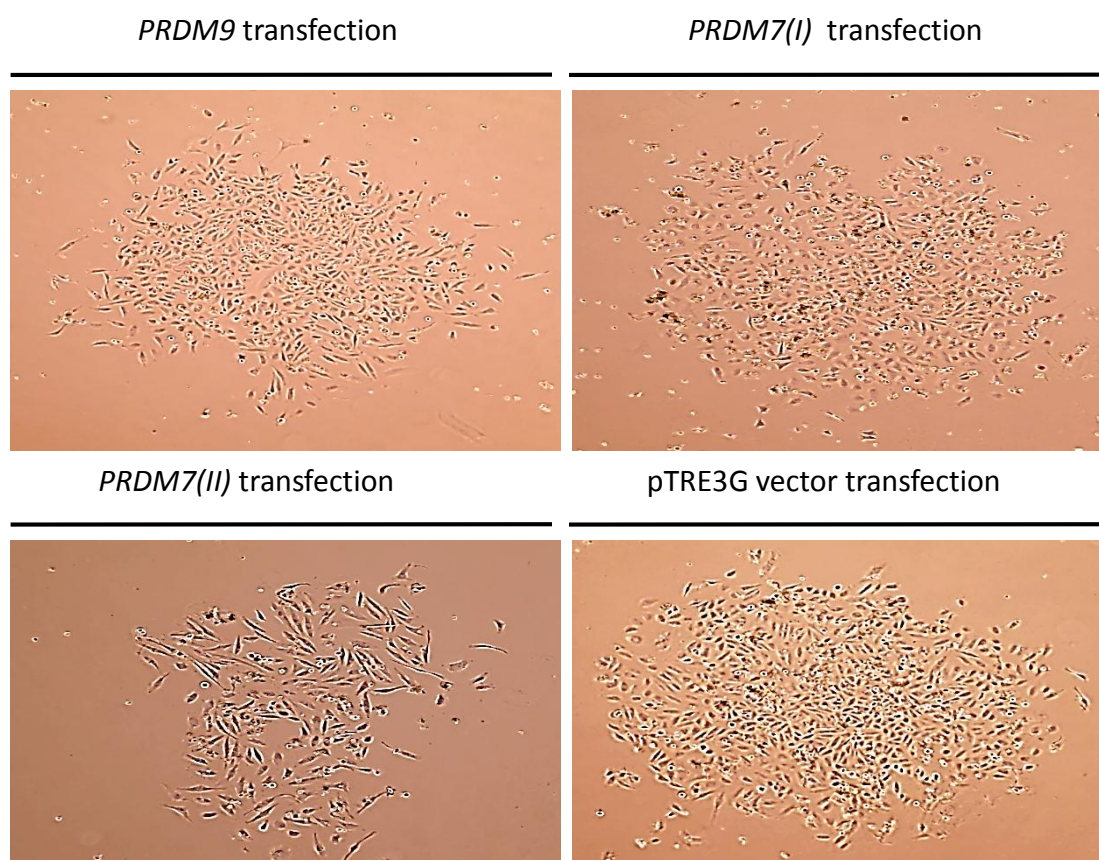
Figure 5.29. HeLa Tet-On 3G cell line transfected with *PRDM9*, *PRDM7(I)* or *PRDM7 (II)* genes after 16 days of cell growth.** Cells were transfected with pTRE3G plasmid containing either the *PRDM9*, *PRDM7(I)* long isoform or *PRDM7(II)* short isoform along with a puromycin linear marker at ratio of 20: 1 (construct: puromycin marker vector). Empty pTRE3G plasmid was transfected into the cells to assess the gene induction. After 16 days of growth, single colonies start to appear and were cultured in the presence of 1 µg/ml puromycin antibiotic.

5.2.9 RT-PCR and qRT-PCR analysis of overexpressed *PRDM9* into double-stable HeLa Tet-on 3G cells

The successful integration of *PRDM9* into HeLa Tet-on 3G cells to create a double-stable cell line was validated by cotransfecting four independent colonies with pTRE3G::*PRDM9* vector and inducing them with 1 µg/ml doxycycline for 24 hours. A fraction from each clone was also grown in the absence of doxycycline. Total RNAs were isolated and cDNAs were synthesized from each clone grown in either the presence or absence of doxycycline.

Two colonies showed expression of *PRDM9* after cells induction with doxycycline. In contrast, no expression was observed in the absence of doxycycline (Figure 5.30). Colonies (C) and (D) show no expression in the presence or absence of doxycycline and were excluded from this study. Two sets of *PRDM9* primers were used to confirm *PRDM9* expression and testis cDNA was used as a positive control.

PRDM9 expression was analysed in two independent colonies following integration of the pTRE3G vector only. No expression of *PRDM9* was displayed in transfected HeLa Tet-on 3G cells with pTRE3G plasmids either in presence or absence of doxycycline after 24 hours (Figure 5.31).

For further analysis, qRT-PCR was carried out to quantify the level of *PRDM9* expression in colony (A) and (B) after induction with doxycycline. Commercial (Qiagen) and in house designed qRT-PCR primers for *PRDM9* were used to carry out SYBR Green-based real time. *PRDM9* result was normalised against the qRT-PCR results of two reference genes *Actin* and *GAPDH* and the results are displayed in Figure 5.32. The qRT-PCR results showed no expression of *PRDM9* was observed in the absence of doxycycline. In contrast, a high level of expression of *PRDM9,* particularly in clone (B), was observed after doxycycline induction. No quantification cycle (Cq) reading was shown in the negative controls NTC (no template control) or NRT (no reverse transcriptase).

Figure 5.30. **RT-PCR analysis confirming expression of *PRDM9* gene after induction in HeLa Tet-on 3G system**. Four independent colonies of *PRDM9* into HeLa Tet-on 3G were induced with 1 µg/ml doxycycline for 24 hours. (I) β-*Actin* served as a positive control for the cDNA samples. (II) The expression profile of the *PRDM9* gene in the presence or absence of doxycycline. Two pairs of primers were designed for *PRDM9* gene and it was expressed in colonies (A) and (B) in the presence of doxycycline. In the absence of doxycycline, no expression was observed. Colonies (C) and (D) showed no expression in the presence or absence of doxycycline. Lane T: positive testis control. 1% agarose gels stained with ethidium bromide were used to visualise the PCR products.

Figure 5.31. **RT-PCR analysis of *PRDM9* gene in two independent colonies transfected with only pTRE3G empty vector.** HeLa Tet-on 3G cells were transfected with pTRE3G plasmids and induced with 1 µg/ml doxycycline for 24 hours. RT-PCR was performed to check the expression of the *PRDM9* gene after induction with only the pTRE3G vector. (II) No expression was determined for the *PRDM9* gene in both colonies, either with or without doxycycline. (I) β-*Actin* served as a positive control for the cDNA samples. Lane T: positive testis control. 1% agarose gels stained with ethidium bromide were used to visualise the PCR products.

(I)
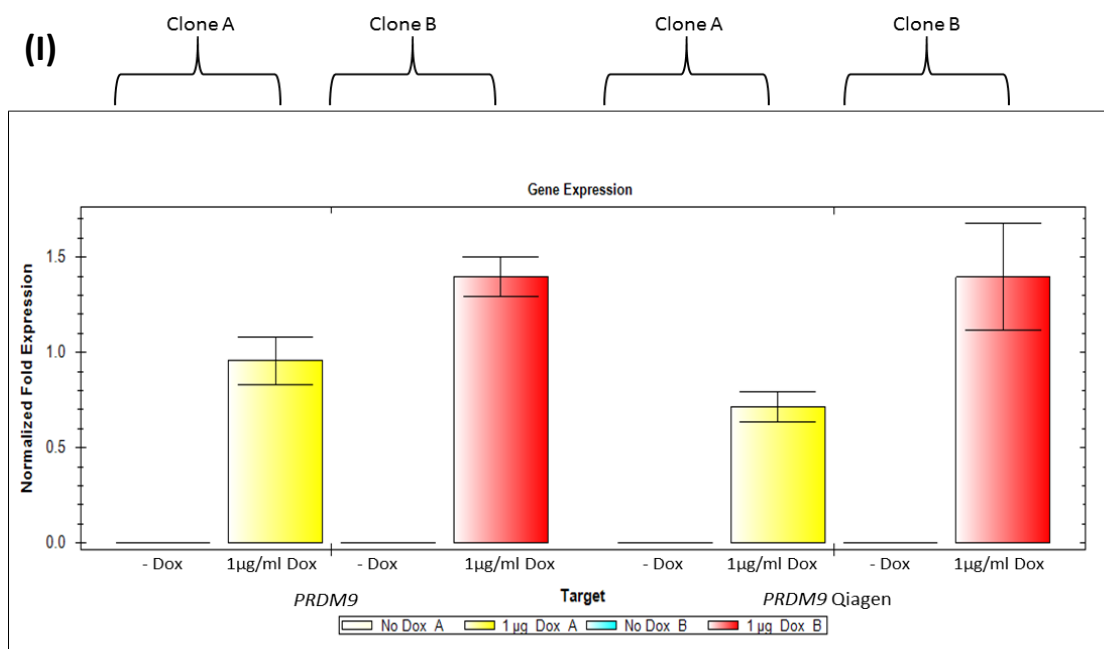
(II)

Target	Sample	Cq Reading 1	Cq Reading 2	Cq Reading 3	Cq Mean	Cq Std. Dev
Actin	No Dox A	17.74	17.75	17.79	17.76	0.025
Actin	1 µg Dox A	18.50	18.10	18.03	18.21	0.256
Actin	No Dox B	18.09	18.14	18.10	18.11	0.029
Actin	1 µg Dox B	18.57	18.51	18.44	18.50	0.065
GAPDH	No Dox A	20.47	20.45	20.37	20.43	0.053
GAPDH	1 µg Dox A	20.49	20.25	20.48	20.41	0.136
GAPDH	No Dox B	20.38	20.18	20.32	20.29	0.105
GAPDH	1 µg Dox B	20.39	20.54	20.27	20.40	0.137
PRDM9 **Qigene**	No Dox A	38.84	36.40	37.36	37.53	1.231
PRDM9 **Qigene**	1 µg Dox A	26.06	26.51	26.13	26.23	0.242
PRDM9 **Qigene**	No Dox B	37.94	37.64	36.19	37.26	0.936
PRDM9 **Qigene**	1 µg Dox B	26.15	25.12	25.24	25.50	0.561
PRDM9	No Dox A	36.22	36.27	36.18	36.23	0.046
PRDM9	1 µg Dox A	25.38	24.84	24.81	25.01	0.320
PRDM9	No Dox B	36.08	36.32	36.46	36.29	0.196
PRDM9	1 µg Dox B	24.44	24.77	24.77	24.66	0.190
Actin	No Dox A (NRT)	N/A	N/A	N/A	0.00	0.000
GAPDH	No Dox A (NRT)	N/A	N/A	N/A	0.00	0.000
PRDM9 **Qigene**	No Dox A (NRT)	N/A	N/A	N/A	0.00	0.000
PRDM9	No Dox A (NRT)	N/A	N/A	N/A	0.00	0.000
Actin	1 µg Dox A (NRT)	N/A	N/A	N/A	0.00	0.000

GAPDH	1 µg Dox A (NRT)	N/A	N/A	N/A	0.00	0.000
PRDM9 Qigene	1 µg Dox A (NRT)	N/A	N/A	N/A	0.00	0.000
PRDM9	1 µg Dox A (NRT)	N/A	N/A	N/A	0.00	0.000
Actin	No Dox B (NRT)	N/A	N/A	N/A	0.00	0.000
GAPDH	No Dox B (NRT)	N/A	N/A	N/A	0.00	0.000
PRDM9 Qigene	No Dox B (NRT)	N/A	N/A	N/A	0.00	0.000
PRDM9	No Dox B (NRT)	N/A	N/A	N/A	0.00	0.000
Actin	1 µg Dox B (NRT)	N/A	N/A	N/A	0.00	0.000
GAPDH	1 µg Dox B (NRT)	N/A	N/A	N/A	0.00	0.000
PRDM9 Qigene	1 µg Dox B (NRT)	N/A	N/A	N/A	0.00	0.000
PRDM9	1 µg Dox B (NRT)	N/A	N/A	N/A	0.00	0.000
Actin	NTC	N/A	N/A	N/A	0.00	0.000
GAPDH	NTC	N/A	N/A	N/A	0.00	0.000
PRDM9 Qigene	NTC	N/A	N/A	N/A	0.00	0.000
PRDM9	NTC	N/A	N/A	N/A	0.00	0.000

Figure 5.32. **SYBR® Green-based quantitative real time PCR for overexpressed *PRDM9* in double-stable HeLa Tet-on 3G cells.** Comparison of the expression of *PRDM9* in two independent colonies with and without 1 µg/ml doxycycline induction. **(I)** Bar chart presenting the expression of the *PRDM9* gene in two independent colonies of HeLa Tet-on 3G, induced and not induced with 1 µg/ml doxycycline. Two sets of primers were used to compare the expression of *PRDM9* before and after induction. *Actin* and *GAPDH* primers were used to normalise the data and Bio-RAD CFX Manager was used to the data analysis. The error bars refer to the standard error for triplicate repeats. **(II)** Table showing the readings of a quantification cycle (Cq) and standard deviation for the reference genes *Actin* and *GAPDH,* which were used to normalise the *PRDM9* readings. Negative control readings (Cq) for NTR (no reverse transcriptase) and NTC (no template control) are presented in the table.

5.2.2   RT-PCR analysis of overexpression of *PRDM7* (*I*) or *PRDM7* (*II*) in double-stable HeLa Tet-on 3G cells

The successful integration of *PRDM7* (*I*) and *PRDM7* (*II*) into HeLa Tet-on 3G cells was determined by transfecting six independent colonies with pTRE3G::*PRDM7*(*I*) vector or three colonies with pTRE3G::*PRDM7*(*II*) vector and induction with 1 µg/ml doxycycline for 24 hours. In addition, a fraction of each clone was grown in the absence of doxycycline. RNAs were isolated and cDNAs were synthesized from each clone either with presence or absence of doxycycline.

The RT-PCR analysis revealed that *PRDM7* (*I*) was expressed in three colonies, (B), (C) and (E), after doxycycline induction (Figure 5.33). No expression of *PRDM7* (*I*) was displayed in the absence of doxycycline with exception of clone (B), which might have had a leaky promoter. The band intensity was stronger for clone (C) than for clone (E) and no expression was shown in clones (A),(D) and (F), which were excluded from further analysis.

The *PRDM7* (*I*) gene was further analysed in two independent colonies with integration of pTRE3G vector only. No expression of *PRDM7* (*I*) was shown in cells transfected with pTRE3G plasmid either in presence or absence of doxycycline after 24 hours (Figure 5.34).

In addition to *PRDM7* (*I*) analysis, RT-PCR analysis of *PRDM7* (*II*) showed only one promising clone out of three after induction of *PRDM7* (*II*) by doxycycline (Figure 5.35). Faint band intensity, indicating low expression of *PRDM7* (*II*), was determined in clones (A) and (B) in the absence of doxycycline, which might be the result of a doxycycline leak. The *PRDM7* (*II*) gene was also analysed in two independent colonies with integration of the pTRE3G vector only. No expression of *PRDM7* (*II*) was displayed in cells transfected with empty pTRE3G plasmid either in presence or absence of doxycycline after 24 hours (Figure 5.36).
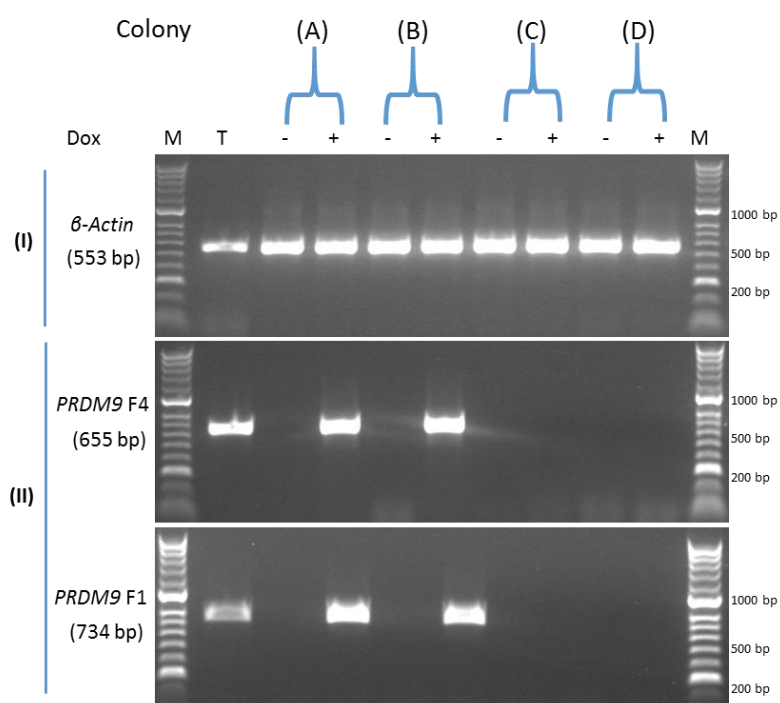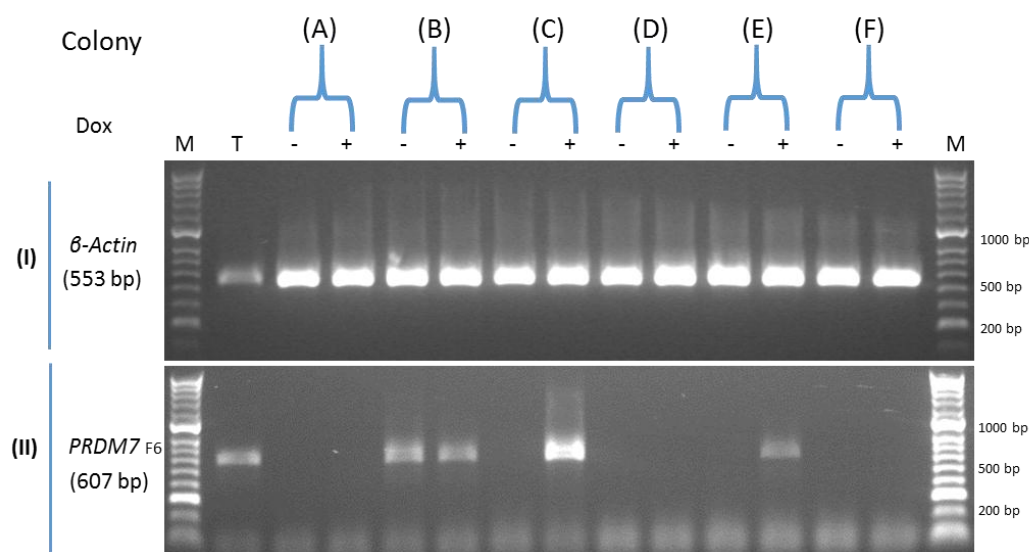
Figure 5.33. **RT-PCR analysis confirming expression of *PRDM7* (*I*) after induction in HeLa Tet-On 3G system.** Six independent colonies of HeLa Tet-On 3G were transfected with a *PRDM7*:pTRE3G construct. Cells were cultured in absence or presence of 1 μg/ml doxycycline for 24 hours. (I) *β-actin* served as a positive control for the cDNA samples. (II) The profile expression of *PRDM7* distinguishes primer (F6) in the presence or absence of doxycycline. Colonies (C) and (E) showed promising expression of *PRDM7* (*I*) in the presence of doxycycline. No expression was observed in any colony, except for colony (B) in the absence of doxycycline. Lane T is a testis sample that served as a positive control and M is a HyperLadder II marker.
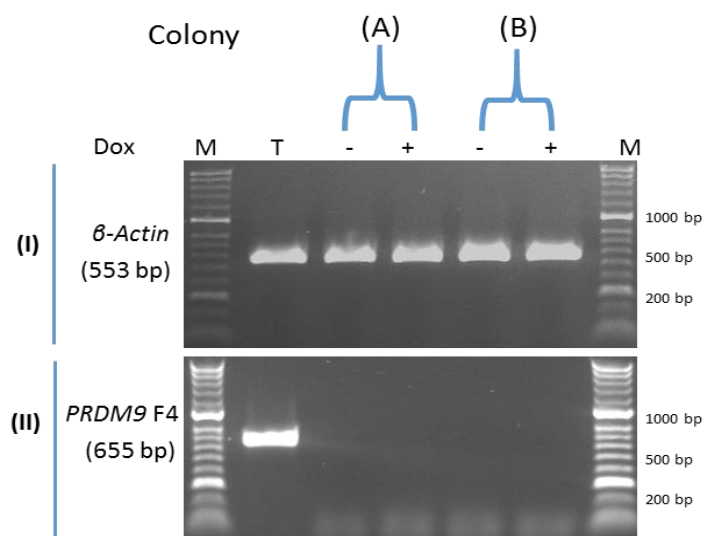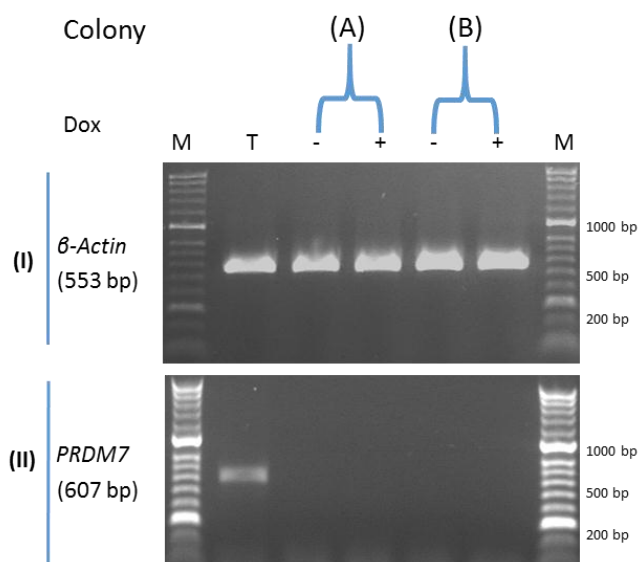
Figure 5.34. **RT-PCR analysis of *PRDM7* in two independent colonies transfected with pTRE3G empty vector only.** HeLa Tet-on 3G cells were transfected with pTRE3G plasmid and induced with 1 µg/ml doxycycline for 24 hours. RT-PCR was performed to check the expression of *PRDM7* (*I*) after induction of pTRE3G vector only. (II) No expression was determined for *PRDM7* (*I*) in both colonies either with or without doxycycline. (I) *β-Actin* served as a positive control for the cDNA samples. Lane T is a positive testis control and M is a HyperLadder II marker. 1% agarose gels stained with ethidium bromide were used to visualise the PCR products.
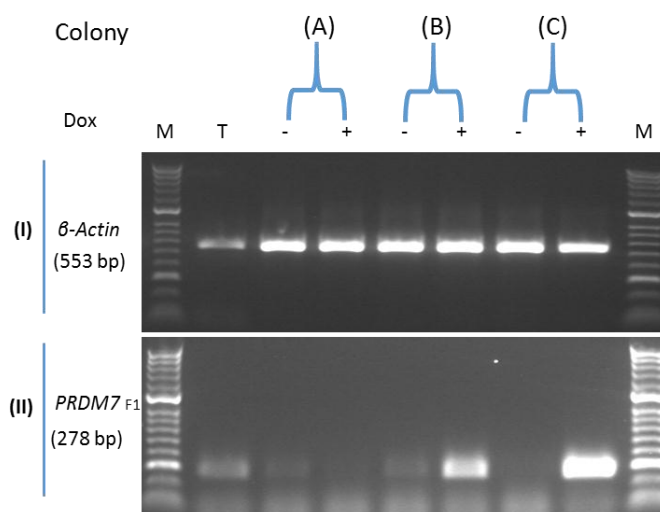
Figure 5.35. **RT-PCR analysis confirming expression of *PRDM7(II)* after induction in HeLa Tet-on 3G system.** Three independent colonies of HeLa Tet-on 3G transfected with *PRDM7*:pTRE3G construct. Cells were cultured in absence or presence of 1 µg/ml doxycycline for 24 hours. (I) β-actin served as a positive control for the cDNA samples. (II) The profile expression of *PRDM7* (*II*) in the presence or absence of doxycycline. Colonies (B) and (C) showed strong expression of the *PRDM7* (*II*) in the presence of doxycycline. In the absence of doxycycline, no expression was detected in colony (C) but a faint band intensity was present for colony (B). Lane T is a testis positive control and M is a HyperLadder II marker.
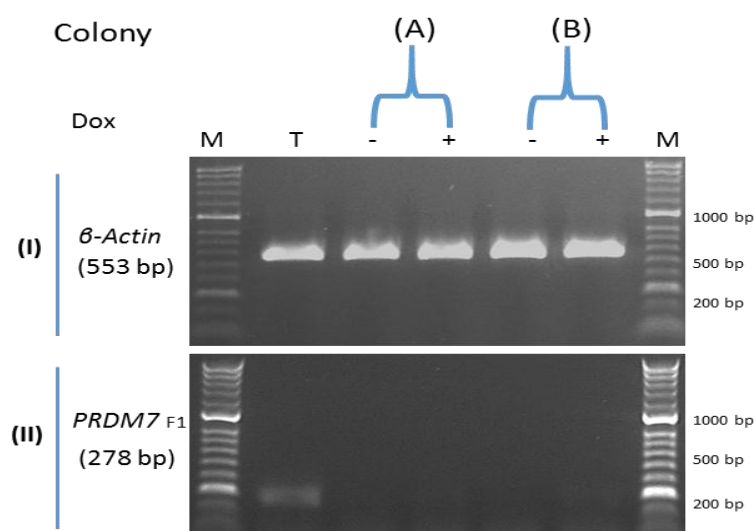
Figure 5.36. **RT-PCR analysis of** *PRDM7* **(***II***) in two independent colonies after transfection with pTRE3G empty vector.** HeLa Tet-on 3G cells were transfected with pTRE3G plasmid and induced with 1 μg/ml doxycycline for 24 hours. RT-PCR was performed to check the expression of *PRDM7* (II) after induction of the pTRE3G vector only. (II) No expression was determined for *PRDM7* (II) in either colony with or without doxycycline. (I) *β-Actin* served as a positive control for the cDNA samples. Lane T is a positive testis control and M is a HyperLadder II marker. 1% agarose gels stained with ethidium bromide were used to visualise the PCR products.

5.2.10   RT-PCR analysis of *PRDM9* overexpression on influence the expression of meiosis genes

Several members of the PRDM family of proteins were suggested to have an oncogenic protein functions and might drive the expression of different genes via modification the chromatin stat at the target genes promoters (Fog et al., 2012; Hohenauer and Moore, 2012). Prdm9 was found to play a critical role in the transcriptional control the expression of the *RiK* (*morc2b*) gene in mice (Hayashi *et al.*, 2005).

Here, total RNA preparations were isolated from two independent HeLa Tet-on 3G cells transfected with the pTRE3G::*PRDM9* vector or transfected with pTRE3G only, either in the absence or presence of 1 µg/ml doxycycline for 24 hours. The cDNA was synthesised from each clone and *β-actin* primers were used to check the cDNA quality.

The influence of *PRDM9* overexpression in the *MORC* family genes, was investigated by RT-PCR analysis using *MORC1*, *MORC2*, *MORC3* and *MORC4* primers (Figure 5.37). No expression of *MORC1* was observed either with or without *PRDM9* overexpression. *MORC2*, *MORC3* and *MORC4* showed expression in both induced and uninduced HeLa Tet-on 3G cells; this feature requires further investigation using qRT-PCR. Cells transfected with only pTRE3G as well showed similar results for these genes.

PRDM9 has an essential role as a meiotic hotspot regulator and we asked whether overexpression of *PRDM9* might influence the expression of known meiosis genes. RT-PCR analyses were performed on 13 known meiosis genes using cDNA generated from HeLa Tet-on 3G cells with or without *PRDM9* overexpression. No influence of *PRDM9* overexpression was observed on *DMC1*, *HORMAD2*, *RAD21L* and *REC8* genes. The *HORMAD1* gene also showed expression in both HeLa Tet-on 3G cells with or without *PRDM9* overexpression, which also might need further analysis by qRT-PCR (Figure 5.38 A).

In addition, no significant effects of PRDM9 overexpression were observed in either *SMC1β* or *SPO11* expression. Two sets of *SPO11* primers were used and showed similar results. Two sets of *STAG3* primers were also used and displayed faint band expression after addition of 1 µg/ml doxycycline, even with empty transfected pTRE3G plasmid only (Figure 5.38 B). Furthermore, no influence of *PRDM9* overexpression was found for the expression of

synaptonemal complex genes *SYCE1*, *SYCE2*, *SYCP1* and *SYCEP3* or the meiotic regulator gene *STRA8* (Figure 5.38 C).

RT-PCR was performed using *PRDM1*, *PRDM4*, *PRDM6*, *PRDM7* and *PRDM11* primers to determine whether overexpression of *PRDM9* can regulate the expression of PRDM family genes (Figure 5.39). RT-PCR analysis results showed strong band intensity for *PRDM7* expression in *PRDM9* overexpressed HeLa Tet-on 3G cells. No expression of *PRDM7* was displayed either in the absence of doxycycline or with transfected cells with pTRE3G only. Two sets of distinguishing *PRDM7* primers were used and the results suggested that induction of *PRDM9* might up-regulate the expression of *PRDM7*. In contrast, *PRDM1*, *PRDM4*, *PRDM6* and *PRDM11* showed expression in both induced and uniduced *PRDM9* HeLa Tet-on 3G cells, which needs to be investigated further using qRT-PCR to measure the gene expression.

RT-PCR was also carried out to investigate the influence of *PRDM9* overexpression on some cancer testis genes identified in Chapter 3.0 expression of testis-restricted *ARRD5*, *C12orf12*, *DDX4* and *NT5C1B* was not affected by overexpression of *PRDM9* in HeLa Tet-on 3G cells (Figure 5.40 A). The expressions of cancer/testis *NUT*, *ODF4*, *SEPT12* and *TDRD12* genes were also not influenced (Figure 5.40 B). Finally, no significant expression was observed on X-CT *MAGE-B5*, *SSX2* and *GAGE1* expression in either induced or uninduced cells (Figure 5.40 C).
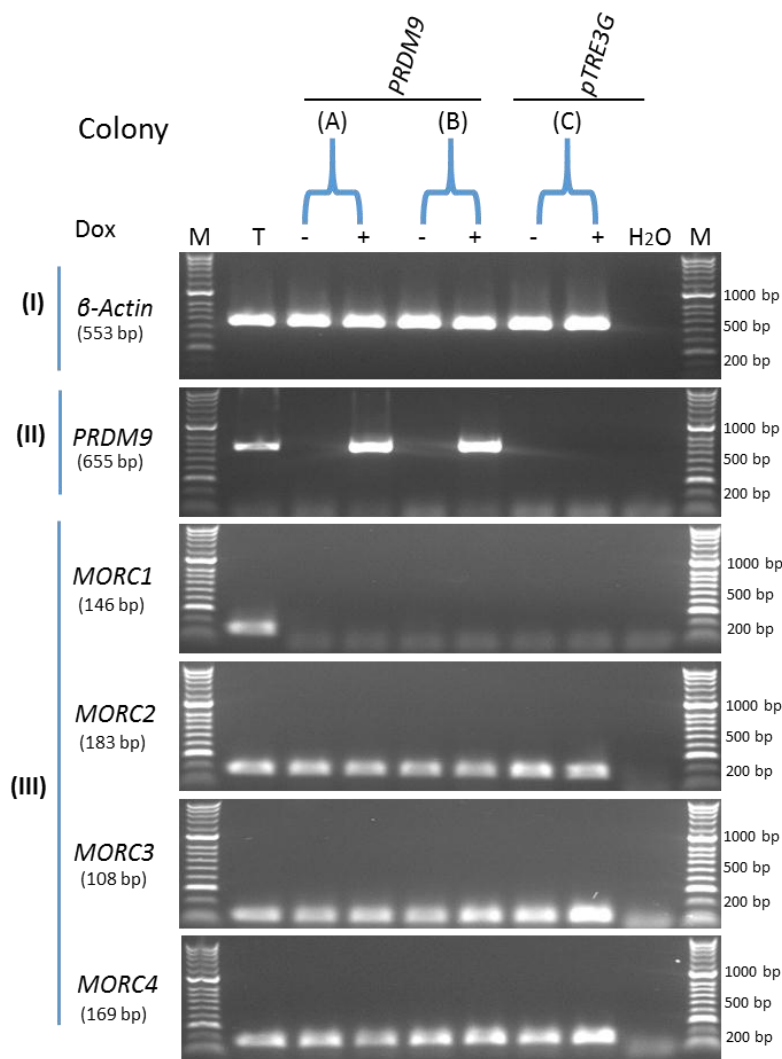
Figure 5.37. **RT-PCR analysis of four *MORC* family genes after overexpression of *PRDM9* in a HeLa Tet-on 3G system.** HeLa Tet-on 3G cells were transfected with *PRDM9*:pTRE3G construct (A and B) or pTRE3G plasmid alone (C) and induced with 1 µg/ml doxycycline for 24 hours. (I) *β-Actin* served as a positive control of the cDNA samples. (II) Expression of *PRDM9* was detected in induced (+) colonies integrated *PRDM9* whereas no expression was seen in uninduced (-) colonies. (III) No expression of the *MORC1* gene was observed either with or without *PRDM9* overexpression. *MORC2, MORC3 and MORC4* showed expression in both induced and uninduced PRDM9 gene. Lane T is a positive testis control and M is a HyperLadder II marker. 1% agarose gels stained with ethidium bromide were used to visualise the PCR products.
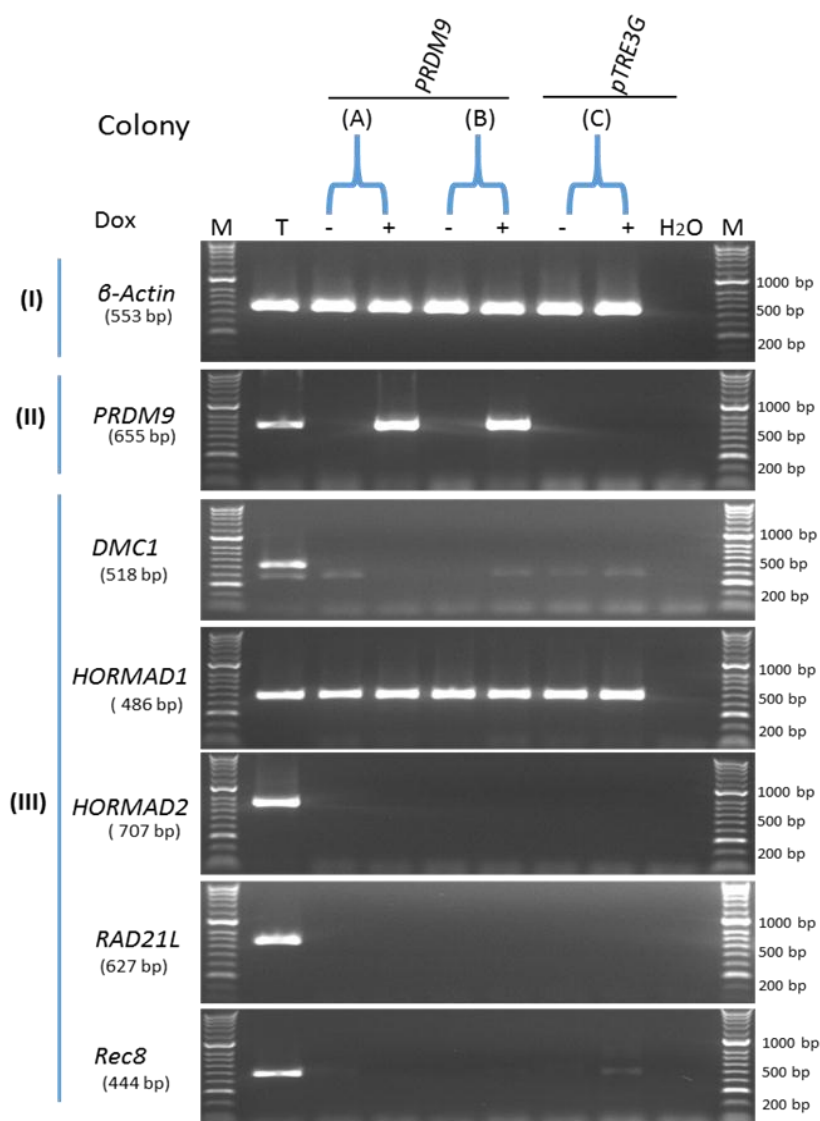
Figure 5.38 A. **RT-PCR analysis of five meiosis specific genes after overexpression of *PRDM9* in a HeLa Tet-on 3G system.** (I) *β-Actin* served as a positive control for the cDNA samples. (II) Expression of *PRDM9* was detected in induced (+) colonies containing integrated *PRDM9* whereas no expression was observed in uninduced (-) colonies. (III) No influence of *PRDM9* overexpression was noted on *DMC1*, *HORMAD2*, *RAD21L* and *Rec8* genes either with or without 1 µg/ml doxycycline for 24 hours. In contrast, *HORMAD1* showed expression in both induced and uniduced *PRDM9* gene. Lane T is a positive testis control and M is a HyperLadder II marker. 1% agarose gels stained with ethidium bromide were used to visualise the PCR products.

Figure 5.38 B. RT-PCR analysis of three meiosis specific genes after overexpression of *PRDM9* in a HeLa Tet-on 3G system. (I) *β-Actin* served as a positive control for the cDNA samples. (II) Expression of *PRDM9* was detected in induced (+) colonies containing integrated *PRDM9* whereas no expression was noted in uninduced (-) colonies. (III) *PRDM9* overexpression had no influence on *SMC1β* and *SPO11* genes either with or without 1 μg/ml doxycycline for 24 hours. *STAG3* showed an expression band after addition of 1 μg/ml doxycycline even with only empty pTRE3G plasmid. Two sets of primers were used for *SPO11* and *STAGE3*. Lane T is a positive testis control and M is a HyperLadder II marker. 1% agarose gels stained with ethidium bromide were used to visualise the PCR products.
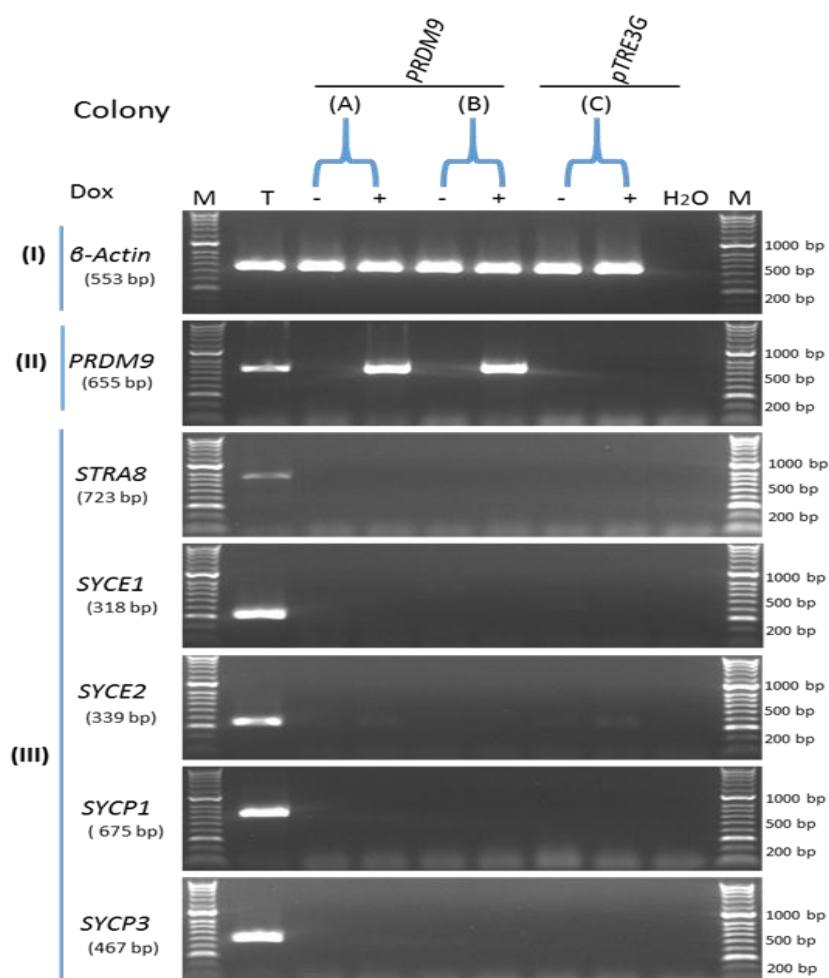
Figure 5.38 C. **RT-PCR analysis of five meiosis specific genes after overexpression of _PRDM9_ in a HeLa Tet-on 3G system.** (I) _β-Actin_ served as a positive control for the cDNA samples. (II) Expression of _PRDM9_ was detected in induced (+) colonies containing integrated _PRDM9,_ whereas no expression was seen in uninduced (-) colonies. (III) No influence of _PRDM9_ overexpression was observed on _STRA8_, _SYCE1_, _SYCE2_, _SYCP1 and SYCEP3_ genes either with or without 1 µg/ml doxycycline for 24 hours. Lane T is a positive testis control and M is a HyperLadder II marker. 1% agarose gels stained with ethidium bromide were used to visualise the PCR products.
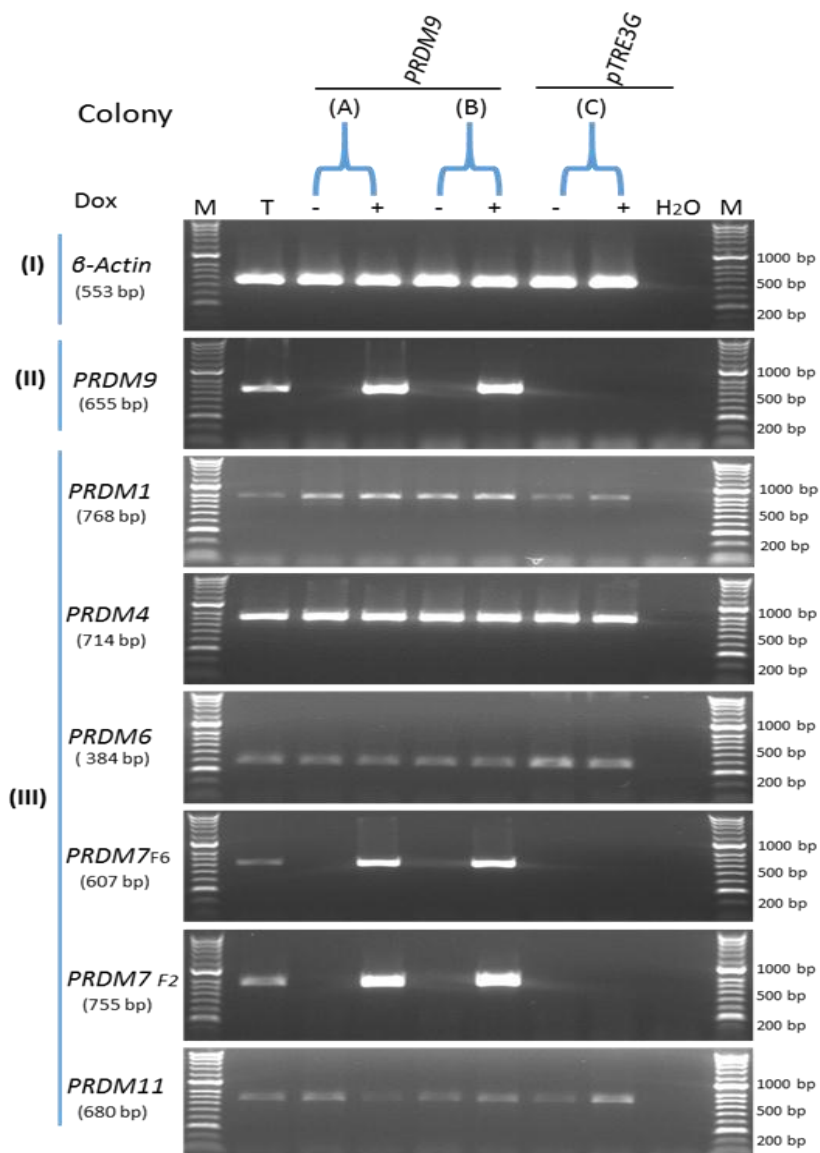
Figure 5.39. **RT-PCR analysis of five *PRDM* family genes after overexpression of *PRDM9* in HeLa Tet-on 3G system.** (I) *β-Actin* served as a positive control for the cDNA samples. (II) Expression of *PRDM9* was detected in induced (+) colonies containing integrated *PRDM9,* whereas no expression was noted in uninduced (-) colonies. (III) *PRDM1*, *PRDM4*, *PRDM6* and *PRDM11* showed expression in both induced and uniduced HeLa Tet-on 3G cells with the *PRDM9* gene. In contrast, *PRDM7* showed a high band intensity in HeLa Tet-on 3G cells with the *PRDM9* gene. No expression of *PRDM7* was observed either in the absence of doxycycline or with transfected cells with pTRE3G only. Lane T is a positive testis control and M is a HyperLadder II marker. 1% agarose gels stained with ethidium bromide were used to visualise the PCR products.
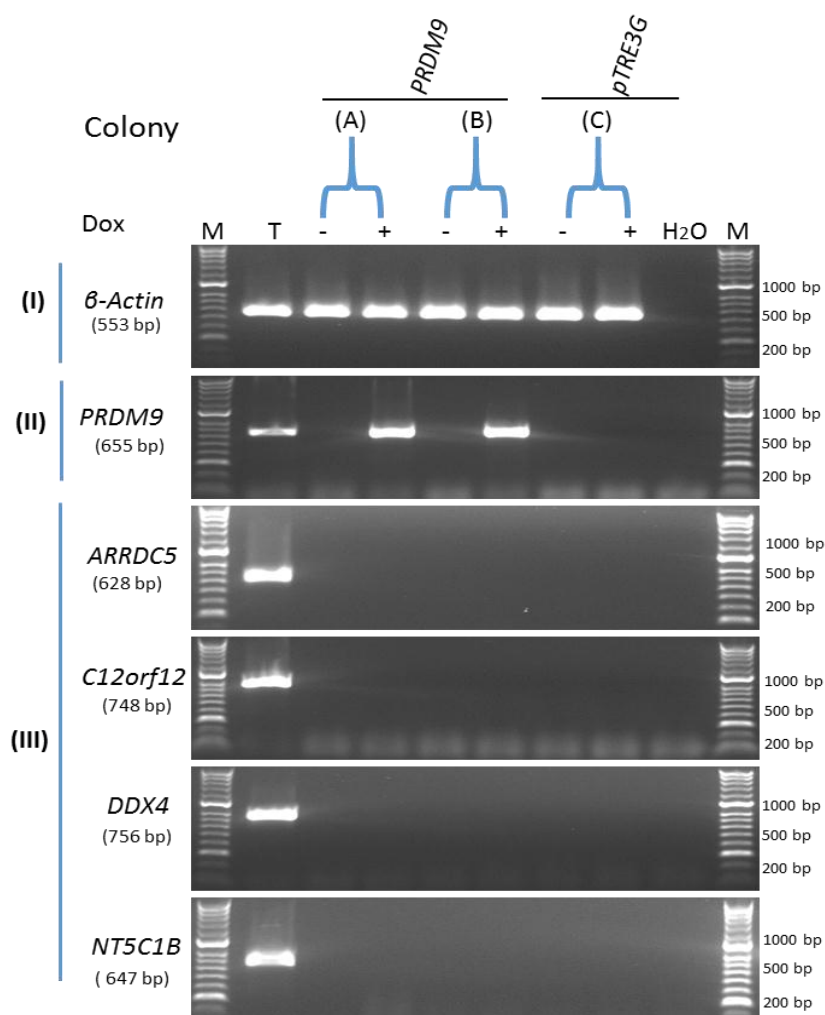
Figure 5.40 A. **RT-PCR analysis of four testis-restricted CT genes after overexpression of *PRDM9* in HeLa Tet-on 3G system.** (I) *β-Actin* served as a positive control for the cDNA samples. (II) Expression of *PRDM9* was detected in induced (+) colonies containing integrated *PRDM9,* whereas no expression was noted in uninduced (-) colonies. (III) No expression was observed for *ARRDC5*, *C12orf12*, *DDX4* and *NT5C1B* in both induced and uninduced HeLa Tet-on 3G cells transfected with either *PRDM9* or pTRE3G only. Lane T is a positive testis control and M is a HyperLadder II marker. 1% agarose gels stained with ethidium bromide were used to visualise the PCR products.
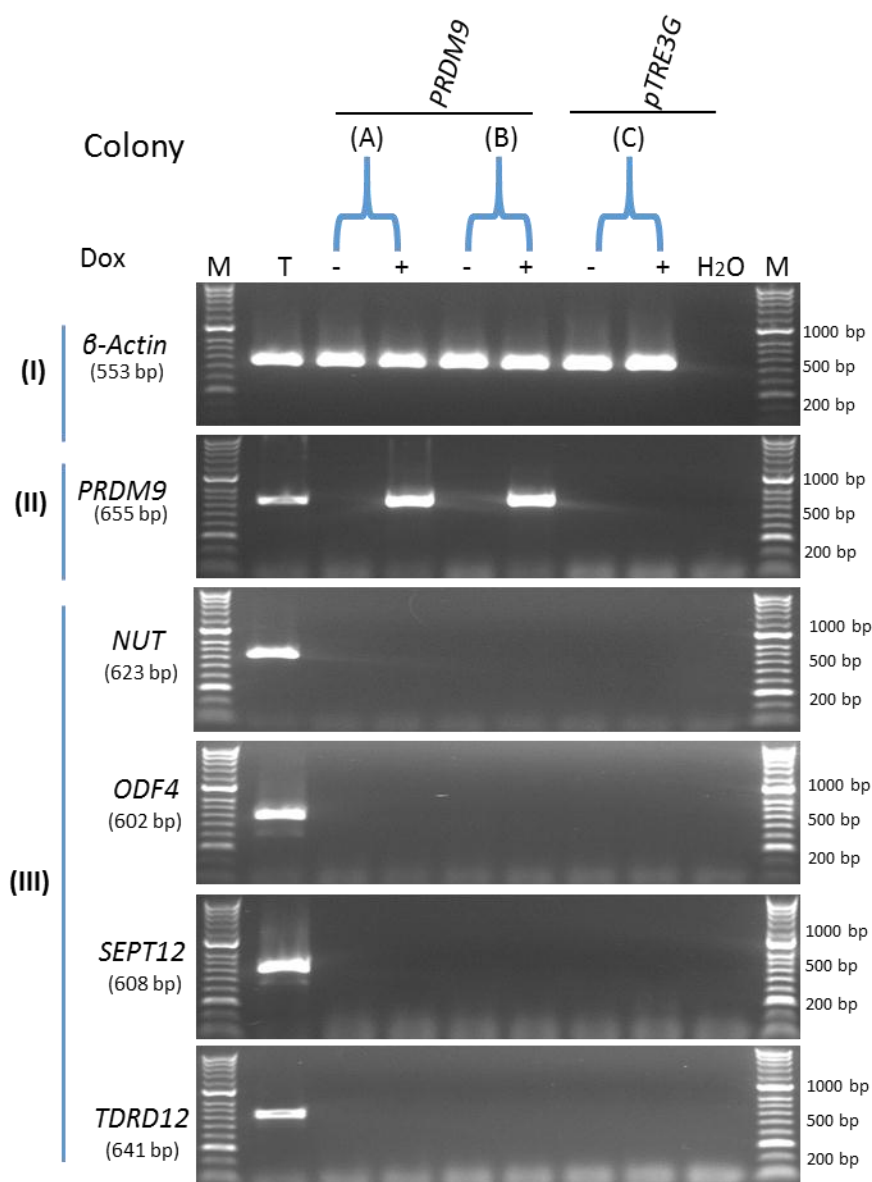
Figure 5.40 B. RT-PCR analysis of four potential cancer/testis-restricted genes after overexpression of *PRDM9* in HeLa Tet-on 3G system. (I) *β-Actin* served as a positive control for the cDNA samples. (II) Expression of *PRDM9* was detected in induced (+) colonies containing integrated *PRDM9,* whereas no expression was noted in uninduced (-) colonies. (III) No expression was observed for *NUT*, *ODF4*, *SEPT12* and *TDRD12* in both induced and uninduced HeLa Tet-on 3G cells transfected with either *PRDM9* or pTRE3G only. Lane T is a positive testis control and M is a HyperLadder II marker. 1% agarose gels stained with ethidium bromide were used to visualise the PCR products.
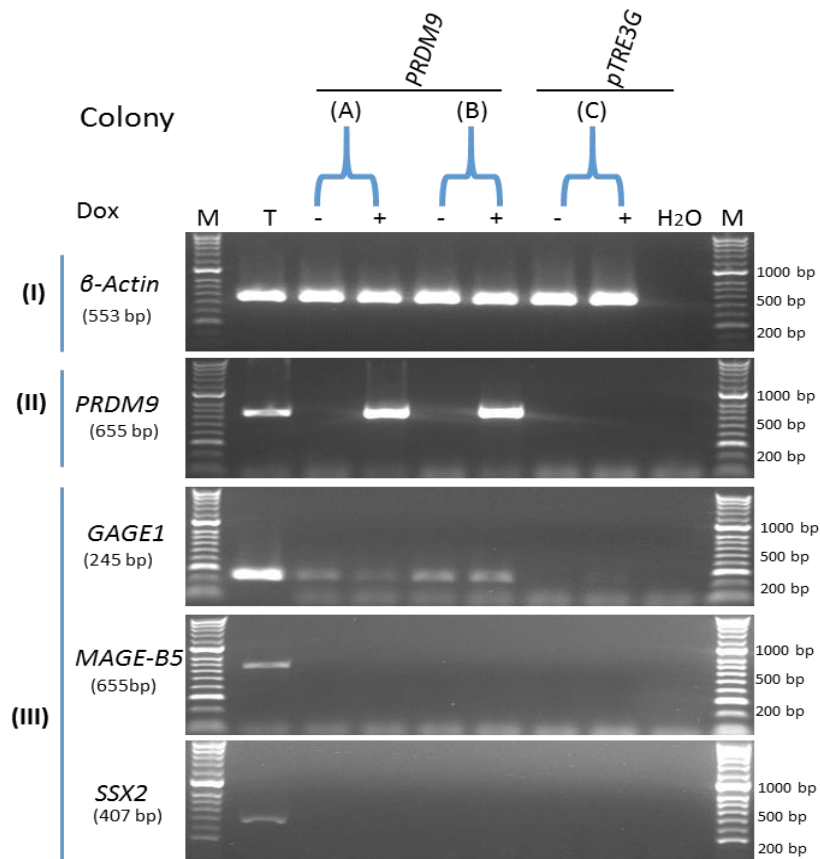
Figure 5.40 C. **RT-PCR analysis of three known X-CT genes after overexpression of *PRDM9* in HeLa Tet-on 3G system.** (I) *β-Actin* served as a positive control for the cDNA samples. (II) Expression of *PRDM9* was detected in induced (+) colonies containing integrated *PRDM9,* whereas no expression was noted in uninduced (-) colonies. (III) No expression was observed for *MAGE-B5* and *SSX2* in both induced and uninduced HeLa Tet-on 3G cells transfected with either *PRDM9* or pTRE3G only. *GAGE1* was expressed in both colonies (A) and (B) with or without *PRDM9* overexpression. Lane T is a positive testis control and M is a HyperLadder II marker. 1% agarose gels stained with ethidium bromide were used to visualise the PCR products.
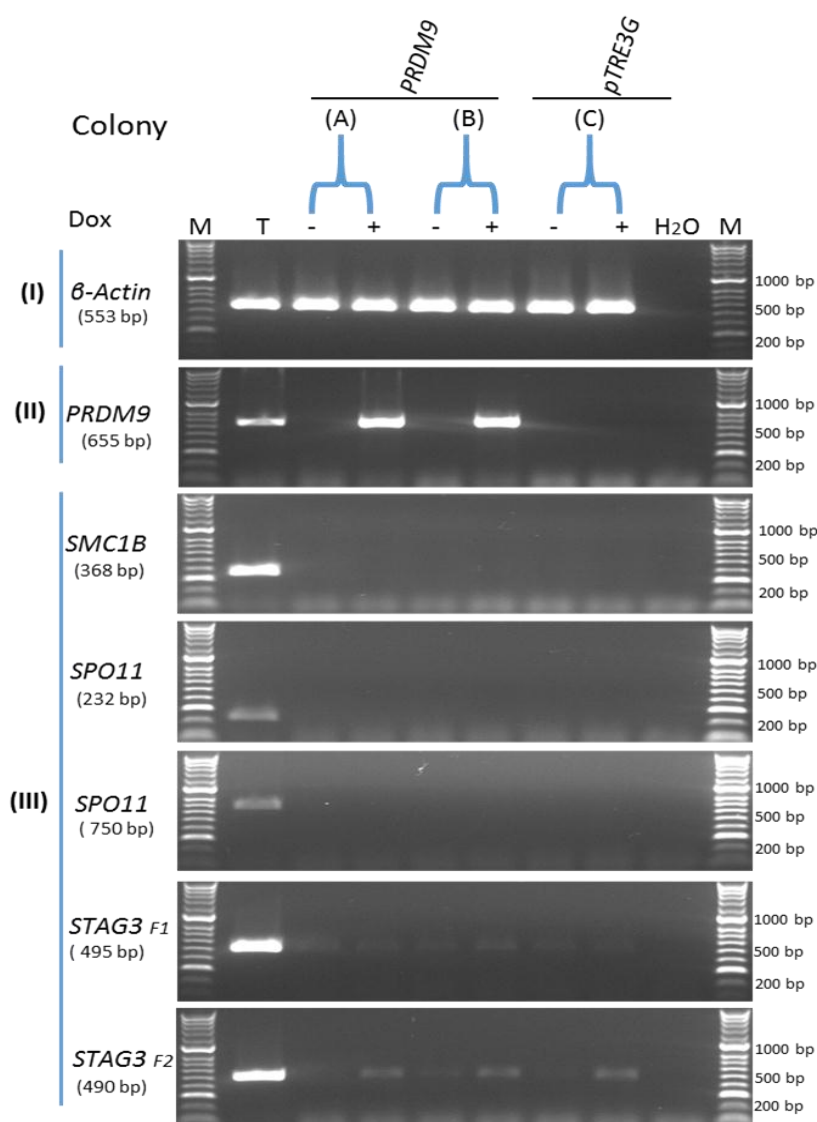
5.3   Discussion

5.3.1 The relationship between *PRDM9* expression in normal and cancer cells compared to expression of *PRDM7* and *MORC* family genes

Aberrant production expression of PRDM family proteins is often associated with cancer development, particularly in breast cancer, cervical cancer, colorectal cancer and leukaemia (reviewed in Hohenauer and Moore, 2012). PRDM9 is the only meiosis-specific protein of this family and has been identified to regulate meiotic recombination hotspot locations (Myers *et al.*, 2010). RT-PCR analysis has also confirmed *PRDM9* expression in different cancer types including colon cancer, breast cancer, ovarian cancer, melanoma and leukaemia (Feichtinger *et al.*, 2012a). In addition, a link was reported between a rare allele of *PRDM9* and the development of leukaemogenesis in children (Hussin *et al.*, 2013).

In the present study, *PRDM9* was found to be co-expressed with *PRDM7* in leukaemia only. This might raise a question regarding the relationship between *PRDM7* expression in leukaemia and the rare *PRDM9* allele; this needs to be further explored. However, gene alignment of *PRDM7* and *PRDM9* showed high similarity in their DNA sequences. All attempts to design specific primers for the *PRDM9* zinc finger region failed but distinguishing primers were designed using the 89-nucleotide segment present in *PRDM7* only.

The PRDM9 protein in mice was found to directly regulate the expression of the *RIk* (*Morc2b*) gene (Hayashi *et al.*, 2005). However, in the present study, the expression of the human *RIK* (*Morc2b*) orthologue (*MORC1*, *MORC2*, *MORC3* and *MORC4*) showed different expression patterns than *PRDM9* in both normal and cancer tissues. This indicated that *PRDM9* might not regulate the expression of *MORC* family genes in humans.

5.3.2 PRDM9 protein analysis and the influence of PRDM9 knockdown in HCT116 and SW480 growth

PRDM9 protein was detected in NTERA-2 cancer cells but was not detected in the normal prostate tissues. This suggested PRDM9 as a potential CT antigen biomarker (Feichtinger *et al.*, 2012a). Western blot analysis performed to explore the presence of PRDM9 protein in cancer cells revealed it in nearly all cancer cell lysates used in this study. However, the PRDM9 siRNA knockdown results need to be verified. This could be done using a new anti-

PRDM9 antibody or better protein control such as Ponceau S protein stain. This stain total protein in the membrane and can verify the loading and transfer.

In mice and humans, PRDM9 recognises and binds to specific DNA segments called hotspots via its zinc finger domain, which is a nuclear function (Baudat *et al.*, 2010; Myers *et al.*, 2010; Parvanov *et al.*, 2010). Examination of the PRDM9 protein localisation by cell fractionation showed that the PRDM9 protein was present in the nucleus in most of the cancer cells. The protein was determined in the cytoplasm only in NTERA-2 cancer cells. This suggested that PRDM9 could have transcriptional activation functions in cancer cells.

PRDM9 knockdown was successful in some cell lines, but failed in others. Three hits of siRNAs reduced the level of PRDM9 protein in most cancer cells, while only one hit was required for PRDM9 knockdown in NTERA-2 cells. We noted that PRDM9 was detected in most of the cancer cell lines used, so we asked if PRDM9 might be required for cancer cell proliferation.

The ELDA assay was used to examine the influence of PRDM9 knockdown on cancer cell survival. Defects in PRDM9 protein in HCT116 and SW480 cells reduced the proliferation frequency of these cells, with highest effect observed using siRNA-7. Transfection with siRNA-7 reduced the cell proliferation frequency 8-fold in HCT116 cells and 14-fold in SW480 cells, compared to the negative control siRNA transfected cells or untreated cells. This result was compatible with the results from the western blot analysis of PRDM9 knockdown, which also showed preferential knockdown of PRDM9 with siRNA-7 in these two cell lines.

The *PRDM* gene family has been reported to control cell proliferation in normal and cancer cells (Hohenauer and Moore, 2012). For example, PRDM1 (also known BLIMP1) was identified as a transcriptional regulator of cell proliferation and survival through its control of the activity of *P53* transcription (Yan *et al.*, 2007). PRDM1 and PRDM14 co-express and are obligatory for the establishment of germ cell lineage (Yamaji *et al.,* 2008; Di Zazzo *et al.*, 2013). Expression of PRDM1 was identified in different cancer cells and knockdown of PRDM1 in HCT116 cells induced apoptosis and arrested cell growth (Yan *et al.*, 2007). In addition, Nishikawa et al. (2007) showed that PRDM14 was overexpressed in about two thirds of breast cancers, and gene knockdown reduced cell proliferation, induced apoptosis

and increased the sensitivity of chemotherapeutic drugs (Nishikawa *et al*., 2007). PRDM9 plays a critical role in the regulation of hotspot activity during meiosis (Baudat *et al*., 2010; Myers *et al*., 2010; Parvanov *et al*., 2010); consequently, it might be associated with pathological genomic rearrangements (reviewed in Hussin *et al.,* 2013). In the present study, knockdown of PRDM9 reduced the proliferation and survival of HCT116 and SW480 cells. The molecular role of PRDM9 in carcinogenesis remain elusive. It has been demonstrated that germline genes in *Drosophila melanogaster* are required for the oncogenic process and that the human orthologues of these *Drosophila* genes have up-regulated expression in a range of human cancers, although the functional implications for oncogenesis of this up-regulation remains unclear (Janic *et a*l., 2010; Feichtinger *et al*., 2014 b). Interestingly, down-regulation of a number of CT genes in human cancer cells results in perturbation of cellular proliferative potential (for example, see Linley *et al*., 2012; Cappell *et al.,* 2012). These findings open up the exciting possibility that CT genes might encode functions that are required for tumour homeostasis and it has recently been proposed that tumours become 'addicted' to these germline factors (Rousseaux *et al.,* 2013b; McFarlane *et al.,* 2014).

Taking all these findings together raises the possibility that PRDM9, as a zinc finger transcription protein, might also play an oncogenic role that drives cancer cell proliferation in association with other unknown genes.

5.3.3   Cloning and overexpression of PRDM9 in HeLa Tet-on 3G cells

Several members of the PRDM family have been identified to regulate target gene expression via modification of the chromatin state of gene promoters (Hohenauer and Moore, 2012). In mice, Prdm9 was found to directly regulate *RIK* (*Morc2b*) expression (Hayashi *et al.,* 2005). However, the transcriptional activity of PRDM9 in human cancer cells remains poorly understood. We asked if this protein might drive the expression of other meiosis genes in cancer cells.

The full lengths of *PRDM9* and *PRDM7* were cloned and overexpressed using a HeLa Tet-on 3G system. Overexpression of *PRDM9* was demonstrated using both RT-PCR and qRT-PCR analysis. Both variants of *PRDM7* were also cloned and overexpressed in HeLa Tet-on 3G

cells. The main purpose was to generate double-stable HeLa Tet-on 3G cells with transgenes of *PRDM9*, *PRDM7* (*I*) or *PRDM7* (*II*).

We investigated whether overexpression of *PRDM9* in cancer cell lines might influence the expression of other genes. RT-PCR analysis was performed to investigate the expression of 34 selected genes, which included meiosis-specific genes, meiCT genes, *MORC* family genes and *PRDM* family genes. The RT-PCR analysis showed no significant overexpression of 33 genes when validated in HeLa Tet-on 3G cells overexpressing *PRDM9*. RT-PCR might not be sufficient for studying whether *PRDM9* can induce the expression of these genes. However, *PRDM7* showed overexpression with high band intensity after validation in the *PRDM9* overexpressing cell line. Distinguishing primers for *PRDM7* were used to confirm this finding because of the high similarity of the DNA sequences of *PRDM7* and *PRDM9*. Co-expression of these two PRDM genes was detected in normal testis tissues, in HeLa-tet on 3G cells overexpressing *PRDM9* and in a human erythroleukaemic (K-562) cell line. This co-expression needs further confirmation using qRT-PCR and western blotting, which could not be performed in the present study due to time limitations. This might be a promising avenue of research, as some members of the *PRDM* family are co-expressed/activated with other multiple factors as well as with members of the same family (e.g., co-expression of *PRDM1* and *PRDM14*) (Di Zazzo *et al*., 2013).

5.4 Conclusion

RT-PCR analysis of *PRDM7* and *PRDM9* showed a clear restriction of expression to the normal testis tissues and different cancer tissues, with no expression observed in other normal tissues. In this chapter, a lack of relationship between *PRDM9* and *MORC* family genes was identified in human normal and cancer tissues. PRDM9 protein production was detected in different types of cancer cells, with nuclear localisation in most cells, indicating it could be a potential biomarker for cancer immunotherapy. Following overexpression of *PRDM9* genes in HeLa Tet-on 3G cells that led to activation of the expression of the *PRDM7* gene. No overexpression of 33 other genes was observed in HeLa tet-on 3G cells that overexpressed *PRDM9*. Further investigation of these genes using qRT-PCR and western blot analysis is required.

Chapter 6.0 Cloning and expression of the novel cancer testis antigens PRDM9 and PRDM7 in *E. coli*

6.1 Introduction

PRDM family proteins were identified as a class of putative transcriptional regulators involved in the control of gene expression through alteration of the chromatin state in the promoters of target genes. The existence of the PR domain and the presence of multiple zinc finger arrays are the main features of this family (Hohenauer and Moore, 2012). The PR domain is related to the catalytic SET domains that characterise protein-arginine methyltransferases (HMTs) (reviewed in Fog *et al.*, 2012; Hohenauer and Moore 2012). Histone methyltransferase activity was detected in PRDM2 (Kim *et al.*, 2003), PRDM3 (Pinheiro *et al.*, 2012), PRDM6 (Wu *et al*., 2008), PRDM8 (Eom *et al.*, 2009) and PRDM16 (Pinheiro *et al*., 2012).

In addition, mouse Prdm9 (Meisetz) was demonstrated to catalyse the trimethylation of lysine 4 of histone H3 (H3K4me3) but not mono-or dimethylation (Hayashi *et al.*, 2005). This meiosis-specific protein binds to a 13-mer specific DNA sequence through the C2H2 zinc finger domain. Subsequently, the PR/SET domain activates H3K4me3, followed by induction of DSBs by the SPO11 protein (Parvanov *et al.*, 2010; Baudat *et al.*, 2013). *PRDM9* and *PRDM7* share a high similarity in their DNA sequences and structures, and *PRDM7* exists in humans only (Fumasoni *et al.*, 2007). To date, no methyltransferase activity has been demonstrated for PRDM7.

Recombinant DNA technology using the glutathione S-transferase (GST) gene fusion system has been used to identify the specific protein function of a range of PRDM family proteins including PRDM2 (Kim *et al.*, 2003), PRDM3, PRDM16 (Pinheiro *et al*., 2012), PRDM8 (Eom *et al*., 2009) and Prdm9 (Meisetz) (Hayashi *et al.*, 2005).

Protein production using the GST fusion system is a versatile system for expression, purification and detection in *E.coli* cells. It is one of the most extensively used systems for

obtaining high levels of induced desired proteins. The aim of this chapter was to subclone and express the full lengths of human *PRDM7* and *PRDM9* genes into pGEX-2T plasmids to produce recombinant proteins *in vitro*. The goal was to produce high purity PRDM9 and PRDM7 proteins which could then be examined for histone methyltransferase activity.

6.2   Results

6.2.1   Subcloning of *PRDM9* and *PRDM7* genes into a protein production pGEX-2T vector

The full lengths of *PRDM9* and *PRDM7* variants (*I*) and (*II*) were subcloned into the protein production pGEX-2T vector containing an upstream GST tag for protein function analysis (Figure 6.1). Genes were released from constructed plasmids (see Section 5.2.6) pGEM-3ZF (+)::*PRDM9* (pAMO1); pUC57::*PRDM7* (*I*) (pAMO5) and pNEB193::*PRDM7*(*II*) (pAMO6)   by digestion with the *Bam*HI restriction enzyme. Digested and purified pGEX-2T plasmid with *Bam*HI (≈ 4948 bp) was used to ligate purified *PRDM9* ≈2685 bp; *PRDM7* (*I*) ≈1479 bp; and *PRDM7* (*II*) ≈516 bp individually to generate recombinant plasmids (Figure 6.2).

Ligated plasmids pGEX-2T:: *PRDM9*; pGEX-2T::*PRDM7* (*I*) and pGEX-2T::*PRDM7* (*II*) were then transformed into BL21 (DE3) and/or SoluB21 competent cells under ampicillin resistance selection. Random white colonies were selected as the pGEX-2T plasmid lacks the blue/white detection system. PCR reactions were then performed using internal primers for *PRDM9* or *PRDM7* genes to confirm that the colonies with the correct insertion were chosen. Colonies with positive PCR were selected for overnight growth (Figure 6.3).

Recombinant plasmid *PRDM9*: pGEX-2T (pAMO2); *PRDM7* (*I*): pGEX-2T (pAMO3); and *PRDM7* (*II*): pGEX-2T (pAMO4)   were digested successfully with the *Bam*HI enzyme. Analysis of the digestion on 0.8% agarose gels showed the presence of a full-length insertion of *PRDM9* (≈ 2685 bp), *PRDM7* (*I*) (≈ 1479 bp) and *PRDM7* (*II*) (≈ 516 bp). Undigested and digested pGEX-2T (≈ 4948 bp) were used for comparison with the recombinant plasmids (Figure 6.4).

The recombinant plasmids were confirmed by DNA sequencing to check for unwanted mutations and the correct gene orientation. Universal pGEX primers and internal primers for *PRDM9* and *PRDM7* were used for the DNA sequencing listed in Tables 2.3 and 2.11 (see the Materials and Methods section).
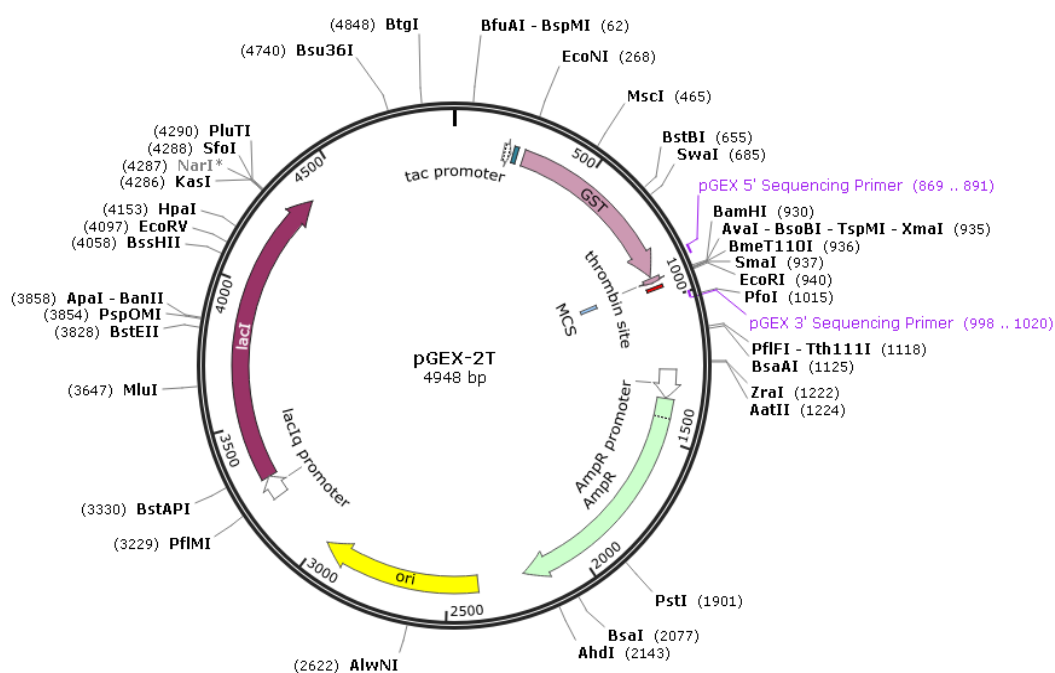
Figure 6.1. **Protein expression pGEX-2T plasmid was used to subclone *PRDM9* and *PRDM7* genes.** The plasmid was digested with *Bam*HI restriction enzyme and genes were cloned into the *Bam*HI site.
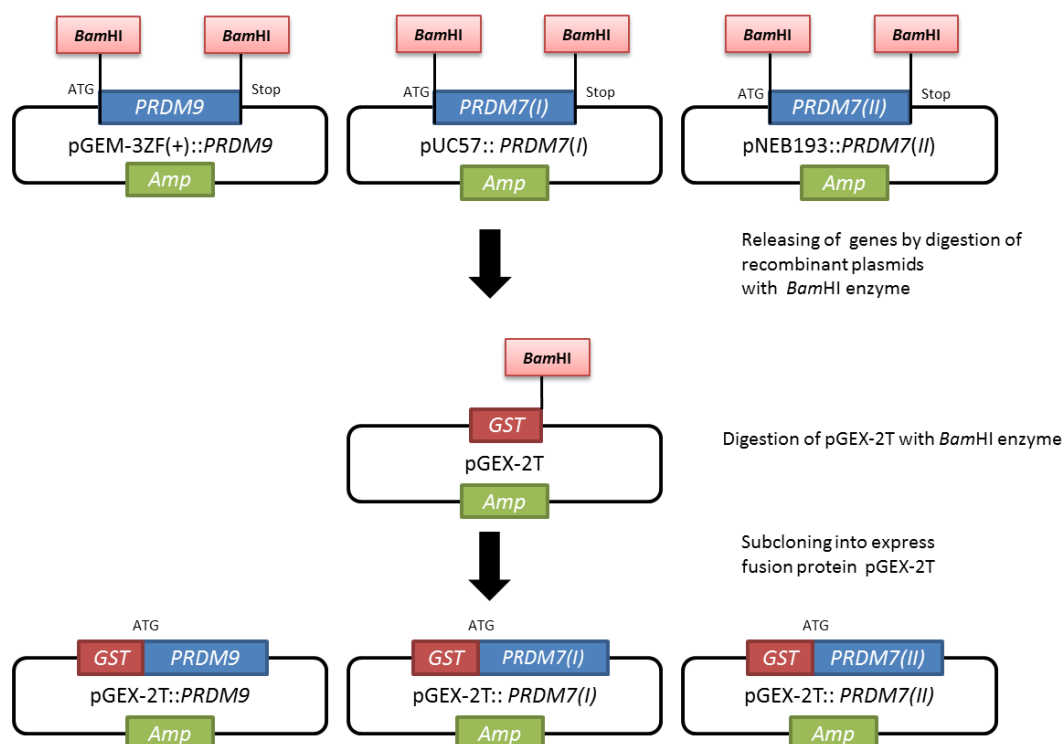
Figure 6.2. **Schematic diagram of the subcloning strategy for insertion of _PRDM9_ and _PRDM7_ genes into the fusion expression vector pGEX-2T.** Step 1: the gene of interest was released from the recombinant plasmids using the _Bam_HI restriction enzyme. Step 2: pGEX-2T plasmid was digested with _Bam_HI at the multiple cloning sites (MCS) after the GST taq site. Step 3: genes of interest were cloned into pGEX-2T; the proteins have an N-terminal GST tag which aids in recombinant protein purification.
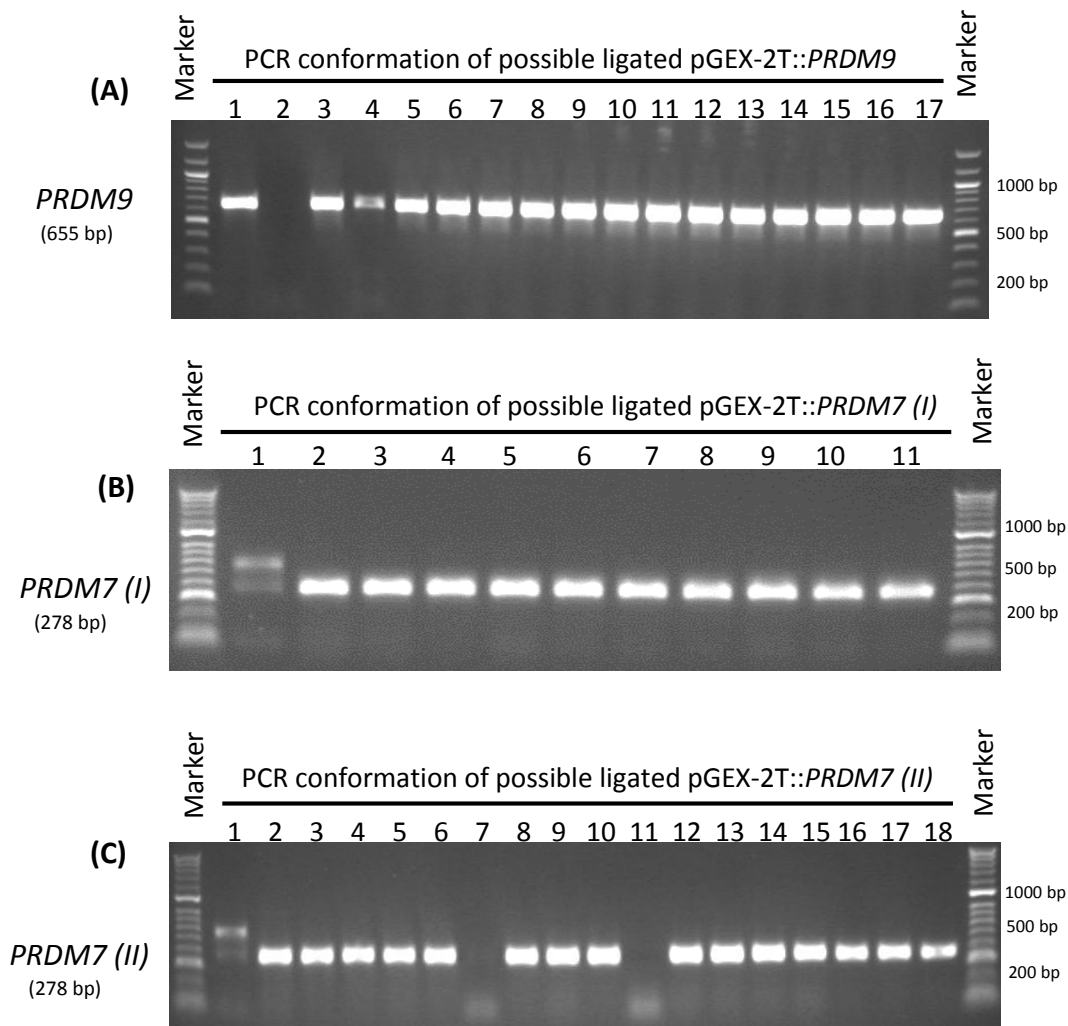
Figure 6.3. **PCR profile analyses of *PRDM9* and *PRDM7* genes after cloning into pGEX-2T expression vector and transformed into BL21 expression *E.coli*.** Genes were ligated at the *Bam*HI site with ratio of 1:6 and/or 1:9 (plasmid: insert) and transformed into BL21 expression *E. coli*. Random large white colonies were picked and PCR was performed to check the possibility of transformation using internal primers for *PRDM9* and *PRDM7* genes. PCR products were fractionated on 1% agarose gels stained with ethiduim bromide. (A) *PRDM9* gene; Lane 1: positive control testis; Lanes 2-17: colonies containing the *PRDM9* gene. (B) *PRDM7* gene long isoform; and (C) *PRDM7* gene short isoform; Lane 1: positive control testis; (B) Lanes 2-11: colonies containing *PRDM7* genes in the long isoform and (C) Lane 2-18: short isoform. 5 μl of HyperLadder II used as a marker.
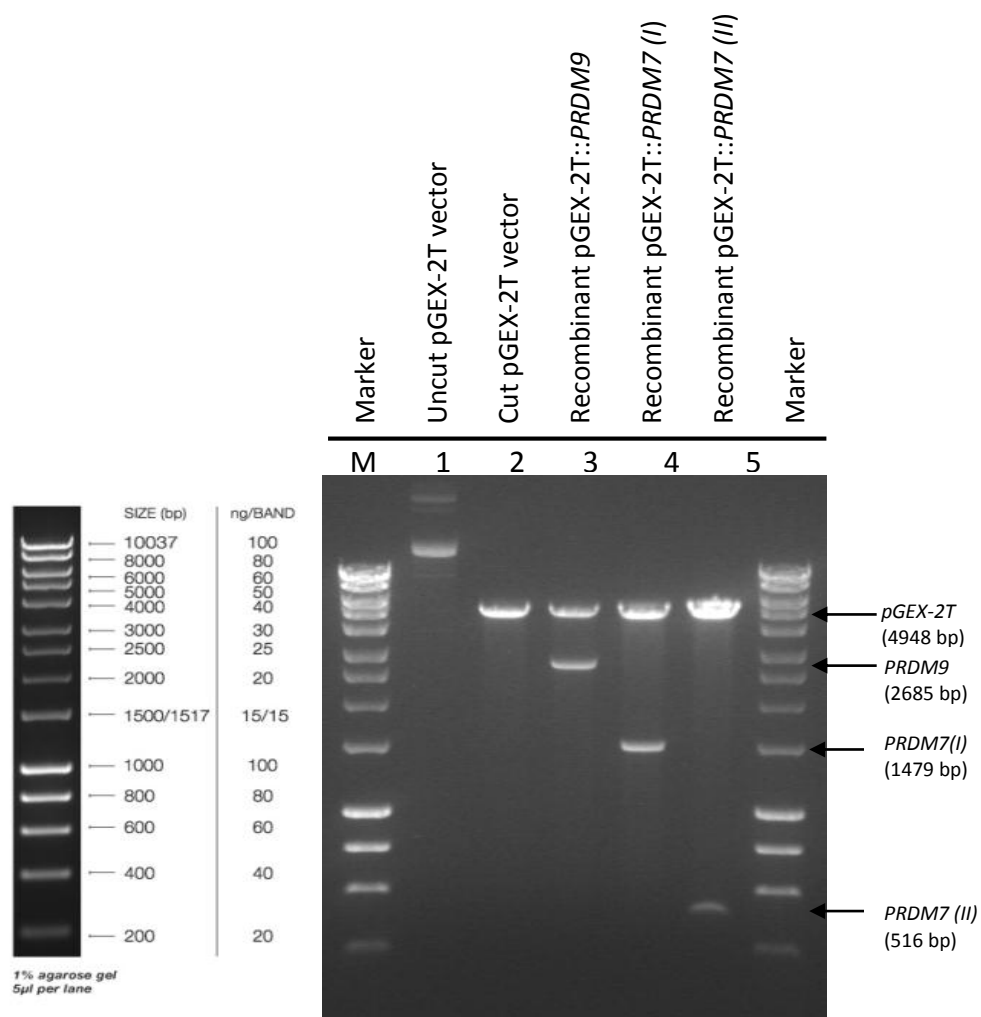
Figure 6.4. **Digestion of recombinant pGEX-2T expression plasmids with either *PRDM9* and/or *PRDM7* genes.** 0.8 % agarose gel stained with ethidium bromide confirm the cloning of *PRDM9*, *PRDM7* isoform (*I*) and (*II*) into the pGEX-2T plasmid. From the left, lane 1:uncut pGEX-2T plasmid; Lane 2: cut pGEX-2T plasmid with 4948 bp size; Lane3: digestion of recombinant *PRDM9:* pGEX-2T; Lane 4: digestion of recombinant *PRDM7 (I)* :pGEX-2T; Lane 5: digestion of recombinant *PRDM7 (II)*: pGEX-2T. 5 µl of HyperLadder 1 kb used as a marker.

6.2.2 Protein production

6.2.2.1 Optimising the overexpression of PRDM9 and PRDM7 with GST fusion proteins

Recombinant plasmids pGEX-2T:: *PRDM9*; pGEX-2T::*PRDM7* (*I*) and pGEX-2T:: *PRDM7* (*II*) were successfully transformed individually into the *E. coli* strain BL21 (DE3). Similarly, wild-type pGEX-2T plasmid was transformed into *E. coli* strain BL21 (DE3) to assess the protein production. The pGEX-2T plasmid contains of a 26 kDa glutathione-S-transferase (GST) which can express under the control of the IPTG-inducible *tac* promoter (Smith and Johnson, 1988). Production of recombinant proteins was induced in cells growing in LB media at 37°C by treatment with 100 µM IPTG for 2, 3 and 5 hours, as described in the Materials and Methods (Section 2.14.8). Total cell lysates of IPTG-induced and uninduced cells were resolved on 4-12% SDS-PAGE gels. Analysis of the induced PRDM9 fusion protein showed the expected size of 128 kDa corresponding to GST (≈ 26 kDa) fused to PRDM9 (≈ 102 kDa) (Figure 6.5 A; lanes 4, 5, 7, 8, 10 and 11). Similarly, the induction of fusion proteins PRDM7 (I) and PRDM7 (II) showed the correct expected sizes of approximately 81.7 kDa for fusion PRDM7(I): GST (Figure 6.5 B; lane 3, 5 and 7) and 44.8 kDa for fusion PRDM7(II): GST (Figure 6.5 C; lanes 4, 5, 7, 8, 10 and 11).

As a positive control, empty pGEX-2T plasmid was induced at 37°C with 100 µM IPTG for 2, 3 and 5 hours to assess the experimental conditions. The yield of fusion GST alone (≈ 26 kDa) and/or with fusion gene proteins showed a slight increase after 5 hours of induction with IPTG. No fusion protein was observed either with pGEX-2T alone or with the recombinant insert in the absence of IPTG (Figure 6.5).

Western blot analysis using anti-GST antibody was performed to assess the positive induction of recombinant plasmids. However, due to high background problems in western blot Coomassie blue stained gels were chosen to show the protein production in this chapter. Several conditions were optimised including induction under different temperatures (18°C, 25°C and 30°C), different IPTG concentrations and different lysis buffers to produce a high yield of recombinant proteins. The best inductions of PRDM9 and PRDM7 fusion proteins were obtained by treatment at 37°C with 100 µM IPTG for 5 hours (Figure 6.6).
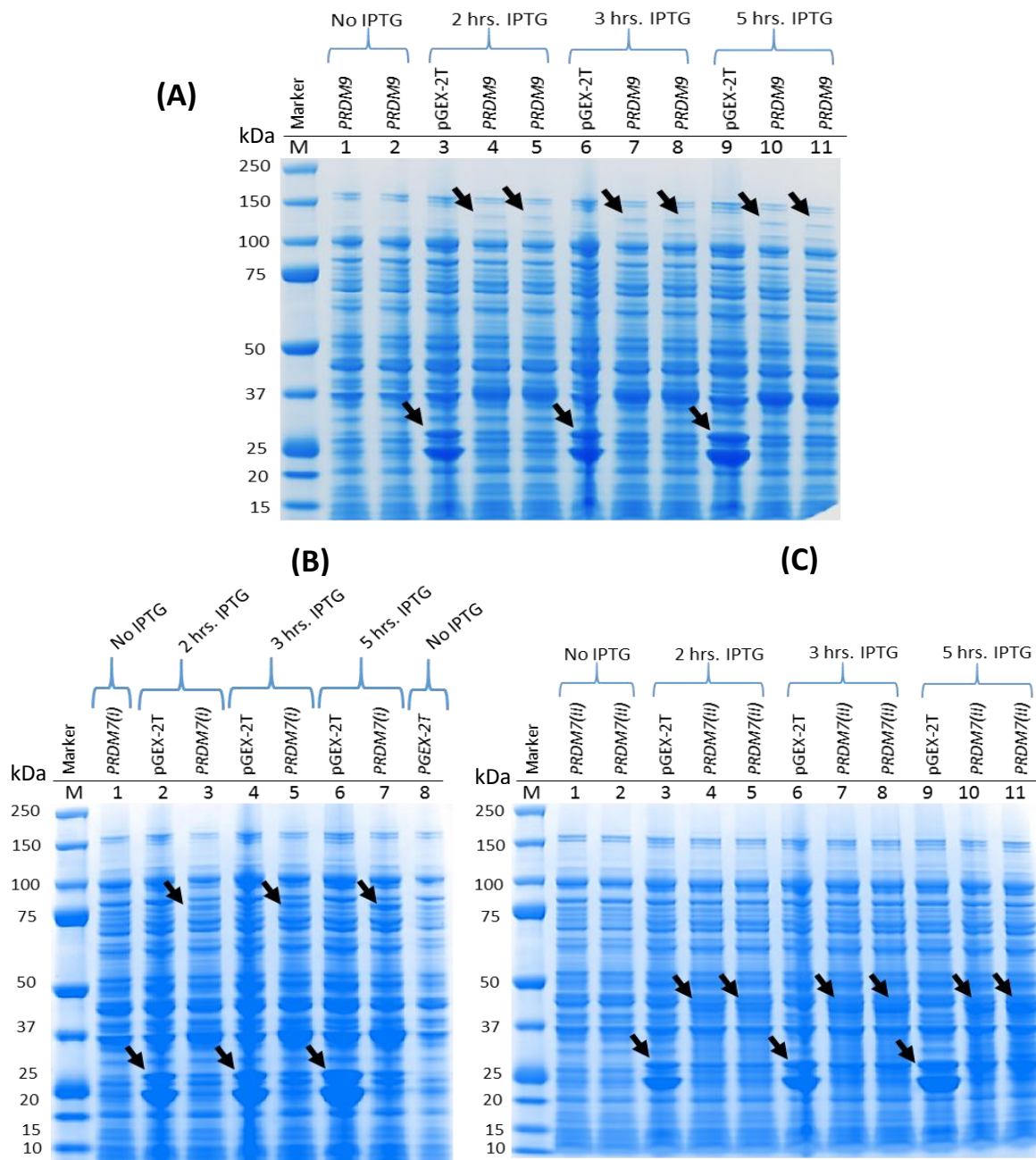
Figure 6.5. **Coomassie blue stained PAGE gel analysis of induction of the PRDM9 and PRDM7 proteins with 100 μM IPTG in BL21 *E. coli* cells.** Cells were grown in LB media and induced with 100 μM IPTG at 37°C for 2, 3 and 5 hours. (A) Fusion PRDM9 ≈128 kDa (B) PRDM7 (I) ≈ 81.7 kDa (C) PRDM7 (II) ≈ 44.81 kDa. Uninduced pGEX-2T alone or with inserts was used to assess the analysis. In addition, induced pGEX-2T without cloning (≈26 kDa) was added to assess the induction. M: protein marker; molecular weights are shown on the left in kDa.
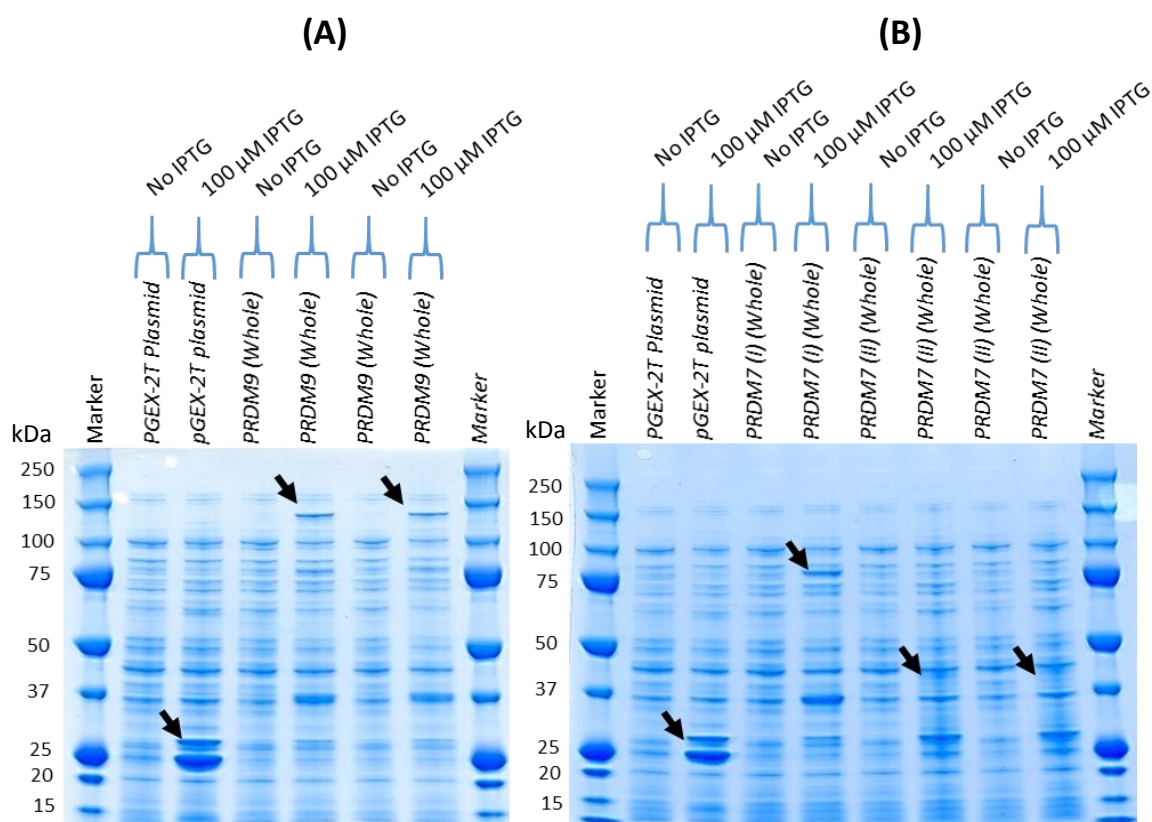
Figure 6.6. Optimised induction of fusion PRDM9 and PRDM7 into BL21 *E.coli* cells. Coomassie blue stained gels show the best optimised induction condition of PRDM9 and PRDM7 protein is induction with 100 µM IPTG at 37°C for 5 hours. A) Fusion PRDM9 ≈128 kDa (B) PRDM7 (I) ≈ 81.7 kDa and PRDM7 (II) ≈ 44.81 kDa. Uninduced pGEX-2T alone or with inserts was used to assess the analysis. In addition, induced pGEX-2T without cloning (≈26 kDa) was added to assess the induction. M: protein marker; molecular weights are shown on the left in kDa.

6.2.3 Determination of the solubility of overexpressed PRDM9 and PRDM7 GST fusion proteins

The solubility of overexpressed fusion proteins was examined under the best optimised condition (37°C with 100 µM IPTG for 5 hours) for both PRDM9 and PRDM7 (I) or with 200 µM IPTG for PRDM7(II), as described in the Material and Methods (Section 2.14.8). When both the pellet (insoluble) and the supernatants (soluble) of induced-IPTG cells were run on SDS-PAGE gels, the majority of the PRDM9 and PRDM7 (I) was observed in the insoluble fractions (data not shown).

Therefore, before starting the purification steps for PRDM9, PRDM7 (I) and PRDM7 (II) fusion proteins, different conditions were examined in order to solubilise more of the produced proteins. These conditions included different temperatures for fusion induction, different IPTG concentrations, different times of sonication, different lysis buffers and different host strains of *E. coli*. The majority of PRDM9 and PRDM7 (I) fusion proteins remained insoluble under all conditions tested.

PRDM9 fusion protein was induced with a range of IPTG concentrations from 40 to 400 µM for 5 hours at 37°C. The cell lysate was resuspended in sodium chloride-Tris-EDTA (STE) buffer containing 100 µg/ml lysozyme to increase the solubility of the produced protein, as described in the Materials and Methods (Section 2.14.8.1). However, the majority of PRDM9 fusion protein was still found in the insoluble fraction (Figure 6.7; lanes 4, 7, 10 and 14). Different IPTG-induction temperatures of 18°, 25° and 30°C were used to solubilise PRDM9 in the BL21 (DE3) *E.coli* strain but the bulk of the fusion protein remained in the insoluble fraction (data not shown).

Similarly, reductions in the IPTG concentration and temperature to 25° or 30°C for fusion of PRDM7 (I) were tested to increase the amount of soluble protein before the purification step (Figure 6.8). Decreasing the induction temperature to 30°C with 100 µM IPTG increased the amount of recovered soluble protein in the supernatant but a lot of protein remained in the insoluble fraction (Figure 6.8; lanes 5, 8 and 11). The solubility of the fusion PRDM7 (II) protein was also tested under different temperatures and IPTG induction. Using 40 to 200

µM IPTG for induction at 37°C for 5 hours was sufficient to produce sufficient soluble PRDM7 (II) to use in the purification step (Figure 6.9; lanes 5, 8, 11 and 15).

SoluBL21 chemically competent cells (Genlantis) have been proposed to enhance the solubility of insoluble fusion proteins and limit the toxicity of recombinant proteins (Genlantis; http://www.genlantis.com). Recombinant plasmids as well as the pGEX-2T plasmid were transformed individually into SoluBL21 cells, which were then induced with 100 µM IPTG in Minimal M9 medium at room temperature in a shaking incubator at 200 r.p.m for 24 hours (as per the manufacturer's protocol). However, no induction of overexpression of fusion PRDM9 protein was observed in the SoluBL21 cells (Figure 6.10 A). The majority of the fusion PRDM7 (I) and PRDM7 (II) proteins were again produced in an insoluble protein form in these competent cells (Figure 6.10 B). Analysis of SoluBL21 cells with transformed pGEX-2T plasmid only showed a high level of GST protein (approximately ≈26 kDa) in both soluble and insoluble protein fractions. No production of GST fusion protein was detected in the absence of IPTG induction.
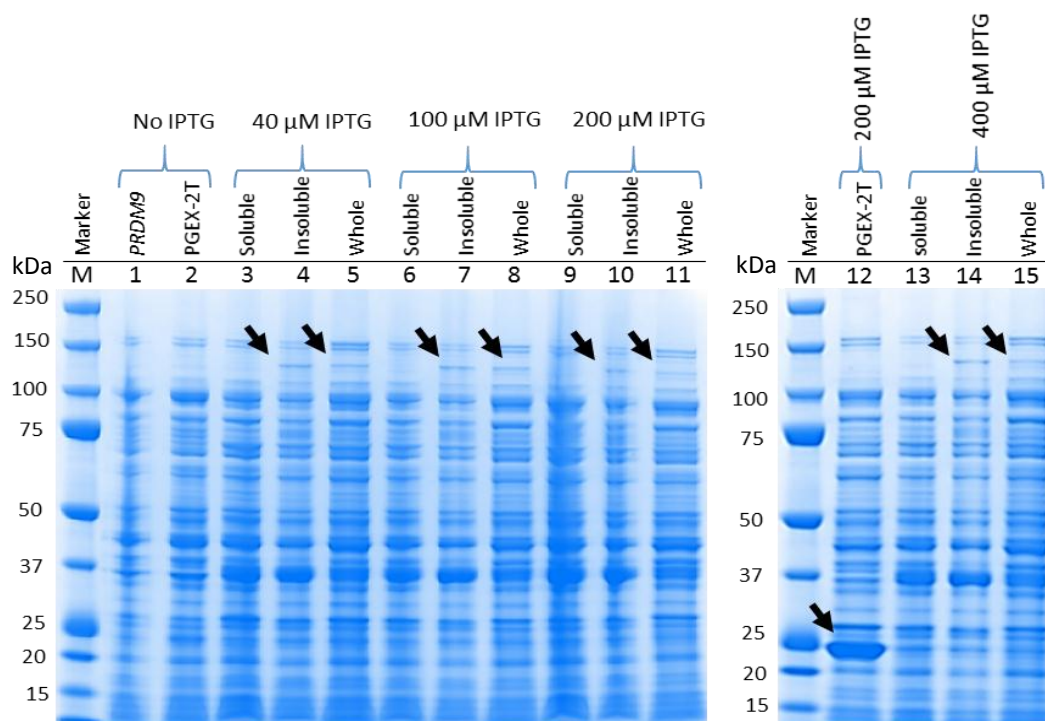
Figure 6.7. Determination of the solubility of PRDM9 fusion protein using different concentrations of IPTG in BL21 *E. coli* cells. Coomassie blue stained gels show the induction of fusion PRDM9 protein ≈128 kDa following induction with 40, 100, 200 and 400 µM IPTG at 37°C for 5 hours. Fusion PRDM9 shows an insoluble fraction (arrow) at all IPTG concentrations. Uninduced pGEX-2T alone or with PRDM9 was used to assess the analysis. In addition, induced pGEX-2T without the insert (≈26 kDa) was added to assess the induction. M: protein marker; molecular weights are shown on the left in kDa.
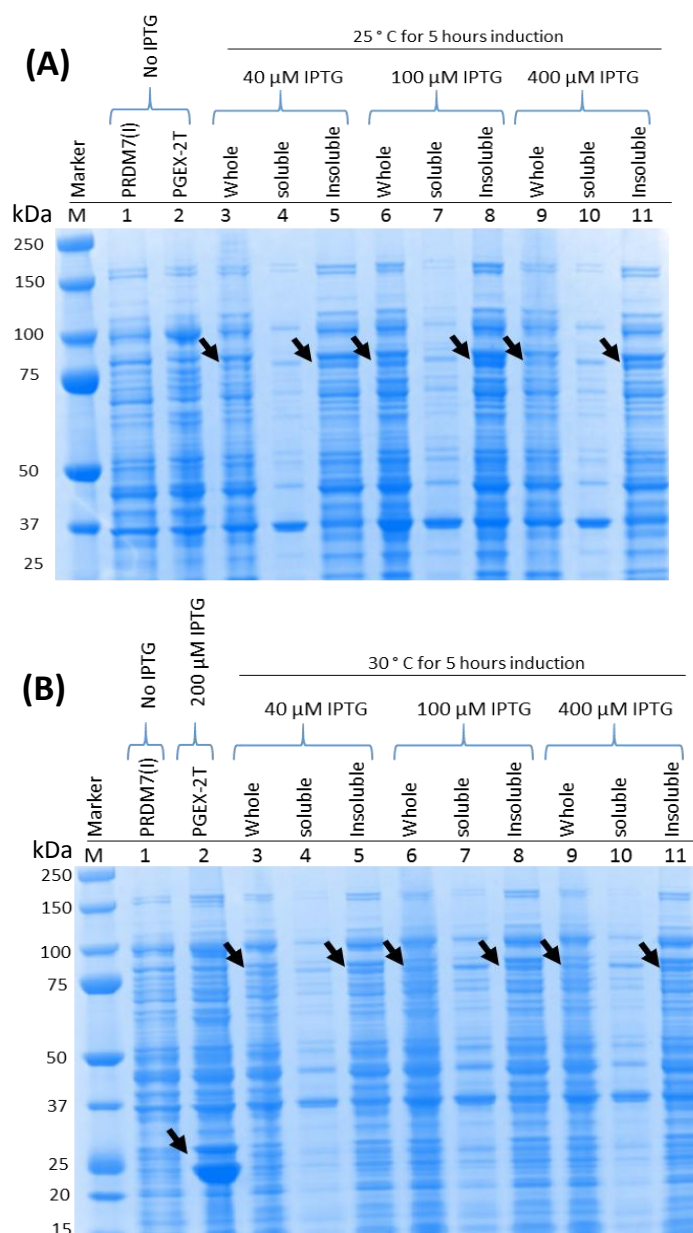
Figure 6.8. **Determination of the solubility of PRDM7(I) fusion protein at 25°C and 30°C using different concentrations of IPTG into *E.coli* (BL21).** Coomassie blue stained gels show the induction of fusion PRDM7 (I) ≈ 81.7 kDa protein with 40 μM, 100 μM and 400 μM of IPTG. The majority of the IPTG-induced protein remained insoluble at (A) 25°C or (B) 30°C. pGEX-2T alone or with PRDM9 in uninduced cells was used to assess the analysis. In addition, induced pGEX-2T without insert (≈26 kDa) was added to assess the induction. M: protein marker; molecular weights are shown on the left in kDa.
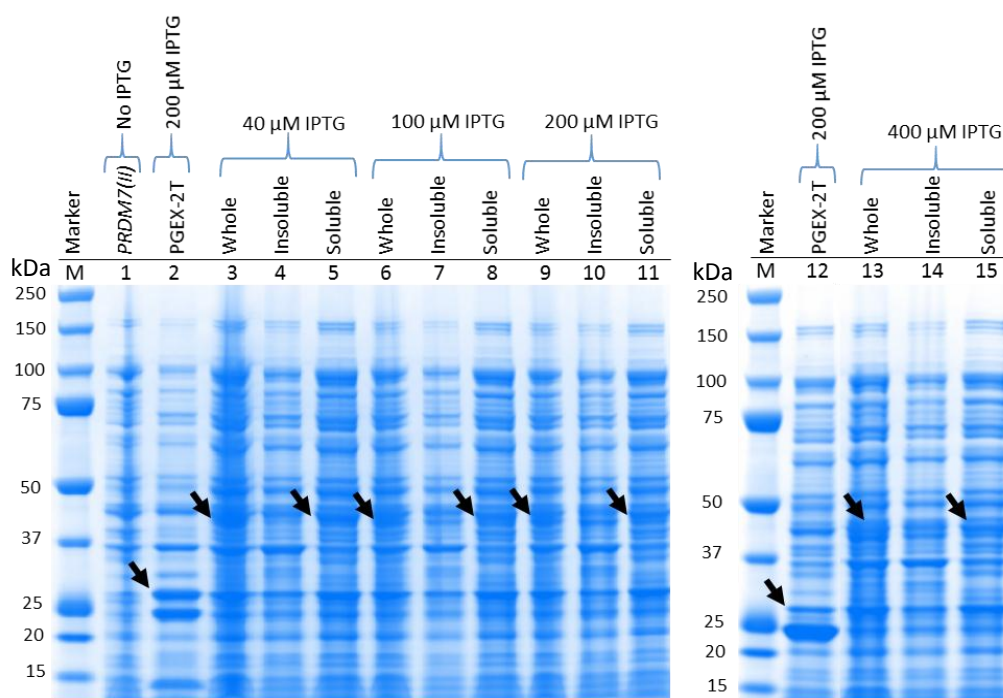
Figure 6.9. **Determination of the solubility of fusion PRDM7 (II) protein using different concentration of IPTG into *E.coli* (BL21).** Coomassie blue stained gels show the induction of fusion PRDM7 (II) protein ≈ 44.81 kDa with 40, 100, 200 and 400 µM of IPTG at 37°C for 5 hours. Fusion PRDM7 (II) shows soluble fraction (arrow) with 40, 100 and 200 µM of IPTG concentration. pGEX-2T alone or with PRDM9 in uninduced cells was used to assess the analysis. In addition, pGEX-2T without insert (≈26 kDa) was added to assess the induction. M: protein marker; molecular weights are shown on the left in kDa.
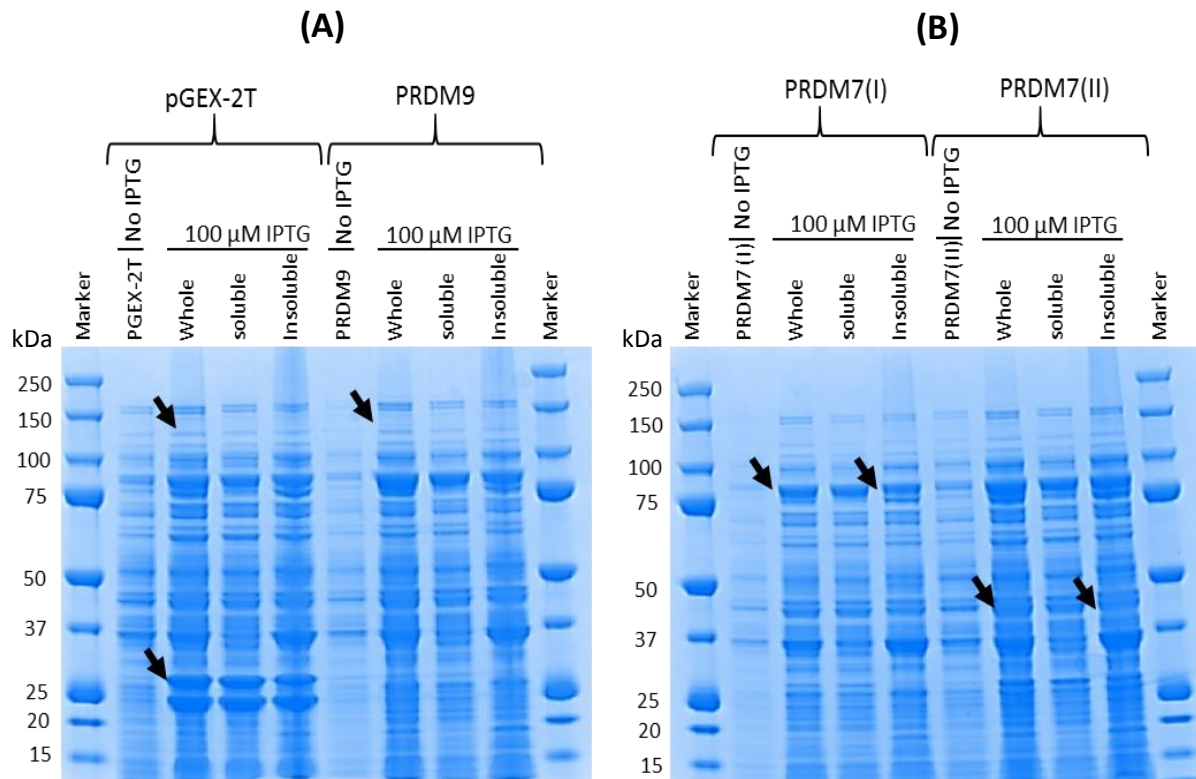
Figure 6.10. **Solubility analysis of fusion PRDM9 and PRDM7 proteins with 100 µM IPTG into SoluBL21 *E.coli* cells**. Coomassie blue stained gels showed (A) no expression of fusion PRDM9 protein after 24 hours of induction with 100 µM of IPTG at room temperature. (B) Fusion PRDM7 (I) ≈ 81.7 kDa and PRDM7 (II) ≈ 44.81 kDa appeared to have an insoluble fraction. Cells were grown in M9 medium at room temperature for one day and diluted in fresh M9 medium. Induction of the cells was performed at OD_{600} ≈0.4 with 100 µM of IPTG at room temperature for 24 hours (as per the manufacturer's protocol). pGEX-2T alone or with PRDM9 in uninduced cells was used to assess the analysis. In addition, induced pGEX-2T without insert (≈26 kDa) was added to assess the induction. M: protein marker; molecular weights are shown on the left in kDa.

6.2.4 The purification of produced fusion proteins

Solubility analysis of the full lengths protien of PRDM9 and PRDM7 (I) under different culture conditions revealed that the vast majority of the fusion protein was located in the insoluble fraction. Purification of the soluble fusion proteins of PRDM9, PRDM7 (I) and PRDM7 (II) was performed using Pierce GST spin purification kits (Thermo) and GST spin traps (GE Healthcare) as described in the Materials and Methods (Section 2.14.8). Purification of these soluble fractions (the supernatant) resulted in no purified proteins.

We next examined the possibility of purification of the whole induced *E. coli* lysates. IPTG induced cell lysates were re-suspended in STE buffer containing 100 µg/ml of lysozyme. After sonication, the soluble materials were recovered by centrifugation and Triton X-100 was added to a final concentration of 2%. Purification of induced pGEX-2T without the insert showed only a clear band with a size of ≈26 kDa (Figure 6.11 A; lanes 7, 8, 9 and 10). No fusion PRDM9 protein was purified after elution steps with glutathione elution buffer (Figure 6.11 B; lane 8, 9 and 10). Some of the fusion PRDM9 protein was observed to pass through the first flow-through centrifugation of the whole sample (Figure 6.11 B; lane 3).

Similarly, no purified PRDM7 (I) fusion protein was observed in the elution buffer even with extensive elution (Figure 6.11 C; lane 8, 9 and 10). However, some of the fusion PRDM7 (I) protein was noticed in the first flow-through centrifugation (Figure 6.11 C; lane 3). The PRDM7 (II) purification displayed some purified fusion protein but also showed some degraded protein (Figure 6.11 D; lanes 8, 9 and 10). Different approaches were examined to reduce the protein degradation of fusion PRDM7 (II) including low temperature growth and various type of inhibitor cocktails and lysis buffer but none were successful.

The glutathione resins of the purification column were collected and boiled to test if the fusion proteins were still bound to the resins after the elution steps. The vast majority of fusion PRDM9, PRDM7 (I) and PRDM7 (II) remained associated with the column resins (Figure 6.12). Therefore, formic acid was added to the elution buffer at a final concentration of 2%. Formic acid was chosen because it has been shown to enhance the elution of insoluble fractions without altering protein structure and function (Lee *et al.*, 2008). Unfortunately, extensive elution steps using elution buffer containing 2% formic acid did not release the

fusion PRDM9 protein and PRDM7 (II) protein (Figure 6.13). Formic acid elution slightly enhanced the purification of fusion PRDM7 (I), with a small amount obtained after the first elution. Clear bands with estimated size ≈26 kDa were observed, indicating purification of induced pGEX-2T plasmid only in formic acid elution buffer (Figure 6.13).

The use of mild detergents such as sarkosyl with a combination of Triton X-100 (1:200 ratio; S-T) in the lysis buffer has been suggested to facilitate the purification of insoluble proteins in high yields and without effects on the fusion protein functions (Park *et al.*, 2011). However, the use of 1-2% sarkosyl with Triton X-100 at a ratio of 1:200 (sarkosyl:Triton X-100) did not improve the purification of insoluble fractions of PRDM9 and PRDM7. Again, a clear band was obtained with expected size of the induced pGEX-2T plasmid only (Figure 6.14).
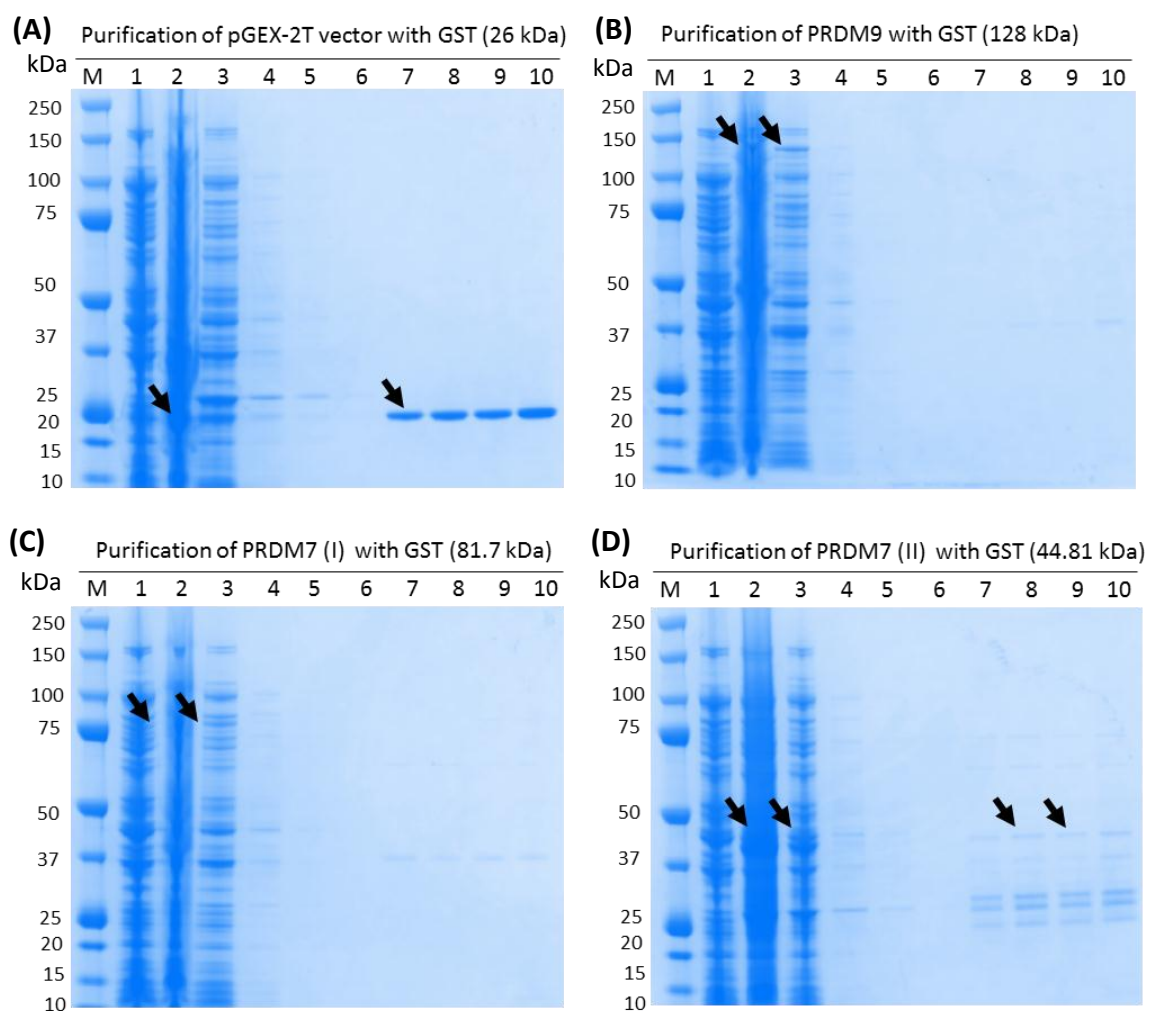
Figure 6.11. **Coomassie blue stained 4-12% SDS-PAGE gels analysis after expression and purification of fusion PRDM9 and PRDM7 proteins in BL21 _E.coli_ cells.** _E. coli_ cells were grown in LB media at 37°C with fusion proteins containing either (A) pGEX-2T only, or with (B) PRDM9, (C) PRDM7 (I) or (D) PRDM7 (II) and induced with 100 μM IPTG for 5 hours. The whole cells lysate were solubilized using 2% Triton X-100 and purified using GST thermo purification columns. From left to right lane M: Protein marker; 1: Non-induced whole BL21 _E.coli_ lysate; 2: Whole cell lysate of cells induced with 100 μM IPTG; 3: Flow-through fractions from the column; 4: First wash with Equilibration/wash buffer; 5: Second wash; 6: Third wash; 7: First elution buffer centrifugation; 8: Second elution buffer; 9: Third elution; 10: Fourth elution. M: protein marker; molecular weights are shown on the left in kDa.
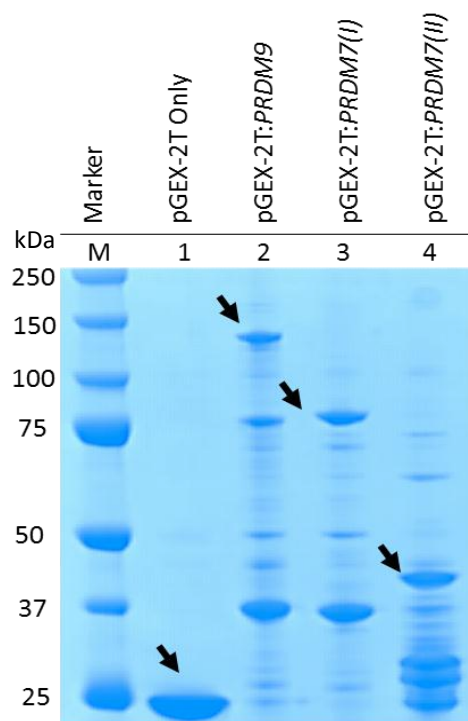
Figure 6.12. **Coomassie blue stained 4-12% SDS-PAGE gel of boiled column resins after purification of PRDM9 and PRDM7 from BL21 *E. coli* lysate.** Resins were collected and boiled after the purification of fusion proteins by the Triton X-100 purification method. The vast majority of produced proteins remained bound to the column resins. From the left to right M: protein marker; 1: GST protein only; 2: Fusion PRDM9 protein; 3: Fusion PRDM7 (I) protein; 4: Fusion PRDM7 (II) protein. M: protein marker; molecular weights are shown on the left in kDa.

(A) Purification of pGEX-2T vector with GST (26 kDa)

(B) Purification of PRDM9 with GST (128 kDa)

(C) Purification of PRDM7 (I) with GST (81.7 kDa)

(D) Purification of PRDM7 (II) with GST (44.81 kDa)
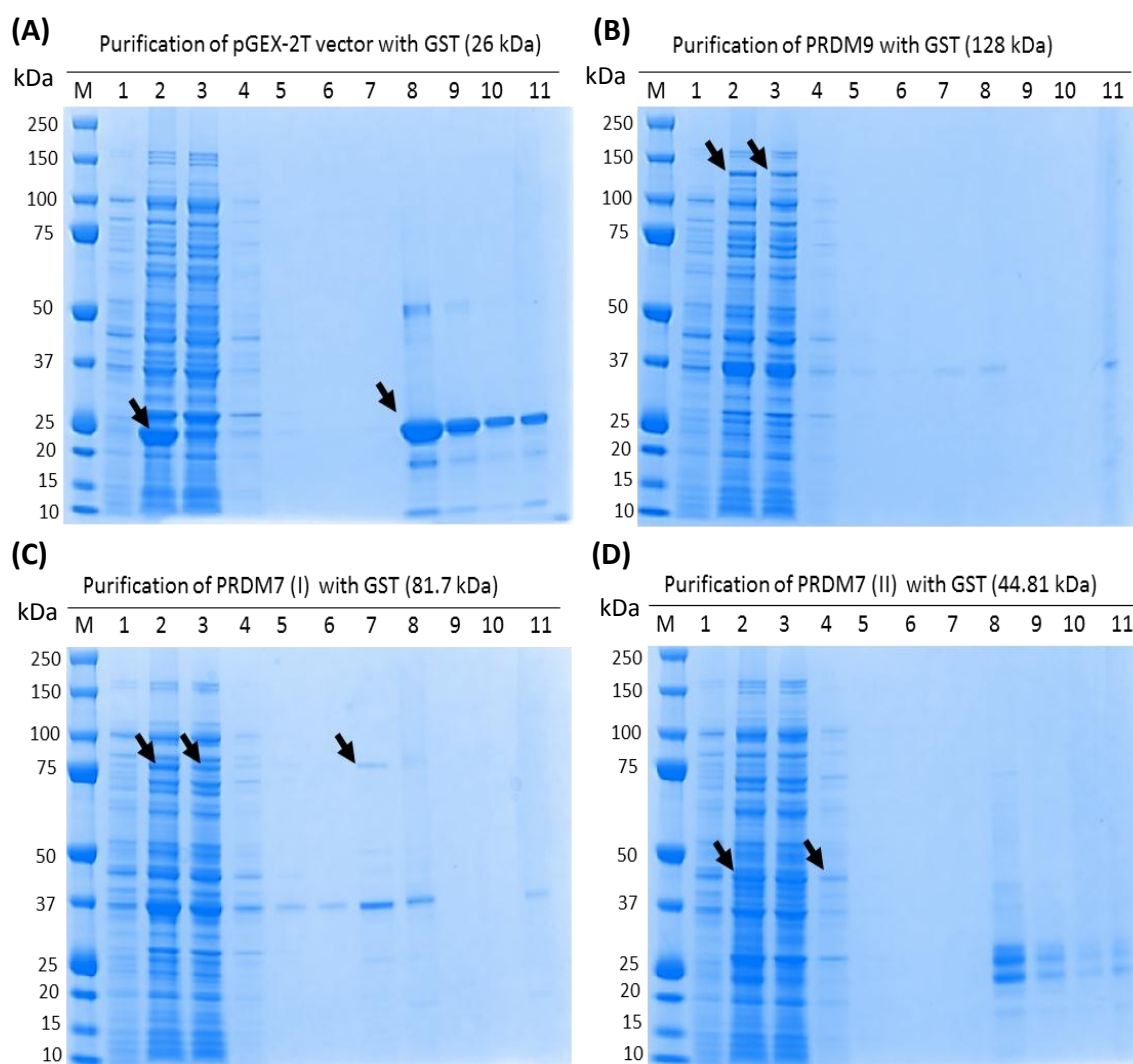
Figure 6. 13. Coomassie blue stained 4-12% SDS-PAGE gels after purification of PRDM9 and PRDM7 genes from BL21 _E.coli_ cells and elution with formic acid. _E. coli_ cells were grown in LB media at 37°C with fusion proteins containing either (A) pGEX-2T only, or with (B) PRDM9, (C) PRDM7 (I) or (D) PRDM7 (II) and induced with 100 µM IPTG for 5 hours. The whole cell lysates in STE buffer were solubilized using 2% Triton X-100 and purified using GST thermo purification columns. An extensive elution with 2% of formic acid was used with the elution buffer to enhance the elution of insoluble fractions bound to the GST column. From left to right lane M: Protein marker; 1: Non-induced whole BL21 _E.coli_ lysate; 2: Whole cell lysate of cells induced with 100 µM IPTG; 3: Flow-through fractions from the column; 4: First wash with Equilibration/wash buffer; 5: Second wash; 6: Third wash; 7: Pre-elution 8: First elution with 2% of formic acid buffer; 9: Second elution; 10: Third elution; 11 Fourth elution. M: protein marker; molecular weights are shown on the left in kDa.
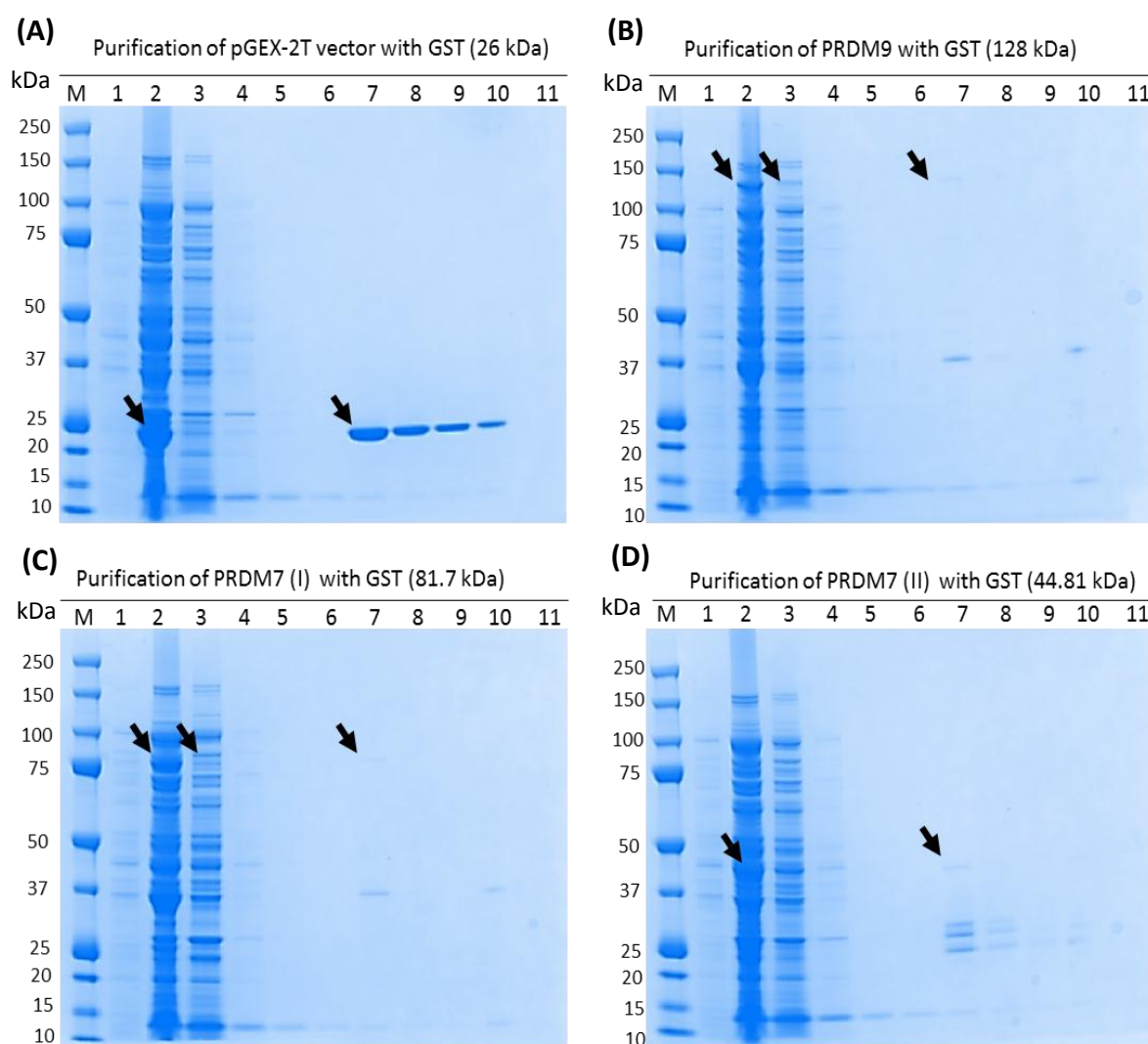
Figure 6.14. **Purification of PRDM9 and PRDM7 genes from BL21 *E.coli* cells using sarkosyl detergent.** *E. coli* cells were grown in LB media at 37°C with fusion proteins containing either (A) pGEX-2T only, or with (B) PRDM9, (C) PRDM7 (I) or (D) PRDM7 (II) and induced with 100 µM of IPTG for 5 hours. The whole cell lysates in STE buffer were solubilized in 2% sarkosyl:Triton X-100 (1:200) and purified using GST thermo purification columns. An extensive elution with 2% of formic acid was used with the elution buffer to enhance the elution of bound insoluble fractions to the GST column. From left to right: Lane M: Protein marker; 1: Non-induced whole BL21 *E. coli* lysate; 2: Whole cell lysate of cells induced with 100 µM IPTG; 3: Flow-through fractions from the column; 4: First wash with Equilibration/wash buffer; 5: Second wash; 6: Third wash; 7: First elution; 8: Second elution; 9: Third elution; 10: Fourth elution; 11: Fifth elution. M: protein marker; molecular weights are shown on the left in kDa.

6.2.5 Histone methyltransferase assay

Several attempts were made to purify the fusion PRDM9 (≈128 kDa) and PRDM7 (≈81.7 kDa and 44.81 kDa) proteins from *E. coli* cells but all attempts failed. Generally, *E. coli* cells did not express very well or did not promote the correct folding of very large proteins (>70 kDa). In this case, PRDM9 and PRDM7 were often found as insoluble proteins and accumulated as inclusion bodies.

In vitro assays of histone methyltransferases activity (HMTs) are preferably conducted using purified proteins to eliminate reactions due to contamination with endogenous histone methyltransferases. However, an attempt was made to use the whole bacterial lysates of induced PRDM9 or PRDM7 cells using the EpiQuik™ Histone Methyltransferase Activity/Inhibition Assay Kit (H3-K4) (Epigentek; cat no. P-3002). This kit was suggested for the detection of histone methyltransferase activity using whole cell lysates (Personal communication with the provider company).

The principle of this kit is a colorimetric assay based on measuring the HMT activity that targeted a particular histone H3 at lysine K4. In this assay, each strip well consists of a strongly captured histone substrate. A methyl group transfer from occurs from S-adenosylmethionine (Adomet) onto the histone H3 at lysine 4 by HMT enzymes. High-affinity antibody is then used to detect the methylated histone H3-K4. Then, HRP conjugated secondary antibody is used to determine the amount of methylated H3-K4 based on colour development (Figure 6.15). The HMT activity can be then calculated based on the amount of methylated H3-K4 converted by the HMTs using the following formula:

$$\text{Activity (ng/h/mg)} = \frac{\text{OD (Sample} - \text{blank} *)}{\text{Protein Amount (µg) x hour x slope} **} \text{ x } 1000$$

*   For the blank, only histone buffer is added instead of whole cell lysate
** Slope is calculated using a standard curve of known concentration provided by the manufacturer

Based on the data results, the assay appears to have worked since the OD of the blank was ≈ 0.07 ng and the positive control SET7/9 protein OD was ≈ 1.131 ng (activity ≈ 404.96) within

the expected range and the sample values were higher than the blank (Table 6.1). Whole lysates of *E. coli* cells with empty pGEX-2T had the highest activity, the reasons for this are unclear as *E. coli* should not possess histone methyltransferase activity. However, PRDM7 and PRDM9 lysates showed less activity than the induced empty plasmid that could be detected by the capture antibody. Induction of PRDM7 and PRDM9 with IPTG resulted in a slight increase in activity, but it was still less than that seen for the induced pGEX-2T alone. The capture antibody was designed to detect dimethyl H3K4 but not trimethylated H3K4 (Personal communication with the company, as they did not mention this information on the website and/or in the manufacturer's instructions). Therefore, the lower OD might reflect higher activity of PRDM7 and PRDM9 since the dimethylated products could have been converted to trimethylated products which cannot be detected by the capture antibody. The OD value was inversely proportional to the activity of PRDM7 or PRDM9.

PRDM9 was demonstrated to catalyse the production of tri-methylated lysine 4 of histone H3 (H3K4me3) but did not catalyse mono-or di-methylation in mice (Hayashi *et al.*, 2005). However, very recently, different groups demonstrated that PRDM9 *in vitro* has the ability to catalyse the mono-, di- and trimethylation of lysine 4 of histone H3 histone in mice (Wu *et al.*, 2013; Xiaoying *et al.*, 2014). Furthermore, PRDM9 was identified to have the ability to catalyse the mono-, di- and trimethylation of H3K4 as well as H3K36 in humans (Eram *et al.*, 2014). All these groups used only the PR/SET domain fragment of PRDM9 (195-385 amino acids) instead of the full-length protein which were used in this project.

Nuclear Extract

↓

Incubate with substrate and assay buffer for 60-90 minutes

↓

Add capture antibody after wash

↓

Add detection antibody

↓

Add developing solution for developing colour and measure absorbance
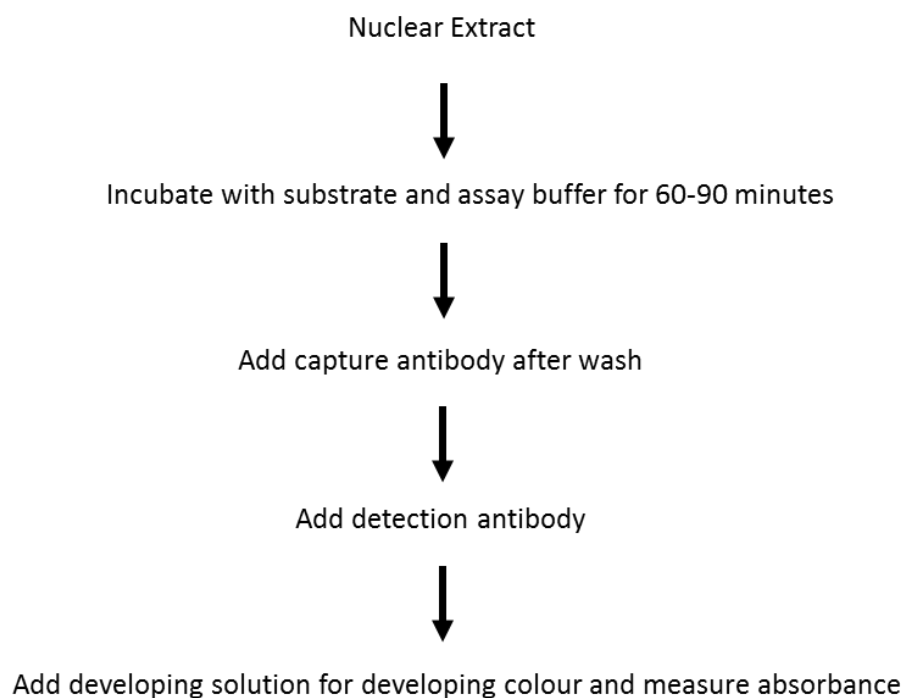
Figure 6.15. Summary of the protocol for determining the histone methyltransferase activity of PRDM9 and PRDM7 using the EpiQuik histone methylation assay kit.

__Table 6.1.__ **Calculation of the histone methyltransferase activity of PRDM7 and PRDM9**

	pGEX-2T no IPTG	pGEX-2T 100 μM IPTG	*PRDM9* no IPTG	*PRDM9* 100 μM IPTG
Sample (OD)	0.883	0.981	0.433	0.705
Blank (OD)	0. 07	0. 07	0. 07	0. 07
x1000	1000	1000	1000	1000
protein amount (μg)	10	10	10	10
Hour	1	1	1	1
Slop	0.262	0.262	0.262	0.262
Activity	310.3053435	347.7099237	138.5496183	242.3664122
	PRDM7 (I) no IPTG	*PRDM7 (I)* 100 μM IPTG	*PRDM7 (II)* no IPTG	*PRDM7 (II)* 100 μM IPTG
Sample (OD)	0.496	0.534	0.596	0.755
Blank (OD)	0. 07	0. 07	0. 07	0. 07
x1000	1000	1000	1000	1000
protein amount (μg)	10	10	10	10
Hour	1	1	1	1
Slop	0.262	0.262	0.262	0.262
Activity	162.5954198	177.0992366	200.7633588	261.4503817

6.3 Discussion

Genetic recombination is an important feature of meiotic prophase I and a key source of genetic information exchange for the generation of new combinations of alleles in a population (Baudat *et al*., 2013). Recombination events occur in a small segment of the DNA called hotspots, which are approximately 1 to 2 kb long (Jeffreys *et al.*, 2001). These specific sites are encoded for trimethylation of lysine 4 of Histone H3 (H3K4), which is a mark of recombination initiation (Borde *et al*., 2009; Buard *et al.,* 2009; Smagulova *et al*., 2011). In mice and humans, the PRDM9 protein has been identified to have a critical role in regulation of recombination by binding to the hotspots via its long zinc finger array (Parvanov *et al.,* 2010*;* Baudat *et al*., 2010; Myers *et al*., 2010). In mice, PRDM9 shows *in vitro* catalytic activity for tri-methylation of lysine 4 of histone H3 (H3K4me3) but not mono-or di-methylation. It also has a transcriptional activity that relies on its methylation activity (Hayashi *et al*., 2005). PRDM7 was reported to arise as a duplicate of PRDM9, where both genes share highly similar DNA sequences and structures (Fumasoni *et al.*, 2007).

The aim of this chapter was to subclone, express and purify *PRDM9*, *PRDM7* (*I*) and (*II*) in the protein production pGEX-2T plasmid with the goal of studying the histone methyltransferase activity of purified PRDM9 and PRDM7 *in vitro*. The full lengths of human *PRDM9* and *PRDM7* (*I*) and *(II)* genes were successfully cloned into the pGEX-2T vector and overexpressed using IPTG. Several conditions were optimised to obtain a high yield of recombinant proteins. These included different induction times and temperatures, different IPTG concentrations, different cell lysis buffers and different host cells.

Fusion PRDM7 and PRDM9 proteins were produced in insoluble form and accumulated as inclusion bodies. Several attempts were made to purify the insoluble fraction of PRDM7 and PRDM9 proteins, including using different GST elution buffers in combination with formic acid or mild detergents such as sarkosyl, but all attempts failed. One additional concern is that the use of detergent to solubilise the inclusion bodies also might inactivate the enzymatic activity.

Failure to produce soluble recombinant PRDM9 (≈128 kDa) and PRDM7 (≈81.7 kDa) proteins might be due to improper protein folding that occurred during the protein production.

Different factors can lead to recombinant protein misfolding including recombinant protein size (>70 kDa), use of high concentrations of IPTG and use of strong host promoters, as well as use of a high expression temperature (Baneyx and Mujacic, 2004).

The histone methyltransferase activities of PRDM9 and PRDM7 were measured using a colorimetric assay kit. This assay is able to detect the dimethylation of H3K4 using the capture antibody provided. The measurements for the positive control SET7/9 protein and the blank samples were in the expected range. However, low activity was detected for PRDM7 and PRDM9 after protein induction with IPTG. This may be a result of different factors, including the specificity of the capture antibody and the real amount of protein used since the whole cell lysate was applied.

Mouse PRDM9 was recently reported to catalyse not only tri-methylation but also mono- and dimethylation of H3K4 *in vitro* (Wu *et al*., 2013; Xiaoying *et al*., 2014). PRDM9 also methylated the other core histones H2A, H2B and H4, but with less modification than seen for H3. A single C321P mutation in the PR/SET domain led to weak activity of PRDM9 through inhibition of interaction with S-adenosylmethionine (SAM) (Wu *et al*., 2013; Xiaoying *et al*., 2014). In addition, very recently, human PRDM9 was demonstrated to catalyse mono-, di- and trimethylation of H3K4 as well as H3K36 *in vitro* and in HEK293 cells (Eram *et al*., 2014).

Several transcriptional regulator proteins have been demonstrated to reactivate the expression of some CTA genes, including those that are aberrantly activated in cancer cells (e.g., BORIS, BRDT and ATAD2). These genes encode proteins that act as transcriptional factor regulators and/or potentially affect chromatin structures (reviewed in De Smet and Loriot 2013; Whitehurst, 2014). PRDM9 also encodes a protein that is able to bind to 13-mer specific DNA sequences and has histone methyltransferase activity with H3K4 and H3K36 (Eram *et al*., 2014).

The remaining question is how the aberrant expression of PRDM9 in cancer cells might activates the gametogenesis programme in cancer cells that, in turn, could drives the expression of gametogenic genes in tumour cells. A second unaddressed question is whether aberrant expression of PRDM9 might lead to unstable chromatin lesions that then activate CTA gene expression in cancer cells. These questions need further investigation to identify

the relationship between activation of PRDM9 and aberrant expression of CTA genes (or other genes) in cancer cells.

However, small molecule drugs that target methylation-dependent pathways are now attracting considerable interest in cancer therapy. Selective inhibitors of DOT1L and EZH2 histone lysine methyltransferases have been shown to target cancer cells with minimal influence on the proliferation of normal cells. These inhibitors have now entered Phase I clinical trials for MLL-rearranged leukaemia and diffuse large B-cell and follicular lymphoma (reviewed in Copeland, 2013; Wigle and Copeland, 2013; Xiaoying *et al.*, 2014). The successful use of these inhibitors indicates the possible application of other small molecule drugs that target histone lysine methyltransferase proteins, including the PRDM protein family (Xiaoying *et al.*, 2014). These genes, and particularly PRDM7 and PRDM9, might be especially attractive as potential targets for epigenetic drugs due to their aberrant expression in different types of cancers and their lack of expression in normal human tissues.

6.4 Future Work

Recombinant production of protein is one the most powerful tools in biological study and has a wide range of applications. All previous studies of PRDM9 were based on cloning of the PR/SET domain with an additional flanking sequence, either in mice or humans, instead of the full-length protein. Production of the full length of human PRDM9 and PRDM7 using another induction system such as baculovirus is required for investigation of their catalytic activity. To date, the activity of PRDM7 has not been characterised. In this case, the PR/SET domain of PRDM7 must be subcloned and induced to investigate its ability for histone methyltransferase activity.

PRDM9 was proposed to bind to specific DNA fragments called hotspots through its zinc finger domain. Gel shift assays could be performed to study the interaction between purified PRDM9 protein and a specific DNA sequence. Use of another vector such as the pET expression system with His tagging rather than the GST system might induce soluble PRDM7/9.

The effect of single or multiple mutations in the PR/SET domain of PRDM7 and PRDM9 must be studied to determine the influence of mutation on the interaction with SAM.

Furthermore, the activity of histone methyltransferase should be studied in HeLa Tet-on 3G cells overexpressing PRDM9 and PRDM7 (Chapter 5.0). Finally, the HMT activity must be determined using a radioactive assay to measure the activity of PRDM9 and PRDM7.

Chapter 7.0 General Discussion

7.1 RT-PCR analysis of potential CT genes in normal and cancer tissues

The identification and characterisation of novel cancer-specific biomarkers and/or targets is a key challenge for the successful development of diagnostic, prognostic and immunotherapeutic strategies aimed at tumour rejection (Feichtinger *et al.,* 2012a; Krishnadas, Bai and Lucas, 2013). These factors are of particular interest because cancer is typically diagnosed and treated in its late stages, when cancer cells have spread within the body and become metastatic (Suri, 2006; Aly, 2012).

Cancer-germline (CG) genes or CT antigen genes have been identified as an attractive group of genes which encode proteins that are restricted in their expression to the human testis and malignant tumours, and are not expressed in normal somatic cells (Old, 2001; Simpson *et al.,* 2005; Feichtinger *et al.,* 2012a; Whitehurst, 2014). Rousseaux and co-workers have shown an association between aberrant expression of a sub class of 26 testis-specific/placenta specific (TS/PS) genes in 293 lung cancer patients and aggressive, metastasis prone tumours with poor prognosis in all stages of disease (For more details, see Rousseaux *et al*., 2013b). A study by Whitehurst and colleagues (2007) further identified a number of CT genes capable of driving chemoresistance of lung cancer cells, suggesting that CT genes might be involved in chemotherapeutic resistance (see details in Whitehurst *et al*., 2007). A study by Janic *et al*. (2010) in *Drosophila melanogaster* showed that inactivation of some of cancer-germline (CG) genes reduced the tumour-suppressing capability, indicating that activation of these genes in cancer cells might be associated with tumour development (Janic *et al.,* 2010). Additionally, a recent study has shown that cancer cells maintain telomeres via induction of a set of meiosis-specific functions (Cho *et al*., 2014). For these reasons, we hypothesis that CT genes represent an excellent source for early cancer diagnostic and/or therapeutic strategies. Consequently, new candidate CT genes need to be identified to improve cancer immunotherapy.

In the present study, two approaches were taken to identify new CTA genes. The first approach was based on a manual search of the literature for meiosis-specific genes, which were considered to be a rich source of possible CTA genes. The second approach was based on a systematic approach that used a bioinformatics pipeline (developed by J. Feichtinger; Feichtinger *et al.*, 2012b; 2014b).

A total of 24 candidate genes were tested using RT-PCR analysis in a range of normal and cancer human tissues. The RT-PCR analysis of these genes identified 7 testis-restricted genes; 5 potential CT antigen genes; 2 CT/CNS-restricted genes; one CT-selective gene; and 2 CT/CNS-selective genes. Of these 24 candidate genes, 7 were identified to express in more than two normal human tissues and were excluded from further analysis in cancer tissues.

The seven testis-restricted genes were expressed only in normal human testis but were not detected in other normal and cancer tissues. (These genes, however, might be expressed in other cancer tissues not included in the present project). These seven genes were investigated in clinical samples derived from patients using a clinical meta-analysis microarray web tools (CancerMA, http://www.cancerma.org.uk/; Feichtinger *et al.*, 2012b). Five of the seven testis-restricted genes (*C4orf17*, *C16orf92*, *DDX4*, *IQCF3* and *NT5C1B*) were expressed in different types of clinical cancer samples, whereas *ARRDC5* and *TMEM202* were not present on this array.

Importantly, 10 of the 24 identified genes in the present study were identified as potential novel CT genes with restricted expression to the normal testis and a wide range of cancer cell lines/tissues. Five of the 10 novel CT genes were expressed in no more than two non-testis tissues. The aberrant expression of these genes in some normal tissues might reflect the "normal" RNA. A study by Chen *et al.*, (2005) showed different CT gene expression in similar normal human tissues but obtained from different sources (Chen *et al.*, 2005b; Feichtinger *et al.*, 2012a). The other explanation is that these genes might have a mitotic function in additional to their meiosis-specific function.

These 10 genes were also analysed in clinically normal *vs.* cancer samples using meta-analysis microarray tools. Four genes (*C9orf11*, *ODF4*, *PRDM7* and *PRSS5*) showed up-regulation in different types of clinical samples, which could mean that they are useful target

genes for cancer therapeutic strategies. All genes identified in the present project are located on autosomes, with exception of *MAGE-B5,* which is X encoded. This opens up a new possibility for their oncogenic activities and tumour distribution, as they are not subjected to meiotic X inactivation (Feichtinger et al., 2012a).

Several of the genes identified in the current project encode proteins with uncharacterised functions. Therefore, further investigations are required to determine the biological function of these genes in cancer cells (e.g., *C12orf12*). In contrast, other identified genes have known functions in meiosis or belong to families involved in oncogenic activity. For example, in the present work, RT-PCR analysis revealed that *TDRD12* was expressed only in the embryonic carcinoma cell line (NT2) as well as in the normal testis tissue. In mice, an involvement of TDRD12 was identified in Piwi-interacting RNAs (piRNAs) as an association between TDRD1 and Piwi protein MILI (PIWIL2)(Pandey *et al*., 2013). PIWIL2 was demonstrated in precancerous stem cells (pCSCs) and also as CT gene expressed in a wide range of human and mouse cancer cells. It might play a role as putative proto-oncogenes promoting cell proliferation, damage repair and differentiation of cancer cells (Ye *et al*., 2010; Shi *et al*., 2013). Taking all these facts together, TDRD12 showed potential as a target for early diagnostics and anticancer drug development in the treatment of (embryonal carcinoma) germ line tumours. However, further analysis is required to investigate its association with PIWIL2 in promoting tumorigenesis.

In the current study, the expression of potential CT genes were observed in breast cancer, colon cancer and melanoma tissues. However, the vast majority of these genes were identified to express in ovarian cancer either by RT-PCR analysis or meta-analysis. Ovarian cancers are often treated using cytoreductive surgery followed by chemotherapy or radiation treatment (reviewed in Mirandola *et al.,* 2011; Sammut *et al*., 2014; Tse *et al.*, 2014). High grade tumours have less probability of eradication by surgical treatment. These genes might be used as an early detection biomarker and also might serve as therapeutic targets.

In the present study, *DDX4* for example was identified as testis-restricted genes using RT-PCR analysis in normal and cancer tissues. Use of the clinical meta-analysis microarray dataset identified up-regulation of *DDX4* in patients with ovarian cancer. *DDX4* (known as *Vasa*) is

germ cell marker encoded protein that is restricted to the germline cells (Castrillon *et al.*, 2000). It is an RNA helicase involved in different cellular processes including alteration of RNA secondary structure (reviewed in Kim *et al.*, 2014). A study by Hashimoto and co-worker (2008) identified the over-expression of *DDX4* in 21 of 75 epithelial ovarian cancers. Overexpression of *DDX4* in ovarian cancer was found to disturb the G2 checkpoint and induce DNA damage through down-regulation of 14-3-3 σ expression (Hashimoto *et al.*, 2008). Recent work found that DDX4 also serves as a potential ovarian cancer stem cell marker. It co-localized with and had approximately the same expression patterns as the ovarian cancer stem marker CD133 protein (reviewed in Kim *et al.*, 2014).

Another study in *D. melanogaster* showed that an oncogenic soma-to-germline transition of germline genes that are expressed in brain tumours is required to promote tumourigenesis (Janic *et al.*, 2010). Furthermore, Feichtinger and co-workers (2014) investigated the human orthologues of these germline genes (e.g., *DDX4*) and identified extensive up-regulation of these genes in different types of human tumours, although the functional implications for oncogenesis of this up-regulation remains unclear (Feichtinger *et al.*, 2014 b).

7.2 Regulation of the expression of CTA genes in colorectal cancer cells using epigenetic inhibitors

Epigenetic mechanisms regulate gene expression and normal mammalian cell development (Nottke *et al.*, 2009; Meikar et al., 2013). CT genes are usually co-expressed in positive tumours, which suggests that their activation in cancers is not due to individual events but arises due to global gene activation mechanisms (De Smet *et al.*, 1999; De Smet and Loriot, 2013). One interesting theory suggested that activation of a gametogenic programme during tumourigenesis might be a result of switching on of master germline genes, particularly those that directly control genome demethylation. This, in turn, might lead to widespread activation of silenced CT genes in somatic tissues (Old, 2001; Simpson *et al.*, 2005).

Regulation of the CT genes is proposed to occur at the transcriptional level by epigenetic signals that affect repeated sequences as well as single copy genes, thereby potentially leading to alteration of gene expression in cancers (De Smet *et al.*, 1999; De Smet and Loriot, 2013). Understanding CT gene expression is critical as drug-induced augmentation of

expression has also been postulated to be a potential enhancer of immunotherapeutics, the rationale being that further up-regulation of a tumour-specific antigen will result in enhanced immunological targeting of the tumour (Almatrafi *et al.*, 2014). For example, the expression of *NY-ESO-1* was reactivated in different types of cancers, using the DNA demethylation inhibitor 5-AZA-CdR (reviewed in Karpf, 2006). *NY-ESO-1* gene product has been successfully targeted in an adoptive therapeutic approaches to melanoma therapy (for details, see Hunder et al., 2008).

In this study, we examined the influence of DNA demethylation and/or histone deacetylation on the expression of the 11 potential novel CT genes (meiCT genes) identified in Chapter 3.0 and five genes previously published by Feichtinger *et al.*, (2012a). Treatment of HCT116 and SW480 cells with epigenetic inhibitors appeared to induce the expression of 10 out of the 16 candidate CTA genes. These genes were previously classified, based on their RT-PCR results in Chapter 3.0, into different classes. The treatment led to activated expression of two of the six testis-restricted genes; five of the seven cancer/testis-restricted genes; and all three cancer/testis-CNS restricted genes. However, six genes showed no change in their expression, suggesting that regulation of these genes occurs by an as yet unknown mechanism. In addition, treatment of HCT116 and SW480 cells with 5-AZA-CdR, followed by culture in drug-free medium for 10 days, showed that the activation of *TDRD12* and *STRA8* continued to be observed after 3, 6 and 9 days of cell growth in drug free medium, whereas *C20orf201* and *DDX4* disappeared after 3 days of cell culture in drug free medium.

The sensitivity of CT genes to epigenetic modulator inhibitors, such as methylation drugs, probably involves the CpG-rich islands that frequently occur in their promoter regions (De Smet *et al.*, 1999; De Smet and Loriot, 2013). DNA methylation has little effect on genes with promoter regions containing low densities of CpG islands. These latter genes frequently show no consistent relationship between promoter methylation and transcriptional silencing (Boyes and Bird 1992; De Smet and Loriot, 2013).

The relationship between CpG islands and promoter gene methylation was identified by using the UCSC Genome Browser bioinformatics website (https://genome.ucsc.edu/; Gardiner-Garden and Frommer, 1987) to predict the CpG islands of the 16 CTAs in this study.

Six of 16 genes showed rich CpG islands, whereas 10 genes showed no CpG islands. A search for CpG islands upstream of the *TDRD12* and *STRA8* promoter regions predicted two CpG islands in the TSS of both *TDRD12* and *STRA8*. This observation supports the idea that DNA hypomethylation activates *TDRD12* and *STRA8* via their CpG-rich islands (Almatrafi *et al*., 2014).

The regulation of CTA expression does not simply rely on the presence or absence of methylation target CpG islands within promoter regions. Analysis of the TSS upstream of an inactivated CT gene after 5-AZA-CdR treatment revealed at least two genes, *ARRDC5* and *NT5C1B*, reported to have CpG islands in their transcriptional promoter regions [https://genome.ucsc.edu/]. Therefore, a very broad range of regulatory mechanisms likely exists for the control of CT gene regulation, which has implications for CT gene selection for clinical targeting strategies. The mechanistic regulatory pathways might also indicate the existence of co-regulated sub-groups of CT genes, which would have implications for the study of these genes as biomarkers, potential oncogenes and/or encoders of drug targets. Some CT genes are also required for tumour cell proliferation; therefore, turning off these genes could reduce the proliferation-mediated burden of tumours, thereby restricting their disease effect and/or enhancing other therapeutic approaches (Almatrafi *et al*., 2014).

7.3   Function analysis of the PRDM7 and PRDM9 genes

7.3.1 Comparing PRDM7 and PRDM9 expression

In humans and mice, PRDM9 has been identified as the major protein that binds to the degenerated 13-mer motifs hotspots through its zinc finger array (Baudat *et al*., 2010; Myers *et al*., 2010; Parvanov *et al*., 2010). In mice, disruption of *Prdm9*/*Meisetz* leads to hybrid sterility in both sexes resulting from a severe deficiency in the DSB repair pathway, deficient pairing of homologues and failure of sex body formation (Hayashi et al., 2005). Recently, Feichtinger and co-worker (2012) demonstrated that *PRDM9* was expressed in a wide range of cancer cell lines/tissues, but was not expressed in normal human tissues and classified as a meiCT gene encoded protein that could serve as a potential CT antigen biomarker/target (Feichtinger *et al.*, 2012a).

In the current study, *PRDM7*, the paralogous gene of *PRDM9*, was identified as a CT gene with restricted expression to the testis, melanoma and leukaemia. Gene alignment of *PRDM7/9* identified a high similarity in the DNA sequence, with the key difference being an 89-nucleotide duplicon that occurred only in *PRDM7*. RT-PCR analysis carried out on *PRDM9* and *PRDM7* using distinguishing primers showed that both genes were co-expressed in leukaemia K-562 cells, with no expression observed in non-testis normal human tissues.

A study by Berg and colleagues (2010) showed that variation within the PRDM9 zinc finger array strongly affected the human sperm hotspot activity. A sub-change in the zinc finger sequence lead to activation of new hotspots and creation of some nonactivated hotspots. The study suggested that PRDM9 might be a risk factor for creating genomic instability due to it roles in meiotic recombination (Berg *et al.*, 2010). Recent research found that a rare *PRDM9* allele was associated with the development of childhood B- cell precursor acute lymphoblastic leukaemia (B-ALL). The study found that parents carrying the rare *PRDM9* alleles transmitted them to half of their affected children, which indicated these alleles may act in meiosis (Hussin *et al*. 2013). Further genome-wide association studies showed that a few germline single nucleotide polymorphisms (SNPs) were associated with childhood ALL (reviewed in Woodward *et al*. 2014).

7.3.2 Is PRDM9 a transcriptional activator in cancer cells?

In addition to its function as a hotspot activator, PRDM9 was suggested to serve as a transcriptional regulation protein that regulates the expression of a group of meiosis prophase genes through its H3K4 trimethylation reaction. The evidence came from a study in mice showed that disruption of *Prdm9* $^{-/-}$ led to direct regulation of the expression of the testis-specific *RIK* gene (also term *Morc2b*) (Hayashi *et al.*, 2005). *Morc2b* has a GHKL ATPase domain in its N-terminal region, which is conserved among topoisomerase, histidine kinase, DNA mismatch repair and HSP90 proteins (Dutta and Inouye, 2000; Hayashi *et al.*, 2005). In this study, the result of RT-PCR analysis of the human *MORC* family genes revealed expression patterns for these genes that differed from those seen for *PRDM9* in the normal and cancer tissues. Therefore, PRDM9 may have no involvement in the direct regulation of these genes in human normal and cancers.

The full open reading frames of *PRDM7* and *PRDM9* were cloned and overexpressed in a HeLa Tet-on 3G system to identify their transcriptional activity. The transcriptional regulation of *PRDM9* was identified by analysing 33 genes using the cDNA of overexpressed *PRDM9* gene. No significant result was observed in the RT-PCR analysis of 32 genes, which included 4 genes of the *MORC* family, 13 known meiosis-specific genes, 5 genes of the PRDM family and the 11 CTA genes. The RT-PCR analysis used in this study was perhaps not sufficient to determine whether overexpression of PRDM9 can promote the expression of this group of genes. However, the expression of *PRDM7* was promoted with high band intensity in overexpressing HeLa Tet-on 3G cells expressing *PRDM9*. Distinguishing primers between *PRDM7* and *PRDM9* were used to detect the expression and showed similar results. Taking all these factors into account, *PRDM7* and *PRDM9* showed similar patterns of expression in normal testis, leukaemia and overexpressed PRDM9 HeLa cells using RT-PCR analysis and might also have the same similarity at the protein level.

One of the main features of PRDM7 and PRDM9 proteins among the other PRDM family proteins is the existence of a KRAB (Kruppel-association box) protein-protein binding domain and an SSX repression domain. The KRAB/SSXRD domain belongs to a synovial sarcoma X breakpoint protein family, which is normally associated with transcriptional repressors (Birtle and Ponting, 2006; Oliver *et al.*, 2009). The KRAB/SSXRD domain was identified in one the most well-known CT gene families, the SSX family. The expression of *SSX* family genes was observed in a wide range of cancer cells and associated with advanced stages of tumours and worse patient prognosis (reviewed in Smith and McNeel, 2010).

PRDM9 protein was detected in nearly all the cancer cell lines investigated in this study. PRDM9 was also shown to localise in the nucleus of three cancer cells lines and in the cytoplasm in NT2 cells. Unfortunately, the lack of a specific antibody to PRDM7 protein precluded any further evaluation of PRDM7 in this study. PRDM9 knockdown was shown in some cancer cell lines but it failed in others. However, the PRDM9 siRNA knockdown results need to be verified. This could be done using a new anti-PRDM9 antibody or better protein control such as Ponceau S protein stain. This stain total protein in the membrane and can verify the loading and transfer.

The aberrant expression of PRDM9 in cancer cells raises a question regarding its function in cancer and suggests that it might play a role as an oncogenic activator protein. Cancer cells and germline cells typically exhibit sharing of some common features including rapid proliferation, enabling replicative immortality and the ability to exist in an undifferentiated state (Wu and Ruvkun, 2010). A study by Janic *et al*. (2010) in *D. melanogaster* showed that mutation in the tumour suppressor gene *lethal (3) malignant brain tumor* [*1(3)mbt*] led to the formation of brain tumours in *Drosophila.* They found that one-fourth of the activated genes were associated with germline function. They suggested that activation of some of these germline genes led to tumour development (Janic *et al*., 2010). Subsequently, the human orthologues of these germline genes were identified as also being extensively activated in different types of human tumours (Feichtinger *et al*., 2014b). Both studies proposed a soma-to-germline transition in *Drosophila* and human cancers that drives oncogenic mitotic proliferation cells, resulting in activation of master germline genes in cancer cells.

7.3.3 Histone methyltransferase activity of human PRDM7 and PRDM9 proteins

In addition to the multiple zinc finger arrays and the KRAB (Kruppel-association box) protein-protein binding domain, the PRDM7 and PRDM9 proteins are characterized as having a PR/SET domain, which consists of histone lysine methyltransferase activity (reviewed in Fumasoni *et al*., 2007; Baudat *et al*., 2013). A study in mice demonstrated that *Prdm9* encodes a histone methyltransferase activity at the PR/SET domain for trimethylation of lysine 4 of Histone H3 (H3K4) and is associated with transcriptional activity modification (Hayashi *et al*., 2005). In the current study, the full open reading frames of *PRDM7* and *PRDM9* were successfully cloned and overexpressed into *E. coli*. Several attempts were made to purify the fusion PRDM7 and PRDM9 proteins but all attempts to produce soluble proteins failed and led to production of insoluble proteins that were difficult to purify. Generally, failure of *E. coli* to produce soluble proteins might be a result of improper protein folding that occurred during the protein production (Baneyx and Mujacic, 2004). *E. coli* has limited tolerance of proteins larger than >70 kDa as this leads to misfolding of the large recombinant

protein (Baneyx and Mujacic, 2004). This might explain why PRDM7 and PRDM9 were produced as inclusion bodies, due to their sizes [PRDM9 (≈128 kDa) and PRDM7 (≈81.7 kDa)]. However, more recently, human PRDM9 protein was demonstrated *in vitro* to have high activity of mono-, di-, and trimethylation of the H3K4 marker. In addition, PRDM9 was identified to have histone 3 lysine 36 (H3K36) activity as a second histone residue, which might also have mono-, di- and trimethylation of H3K36 activity. The study also identified that overexpression of PRDM9 in HEK293 cells led to significant increases in trimethylated H3K4 and H3K36 (Eram *et al.*, 2014). Furthermore, PRDM9 was found to methylate the other core histones, H2A, H2B and H4, although to a lesser degree than with H3. A single C321P mutation in the PR/SET domain led to weak activity of PRDM9 through inhibition of interaction with S-adenosylmethionine (SAM) (Wu *et al*., 2013, Xiaoying *et al*., 2014).

The trimethylated H3K36 is associated with transcription of active euchromatin and is involved in different cell processes such as alternative splicing and elongation, regulation of DNA replication and DNA mismatch repair (reviewed in Wagner and Carpenter, 2012). For example, the SET domain containing protein 2 (STED2) was found to have trimethylated H3K36 activity (Edmunds *et al.*, 2008). STED2 behaves like a tumour suppressor in breast cancer (Newbold and Mokbel, 2010) and renal cell carcinoma (Duns *et al*., 2010) and is expressed in the testis (Anand and Prasad 2012). This raises many questions about the role of PRDM9 in cancer cells. Is PRDM9 activate of the gametogenesis program in cancer cells? Does induction increase the trimethylation of H3K4 and H3K36, which control the transcriptional activity of many genes? Is the expression of PRDM7 and PRDM9 in different types of cancer a result of single/multiple mutation or could it be epigenetic modification? Further investigation should be carried out as these might be potential biomarkers/targets for immunotherapeutic strategies.

In summary, this study focused on the identification of novel CTA genes as potential target/biomarkers for early diagnostic, prognostic and immunotherapeutic strategies for cancer treatment. In total, 15 CT genes were identified as promising biomarkers/targets for therapeutic strategies. However, further study should be carried out in the future to identify their biological roles in cancer cells. In addition, this work identified a sub-class of CT genes

that show transcriptional silencing that is refractory to the hypomethylation drug 5-aza-2´-deoxycytidine. We also established an *in vivo* overexpression system for PRDM7 and PRDM9 using a HeLa Tet on 3G system, as well as an *in vitro* overexpression system for these genes in *E. coli* cells. Further work using these two systems is needed for *in vivo* or *in vitro* identification of the biological roles of PRDM7 and PRDM9.

7.4 Future directions

Due to time limitations, further studies regarding the present project are recommended for investigation in the future:

1. qRT-PCR, western blot analysis and immunohistochemistry should be used to identify the encoded proteins of CT genes identified in Chapter 3.0. These CT genes could be new potential targets/biomarkers for early diagnostic and therapeutic strategies.

2. Potential CT genes should also validated in tumours derived from the patients, as these will be more clinically applicable compared to cancer cell lines.

3. Methylation specific PCR (MSP) and/or Bisulphite-sequencing PCR (BSP) might be carried out on the candidate genes studied in Chapter 4.0 to investigate the methylation of CpG islands in their promoters.

4. Western blot analysis and immunohistochemistry should be carried out in normal and cancer human cells using specific antibodies that have no cross reactivity between PRDM7 and PRDM9. Anti 6x His antibody should also be used on HeLa Tet-on 3G with transgenic PRDM9 or PRDM7 (I)/ (II), as these clones contain a 6x His tag sequence at their C-terminal.

5. Investigations using qRT-PCR should be carried out to determine the effects of *PRDM7* overexpression on *PRDM9* gene expression. These qRT-PCR studies might also be carried out in overexpressed PRDM9 cells to determine if any effect occurs in the expression of other genes.

6. RNA sequencing should performed in overexpressed HeLa Tet-on 3G cells to identify the transcriptional activity of PRDM9, if it exists.

Chapter 8.0 References

Aarts, M., Linardopoulos, S. & Turner, N.C. 2013, "Tumour selective targeting of cell cycle kinases for cancer treatment", *Current Opinion in Pharmacology,* vol. 13, no. 4, pp. 529-35.

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.

Abelev, G.I., Perova, S., Khramkova, N., Postnikova, Z. & Irlin, I. 1963, "PRODUCTION OF EMBRYONAL [alpha]-GLOBULIN BY TRANSPLANTABLE MOUSE HEPATOMAS", *Transplantation,* vol. 1, no. 2, pp. 174-180.

Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M., Utterback, T.R., Nagle, J.W., Fields, C. & Venter, J.C. 1992, "Sequence identification of 2,375 human brain genes", *Nature,* vol. 355, no. 6361, pp. 632-634.

Aly, H.A. 2012, "Cancer therapy and vaccination", *Journal of immunological methods,* vol. 382, no. 1, pp. 1-23.

Anand, P., Kunnumakkara, A.B., Kunnumakara, A.B., Sundaram, C., Harikumar, K.B., Tharakan, S.T., Lai, O.S., Sung, B. & Aggarwal, B.B. 2008, "Cancer is a preventable disease that requires major lifestyle changes", *Pharmaceutical Research,* vol. 25, no. 9, pp. 2097-116.

Auton, A., Li, Y.R., Kidd, J., Oliveira, K., Nadel, J., Holloway, J.K., Hayward, J.J., Cohen, P.E., Greally, J.M. & Wang, J. 2013, "Genetic recombination is targeted towards gene promoter regions in dogs", *PLoS Genetics,* vol. 9, no. 12, pp. e1003984.

Axelsson, E., Webster, M.T., Ratnakumar, A., LUPA Consortium, Ponting, C.P. & Lindblad-Toh, K. 2012, "Death of PRDM9 coincides with stabilization of the recombination landscape in the dog genome", *Genome Research,* vol. 22, no. 1, pp. 51-63.

Bacac, M. & Stamenkovic, I. 2008, "Metastatic cancer cell", *Annual Review of Pathology,* vol. 3, pp. 221-247.

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.

Ballestar, E. 2011, "An introduction to epigenetics" in *Epigenetic Contributions in Autoimmune Disease* Springer, , pp. 1-11.

Baneyx, F. & Mujacic, M. 2004, "Recombinant protein folding and misfolding in *Escherichia coli*", *Nature Biotechnology,* vol. 22, no. 11, pp. 1399-1408.

Barrio, M.M., Abes, R., Colombo, M., Pizzurro, G., Boix, C., Roberti, M.P., Gélizé, E., Rodriguez-Zubieta, M., Mordoh, J. & Teillaud, J. 2012, "Human macrophages and dendritic cells can equally present MART-1 antigen to CD8(+) T cells after phagocytosis of gamma-irradiated melanoma cells", *PloS one,* vol. 7, no. 7, pp. e40311-e40311.

Baudat, F., Buard, J., Grey, C., Fledel-Alon, a., Ober, C., Przeworski, M., Coop, G. & de Massy, B. 2010, "PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice", *Science,* vol. 327, no. 5967, pp. 836-40.

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.

Bechmann, I. & Woodroofe, N. 2014, "Immune privilege of the brain", *Neuroinflammation and CNS Disorders,* , pp. 1-8.

Behrends, U., Schneider, I., Rössler, S., Frauenknecht, H., Golbeck, A., Lechner, B., Eigenstetter, G., Zobywalski, C., Müller-Weihrich, S. & Graubner, U. 2003, "Novel tumor antigens identified by autologous antibody screening of childhood medulloblastoma cDNA libraries", *International Journal of Cancer,* vol. 106, no. 2, pp. 244-251.

Bellani, M.A., Boateng, K.A., McLeod, D. & Camerini-Otero, R. 2010, "The expression profile of the major mouse SPO11 isoforms indicates that SPO11beta introduces double strand breaks and suggests that SPO11alpha has an additional role in prophase in both spermatocytes and oocytes", *Molecular and Cellular Biology,* vol. 30, no. 18, pp. 4391-403.

Berg, I.L., Neumann, R., Lam, K.G., Sarbajna, S., Odenthal-Hesse, L., May, C.A. & Jeffreys, A.J. 2010, "PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans", *Nature Genetics,* vol. 42, no. 10, pp. 859-863.

Bergerat, A., de Massy, B., Gadelle, D., Varoutas, P.C., Nicolas, A. & Forterre, P. 1997, "An atypical topoisomerase II from Archaea with implications for meiotic recombination", *Nature,* vol. 386, no. 6623, pp. 414-7.

Birtle, Z. & Ponting, C.P. 2006, "Meisetz and the birth of the KRAB motif", *Bioinformatics,* vol. 22, no. 23, pp. 2841-2845.

Bishop, D.K. & Zickler, D. 2004, "Early decision; meiotic crossover interference prior to stable strand exchange and synapsis", *Cell,* vol. 117, no. 1, pp. 9-15.

Boël, P., Wildmann, C., Sensi, M.L., Brasseur, R., Renauld, J., 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.

Borde, V., Robine, N., Lin, W., Bonfils, S., Géli, V. & Nicolas, A. 2009, "Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites", *The EMBO Journal,* vol. 28, no. 2, pp. 99-111.

Boyes, J. & Bird, A. 1992, "Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein", *The EMBO Journal,* vol. 11, no. 1, pp. 327-333.

Buard, J., Barthès, P., Grey, C. & de Massy, B. 2009, "Distinct histone modifications define initiation and repair of meiotic recombination in the mouse", *The EMBO Journal,* vol. 28, no. 17, pp. 2616-2624.

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

Cai, R.L. 1998, "Human CART1, a Paired-Class Homeodomain Protein, Activates Transcription through Palindromic Binding Sites", *Biochemical and Biophysical Research Communications,* vol. 250, no. 2, pp. 305-311.

Cameron, E.E., Bachman, K.E., Myöhänen, S., Herman, J.G. & Baylin, S.B. 1999, "Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer", *Nature Genetics,* vol. 21, no. 1, pp. 103-107.

Cappell, K.M., Sinnott, R., Taus, P., Maxfield, K., Scarbrough, M., Whitehurst, AW., 2012," Multiple cancer testis antigens function to support tumour cell mitotic fidelity". *Mol Cell Biol*, vol 32, pp. 4131-4140.

Carelle, N., Piotto, E., Bellanger, A., Germanaud, J., Thuillier, A. & Khayat, D. 2002, "Changing patient perceptions of the side effects of cancer chemotherapy", *Cancer,* vol. 95, no. 1, pp. 155-163.

Castrillon, D.H., Quade, B.J., Wang, T.Y., Quigley, C. & Crum, C.P. 2000, "The human VASA gene is specifically expressed in the germ cell lineage", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 97, no. 17, pp. 9585-90.

Cedar, H. & Bergman, Y. 2009, "Linking DNA methylation and histone modification: patterns and paradigms", *Nature Reviews Genetics,* vol. 10, no. 5, pp. 295-304.

Chaffer, C.L. & Weinberg, R. 2011, "A perspective on cancer cell metastasis", *Science,* vol. 331, no. 6024, pp. 1559-64.

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.

Chen YT, Scanlan MJ, Venditti CA, Chua R, Theiler G, Stevenson BJ, Iseli C, Gure AO, Vasicek T, Strausberg RL, Jongeneel CV, Old LJ, Simpson AJ. 2005b, "Identification of cancer/testis-antigen genes by massive parallel signature sequencing". *Proc Natl Acad Sci* USA, 102:7940–7945.

Chen, H., Tsai, S. & Leone, G. 2009, "Emerging roles of E2Fs in cancer: an exit from cell cycle control", *Nature Reviews, Cancer,* vol. 9, no. 11, pp. 785-97.

Chen, Y., Venditti, C.A., Theiler, G., Stevenson, B.J., Iseli, C., Gure, A.O., Jongeneel, C.V., Old, L.J. & Simpson, A. 2005a, "Identification of CT46/HORMAD1, an immunogenic cancer/testis antigen encoding a putative meiosis-related protein", *Cancer Immun,* vol. 5, no. 9.

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, Z.X. & Riggs, A.D. 2011, "DNA methylation and demethylation in mammals", *The Journal of Biological Chemistry,* vol. 286, no. 21, pp. 18347-18353.

Cho, N. W., Dilley, R. L., Lampson, M. A., & Greenberg, R. A. 2014,."Interchromosomal homology searches drive directional ALT telomere movement and synapsis". *Cell*, 159(1), 108-121.

Copeland, R.A. 2013, "Molecular pathways: protein methyltransferases in cancer", *Clinical cancer research : an official journal of the American Association for Cancer Research,* vol. 19, no. 23, pp. 6344-6350.

Coppedè, F. 2014, "Epigenetic biomarkers of colorectal cancer: Focus on DNA methylation", *Cancer Letters,* vol. 342, no. 2, pp. 238-247.

Costa, Y., Speed, R., Ollinger, R., Alsheimer, M., Semple, C.a., Gautier, P., Maratou, K., Novak, I., Höög, C., Benavente, R. & Cooke, H.J. 2005, "Two novel proteins recruited by synaptonemal complex protein 1 (SYCP1) are at the centre of meiosis", *Journal of Cell Science,* vol. 118, pp. 2755-62.

Croce, C.M. 2008, "Oncogenes and cancer", *The New England journal of medicine,* vol. 358, no. 5, pp. 502-11.

Cronwright, G., Le Blanc, K., Gotherstrom, C., Darcy, P., Ehnman, M. & Brodin, B. 2005, "Cancer/testis antigen expression in human mesenchymal stem cells: down-regulation of *SSX* impairs cell migration and matrix metalloproteinase 2 expression", *Cancer Research,* vol. 65, no. 6, pp. 2207-2215.

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

Davies, O.R., Maman, J.D. & Pellegrini, L. 2012, "Structural analysis of the human SYCE2-TEX12 complex provides molecular insights into synaptonemal complex assembly", *Open Biology,* vol. 2, no. 7, pp. 120099-120099.

Davuluri, RV., Grosse, I., & Zhang, MQ. 2001 "Computational identification of promoters and first exons in the human genome." *Nature Genetics* vol 29, no. 4, pp.412-417.

De Backer, O., Arden, K.C., Boretti, M., Vantomme, V., De Smet, C., Czekay, S., Viars, C.S., De Plaen, E., Brasseur, F., Chomez, P., Van den Eynde, B., Boon, T. & van der Bruggen, P. 1999, "Characterization of the GAGE genes that are expressed in various human cancers and in normal testis", *Cancer Research,* vol. 59, no. 13, pp. 3157-3165.

de Bruijn, D.R., dos Santos, N.R., Kater-Baats, E., Thijssen, J., van den Berk, L., Stap, J., Balemans, M., Schepens, M., Merkx, G. & Geurts van Kessel, A. 2002, "The cancer-related protein SSX2 interacts with the human homologue of a Ras-like GTPase interactor, RAB3IP, and a novel nuclear protein, SSX2IP", *Genes, Chromosomes and Cancer,* vol. 34, no. 3, pp. 285-298.

de Massy, B. 2013, "Initiation of meiotic recombination: how and where? Conservation and specificities among eukaryotes", *Annual Review of Genetics,* vol. 47, pp. 563-99.

De Smet, C. & Loriot, A. 2013, "DNA hypomethylation and activation of germline-specific genes in cancer" in *Epigenetic Alterations in Oncogenesis* Springer, vol 754, pp. 149-166.

De Smet, C., Courtois, S.J., Faraoni, I., Lurquin, C., Szikora, J., De Backer, O. & Boon, T. 1995, "Involvement of two Ets binding sites in the transcriptional activation of the MAGE1 gene", *Immunogenetics,* vol. 42, no. 4, pp. 282-290.

De Smet, C., Loriot, A. & Boon, T. 2004, "Promoter-dependent mechanism leading to selective hypomethylation within the 5' region of gene MAGE-A1 in tumor cells", *Molecular and Cellular Biology,* vol. 24, no. 11, pp. 4781-4790.

De Smet, C., Lurquin, C., Lethe, B., Martelange, V. & Boon, T. 1999, "DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter", *Molecular and Cellular Biology,* vol. 19, no. 11, pp. 7327-7335.

Dettman, E., Simko, S.J., Ayanga, B., Carofino, B., Margolin, J., Morse, H. & Justice, M.J. 2011, "Prdm14 initiates lymphoblastic leukemia after expanding a population of cells resembling common lymphoid progenitors", *Oncogene,* vol. 30, no. 25, pp. 2859-2873.

Di Zazzo, E., De Rosa, C., Abbondanza, C. & Moncharmont, B. 2013, "PRDM Proteins: Molecular mechanisms in signal transduction and transcriptional regulation", *Biology,* vol. 2, no. 1, pp. 107-141.

Dray, E., Dunlop, M.H., Kauppi, L., San Filippo, J., Wiese, C., Tsai, M.S., Begovic, S., Schild, D., Jasin, M., Keeney, S. & Sung, P. 2011, "Molecular basis for enhancement of the meiotic DMC1 recombinase by RAD51 associated protein 1 (RAD51AP1)", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 108, no. 9, pp. 3560-3565.

Duesberg, P. & Li, R. 2003, "Multistep carcinogenesis: a chain reaction of aneuploidizations", *Cell Cycle,* vol. 2, no. 3, pp. 201-209.

Dutta, R. & Inouye, M. 2000, "GHKL, an emergent ATPase/kinase superfamily", *Trends in biochemical sciences,* vol. 25, no. 1, pp. 24-28.

Edmunds, J.W., Mahadevan, L.C. & Clayton, A.L. 2008, "Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation", *The EMBO journal,* vol. 27, no. 2, pp. 406-420.

Ehrlich, M. & Lacey, M. 2013, "DNA hypomethylation and hemimethylation in cancer" in *Epigenetic Alterations in Oncogenesis. Springer* New York , 2013. pp. 31-56.

Eichler, A.F. & Plotkin, S.R. 2008, "Brain metastases", *Current treatment options in neurology,* vol. 10, no. 4, pp. 308-314.

Eom, G.H., Kim, K., Kim, S., Kee, H.J., Kim, J., Jin, H.M., Kim, J., Kim, J.H., Choe, N. & Kim, K. 2009, "Histone methyltransferase PRDM8 regulates mouse testis steroidogenesis", *Biochemical and biophysical research communications,* vol. 388, no. 1, pp. 131-136.

Eram, M.S., Bustos, S.P., Lima-Fernandes, E., Siarheyeva, A., Senisterra, G., Hajian, T., Chau, I., Duan, S., Wu, H., Dombrovski, L., Schapira, M., Arrowsmith, C.H. & Vedadi, M. 2014, "Trimethylation of Histone H3 Lysine 36 by Human Methyltransferase PRDM9 Protein", *The Journal of biological chemistry,* vol. 289, no. 17, pp. 12177-88.

Feichtinger, J., Aldeailej, I., Anderson, R., Almutairi, M., Almatrafi, A., Alsiwiehri, N., Griffiths, K., Stuart, N., Wakeman, J.A., Larcombe, L. & McFarlane, R.J. 2012a, "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., McFarlane, R.J. & Larcombe, L.D. 2012b, "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.

Feichtinger, J., Thallinger, G.G., McFarlane, R.J. & Larcombe, L.D. 2012c, *Microarray meta-analysis: From data to expression to biological relationships,* In: *Computational Medicine.* (Trajanowski, Z., ed.). pp. 59-77, *Springer*, Heidelberg.

Feichtinger, J., Larcombe, L. & McFarlane, R.J. 2014b, "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*, vol. 134, no. 10, pp. 2359-2365.

Feichtinger, J., McFarlane, R.J. & Larcombe, L.D. 2014a, "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.

Ferlay, J., Shin, H.R., Bray, F., Forman, D., Mathers, C. & Parkin, D.M. 2010, "Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008", *International journal of cancer.Journal international du cancer,* vol. 127, no. 12, pp. 2893-2917.

Fog, C.K., Galli, G.G. & Lund, A.H. 2012, "PRDM proteins: Important players in differentiation and disease", *Bioessays,* vol. 34, no. 1, pp. 50-60.

Fratta, E., Coral, S., Covre, A., Parisi, G., Colizzi, F., Danielli, R., Marie 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. 1340-6.

Fraune, J., Wiesner, M. & Benavente, R. 2014, "The synaptonemal complex of basal metazoan hydra: more similarities to vertebrate than invertebrate meiosis model organisms", *Journal of genetics and genomics,* vol. 41, no. 3, pp. 107-15.

Fumasoni, I., Meani, N., Rambaldi, D., Scafetta, G., Alcalay, M. & Ciccarelli, F.D. 2007, "Family expansion and gene rearrangements contributed to the functional specialization of PRDM genes in vertebrates", *BMC evolutionary biology,* vol. 7, no. 1, pp. 187.

Garcia, V., Phelps, S.E.L., Gray, S. & Neale, M.J. 2011, "Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1", *Nature,* vol. 479, no. 7372, pp. 241-4.

Gardiner-Garden, M. & Frommer, M. 1987, "CpG Islands in vertebrate genomes", *Journal of Molecular Biology,* vol. 196, no. 2, pp. 261-282.

Gartler, S.M. & Goldman, M.A. 2005, "X-chromosome Inactivation", *Encyclopedia of life*. New York: Wiley Interscience. pp. 1–6.

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. 2009, "Cancer-testis antigens: potential targets for cancer immunotherapy.", *Archives of Iranian Medicine (AIM),* vol. 12, no. 4.

Ghafouri-Fard, S. & Modarressi, M. 2012, "Expression of splice variants of cancer-testis genes ODF3 and ODF4 in the testis of a prostate cancer patient", *Genetics and molecular research,* vol. 11, no. 4, pp. 3642-3648.

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, pp. 256-261.

Glazer, C.A., Smith, I.M., Ochs, M.F., Begum, S., Westra, W., Chang, S.S., Sun, W., Bhan, S., Khan, Z. & Ahrendt, S. 2009, "Integrative discovery of epigenetically derepressed cancer testis antigens in NSCLC", *PLoS One,* vol. 4, no. 12, pp. e8189.

Gold, P. & Freedman, S.O. 1965, "Specific carcinoembryonic antigens of the human digestive system", *The Journal of experimental medicine,* vol. 122, no. 3, pp. 467-481.

Gossen, M., & Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proceedings of the *National Academy of Sciences*, 89(12), 5547-5551.

Govindan, R., Ding, L., Griffith, M., Subramanian, J., Dees, N.D., Kanchi, K.L., Maher, C.A., Fulton, R., Fulton, L., Wallis, J., Chen, K., Walker, J., McDonald, S., Bose, R., Ornitz, D., Xiong, D., You, M., Dooling, D.J., Watson, M., Mardis, E.R. & Wilson, R.K. 2012, "Genomic landscape of non-small cell lung cancer in smokers and never-smokers", *Cell,* vol. 150, no. 6, pp. 1121-34.

Gray, Steven G., Antonio H. Iglesias, Fernando Lizcano, Raul Villanueva, Sandra Camelo, Hisaka Jingu, Bin T. Teh et al. 2005 "Functional characterization of JMJD2A, a histone deacetylase- and retinoblastoma-binding protein." *Journal of Biological Chemistry* 280, no. 31, pp. 28507-28518.

Gregan, J., Polakova, S., Zhang, L., Tolić-Nørrelykke, I.,M. & Cimini, D. 2011, "Merotelic kinetochore attachment: causes and effects", *Trends in cell biology,* vol. 21, no. 6, pp. 374-81.

Gruber, S., Haering, C.H. & Nasmyth, K. 2003, "Chromosomal Cohesin Forms a Ring", *Cell,* vol. 112, no. 6, pp. 765-777.

Hamer, G., Gell, K., Kouznetsova, A., Novak, I., Benavente, R. & Höög, C. 2006, "Characterization of a novel meiosis-specific protein within the central element of the synaptonemal complex", *Journal of cell science,* vol. 119, pp. 4025-32.

Hamer, G., Wang, H., Bolcun-Filas, E., Cooke, H.J., Benavente, R. & Höög, C. 2008, "Progression of meiotic recombination requires structural maturation of the central element of the synaptonemal complex", *Journal of cell science,* vol. 121, no. 15, pp. 2445.

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

Hanahan, D. & Weinberg, R.a. 2011, "Hallmarks of cancer: the next generation", *Cell,* vol. 144, no. 5, pp. 646-74.

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. 124-36.

Harris, T.J. & Drake, C.G. 2013, "Primer on tumor immunology and cancer immunotherapy", *Journal for ImmunoTherapy of Cancer,* vol. 1, no. 1, pp. 12.

Hashimoto, H., Sudo, T., Mikami, Y., Otani, M., Takano, M., Tsuda, H., Itamochi, H., Katabuchi, H., Ito, M. & Nishimura, R. 2008, "Germ cell specific protein VASA is over-expressed in epithelial ovarian cancer and disrupts DNA damage-induced G2 checkpoint", *Gynecologic oncology,* vol. 111, no. 2, pp. 312-319.

Hassold, T. & Hunt, P. 2001, "To err (meiotically) is human: the genesis of human aneuploidy", *Nature reviews.Genetics,* vol. 2, no. 4, pp. 280-91.

Hassold, T., Hall, H. & Hunt, P. 2007, "The origin of human aneuploidy: where we have been, where we are going", *Human molecular genetics,* vol. 16 Spec No, pp. R203-8.

Hayashi, K., Yoshida, K. & Matsui, Y. 2005, "A histone H3 methyltransferase controls epigenetic events required for meiotic prophase", *Nature,* vol. 438, no. 7066, pp. 374-378.

Heaney, J.D., Anderson, E.L., Michelson, M.V., Zechel, J.L., Conrad, P.A., Page, D.C. & Nadeau, J.H. 2012, "Germ cell pluripotency, premature differentiation and susceptibility to testicular teratomas in mice", *Development (Cambridge, England),* vol. 139, no. 9, pp. 1577-1586.

Herman, J.G., Graff, J.R., Myohanen, S., Nelkin, B.D. & Baylin, S.B. 1996, "Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 93, no. 18, pp. 9821-9826.

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.

Hohenauer, T. & Moore, A.W. 2012, "The Prdm family: expanding roles in stem cells and development", *Development (Cambridge, England),* vol. 139, no. 13, pp. 2267-2282.

Holthausen, J.T., Wyman, C. & Kanaar, R. 2010, "Regulation of DNA strand exchange in homologous recombination", *DNA repair,* vol. 9, no. 12, pp. 1264-72.

Hong, J.A., Kang, Y., Abdullaev, Z., Flanagan, P.T., Pack, S.D., Fischette, M.R., Adnani, M.T., Loukinov, D.I., Vatolin, S., Risinger, J.I., Custer, M., Chen, G.A., Zhao, M., Nguyen, D.M., Barrett, J.C., Lobanenkov, V.V. & Schrump, D.S. 2005, "Reciprocal binding of CTCF and BORIS to the NY-ESO-1 promoter coincides with derepression of this cancer-testis gene in lung cancer cells", *Cancer research,* vol. 65, no. 17, pp. 7763-7774.

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", *New England Journal of Medicine,* vol. 358, no. 25, pp. 2698-2703.

Hussin, J., Sinnett, D., Casals, F., Idaghdour, Y., Bruat, V., Saillour, V., Healy, J., Grenier, J., de Malliard, T., Busche, S., Spinella, J., Larivière, M., Gibson, G., Andersson, A., Holmfeldt, L., Ma, J., Wei, L., Zhang, J., Andelfinger, G., Downing, J.R., Mullighan, C.G. & Awadalla, P. 2013, "Rare allelic forms of PRDM9 associated with childhood leukemogenesis", *Genome Research,* vol. 23, no. 3, pp. 419-30.

Ikehata, M., Ueda, K. & Iwakawa, S. 2012, "Different involvement of DNA methylation and histone deacetylation in the expression of solute-carrier transporters in 4 colon cancer cell lines", *Biological and Pharmaceutical Bulletin,* vol. 35, no. 3, pp. 301-307.

Irie, S., Tsujimura, A., Miyagawa, Y., Ueda, T., Matsuoka, Y., Matsui, Y., Okuyama, A., Nishimune, Y. & Tanaka, H. 2009, "Single-nucleotide polymorphisms of the PRDM9 (MEISETZ) gene in patients with nonobstructive azoospermia", *Journal of Andrology,* vol. 30, no. 4, pp. 426-431.

James, S., Link, P. & Karpf, A. 2006, "Epigenetic regulation of X-linked cancer/germline antigen genes by DNMT1 and DNMT3b", *Oncogene,* vol. 25, no. 52, pp. 6975-6985.

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

Jeffreys, A.J., Kauppi, L. & Neumann, R. 2001, "Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex", *Nature genetics,* vol. 29, no. 2, pp. 217-22.

Jemal, A., Bray, F., Center, M.M., Ferlay, J., Ward, E. & Forman, D. 2011, "Global cancer statistics", *CA: a Cancer Journal for Clinicians,* vol. 61, no. 2, pp. 69-90.

Kagawa, W. & Kurumizaka, H. 2010, "From meiosis to postmeiotic events: uncovering the molecular roles of the meiosis-specific recombinase Dmc1", *The FEBS Journal,* vol. 277, no. 3, pp. 590-8.

Kang, Y., Hong, J., Chen, G., Nguyen, D. & Schrump, D. 2007, "Dynamic transcriptional regulatory complexes including BORIS, CTCF and Sp1 modulate NY-ESO-1 expression in lung cancer cells", *Oncogene,* vol. 26, no. 30, pp. 4394-4403.

Karpf, A.R. 2006, "A potential role for epigenetic modulatory drugs in the enhancement of cancer/germ-line antigen vaccine efficacy", *Epigenetics,* vol. 1, no. 3, pp. 116-120.

Karpf, A.R., Lasek, A.W., Ririe, T.O., Hanks, A.N., Grossman, D. & Jones, D.A. 2004, "Limited gene activation in tumor and normal epithelial cells treated with the DNA methyltransferase inhibitor 5-aza-2'-deoxycytidine", *Molecular pharmacology,* vol. 65, no. 1, pp. 18-27.

Karube, K., Nakagawa, M., Tsuzuki, S., Takeuchi, I., Honma, K., Nakashima, Y., Shimizu, N., Ko, Y.H., Morishima, Y., Ohshima, K., Nakamura, S. & Seto, M. 2011, "Identification of FOXO3 and PRDM1 as tumor-suppressor gene candidates in NK-cell neoplasms by genomic and functional analyses", *Blood,* vol. 118, no. 12, pp. 3195-3204.

Kauppi, L., Barchi, M., Baudat, F., Romanienko, P.J., Keeney, S. & Jasin, M. 2011, "Distinct properties of the XY pseudoautosomal region crucial for male meiosis", *Science (New York, N.Y.),* vol. 331, no. 6019, pp. 916-20.

Kaur, S., Momi, N., Chakraborty, S., Wagner, D.G., Horn, A.J., Lele, S.M., Theodorescu, D. & Batra, S.K. 2014, "Altered expression of transmembrane mucins, MUC1 and MUC4, in bladder cancer: pathological implications in diagnosis", *PloS one,* vol. 9, no. 3, pp. e92742-e92742.

Kee, Y., Yang, K., Cohn, M.A., Haas, W., Gygi, S.P. & D'Andrea, A.D. 2010, "WDR20 regulates activity of the USP12 x UAF1 deubiquitinating enzyme complex", *Journal of biological chemistry,* vol. 285, no. 15, pp. 11252-11257.

Keeney, S., Giroux, C.N. & Kleckner, N. 1997, "Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family", *Cell,* vol. 88, no. 3, pp. 375-384.

Kim, B., Lee, H.J., Choi, H.Y., Shin, Y., Nam, S., Seo, G., Son, D.S., Jo, J., Kim, J., Lee, J., Kim, J., Kim, K. & Lee, S. 2007, "Clinical validity of the lung cancer biomarkers identified by bioinformatics analysis of public expression data", *Cancer research,* vol. 67, no. 15, pp. 7431-7438.

Kim, K.C., Geng, L. & Huang, S. 2003, "Inactivation of a histone methyltransferase by mutations in human cancers", *Cancer research,* vol. 63, no. 22, pp. 7619-7623.

Kim, K.H., Kang, Y., Jo, J., Ock, M.S., Moon, S.H., Suh, D.S., Yoon, M.S., Park, E., Jeong, N. & Eo, W. 2014, "DDX4 (DEAD box polypeptide 4) colocalizes with cancer stem cell marker CD133 in ovarian cancers", *Biochemical and biophysical research communications,* vol. 447, no. 2, pp. 315-322.

Kim, K.P., Weiner, B.M., Zhang, L., Jordan, A., Dekker, J. & Kleckner, N. 2010, "Sister cohesion and structural axis components mediate homolog bias of meiotic recombination", *Cell,* vol. 143, no. 6, pp. 924-37.

Kitamura, Kouichi, Yasushi Miyagawa, Naoko Iguchi, Hiromi Nishimura, Hiromitsu Tanaka, and Yoshitake Nishimune. "Molecular cloning and characterization of the human orthologue of the oppo 1 gene encoding a sperm tail protein." *Molecular human reproductio*n 9, no. 5 (2003): 237-243.

Klein, F., Mahr, P., Galova, M., Buonomo, S.B., Michaelis, C., Nairz, K. & Nasmyth, K. 1999, "A central role for cohesins in sister chromatid cohesion, formation of axial elements, and recombination during yeast meiosis", *Cell,* vol. 98, no. 1, pp. 91-103.

Koos, B., Bender, S., Witt, H., Mertsch, S., Felsberg, J., Beschorner, R., Korshunov, A., Riesmeier, B., Pfister, S., Paulus, W. & Hasselblatt, M. 2011, "The transcription factor evi-1 is overexpressed, promotes proliferation, and is prognostically unfavorable in infratentorial ependymomas", *Clinical cancer research : an official journal of the American Association for Cancer Research,* vol. 17, no. 11, pp. 3631-3637.

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.

Kozłowska, A., Mackiewicz, J. & Mackiewicz, A. 2013, "Therapeutic gene modified cell based cancer vaccines", *Gene,* vol. 525, no. 2, pp. 200-207.

Krishnadas, D.K., Bai, F. & Lucas, K.G. 2013, "Cancer testis antigen and immunotherapy.", *ImmunoTargets & Therapy,* vol. 2.

Kronja, I. & Orr-Weaver, T. 2011, "Translational regulation of the cell cycle: when, where, how and why?", *Philosophical transactions of the Royal Society of London.Series B, Biological sciences,* vol. 366, no. 1584, pp. 3638-52.

Kuo, P., Chiang, H., Wang, Y., Kuo, Y., Chen, M., Yu, I., Teng, Y., Lin, S. & Lin, Y. 2013, "SEPT12-Microtubule Complexes Are Required for Sperm Head and Tail Formation", *International journal of molecular sciences,* vol. 14, no. 11, pp. 22102-22116.

Kuo, Y., Lin, Y., Chen, H., Wang, Y., Chiou, Y., Lin, H., Pan, H., Wu, C., Su, S. & Hsu, C. 2012, "SEPT12 mutations cause male infertility with defective sperm annulus", *Human mutation,* vol. 33, no. 4, pp. 710-719.

Kurashige, T., Noguchi, Y., Saika, T., Ono, T., Nagata, Y., Jungbluth, A., Ritter, G., Chen, Y.T., Stockert, E., Tsushima, T., Kumon, H., Old, L.J. & Nakayama, E. 2001, "Ny-ESO-1 expression and immunogenicity associated with transitional cell carcinoma: correlation with tumor grade", *Cancer research,* vol. 61, no. 12, pp. 4671-4674.

Lammers, J.H., Offenberg, H.H., van Aalderen, M., Vink, A.C., Dietrich, A.J. & Heyting, C. 1994, "The gene encoding a major component of the lateral elements of synaptonemal complexes of the rat is related to X-linked lymphocyte-regulated genes", *Molecular and cellular biology,* vol. 14, no. 2, pp. 1137-46.

Lee, S., Han, X., Choi, K.J., Ding, Y., Choi, T., Tak, E., Lee, J., Ha, J., Kim, S.S. & Lee, J. 2008, "A new method for purification of functional recombinant GST-cyclophilin A protein from E. coli", *Indian journal of biochemistry & biophysics,* vol. 45, no. 6, pp. 374-8.

Li, Y., Hu, X., Zhang, K., Guo, J., Hu, Z., Tao, S., Xiao, L., Wang, Q., Han, C. & Liu, Y. 2006, "Afaf, a novel vesicle membrane protein, is related to acrosome formation in murine testis", *FEBS letters*, vol. 580, no. 17, pp. 4266-4273.

Li, Y., Sosnik, J., Brassard, L., Reese, M., Spiridonov, N.A., Bates, T.C., Johnson, G.R., Anguita, J., Visconti, P.E. & Salicioni, A.M. 2011, "Expression and localization of five members of the testis-specific serine kinase (Tssk) family in mouse and human sperm and testis", *Molecular human reproduction*, vol. 17, no. 1, pp. 42-56.

Lichten, M. & de Massy, B. 2011, "The impressionistic landscape of meiotic recombination", *Cell,* vol. 147, no. 2, pp. 267-70.

Lim, J.H., Kim, S., Gabrielson, E., Park, Y.B., Park, J. & Kwon, T.K. 2005, "Activation of human cancer/testis antigen gene, XAGE-1, in tumor cells is correlated with CpG island hypomethylation", *International journal of cancer,* vol. 116, no. 2, pp. 200-206.

Lin, W., Jin, H., Liu, X., Hampton, K. & Yu, H. 2011, "Scc2 regulates gene expression by recruiting cohesin to the chromosome as a transcriptional activator during yeast meiosis", *Molecular biology of the cell,* vol. 22, no. 12, pp. 1985-96.

Lindsey, S.F., Byrnes, D.M., Eller, M.S., Rosa, A.M., Dabas, N., Escandon, J. & Grichnik, J.M. 2013, "Potential role of meiosis proteins in melanoma chromosomal instability", *Journal of skin cancer,* vol. 2013, pp. 190109-190109.

Link, P.A., Gangisetty, O., James, S.R., Woloszynska-Read, A., Tachibana, M., Shinkai, Y. & Karpf, A.R. 2009, "Distinct roles for histone methyltransferases G9a and GLP in cancer germ-line antigen gene regulation in human cancer cells and murine embryonic stem cells", *Molecular cancer research : MCR,* vol. 7, no. 6, pp. 851-862.

Linley AJ, Mathieu MG, Miles AK, Rees RC, McArdle SE, Regad T: The helicase HAGE expressed by malignant melanoma-initiating cells is required for tumour cell proliferation in vivo. *J Biol Chem* 2012, 287:13633-13643.

Liu, J.G., Yuan, L., Brundell, E., Bjorkroth, B., Daneholt, B. & Hoog, C. 1996, "Localization of the N-terminus of SCP1 to the central element of the synaptonemal complex and evidence for direct interactions between the N-termini of SCP1 molecules organized head-to-head", *Experimental cell research,* vol. 226, no. 1, pp. 11-9.

Loew, R., Heinz, N., Hampf, M., Bujard, H., & Gossen, M. (2010). Improved Tet-responsive promoters with minimized background expression. *BMC biotechnology*, 10(1), 81.

Lok, B.H. & Powell, S.N. 2012, "Molecular pathways: understanding the role of Rad52 in homologous recombination for therapeutic advancement", *Clinical cancer research : an official journal of the American Association for Cancer Research,* vol. 18, no. 23, pp. 6400-6.

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.

Longhese, M.P., Bonetti, D., Manfrini, N. & Clerici, M. 2010, "Mechanisms and regulation of DNA end resection", *The EMBO journal,* vol. 29, no. 17, pp. 2864-74.

Lord, C.J. & Ashworth, A. 2012, "The DNA damage response and cancer therapy", *Nature,* vol. 481, no. 7381, pp. 287-94.

Loukinov, D.I., Pugacheva, E., Vatolin, S., Pack, S.D., Moon, H., Chernukhin, I., Mannan, P., Larsson, E., Kanduri, C., Vostrov, A.A., Cui, H., Niemitz, E.L., Rasko, J.E., Docquier, F.M., Kistler, M., Breen, J.J., Zhuang, Z., Quitschke, W.W., Renkawitz, R., Klenova, E.M., Feinberg, A.P., Ohlsson, R., Morse, H.C.,3rd & Lobanenkov, V.V. 2002, "BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 99, no. 10, pp. 6806-6811.

Lucas, S., De Plaen, E. & Boon, T. 2000, "MAGE-B5, MAGE-B6, MAGE-C2, and MAGE-C3: four new members of the MAGE family with tumor-specific expression", *International journal of cancer,* vol. 87, no. 1, pp. 55-60.

Mandelbaum, J., Bhagat, G., Tang, H., Mo, T., Brahmachary, M., Shen, Q., Chadburn, A., Rajewsky, K., Tarakhovsky, A. & Pasqualucci, L. 2010, "*BLIMP1* Is a Tumor Suppressor Gene Frequently Disrupted in Activated B Cell-like Diffuse Large B Cell Lymphoma", *Cancer cell,* vol. 18, no. 6, pp. 568-579.

Marchal, R., Chicheportiche, A., Dutrillaux, B. & Bernardino-Sgherri, J. 2004, "DNA methylation in mouse gametogenesis", *Cytogenetic and genome research,* vol. 105, no. 2-4, pp. 316-324.

Marston, A.L. & Amon, A. 2004, "Meiosis: cell-cycle controls shuffle and deal", *Nature Reviews Molecular Cell Biology,* vol. 5, no. 12, pp. 983-997.

Mazin, A.V., Mazina, O.M., Bugreev, D.V. & Rossi, M.J. 2010, "Rad54, the motor of homologous recombination", *DNA repair,* vol. 9, no. 3, pp. 286-302.

McFarlane, R.J., Feichtinger, J. & Larcombe, L. 2014, "Cancer germline gene activation: Friend or foe?", Cell Cycle, vol. 13, no. 14, pp. 0--1.

Mehta, G.D., Rizvi, S.M.A. & Ghosh, S.K. 2012, "Cohesin: a guardian of genome integrity", Biochimica et biophysica acta, vol. 1823, no. 8, pp. 1324-42.

Meikar, O., Da Ros, M. & Kotaja, N. 2013, "Epigenetic regulation of male germ cell differentiation" in *Epigenetics: Development and Disease.* Springer Netherlands, 2013. pp. 119-138.

Mellman, I., Coukos, G. & Dranoff, G. 2011, "Cancer immunotherapy comes of age", *Nature*, vol. 480, no. 7378, pp. 480-9.

Menendez, L., Walker, D., Matyunina, L.V., Dickerson, E.B., Bowen, N.J., Polavarapu, N., Benigno, B.B. & McDonald, J.F. 2007, "Identification of candidate methylation-responsive genes in ovarian cancer", *Molecular cancer,* vol. 6, no. 10.

Metzler-Guillemain, C. & de Massy, B. 2000, "Identification and characterization of an SPO11 homolog in the mouse", *Chromosoma,* vol. 109, no. 1-2, pp. 133-8.

Michor, F., Iwasa, Y., Vogelstein, B., Lengauer, C. & Nowak, M.a. 2005, "Can chromosomal instability initiate tumorigenesis?", *Seminars in cancer biology,* vol. 15, no. 1, pp. 43-9.

Miller, M.P., Amon, A. & Ünal, E. 2013, "Meiosis I: when chromosomes undergo extreme makeover ", *Current opinion in cell biology,* vol. 25, no. 6, pp. 687-96.

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.

Miyamoto, T., Koh, E., Sakugawa, N., Sato, H., Hayashi, H., Namiki, M. & Sengoku, K. 2008, "Two single nucleotide polymorphisms in PRDM9 (MEISETZ) gene may be a genetic risk factor for Japanese patients with azoospermia by meiotic arrest", *Journal of assisted reproduction and genetics,* vol. 25, no. 11-12, pp. 553-557.

Modarressi, M.H. & GhafouriFard, S. 2011, "Potential of cancer-testis antigens as targets for cancer immunotherapy", *Bridging cell biology and genetics to the cancer clinic.Kerala, India,* , pp. 83-98.

Modarressi, M.H., Behnam, B., Cheng, M., Taylor, K.E., Wolfe, J. & van der Hoorn, F.A. 2004, "Tsga10 encodes a 65-kilodalton protein that is processed to the 27-kilodalton fibrous sheath protein", *Biology of reproduction,* vol. 70, no. 3, pp. 608-615.

Monte, M., Simonatto, M., Peche, L.Y., Bublik, D.R., Gobessi, S., Pierotti, M.A., Rodolfo, M. & Schneider, C. 2006, "MAGE-A tumor antigens target p53 transactivation function through histone deacetylase recruitment and confer resistance to chemotherapeutic agents", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 103, no. 30, pp. 11160-11165.

Morishita, K. 2007, "Leukemogenesis of the EVI1/MEL1 gene family", *International journal of hematology,* vol. 85, no. 4, pp. 279-286.

Mossman, D., & Scott, R.J. 2011, " Long term transcriptional reactivation of epigenetically silenced genes in colorectal cancer cells requires DNA hypomethylation and histone acetylation", *PLoS One*, vol. 6, pp. e23127.

Mossman, D., Kim, K. & Scott, R.J. 2010, "Demethylation by 5-aza-2'-deoxycytidine in colorectal cancer cells targets genomic DNA whilst promoter CpG island methylation persists", *BMC cancer,* vol. 10, pp. 366-366.

Muller, P.A.J., Vousden, K.H. & Norman, J.C. 2011, "p53 and its mutants in tumor cell migration and invasion", *The Journal of cell biology,* vol. 192, no. 2, pp. 209-18.

Myers, S., Bowden, R., Tumian, A., Bontrop, R.E., Freeman, C., MacFie, T.S., McVean, G. & Donnelly, P. 2010, "Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination", *Science (New York, N.Y.),* vol. 327, no. 5967, pp. 876-9.

Nasmyth, K. 2011, "Cohesin: a catenase with separate entry and exit gates?", *Nature cell biology,* vol. 13, no. 10, pp. 1170-7.

Naumov, G.N., Folkman, J. & Straume, O. 2009, "Tumor dormancy due to failure of angiogenesis: role of the microenvironment", *Clinical & experimental metastasis,* vol. 26, no. 1, pp. 51-60.

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. 220-8.

Nguyen, P., Bar-Sela, G., Sun, L., Bisht, K.S., Cui, H., Kohn, E., Feinberg, A.P. & Gius, D. 2008, "BAT3 and SET1A form a complex with CTCFL/BORIS to modulate H3K4 histone dimethylation and gene expression", *Molecular and cellular biology,* vol. 28, no. 21, pp. 6720-6729.

Ningaraj, N.S., Salimath, B.P., Sankpal, U.T., Perera, R. & Vats, T. 2007, "Targeted brain tumor treatment-current perspectives", *Drug target insights,* vol. 2, pp. 197-207.

Nishikawa, N., Toyota, M., Suzuki, H., Honma, T., Fujikane, T., Ohmura, T., Nishidate, T., Ohe-Toyota, M., Maruyama, R., Sonoda, T., Sasaki, Y., Urano, T., Imai, K., Hirata, K. & Tokino, T. 2007, "Gene amplification and overexpression of PRDM14 in breast cancers", *Cancer research,* vol. 67, no. 20, pp. 9649-9657.

Nottke, A., Colaiacovo, M.P. & Shi, Y. 2009, "Developmental roles of the histone lysine demethylases", *Development (Cambridge, England),* vol. 136, no. 6, pp. 879-889.

O'Donovan, P.,J. & Livingston, D.M. 2010, "BRCA1 and BRCA2: breast/ovarian cancer susceptibility gene products and participants in DNA double-strand break repair", *Carcinogenesis,* vol. 31, no. 6, pp. 961-7.

Offenberg, H.H., Schalk, J.a., Meuwissen, R.L., van Aalderen, M., Kester, H.a., Dietrich, a.J. & Heyting, C. 1998, "SCP2: a major protein component of the axial elements of synaptonemal complexes of the rat", *Nucleic acids research,* vol. 26, no. 11, pp. 2572-9.

Ohlsson, R., Renkawitz, R. & Lobanenkov, V. 2001, "CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease", *TRENDS in Genetics,* vol. 17, no. 9, pp. 520-527.

Old, L.J. 2001, "Cancer/testis (CT) antigens-a new link between gametogenesis and cancer", *Cancer Immun,* vol. 1, no. 1, pp. 1-7.

Oliver, P.L., Goodstadt, L., Bayes, J.J., Birtle, Z., Roach, K.C., Phadnis, N., Beatson, S.A., Lunter, G., Malik, H.S. & Ponting, C.P. 2009, "Accelerated evolution of the Prdm9 speciation gene across diverse metazoan taxa", *PLoS genetics,* vol. 5, no. 12, pp. e1000753.

Page, S.L. & Hawley, R.S. 2004, "The genetics and molecular biology of the synaptonemal complex", *Annual Review of Cell and Developmental Biology,* vol. 20, pp. 525-58.

Pandey, R.R., Tokuzawa, Y., Yang, Z., Hayashi, E., Ichisaka, T., Kajita, S., Asano, Y., Kunieda, T., Sachidanandam, R., Chuma, S., Yamanaka, S. & Pillai, R.S. 2013, "Tudor domain containing 12 (TDRD12) is essential for secondary PIWI interacting RNA biogenesis in mice", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 110, no. 41, pp. 16492-16497.

Pardoll, D. 2003, "Does the immune system see tumors as foreign or self?", *Annual Review of Immunology,* vol. 21, no. 1, pp. 807-839.

Park, D., Kim, S., Nam, M., Kim, G., Kim, J. & Rhim, H. 2011, "Improved recovery of active GST-fusion proteins from insoluble aggregates: solubilization and purification conditions using PKM2 and HtrA2 as model proteins", *BMB reports,* vol. 44, no. 4, pp. 279-84.

Parvanov, E.D., Petkov, P.M. & Paigen, K. 2010, "Prdm9 controls activation of mammalian recombination hotspots", *Science (New York, N.Y.),* vol. 327, no. 5967, pp. 835-835.

Petronczki, M., Siomos, M.F. & Nasmyth, K. 2003, "Un menage a quatre: the molecular biology of chromosome segregation in meiosis", *Cell,* vol. 112, no. 4, pp. 423-440.

Pinheiro, I., Margueron, R., Shukeir, N., Eisold, M., Fritzsch, C., Richter, F.M., Mittler, G., Genoud, C., Goyama, S. & Kurokawa, M. 2012, "Prdm3 and Prdm16 are H3K9me1 methyltransferases required for mammalian heterochromatin integrity", *Cell,* vol. 150, no. 5, pp. 948-960.

Pylayeva-Gupta, Y., Grabocka, E. & Bar-Sagi, D. 2011, "RAS oncogenes: weaving a tumorigenic web", *Nature reviews.Cancer,* vol. 11, no. 11, pp. 761-74.

Ramírez, LM., Spouge, JL., Kanga, GC., & Landsman, D. 2004 "Statistical analysis of over-represented words in human promoter sequences." *Nucleic Acids Research* vol. 32, no. 3, pp. 949-958.

Ranzani, M., Annunziato, S., Adams, D.J. & Montini, E. 2013, "Cancer gene discovery: exploiting insertional mutagenesis", *Molecular cancer research : MCR,* vol. 11, no. 10, pp. 1141-58.

Reid, A. & Nacheva, E. 2003, "A potential role for PRDM12 in the pathogenesis of chronic myeloid leukaemia with derivative chromosome 9 deletion", *Leukemia,* vol. 18, no. 1, pp. 178-180.

Remeseiro, S., Cuadrado, A., Carretero, M., Martínez, P., Drosopoulos, W.C., Cañamero, M., Schildkraut, C.L., Blasco, M.a. & Losada, A. 2012, "Cohesin-SA1 deficiency drives aneuploidy and tumourigenesis in mice due to impaired replication of telomeres", *The EMBO journal,* vol. 31, no. 9, pp. 2076-89.

Restifo, N.P., Dudley, M.E. & Rosenberg, S.A. 2012, "Adoptive immunotherapy for cancer: harnessing the T cell response", *Nature Reviews Immunology,* vol. 12, no. 4, pp. 269-281.

Rivera, E. & Gomez, H. 2010, "Chemotherapy resistance in metastatic breast cancer: the evolving role of ixabepilone", *Breast Cancer Res,* vol. 12, no. Suppl 2, pp. S2.

Rogozin, I.B., Basu, M.K., Jordan, I.K., Pavlov, Y.I. & Koonin, E.V. 2005, "Report APOBEC4, a New Member of the AID/APOBEC Family of Polynucleotide (Deoxy) cytidine Deaminases Predicted by Computational Analysis", *Cell Cycle,* vol. 4, no. 9, pp. 1281-1285.

Romanienko, P.J. & Camerini-Otero, R. 2000, "The mouse Spo11 gene is required for meiotic chromosome synapsis", *Molecular cell,* vol. 6, no. 5, pp. 975-87.

Rousseaux S, Debernardi A, Jacquiau B, Vitte AL, Vesin A, Nagy-Mignotte H, Moro-Sibilot D, Brichon PY, Lantuejoul S, Hainaut P, Laffaire J, de Reynies A, Beer DG, Timsit JF, Brambilla C, Brambilla E, Khochbin S 2013a, "Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers", *Sci Trans Med*, 5:pp. 186ra66.

Rousseaux S, Wang J, Khochbin S 2013b." Cancer hallmarks sustained by ectopic activations of placenta/male germline genes". *Cell Cycle*, 12:2331-2332.

Russo, G., Zegar, C. & Giordano, A. 2003, "Advantages and limitations of microarray technology in human cancer", *Oncogene,* vol. 22, no. 42, pp. 6497-6507.

Sahin, U., Türeci, Ö. & Pfreundschuh, M. 1997, "Serological identification of human tumor antigens", *Current opinion in immunology,* vol. 9, no. 5, pp. 709-716.

Sahin, U., Türeci, Ö., Chen, Y., Seitz, G., Villena-Heinsen, C., Old, L.J. & Pfreundschuh, M. 1998, "Expression of multiple cancer/testis (CT) antigens in breast cancer and melanoma: basis for polyvalent CT vaccine strategies", *International journal of cancer,* vol. 78, no. 3, pp. 387-389.

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.

Scanlan, M.J., Altorki, N.K., Gure, A.O., Williamson, B., Jungbluth, A., Chen, Y. & Old, L.J. 2000, "Expression of cancer-testis antigens in lung cancer: definition of bromodomain testis-specific gene (BRDT) as a new CT gene, CT9", *Cancer letters,* vol. 150, no. 2, pp. 155-164.

Schramm, S., Fraune, J., Naumann, R., Hernandez-Hernandez, A., Höög, C., Cooke, H.J., Alsheimer, M. & Benavente, R. 2011a, "A novel mouse synaptonemal complex protein is

essential for loading of central element proteins, recombination, and fertility", *PLoS genetics,* vol. 7, no. 5, pp. e1002088-e1002088.

Schwacha, A. & Kleckner, N. 1995, "Identification of double Holliday junctions as intermediates in meiotic recombination", *Cell,* vol. 83, no. 5, pp. 783-791.

Segurel, L. 2013, "The complex binding of PRDM9", *Genome biology,* vol. 14, no. 1, pp. 112.

Sharma, S., Kelly, T.K. & Jones, P.A. 2010, "Epigenetics in cancer", *Carcinogenesis,* vol. 31, no. 1, pp. 27-36.

Shen, L., Kondo, Y., Guo, Y., Zhang, J., Zhang, L., Ahmed, S., Shu, J., Chen, X., Waterland, R.A. & Issa, J.J. 2007, "Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters", *PLoS genetics,* vol. 3, no. 10, pp. e181.

Shi, R., Liu, J., Zou, Z., Qi, Y., Zhai, M., Zhai, W. & Gao, Y. 2013, "The immunogenicity of a novel cytotoxic T lymphocyte epitope from tumor antigen PL2L60 could be enhanced by 4-chlorophenylalanine substitution at position 1", *Cancer Immunology, Immunotherapy,* vol. 62, no. 11, pp. 1723-1732.

Shibata, A., Moiani, D., Arvai, A.S., Perry, J., Harding, S.M., Genois, M., Maity, R., van Rossum-Fikkert, S., Kertokalio, A., Romoli, F., Ismail, A., Ismalaj, E., Petricci, E., Neale, M.J., Bristow, R.G., Masson, J., Wyman, C., Jeggo, P.A. & Tainer, J.A. 2014, "DNA double-strand break repair pathway choice is directed by distinct MRE11 nuclease activities", *Molecular cell,* vol. 53, no. 1, pp. 7-18.

Shin, Y., 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-e1001190.

Shu, X., Geng, H., Li, L., Ying, J., Ma, C., Wang, Y., Poon, F.F., Wang, X., Ying, Y. & Yeo, W. 2011, "The epigenetic modifier PRDM5 functions as a tumor suppressor through modulating WNT/β-catenin signaling and is frequently silenced in multiple tumors", *PLoS one,* vol. 6, no. 11, pp. e27346.

Siegel, R., Naishadham, D. & Jemal, A. 2013, "Cancer statistics, 2013", *CA: a cancer journal for clinicians,* vol. 63, no. 1, pp. 11-30.

Sigalotti, L., Covre, A., Zabierowski, S., Himes, B., Colizzi, F., Natali, P.G., Herlyn, M. & Maio, M. 2008, "Cancer testis antigens in human melanoma stem cells: expression, distribution, and methylation status", *Journal of cellular physiology,* vol. 215, no. 2, pp. 287-291.

Silkworth, W.T. & Cimini, D. 2012, "Transient defects of mitotic spindle geometry and chromosome segregation errors", *Cell division,* vol. 7, no. 1, pp. 19-19.

Silva, W.A.,Jr, Gnjatic, S., Ritter, E., Chua, R., Cohen, T., Hsu, M., Jungbluth, A.A., Altorki, N.K., Chen, Y.T., Old, L.J., Simpson, A.J. & Caballero, O.L. 2007, "PLAC1, a trophoblast-specific cell surface protein, is expressed in a range of human tumors and elicits spontaneous antibody responses", Cancer immunity, vol. 7, pp. 18.

Simpson, A.J., Caballero, O.L., Jungbluth, A., Chen, Y. & Old, L.J. 2005, "Cancer/testis antigens, gametogenesis and cancer", *Nature Reviews Cancer,* vol. 5, no. 8, pp. 615-625.

Singh, A.P., Chauhan, S.C., Bafna, S., Johansson, S.L., Smith, L.M., Moniaux, N., Lin, M. & Batra, S.K. 2006, "Aberrant expression of transmembrane mucins, MUC1 and MUC4, in human prostate carcinomas", *The Prostate*, vol. 66, no. 4, pp. 421-9.

Sjögren, C. & Nasmyth, K. 2001, "Sister chromatid cohesion is required for postreplicative double-strand break repair in Saccharomyces cerevisiae. *Current Biology*, vol. 11, no. 12, pp. 991-995.

Skibbens, R.V., Colquhoun, J.M., Green, M.J., Molnar, C.A., Sin, D.N., Sullivan, B.J. & Tanzosh, E.E. 2013, "Cohesinopathies of a feather flock together", *PLoS genetics*, vol. 9, no. 12, pp. e1004036.

Smagulova, F., Gregoretti, I.V., Brick, K., Khil, P., Camerini-Otero, R. & Petukhova, G.V. 2011, "Genome-wide analysis reveals novel molecular features of mouse recombination hotspots", *Nature*, vol. 472, no. 7343, pp. 375-8.

Smith, D.B. & Johnson, K.S. 1988, "Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase", *Gene,* vol. 67, no. 1, pp. 31-40.

Smith, H.a. & McNeel, D.G. 2010, "The SSX family of cancer-testis antigens as target proteins for tumor therapy", *Clinical & developmental immunology,* vol. 2010, pp. 150591-150591.

Smith, H.A., Cronk, R.J., Lang, J.M. & McNeel, D.G. 2011, "Expression and immunotherapeutic targeting of the SSX family of cancer-testis antigens in prostate cancer", *Cancer research,* vol. 71, no. 21, pp. 6785-6795.

Smith, I.M., Glazer, C.A., Mithani, S.K., Ochs, M.F., Sun, W., Bhan, S., Vostrov, A., Abdullaev, Z., Lobanenkov, V. & Gray, A. 2009, "Coordinated activation of candidate proto-oncogenes and cancer testes antigens via promoter demethylation in head and neck cancer and lung cancer", *PloS one,* vol. 4, no. 3, pp. e4961.

Steele-Perkins, G., Fang, W., Yang, X.H., Van Gele, M., Carling, T., Gu, J., Buyse, I.M., Fletcher, J.A., Liu, J., Bronson, R., Chadwick, R.B., de la Chapelle, A., Zhang, X., Speleman, F. & Huang, S. 2001, "Tumor formation and inactivation of RIZ1, an Rb-binding member of a nuclear protein-methyltransferase superfamily", *Genes & development,* vol. 15, no. 17, pp. 2250-2262.

Stratton, M.R., Campbell, P.J. & Futreal, P.A. 2009, "The cancer genome", *Nature,* vol. 458, no. 7239, pp. 719-24.

Suri, A. 2006, "Cancer testis antigens-their importance in immunotherapy and in the early detection of cancer", *Expert opinion on Biological Therapy*, vol.6, no 4, pp. 379-389.

Suzuki, H., Gabrielson, E., Chen, W., Anbazhagan, R., van Engeland, M., Weijenberg, M.P., Herman, J.G. & Baylin, S.B. 2002, "A genomic screen for genes upregulated by demethylation and histone deacetylase inhibition in human colorectal cancer", *Nature genetics,* vol. 31, no. 2, pp. 141-149.

Takai, D. & Jones, P.A. 2002, "Comprehensive analysis of CpG islands in human chromosomes 21 and 22", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 99, no. 6, pp. 3740-3745.

Thoma, C.R., Toso, A., Meraldi, P. & Krek, W. 2011, "Mechanisms of aneuploidy and its suppression by tumour suppressor proteins", *Swiss medical weekly,* vol. 141, pp. w13170-w13170.

Thompson, S.L., Bakhoum, S.F. & Compton, D.A. 2010, "Mechanisms of chromosomal instability", *Current biology : CB,* vol. 20, no. 6, pp. R285-95.

Tomasetti, C., Vogelstein, B. & Parmigiani, G. 2013, "Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 110, no. 6, pp. 1999-2004.

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.

Tureci, O., Chen, Y., Sahin, U., GUre, A.O., Zwick, C., Villena, C., Tsang, S., Seitz, G., Old, L. & Pfreundschuh, M. 1998a, "Expression of SSX genes in human tumors", *International journal of cancer,* vol. 77, no. 1, pp. 19-23.

Tureci, O., Sahin, U., Zwick, C., Koslowski, M., Seitz, G. & Pfreundschuh, M. 1998b, "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*, *134*(10), 1823-1831.

Van Cutsem, E., Köhne, C., Hitre, E., Zaluski, J., Chang Chien, C., Makhson, A., D'Haens, G., Pintér, T., Lim, R. & Bodoky, G. 2009, "Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer", *New England Journal of Medicine,* vol. 360, no. 14, pp. 1408-1417.

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.

Vatolin, S., Abdullaev, Z., Pack, S.D., Flanagan, P.T., Custer, M., Loukinov, D.I., Pugacheva, E., Hong, J.A., Morse, H.,3rd, Schrump, D.S., Risinger, J.I., Barrett, J.C. & Lobanenkov, V.V. 2005, "Conditional expression of the CTCF-paralogous transcriptional factor BORIS in normal cells results in demethylation and derepression of MAGE-A1 and reactivation of other cancer-testis genes", *Cancer research,* vol. 65, no. 17, pp. 7751-7762.

Verbsky, J. W., Wilson, M. P., Kisseleva, M. V., Majerus, P. W., & Wente, S. R. (2002). The synthesis of inositol hexakisphosphate Characterization of human inositol 1, 3, 4, 5, 6-pentakisphosphate 2-kinase. *Journal of Biological Chemistry*, 277(35), 31857-31862.

Vigna, E., Cavalieri, S., Ailles, L., Geuna, M., Loew, R., Bujard, H., & Naldini, L. (2002). Robust and efficient regulation of transgene expression in vivo by improved tetracycline-dependent lentiviral vectors. *Molecular therapy*, 5(3), 252-261.

Vogelstein, B. & Kinzler, K.W. 2004, "Cancer genes and the pathways they control", *Nature medicine,* vol. 10, no. 8, pp. 789-99.

Vogt, P.K. 2012, "Retroviral oncogenes: a historical primer", *Nature reviews.Cancer,* vol. 12, no. 9, pp. 639-48.

Von Stetina, J.,R. & Orr-Weaver, T. 2011, "Developmental control of oocyte maturation and egg activation in metazoan models", *Cold Spring Harbor perspectives in biology,* vol. 3, no. 10, pp. a005553.

Wade, Tracey D., Scott Gordon, Sarah Medland, Cynthia M. Bulik, Andrew C. Heath, Grant W. Montgomery, and Nicholas G. Martin. 2013 "Genetic variants associated with disordered eating." *International Journal of Eating Disorders* 46, no. 6, 594-608

Walczak, C.E., Cai, S. & Khodjakov, A. 2010, "Mechanisms of chromosome behaviour during mitosis", *Nature Reviews Molecular Cell Biology,* vol. 11, no. 2, pp. 91-102.

Weber, J., Salgaller, M., Samid, D., Johnson, B., Herlyn, M., Lassam, N., Treisman, J. & Rosenberg, S.A. 1994, "Expression of the MAGE-1 tumor antigen is up-regulated by the demethylating agent 5-aza-2'-deoxycytidine", *Cancer research,* vol. 54, no. 7, pp. 1766-1771.

Weiner, L.M., Surana, R. & Wang, S. 2010, "Monoclonal antibodies: versatile platforms for cancer immunotherapy", *Nature Reviews Immunology,* vol. 10, no. 5, pp. 317-327.

Welch, J.S., Ley, T.J., Link, D.C., Miller, C.a., Larson, D.E., Koboldt, D.C., Wartman, L.D., Lamprecht, T.L., Liu, F., Xia, J., Kandoth, C., Fulton, R.S., McLellan, M.D., Dooling, D.J., Wallis, J.W., Chen, K., Harris, C.C., Schmidt, H.K., Kalicki-Veizer, J., Lu, C., Zhang, Q., Lin, L., O'Laughlin, M.,D.,

McMichael, J.F., Delehaunty, K.D., Fulton, L.a., Magrini, V.J., McGrath, S.D., Demeter, R.T., Vickery, T.L., Hundal, J., Cook, L.L., Swift, G.W., Reed, J.P., Alldredge, P.a., Wylie, T.N., Walker, J.R., Watson, M.a., Heath, S.E., Shannon, W.D., Varghese, N., Nagarajan, R., Payton, J.E., Baty, J.D., Kulkarni, S., Klco, J.M., Tomasson, M.H., Westervelt, P., Walter, M.J., Graubert, T.a., DiPersio, J.F., Ding, L., Mardis, E.R. & Wilson, R.K. 2012, "The origin and evolution of mutations in acute myeloid leukemia", *Cell*, vol. 150, no. 2, pp. 264-78.

Whitehurst, A.W. 2014, "Cause and consequence of cancer/testis antigen activation in cancer", *Annual Review of Pharmacology and Toxicology*, vol. 54, pp. 251-72.

Whitehurst, A.W., Bodemann, B.O., Cardenas, J., Ferguson, D., Girard, L., Peyton, M., Minna, J.D., Michnoff, C., Hao, W. & Roth, M.G. 2007, "Synthetic lethal screen identification of chemosensitizer loci in cancer cells", *Nature*, vol. 446, no. 7137, pp. 815-819.

Wigle, T.J. & Copeland, R.A. 2013, "Drugging the human methylome: an emerging modality for reversible control of aberrant gene transcription", *Current Opinion in Chemical Biology,* vol. 17, no. 3, pp. 369-378.

Wischnewski, F., Pantel, K. & Schwarzenbach, H. 2006, "Promoter demethylation and histone acetylation mediate gene expression of MAGE-A1, -A2, -A3, and -A12 in human cancer cells", *Molecular Cancer Research : MCR*, vol. 4, no. 5, pp. 339-349.

Woloszynska-Read, A., James, S.R., Link, P.A., Yu, J., Odunsi, K. & Karpf, A.R. 2007, "DNA methylation-dependent regulation of BORIS/CTCFL expression in ovarian cancer", *Cancer Immunity,* vol. 7, pp. 21.

Woodward, E.L., Olsson, M.L., Johansson, B. & Paulsson, K. 2014, "Allelic variants of PRDM9 associated with high hyperdiploid childhood acute lymphoblastic leukaemia", *British Journal of Haematology*.

Wu, H., Mathioudakis, N., Diagouraga, B., Dong, A., Dombrovski, L., Baudat, F., Cusack, S., de Massy, B. & Kadlec, J. 2013, "Molecular Basis for the Regulation of the H3K4 Methyltransferase Activity of PRDM9", *Cell Reports,* vol. 5, no. 1, pp. 13-20.

Wu, X. & Ruvkun, G. 2010, "Cancer. Germ cell genes and cancer", *Science,* vol. 330, no. 6012, pp. 1761-1762.

Wu, Y., Ferguson III, J.E., Wang, H., Kelley, R., Ren, R., McDonough, H., Meeker, J., Charles, P.C., Wang, H. & Patterson, C. 2008, "PRDM6 is enriched in vascular precursors during development and inhibits endothelial cell proliferation, survival, and differentiation", *Journal of Molecular and Cellular Cardiology,* vol. 44, no. 1, pp. 47-58.

Xiaoying, K., Joma, J., Anders, P., Rong, L., Yvonne, T., Yoonjung, S., Jung-Hyun, M., Liling, W., Anna, N. & Jianhe, P. 2014, "Characterization of the histone methyltransferase PRDM9 using biochemical, biophysical and chemical biology techniques", *Biochemical Journal,* vol. 461, no. 2, pp. 323-334.

Yamaji, M., Seki, Y., Kurimoto, K., Yabuta, Y., Yuasa, M., Shigeta, M., Yamanaka, K., Ohinata, Y. & Saitou, M. 2008, "Critical function of Prdm14 for the establishment of the germ cell lineage in mice", *Nature Genetics,* vol. 40, no. 8, pp. 1016-1022.

Yan, J., Jiang, J., Lim, C.A., Wu, Q., Ng, H.H. & Chin, K.C. 2007, "BLIMP1 regulates cell growth through repression of p53 transcription", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 104, no. 6, pp. 1841-1846.

Ye, Y., Yin, D., Chen, L., Zhou, Q., Shen, R., He, G., Yan, Q., Tong, Z., Issekutz, A.C. & Shapiro, C.L. 2010, "Identification of Piwil2-like (PL2L) proteins that promote tumorigenesis", *PLoS One,* vol. 5, no. 10, e13406.

Yoshida, M., Kijima, M., Akita, M. & Beppu, T. 1990, "Potent and specific inhibition of mammalian histone deacetylase both in vivo and in vitro by trichostatin A", *The Journal of Biological Chemistry,* vol. 265, no. 28, pp. 17174-17179.

Yu, X., Jacobs, S.A., West, S.C., Ogawa, T. & Egelman, E.H. 2001, "Domain structure and dynamics in the helical filaments formed by RecA and Rad51 on DNA", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 98, no. 15, pp. 8419-24.

Yuan, X., He, J., Sun, F. & Gu, J. 2013, "Effects and interactions of MiR-577 and TSGA10 in regulating esophageal squamous cell carcinoma", *International Journal of Clinical and Experimental Pathology,* vol. 6, no. 12, pp. 2651.

Zakharyevich, K., Tang, S., Ma, Y. & Hunter, N. 2012, "Delineation of joint molecule resolution pathways in meiosis identifies a crossover-specific resolvase", *Cell,* vol. 149, no. 2, pp. 334-47.

Zendman, A.J., Ruiter, D.J. & Van Muijen, G.N. 2003, "Cancer/testis-associated genes: Identification, expression profile, and putative function", *Journal of Cellular Physiology,* vol. 194, no. 3, pp. 272-288.

Zendman, A.J.W., Zschocke, J., van Kraats, A.A., de Wit, N.J.W., Kurpisz, M., Weidle, U.H., Ruiter, D.J., Weiss, E.H. & van Muijen, G.N.P. 2003b, "The human SPANX multigene family: genomic organization, alignment and expression in male germ cells and tumor cell lines", *Gene,* vol. 309, no. 2, pp. 125-133.

Zhao, G.Q., Zhou, X., Eberspaecher, H., Solursh, M. & de Crombrugghe, B. 1993, "Cartilage homeoprotein 1, a homeoprotein selectively expressed in chondrocytes", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 90, no. 18, pp. 8633-8637.

Zickler, D. & Kleckner, N. 1998, "The leptotene-zygotene transition of meiosis", *Annual Review of Genetics,* vol. 32, pp. 619-97.

Zickler, D. & Kleckner, N. 1999, "Meiotic chromosomes: integrating structure and function", *Annual Review of Genetics,* vol. 33, pp. 603-754.

Zickler, D. 2006, "From early homologue recognition to synaptonemal complex formation", *Chromosoma,* vol. 115, no. 3, pp. 158-74.

Appendix

List of Publication

1. Meta-analysis of clinical data using human meiotic genes identifies a novel cohort of highly restricted cancer-specific marker genes. (Oncotarget Journal)

Authors: Feichtinger J, Aldeailej I, Anderson R, Almutairi M, Almatrafi A, Alsiwiehri N, Griffiths K, Stuart N, Wakeman JA, Larcombe L, McFarlane RJ.

2. Identification of a class of human cancer germline genes with transcriptional silencing refractory to the hypomethylating drug 5-aza-2˙-deoxycytidine. (In process Manuscript Oncosciecne)

Authors: Ahmed Almatrafi, Julia Feichtinger, Jane A. Wakeman, Lee D. Larcombe, Ramsay J. McFarlane.