Investigating the potential role of recombination regulator PRDM9 in mitochondria.

Emily McIlwaine (BSc MRes)

Submitted for degree of Doctor of Philosophy

September 2016

Institute of Genetic Medicine

Faculty of Medical Sciences

Newcastle University

ii

Abstract

At present, 805 mitochondrial DNA (mtDNA) deletions have been described. Short direct repeat regions of DNA flank many of these deletions, suggesting that specific regions of the mtDNA molecule have a susceptibility to deletion formation. Despite this, the exact underlying cellular mechanisms facilitating mtDNA deletions are unclear.

PR domain 9 (PRDM9) is a meiotic-specific protein responsible for determining the site of recombination in the nuclear genome. Through its zinc finger repeat region, PRDM9 binds a specific DNA consensus sequence, and acts as a methyl transferase, opening chromatin for DNA crossover events to occur. This is of interest as mitochondrial DNA also contains PRDM9 binding motif sites.

This thesis outlines the experimental steps taken to determine if PRDM9 has any involvement in mtDNA maintenance and viability.

Firstly, an *in silico* approach was used to screen mtDNA sequences from 31,551 individuals for the presence of the PRDM9 binding motif, identifying multiple putative binding sites in and around known deletion forming flanking regions. In addition, population and phylogenetic stratification showed differential mtDNA binding motif patterns, potentially explaining the variable deletion frequencies between mtDNA haplogroups and populations.

Secondly, to test the potential interaction between PRDM9 and mtDNA, complete genotyping of the *PRDM9* zinc finger repeat region in a cohort of 48 mitochondrial single deletion patients and 50 healthy controls was performed. However, there was no association between *PRDM9* haplotype and the formation of mtDNA deletions. Heterozygous individuals were significantly increased in the patient cohort compared to controls although no particular allele was associated with mtDNA deletion.

Finally, PRDM9 protein levels were interrogated in cell lines and tissue samples. However, due to timing of expression it was not possible to reliably detect nascent protein using commercially available antibodies. To overcome this, stable cell lines overexpressing Flag-tagged PRDM9 were created. Low levels of PRDM9 expression were detected by immunoblotting indicating overexpression had worked but also indicating that PRDM9 turnover in cells is likely rapid.

Given the data presented, and despite the presence of multiple putative PRDM9 binding sites in almost all mitochondrial genomes studied, we conclude that it is unlikely that PRDM9 has a significant effect on the maintenance of mtDNA. However, to the best of my knowledge this is the first stable PRDM9 overexpression model created and it has provided a unique insight into some of the functions of this protein.

Acknowledgements

Firstly, I am grateful to my supervisors Professors Patrick Chinnery and Rita Horvath for the opportunity to join their lab groups and for their support throughout the course of my PhD. I also wish to thank Dr Gavin Hudson and Dr Angela Pyle for their helpful comments and discussions about my research. Completion of this thesis would not have been possible without the extensive experimental supervision and guidance of Dr Aurora Gómez-Durán, thank you for teaching me to become a better researcher and scientist.

I would also like to thank Dr Erika Fernández-Vizarra and Professor Massimo Zeviani for hosting me in their laboratory for a short time. My thanks also extend to everyone in the lab group, you have been a great support network throughout the last few years. Your suggestions, technical knowhow and feedback have helped shape how I approach new challenges and I am very grateful to have worked with you all. Many of you have become great friends and I look forward to many more pub visits in the future, no doubt all over the world.

Lastly, I wish to thank my family for being there for me during my studies and for always supporting me.

Table of Contents

Table of Figures

xiv

Table of Tables

Abbreviations

ACTB   β actin

ADP   adenosine diphosphate

AP2γ   AP-2 gamma

ATFS-1   activating transcription factor associated with stress

ATP   adenosine triphosphate

B-ALL   B-cell acute lymphoblastic leukaemia

BGH   bovine growth hormone

BLAST   basic local alignment search tool

BSA   bovine serum albumin

143B   osteosarcoma cell line

CD   cardiac tissue

cDNA   complementary deoxyribonucleic acid

CMV   cytomegalovirus

CPEO   chronic progressive external ophthalmoplegia

CpG   5' - C – phosphate- G -3'

Ct   threshold cycle

CT   cancer testis antigen

CytC   cytochrome c

DAPI   4',6-Diamidino-2-Phenylindole

DNA   deoxyribonucleic acid

DNMT1   DNA cytosine-5 methyltransferase 1

dNTP   deoxy nucleoside triphosphate

dH$_2$O   distilled water

DICER   RNAseIII-like protein

DMD   Duchenne muscular dystrophy

DMEM   Dulbecco's modified eagle medium

DOK7   docking protein 7

DPX	dibutyl phthalate xylene
DSB	double strand break
EDTA	ethylene diamine tetraacetic acid
EGTA	ethylene glycol tetraacetic acid
EtBr	ethidium bromide
ETC	electron transport chain
EtOH	ethanol
FADH2	flavin adenine dinucleotide
FBS	foetal bovine serum
F(ab)	fragment antibody
FRET	fluorescent resonance energy transfer
FRT	Flp recombinase target
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
GFP	green fluorescent protein
HeH-ALL	high hyperploidy acute lymphoblastic leukaemia
HEK293	human embryonic kidney 293 cells
hESC	human embryonic stem cell
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid
H3K4	histone 3 lysine 4
H3K4me1	histone 3 lysine 4 monomethylation
H3K4me2	histone 3 lysine 4 dimethylation
H3K4me3	histone 3 lysine 4 trimethylation
H_2O	dihydrogen oxide (water)
HPLC	high performance liquid chromatography
HRP	horseradish peroxidase
HSP-5	heat shock protein family A member 5
HSP-60	heat shock protein family D member 1
Hstx2	hybrid sterility X chromosome 2
IgG	immunoglobulin G

IGV	integrated genome viewer
IP	immunoprecipitation
KRAB	Krüpple associated box
KSS	Kearns-Sayer syndrome
LCMSMS	liquid chromatography mass spectrometry
LGT	lateral gene transfer
LHON	Lebers hereditary optic neuropathy
MAM	mitochondrial associated membrane
5mC	5-methylcytosine
MCS	multiple cloning site
MELAS	mitochondrial encephalomyopathy lactic acidosis and stroke like episodes
MERRF	myoclonic epilepsy with ragged red fibres
MRB	mitochondria resuspension buffer
MRE11	homologue A double strand break repair nuclease
mRNA	messenger ribonucleic acid
MT-CO1	mitochondrial encoded cytochrome c oxidase I
MT-CO2	mitochondrial encoded cytochrome c oxidase II
MT-CYB	mitochondrial complex III cytochrome b
mtDNA	mitochondrial DNA
MT-ND1	mitochondrial nicotinamide adenosine dinucleotide dehydrogenase subunit 1
NADH	nicotinamide adenosine dinucleotide
NANOG	nanog homeobox
NCBI	National Centre for Biotechnology Information
NGS	next generation sequencing
NT2	neuron-committed teratocarcinoma cell line
NuMt	nuclear-mitochondrial DNA
OCT	optimal cutting temperature compound
OPA1	optic atrophy 1
OXPHOS	oxidative phosphorylation; PBS- phosphate buffered saline

PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEO1	twinkle helicase
PFA	paraformaldehyde
PGC	primordial germ cells
Pi	phosphate
POLG	DNA polymerase-ɣ
POLRMT	polymerase RNA mitochondria
PRDM1	PR domain containing 1
PRDM9	PR domain containing 9
PRDM14	PR domain containing 14
PVDF	polyvinylidene difluoride
qPCR	quantitative polymerase chain reaction
RAD51	RAD51 recombinase
rCRS	Revised Cambridge Reference Sequence
RISC	RNA induced silencing complex
RNA	ribonucleic acid; RNASEP- ribonuclease P
rRNA	ribosomal RNA
RNR1	mitochondrial 12S RNA
RNR2	mitochondrial 16S RNA
RT-qPCR	real time or reverse transcription quantitative polymerase chain reaction
SAM	S-adenosylmethionine
SDHA	succinate dehydrogenase subunit A
SDS	sodium dodecyl sulphate
SKM	skeletal muscle
siRNA	short-interfering ribonucleic acid
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
SPO11	initiator of meiotic double strand breaks

SSXRD	SSX repression domain
TCA	tricarboxylic acid
TFAM	mitochondrial transcription factor A
tRNA	transfer RNA
TOMM20	translocase of outer mitochondrial membrane 20
TTBS	tween-tris buffered saline
UPR^{mt}	unfolded protein response mitochondrial
UQ	ubiquinone
UQH2	ubiquinol
VDAC1	voltage dependant anion channel 1
ZnF	zinc finger

Units of Measurement

bp	base pair
°C	degrees Celsius
cm	centimetre
G	relative centrifugal force
g	gram
Kb	kilobase pair
kDa	kilo Daltons
L	litre
%	percentage
M	molar
mg	milligram
mm	millimetre
nM	nanomolar
pg	picogram
pM	picomole
rcf	relative centrifugal force
rpm	revolutions per minute
SD	standard deviation from the mean
μL	micro litre
μM	micro molar
μm	micro meter
U	unit

Chapter 1 Introduction

1.1 Mitochondrial biology

1.1.1 Evolutionary origins of the mitochondria

Mitochondria are double membrane bound subcellular organelles present in all eukaryotic cells. They act as a host for many critical cellular processes such as the tricarboxylase acid cycle (TCA), fatty-acid β-oxidation, branched chain amino acid degradation, oxidative phosphorylation (OXPHOS), glycolysis, iron-sulphur cluster formation and apoptosis. The evolutionary origins of the eukaryotic cell are thought to be the result of an internalisation event around 1.7 billion years ago, when an α-proteobacteria entered an archaeon, thus establishing a symbiotic relationship (Parfrey *et al.*, 2011; Williams *et al.*, 2013; Raymann *et al.*, 2015; Spang *et al.*, 2015). Whilst the presence of an endosymbiont could be detrimental to the host cell, this particular event introduced an attractive trade off; production of cellular energy in the form of adenosine triphosphate (ATP).

Three main advantages are thought to have been vital for the success of the eukaryotic ancestor. Firstly, the switch from asexual (clonal) to sexual reproduction allowed greater genetic diversity and limited the accumulation of deleterious mutations from generation to generation (Muller, 1964; Felsenstein, 1974). This key difference allowed eukaryotes to switch from unidirectional lateral gene transfer (LGT) to genome recombination and segregation via meiosis and mitosis (Ku *et al.*, 2015; Speijer *et al.*, 2015). Secondly, eukaryotic cells became compartmentalised by the establishment of a nuclear membrane which allowed gene regulation to be segregated from a dedicated translation compartment, the cytosol (French *et al.*, 2007). Importantly, this physical separation allows production of pre-mRNA species before maturation and transfer to the cytosol where protein synthesis can occur. Finally, by allowing the cell to produce large amounts of ATP, the eukaryotic cell was able to provide an environment where high rates of protein translation could occur, providing the cell with a diverse selection of protein species (das Neves *et al.*, 2010). In addition, both the archaeal host and the α-proteobacteria brought to the table two independent genomes. Studies have shown that in the 1.7 billion years since this endosymbiotic event, large portions of the mitochondrial DNA molecule have been transferred into the host nuclear genome and that this process is ongoing (Huang *et al.*, 2005; Ju *et al.*, 2015).

1.1.2 Mitochondrial OXPHOS

The success of the eukaryotic cell is also due to the ability to perform aerobic respiration using molecular oxygen. Located on the inner mitochondrial membrane are the five protein complexes required to perform OXPHOS, utilising oxygen to generate cellular energy in the form of ATP. These protein complexes are comprised of both nuclear and mitochondrially encoded core subunits as well as a host of assembly factors. The electron transport chain (ETC) comprises Complexes I-IV and is responsible for pumping protons across the inner membrane by the transport of electrons (for review see Papa *et al.* (2012)).

Electrons are donated to Complex I by nicotinamide adenine dinucleotide (NADH) and Complex II by flavin adenine dinucleotide (FADH2) (Ziegler *et al.*, 1959; Sumegi and Srere, 1984). Next, electrons from both complexes I and II are transferred to Complex III via reduction of the lipid molecule ubiquinone (UQ) (Crane *et al.*, 1959), which in turn donates electrons to Complex III (de Vries *et al.*, 1981). Electrons flow from Complex III to Complex IV via reduction of the hemeprotein cytochrome c (CytC) (Lester *et al.*, 1959). The energy released as electrons are transferred along the ETC is used to establish a proton gradient between the inner membrane space and the mitochondrial matrix through Complexes I, III and IV (Baum and Rieske, 1966). To complete OXPHOS, Complex V utilises this transmembrane proton motive force to drive ATP production (Zoratti *et al.*, 1982). Phosphate (Pi) and adenosine diphosphate (ADP) available in the mitochondrial matrix are converted to ATP by the rotary mechanism of Complex V (Cooper and Lehninger, 1957; Stock *et al.*, 1999). Figure 1.1 shows the classical view of the five respiratory chain complexes, however recent data has demonstrated that the localisation and interaction of these protein complexes are highly plastic (Lapuente-Brun *et al.*, 2013). High molecular weight complexes termed 'supercomplexes' are formed by different combinations of Complex I, III and IV, creating a more dynamic ETC than previously described (for review see Chaban *et al.* (2014)).
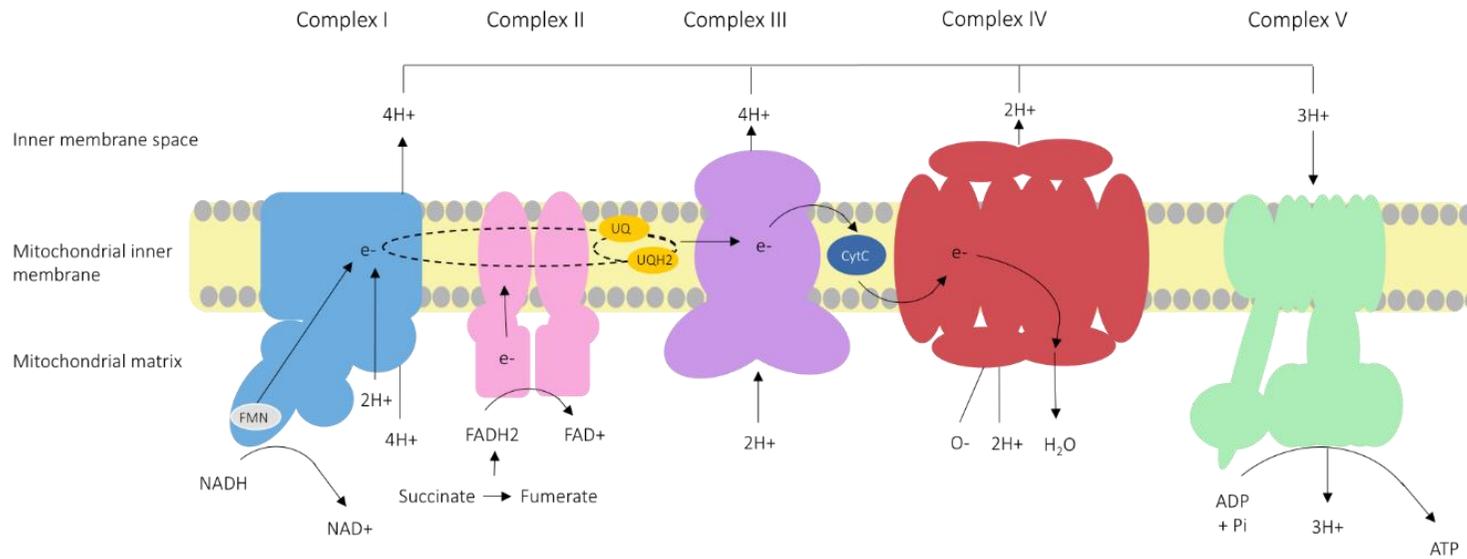
Figure 1.1 Schematic of the five protein complexes required for OXPHOS, located on the inner mitochondrial membrane. Complexes I-IV comprise the electron transport chain (ETC) required for respiration through production of oxygen. The ETC also relies on prosthetic groups; flavin mononucleotide (FMN), ubiquinone (UQ) and cytochrome c (CytC). Complex V (ATP Synthase) is required for oxidative phosphorylation as it utilises available protons generated by the ETC to catalyse the conversion of ADP+Pi to ATP.

1.2 Mitochondrial DNA

1.2.1 Structure of mtDNA

In humans, mitochondrial DNA (mtDNA) is a ~16.5 Kb plasmid molecule contained within the organelle, physically separated from the nuclear genome (Corneo *et al.*, 1966). mtDNA encodes 13 protein subunits of the OXPHOS complexes as well as 22 tRNA's and 2 rRNA's required for mtDNA synthesis and mitochondrial protein synthesis respectively (Andrews *et al.*, 1999). Human mtDNA is double stranded with the majority of the coding regions located on the heavy strand and only one protein encoding gene on the light strand, *MT-ND6* (Figure 1.2).
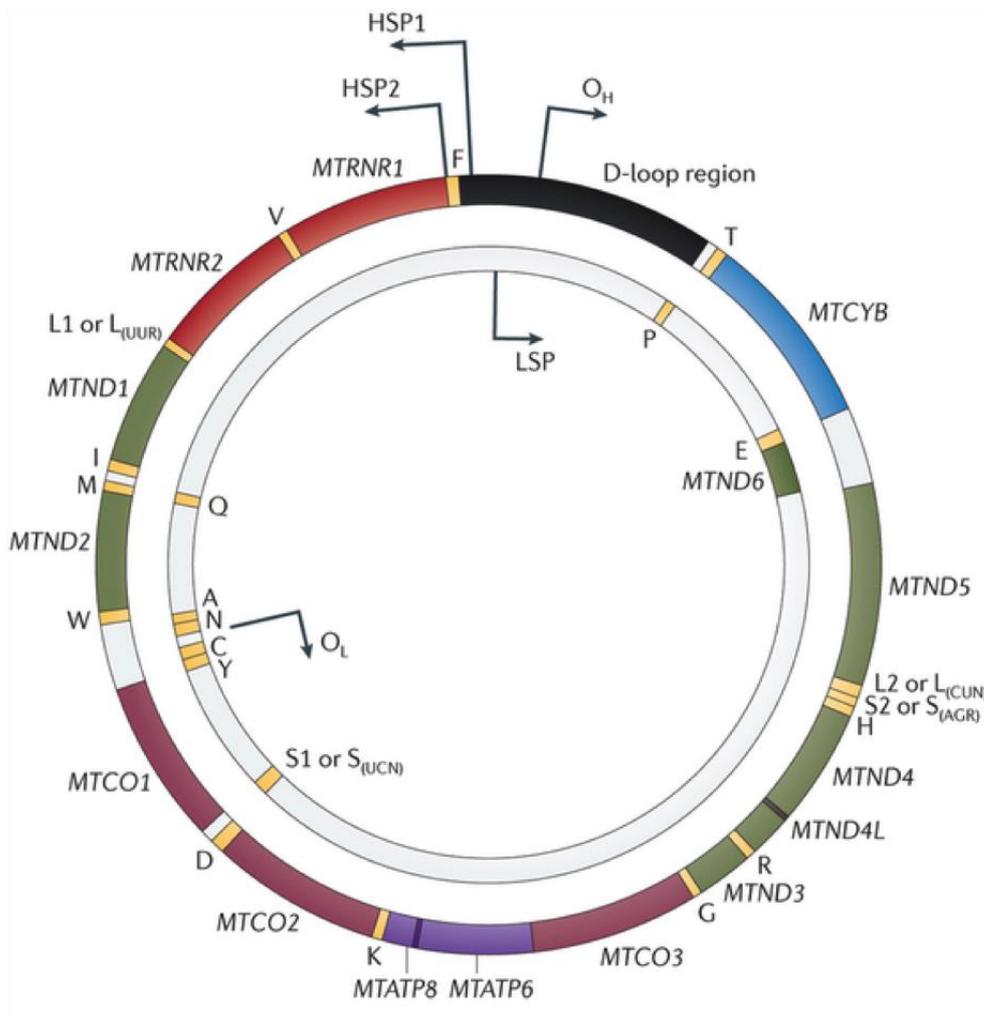
Figure 1.2 Map of the human mitochondrial DNA. Both strands are shown; the heavy strand is the outer circle and the light strand is the inner. All 13 protein coding genes, 22 tRNAs, 2 rRNAs and replication start sites are shown. The 'major arc' of the molecule comprises m.5721 (O_L) to m.15887 (*MT-TT*). Adapted from (Stewart and Chinnery, 2015).

Within the inner mitochondrial matrix, mtDNA molecules are encased in proteinaceous structures called nucleoids, allowing some protection from reactive oxygen species generated by the ETC (Bogenhagen, 2012; Logan *et al.*, 2014; Kukat *et al.*, 2015; Rajala *et al.*, 2015). Despite this protective barrier, mtDNA has a significantly higher mutation rate than the nuclear genome (Hunter *et al.*, 2012).

Replication of mtDNA is independent of that of the nuclear genome and the cell cycle, therefore mtDNA content can be increased at any time in response to extra- or intracellular demand. Several proteins required for mtDNA replication and maintenance have been uncovered, all encoded by nuclear genes. The exact mechanism of mtDNA replication is still debated with two main theories supported by experimental data; the strand displacement model and the strand coupled model. During strand displacement, replication occurs unidirectionally from the origin of heavy strand replication site (O_H) and occurs continuously until the light strand replication site is exposed (O_L) (Bogenhagen *et al.*, 1979; Kang *et al.*, 1997; Brown *et al.*, 2005) mitochondrial RNA polymerase (POLRMT) then synthesises RNA primers which are utilised by mitochondrial DNA polymerase gamma (POLG) to begin light strand synthesis (Fuste *et al.*, 2010). In the strand coupled model, unidirectional replication occurs at O_H but in this model RNA/DNA hybrid molecules are created (Pohjoismaki *et al.*, 2010). The lagging strand is comprised of RNA primers which are then replaced by DNA as replication proceeds, eventually creating the new mtDNA molecule (Reyes *et al.*, 2013).

The plasmid nature of the mtDNA molecule means that it is transcribed firstly as a polycistronic transcript which is then cleaved to allow translation of the mtDNA encoded proteins (Ojala *et al.*, 1981). Although replication and translation of mtDNA are two independent processes, many of the proteins involved are required for both, suggesting that they are intrinsically linked. In addition, the O_H site within the D-loop region of mtDNA is used for both replication and translation initiation (Chang *et al.*, 1985) and recently POLRMT has been shown to be required for both processes (Kuhl *et al.*, 2016). This is not surprising, as mtDNA replication would primarily occur to increase transcript levels and therefore protein expression, requiring these two processes to occur within a close time frame.

1.2.2 Inheritance of mtDNA

The inheritance of mtDNA is uniparental, only transmitted from the maternal gamete, as paternal mitochondria are removed from the zygote shortly after fertilisation (Al Rawi *et al.*, 2011; Sato and Sato, 2011). Interestingly, evidence of paternal mtDNA transmission has been described in a patient with mitochondrial myopathy (Schwartz and Vissing, 2002) as well as after strain crossings of mice, opening the possibility of paternal 'leakage' in rare cases (Luo *et al.*, 2013). It was suggested that levels of paternal mtDNA are present within the population but are low-level and have not been detected due to inadequate sensitivity of sequencing analysis (Nunes *et al.*, 2013). However, the recent use of next-generation sequencing technologies on parent-offspring trios has shown no evidence of paternal transmission (Pyle *et al.*, 2015). New research suggests that mitochondrial endonuclease G mediates paternal mtDNA degradation (Zhou *et al.*, 2016) most likely through activation of autophagic degradation pathways (Song *et al.*, 2016b). It is therefore widely accepted that the mtDNA population within a cell is solely maternally inherited.

Each mitochondrion contains several mtDNA molecules and in turn, each cell contains many mitochondria. Therefore, depending on the energy demand of the cell and tissue, there can be up to ∼100,000 copies of mtDNA per cell (Yu-Wai-Man and Chinnery, 2012; Smeets, 2013). This polyploidy genome, coupled with a high mutation rate, can lead to the phenomenon of heteroplasmy; where a single mitochondria, cell, tissue or organ can have several distinct 'populations' of mtDNA molecules. In contrast, homoplasmy occurs when all of the mtDNA molecules are identical. This high mutation rate leads to the occurrence of both pathogenic mutations and neutral variants. Importantly, the heteroplasmy level of pathogenic mutations can be quantified and correlated with mitochondrial function and disease severity (Payne *et al.*, 2013).

1.2.3 Mitochondrial population haplogroups

Due to the polymorphic nature of mtDNA, human populations contain several distinct and heritable mtDNA sequences known as mtDNA haplogroups (Wallace *et al.*, 1985). The high mutation rate of mtDNA has resulted in rapid divergence of mtDNA sequences amongst human populations. These population specific single nucleotide polymorphisms (SNPs) are non-pathogenic and therefore persist within the population

allowing ancestry to be traced back through the maternal lineage. Haplogroup status has also been used to determine where populations originated and to trace the migration patterns of early humans across continents (Cann *et al.*, 1987; Torroni *et al.*, 1992; Torroni *et al.*, 1993). A phylogenetic tree of mtDNA variants can be found on the PhyloTree website (PhyloTree); combining extensive sequencing analysis of mtDNA from several populations, from the most ancestral African haplogroup L0 to the most recent variants, as this divergence is still occurring. African populations harbour the most ancient of mtDNA sequences, consistent with evidence that the last common ancestor of *Homo Sapiens* originated in this region (Cann *et al.*, 1987; Chen *et al.*, 1995). Asian haplogroups have been well defined due to the migration of founder populations into northern America and Australasia (Brown *et al.*, 1998; Starikovskaya *et al.*, 1998; Derenko *et al.*, 2000; Kong *et al.*, 2003). European haplogroups are diverse due to several different migration patterns of founder groups into this region, as shown in Figure 1.3 (Torroni *et al.*, 1996; Richards *et al.*, 2000; Torroni *et al.*, 2000).
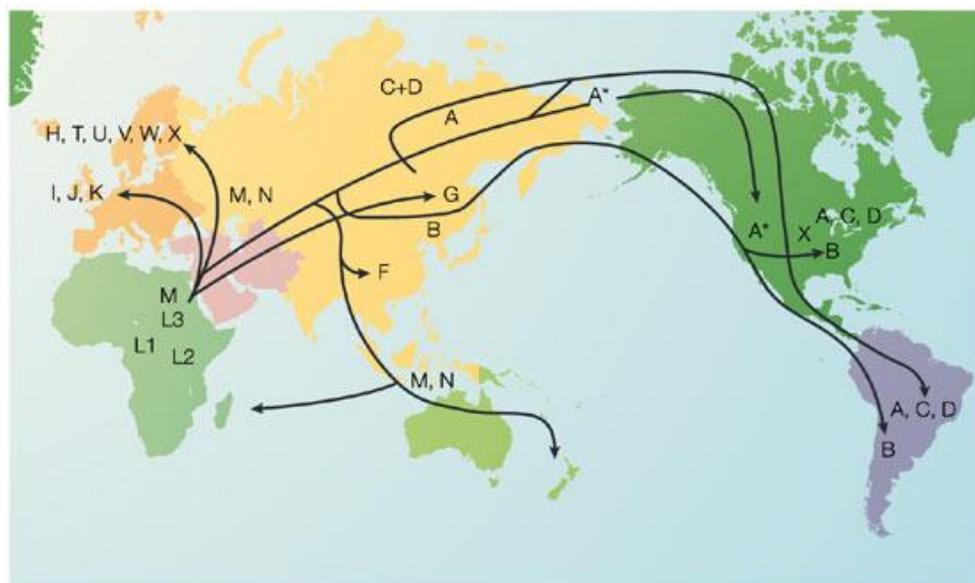
Figure 1.3 mtDNA population haplogroups shown on a map of the continents of the earth. This map indicates that *Homo sapiens* migration patterns correlate with the divergence of mtDNA sequence from the most ancestral African haplogroups L1, L2 and L3. Adapted from (Shriver and Kittles, 2004).

The influence of mtDNA sequence variants in human disease is complex, with several studies showing that mtDNA haplogroup status can perpetuate or protect against certain disease states. Mitochondrial haplogroups have been associated with complex

disorders such as obesity (Ebner *et al.*, 2015), pulmonary arterial hypertension (Farha *et al.*, 2016) and multiple sclerosis (Tranah *et al.*, 2015) amongst others. A recent large scale analysis of mtDNA sequences from patients with complex diseases confirmed previous observations that there were correlations between haplogroup status and increased or decreased risk of developing diseases including schizophrenia, multiple sclerosis and Parkinson's (Hudson *et al.*, 2014) although the mechanisms underlying mtDNA involvement *in vivo* remain unknown. Diseases caused by mtDNA mutations are also influenced by the mitochondrial haplogroup. For example, Lebers hereditary optic neuropathy (LHON) is caused by three known pathogenic mtDNA mutations (Johns *et al.*, 1992; Mackey and Howell, 1992; P *et al.*, 2003) however, the disease penetrance in carriers of these mutations varies widely. Haplogroups J and UK were found to be associated with increased risk of developing LHON (Brown *et al.*, 1997; Howell *et al.*, 2003; Carelli *et al.*, 2006; Hudson *et al.*, 2007; Gomez-Duran *et al.*, 2012). This is assumed to be influenced by two main factors; the metabolic profile of the mtDNA haplogroup itself and nuclear mitochondrial complementation and crosstalk (Gomez-Duran *et al.*, 2012; Giordano *et al.*, 2014). Recently, data has shown the emerging importance of compatibility between the nuclear and mitochondrial genomes for the success of reproduction in mouse and pig embryos (Park *et al.*, 2015; Ma *et al.*, 2016). Similarly, comparisons of transmitochondrial cybrid cell lines with identical nuclear backgrounds have highlighted the important differences in OXPHOS between haplogroups (Gomez-Duran *et al.*, 2010; Kenney *et al.*, 2014).

1.3 Mitochondria and disease

Mitochondria, and mtDNA, are of particular research interest for their involvement in human disease. Mitochondrial dysfunction and dysregulation underlies several human disorders affecting tissues with high energy demand such as brain, liver, heart and skeletal muscle. These conditions often present as complex multisystem phenotypes however, mitochondrial disorders are a heterogeneous category with some patients severely affected whilst others have very mild clinical presentations (for review see Lightowlers *et al.* (2015); Magner *et al.* (2015)). Given the intricate relationship of the nuclear and mitochondrial genomes within the cell, it is not surprising that mitochondrial pathologies can be the result of mutations in either the mtDNA or nuclear encoded mitochondrial genes. In recent years, mitochondrial dysfunction has

increasingly been implicated in several categories of human disease including immunological, metabolic and neurodegenerative (for review see Koopman *et al.* (2013)). Whether mitochondrial dysfunction and dysregulation are contributing to or result as a consequence of these disease states is still to be elucidated.

1.3.1 mtDNA mutations

Mutations in the mtDNA molecule itself can occur through two main events; point mutations or large-scale rearrangements such as deletions or insertions. Interestingly, very little is known about insertions within the mtDNA, most likely due to lack of clinical presentation (Poulton *et al.*, 1993; Krishnan and Birch-Machin, 2006). Point mutations in the mtDNA can present as several clinical phenotypes depending on the position and type of mutation. For example, there are three well characterised point mutations which cause LHON in 95% of cases at positions m.11778G>A, m.3460G>A and m.14484T>C (P *et al.*, 2003). Other examples include multiple mutations causing microencephalopathy lactic acidosis and stroke like episodes (MELAS), at positions m.3243A>G and m.3271T>C, and myoclonic epilepsy with ragged red fibres (MERRF) at position m.8344A>G (Pavlakis *et al.*, 1984; Goto *et al.*, 1990; Shoffner *et al.*, 1990). Several other mtDNA mutations are also associated with these clinical phenotypes, in genes encoding members of the OXPHOS subunits, rRNAs or tRNAs, giving rise to a spectrum of different phenotypes. Interestingly, due to the phenomenon of heteroplasmy, mtDNA mutation level must be sufficiently high to result in a biochemical defect (Schon *et al.*, 2012).

The transmission of mtDNA mutations is still not clearly understood. Remarkably, *de novo* point mutations are a frequent cause of mitochondrial disorders (~25%) but appear to have a low recurrence risk (Sallevelt *et al.*, 2016). One hypothesis as to how deleterious mtDNA mutations are removed from the germline is that mtDNA molecules undergo a phenomenon termed 'clonal expansion' after a genetic bottleneck effect during development of female primordial germ cells (PGCs) (Cao *et al.*, 2007; Cree *et al.*, 2008). This is hypothesised to occur after depletion of mtDNA during the early specification of PGCs resulting in a reduced population of molecules which then proceed to expand and repopulate the cell. Murine PGCs sampled during early development were found to have variable levels of mtDNA heteroplasmy whereas PGCs sampled later in development showed less variation in heteroplasmy levels, suggesting a highly

regulated selection mechanism during embryogenesis (Freyer *et al.*, 2012). In concordance with this, mtDNA copy number measurements in zebrafish oocytes revealed high variation in oocytes isolated from the same female (Otten *et al.*, 2016). Interestingly, comparisons between PGCs and somatic cells in developing zebrafish embryos showed that mtDNA copy number decreased in both cell types but that there were differences in the timing of mtDNA depletion. Nevertheless, mtDNA mutations do persist within the human population as evidenced by paediatric cases of mitochondrial disease (Ardissone *et al.*, 2014). Moreover, different mtDNA mutations have different rates of bottleneck segregation, as shown by *in silico* prediction models (Wilson *et al.*, 2016). Together, this suggests that both the timing of the mtDNA bottleneck during development as well as the type of mutation drives the differences observed in mature oocytes. This goes some way towards explaining the differences in mutation load and disease severity observed between offspring from the same mother.

1.3.2 mtDNA deletions

Deletions within the mtDNA can be of any size and occur at any position on the molecule, although these events mainly occur within the major arc. Patients can present with multiple mtDNA deletions in the cell or tissue, often due to mutations in genes essential for mtDNA replication and maintenance (Campbell *et al.*, 2014). Many patients harbour a single 4977bp deletion, known as the 'common' deletion, thought to occur from clonal expansion of one mutant molecule (Brierley *et al.*, 1998). This deletion is the underlying mtDNA rearrangement in sporadic mitochondrial disorders such as Kearns-Sayre syndrome (KSS), Pearson's syndrome and chronic progressive external ophthalmoplegia (CPEO), with heteroplasmy levels and deletion location differing between the conditions (Lopez-Gallardo *et al.*, 2009). The 'common' deletion heteroplasmy level is also known to increase with age and is used as a biomarker of mitochondrial dysfunction in several age-related conditions (for review see Krishnan *et al.* (2008)). More recently, interrogation of a large patient cohort has shown that heteroplasmy as well as deletion size and location within the mtDNA can be used as predictors of disease progression and severity (Grady *et al.*, 2014).

The 'common' deletion occurs between positions m.8455 and m.13446 on the mtDNA molecule, and is flanked by 13bp direct repeat sequences (Mita *et al.*, 1990; Samuels *et*

al., 2004). It is hypothesised that during mtDNA maintenance, these regions, which are complimentary to one another, align and allow the 4977bp region in-between to be excised resulting in two smaller mtDNA molecules (Figure 1.4) (Krishnan *et al.*, 2008). Although this hypothesis is probable, there is a lack of experimental evidence and isolating mtDNA to study such mechanisms has proved extremely difficult (Yamashita *et al.*, 2008; Sadikovic *et al.*, 2010; Tang *et al.*, 2013).
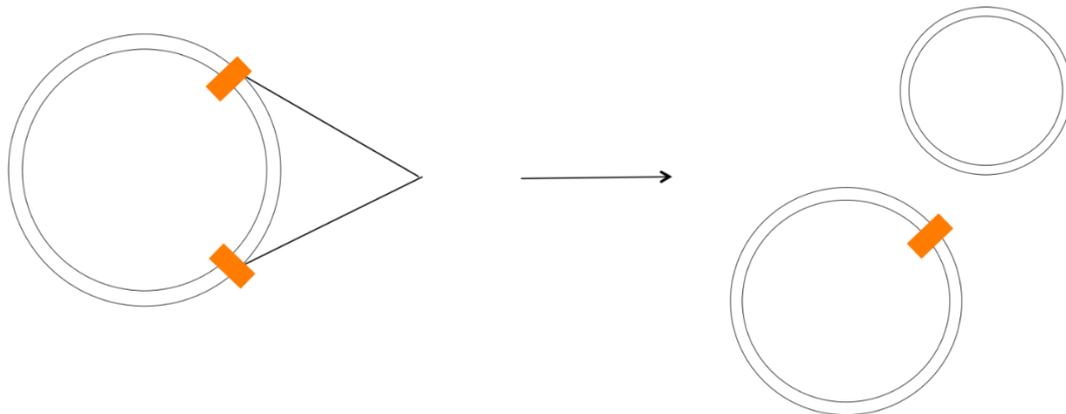
Figure 1.4 Schematic of mtDNA deletion formation. Direct repeats are depicted by orange boxes on the circular mtDNA molecule. Aberrant replication leads to formation of two smaller molecules shown on the right hand side.

Deletion formation is thought to be caused by a decline in the efficiency of mtDNA replication (Copeland and Longley, 2014). The underlying biochemical mechanisms thought to be responsible for this decline fall into three main categories; the error rate and activity of POLG, nucleotide metabolism and mitochondrial network dynamics. During mtDNA replication, spontaneous errors made by POLG can lead to both point mutations and deletions (Longley *et al.*, 1998; Zheng *et al.*, 2006). Additionally, POLG possesses exonuclease activity which has been shown to be disrupted when mutated, thus leading to an increased risk of mtDNA deletion formation (Zhang *et al.*, 2000). Discrimination between ribonucleic and deoxyribonucleic acids also relies upon the proofreading ability of POLG. In the strand coupled model of mtDNA replication, RNA/DNA hybrid molecules are utilised by POLG to create the daughter DNA molecule and therefore is reliant upon the ability for the replication machinery to distinguish between the two nucleic acid species (Yang *et al.*, 2002). Loss of this essential function may lead to mutation or deletion of the molecule. Lastly, mitochondrial dynamics plays

an important role in ensuring that deleterious mtDNA is removed from the organelle and cell, presumed to be through mitophagy (Nunnari and Suomalainen, 2012). This cellular process requires several key proteins and the aberrant activity of many of these has been associated with impaired mitophagy and therefore an accumulation of mutated mtDNA species (Chen *et al.*, 2010). Several other pathways have been implicated in the formation of deletions including inefficient DNA repair mechanisms and mutations in exonucleases, helicases and nucleases within the mitochondria (Copeland and Longley, 2014).

Deletions also occur in childhood suggesting that these molecules can persist through the germline genetic bottleneck described in Section 1.2.2 (Poulton *et al.*, 1991). Although this mechanism has not yet been proven, clinical data shows that the risk associated with transmitting mtDNA deletions from an affected mother to child is 1 in 24 births, suggesting that deletions can be transmitted but that deletion levels in oocytes are most likely variable (Chinnery *et al.*, 2004).

The mechanisms as to how clonal expansion of mtDNA deletions occur within tissues are still debated. One hypothesis is that smaller molecules have a replicative advantage due to their reduced size (Wallace, 1992). This theory is controversial and new evidence suggests that in human muscle fibres there is no clear bias for large or small scale deletion molecules with regards to clonal expansion rates (Campbell *et al.*, 2014). Recently, the involvement of the mitochondrial uncoupled protein response (UPR^{mt}) in deletion propagation has been demonstrated in *C.elegans* harbouring a large scale mtDNA deletion (Gitschlag *et al.*, 2016; Lin *et al.*, 2016). Persistence and stabilisation of this mtDNA deletion under cellular stress conditions suggests that the normal mitochondrial degradation response, mitophagy, is dampened to allow propagation of the deleterious genome (Gitschlag *et al.*, 2016) possibly through the signalling pathway mediated by the activating transcription factor associated with stress (ATFS-1) (Lin *et al.*, 2016). Although germline transmission has not been fully explored in this model, it is an interesting observation that hijacking of cellular homeostatic responses can allow for expansion of such molecules.

1.3.3 Nuclear-mitochondrial genes

Over the last ~1.5 billion years there has been transfer of mtDNA into the nuclear genome of the cell most likely driven by the need to limit mutation accumulation caused in part by free radicals from the ETC. It is estimated that ~98% of the mitochondrial genome has now been transferred, as evidenced by the vast number of proteins which are nuclear encoded and imported into the mitochondria within mammalian cells (Wessels *et al.*, 2013). Nuclear DNA sequences with homology to the mtDNA molecule, known as NuMts, present a problem in extraction of mtDNA sequences from whole genome sequencing datasets as it is unclear whether such sequences map to the mtDNA or to the NuMt region (Gould *et al.*, 2015; Malik *et al.*, 2016). However, it is clear that NuMt sequences are a common phenomenon as they are present in almost every genome sequenced from animal and plant species (Soto-Calderon *et al.*, 2012; Dayama *et al.*, 2014). Increased genomic incorporation of mtDNA has been described in several cancer types as well as during the normal ageing process suggesting that DNA transfer events are either a trigger for or a consequence of cellular dysregulation (Ju *et al.*, 2014; Ju *et al.*, 2015).

It is widely acknowledged that mtDNA is utilised by the cell to repair double strand breaks (DSBs) within the nuclear genome (Ricchetti *et al.*, 1999). The precise molecular mechanism of such events has been debated with some studies suggesting that non-homologous end joining during cell division is the mechanism by which mtDNA fragments are incorporated into transcriptionally silent regions of the genome (Hazkani-Covo and Covo, 2008) whilst others suggest NuMt incorporation is preferentially in regions with open chromatin conformation, as shown in germline cells (Tsuji *et al.*, 2012). A recent study has demonstrated further evidence that NuMt integration is non-random and occurs mostly within repetitive DNA elements, such as centromeric regions (Doynova *et al.*, 2016). Regardless of the method of transfer, it is clear that mtDNA integration into the nuclear genome is prevalent and ongoing.

1.4 Mitochondria and methylation

1.4.1 mtDNA methylation

Modifications to the transcriptional control of DNA not by changes in the DNA sequence itself, termed epigenomics, has been well characterised as an additional layer of

regulation within the nuclear genome (Berger *et al.*, 2009). Primarily, epigenetic control is mediated by methylation and acetylation of histone protein complexes but can also occur through the presence of CpG islands. These C-G rich regions are common near the promoter region of genes and can be activated or silenced depending on methylation status and recruitment of protein complexes. Although the epigenetic landscape is widely accepted as a regulatory mechanism within the nucleus, the importance of DNA methylation and acetylation within the mtDNA remains controversial. Of course, mtDNA is not organised in the same way as nDNA around histones, the primary mechanism of global epigenetic regulation within the nucleus, or wrapped up into higher order chromatin. However, mtDNA is packaged by proteinaceous nucleoids which could offer a similar regulatory mechanism as the nuclear histone proteins (Bogenhagen, 2012). In addition to histone modification, direct methylation of DNA plays a key role in epigenetic regulation and could be influencing expression patterns of mtDNA.

Methylation of mtDNA was characterised as far back as the 1970's and 80's (Nass, 1973; Pollack *et al.*, 1984) but was not extensively investigated due to contradictory reports that there was no methylation of mtDNA (Dawid, 1974). Technical limitations also contribute to a lack of mtDNA methylation data as the mtDNA copy number varies depending on the cell type used and can be contaminated with NuMt sequences as described in Section 1.2.6. Levels of methylated cytosines (5mC) within the mtDNA have been investigated as these marks have been proposed to modify transcription of the molecule from the D-loop regulatory region (Bellizzi *et al.*, 2013). The presence of CpG tracts within the mtDNA have been mapped and show a similar rate of suppression to nDNA (Cardon *et al.*, 1994; Ghosh *et al.*, 2014). A comprehensive study using bisulphite pyrosequencing has shown that mtDNA CpG sites occur at ~2% of the mtDNA and can be mapped to regulatory regions in the D-loop, mitochondrially encoded 12S RNA (RNR1) and mitochondrial encoded 16S RNA (RNR2) as well as within genes encoding several OXPHOS proteins (Liu *et al.*, 2016). Recently, the DNA cytosine-5 methyltransferase1 (DNMT1) protein was shown to have mitochondrial localisation (Shock *et al.*, 2011) and has been proposed as the major regulator of mtDNA methylation. The presence of DNMT1 as well as characterisation of methylated cytosines by bisuphite sequencing suggests that mtDNA methylation may play an important regulatory role within the organelle.

1.4.2 Mitochondrial regulation of nuclear DNA methylation

Recently, studies have shown that metabolic pathways contained within the mitochondria directly influence nDNA methylation status. This coupling of metabolism to genome methylation allows for transcriptional regulation based on nutrient availability and sensing. One recent study showed that brain samples from patients with multiple sclerosis had lower altered histone H3 methylation marks compared to controls. This altered methylation pattern was linked to mitochondrial dysfunction via the methionine metabolism pathway (Singhal *et al.*, 2015).

Methylation of DNA is reliant upon cellular availability of the methyl group donor S-adenosylmethionine (SAM), synthesised from methionine (Figure 1.5). Methyltransferase enzymes utilise available SAM to carry out methylation of DNA, RNA and protein. Delivery of SAM to the mitochondrial matrix has also now been shown to occur through the carrier protein SLC25A26 (Kishita *et al.*, 2015). SAM availability is directly affected by 1-carbon folate metabolism, which was recently demonstrated to increase hypermethylation of 5mC in the mtDNA of oocytes isolated from patients with polycystic ovaries (Jia *et al.*, 2016). Defective mtDNA replication has also been shown to alter the 1-carbon folate dependent pathways of the cell including dNTP synthesis, SAM production and methylation (Nikkanen *et al.*, 2016).
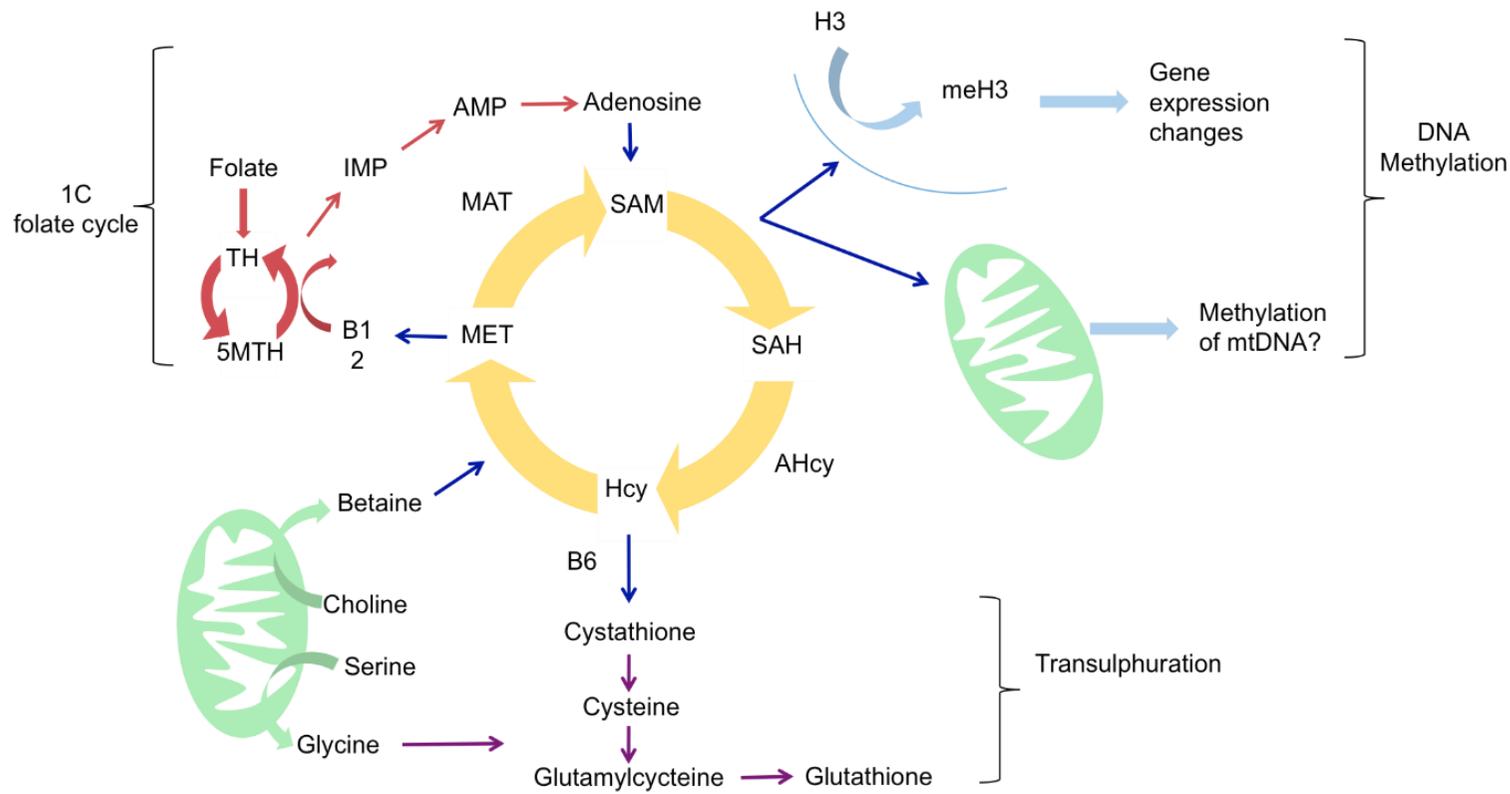
Figure 1.5 Metabolic pathways involved in DNA methylation. S-adenosylmethionine (SAM) is synthesised in the cytosol from adenosine and can be transported to the nucleus or mitochondria where it is used as a methyl group donor. Transulphuration and 1-carbon folate pathways are directly linked to SAM synthesis. Mitochondria metabolise choline within the inner matrix to produce betaine, which is utilised to convert homocysteine (Hcy) to methionine (MET).

In addition, mitochondrial content has been linked to nDNA methylation patterns in single cells further suggesting that cellular ATP production can drive alterations in transcription levels (Guantes *et al.*, 2015). Uncovering such regulation demonstrates the importance of tightly regulating the balance between substrate availability and transcriptional control within the cell. Additional proteins involved in these mechanisms remain to be elucidated, such as transcription factors and methyltransferases that mediate mitochondrial-nuclear methylation status.

1.5 PR-domain containing 9

1.5.1 The PRDM family and gametogenesis

For vertebrates to reproduce, primordial germ cells (PGCs) must undergo both specification and meiosis to produce haploid gametes. This process is controlled by several transcriptional activators and repressors which alter the global gene expression profile via epigenetic reprogramming. In mammals, PGCs are derived from the mesoderm lineage and are specified by repression of the somatic program and re-activation of pluripotency potential (Fuhrmann *et al.*, 2001). One group of proteins required during germ cell development is the PR-domain containing (PRDM) family. In primates, there are 17 PRDM family members, all containing the characteristic N-terminal PR domain, which is highly similar to the SET domain (Huang *et al.*, 1998). This domain is associated with methyltransferase activity, although only four PRDMs have so far been shown to possess intrinsic methyltransferase activity (Kim *et al.*, 2003; Hayashi *et al.*, 2005; Ancelin *et al.*, 2006; Eom *et al.*, 2009; Blazer *et al.*, 2016). All PRDM family members (excluding PRDM11) have a variable number of zinc finger repeats, known to be responsible for protein-DNA binding. PRDM11, -4, -6, -7, -9, -10 and -15 also have zinc knuckle domains, most likely to allow protein-protein interactions. These proteins have diverse cellular functions from the differentiation of cardiac smooth muscle cells to tumour suppressor functions (for review see Fog *et al.* (2012)).

Pluripotent cells differentiate to PGCs by the activation of pathways controlled by expression of *PRDM1* (*BLIMP1*) and *PRDM14*. In *PRDM1* knockout mice, PGCs do not repress mesodermal homeobox genes and fail to migrate to the genital ridge, remaining clustered as PGC-like cells (Ohinata *et al.*, 2005; Vincent *et al.*, 2005). Recently, PRDM1 protein expression has been harnessed to experimentally derive germ cells from human

embryonic stem cells (hESCs) (Lin *et al.*, 2014; Sasaki *et al.*, 2015). The PGCs of *PRDM14* knockout mice do not reactivate pluripotency and also fail to repress histone 3 lysine 9 di-methylation (H3K9me2) (Yamaji *et al.*, 2008). The functions of both of these proteins are critical, independent of each other, for the correct specification of PGCs. Remarkably, overexpression of PRDM1 and -14 in the presence of transcription factor AP-2 gamma (AP2γ) allowed pre-cursor germ cells to specify as PGCs in the absence of cytokines (Magnusdottir *et al.*, 2013) suggesting that activation of these proteins is enough to prompt differentiation, overriding the cell cycle. More recently, the core pluripotency marker Nanog homeobox (NANOG) has been shown to induce differentiation of ESCs in culture to epiblast-like cells with PGC-like fate through binding and activation of PRDM1 and PRDM14 (Murakami *et al.*, 2016).

The timing of expression of PRDM1 and -14 are known to occur in early PGCs to prompt specification of these cells (Ohinata *et al.*, 2005). After specification, PGCs must undergo meiotic recombination regulated by the *PRDM9* (*MEISETZ*) gene. *PRDM9* was first described as a gene required for hybrid sterility in mouse (Forejt *et al.*, 1991). Scientific interest in this protein increased as the PRDM9 protein was found to have methyltransferase activity, driving global epigenetic changes in the cell and prompting progression of meiotic prophase I (Hayashi *et al.*, 2005). It is important to note that the development of gonadal tissue and PGCs differs between males and females. Female PGCs arrest during prophase I and remain in this state until female puberty is reached. Male PGCs arrest after mitosis and meiosis is only resumed in spermatocytes following puberty. Whilst specification of PGCs and migration of these cells to the genital ridge is the same for both sexes, PRDM9 expression is assumed to differ between the sexes based on timing of meiosis (Figure 1.6).
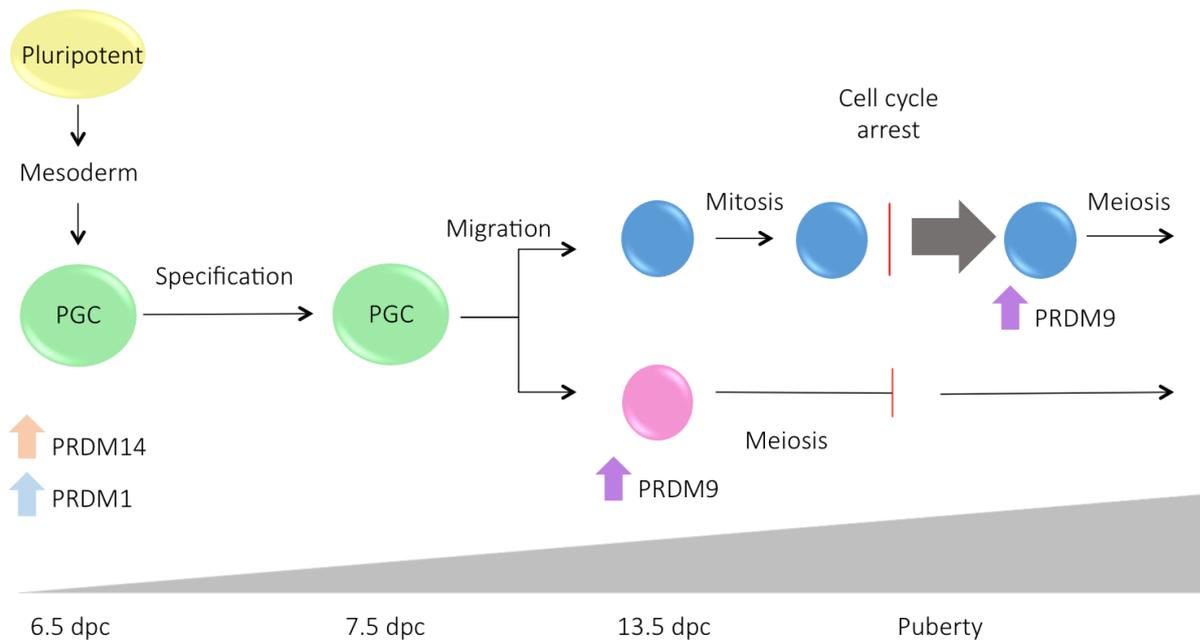
Figure 1.6 Specification of primordial germ cells (PGCs) during human development from pluripotent stem cell to germ cell. Male germ cells are shown in blue and female in pink. A developmental timeline is given along the bottom of the figure in days post conception (dpc).

1.5.2 PRDM9 and meiosis

Meiotic recombination occurs via chromatin remodelling, subsequent DSB formation and resolution by homologous recombination of sister chromatids (McClintock, 1939). Human recombination takes place in regions close to genes but not within transcribed regions, termed 'hotspots', which have been characterised as more likely to contain specific genomic factors such as the retroviral-like retrotransposons *THE1A* and *THE1B*, as well as CT-rich and GA-rich repeats when compared to recombination 'coldspots'. Hotspots are present across the genome at a density of around one hotspot per 50 kb (Myers *et al.*, 2005).

Crossover events during meiosis are critical for sexually reproducing organisms to exchange genetic information; in fact, without recombination during meiosis the resulting offspring are in most cases sterile (Dobzhansky, 1936). *PRDM9* knockout mice are only able to reproduce as heterozygotes because homozygous null mutants of both sexes are completely sterile (Mihola *et al.*, 2009), attributed to a loss of histone 3 lysine 4 trimethylation (H3K4me3) that leads to meiotic arrest of gametocytes (Hayashi *et al.*, 2005). Interestingly, inter-species hybrid mice show male sterility, and semi-sterility, that can be rescued by adding copies of *PRDM9* to the appropriate genetic background

(Flachs *et al.*, 2012). Genetic variation in *PRDM9* between mouse strains crossed was undetectable by sequencing analysis and is therefore not solely responsible for hybrid sterility (Flachs *et al.*, 2014). This dosage dependent effect on fertility suggests that *PRDM9* not only plays an important role in gametogenesis but must also be interacting with other hybrid sterility genes. Hybrid sterility gene interactions are still not fully understood, however, a recent study using mouse interspecies crossings has identified one such modifier locus, hybrid sterility X chromosome 2 (Hstx2/Meir1) (Balcova *et al.*, 2016) and it is likely there are several genomic loci harbouring important modifiers which have subtle effects not yet identifiable by mapping and linkage analysis.

1.5.3 PRDM9 protein function

The PRDM9 protein is responsible for determining where a large proportion of chromosome crossover events occur within the mouse (Baudat *et al.*, 2010), human (Sarbajna *et al.*, 2012) and chimpanzee (Auton *et al.*, 2012) genomes. PRDM9 functions as a methyltransferase by binding to DNA through its tandem-repeat zinc finger (ZnF) domain (Brick *et al.*, 2012). The protein consists of several domains; an N-terminal Kruppel-associated box (KRAB), an SSX repression domain (SSXRD), a zinc-knuckle and zinc-finger surround the PR/SET domain with a zinc-finger DNA binding array in the C-terminus (Figure 1.7).
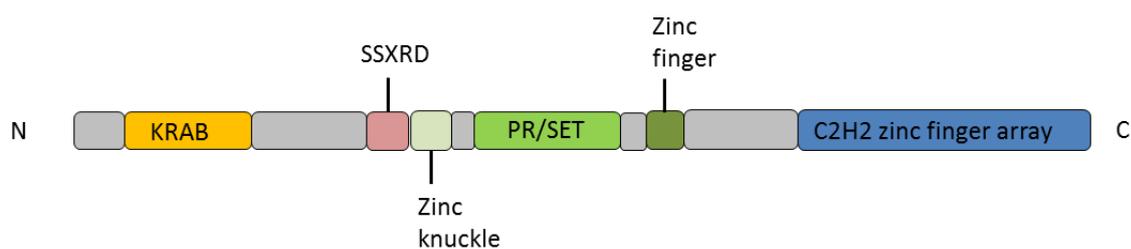
Figure 1.7 PRDM9 protein structure. Important regulatory and binding regions are shown; KRAB, SSXRD, PR/SET, zinc finger and knuckle and the C2H2 ZnF array domains.

Proteins containing KRAB and C2H2 zinc finger domains make up a subset of transcription factors with the ability to repress genomic targets through binding of specific DNA motifs. PRDM9 recognises the so called 'Myers motif' (*CCNCCNTNNCCNC*), a

13bp degenerate motif estimated to be present at ∼40% of human recombination hotspots (Myers *et al.*, 2008). Initial bioinformatic prediction analysis identified PRDM9 as a candidate binding protein for the motif (Myers *et al.*, 2010). Therefore, PRDM9 functions as a trans-acting factor; in other words it influences the stability of DNA molecules other than the one that encodes it. Purification of a partial ZnF region of human PRDM9 showed that the ZnF binds within the major groove of the DNA helix and forms hydrogen bonds (Patel *et al.*, 2016). In addition, analysis of mouse PRDM9 showed that ZnF binding affinity for DNA sequence is dependent on the position of the ZnF within the array; ZnF's at the beginning and end of the array have variable binding intensities whereas internal ZnF's have selective binding affinity, depicted in Figure 1.8 (Billings *et al.*, 2013). This characteristic allows PRDM9 to bind a number of DNA motifs which differ subtly.
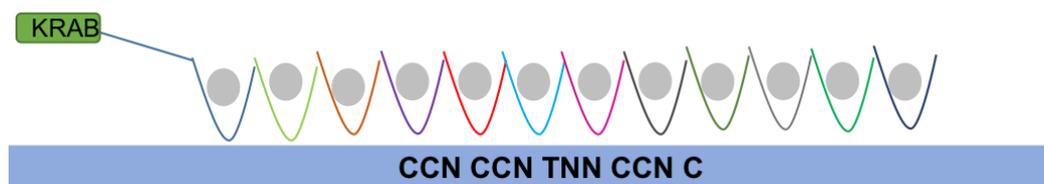
Figure 1.8 Schematic of PRDM9 KRAB (orange box) mediated ZnF binding to DNA motifs. Each coloured curve represents an individual ZnF in the PRDM9 array. Grey circles represent zinc ions held in pockets to create the 'finger' structure. Each finger contacts a codon of DNA bases.

X-ray crystallography has revealed the native protein conformation of both human and mouse PRDM9 (Wu *et al.*, 2013). PRDM9 binds the H3K4me2 peptide substrate in a cleft similar to other SET proteins and in the absence of a substrate, PRDM9 exists in an autoinhibited state, whereby the binding sites are closed off by the conformation of the protein. Binding of substrate peptides along with a cofactor induces a conformational change in the proteins pre- and post-SET domains. This is carried out by a hinge mechanism allowing for complete rearrangement of the binding sites. Electron density revealed that PRDM9 binds H3K4me2 through interactions with the backbone of the β-sheet and α-helix in the cleft as well as side chain interactions, possibly explaining substrate specificity. This suggests that PRDM9 inhibition is tightly controlled by other factors or post-translational modifications (Wu *et al.*, 2013). During meiosis, PRDM9 protein functions by firstly binding the 13bp DNA motif through its ZnF array, followed

by methylation of lysine residues on histone H3 protein tails and finally allowing double strand break complexes to bind and prompt chromosome crossover events (Figure 1.9) (Grey *et al.*, 2011; Brick *et al.*, 2012). New analysis of meiotic hotspot data suggests that this binding occurs in 'valleys' between the histone methylation signals created by PRDM9, although precise binding sites are difficult to ascertain due to variation between generations in any given pedigree (Lange *et al.*, 2016).
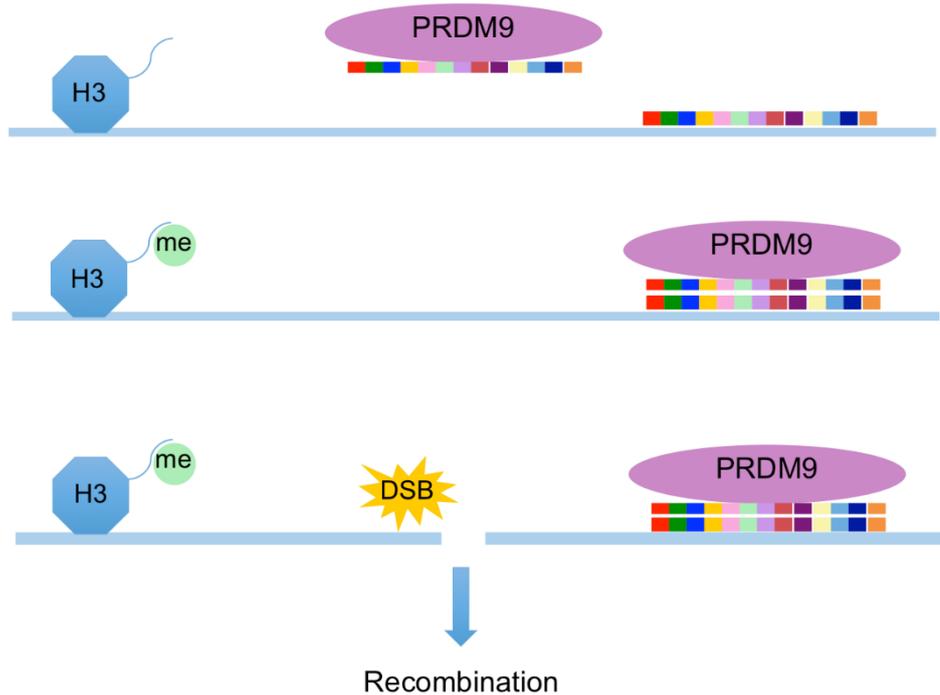
Figure 1.9 Schematic of the role of PRDM9 during meiosis. PRDM9 protein binds DNA motifs through the zinc finger array. It also methylates histone H3 residues to allow for chromatin remodelling. Double strand break machinery can then recognise and bind to the DNA sites directed by PRDM9 and allow for meiotic recombination to occur.

The extent of PRDM9 methyltransferase activity has now been extended to histone H3 lysine residues other than H3K4. Crystal structure analysis suggested that PRDM9 might also methylate H3K36 and H3K9 (Wu *et al.*, 2013). Analysis of enzyme kinetics has revealed that human PRDM9 is active in the presence of un-, mono- and di-methylated H3K4 and H3K36 peptide substrates, and is in fact, one of the most active histone methyltransferases (Eram *et al.*, 2014). In contrast, PRDM9 does not react with trimethylated K4 or K36 confirming that it specifically methylates these residues on H3 peptides. This opens up a new category of PRDM9 methylation targets previously

unknown, as levels of both H3K4 and H3K36 trimethylation were increased upon overexpression of partial PRDM9 in HEK293 cells (Eram *et al.*, 2014). This was also later confirmed *in vivo* using mouse spermatocytes (Powers *et al.*, 2016). The ability for PRDM9 to bind to several genomic targets through ZnF variability and target H3K4, H3K9 and H3K36 residues for trimethylation suggests that PRDM9 may be regulating a much higher number of recombination hotspots than previously thought. The full extent of the role of PRDM9 in meiosis is therefore still largely unknown, however, mutations within the SET domains or ZnF domains affecting protein function could lead to activation or inactivation of PRDM9 at a very critical point in germ cell development.

1.5.4 PRDM9 gene function

The human *PRDM9* gene (NC_000005.10) is located on chromosome 5 and encodes three differentially spliced transcripts; 001, 003 and 004. Only one of the transcripts is known to encode a protein product, transcript 001, which forms the 894aa full length PRDM9 protein (Figure 1.10). It is not yet known if the other two transcripts play a functional role within the cell.
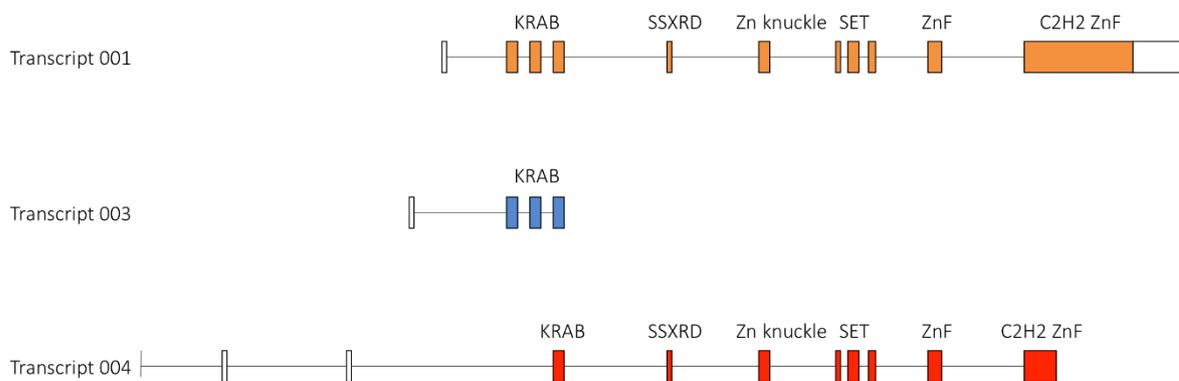
Figure 1.10 Schematic of the three transcripts produced by the PRDM9 gene. Transcript 001 is the only known protein coding transcript, containing all of the functional elements present in the 894aa full length PRDM9 protein.

The region responsible for DNA binding activity, the zinc finger array, is encoded by a minisatellite located in the final exon of the *PRDM9* gene (Hayashi *et al.*, 2005). In humans, *PRDM9* alleles have between 8 and 18 ZnF repeats within the array (Berg *et al.*, 2010). Each repeat is 83bp in length and due to the highly polymorphic nature of this

region, repeats can differ at the nucleotide level creating 29 distinct *PRDM9* alleles within the population (Figure 1.11). Each allele is comprised of distinct repeat combinations (e.g. ABCDD…) based on the number of repeats within the ZnF array region. Allele 'A' (ABCDDECFGHFIJ) is the most common allele, present at a high frequency in European (86%) and African (50%) populations (Berg *et al.*, 2010).

Allele	No. of repeats	Structure	Eu freq.	Af freq.
A	13	ABCDDECFGHFIJ	0.857	0.504
B	13	ABCDDCCFGHFIJ	0.019	0.034
C	14	ABCDDCCFKHLHIJ	0.01	0.128
D	14	ABCDDECFKGHFIJ	0.01	0
E	8	ABCDHFIJ	0.019	0
L1	9	ABCDGHFIJ	0.003	0
L2	12	ABCDDCFGHFIJ	0.003	0
L3	12	ABCDECFGHFIJ	0.003	0
L8	15	ABCDDECCFHLHFIJ	0.003	0
L9	13	ABCDDECFGPFQJ	0.013	0
L10	12	ACDDECFGHFIJ	0.003	0
L20	13	ABCDDECFGKFQJ	0.044	0
L24	13	ABCDDECFTPFQJ	0.013	0
L4	18	ABCDDCCCDDCFKHLHIJ	0	0.014
L5	10	ABCDDECFIJ	0	0.007
L6	15	ABCDDCCCFKHLOIJ	0	0.061
L7	12	ABCDDCCFGHFJ	0	0.02
L11	12	ABCDDECFGHIJ	0	0.027
L12	13	ABCDDCCCFGHFJ	0	0.008
L13	13	ABCDDECFGHQIJ	0	0.008
L14	14	ABCDDCCFKHLOIJ	0	0.047
L15	13	ABCDDCCFKHLHI	0	0.025
L16	13	ABCDRCFKHLHIJ	0	0.025
L17	15	ABCDDKCCFKHLHIJ	0	0.007
L18	13	ABDDSCFKHLOIJ	0	0.008
L19	15	ABCDDCCFKHLHQIJ	0	0.014
L21	14	ABCDDEECFGHFIJ	0	0.014
L22	14	ABCDDECFGHHFIJ	0	0.014
L23	13	ABCDDECFFHFIJ	0	0.008

Figure 1.11 *PRDM9* alleles found identified in European and African populations. Allele structure is shown along with repeat length and frequency within the population. Binding motif for each allele is also shown (Berg *et al.*, 2010).

The effect of different *PRDM9* alleles on global human recombination hotspot patterns is complex. Variation in *PRDM9* is rapid and contributes to the differences seen in global genome recombination patterns over the course of human evolution (Lesecque *et al.*,

2014) as well as between human and chimpanzee (Myers *et al.*, 2008; Groeneveld *et al.*, 2012). Analysis of sperm DNA has shown that men carrying the common 13 repeat length allele 'A' have more crossovers at hotspots containing the Myers motif; whereas men with non-'A' variant *PRDM9* alleles have changes in hotspot usage (Berg *et al.*, 2010). This suggests two things; firstly, that allele A has higher binding affinity for the motif and secondly, some *PRDM9* alleles are able to recognise and bind to variant hotspot motifs. For example, the *PRDM9* allele 'C' has 14 ZnF repeats and recognises a longer motif, resulting in a shift in recombination pattern for individuals carrying this allele (Berg *et al.*, 2011; Hinch *et al.*, 2011). In mice, inter-species crosses have demonstrated that different *PRDM9* alleles influence the erosion of recombination hotspots over evolutionary time due to the different binding properties of each allele (Baker *et al.*, 2015; Smagulova *et al.*, 2016). Interestingly, the *PRDM9* allele was successfully 'humanised' in C57BL/6 mice, revealing that altering one allele affects crossovers controlled by the other allele (Davies *et al.*, 2016). Such modifications were also linked to the fertility of the mice suggesting that *PRDM9* allele interactions themselves are directly influencing hybrid sterility, further supporting the role of this protein in mammalian speciation events. This study also demonstrated that PRDM9 protein species encoded by different *PRDM9* alleles differed in their binding affinity for the Myers motif supporting the idea that there might be dominance over hotspot location based on the *PRDM9* alleles present in the genome, previously suggested to also be the case in the human genome (Thomas *et al.*, 2009; Jeffreys *et al.*, 2013). The proposed mechanism for this hotspot erosion is shown in Figure 1.12.
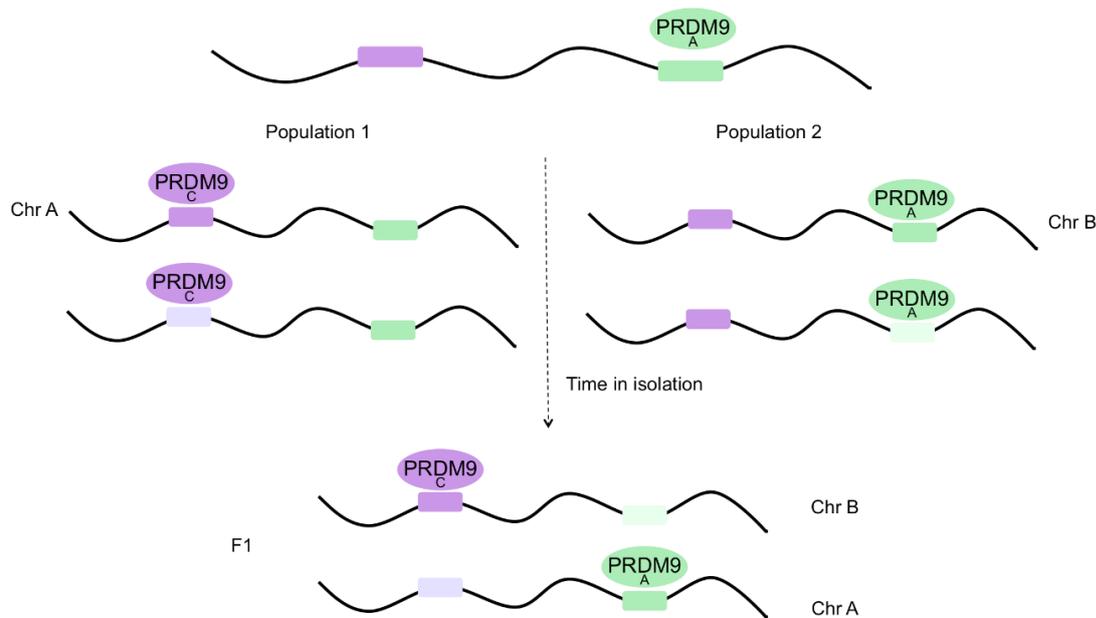
Figure 1.12 Schematic of the proposed mechanism by which PRDM9 influences hotspot erosion in mouse and human genomes. As two populations become isolated, a new *PRDM9* allele is created in one (PRDM9^C) whilst the other still maintains the ancestral allele (PRDM9^A). The two alleles have different binding motifs within the genome. Over time, the allele begins to erode the hotspot on the 'self' chromosomes, leaving other binding sites in-tact. Thus, when the two populations interbreed (F1) the *PRDM9* alleles will still bind to their corresponding motif that has not been eroded. Asymmetric hotspots result in hybrid sterility, therefore driving the creation of new *PRDM9* alleles and new motifs within the population to regain symmetry of recombination (Zelazowski and Cole, 2016).

Moreover, sperm genotyping indicates that many recombined and lengthened *PRDM9* alleles exist within sperm cells that are not represented in somatic cells within the human population (Jeffreys *et al.*, 2013). Lack of these alleles within the population suggests a unique mechanism whereby *PRDM9* influences its own stability at the meiotic level leading to a phenomenon where diversity is driven by the most stable alleles. Perhaps this is a mechanism to limit the shifting of recombination patterns and combat deleterious effects (Boulton *et al.*, 1997). The evolutionary pressures upon *PRDM9* remain to be fully understood and could be due to several reasons; a high mutation rate due to minisatellite instability, purifying or positive selection controlling DNA binding specificity, the erosion of binding sites as described in Figure 1.12, or epistatic interactions which cause incompatibility after population divergence (Buard *et al.*, 2014).

Additionally, there is an argument that other proteins act in a functionally redundant manner with respect to PRDM9 driven recombination. The entire canid lineage lacks active PRDM9 (Munoz-Fuentes *et al.*, 2011; Axelsson *et al.*, 2012; Campbell *et al.*, 2016),

whilst in equids no association between *PRDM9* genetic variation and hybrid sterility was found (Steiner and Ryder, 2013). Recently, there has been a reported case in humans of an individual who lacks PRDM9 expression due to a homozygous mutation rendering both alleles non-functional (Narasimhan *et al.*, 2016). This individual has no clinical phenotype and has two daughters, suggesting that in this case PRDM9 activity was not required for successful meiosis. Tissue biopsies or cultured cells from this individual would help aid functional investigations. These observations suggest that although PRDM9 might be the master regulator of meiotic recombination in some species, it is not required in others. In addition, other as yet unidentified proteins must be facilitating crossover events in the absence of PRDM9. Nevertheless, PRDM9 clearly plays an important functional role during human meiosis.

1.6 PRDM9 and disease

1.6.1 PRDM9 and genomic rearrangement disorders

In recent years, the role of PRDM9 in human disease has become an area of great interest. Firstly, mutations in the *PRDM9* gene have been associated with infertility characterised by nonobstructive azoospermia, which is perhaps not surprising as *PRDM9* knockout mice are sterile (Miyamoto *et al.*, 2008; Irie *et al.*, 2009; He *et al.*, 2013). A mutagenesis screen in mice also showed that *PRDM9* gene mutations caused a severe depletion of germ cells, suggesting that human mutations might have a similar effect (Weiss *et al.*, 2012). However, due to the polymorphic nature of the *PRDM9* gene it is challenging to prove the true pathogenic nature of these SNPs. It is more likely that *PRDM9* influences fertility by the hybrid paradox outlined in Section 1.5.4 due to incompatibility of different *PRDM9* alleles during recombination (Flachs *et al.*, 2014). Further studies into the effects of *PRDM9* on human fertility are required, especially in female oocytes. However, such samples are difficult to obtain and would therefore most appropriately be tested in an animal model such as mouse or zebrafish.

Genomic rearrangement disorders are characterised as diseases resulting from deletions, insertions and copy number changes within the genome. Sequencing of the ZnF region of the *PRDM9* gene in patients with genomic rearrangement disorders and their parents showed an association of ZnF variants with Williams-Beuren disease, caused by microdeletions in a region of chromosome 7 (Borel *et al.*, 2012). PRDM9 has

also been implicated in complex duplication-triplication rearrangements of the Duchenne muscular dystrophy (DMD) gene (Ishmukhametova *et al.*, 2013). This event was studied in a male patient with DMD, where it was shown that inverted repeats allow rearrangement of regions spanning several exons of the *DMD* gene; leading to large scale exon shuffling and resulting in a disease phenotype. The regions flanking the rearrangement break point were visualised and the PRDM9 binding motif was found ~300 bp upstream. This was not in a region known to be a meiotic recombination hotspot, however, Ishmukhametova *et al* hypothesised that LD-based hotspots are only representative of hotspot activity in female meiosis, leading them to believe that this might be a male meiosis specific event. *PRDM9* recruitment of initiator of meiotic double strand breaks (SPO11) could initiate the strand break needed to facilitate incorrect strand invasion and therefore a duplication-triplication event in *DMD*. However, the authors did not sequence *PRDM9* in the patient or parents to determine if rare *PRDM9* alleles were present in the family and further analysis needs to be carried out to confirm that this putative binding site is actually bound by PRDM9 and if this is indeed the causative molecular mechanism.

Disorders influenced by epigenomic changes such as schizophrenia have also been associated with PRDM9 function (Wockner *et al.*, 2015). Whilst this is an intriguing hypothesis, the effects of PRDM9 on epigenetic methylation patterns is not clear. As discussed, in order for PRDM9 to bind to DNA motifs it methylates histone H3 lysine residues, inducing chromatin unwinding and DNA exposure. Whilst this process is important for meiotic recombination during development of germ cells, there is not yet any evidence to support a functional link between PRDM9 methyltransferase activity and global epigenetic regulation.

1.6.2 PRDM9 as a cancer testes antigen

The SSX region in the PRDM9 N-terminus along with the KRAB domain indicates that PRDM9 has homology to the SSX class of proteins, which are suggested to have testis specific functions, although this remains to be explored along with the full functional role of PRDM9. In recent years, the role of a class of biomarkers known as cancer testes (CT) antigens has emerged and is of particular interest in the field of cancer immunotherapy. These CT antigens are defined as proteins expressed specifically in the

testes of adult males but not in other somatic tissues. Interestingly, these proteins have been found to be expressed in tumours from various tissues and are therefore biomarkers of tumourigenesis with the potential to be targeted through antibody based immunotherapy strategies (for review see Suzuki *et al.* (2013)). PRDM9 has been identified as a CT antigen as its expression was detected in several cancer cell lines and tissues by RT-PCR (Feichtinger *et al.*, 2012). Under normal circumstances, PRDM9 is only expressed in testes and at very low levels in ovaries. This suggests that PRDM9 might be upregulated in cells to induce chromatin remodelling and genomic rearrangements, promoting a cancer phenotype.

Cancer genomics has largely focussed on identifying genomic mutations through linkage association analysis. Paediatric cancers are an interesting group as mutations causing these disorders would be assumed to be lost from the population quickly. However, emerging evidence suggests that parental alleles could be risk factors for childhood cancers even if the affected individual themselves does not inherit the mutated allele. For example, in around 25-30% of B-cell lymphoblastic leukaemias there is a non-random gain of chromosomes leading to hyperploid cells (for review see Paulsson *et al.* (2009)). Rare *PRDM9* alleles were found to be enriched in a group of parents with children diagnosed with B-cell precursor acute lymphoblastic leukaemia (B-ALL) compared to ethnically matched controls (Hussin *et al.*, 2013). Mothers of B-ALL children were significantly more likely to carry the rare *PRDM9* allele than fathers. Transmission disequilibrium may be a factor in B-ALL cases, with only half of all affected children carrying the rare *PRDM9* allele. More recently, *PRDM9* alleles were investigated in a parental cohort of patients with high hyperploidy acute lymphoblastic leukaemia (HeH ALL) (Woodward *et al.*, 2014). Again, rare *PRDM9* alleles were overrepresented in both the parental and patient cohort when compared to controls, suggesting that PRDM9 activity may cause genomic rearrangements at the chromosomal level in these patients. However, it is unclear when and in which cell types this happens. Perhaps rare *PRDM9* alleles are modifying the activity of PRDM9 protein in the somatic cells to cause rearrangements. The transmission of the 'risk' allele to only ~60% of offspring suggests that the effect of *PRDM9* is actually happening at the gamete stage, when the cells are somehow predisposed to developing ALL.

1.6.3 PRDM9 and mtDNA deletions: the missing link?

PRDM9 recognises and binds a 13 bp motif enriched in human hotspots. Surprisingly, in the bioinformatics screen used to identify this PRDM9 recognition motif, it was identified within the human mtDNA sequence (Myers *et al.*, 2008). The motif was suggested by the authors to flank the large scale pathogenic 4977bp 'common deletion'. Although the authors do not suggest any possible functional role of this motif site, there remains the possibility that PRDM9 is localising to mitochondria and binding mtDNA. As PRDM9 is expressed solely during meiosis, it could be involved in regulating deletion heteroplasmy levels in oocytes and potentially predispose individuals to mtDNA deletion disorders.

Deletions in mtDNA are often flanked by direct repeat regions (Larsson and Holme, 1992; Samuels *et al.*, 2004) leading to the hypothesis that these repeats align during replication and allow deletions to occur within the molecule. The use of inverted repeats during recombination repair is not new and has been described within the nuclear genome, in disorders with duplication-triplication gene events (Shimojima *et al.*, 2012; Beri *et al.*, 2013; Dittwald *et al.*, 2013). Recent evidence that *PRDM9* alleles are associated with rearrangement disorders raises the possibility that such mechanisms could also be creating mtDNA rearrangements. Mutations in genes known to be involved in mtDNA maintenance and replication cause mtDNA deletion via aberrant repair however, these known mutations only partially explain the prevalence of mtDNA deletions within the population. It is therefore important to try and elucidate factors which predispose the repeat regions of the mtDNA to undergo such deletions, as the mechanism cannot solely hinge on defects in the already characterised mitochondrial replication machinery.

Any possible involvement of PRDM9 in mtDNA maintenance must be proven by directly assessing the mitochondrial localisation of this protein and it's ability to bind mtDNA.

Chapter 2 Aims and Objectives

2.1 Hypothesis

Due to a previous report that PRDM9 binding motifs are present within the human mtDNA sequence, this thesis aims to identify whether this is the case at a population level and whether such motifs could be of functional significance within the mitochondria. More specifically the following aims will be tested;

1. PRDM9 motifs will be present in the majority of mtDNA sequences, at known sites of mtDNA deletions.
2. The PRDM9 protein will localise to the mitochondria, as this must be where any binding events occur.

2.2 Aims and Objectives

Two main approaches will be used to study this hypothesis. Firstly, in Chapters 4 and 5 the genetic aspects of PRDM9 will be explored in relation to mtDNA. More specifically, the aims of these chapters will be;

1. To determine whether PRDM9 binding sites exist within the mtDNA sequence of a large cohort of individuals from several populations.
2. Sequence the ZnF region of the *PRDM9* alleles in a cohort of single deletion patients compared to controls.

Secondly, in Chapters 6, 7 and 8 the localisation and function of PRDM9 protein will be explored. The aims of these three chapters are;

1. To identify an appropriate model to experimentally determine the function of the PRDM9 protein.
2. Modify the expression of PRDM9 in a cell culture system to understand how this protein is potentially interacting with mitochondria.
3. Create a cell culture model, using overexpression, to determine whether PRDM9 expression affects mitochondrial protein levels or mtDNA copy number.

Chapter 3 Materials and Methods

3.1 Sample cohort

Total DNA was extract from 48 patients, known to harbour a single mtDNA deletion, using a DNAeasy extraction Kit (Qiagen, Manchester, UK), according to manufacturers recommended protocol. DNA aliquots were kindly provided by Professor R.W. Taylor (Newcastle Mitochondrial NSCT Diagnostic Service). Further patient details are provided in Chapter 5 Table 5.1. In addition, control cohort (WTCCC 1958 Birth Cohort) total DNA was extracted from whole blood (using same method as above).

3.2 Polymerase chain reaction

Polymerase chain reaction (PCR) for amplification of the gene of interest was carried out using TaKaRa LA Taq, TaKaRa GXL Taq (TaKaRa Bio Europe/Clontech, France) or MyTaq (Bioline, London, UK) polymerases. Primer pairs (Integrated DNA Technologies, IA, USA) were designed using PrimerBLAST or from previous publications. Optimisation of annealing temperatures (by temperature gradient PCR) and DNA concentrations (by concentration titration) was carried out to determine optimal conditions for product amplification.

3.3 Sanger sequencing

Cycle sequencing reactions was carried out using the following reaction mix; 2 µL Big-Dye Terminator v3.1, 1.5 µL Big-Dye reaction buffer (both Applied Biosystems, Thermo Fisher Scientific, Loughborough, UK), 3.2 pmol product specific primer, 10 % betaine and 25 ng of PCR product. Reactions were made to volume (20 µL) using nanopure water. Thermocycling conditions were; 96 °C for 1 min, followed by 25 cycles of 96 °C for 10 sec, 50 °C for 5 sec, 60 °C for 4 min.

To remove excess dye labelled nucleotides, PCR products were cleaned by ethanol precipitation. Briefly, 3 M sodium acetate125 mM EDTA and 70 µL 100% ethanol were added to each reaction in a 96-well plate. The plate was spun at 2000 x rcf for 30 minutes and then inverted on tissue paper. The plate was centrifuged to 100 x rcf, 70 µL

of 70% ethanol was added and the plate spun at 1650 x rcf for 15 min. The supernatant was removed by inversion and the plate allowed to air dry in the dark. The pellets were resuspended in Hi-Di formamide (Applied Biosystems, Thermo Fisher Scientific, Loughborough, UK).

The Big-Dye suspension was then transferred to an Applied Biosystems 3130xl genetic analyser for analysis. Analysis of electropherogram data was performed using SeqScape software V2.6 (Thermo Fisher Scientific, Loughborough, UK) via alignment with the gene of interest reference sequence.

3.4 Cell Culture

3.4.1 Maintenance of HEK293 cell line

Human Embryonic Kidney cells (HEK293), kindly provided by Professor R.N. Lightowlers (Institute of Cell and Molecular Biosciences Newcastle University) were cultured in Dulbecco's modified eagle medium (DMEM) supplemented with 10% foetal bovine serum (FBS) and non-essential amino acids (all GIBCO, Thermo Fisher Scientific, Loughborough, UK). Cells were grown under normal culture conditions (37 °C in 5% CO_2) in T75 flasks. Media was changed every 48-72 hours.

HEK293 cells were harvested every 3-5 days and split in a 1 to 3 basis. Briefly, media was aspirated; cells were washed using 1X phosphate buffered solution (PBS, Oxiod, Thermo Fisher Scientific, Loughborough, UK) and trypsinised by incubation with PBS-EDTA (0.05%) at room temperature for 5 minutes. Fresh media was added and the sides of the flasks were sluiced to ensure all cells were suspended in the media. Cells were centrifuged at 1300 x rcf for 5 minutes and the pellet resuspended in fresh media. To split 1:3, the resuspended pellet in media was split evenly between three flasks. The cells were then incubated and inspected daily.

3.4.2 Freezing/thawing cells

Once obtained (as described in Section 3.4.1) cell pellets were resuspended in freezing media (DMEM supplemented with 20% FBS (GIBCO) and 10% DMSO) at $3-5 \times 10^6$ cells per cryovial and frozen using the freezing container Nalgene® Mr. Frosty (Sigma Aldrich, Dorset, UK) at -80 °C. In this system, the cooling rate is 1 °C/min promoting a successful

cryopreservation. After 24 hours the cells were transferred to liquid nitrogen for long-term storage.

To defrost cells from liquid nitrogen, cells were quickly resuspended in appropriate media, transferred into a falcon tube and centrifuged at 1300 x rcf for 5 minutes. The supernatant was discarded and the pellet resuspended in 5 mL of culture media and seeded.

3.4.3 Limited dilutions of cultured cells

To obtain single cell colonies, cells were harvested as described (Section 3.4.2), counted using a haemocytometer and resuspended in DMEM to a concentration of 2×10^4 cells/mL. In a 96 well plate, 4000 cells were added to well A1 and were sequentially diluted across the plate, giving rise to 1 cell per well in the most diluted regions of the plate (as shown in Figure 3.1). After 14 days of culture the cells were scored by the presence of a single colony by microscopy. On day 18 of culture the colonies were trypsinised and moved to a 24 well tissue culture dish for expansion of the colony.
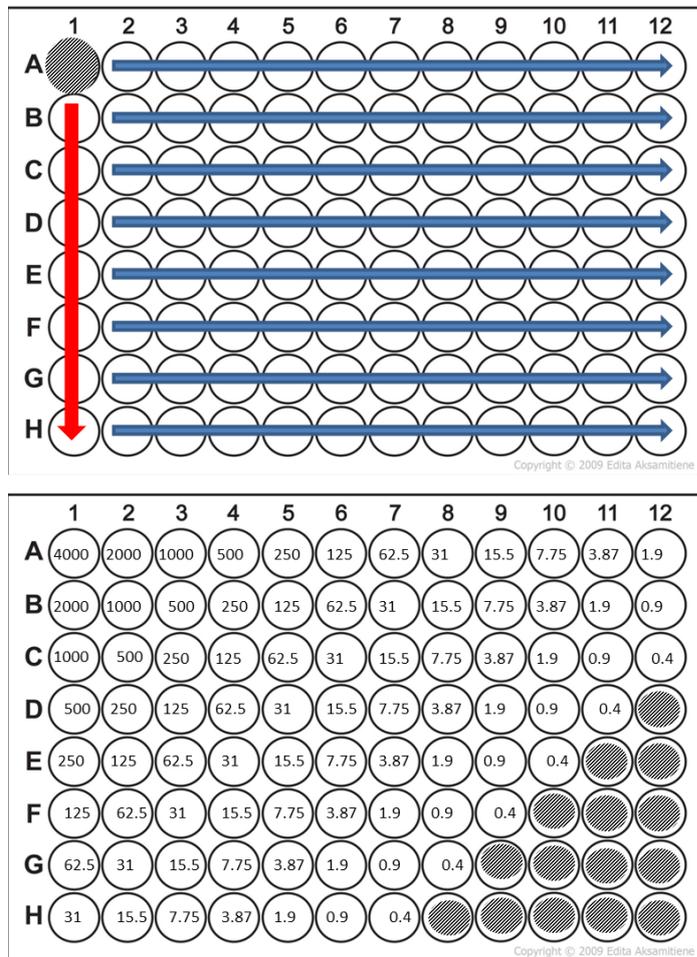
Figure 3.1 Dilution series of cells set up in a 96 well plate format. The top plate shows the direction of pipetting from the initial 4000 cell well (block grey). The bottom plate shows the corresponding cell number in each well. Block grey depicts wells where the cell number is >0.4 cell/well.

3.5 DNA and RNA extraction from cultured cells

Total DNA was extracted from cell culture pellets using a DNeasy Extraction Kit (Qiagen, Manchester, UK), according to manufacturers recommended protocol. Total RNA was extracted from cell culture pellets using the RNeasy extraction kit (Qiagen, Manchester, UK) according to the manufacturers recommended protocol. Cells were pelleted by centrifuging at 1300 rcf for 5 minutes and washed 3 times with PBS before DNA or RNA was extracted. DNA concentration was measured using the Nanodrop2000 UV-vis Spectrophotometer (Thermo Fisher Scientific, Loughborough, UK).

3.6 Protein Extraction

Different protein extractions were carried out during this project (see Sections 3.7.1 – 3.7.5).

3.6.1 Whole cell protein lysis

Cells were harvested as previously described (Section 3.4.1) and washed twice with cold PBS. On ice, the pellet was resuspended in an appropriate volume of Laemmli buffer (50 mM Tris pH7.5 (Sigma), 130 mM NaCl (BDH AnalaR), 2 mM $MgCl_2$ (BDH AnalaR), 0.1% Triton X-100 (Sigma), 1 tablet protease inhibitor (Roche, West Sussex, UK), dH_2O) or PathScan ELISA sandwich lysis buffer (Cell Sinaling, Boston, MA, USA) with protease inhibitor and vortexed for 30 seconds every 5 minutes for 15 minutes. The lysate was sonicated for 7 seconds on ice using a microtip, centrifuged at 2500 x rcf for 5 minutes at 4 °C. Supernatant was kept and the pellet discarded (Laemmli, 1970).

3.6.2 Tissue lysis

Tissue was cut and weighed on dry ice. The tissue sample was ground using a pestle and mortar on dry ice. To the ground tissue, 1 mL of lysis buffer with protease inhibitor per 100 mg of tissue was added and the mix transferred to a 2 mL Dounce homogeniser (Wheaton) on ice. Cells were disrupted with 10-20 strokes on the homogeniser drill at 4 °C. A small volume of lysate (~5 μL) was checked under a light microscope to ensure the cell membranes were disrupted. The lysate was sonicated for 7 seconds on ice using a microtip, centrifuged at 2500 x rcf for 5 minutes at 4 °C. Supernatant was kept and the pellet discarded (Laemmli, 1970).

3.6.3 Intact mitochondrial isolation

The cell pellet or tissue sample was suspended in 1 mL of Medium B; 250 mM Sucrose, 2 mM HEPES, 0.1 mM EGTA (Kirby *et al.*, 2007) and transferred into a 2 mL Dounce homogeniser (Wheaton) where the cells were disrupted with 20-60 strokes on the homogeniser drill at 4 °C. The sample was then transferred into a 1.5 mL micro-centrifuge tube and centrifuged for 10 minutes at 400 x rcf at 4 °C. The supernatant was removed and kept as the cytosolic fraction containing mitochondria. The pellet was resuspended in 800 μL Medium B and homogenised again to increase the amount of mitochondria harvested. The sample was centrifuged as before. The two supernatants were then combined and centrifuged for 10 minutes at 10,000 x rcf at 4 °C. The supernatant was removed and stored and comprises the cytosolic fraction. The final

pellet comprises the mitochondrial fraction (Fernandez-Vizarra *et al.*, 2010). All samples were stored at -80 °C.

To isolate mitochondria from the mitochondrial associated membrane (MAM) a percoll gradient method was used (Wieckowski *et al.*, 2009). The cell pellet was suspended in 30 mL PBS and centrifuged for 5 minutes at 600 x rcf at 4 °C. Pellet was resuspended in 5 mL PBS and centrifuged as before. Pellet was then suspended in 20 mL buffer $IB_{cells}1$ (225-mM mannitol, 75-mM sucrose, 0.1-mM EGTA and 30-mM Tris–HCl pH 7.4) and transferred to a 30 mL Dounce homogeniser using a 1 mL Pasteur pipette where cells were disrupted with 100 strokes using a drill at 4 °C. Cells were transferred back into a plastic tube and centrifuged as before. The pellet was stored as the nuclear fraction and the supernatant was centrifuged for 10 minutes at 7000 x rcf at 4 °C. The supernatant was stored as the cytosolic fraction and the pellet resuspended in 20 mL buffer $IB_{cells}2$ (225 mM mannitol, 75 mM sucrose and 30 mM Tris–HCl pH 7.4) before centrifugation as before. The supernatant was discarded and the mitochondrial pellet resuspended in 20 mL $IB_{cells}2$ before centrifugation for 10 minutes at 10,000 x rcf at 4 °C. The supernatant was discarded and the mitochondria suspended in 2 mL mitochondrial resuspension buffer (MRB; 250 mM mannitol, 5 mM HEPES (pH 7.4) and 0.5 mM EGTA). A small amount of this was stored as the 'crude' mitochondrial fraction. Next, 8 mL of percoll medium (225 mM mannitol, 25 mM HEPES (pH 7.4), 1 mM EGTA and 30% Percoll (vol/vol)) was placed in a 15 mL thin wall polyethylene ultracentrifuge tube. The mitochondrial suspension was layered on top of the percoll medium followed by 3.5 mL MRB gently layered on the top. This suspension was centrifuged for 30 minutes at 95,000 x rcf at 4 °C in a Sorvall ultracentrifuge (Thermo Fisher Scientific, Loughborough, UK). The mitochondria appear as a small pellet at the bottom of the tube whilst the MAM is a white band suspended in the layer above the pellet, this is collected by gentle pipetting using a 1 mL Pasteur pipette and diluted with 10 X the volume of MRB. The mitochondrial pellet is diluted in 10X the volume of MRB. Both suspensions were centrifuged for 10 minutes at 6,300 x rcf at 4 °C, the mitochondrial pellet suspended in 20 mL MRB and centrifuged again. This pellet was finally resuspended in 200 μL MRB and stored as the 'pure' mitochondrial fraction. MAM fraction was centrifuged for 1 hour at 100,000 x rcf at 4 °C and then suspended in a final volume of 200 μL MRB.

3.6.4 Mitochondrial enrichment

Native protein preparations are used to isolate multiprotein complexes from cells using a gentler cell lysis procedure than traditional SDS containing lysis buffers. After treatment with non-ionic detergents and centrifugation a final pellet is created which is enriched for mitochondria and can therefore provide a crude mitochondrial enrichment from a small number of cells.

5×10^6 HEK293 cells were harvested and the pellet washed three times with PBS. The pellet was resuspended in 100 µL PBS + protease inhibitor (Roche Diagnostics Limited, West Sussex, UK), 10 µL was diluted in 90 µL dH_2O and sonicated for 4 seconds. Protein concentration was determined as described in Section 3.7.5 by maxi Bradford. Digitonin is a mild non-ionic detergent which solubilises receptors and permeabilises the cell and nuclear membranes and was added to the non-sonicated cell suspension in a 1:6 ratio based on the formula below:

$$Y = Concentration\ of\ protein\ x\ 1.6\ (ratio)$$

$$Volume\ of\ digitonin = \frac{Y\ x\ Volume\ of\ PBS}{Concentration\ of\ digitonin}$$

Stock concentration of 6 mg/mL digitonin diluted in PBS was used. The digitonin treated lysate was made up to a final volume of 200µL with PBS and centrifuged at 10,000 x rcf for 10 minutes at 4 °C. The mitochondrial enriched pellet was resuspended in 30µL of MB2 buffer (500µL 3XGB (0.5 M aminocaproic acid, 50 mM Tris-HCL pH7), 1M aminocaproic acid, 20mM EDTA), 1% laurel malthoside added and incubated on ice for 15 minutes. The sample was centrifuged at 20,000 x rcf for 20 minutes at 4 °C and the supernatant was placed in a new 1.5 mL tube. Protein concentration was determined using the supernatant to perform a Bradford assay (Section 3.7.5).

3.6.5 Determining protein concentration

The Bradford Assay was used to determine the protein concentration (Bradford, 1976). Two methods were used depending on the absorbance being measured.

Maxi Bradford was performed as follows, a calibration curve was created using bovine serum albumin (BSA) at known concentrations of 0.05 µg, 0.1 µg, 0.2 µg, 0.3 µg, 0.4 µg

and 0.5 µg/µL solution. An example of this curve is shown in Figure 3.2. A blank was also set up to correct for any background absorbance in the buffer itself. Samples were run at different dilutions depending on the type of lysate e.g. cellular or tissue. In a 96 well plate, 10 µL of each calibration standard plus samples were measured in triplicate. To this, 200 µL of 1:5 dilution of Bradford solution (BioRad Laboratories, Hemel Hempsted, UK) was added. Detection was carried out on a plate reader. Analysis was carried out using Microsoft Office Excel package to determine standard curve and final protein concentration.
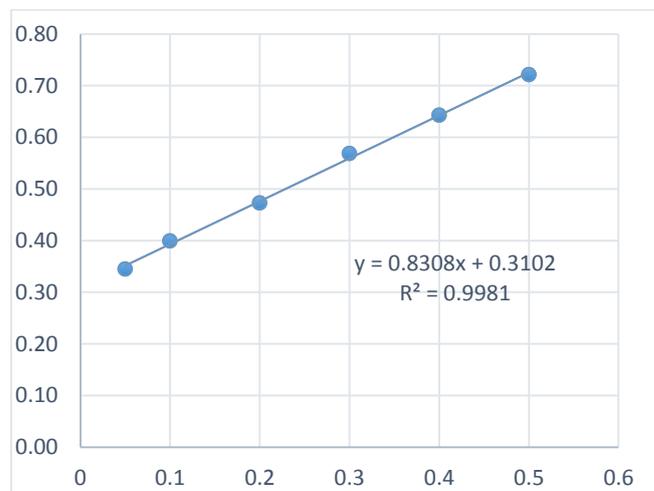
Figure 3.2 Example of calibration curve used to determine protein concentration of lysate samples in this study.

Mini Bradford was performed for protein lysis where a commercial lysis buffer was used. This allows for accurate protein detection in the presence of glycerol and NP-40. BSA protein standards were set up as shown in Table 3.1. Standards were set up in semi-micro polystyrene cuvettes (VWR International, Leicestershire, UK). To this, 200 µL of Bradford solution was added and values were measured on the Eppendorf BioPhotometer. Final protein concentrations were analysed as outlined previously.

Concentration of BSA (mg/mL)	Volume (µL) BSA (1mg/mL)	Volume (µL) dH₂O
0.01	1	799
0.03	3	797
0.05	5	795
0.1	10	790
0.15	15	785
0.2	20	780
Blank	0	800

Table 3.1 Dilutions for setting up a standard curve for a mini-Bradford used in this study.

3.7 Protein Investigations

Western blot method was used to quantify proteins. Briefly, a 10 µL sample volume was made by adding 1 µL 10X Sample Reducing Agent, 2.5 µL 4X LDS Sample Buffer (both Novex®, Thermo Fisher Scientific, Loughborough, UK) and the appropriate volume of sample lysate based on the protein concentration determined as described in Section 3.7.5. Samples were heated at 70 °C for 10 minutes.

3.7.1 Polyacrylamide gel electrophoresis (SDS-PAGE)

Protein electrophoresis was performed using SDS-PAGE denaturing acrylamide gels (4-20% NuPAGE® Bis-Tris Precast Gels or 4-20% MiniPROTEAN TGX Precast Gels, BioRad Laboratories, Hemel Hempsted, UK). The electrophoresis was carried out in MES buffer with 500 µL antioxidant added (both NuPAGE®, ThermoFisher Scientific, Loughborough, UK) at 120 V for 30 minutes and 160 V thereafter using the X-Cell system (Novex®, Thermo Fisher Scientific, Loughborough, UK). Two molecular markers were run in parallel with the samples, one coloured marker (See Blue Plus 2, Thermo Fisher Scientific, Loughborough, UK) to ensure the correct molecular weight was present in gel and one biotinylated marker (Biotinylated Protein Marker, Cell Signaling, Boston, MA, USA) to determine molecular weight during western blot detection.

3.7.2 Western Blotting

Transfer of protein to a membrane was carried out in the iBlot™ 2 Semi Dry Transfer System (Thermo Fisher Scientific, Loughborough, UK). Briefly, the gel was removed from its case and soaked in 10% ethanol for 10 minutes. Pre-prepared stacks were soaked in dH₂O. The components were assembled as follows; stack-membrane-gel-stack, a

protocol for mixed molecular weight was used for 7 minutes at 1.3 A constant. Once transferred, the membrane was removed and blocked in 5% buffer (2.5 g non-fat milk in 50 mL 1X TTBS) for 1 hour. After this, the primary antibody at an appropriate concentration (Table 3.2) was incubated overnight at 4 °C. Three washes with TTBS buffer (20 mM Tris-HCL pH7.5 (Sigma Aldrich, Dorset, UK), 0.5 M NaCl (BDH AnalR, VWR, Leicestershire, UK), 0.1% Tween 20 (Sigma Aldrich, Dorset, UK) for 10 minutes each were followed by incubation with a secondary antibody and anti-biotin antibody (Cell Signaling, Boston, MA, USA) diluted in blocking buffer for 1 hour at room temperature. Membrane was then washed 4 times for 5 minutes each in TTBS before developing with 1 mL (1:40 dilution) Pierce ECL Blotting Substrate (Fisher Scientific, Loughborough, UK Scientific) for 5 minutes in the dark. Membrane was exposed and imaged using the Amersham Imager 600 (GE Healthcare, Amersham, London, UK).

Antibody	Manufacturer	Catalog #	Dilution
Anti-PRDM9	Abcam, Cambridge, UK	ab85654	1:500
Anti-PRDM9	Novex, Thermo Fisher Scientific	730062	1:500
Anti-PRDM9	Human Protein Atlas	HPA063372	1:100
Anti-PRDM9	Human Protein Atlas	HPA059555	1:100
Anti-GAPDH	Santa Cruz, Dallas, Texas, USA	sc-25778	1:1000
Anti-MTCO1	Abcam, Cambridge, UK	ab14705	1:1000
Anti-β actin	Sigma Aldrich, Dorset, UK	A1978	1:2000
Anti-HSPA5 (Bip1)	Abnova, Cambridge, UK	PAB2462	1:1000
Anti-SDHA	Abcam, Cambridge, UK	ab14715	1:5000
Anti-MTCO2	Abcam, Cambridge, UK	ab3298	1:1000
Anti-TFAM	Source Bioscience, Manchester, UK	LS-C143233	1:1000
Anti-α tubulin	Abcam, Cambridge, UK	ab59680	1:1000
Anti-VDAC1	Abcam, Cambridge, UK	ab14734	1:250
Anti-HSP60	GeneTex, California, USA	GTX110089	1:5000
Anti-histone H3	Abcam, Cambridge, UK	ab1791	1:1000
Anti-H3K4me1	Cell Signaling, Boston, MA, USA	#5326	1:1000
Anti-H3K4me2	Cell Signaling, Boston, MA, USA	#9725	1:1000
Anti-H3K4me3	Cell Signaling, Boston, MA, USA	#9751	1:1000
Anti-OXPHOS cocktail	MitoSciences, Oregon, USA	MS604	1:250
Anti-biotin	Cell Signaling, Boston, MA, USA	#7727	1:1000
Rabbit anti-mouse	DAKO, Cambridge, UK	P0260	1:1000
Swine anti-rabbit	DAKO, Cambridge, UK	P0399	1:1000
Biotinylated protein ladder	Cell Signaling, Boston, MA, USA	#7727	1:2000
See Blue Plus2 ladder	Fisher Scientific, Loughborough, UK	LC5925	5 μL
PRDM9 synthetic protein	Abcam, Cambridge, UK	ab87806	1 μg/mL

Table 3.2 List of antibodies and protein ladders used in this study. Manufacturer is provided along with the working dilutions used in all experiments.

3.7.4 Immunoprecipitation

Immunoprecipitation (IP) was performed using an IP kit (Pierce Classic Magnetic #88804) following the manufacturers guidelines. Cell pellets were washed with PBS and collected on ice. For each pellet, 500 μL lysis buffer (Pierce) with 2% proteinase inhibitor cocktail (Roche Diagnostics Limited, West Sussex, UK) was added and after an

incubation of 5 minutes, the samples were centrifuged at 13,000 x rcf for 10 minutes at 4 °C. Lysate was analysed for protein concentration via mini Bradford outlined above. Antibody (Table 3.3) was added to the lysate and incubated overnight at 4 °C with constant rotation to form protein-antibody complexes. The protein-lysate complexes were incubated with A/G magnetic agarose beads for 2 hours at room temperature with constant rotation. Beads were then collected on a magnetic stand and washed several times in lysis buffer and dH$_2$O to remove any unbound protein. Final elution of bound protein was performed with 25 μL of elution buffer with 10% neutralisation buffer. Protein concentration was determined for each wash step plus final elution lysate via mini Bradford. SDS-PAGE was performed using the X-Cell electrophoresis unit (Novex) with 4-20% precast gels (Novex). After electrophoresis, gels were removed and incubated for 1 hour in InstantBlue solution (Expedeon). Gels were washed in warm dH$_2$O until background staining was eliminated and only protein bands were stained.

Antibody	Manufacturer	Quantity used
Anti-PRDM9	Abcam, Cambridge, UK	10 mg
Anti-PRDM9	Novex	10 mg
Anti-PRDM9 HPA059555	Human Protein Atlas	1.7 mg
Anti-PRDM9 HPA063372	Human Protein Atlas	3.5 mg

Table 3.3 Antibodies used for PRDM9 immunodetection.

3.7.5 Mass spectrometry analysis

Mass spectrometry analysis was performed at the Newcastle University Protein and Proteome Analysis Unit. Briefly, bands were cut out and reduction of the proteins performed in gel. Alkylation with iodoacetamide was carried out before in gel digestion with modified trypsin. Resulting peptides were extracted from the gel and purified on custom reversed phase columns. Analysis was performed by LCMSMS on a nano-HPLC system (column: 25 cm x 75 μm ID, flow rate 300 nL/min). Data was processed using the open source search engine X!Tandem and the Global Proteome Machine interface (http://www.thegpm.org/).

3.7.6 siRNA experiments

HEK293 cells were transfected with pre-designed PRDM9 siRNA (Silencer Select, Thermo Fisher Scientific, Loughborough, UK) following manufacturer's specifications with small modifications to the protocol. Briefly, cells were seeded at 10×10^4 cells/flask and incubated overnight in appropriate media. siRNA was diluted to 20 nM concentration in Opti-MEM media (GIBCO, Thermo Fisher Scientific, Loughborough, UK). Lipofectamine (RNAiMAX, Thermo Fisher Scientific, Loughborough, UK) was diluted to 0.5% in Opti-MEM. Reactions were set up as shown in Table 3.4.

siRNA target	10µM siRNA	Opti-MEM	Lipofectamine	Opti-MEM
PRDM9 (20nM)	100 µL	400 µL	2.5 µL	497.5 µL
Blank control	-	-	-	1000 µl
Lipofectamine control	-	500 µL	2.5 µL	497.5 µL
Non-targeting control (20nM)	100 µL	400 µL	2.5 µL	497.5 µL

Table 3.4 siRNA and lipofectamine dilutions. Controls were also used including a blank (Opti-MEM only), lipofectamine control (no siRNA) and a non-targeting siRNA control (Negative control #2, Thermo Fisher Scientific, Loughborough, UK).

The lipofectamine and siRNA dilutions were then combined and incubated at room temperature for 20 minutes. The dilutions were then added to the appropriate flasks and cells were left to transfect for 2-3 days depending on confluency. Forward transfection was performed every 2-3 days for the duration of the experiment.

3.8 Immunostaining and Imaging

3.8.1 CryoSectioning

Human tissue samples were collected by the relevant tissue bank, snap frozen in liquid nitrogen and stored at -80 °C. CryoSectioning was performed by mounting the tissue block in optimum cutting temperature formulation (OCT) and allowing the sample to adjust to -20 °C in the cryostat chamber. Sections were cut 10 μm thick, collected on Superfrost glass slides and allowed to dry for 30 minutes at room temperature to avoid freezing artefacts.

3.8.2 H&E staining of human tissue Sections

Slides were washed for 2 minutes in dH_2O and were placed in Haemalum solution for 1 minute. Slides were washed again for 2 minutes in dH_2O and placed in Eosin solution for 1 minute. Slides were washed for 2 minutes in dH_2O before being dehydrated in graded alcohol: 70% EtOH 30 seconds, 95% EtOH for 30 seconds, 100% EtOH for 2 minutes. Sections were covered in DPX solution and a glass coverslip was placed on top.

3.8.3 Immunostaining human tissue sections

Tissue sections on slides were fixed using 4% paraformaldehyde (PFA, Sigma Aldrich, Dorset, UK Aldrich) for 10 minutes. Slides were washed 3 times with PBS to completely remove the PFA. Sections were permeabilised by adding 0.5% Triton X-100/PBS for 10 minutes. After washing 3 times with PBS the Sections were blocked in buffer (5% FBS/0.1% Tween20/PBS) for 1 hour at room temperature. Primary antibodies were diluted in 2% FBS/PBS and were added to the Sections and left to incubate at 4 °C overnight.

Sections were washed twice for 5 minutes with cold PBS before the addition of the secondary fluorochrome-conjugated antibody diluted in 2% FBS/0.1% Tween20/PBS. Slides were incubated for 1 hour at room temperature in the dark. Slides were washed 3 x 5 minutes in PBS. Each Section was covered with VECTASHIELD Anti-fade Mounting Medium with DAPI (Vectorlabs).

3.8.4 Immunostaining cultured cells

Cells were grown on glass coverslips pre-coated with poly-L-lysine (Sigma Aldrich, Dorset, UK). Briefly, coverslips were washed in 70% EtOH before being placed into the well of a 6 well tissue culture plate and subjected to UV light for 20 minutes. 1 mL

poly-L-lysine was added to each well to completely cover the surface area. The plate was incubated at room temperature for 30 minutes. The poly-L-lysine solution was removed and the wells were washed twice with sterile PBS. Plates were covered with parafilm and stored at 4 °C until required or were used immediately by washing once with DMEM before cells were plated.

When the cells were ready to image, growth media was removed and cells were fixed in 4% PFA at room temperature for 10 minutes. Cells were then washed twice with PBS for 2 minutes each time. Blocking buffer (PBS/5% normal goat serum/0.3% Triton X-100) was added to the cells for 1 hour at room temperature before the addition of primary antibody in dilution buffer (PBS/1% BSA/0.3% Triton X-100) overnight at 4 °C. Cells were washed three times with PBS before addition of the secondary fluorochrome-conjugated antibody for 1 hour in the dark. Cells were washed three times with PBS before the addition of VECTASHIELD Anti-fade Mounting Medium with DAPI (Vectorlabs). All antibodies used are detailed in Table 3.5.

Antibody	Manufacturer	Catalog #	Dilution
Anti-PRDM9	Abcam, Cambridge, UK	ab85654	1:500
Anti-FLAG M2	Sigma Aldrich	F1804	1:500
Anti-OPA1	Abcam, Cambridge, UK	ab42364	1:250
Anti-TFAM	Source Bioscience, Manchester, UK	LS-C143233	1:250
Anti-HSP60	Abcam, Cambridge, UK	ab46798	1:400
Anti-TOMM20	Santa Cruz, Dallas, Texas, USA	sc-17764	1:1000
Goat anti-mouse Alexa Fluor® 488	Fisher Scientific, Loughborough, UK	A-11001	1:1000
Goat anti-rabbit Alexa Fluor® 488	Fisher Scientific, Loughborough, UK	A-11034	1:1000
Goat anti-mouse Alexa Fluor® 594	Fisher Scientific, Loughborough, UK	A-11032	1:1000

Table 3.5 List of antibodies used for immunostaining in this study.

3.8.5 Microscopy

Cells were imaged using the Nikon Confocal Microscope and or the AxioImager (Zeiss).

3.9 Cloning

Cloning was performed using the pGEM-T easy vector system (Promega, Southampton UK). Incubations carried out at 37 °C were performed in a temperature controlled microbiology incubator (Thermo Fisher Scientific, Loughborough, UK).

3.9.1 Product preparation

cDNA sequences of interest were provided already integrated into a vector DNA backbone. In order to subclone insert sequences into another vector system, restriction enzymes (Table 3.6) were utilised to digest the insert from the pcDNA3.1(+) vector sequence. Reaction mixture was as follows: 50 ng plasmid DNA, 10 U each restriction enzyme, 2 µL CutSmart buffer (New England Biolabs inc.), 3 µL10X BSA made up to a final volume of 50 µL with dH$_2$O. Reactions were incubated at 37 °C for 1 hour. The total volume of digested sample was run on a size selection gel as previously described. The product of interest was gel extracted using the QIAquick gel extraction kit following manufacturers instructions (Qiagen, Manchester, UK, Manchester UK).

Restriction enzyme	Manufacturer	5'-3' Cut site
ApaI	New England Biolabs Inc.	GGGCCC
BamHI	New England Biolabs Inc.	GGATCC
EcoRV	New England Biolabs Inc.	GATATC
HindIII	New England Biolabs Inc.	AAGCTT
PmeI	New England Biolabs Inc.	GTTTAAAC
XhoI	New England Biolabs Inc.	CTCGAG

Table 3.6 List of restriction enzymes used in this study.

3.9.2 Ligation

Next, a ligation reaction was set up using the ligase provided in the pGEM-T easy kit (Promega, Southampton UK). Reaction mixture was as follows: 2X Rapid ligation buffer,

50 ng pcDNA5 vector, 3U T4 DNA ligase, dH$_2$O and the volume of insert DNA from the restriction enzyme digest as calculated by the following equation:

$$insert{:}\,vector\ molar\ ratio = \frac{ng\ vector\ x\ size\ of\ insert\ (Kb)}{size\ of\ vector\ (Kb)}$$

To maximise the number of positive transformants from the reaction, ligation was performed at 16 °C for 12 hours.

3.9.3 Transformation

Ligation transformation was performed using JM109 high efficiency competent cells (Promega, Southampton UK), 2X Yeast Tryptone broth (2YT) and 2YT agar plates were used as growth medium. 2YT agar was autoclaved and allowed to cool to approximately 55 °C before being supplemented with 100 mg/mL ampicillin, 75 µg/mL X-Gal and 0.5 mM IPTG. Agar was poured into 90 mm plates under aseptic conditions and left to cool and solidify at room temperature. Plates were either dried at 37 °C or stored at 4 °C for up to 1 month. 5 µL of ligation product was added to pre-aliquoted (25 µL) competent cells on ice, gently agitating the tube to ensure the ligation product and cells were mixed. Competent cells used in this study are detailed in Table 3.7. The mixture was incubated on ice for 20 minutes before being subjected to heat shock for 45 seconds at 42 °C using a dry heat bath with 200 µL water added to the well. The tube was then returned to ice for 2 minutes before 470 µL autoclaved 2YT broth was added to the transformants and incubated at 37 °C for 1 hour at 150 rpm. To plate the cells, 200 µL of the cell suspension was spread onto an agar plate under aseptic conditions using a glass spreader dipped in ethanol and flamed. Plates were incubated for 16 hours at 37 °C.

White colonies were picked for colony PCR and sequencing. Growth of white colonies indicates a successful ligation and transformation. Blue colonies indicate a successful transformation but inefficient ligation due to the activation of β-galactosidase. To perform the PCR 10 µL dH$_2$O was added to each tube in a sterile 8 strip PCR tube. Autoclaved pipette tips were used to pick single colonies from the growth plate and were dipped in the PCR tube before being streaked onto a new growth plate and placed in 5 mL fresh 2YT broth, supplemented with 100 mg/mL ampicillin. The plate was

incubated at 37 °C to expand the colonies and the 5 mL inoculated 2YT broth was incubated at 37 °C at 150 rpm for 16 hours. Due to the size of the insert DNA sequences used in this study, long range PCR was performed as follows; to the 10 μL dH$_2$O colony pick, 10% betaine, 10 %10X LA PCR Buffer II (Mg^{2+}), 200 μM dNTP, 0.2 μM forward primer, 0.2 μM reverse primer and 1.25 U TaKaRa LA Taq was added to a final reaction volume of 25 μL. Thermocycling conditions were performed in 30 cycles of; 94 °C for 1 minute, denaturing at 98 °C for 10 seconds, primer annealing at 62 °C for 15 minutes followed by final extension at 72 °C for 10 minutes. Sequencing was performed as previously described. Once the desired product was confirmed by running PCR products on size separation gels and sequencing, 1 mL of the 5 mL inoculated broth was used to inoculate 100 mL 2YT broth. This broth was incubated at 37 °C for 16 hours at 150 rpm. The remaining 4 mL of original inoculated broth was used to create glycerol stocks of the colony, 150 μL glycerol was added to 850 μL broth in a 2 mL cryovial and stored at -80 °C.

Competent Cell Type	Manufacturer	Application
JM109	Promega	Plasmids <10Kb
One Shot® Stbl3™	Invitrogen	Large unstable plasmids
One Shot® MAX Efficiency® DH5α™	Invitrogen	Large plasmids

Table 3.7 Details of all competent cell types used throughout the study.

3.9.4 Plasmid purification

Growth medium inoculated with colonies carrying the correct plasmid and insert was centrifuged for 15 minutes at 6000 x rcf at 4 °C. Pellets were then processed using the Qiagen, Plasmid Maxi Kit (Qiagen, Manchester, UK) according to the manufacturers instructions. The final pellet was air dried for 10 minutes to allow any remaining ethanol to evaporate and the DNA was resuspended in 500 μL dH$_2$O. DNA concentration was measured using the Nanodrop2000 UV-vis Spectrophotometer (Thermo Fisher Scientific, Loughborough, UK).

3.10 Transfection of cultured cells

3.10.1 Transient overexpression system

Transfection of HEK293 cells was carried out in 6 well tissue culture plates. Briefly, HEK293 cells were harvested as previously described (Section 3.4.1), cell pellets were resuspended in 10 mL DMEM and counted using a haemocytometer. Cells were seeded at 30, 000 cells/well in 4 mL fresh DMEM and were left to adhere for 24 hours. The transfection mixture was prepared by adding 97 µL Optimem (Invitrogen, Thermo Fisher Scientific, Loughborough, UK) serum free media and 3 µL Genejuice (Novagen) to a 1.5 mL tube, mixed by vortexing and left to incubate at room temperature for 5 minutes. To this 1 µg plasmid DNA (cloned and purified as previously described in Section 3.10) was added and the mixture was incubated at room temperature for 20 minutes. The transfection mixture was added dropwise to the cell culture well containing the cells seeded previously. The cells were incubated at 37 °C for 3 days before harvesting and being processed for further analysis. For imaging protocols, 22 mm glass cover slips were placed in the wells of the tissue culture plates and were coated with 1 mL poly-L-lysine for 30 minutes at room temperature, washed twice with PBS, washed once with DMEM and then seeded with cells as described above. After transfection for 3 days, the cover slips were removed from the plates and processed for immunostaining as described in Section 3.8.4.

3.10.2 Stable overexpression cell line

The Flp-In T-Rex system (Invitrogen, Thermo Fisher Scientific, Loughborough, UK) was used to selectively induce overexpression of genes of interest in HEK293 cells previously modified to contain a stable FRT site and the tetracycline repressor in the genome (a kind gift from Professor Robert Lightowlers, Institute of Cell and Molecular Biology Newcastle University). Cells were transfected with a reaction mixture comprising; 91 µL OptiMem (GIBCO, Thermo Fisher Scientific, Loughborough, UK), 9 µL Genejuice (Novagen, EMD Biosciences, USA), a range of 2 µg, 4 µg and 6 µg pcDNA5/FRT/TO plasmid and 1.8 µg pOG44 plasmid expressing Flp recombinase. Expression of Flp recombinase enzyme is required to catalyse the transfer of the insert DNA sequence from the pcDNA5 plasmid into the genome of the cell through homologous recombination. This occurs at specific recognition sites present in the pcDNA5 construct sequence and the genome. Stable integration of the gene of interest confers hygromycin resistance to the cell.

After 24 hours the cells were split to 50% confluency and stably integrated colonies were selected for by adding 50 mg/mL hygromycin to growth medium. To ensure the stability of the tetracycline promoter in the HEK293 genome, 100 mg/mL blasticidin was added to appropriate growth medium and added to the cells every third media feed. After several weeks of colony selection, pure cultures of transfected cells were obtained. Tetracycline (1 µg/mL) was added to the growth media to induce gene expression before cells were harvested for further investigation.

3.11 Quantitative polymerase chain reaction

3.11.1 Reverse transcription of RNA to cDNA

Total RNA was extracted from HEK293 cells as described in Section 3.6. RNA was converted to cDNA using the High Capacity Reverse Transcription Kit (Thermo Fisher Scientific, Loughborough, UK). RNA was diluted to a final concentration of 2 µg in a 10 µL volume using nuclease free water before the reaction mixture was added. The reaction mixture was as follows: 2 µL 10X RT buffer, 4 mM dNTPs, 2 µL 10X random primers, 1 µL MultiScribe reverse transcriptase made to a final volume of 10 µL with nuclease free water. The 20 µL reactions were incubated at 25 °C for 10 minutes, 37 °C for 120 minutes and finally 85 °C 5 minutes. Samples were stored at -20 °C until required.

3.11.2 PRDM9 gene expression

Quantification of the relative transcription of specific genes was carried out by real time quantitative PCR (RT-qPCR) using either a double stranded DNA intercalating dye (SYBR) or the TaqMan® gene Expression Assay (Applied Biosystems, Thermo Fisher Scientific, Loughborough, UK). The SYBR® reactions were comprised as follows; 1X iQ™ SYBR® green, 0.5 µM forward and reverse primer, 1-5 µg sample cDNA, made up to 25 µL with dH$_2$O. All reactions were carried out on the MyIQ™ thermocycler (BioRad Laboratories, Hemel Hempsted, UK) with the following program: initial denaturation at 95 °C for 3 minutes followed by 40 cycles of; 95 °C for 10 seconds and 62.5 °C for 1 minute with a final 95 °C for 1 minute and a melt curve. A melt curve was performed by measuring loss of fluorescence after each 10 second incubation step, rising from 62 – 95 °C, in 0.5 °C increments.

For the TaqMan® gene Expression Assay, target specific probes are designed, nested in between target specific primers, and covalently attached to a fluorescent reporter dye on the 5' end and a non-fluorescent quencher dye on the 3' end. When the reporter and quencher are in close proximity the excitation energy of the fluorescent reporter is transferred to the quencher through FRET chemistry (fluorescent resonance energy transfer). During the PCR reaction when the amplicon extends, the fluorescent reporter dye is cleaved from the probe by Taq polymerase exonuclease activity and the emission spectra of the reporter can be measured.

RT-qPCR was carried out on a 7500 Fast Real Time PCR System. Reaction mix was as follows, 1X TaqMan® Gene Expression Master Mix, 250 nM TaqMan® probe, 4 µg cDNA made up to a final volume of 20 µL nuclease free water. The reaction was as follows, initial incubation at 50 °C for 2 minutes, denaturing at 95 °C for 20 seconds and 40 cycles of primer annealing at 60 °C for 30 seconds and 95 °C for 3 seconds when fluorescence was detected. Serial 1:10 dilutions of cDNA in nuclease free water were used to ensure linearity and efficiency of the reaction and for standard curve quantification. Each sample was measured in triplicate and contamination of reagents excluded by the use of cDNA negative controls. Targets and their probes are listed in Table 3.8.

Target	Probe	Location	Length (bp)	Supplier
PRDM9	Hs01633270_s1	3569	179	Applied Biosystems
PRDM9	Hs00360639_m1	360	143	Applied Biosystems
β-actin	Hs01060665	208	63	Applied Biosystems

Table 3.8 Probes used in this study.

3.11.3 Mitochondrial DNA copy number

Quantification of mtDNA copy number and mtDNA deletion level was carried out using an RT-qPCR assay. This method allows detection of multiple target specific fluorescent probes in the same reaction using Taqman® qPCR amplification as described in Section 3.11.2. Each probe was designed so that the emission spectra of the fluorophore did not overlap with the other probes therefore allowing accurate detection of each amplicon. Quencher dyes were also designed to have an absorbance spectrum which overlapped

the emission spectra of the fluorophore, limiting background signal from the probe itself. Serial 1:10 dilutions of DNA from non-transfected HEK293 cells were used to ensure linearity of the reaction and for standard curve quantification. The reaction mixture was as follows: 0.6 μL of each primer, 0.4 μL of each probe, 10 μL iTaq SuperMix reaction buffer (BioRad Laboratories, Hemel Hempsted, UK) and 1 μL of sample DNA made up to 20 μL with nuclease free water. Probes used in this study are outlined in Table 3.9.

Target	Forward Primer 5'-3'	Reverse Primer 5'-3'	Dye	Quencher	Probe 5'-3'
RNase P	AGATTTGGACCTGCGAGCG	GAGCGGCTGTCTCCACAAGT	HEX	BHQ-2	TTCTGACCTGAAGGCTCTGCGCG
B2M	CACTGAAAAAGATGAGTATGCC	AACATTCCCTGACAATCCC	FAM	BHQ_1	CCGTGTGAACCATGTGACTTTGTC
MT-CYB	ATGACCCCAATACGCAAAA	CGAAGTTTCATCATGCGGAG	CY5	BHQ-2	CATTCATCGACCTCCCCACCC
MT-CO3	ATGACCCACCAATCACATGC	ATCACATGGCTAGGCCGGAG	FAM	BHQ-1	ACCCAGCCCATGACCCCTAAC
MT-ND1	ACGCCATAAAACTCTTCACCAAAG	GGGTTCATAGTAGAAGAGCGATGG	HEX	BHQ_1	ACCCGCCACATCTACCATCACCCTC
MT-ND4	ACCTTGGCTATCATCACCCGAT	AGTGCGATGAGTAGGGGAAGG	Cy5	BHQ_2	CAACCAGCCAGAACGCCTGAACGCA

Table 3.9 Primer and probe details used for mtDNA copy number quantification.

Chapter 4 Analysis of PRDM9 binding motifs in mtDNA sequences.

4.1 Overview

The role of PRDM9 in creating recombination hotspots within the human genome is predicted to occur when the protein binds to the DNA recognition motif through its zinc finger domain (Myers *et al.*, 2008; Baudat *et al.*, 2010; Billings *et al.*, 2013). Due to the polymorphic nature of the PRDM9 zinc finger repeat region, each *PRDM9* allele in the population is predicted to have a subtly different genomic binding motif. A previous study showed that the predicted PRDM9 binding motif is also present in mtDNA at the site flanking the 4977bp 'common' deletion (Myers *et al.*, 2008). To determine whether this predicted recognition motif is present in human mtDNA we used an *in silico* screening approach. A large cohort (n=31,516) of mtDNA sequences downloaded from the Genbank sequence depository were computationally screened for the presence of each of the eleven predicted PRDM9 binding motifs. This is of interest as protein-DNA binding motifs within the mtDNA could explain why certain regions of the molecule are prone to deletion formation (Figure 4.1).
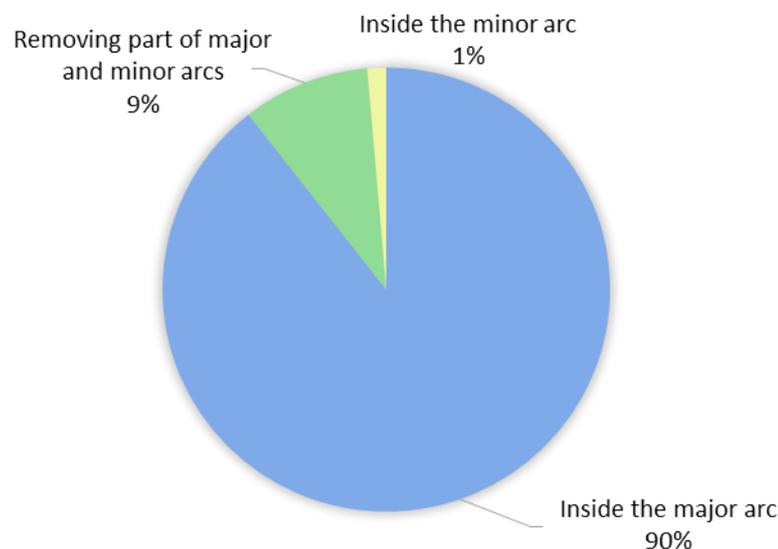
Figure 4.1 Location of mtDNA deletions in respect to the region of the molecule based on the reported cases. Adapted from the Mitobreak mitochondrial deletion database (Damas *et al.*, 2014).

81

In addition, the matrilineal nature of mtDNA can be utilised to define sequences based on their phylogenetic ancestry allowing population differences to be explored (Emery *et al.*, 2015). Associations between haplogroup status and disease risk have been described previously, however no study has interrogated the potential difference in mtDNA deletion susceptibility between population haplogroups. Deletion frequency and haplogroup-associated disease prevalence have been extensively reported in European populations (Taylor and Turnbull, 2005; Bua *et al.*, 2006; Campbell *et al.*, 2014; Komulainen *et al.*, 2015; Mancuso *et al.*, 2015) however there is a lack of data for populations from other geographical regions. The mtDNA sequences screened for the presence of PRDM9 binding sites were also segregated by haplogroup to assess whether there is any correlation between PRDM9 motif presence and haplogroup status.

4.2 Hypothesis

Background mtDNA sequence variation will alter the frequency of the PRDM9 recognition motif potentially giving an insight into the differential frequency of mtDNA deletions between human populations.

4.3 Aims

This study aims to determine whether the PRDM9 recognition motif is present in human mtDNA sequences and identify whether different mtDNA haplogroups have unique PRDM9 motifs.

4.4 Results

4.3.1 Pilot screen of PRDM9 binding motifs in 9, 769 sequences

European mtDNA sequence files were downloaded from the NCBI Nucleotide Database (http://www.ncbi.nlm.nih.gov/nuccore/) in FASTA format. In order to quickly screen for the presence of the PRDM9 recognition motif, *Perl* was used to search sequence files for the most common motif (motif A) '*CCNCCNTNNCCNC*'. The script used is shown in Figure 4.2.

```perl
#!/usr/bin/perl
use strict;
use warnings;
use autodie;
# Finding motif in DNA sequencing file
# define variables
my $motif = "CC.CC.T..CC.C";
my @c; #array variable to store DNA sequence
my $fh; #filehandle to use in subroutine
my $output = "motif_result.txt";
my $currdir= "/home/b2050897/Euroseq";
my $Bigsequence;
#chomp $currdir;
my @DirContent=`ls $currdir`;
my $ID;
# write output to a file
open OUT, ">$output" or die "CANNOT OPEN OUTPUT FILE";
#loop to open each file in file directory
foreach my $f (@DirContent){
chomp $f;
if($f=~/(\S+).fas/){  #\S+ matches pattern for any file ending .fas in folder
$ID=$1; #assign value of 1 to each file matched
# tell the programme the filepath and open the input file
my $fullpath = $currdir."/".$f;
open IN, $fullpath or die "CANNOT OPEN IN FILE";
my $header = <IN>;
# loop to create one line of DNA sequence from the input file
while (<IN>){
                my $line =$_;
                chomp $line;
                 $Bigsequence = $Bigsequence.$line;
 }
close(IN);
#IF loop to search the string for the motif
if ($Bigsequence =~/$motif/){
                @c = split ("", $Bigsequence);
                #loop to count the motif match and place in scalar array @c
                for (my $count1 = 0; $count1 <= ((scalar @c) -13);$count1++){
                            my $tempseq = ""; #place the counted sequences in a my variable
                            for (my $count2 = 0; $count2<13; $count2++){
                                        my $final_count = $count1 + $count2;
                                        $tempseq= $tempseq.$c[$final_count];
 }
#new loop to search the extracted split sequence for the motif
if ($tempseq =~/$motif/){
                my $print_count = $count1+1; #must add 1 as the count starts from 0 but DNA base starts from 1
                print OUT "\nSTART POSITION DNA: $print_count\nMOTIF SEQUENCE: $tempseq\nFILENAME:
$ID\n";
 }
 }
 }
 }
 }
close OUT;
exit;
```

Figure 4.2 Custom *Perl* script to screen sequence files for the PRDM9 recognition motif 'A' Output file will contain; the filename for each sequence file containing a matching sequence, the sequence matched and the start position in the sequence file (when the first base of the sequence file is counted as 1).

Using a simple find loop, regions of the sequences matching the motif were counted and given as total matches per sequence along with the base position where the match occurred (Table 4.1). In contrast to the site described by Myers et al, motif A was found at position m.5327 in 99.9% of sequences screened (Table 4.1). Several sequences contained the motif in more than one position. Interestingly, nine (0.00092%) of the sequences did not contain this recognition site due to the polymorphism m.5333T>C in the gene coding for the OXPHOS subunit mitochondrial NADH dehydrogenase 2 (*MT-ND2*). This SNP is not a population haplogroup marker and is a synonymous mutation not previously reported in the NCBI dbSNP database (http://www.ncbi.nlm.nih.gov/variation/view/), suggesting that the population frequency is very low.

Sixteen (0.0016%) of the sequences screened had a motif starting at position m.14481 caused by the transition m.14484T>C, one of three pathogenic mutations found in Leber's hereditary optic neuropathy (LHON) (Johns *et al,* 1992*;* Mackey and Howell, 1992; Man *et al,* 2003). This mutation causes a non-synonymous change in the gene encoding mitochondrial NADH dehydrogenase 6 (*MT-ND6*, rs199476104).

Position m.304 contains a variable poly-C tract with a tyrosine nucleotide in the middle that differs in length from 12-18 bp within the population (Marchington *et al.*, 1997). An insertion of an extra cytosine nucleotide leads to the creation of a PRDM9 recognition motif in twelve (0.0012%) of the sequences screened (Table 4.1). In twenty-eight (0.0029%) of the sequences screened there was an insertion of a tyrosine nucleotide at position m.311 leading to the creation of a PRDM9 motif shifted by one base pair to begin at m.304.

Frequency	Motif Start	Motif End	Motif Match	Array ID	SNP
			CCNCCNTNNCCNC		
9760	5327	5340	CCTCCTTAACCTC	rCRS	rCRS
9	5327	5340	CCTCCTCAACCTC	5333	m.5332T>C
1	303	316	CCCCC-TCCCCCC	cTRAC	-
8	303	316	CCCCCCTCCCCCC	cTRAC	316insC
3	304	317	CCCCCCT-CCCCC	cTRAC	-
28	304	317	CCCCCCTTCCCCC	cTRAC	311insT
3	8131	8144	CCACCTTCACCGC	8134	m.8133T>C
3	8273	8286	CCCCCCTACCCCC	8277	m.8276T>C
6	14054	14067	CCTCCATCACCAC	14063	m.14062T>C
1	14058	14071	CCACCATCACCTC	14060	m.14059T>C
16	14481	14494	CCACCATTCCCCC	14485	m.14484T>C
1	14481	14494	CCACCATCCCCCC	14485 14488	m.14484T>C//m.14487T>C
			GGNGGNAGGGG		
9655	9529	9540	CCCCTACCCCC	rCRS	rCRS
11	9529	9540	CCCCTGCCCCC	9531	m.9530A>G
28	308	319	CCCCTTCCCCC	cTRAC	311insT
9	308	319	CCCCTCCCCCC	cTRAC	316insC
3	308	319	CCCCT-CCCCC	cTRAC	-
7	459	470	CCCCTCCCACC	467	m.466T>C
1	5441	5452	CCCCTTCCTCC	5441	m.5440A>C
2	6224	6235	CCCCTCCCTCC	6221 6227	m.6220T>C m.6226T>C

Table 4.1 PRDM9 motif search in 9769 mtDNA sequences. The frequency of each motif match is shown in column 1, this corresponds to the number of mtDNA sequences that have the motif. The matched sequence start and end positions are shown in columns 2 & 3 and correspond to the mtDNA position 5'-3'. The DNA sequence matched is shown in column 4. The revised Cambridge Reference Sequence (rCRS) is used as the 'wild type' reference sequence. SNP changes from rCRS causing a motif site to be created are shown in column 6.

The sequence files were screened for an alternative A motif (GGnGGnAGGGG) suggested in the literature to be a PRDM9 binding motif (Berg *et al*, 2011). When the mtDNA was screened for the antisense strand sequence of this motif there was a positive hit in almost every sequence (98.8%) at position m.9529 (Table 4.1). A SNP at m.9531 (A>G) changes the binding motif in eleven of the sequences screened (0.0011%). This SNP gives place to a non-synonymous change m.9531A>G/p.MT-CO3:T109A (rs386829082).

Having determined that the PRDM9 'A' motif binding site was present in human mtDNA, the search was scaled up to include a larger number of sequences.

4.3.2 Expanded search for PRDM9 binding motifs n=31, 516 sequences

Sequence files were downloaded as above and were converted from FASTA to BAM format. BAM files were aligned using the Integrated Genome Viewer (IGV) and Haplogrep was used to determine mtDNA haplogroup. Haplogrep searches a defined set of mitochondrial haplogroup specific SNPs continuously updated and documented in PhyloTree (van Oven and Kayser, 2009; Kloss-Brandstatter *et al.*, 2011). The files were then screened using a custom *Perl* script written by Dr H. Griffin (Institute of Genetic Medicine, Newcastle University) for all eleven possible PRDM9 motif matches listed in Table 4.2 (Berg et al., 2012). Each sequence was searched for every motif several times to ensure that all possible matches were counted as hits at specific nucleotide positions.

PRDM9 Allele Match	Motif Sequence	Motif Search No.
A/B	CCNCCNTNNCCNC	1
A/B 2	CCNCCNTNNCCNCC	2
C	CCNCNNTNNNCNTNNC	3
D	CCNCNNNCNTNNCCNC	4
E	CNNNNCNNNTNNCCNCC	5
L1	CNNNCCNTNNCCNC	6
L4/L8/L14/L16/L17/L18	CCNCNNTNNNCNTNNC	7
L5	CNNCNNNNNNNNNCCNC	8
L6	CCNCNNTNNNCNTNNC	9
L7/L11/L12	CCNCCNTNNC	10
L19	NCNNTNNNCNTNNCCNC	11

Table 4.2 Sequences of the eleven PRDM9 recognition motifs used to search the mtDNA sequence files. The corresponding *PRDM9* alleles are shown in the left hand column, the protein products of which are predicted to bind the motif sequence shown in the middle column. Each motif was assigned a motif search number shown in the right hand column. Adapted from Berg et al., 2012.

Sequences were categorised into three main groups based on macro-haplogroup status: European, African and Asian. Out of the total 31, 516 sequences screened; 16, 490 were European (52.32%), 4020 were African (12.75%) and 11, 006 were Asian (34.92%) (Figure 4.3). The nature of maternal inheritance has allowed these sequences to be

further sub-grouped based on a well-established set of mtDNA population SNPs (Torroni *et al.*, 1994; Brown *et al.*, 1998; Finnila *et al.*, 2001).

European haplogroups comprised over half of the sequences in this data set (Figure 4.3 A). Within the European group itself, sub-haplogroups appeared to be present in frequencies representative of the European population observed in other studies for example haplogroup H was present in 39.9% of our samples, mirroring the observed population frequency of ~40% (Figure 4.3 B). Asian sub-haplogroups also showed population distribution as the most frequent haplogroup represented was M at 24.5% of the Asian population (Figure 4.3 C). All African haplogroups were denoted as L due to the most ancestral mtDNA's being extremely similar in sequence and belonging almost exclusively to haplogroups L0, L1, L2 or L3 (Emery *et al.*, 2015).
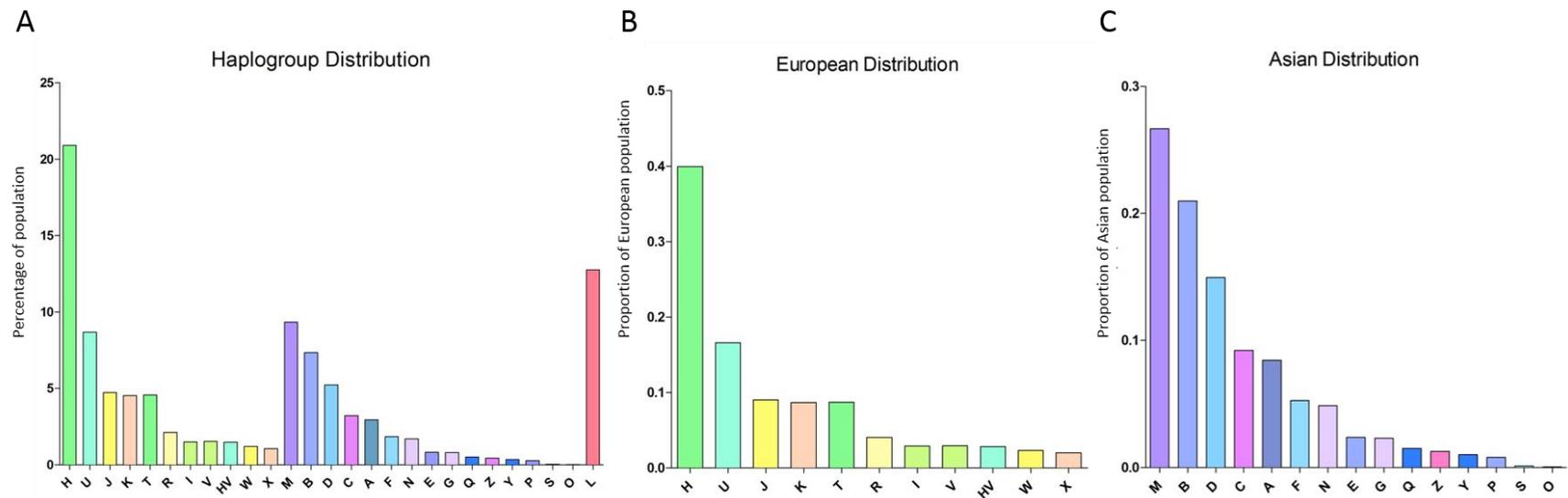
Figure 4.3 Distribution of mitochondrial haplogroup frequencies in the 31,516 sequences analysed. A) Haplogroup frequency as a percentage of the total number of sequences used in the analysis. B) European haplogroup frequency as a proportion of the total number of European sequences used in the analysis. C) Asian haplogroup frequency as a proportion of the total number of Asian sequences used in the analysis.

Firstly, the frequency of binding sites was assessed in each population group. All eleven binding motifs were analysed, however the most informative are the motifs bound by the most common *PRDM9* alleles A, B and C. Slight differences were seen in motif A binding sites between the three population groups (Figure 4.4 A). Most mtDNA sequences contained two motif A binding sites. Asian sequences appeared to have a higher frequency of genomes containing one motif site. European sequences had a higher frequency of mtDNAs with three binding sites. Although it was rare to have mtDNA sequences with no motif A recognition sites, there was a higher number of African sequences lacking any binding motif (Figure 4.4 A).
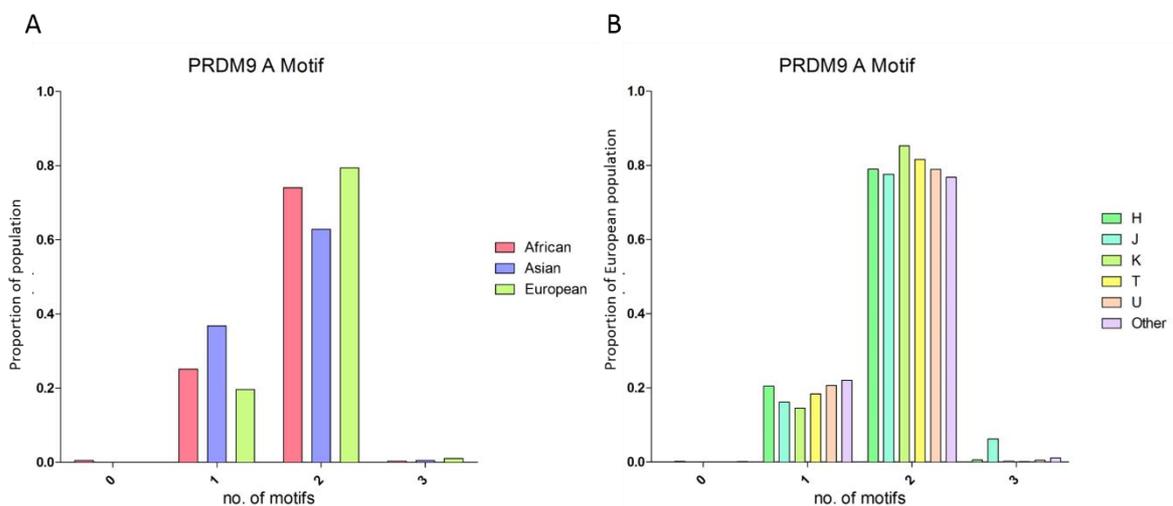
Figure 4.4 PRDM9 'Motif A' binding sites. A) Number of times motif A was found in mtDNA sequences from African, Asian and European mtDNA sequence groups. Proportion of population refers to the percentage of mtDNA sequences from each of the three population groups as a fraction of 1. B) Number of times Motif A was found in mtDNA sequences from each European haplogroup present in the mtDNA sequencing files. Proportion of population refers to the percentage of mtDNA sequences from the European population group as a fraction of 1.

European sequences were then subdivided by haplogroup (Figure 4.4 B). This analysis showed that most of the mtDNAs containing three motif A sites were haplogroup J. There were subtle differences between sub-haplogroups containing one or two motif A sites however the general trend followed that of the population with more European sequences containing two motif A sites (Figure 4.4 A & B).

The second most common *PRDM9* allele in Europeans is B. The motif for allele B was analysed by broad population grouping (Figure 4.5 A). Most sequences analysed did not contain any motif B sites. A small proportion of sequences in each population had one motif B site.
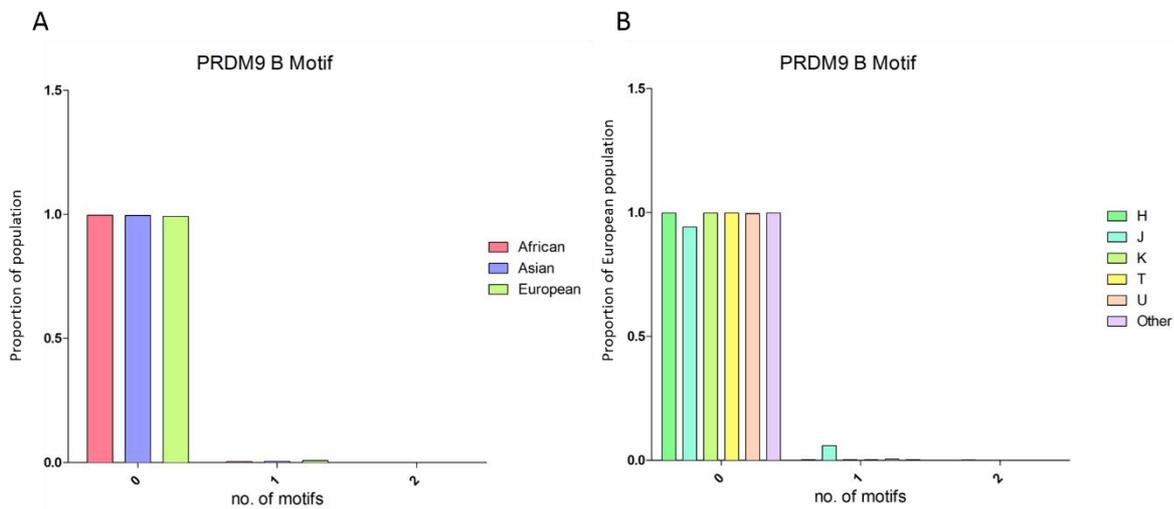
Figure 4.5 PRDM9 'Motif B' binding sites. A) Number of times Motif B was found in mtDNA sequences from African, Asian and European mtDNA sequence groups. Proportion of population refers to the percentage of mtDNA sequences from each of the three population groups as a fraction of 1. B) Number of times Motif B was found in mtDNA sequences from each European haplogroup present in the mtDNA sequencing files. Proportion of population refers to the percentage of mtDNA sequences from the European population group as a fraction of 1.

When the European sequences were analysed by sub-haplogroup, haplogroup J was over represented in the sequences containing one motif B site (Figure 4.5 B). Interestingly, haplogroup J was the only mtDNA sequence to have two motif B sites.

The second most prevalent PRDM9 allele in African populations is the C allele. Several motif C sites were found within mtDNA in all populations analysed with the majority of sequences containing ten sites (Figure 4.6 A). The proportion of African alleles containing eleven motif C sites was higher than that of the European and Asian groups. Conversely, a larger proportion of European and Asian sequences tended to have less motif C sites. A slightly higher proportion of Asian sequences contained nine motif C sites compared to European and African groups. European sequences were further sub-haplogrouped and analysed (Figure 4.6 B). Interestingly, the proportion of sequences containing nine motif C sites were haplogroups I, R, W & X (denoted as 'other' in Figure 4.6 B).
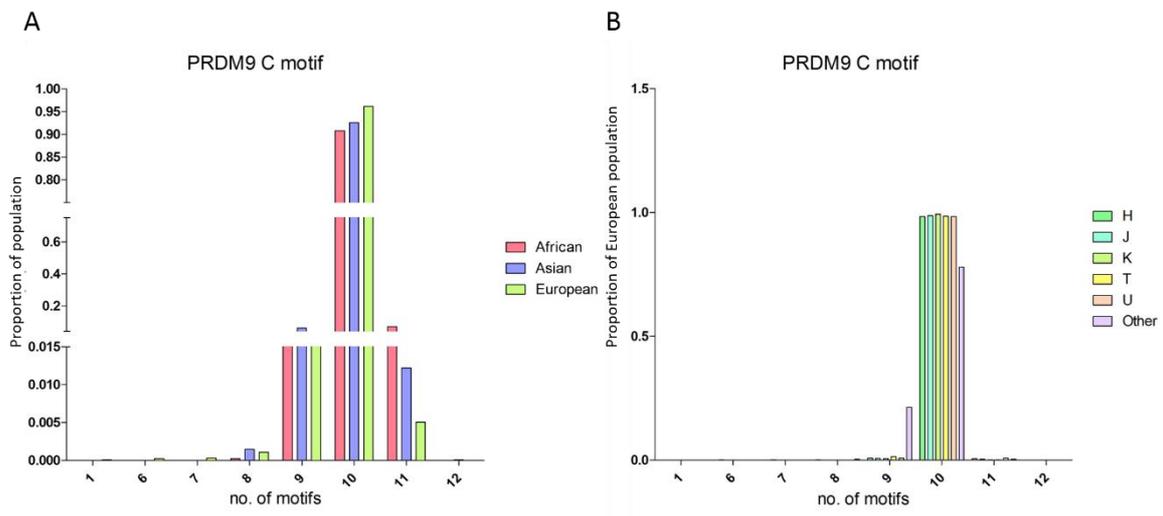
Figure 4.6 PRDM9 'Motif C' binding sites. A) Number of times Motif C was found in mtDNA sequences from African, Asian and European mtDNA sequence groups. Proportion of population refers to the percentage of mtDNA sequences from each of the three population groups as a fraction of 1. B) Number of times Motif C was found in mtDNA sequences from each European haplogroup present in the mtDNA sequencing files. Proportion of population refers to the percentage of mtDNA sequences from the European population group as a fraction of 1.

Next, the positions of the binding motifs along the mtDNA molecule were investigated (Figure 4.7). The analysis was performed by Dr H. Griffin (Institute of Genetic Medicine, Newcastle University). There appear to be clusters of PRDM9 recognition motif sites on the mtDNA molecule as depicted by peaks of dots extending to the outer edge of the circle. The outer circle shows the frequency of the three most common allele binding motifs A, B and C in red, blue and green respectively. Of note are positions m.5327 and m.14484 which appear to be sites of multiple PRDM9 motifs, not just motif A as previously described in the pilot screen in Section 4.3.1. There is also a motif cluster at position m.12400 which appears to be a site containing multiple PRDM9 motifs. This is expected as the eleven motifs searched for are very similar in sequence (Table 4.2) due to the degeneracy required for compatibility of a variety of PRDM9 allele combinations within individuals.
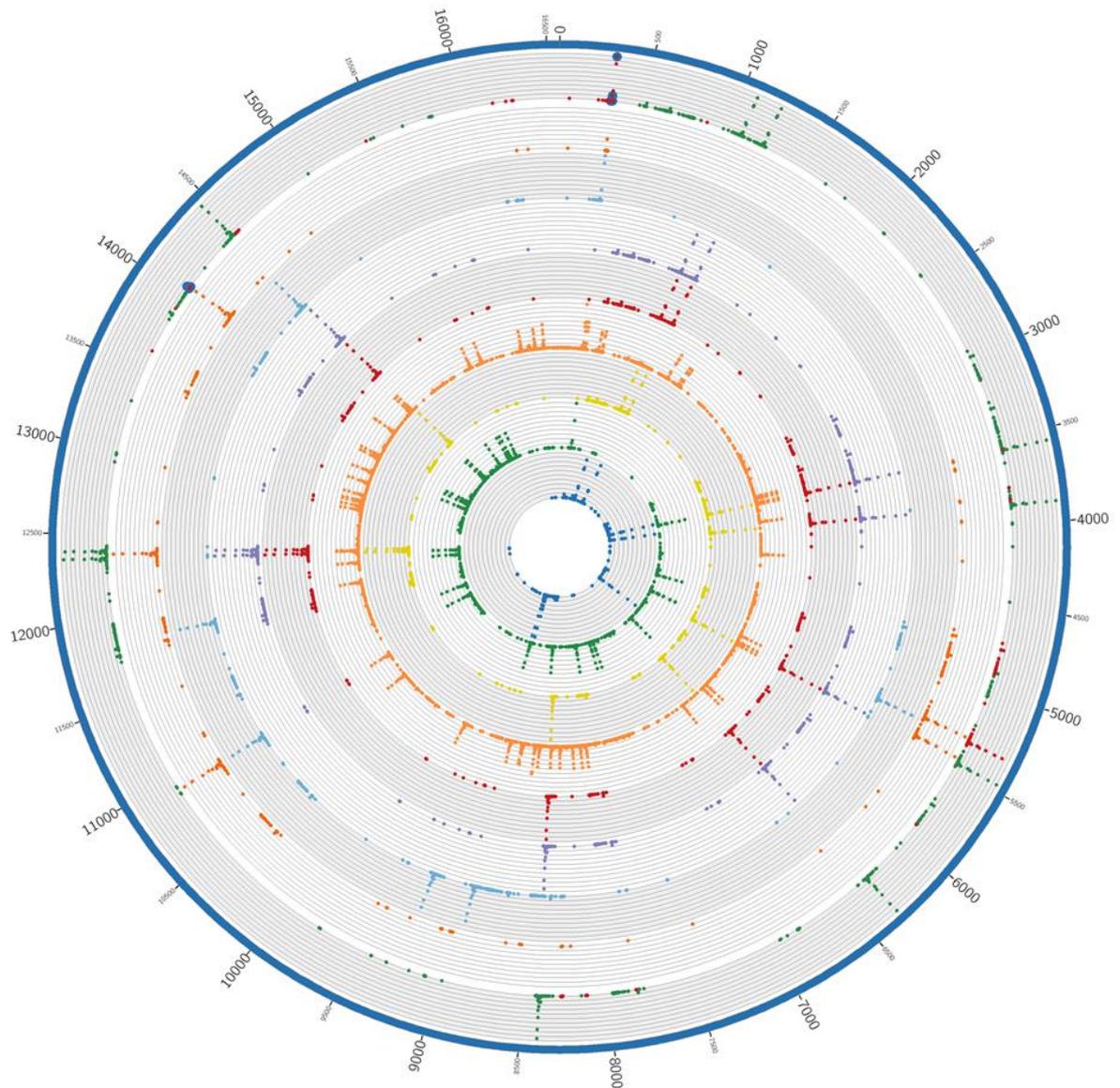
Figure 4.7 Circos plot of all eleven PRDM9 binding motifs searched in this study. Outer circle represents the mtDNA sequence in clockwise orientation with base positions denoted. From outer circle to the centre the order of the motif searches plotted is; 1, 2 & 3 (red, blue, green respectively); 4; 5; 6; 7; 8; 9; 10; 11.

Two of the motifs searched showed sequence matches over almost the entire mtDNA molecule (Figure 4.7). Motif 8, shown in orange on the circos plot, displays coverage of the whole mtDNA sequence. Additionally, motif 10 which is shown in dark green towards the centre of Figure 4.7 also shows sequence matches across a large proportion of the mtDNA molecule. It is likely that this representation is due to the degeneracy of these two motifs. Highly degenerate sequence motifs will by nature match more regions than those with a more specific sequence.

4.3.3 Alignment of known mtDNA deletion breakpoints with PRDM9 motifs

All 807 reported mtDNA deletions are recorded within the MitoBreak database (Damas *et al.*, 2014). Breakpoint data for each deletion is available and the frequencies of each, displayed as number of cases, are shown in Figure 4.8. Both 5' and 3' breakpoints are observed as occurring more frequently within the major arc of the molecule, between positions m.5721 and m.15887. Breakpoints have also been observed within the minor arc but are less frequent (Figure 4.8).
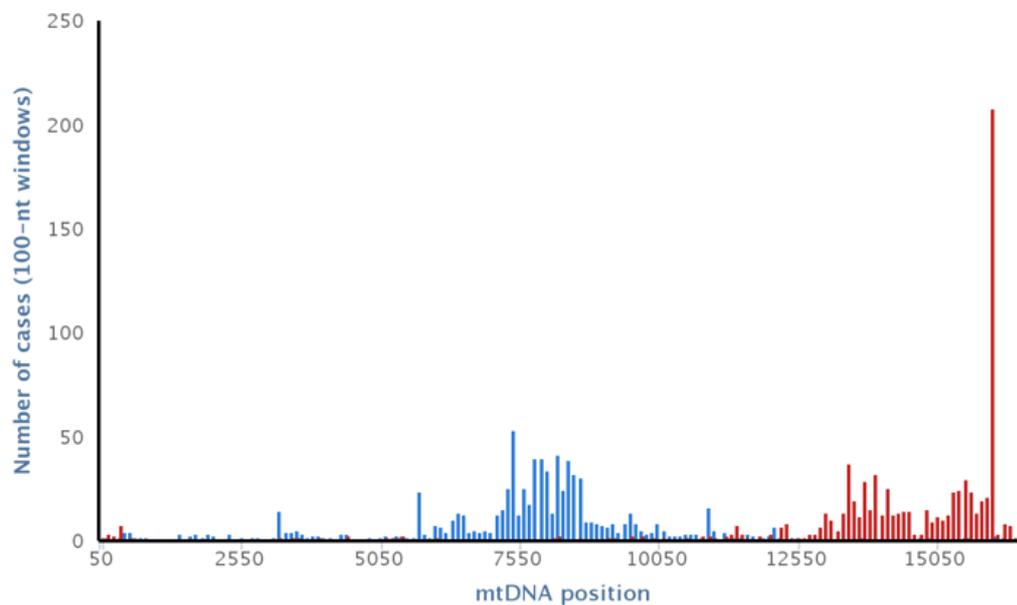
Figure 4.8 mtDNA deletion breakpoint frequencies. The x-axis represents the position along the mtDNA, with values shown every 100 nucleotides. The y-axis represents the number of reported cases in which each deletion breakpoint has been observed. Blue bins represent 5' breakpoint positions, red bins represent 3' breakpoint positions. This image was adapted from www.mitobreak.portugene.com (Damas *et al.*, 2014).

Available breakpoint data were downloaded for all 807 published deletions and subsequently aligned with the PRDM9 motif data presented in Section 4.3.2, to ascertain whether any breakpoints occur at or within the proposed PRDM9 motif sites. Briefly, a simple count command was used to determine whether any of the deletion breakpoint positions were also PRDM9 motif positions. Results were transformed back to the mtDNA base positions and plotted as a histogram. This analysis was performed for all 9 genetically distinct motif positions identified (Table 4.2). The frequencies of breakpoints which coincided with PRDM9 motif sites are shown in Figure 4.9.
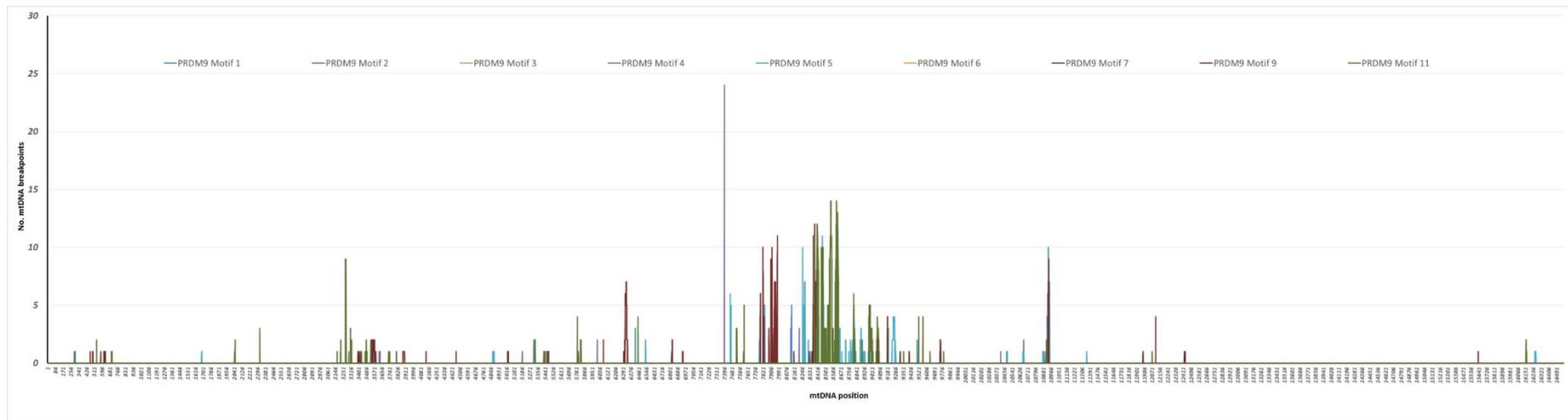
Figure 4.9 Frequency histogram of mtDNA deletion breakpoints aligned with PRDM9 motif sites identified within the mtDNA. All 9 genetically distinct PRDM9 motifs are displayed in different colours (listed at the top of the graph). The x-axis represents the mtDNA position and the y-axis represents the number of breakpoints that were found to align with a particular motif site.

Motifs 8 and 10 were not included in this analysis because they were genetically identical to partial segments of the 9 other PRDM9 motifs. Several mtDNA breakpoints were found either within or immediately adjacent to a PRDM9 motif site. Interestingly, some PRDM9 motif sites were found to coincide with more than one breakpoint position. For example, there were 24 breakpoints at position m.7400, which is inside a PRDM9 motif 4 site (largest peak, Figure 4.9). Frequencies of breakpoint and PRDM9 motif alignment varied across the molecule however, generally, higher frequencies were observed in the major arc of the molecule than in the minor arc. In addition, breakpoints aligning with PRDM9 motif sites were observed more often within the major arc of the molecule, as would be expected given that most breakpoints occur within this region.

Next, the data was analysed visually for each individual PRDM9 motif using the UCSC Genome Browser (www.genome.ucsc.edu). Briefly, the 5' and 3' breakpoints of each mtDNA deletion were saved in BED file format. This file format was also created for the start and end positions of all identified PRDM9 motifs within the mtDNA (described in Section 4.3.2). The reference chromosome used was the mitochondrial genome (ChrM). The data files were uploaded into the Genome Browser as a custom track and aligned with the Human Genome Assembly GRCh38/hg38. This analysis was performed for all 9 genetically distinct PRDM9 motifs. Each PRDM9 motif data file was uploaded to the browser alongside the mtDNA breakpoint data file for direct comparison. A visual representation of the data for PRDM9 motif 1 (*CCnCCnTnnCCnC*) is shown in Figure 4.10. The images in this figure show the entire mtDNA molecule along the top with PRDM9 motif 1 sites shown underneath as vertical dots (more clearly shown in Figure 4.10 C). The number of PRDM9 motif A 'dots' signifies the frequency of that particular motif site in the sequences used to identify the presence of the motifs, as described in Section 4.3.2 and represented in Figure 4.7. Underneath this are the mtDNA deletions, shown as black horizontal lines spanning from their reported 5' to their 3' breakpoint (Figure 4.10).
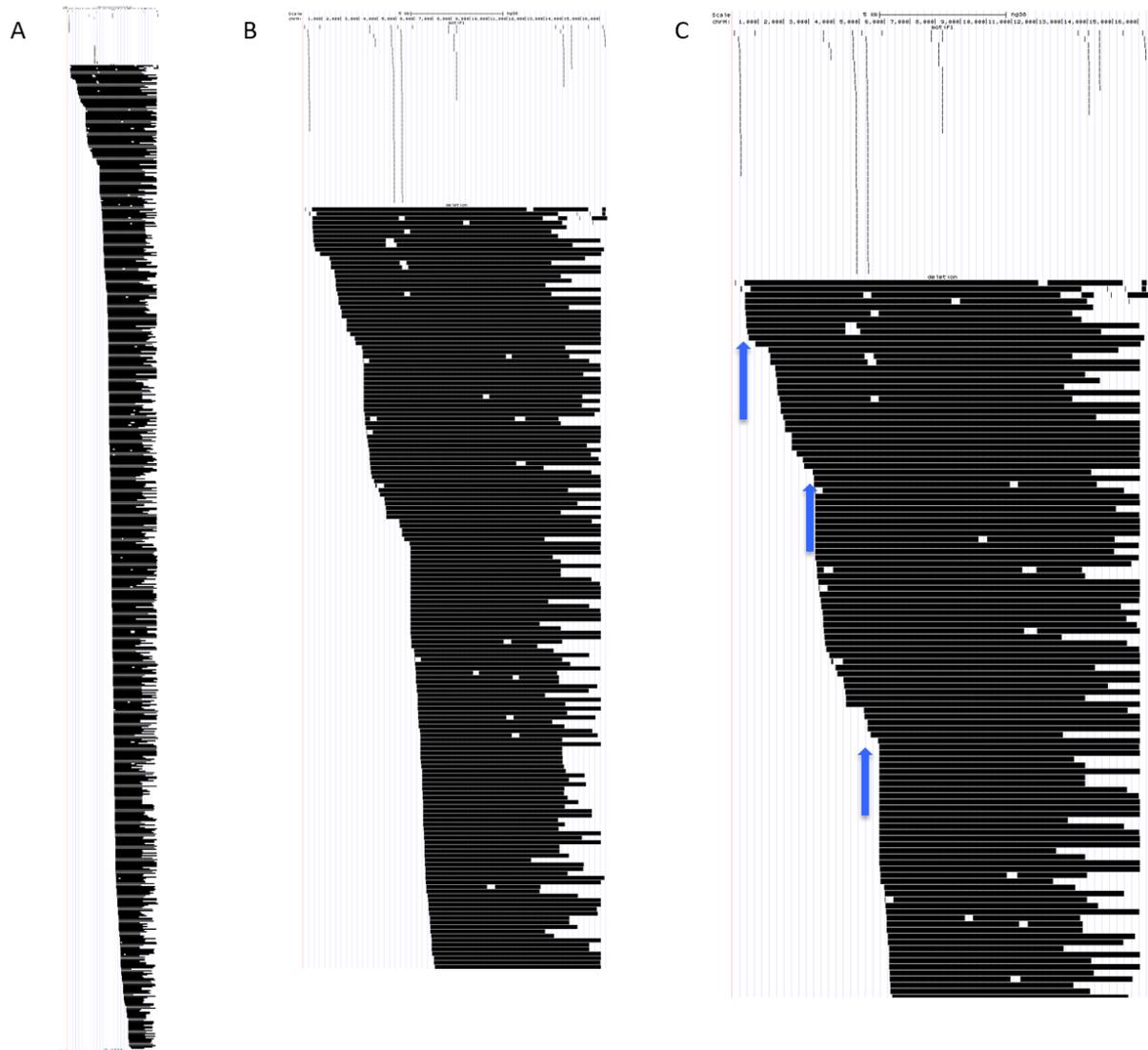
Figure 4.10 Visual representation of the alignment between deletion breakpoints and PRDM9 motif 1 sites within the mtDNA. Images were generated using the UCSC genome browser. A) The whole image generated using this online tool. B) Magnified version of image A. C) Magnified version of image C. Blue arrows represent regions where there is visual alignment of the breakpoints and PRDM9 motif 1 sites.

Images B and C are simply cropped versions of image A to more clearly display the alignment and to view the PRDM9 motif 1 sites. The blue arrows in Figure 4.10 C show regions of the molecule where PRDM9 motif 1 sites align with mtDNA breakpoints. This proof of concept experiment shows that there is a potential relationship between the regions within the mtDNA where PRDM9 motifs have been identified and know mtDNA breakpoint positions. Similar patterns were observed for the other PRDM9 motif sequences searched (Appendix A).

4.5 Discussion

An *in silico* approach was used to investigate whether the PRDM9 binding motif was present in mtDNA sequences from a large dataset. The 13bp degenerate PRDM9 binding motif A (*CCNCCNTNNCCNC*) was identified in almost all mitochondrial genomes searched. In contrast to a previous publication, motif A was not found to exclusively flank the mitochondrial 4977bp 'common' deletion but was present at several positions across the molecule flanking other reported deletion sites (Myers *et al.*, 2008). There were subtle population differences in the motif positions found although the functional significance of these differences remains to be experimentally determined.

A motif site at m.304 is questionable due to the known variability of the poly-C tract present at this position. This poly-C tract is normally 12-18 bp in length with a T nucleotide in the middle at position m.310, known as the D310 variant position. Length variants at D310 have been associated with disease, particularly in a number of cancers (Coto *et al.*, 2011; Alhomidi *et al.*, 2013). Heteroplasmy of D310 length variation has also been described in oocytes from the same mother suggesting that this region is prone to variation probably due to the repetitive nature of the sequence (Marchington *et al.*, 1997).

Interestingly, the motif position at m.5327 was present in 99.9% of sequences searched and flanks a deletion described in the literature (Samuels *et al.*, 2004). Deletions in mtDNA have been reported at positions m.5328, m.5329, m.5330 and m.5331, only a few base pairs downstream of the PRDM9 binding site found in this study. These deletions have all been reported in healthy tissues and remove part of the minor and major arcs with 3' break sites around position m.16076 (Kajander *et al.*, 2000; Samuels *et al.*, 2004). Although these deletions have only been reported for a small number of samples there may be many healthy individuals in the population harbouring these deletions as they do not seem to be associated with any particular disease.

It is unclear as to the role of PRDM9 binding sites in the mtDNA. To ascertain whether there is any correlation between mtDNA deletion breakpoint positions and the PRDM9 motif sites found, analysis was carried out to align these two data sets. Several breakpoints were found to align within or adjacent to the PRDM9 motif sites. When this data was visually displayed using a genome browser tool, there was a pattern of PRDM9 motif sites aligning with the 5' deletion breakpoints.

There are several ways to test the significance of the presence of PRDM9 binding motifs within mtDNA. For example, an algorithm could be constructed and used to test the probability that other randomly generated motif sequences with the same degeneracy as the motif considered here will be found in the mtDNA sequence (Basu *et al.*, 2005; Basha Gutierrez and Nakai, 2016). Another way would be to use a replication cohort. This would require more mitochondrial sequencing data from individuals from several populations, however due to a lack of African mtDNA sequences it would also be beneficial to have more of these sequences in particular. There is an absence of data regarding mtDNA deletion accumulation in African populations due to limited resources as well as difficulties surrounding mitochondrial patient diagnosis. Studies using African mtDNAs have comprised small sample numbers or have been comprised of populations such as South African or African-American which will typically have some representation of European haplogroups and are not representative of the variety of African populations (van der Walt *et al.*, 2012; Silva *et al.*, 2015; Farha *et al.*, 2016; Scheible *et al.*, 2016). The frequency of pathogenic mtDNA mutations, deletions and insertions in African and Asian populations are not as well reported or understood as in European populations (Schaefer *et al.*, 2004).

The discovery of multiple PRDM9 binding motifs within the human mtDNA sequence, as well as the potential pattern of these sites aligning with mtDNA breakpoints, raises the possibility that PRDM9 protein has a functional role within this organelle.

Chapter 5 Genotyping *PRDM9* in patients harbouring mtDNA deletions

5.1 Overview

The mechanisms leading to mtDNA deletion formation remain elusive and are thought to primarily be related to replication or transcription events. Individuals harbouring mtDNA deletions often have mutations in nuclear genes involved in mtDNA maintenance such as *PolG* and twinkle helicase (*PEO1*) (Spelbrink *et al.*, 2001; Van Goethem *et al.*, 2001; Tyynismaa *et al.*, 2004). Due to the frequency of mtDNA deletions and the frequency of variants in these genes within the human population it is reasonable to assume that other genes could be associated with an increased incidence of mtDNA deletions.

The C2H2 zinc finger (ZnF) repeat region of PRDM9 is responsible for binding to DNA motifs through hydrogen bonding between the amino acids of the protein and the nucleic acids present in the major groove of the DNA molecule (Wu *et al.*, 2013). This ZnF region is highly polymorphic in the human population, driven by selection pressures acting to create different recombination sites within individual genomes (Berg *et al.*, 2010).

There are twenty-one *PRDM9* alleles found in the human population (Figure 5.1 B). Allelic differences can be caused by single nucleotide polymorphisms (SNPs) within the ZnF repeat region (Figure 5.1 A) or by the number of repeat regions present (Jeffreys *et al.*, 2013). PRDM9 recognises regions of DNA containing the binding motif which is of interest in a mitochondrial setting due to PRDM9 binding motifs being present in the mtDNA.

A

	Sequence
A	TGTGGACAAGGTTTCAGTGTAAATCAGATGTTATTACACACCAAAGGACACATACAGGGGAGAAGCTCTACGTCTGCAGGGAG
B	TGTGGGCGGGGCTTTAGCTGGAAGTCACACCTCCTCATTCACCAGAGGATACACACAGGGAGAAGCCCTATGTCTGCAGGGAG
C	TGTGGGCGGGGCTTTAGCTGGCAGTCAGTCCTCCTCACTCACCAGAGGACACACACAGGGAGAAGCCCTATGTCTGCAGGGAG
D	TGTGGGCGGGGCTTTAGCCGGCAGTCAGTCCTCCTCACTCACCAGAGGAGACACACAGGGAGAAGCCCTATGTCTGCAGGGAG
E	TGTGGGCGGGGCTTTAGCTGGCAGTCAGTCCTCCTCAGTCACCAGAGGACACACACAGGGAGAAGCCCTATGTCTGCAGGGAG
F	TGTGGGCGGGGCTTTAGCAATAAGTCACACCTCCTCAGACACCAGAGGACACACACAGGGAGAAGCCCTATGTCTGCAGGGAG
G	TGTGGGCGGGGCTTTCGCGATAAGTCACACCTCCTCAGACACCAGAGGACACACACAGGGAGAAGCCCTATGTCTGCAGGGAG
H	TGTGGGCGGGGCTTTAGAGATAAGTCAAACCTCCTCAGTCACCAGAGGACACACACAGGGAGAAGCCCTATGTCTGCAGGGAG
I	TGTGGGCGGGGCTTTCGCAATAAGTCACACCTCCTCAGACACCAGAGGACACACACAGGGAGAAGCCCTACGTCTGCAGGGAG
J	TGTGGGCGGGGCTTTAGCGATAGGTCAAGCCTCTGCTATCACCAGAGGACACACACAGGGAGAAGCCCTACGTCTGCAGGGAG
K	TGTGGGCGGGGCTTTAGAGATAAGTCACACCTCCTCAGTCACCAGAGGACACACACAGGGAGAAGCCCTATGTCTGCAGGGAG
L	TGTGGGCGGGGCTTTAGCTGGCAGTCAGTCCTCCTCAGACACCAGAGGACACACACAGGGAGAAGCCCTATGTCTGCAGGGAG
M	TGTGGGCGGGGCTTTAGAGATAAGTCACACCTCCTCAGACACCAGAGGACACACACAGGGAGAAGCCCTATGTCTGCAGGGAG
N	TGTGGGCGGGGCTTTAGCCGGCAGTCAGTCCTCCTCAGTCACCAGAGGACACACACAGGGAGAAGCCCTATGTCTGCAGGGAG
O	TGTGGGCGGGGCTTTAGAGATAAGTCAAACCTCCTCAGTCACCAGAGGACACACACAGGGACAAGCCCTATGTCTGCAGGGAG
P	TGTGGGCGGGGCTTTAGAGATGAGTCAAACCTCCTCAGTCACCAGAGGACACACACAGGGAGAAGCCCTATGTCTGCAGGGAG
Q	TGTGGGCGGGGCTTTCGCAATAAGTCACACCTCCTCAGACACCAGAGGACACACACAGGGAGAAGCCCTATGTCTGCAGGGAG
R	TGTGGGCGGGGCTTTAGCCGGCAGTCAGTCCTCCTCACTCACCAGAGGACACACACAGGGAGAAGCCCTATGTCTGCAGGGAG
S	TGTAGGCGGGGCTTTAGCTGGCAGTCAGTCCTCCTCACTCACCAGAGGACACACACAGGGAGAAGCCCTATGTCTGCAGGGAG
T	TGTGGGCGGGGCTTTCGCGATAAGTCACACCTCCTCAGACACCAGAGGACACACACAGGGAGAAGCCCTATGTTTGCAGGGAG

B

Allele	Structure	no. repeats
A	ABCDDECFGHFIJ	13
B	ABCDDCCFGHFIJ	13
C	ABCDDCCFKHLHIJ	14
D	ABCDDECFKGHFIJ	14
E	ABCDHFIJ	8
L1	ABCDGHFIJ	9
L2	ABCDDCFGHFIJ	12
L3	ABCDECFGHFIJ	12
L8	ABCDDECCFKHLHIJ	15
L9	ABCDDECFGPFQJ	13
L10	ACDDECFGHFIJ	12
L20	ABCDDECFGKFQJ	13
L24	ABCDDECFTPFQJ	13
L4	ABCDDCCCDDCFKHLHIJ	18
L5	ABCDDECFIJ	10
L6	ABCDDCCCFKHLOIJ	15
L7	ABCDDCCFGHFJ	12
L11	ABCDDECFGHIJ	12
L12	ABCDDCCCFGHFJ	13
L13	ABCDDECFGHQIJ	13
L14	ABCDDCCFKHLOIJ	14
L15	ABCDDCCFKHLHI	13
L16	ABCDRCFKHLHHIJ	13
L17	ABCDDKCCFKHLHIJ	15
L18	ABDDSCFKHLOIJ	13
L19	ABCDDCCFKHLHQIJ	15
L21	ABCDDEECFGHFIJ	14
L22	ABCDDECFGHHFIJ	14
L23	ABCDDECFFHFQJ	13

Figure 5.1 List of all *PRDM9* zinc finger repeats and alleles known in *H.Sapiens*. A) Sequences of each zinc finger, SNPs are highlighted in colour. B) The structure and number of ZnF repeats comprising the ZnF domain for each *PRDM9* allele. Adapted from (Berg *et al.*, 2011).

5.2 Hypothesis

PRDM9 binding sites within the mtDNA are potentially mediating mtDNA deletion formation, therefore, *PRDM9* alleles will be associated with the risk of deletion formation within the population.

5.3 Aims

The principle aim of this study is to amplify, sequence and haplotype the ZnF repeat region of *PRDM9* in patient and control DNA samples and determine if there is an association between any of the *PRDM9* haplotypes and the formation of mtDNA deletions.

5.4 Sample Cohort

Obtaining the case and control DNA cohorts is described in Section 3.1. Further information about the single deletion patient DNA samples is provided in Table 5.1.

ID	DNA number	Tissue Type	DNA conc. (ng)	Genotyped
M0044-13	2013/0049	SKM	66.9	
M005-04	2004/0010	SKM		Yes
M0078-11	2011/0449	SKM	83.6	Yes
M0101-09	2009/0103	SKM	81.8	Yes
M0112-10	2010/0143	SKM	81.8	Yes
M0116-10	2010/0145	SKM	75.6	
M0183-11	2011/0210	SKM		Yes
M0202-09	2009/0175	SKM	47	
M0205-08	2008/0347	NF		Yes
M0216-13	2013/0206	SKM	5.9	Yes
M0221-11	2011/0241	SKM	37.9	Yes
M0223-11	2011/0243	SKM	34	Yes
M0226-12	2012/0289	SKM	192	Yes
M0229-11	2011/0899	SKM	94.2	Yes
M0271-12	2012/0379	CD		Yes
M0271-12	2012/0380	CD	60	Yes
M0271-12	2012/0381	CD	43.1	Yes
M0271-12	2012/0382	CD	58.5	Yes
M0283-03	2003/0319	SKM		Yes
M0283-04	2004/0377	UR		Yes
M0284-09	2009/0234	SKM	31.5	

M0284-12	2012/0329	SKM	136.1	
M0292-11	2011/0924	SKM	72.5	
M0293-11	2011/0923	SKM	142.9	Yes
M0298-10	2010/0317	SKM	144.8	
M0451-09	2009/0412	SKM	95.5	Yes
M0478-11	2011/0527	SKM	49.2	
M0483-11	2011/0499	SKM	44.8	Yes
M0511-10	2010/0496	SKM	65.4	Yes
M0516-11	2011/0502	SKM	14.7	
M0517-11	2011/0503	SKM	23.9	
M0531-09	2009/0498	SKM	46.6	Yes
M0532-08	2008/0461	SKM		Yes
M0560-07	2007/0616	SKM		Yes
M0585-11	2011/0564	SKM	45.5	Yes
M0625-11	2011/0629	SKM	242	
M0626-12	2012/0675	SKM	3.5	
M0653-08	2008/0590	SKM		Yes
M0680-11	2011/0702	SKM	74.9	Yes
M0688-09	2009/0612	SKM	500	Yes
M0734-08	2008/0705	SKM		Yes
M0766-12	2012/0803	SKM	51.6	Yes
M0812-11	2011/0794	SKM		
M0858-11	2011/0827	SKM		
M0936-09	2009/0919	SKM	28.2	Yes

M0941-10	2010/0859	SKM	122	
M0959-12	2012/1020	SKM	53.4	Yes
M0966-10	2010/0940	SKM	53.4	Yes
M0987-12	2012/1022	SKM	59.2	Yes
M1074-10	2010/1051	SKM	14.6	Yes
M1089-09	2010/0023	SKM	153.6	Yes
M1207-11	2011/1381	SKM	137.9	
M1352-12	2012/1360	SKM	7.4	Yes
M1480-12	2012/1465	SKM	6.6	
M1648-12	2012/1627	SKM	14.3	Yes
M0413-04	2010/1253	SKM		Yes
M0124-05	2005/0192	SKM		Yes
M0158-06	2006/0132	SKM		Yes
M0180-03	2003/0244	NF		Yes
M0258-08	2008/0255	SKM		Yes
M0274-05	2005/0352	SKM		Yes
M0288-04	2004/0632	UR		Yes
M0629-07	2007/0686	SKM		Yes
M0919-08	2008/0792	SKM		
M0265-05	2005/0869	SKM		Yes
M0433-10	2010/0172	SKM		Yes
M0999-06	2006/0387	SKM		Yes
M0478-11	2011/0527	SKM	49.2	Yes

Table 5.1 Details of patient DNA samples used in this study. Newcastle Mitochondrial NSCT Diagnostic Service codes are given along with the unique DNA number, tissue type the DNA was extracted from and concentration (if known). CD = cardiac muscle, SKM = skeletal muscle, UR = urine. Successfully sequenced and genotyped samples denoted with 'yes' were included in the analysis.

5.5 Experimental Method

Highly repetitive regions of DNA are known to be difficult to amplify and sequence due to mispriming events leading to PCR artefacts or collapsing of partially amplified oligonucleotides (Stirling, 2003). The ZnF repeat region of the *PRDM9* gene proved to be one such region therefore requiring substantial method optimisation.

5.5.1 Polymerase chain reaction amplification of the zinc finger repeat region of PRDM9

Primers specific to the *PRDM9* zinc finger region (NCBI Reference Sequence: NC_000005.10) were selected based on a previous publication (Berg *et al.*, 2011) which extensively genotyped the *PRDM9* ZnF repeat region in human blood and sperm DNA samples.
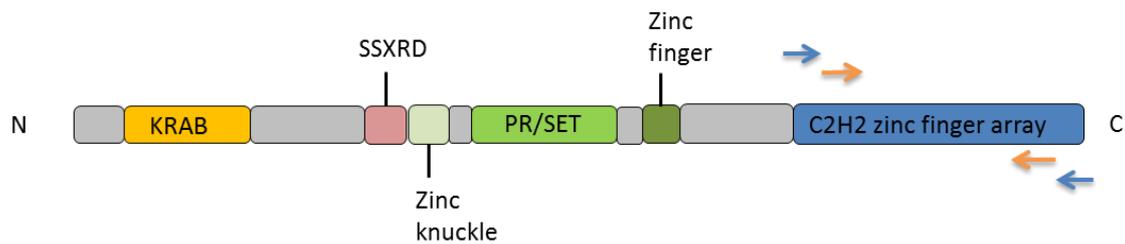
Figure 5.2 Schematic of the functional domains of PRDM9. The C2H2 ZnF domain is at the C-terminus. Primer annealing sites are depicted by blue arrows and nested sequencing primers are in orange.

Primer set 1 (PN0.6F and PN2.5R) was used for initial PCR amplification. Nested primers (PN1.2F and PN2.4R) were then used to sequence the amplicon. Primer annealing sites are shown in the schematic Figure 5.2. Primer sequences are detailed in Table 5.2.

Gene	Primer	Sequence	Application	Product length (bp)	Origin	Annealing temp.
PRDM9	PN0.6F	TGAGGTTACCTAGTCTGGCA	PCR	1864	Berg et al.	62°C
PRDM9	PN2.5R	ATAAGGGGTCAGCAGACTTC	PCR	1864	Berg et al.	62°C
PRDM9	PN1.2F	TGAATCCAGGGAACACAGGC	Sequencing	1257	Berg et al.	50°C
PRDM9	PN2.4R	GCAAGTGTGTGGKGACCACA	Sequencing	1257	Berg et al.	50°C

Table 5.2 Details of primers used to amplify and sequence PRDM9 amplicons.

The *PRDM9* PCR amplicon is 1864 bp in length and is highly repetitive. Initial PCR was attempted using Immolase Taq polymerase protocols (Bioline, London, UK). However, due to the repetitive nature of this DNA fragment, long-range high fidelity PCR techniques were required for sufficient amplification. TaKaRa protocols were optimised using healthy control DNA as template.

Temperature gradients were used to determine the optimum annealing temperature of the *PRDM9* primers (Figure 5.3). PCR reaction mixture was as follows; 0.5 mM betaine, 200 µM dNTPs, 1.25 U TaKaRa LA Taq polymerase, 0.2 µM primers (forward and reverse), dH$_2$O and 100 ng DNA made to final volume of 25 µL with nanopure water. Thermocycling conditions were: denaturation at 94 °C for 1 min, 98 °C for 10 secs, annealing at 62 °C for 15 min, for 30 cycles followed by a final extension for 10 min at 72 °C. PCR conditions were optimised using a control DNA sample (Figure 5.3). *PRDM9* amplicons were visualised via electrophoresis and UV exposure, showing highly specific amplification of one clear band at 1900 bp representing the full length ZnF region of *PRDM9* (Figure 5.3).
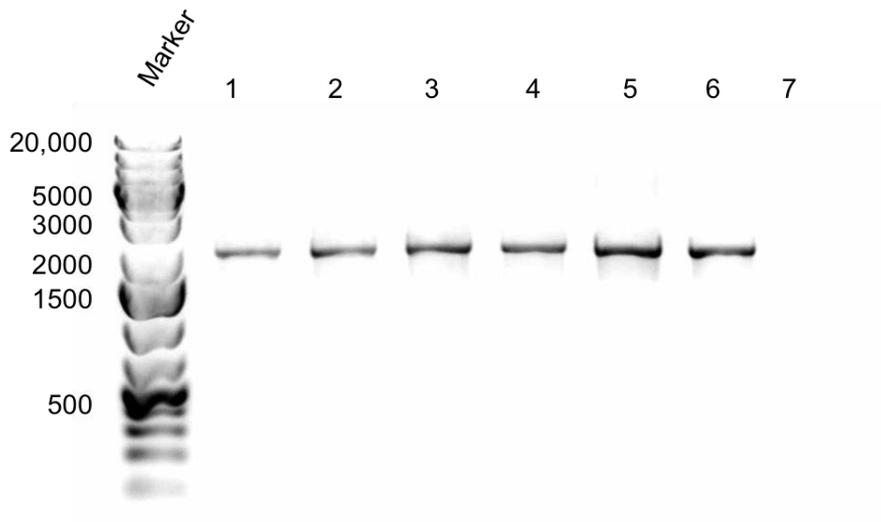
Figure 5.3 Gradient PCR of the *PRDM9* ZnF region. Control DNA was used as a template. Annealing temperatures were as follows; lane 1 55°C, lane 2 57°C, lane 3 59°C, lane 4 60°C, lane 5 62 °C, lane 6 64 °C, lane 7 negative control. *PRDM9* PCR products were electrophoresed on a 1% agarose gel.

Single deletion patient DNA samples were then amplified using this reaction (Figure 5.4). Lanes 1 to 13 of Figure 5.4 show specific PCR amplification of PRDM9. Individuals with ZnF length differences could be distinguished by the presence of additional bands on the electrophoresis, for example lane 9 (Figure 5.4) which appears is heterozygous for the haplotypes A and a shorter allele.
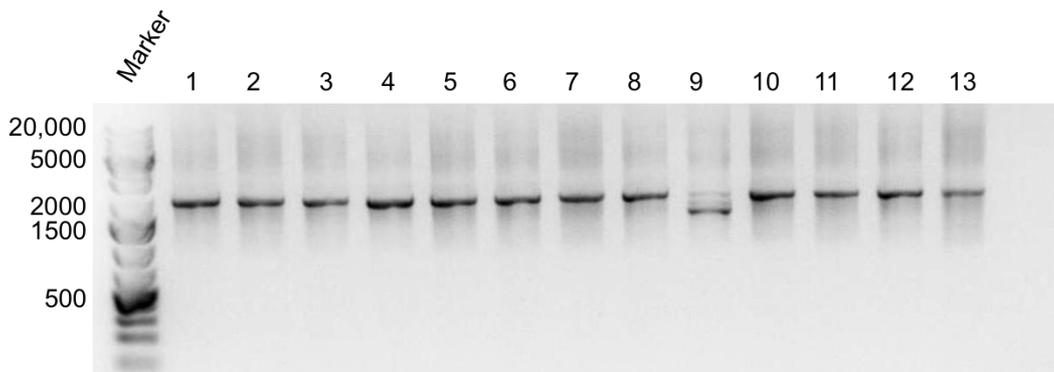
Figure 5.4 PCR of the *PRDM9* ZnF region in single deletion patient samples. *PRDM9* PCR product was visualised after electrophoresis on 1% agarose gels. A molecular weight marker was used to size the products (far left lane). Sizes are denoted in base pairs.

The size of the fragments was used to identify individual samples that could be heterozygous for the ZnF repeat region of *PRDM9.* For example, PCR products in lanes 5 and 19 (Figure 5.5) clearly have additional bands to the 1900 bp expected product. These individuals are potentially heterozygous for haplotypes A and a longer allele.
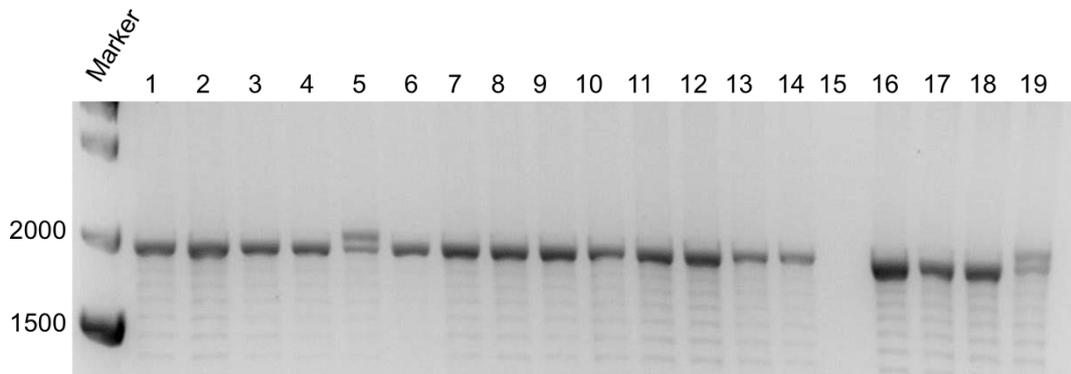
Figure 5.5 PCR of the *PRDM9* ZnF region in single deletion patient samples. *PRDM9* PCR product was visualised after electrophoresis on 1% agarose gels. A molecular weight marker was used to size the products (far left lane). Sizes are denoted in base pairs.

Samples typically had faint bands at lower molecular weight such as the samples in Figure 5.5. These bands are probably fragments which have collapsed during the amplification due to the complexity of this repeat region. Mispriming events may also produce a small amount of shorter product detectable by gel electrophoresis.

5.5.2 Sequencing and molecular cloning of single deletion patients

Nested primers specific to the *PRDM9* amplicon were used to sequence the ZnF repeat region, 1257 bp in size (Table 5.2). Initial Sanger sequencing reads were only 600 bp in length due to the capacity of the capillary used on the ABI3130xl Genetic analyser. Using a forward primer in one sequencing reaction and a reverse primer in a separate, but identical, reaction, 1200 bp reads could be analysed for each individual DNA sample. Sequencing data was analysed using Seqscape sequencing software aligning to the reference *PRDM9* sequence (NCBI Reference Sequence: NC_000005.10).

112
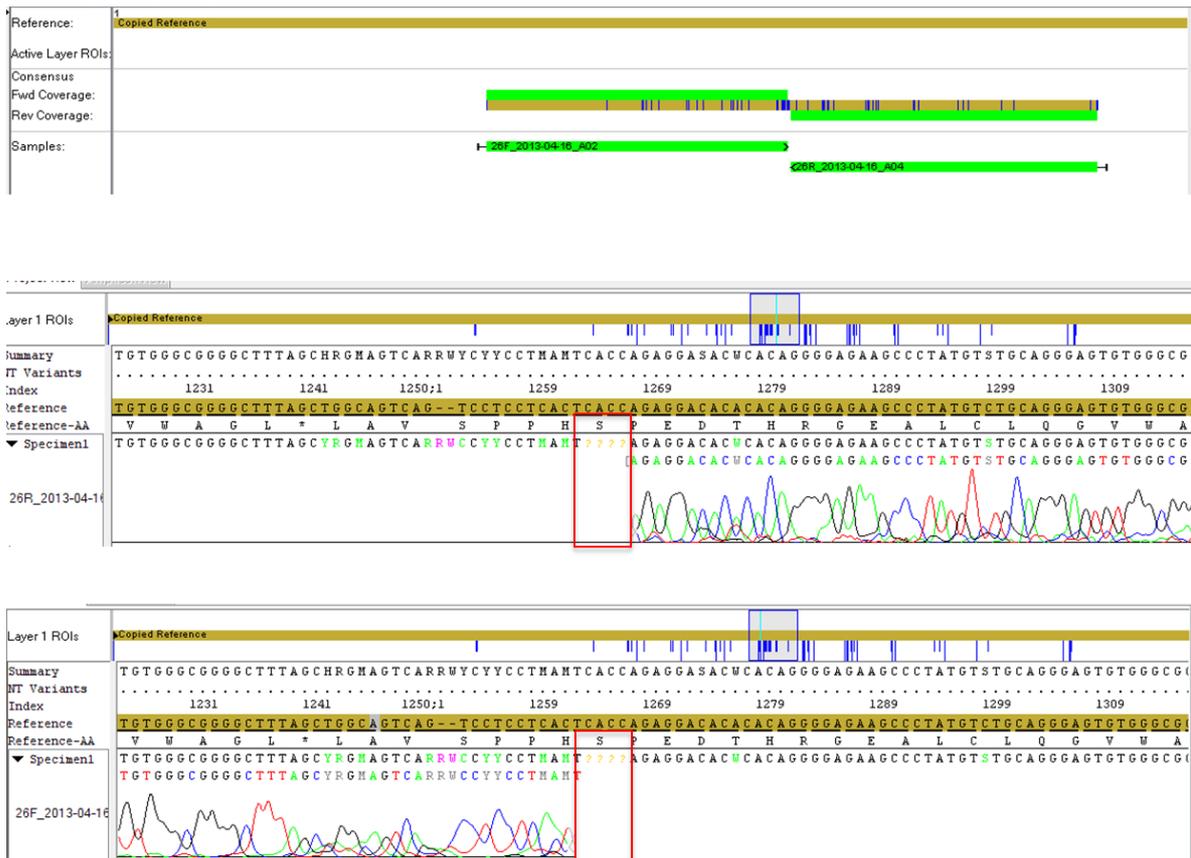
Figure 5.6 Electropherograms of the *PRDM9* ZnF repeat in a representative DNA sample. Top image; alignment of forward and reverse amplicons to the reference sequence entered into the Seqscape software. Middle image; sequencing trace for the reverse primed amplicon. Bottom image; sequencing trace for the forward primed amplicon.

Depending on the quality of the sequencing there was a gap spanning 6 to 56 nucleotides as highlighted by the red box in Figure 5.6. This gap was in the middle of the region of interest where the most allele defining repeats lie. In addition, sequencing traces usually became less efficient towards the end of the amplicon due to technical limitations of the machine in detecting nucleotides effectively. Electropherograms lose quality towards the end of forward and reverse strand reads and therefore much of the middle section of the *PRDM9* amplicon was impossible to genotype. In this instance genotyping was not possible.

To modify this protocol the capillary array length in the sequencing machine was changed to a longer array allowing longer high quality and highly accurate nucleotide calls along the read length. This allowed sequencing of almost the entire amplicon for each primer used and allowed comparisons to be made between the sequence obtained

113

with the forward and reverse primers individually. This array allowed for overlap in the sequencing traces obtained from each individual primer so that variants could be confirmed using both forward and reverse reads (Figure 5.7).
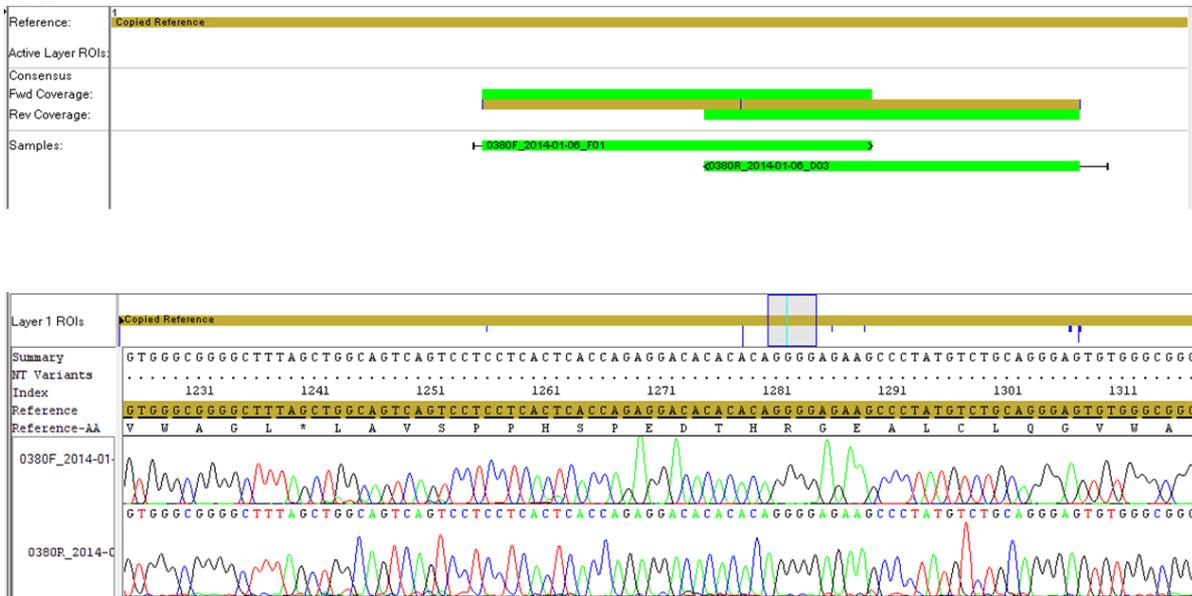
Figure 5.7 Electropherograms of the *PRDM9* ZnF repeat in a representative DNA sample using a longer sequencing capillary. Top image; alignment of the forward and reverse amplicons to the reference sequence entered into the Seqscape software. Bottom image; sequencing traces of forward and reverse amplicons over the middle region of the ZnF repeat domain.

Although electropherograms were of high quality using the longer array, there were still artefacts observed in samples sequenced using this method. Figure 5.8 shows a typical sequencing trace for one of the ZnF repeats used to genotype the individual. Multiple peaks are present at each position making it difficult to define a real heterozygous signal from noise.
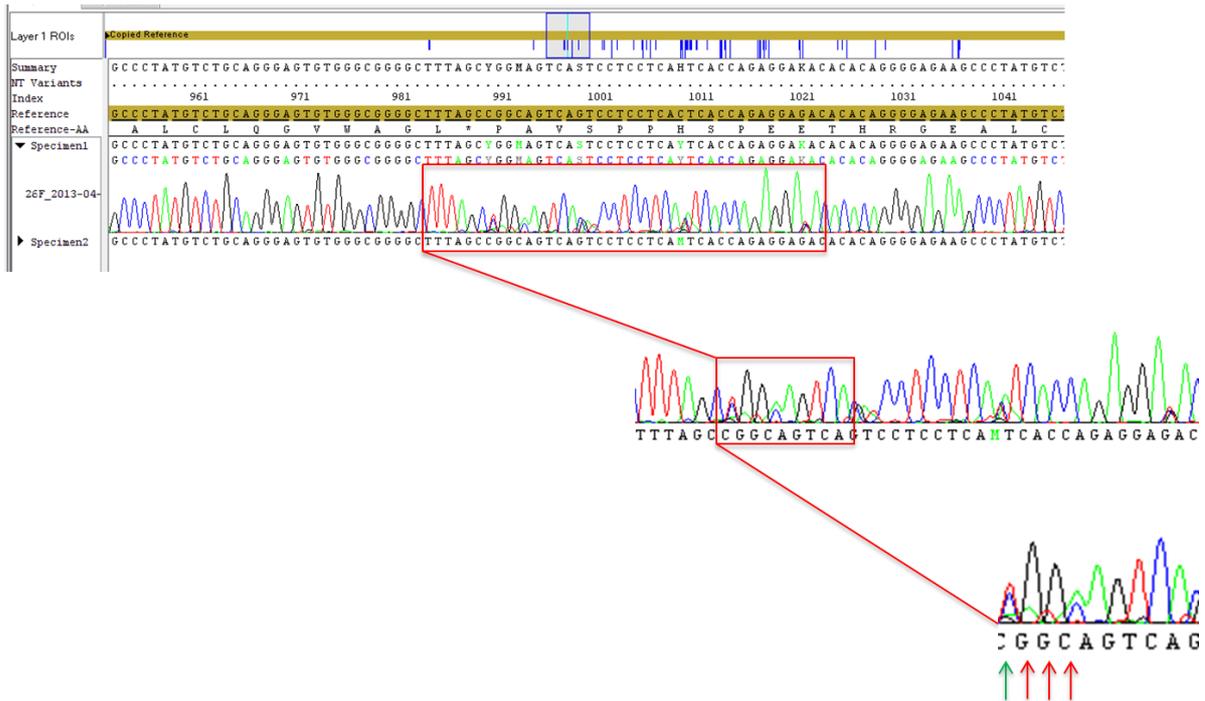
Figure 5.8 Electropherogram of the *PRDM9* ZnF region. Top image; sequencing trace of the PRDM9 ZnF repeat in a representative DNA sample. The 150 base pair window is shown at the top of the image. Below are magnified images of the sequencing traces depicted in the red boxes.

The green arrow in Figure 5.8 shows a base which would be considered as heterozygous T/C due to the peaks being of equal height. The red arrows depict bases where there are two peaks present but the expected bases GG are not reduced in height, or where the expected C nucleotide is under the signal for an A nucleotide.

Aligning DNA sequences from individuals that were heterozygous for repeat length variation in the *PRDM9* allele proved difficult to genotype as the Seqscape software was unaware of the length differences. To overcome this I created reference sequences using only the 'A' or 'J' repeats listed in Figure 5.1. Figure 5.9 shows the misalignment of one such DNA sample that was heterozygous for a 13 repeat length allele and a longer allele. Variant calling in blue shows the number of base pair calls that did not match the reference sequence due to the 'insertion' of an extra ZnF repeat sequence.
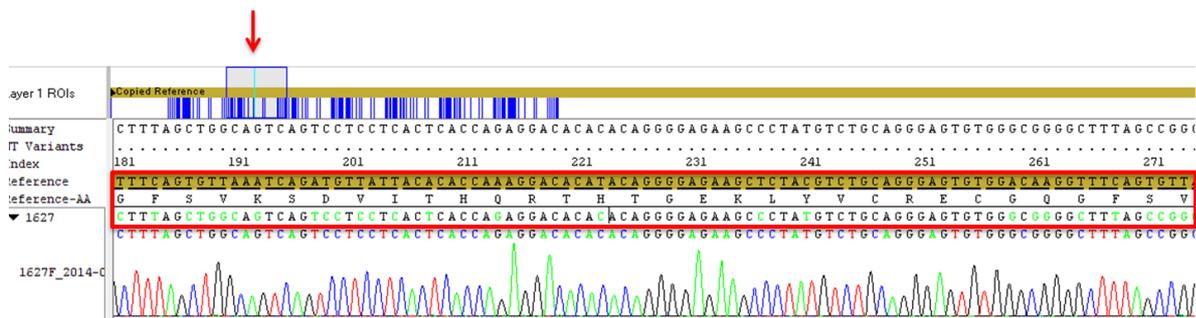
Figure 5.9 Electropherogram of the PRDM9 ZnF region in a patient with a heterozygous extended repeat allele. The red arrow shows variant calls highlighted in blue along the length of the amplicon analysed. The red box shows mismatching of the sequencing calls (green base pairs) to the reference sequence (black base pairs) in the sample.

The 'A' repeat is present in every PRDM9 allele as the first repeat in the sequence and the 'J' allele is always the last repeat of the allele with the exception of allele L15 (Figure 5.1). Forward sequencing files were aligned to a continuous 'A' repeat reference and reverse sequencing files to the continuous 'J' reference. For most individuals harbouring length variants this strategy allowed accurate genotyping.

To validate the allele status assigned to each DNA sample, molecular cloning of each allele was performed. *PRDM9* was amplified in the patient DNA samples. PCR products were electrophoresed, gel extracted and ligated into the pGEM-T Easy Vector. Colonies containing a single allele were picked and *PRDM9* was amplified using primers annealing to the pGEM-T vector sequence (Table 5.3).

Target	Primer Sequence	Annealing Temp	Source
p-GEM vector Reverse	5´-d(TCACACAGGAAACAGCTATGAC)-3´	62 °C	Promega Cat. No. Q5421
p-GEM vector Forward	5´-d(CGCCAGGGTTTTCCCAGTCACGAC)-3´	62 °C	Promega Cat. No. Q5601

Table 5.3 Details of primers used to amplify the *PRDM9* insert sequence in the p-GEM vector sequence used to clone the amplicon.

Colony PCR reaction mixture was as follows; 0.5 mM betaine, 200 µM dNTPs, 1.25 U TaKaRa LA Taq polymerase, 0.2 µM primers (forward and reverse), 10 µL dH2O containing a colony stab. Thermocycling conditions were: denature at 94 °C for 1 min, 98 °C for 10 secs, annealing at 62 °C for 15 min, for 30 cycles followed by a final extension for 10 min at 72 °C.

PRDM9 amplicons were successfully cloned as confirmed by colony PCR (Figure 5.10). Length variants can be seen in lanes 5 and 8 (Figure 5.10).
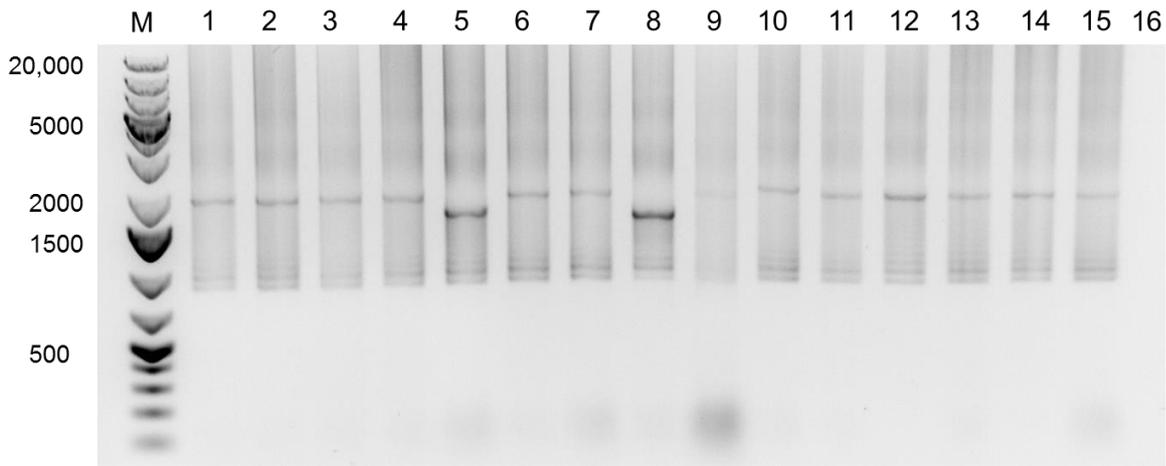
Figure 5.10 Colony PCR followed by gel electrophoresis of PRDM9 amplicons from single deletion patients. PCR products were electrophoresed on a 1% agarose gel. M corresponds to the molecular weight marker. Lanes 1-15 contain patient PRDM9 PCR product, lane 16 is a negative control PCR product.

The PCR products were then sequenced as described above. Sequencing traces appeared to be of good quality (Figure 5.11) and length variants were genotyped using a different reference sequence as described above.
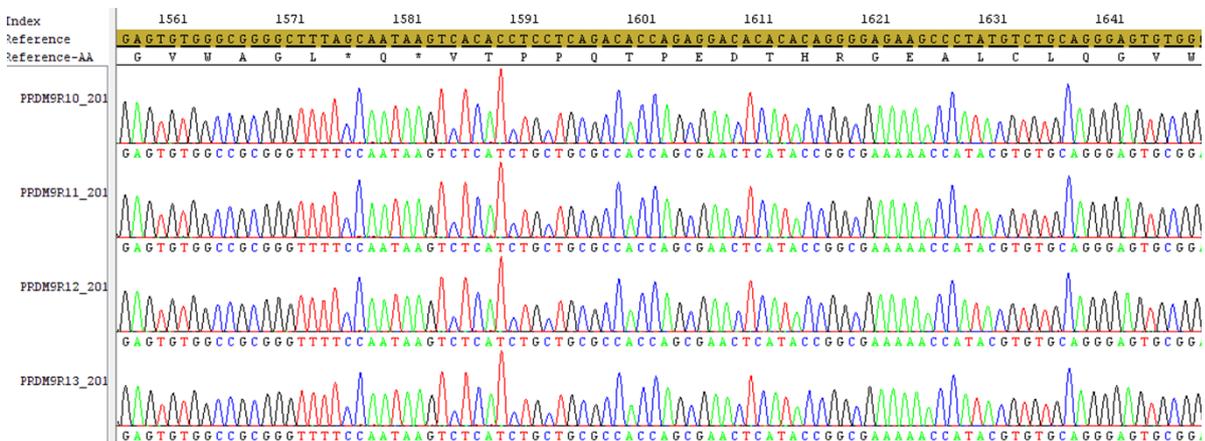
Figure 5.11 Sequencing traces of the ZnF repeat region from colony PCR DNA samples. Representative 150bp section is shown.

Sequencing analysis was performed as before and the results were documented in a separate file to the original results. In addition, a second experimentally blind invigilator carried out the genotyping of the sequencing traces obtained and the allele status results

118

were compared. Both assessors agreed on the vast majority of the genotypes showing that there was no analysis bias.

5.6 Results

5.6.1 Statistical analysis of genotyping results

Overall, 8 *PRDM9* alleles were detected in the case group and 6 in the control group. Chi-squared comparisons were performed on the allele counts to determine if there was any statistically significant difference between the groups.

	Cases		Controls		
Allele		100		96	P value
A	15	85	13	83	0.8397
B	97	4	94	2	0.6833
C	98	2	95	1	1
D	99	1	96	0	1
E	99	1	96	0	1
L9	98	2	94	2	1
L20	99	1	95	1	1
L24	96	4	93	3	1
		100		92	
	Students t-test		0.9462064		

Table 5.4 Distribution of alleles in case and control groups. Total allele counts are shown for each group. The number of each allele type is shown in the right hand columns versus all other alleles (left hand columns). P values were calculated using Chi squared and two-tailed Fishers exact test or a two-tailed Students t-test.

Total allele counts were 100 and 96 for the case and control groups respectively (50 cases versus 48 controls). Firstly, alleles were counted for each group and a Students t-test performed on the allele distribution (Table 5.4). Of the total alleles 85% of case and 87% of control cohort alleles were 'A'. This was representative of the population

frequency of the 'A' allele published by Berg et al., 2012 as 85.7% in Europeans. The distribution of the other alleles found in the cohorts was also similar to those previously published for the population. Cases and controls had a similar distribution of alleles (Two-tailed Students t-test, p = 0.9462).

Alleles were grouped into allele 'A', allele 'B' and 'Rare' (all other alleles). There was no statistically significant difference between the number of A, B or Rare alleles in cases versus the control group (Table 5.5).

	Cases 100		Controls 96		P value
A	15	85	13	83	0.8397
B	96	4	94	2	1.00
Rare	89	11	89	7	0.4609

Table 5.5 Allele counts for case and control groups. Alleles were grouped as 'A', 'B' or 'Rare' referring to any allele that was not A or B genotype. P values were calculated using Chi-squared and two-tailed Fisher's exact test.

Allele length variation was tested by grouping alleles into 'Short' <13 repeats, 'Normal' 13 repeats or 'Long' >13 repeats. There was no statistically significant difference between the number of Short, Normal or Long alleles in cases versus controls (Table 5.6).

		Cases 100		Controls 96		P value
E <13	Short	99	1	96	0	1.00
A	Normal	4	96	1	95	0.3691
C,D >13	Long	97	3	95	1	0.6214

Table 5.6 Allele counts for case and control groups. Alleles were grouped by ZnF repeat length. P values were calculated using Chi-square and two-tailed Fisher's exact test.

Finally, the distribution of individuals who were homozygous 'A', heterozygous 'A/N' or homozygous 'N/N' (where N is any other allele) was tested. All samples apart from one control individual were homozygous or heterozygous for the 'A' allele, consistent with both patient and control DNA samples being collected in the UK and assumed to be of European descent. Differences in genotype were statistically significant in the cases versus controls (Table 5.7). There were significantly more (p=0.0047) heterozygous individuals in the case group.

		AA	AN	NN	P value
Cases	100	85	15	0	0.0047
Controls	96	88	7	1	
Cases	100	85	15		0.0097
Controls	96	88	8		

Table 5.7 Allele counts for case and control groups. Alleles were grouped by genotype 'AA', 'AN' or 'NN'. P values were calculated using Chi squared test.

5.6 Discussion

PRDM9 binding motifs are present at multiple sites along the mtDNA sequence, leading us to hypothesise that there may be a functional role for PRDM9 in the mitochondria. The ZnF region of PRDM9 is responsible for binding to DNA and is a highly polymorphic region. This ability to rapidly alter the nucleotide sequence of the ZnF domain allows PRDM9 protein to adapt to shifts in the recombination hotspots of the nuclear genome between generations. Loss of this ability leads to a lack of viable offspring and ultimately speciation events (Myers *et al.*, 2010; Flachs *et al.*, 2014). The possible role of PRDM9 in mitochondria lead us to hypothesise that *PRDM9* alleles may increase susceptibility to mtDNA deletion formation in individuals. Access to mitochondrial patients harbouring single deletions provided us with a means to perform a gene screen to address whether any *PRDM9* allele is over-represented in this cohort.

The ZnF region of the human *PRDM9* gene was successfully genotyped in 50 mtDNA single deletion patient and 48 healthy control DNA samples. The distribution of *PRDM9* alleles found in both case and control groups was indicative of the population

121

frequencies previously published, validating the European nature of our DNA samples. There was no statistically significant association between *PRDM9* allele status and the presence of mtDNA single deletions. There were significantly more heterozygous individuals in the case group versus controls (p=0.0047). However based on the results in Table 5.4, there does not seem to be any particular *PRDM9* allele associated with this observation. The alleles most common after 'A' are 'B' and 'L24' but the frequency of these alleles was not statistically different between case and control groups (Table 5.4).

There are several reasons for this observation. Firstly, the number of individual samples in both the case and control groups was small (50 and 48 respectively). An association between a particular SNP or genotype and the disease being studied would have to be strong to be significant with such a small sample size. Using the parameters; a significance level (*p* value) of 0.05 and an effect size of 0.5 (Cohen's *d* = medium), statistical power calculations show that 130 individual samples would be required for an 80% chance of statistical significance. A smaller effect size would require even more samples suggesting that this study was underpowered.

Secondly, this genomic region is highly polymorphic and appears to be under selective pressures acting to recombine the genome and allow for higher genetic diversity between individuals within the population. This means that the individual zinc finger repeat units that comprise the *PRDM9* ZnF domain are not highly conserved and are therefore not predicted to have a detrimental effect when mutated. This is also possible because of the ability of *PRDM9* to adapt its amino acid sequence to bind slightly different DNA motif regions. Heterozygous individuals will more than likely have subtle shifts in recombination events at the genome level compared to homozygous individuals but this does not necessarily have a negative impact on the health of the individual. In addition, if PRDM9 protein binds to mtDNA then it must be able to recognise a wide range of mtDNA haplogroups. It is exclusively the maternal mtDNA haplogroup which is inherited by the offspring however, there are four possible *PRDM9* alleles (two maternal and two paternal) leading to potentially dramatic genomic recombination shifts. If PRDM9 could only interact with some subsets of mtDNA sequences then there would be very little chance of viable offspring from individuals with different *PRDM9* allele status and mtDNA haplogroups. Since both *PRDM9* haplotype and mtDNA haplogroup are associated with global population location then this would lead to possible speciation

events within the human population. This is unlikely as genetic theories determine that it is advantageous to outbreed and increase individual and population fitness.

Several patient samples shown in Table 5.1 failed to amplify and therefore could not have a *PRDM9* genotype assigned. This is likely due to the quality of the DNA samples. The *PRDM9* ZnF domain is a highly repetitive PCR product and therefore proved difficult to amplify. Successful amplification relies on good quality starting template DNA in terms of purity as well as concentration. In addition to ensure that the PCR reaction itself is successful, most of the amplicon produced in each amplification cycle must be the full-length product so that miss-priming doesn't occur. The longer and more repetitive a DNA fragment, the more likely it is that the copy of the template DNA will collapse during the extension phase to produce smaller fragments than the intended amplicon.

Analysis of the *PRDM9* ZnF repeat region also proved difficult due to the repetitive nature of this amplicon. Genotype specific SNPs were highlighted in Berg et al., 2011. Due to the polymorphic nature and complexity of this genomic region it was often difficult to assign a genotype to each repeat and sometimes appeared to be somewhat subjective. Mostly this was due to the quality of the DNA sequence obtained for each individual DNA sample. Any heterozygous signals were scrutinized and assessed for likelihood of the assigned nucleotide, that is to say, if there was no reported alternative in the literature or in a database then the wild type nucleotide was assumed to be homozygous. To ensure that the genotyping was not wholly subjective an independent marker was used to verify the sequencing and assign a genotype to the samples. The complexities of the *PRDM9* ZnF repeat region proved technically challenging for genotyping and allele association. Although there was a significant difference between the numbers of heterozygous individuals in the case cohort it is unlikely that this result is truly informative.

Larger cohorts of mtDNA deletion patients are required to confirm this finding in combination with relevant functional studies. It would be particularly interesting to haplotype trios including the parents of the affected individual to identify if there is a trend in the likelihood that one parent might pass on a particular haplotype to their affected offspring. This was the case in a study where mothers were statistically more likely to carry the associated *PRDM9* allele present in her affected child (Woodward *et*

al., 2014). With the increasing availability and affordability of next generation sequencing (NGS) technologies it would perhaps also be more efficient to haplotype *PRDM9* amplicons using a long range sequencing protocol. This would allow for single molecule sequencing reads that have more accurate base calling and can be haplotyped with higher throughput than the method presented in this chapter.

In conclusion, *PRDM9* alleles were genotyped in a single deletion patient cohort and compared to healthy controls. There was a significant increase in the number of individuals carrying heterozygous *PRDM9* alleles in the patient cohort. However, no single allele was over-represented indicating that *PRDM9* genotype is unlikely to affect mtDNA deletion risk. Functional studies are required to experimentally determine whether PRDM9 interacts with mtDNA *in vivo*.

Chapter 6 PRDM9 expression in human tissue and cell lines.

6.1 Overview

In order to understand the relevance of PRDM9 binding motifs found in mtDNA sequences, PRDM9 protein expression was investigated in human cell lines and tissue samples. PRDM9 remains an elusive protein with previous studies focussing on *Mus musculus* or partial recombinant human proteins (Hayashi *et al.*, 2005; Mihola *et al.*, 2009; Eram *et al.*, 2014; Davies *et al.*, 2016; Patel *et al.*, 2016). Although these studies have confirmed the methyltransferase activity of the protein as well as its ability to direct DNA cross-over events, the true functional role of full length human PRDM9, as well as interacting partner proteins, timing of expression and cellular localisation remains unknown. Although murine models are of importance to our understanding of molecular pathways and interactions, the PRDM9 proteins differ between mouse and human at the amino acid sequence level and thus might function differently *in vitro*. In addition, *PRDM9* has been described as the primary genetic candidate for speciation events. This is suspected to be due to its polymorphic nature and motif binding ability, which leads to a genetic complementation barrier resulting in non-viable zygotes (Berg *et al.*, 2010; Jeffreys *et al.*, 2013; Flachs *et al.*, 2014). For this reason, cross species comparisons of PRDM9 protein functions are helpful and necessary but investigations using human PRDM9 are warranted to ascertain its function.

The human PRDM9 protein is not predicted to have a mitochondrial targeting sequence according to an *in silico* prediction tool (MitoP2 Elstner *et al.* (2009)) and does not appear on the extensive mitochondrial proteome database (MitoCarta2.0 Pagliarini *et al.* (2008); Calvo *et al.* (2016)). Due to the discovery of PRDM9 binding sites within the mitochondrial genome, it would be important to explore PRDM9 protein localisation in the context of the mitochondria. To investigate the expression of PRDM9, we had unique access to a range of difficult to obtain human tissue samples as well as both primary and immortalised cell lines.

6.2 Aim

The aim of this study was to identify PRDM9 protein in human tissue and cell lines using a variety of detection methods. This would provide an appropriate model to study the localisation of this protein, in particular, to elucidate any mitochondrial function.

6.3 Tissue samples and cell lines

Details of the tissue samples used in this study are provided in Table 6.1.

Tissue	Log no.	Stage	Storage	Source	Disease
Human Foetal Gonad	HDBR#11942	11pcw	-80 °C	HDBR	N/A
Human Placenta	HDBR#12168	5pcw	-80 °C	HDBR	N/A
Human Placenta	HDBR#12169	6pcw	-80 °C	HDBR	N/A
Human Placenta	HDBR#12170	10pcw	-80 °C	HDBR	N/A
Human Skeletal Muscle	PFC_H_351_126	23yrs	-80 °C	MRG Tissue Resource	N/A
Human Ovary (Left)	PFC_H_02_07	50yrs	-80 °C	MRG Tissue Resource	MELAS
Human Ovary (Right)	PFC_H_01_07	50yrs	-80 °C	MRG Tissue Resource	MELAS

Table 6.1 Details of tissue samples used in the study. Tissue type, identification number, age, storage conditions, source and disease state are provided.

Details of the cell lines used in this study are provided in Table 6.2.

Cell Line	Type	State	Source
HEK293	Human Embryonic Kidney	Immortalised	Prof. R.N. Lightowlers
143B	Osteosarcoma	Immortalised	Dr A. Gómez-Durán
NT2	Neuron-committed Teratocarcinoma	Immortalised	Dr A. Gómez-Durán
H9	Human Embryonic Stem Cell	Stem cell	Prof. M. Lako
AM	Fibroblast	Primary	Newcastle Biobank
F011	Fibroblast	Primary	Newcastle Biobank
M014	Myoblast	Primary	Newcastle Biobank
M062	Myoblast	Primary	Newcastle Biobank
M235	Myoblast	Primary	Newcastle Biobank
M259	Myoblast	Primary	Newcastle Biobank

Table 6.2 Cell lines used in this study. Cell line, cell type, state and source are provided. Professor R.N. Lightowlers, Institute of Neuroscience: Dr A. Gómez-Durán and Professor M. Lako, Institute of Genetic Medicine, all Newcastle University.

Detailed methods and a full list of antibodies used in this study are provided in Methods Sections 3.6, 3.8 and Table 3.2.

6.4 Results

6.4.1 Detecting PRDM9 in cell lines

Available cell lines were investigated for the presence of PRDM9 protein by gel electrophoresis followed by Western blotting with a commercially available PRDM9 antibody (Abcam, Cambridge, UK). To assess which cell model would be most appropriate; primary myoblast, primary fibroblasts and immortalised HEK293 cell lysates were run alongside a human testes (HuT) protein lysate sample (Figure 6.1). Testes tissue lysate was a kind gift from Professor David Elliot (Institute of Genetic Medicine, Newcastle University). After 60 minutes of exposure there was a faint signal detected in all sample lanes however there were three bands visible at ~90, ~65 and ~40 kDa. There was no protein detection in the human testes lysate using 4 µL of lysate of unknown concentration, however GAPDH signal was also low suggesting that there might not have been a high enough quantity of protein used to detect PRDM9.
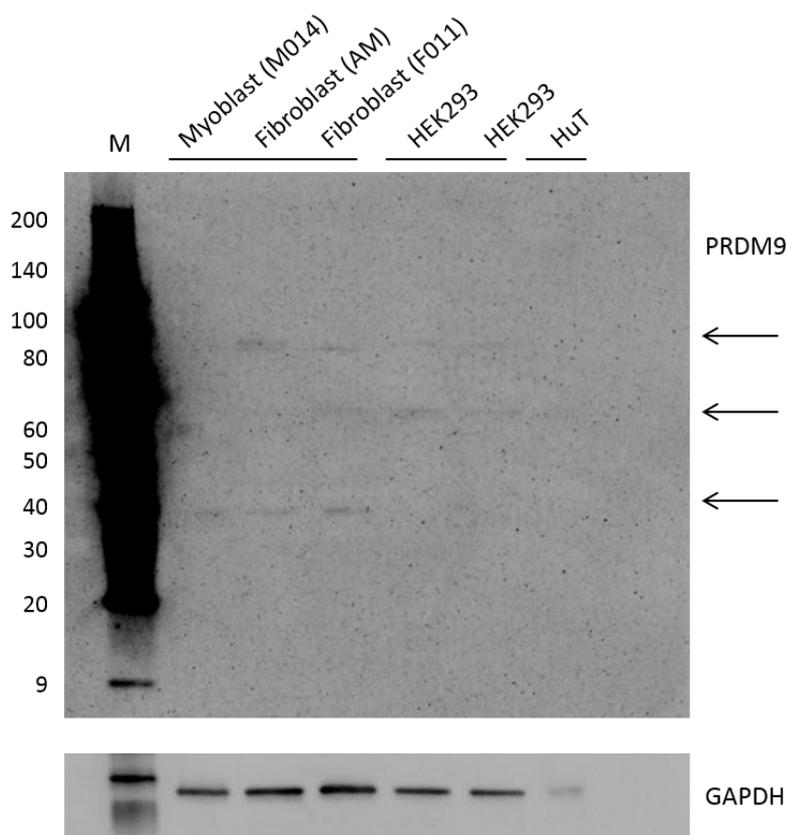
Figure 6.1 Western blot detection of PRDM9 in different cell lines. Protein size marker was included (M), sizes are provided in kDa. 20 µg of myoblast, fibroblast and HEK293 cell lysate was loaded on a 4-20% bis-tris gel alongside 4 µL human testes lysate of unknown protein concentration (HuT). Membrane was incubated with anti-PRDM9 (1:500 dilution) or GAPDH (1:1000 dilution) overnight at 4°C.

In order to determine whether the faint bands observed in Figure 6.1 were PRDM9, larger amounts of protein were required. HEK293 cells were chosen as the most appropriate model because immortalised cells typically divide at a much faster rate without replicative senescence and provide more cellular material than primary cell lines (Hayflick, 1965). Optimal protein concentration for detection was performed by running concentration gradients (Figure 6.2). Double the volume (8 μL) of human testis tissue lysate was used compared to the previous blot (Figure 6.2 A). There was a single band corresponding to a peptide slightly larger than the predicted molecular weight of 103 kDa in both the HEK293 and human testes sample lanes. There was also a band at ~60 kDa in the highest concentration HEK293 sample lanes. Detection of these bands was very weak and required long exposure times. Next, a higher protein concentration gradient of HEK293 cell lysate was tested (Figure 6.2 B). Bands were observed at ~100 and ~60 kDa but additional bands were also seen at ~120 and ~30 kDa. Thus, loading higher amounts of protein onto the gel gave more non-specific banding when using the anti-PRDM9 antibody to detect protein levels.
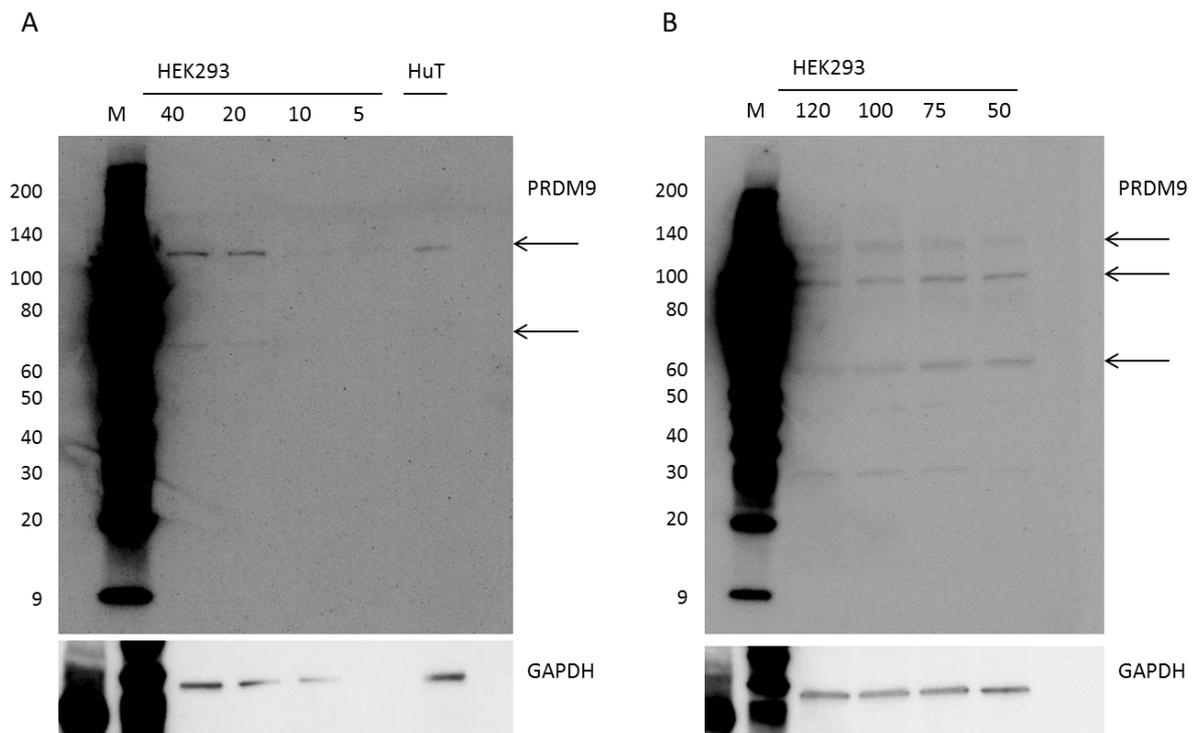
Figure 6.2 Western blot detection of PRDM9 in HEK293 cell and human testes lysate. Protein size marker was included on each blot (M), sizes are denoted in kDa. A) HEK293 cell lysate gradient from 40-5 μg protein was loaded on a 4-20% bis-tris gel alongside 8 μL human testes lysate of unknown protein concentration (HuT). B) HEK293 cell lysate gradient from 120-50 μg was loaded on a 4-20% bis-tris gel. Membrane was incubated with anti-PRDM9 (1:500 dilution) or GAPDH (1:1000 dilution) overnight at 4°C.

Due to the embryonic nature of PRDM9 expression, the human embryonic stem cell line
H9 (H9 hESC) was investigated for PRDM9 protein expression (Figure 6.3). The HEK293
cell lysate was immunoblotted alongside H9 as a comparison. As expected from previous
results, HEK293 lysate gave bands at ~120 and ~60 kDa in size as well as faint bands at
several other molecular weights. Bands were observed at ~60, ~50 and ~20 kDa in the
H9 sample lanes. The band at ~120 kDa and a weaker band at ~80 kDa observed in the
HEK293 samples were not present in the H9 samples. Overall, this suggests that the
HEK293 and H9 cell lines have completely different protein expression profiles using the
commercially available PRDM9 antibody in this study. Importantly, the band detectable
at a molecular weight close to that of PRDM9 (103 kDa) was observed in the HEK293
sample lysate only.
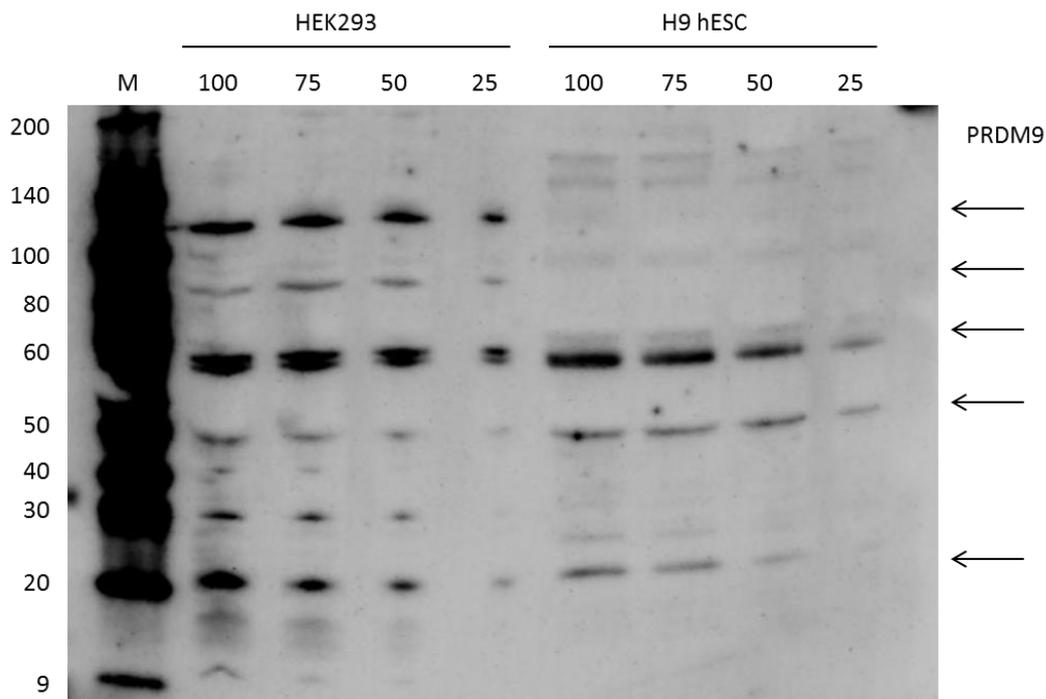
Figure 6.3 Western blot detection of PRDM9 in HEK293 and H9_hESC cell lysates. Protein size marker was
included (M), sizes are denoted in kDa. Protein concentration gradient for each sample lysate was loaded
on a 4-20% bis-tris gel (100 – 25 μg). Membrane was incubated with anti-PRDM9 (1:500 dilution)
overnight at 4°C.

The well-established immortalised cell line Neuron-committed Teratocarcinoma (NT2) is derived from a human embryonic teratoma. As it is of embryonic origin, it was tested for PRDM9 protein expression by immunoblotting alongside three primary myoblast cell lines available in our laboratory (Figure 6.4). Detection of PRDM9 was unsuccessful in all cell lines tested, even after 1 hour of membrane exposure. Incubation of the samples with anti-βactin showed that protein was detectable in all sample wells further confirming that PRDM9 was not present in these cell lines (Figure 6.4).
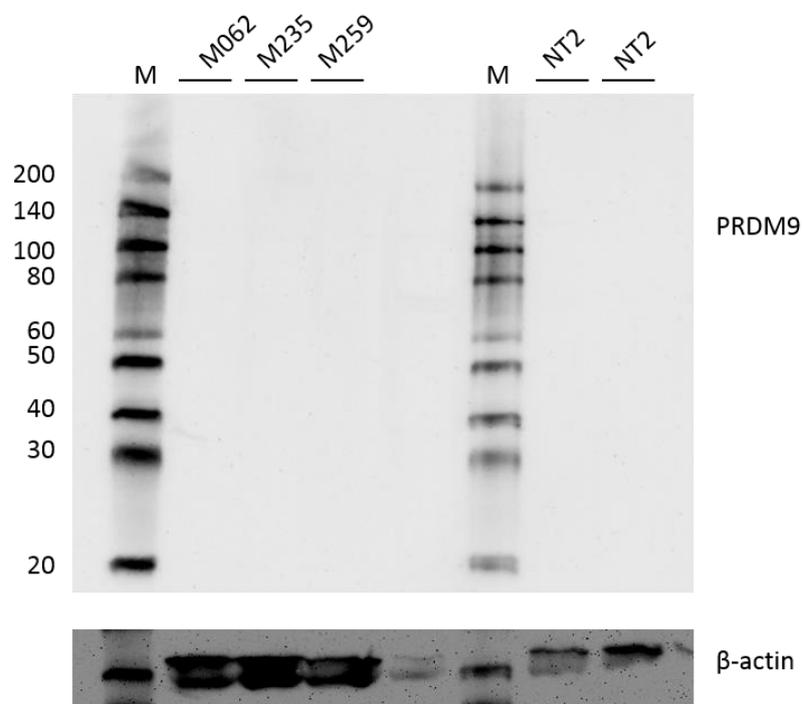
Figure 6.4 Western blot detection of PRDM9 in primary myoblast and NT2 cell lysates. Protein size marker was included (M), sizes are denoted in kDa. 75 µg sample lysate was loaded on a 4-20% bis-tris gel. Membrane was incubated with anti-PRDM9 (1:500 dilution) overnight at 4°C.

To check antibody specificity, two antibodies against PRDM9 were kindly provided by the HPA (AlbaNova, Stockholm, Sweden). These antibodies were used for Western blot detection of PRDM9 in HEK293 cell lysate (Figure 6.5). Bands were detected at ~55, ~45 and ~25 kDa using the Abcam and HPA antibodies respectively (Figure 6.5Figure **6.5**). When compared to levels of β-actin detectable in the lysate (Figure 6.5) the bands detected using the Abcam and HPA059555 antibodies appeared to be very faint compared to HPA063372 (Figure 6.5). No band was observed at the predicted PRDM9 weight, 103 kDa, using any of the antibodies.
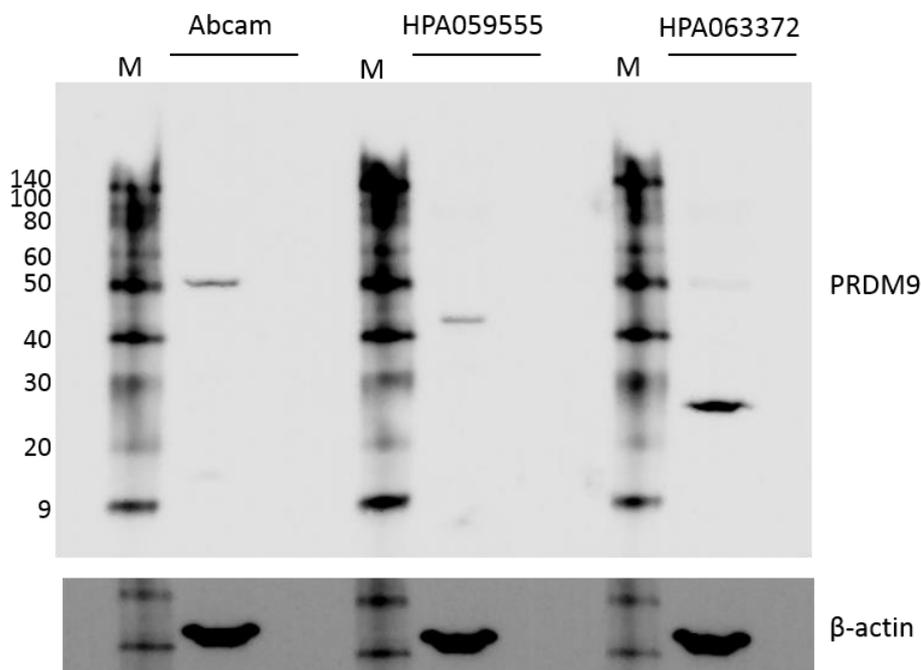
Figure 6.5 Western blot detection of PRDM9 in HEK293 cell lysate. 100 μg total protein was loaded in each sample well and electrophoresed on 4-20% bis-tris gels. Protein size marker was included on each blot (M), sizes are provided in kDa. A) PRDM9 protein was immunoblotted using either Abcam, HPA059555 or HPA063372 antibody against PRDM9. B) β-actin protein was immunoblotted for each sample lane as a loading control. Membrane was incubated with anti-PRDM9 (1:500 dilution) or anti-βactin (1:1000) overnight at 4℃.

The peptide sequence used to generate the antibody against PRDM9 was commercially available (Abcam, Cambridge, UK). This peptide comprises amino acids 432-481 of the human PRDM9 protein sequence and could therefore be used for antigen binding assays to elucidate which of the protein bands detected by blotting could potentially be PRDM9. Briefly, HEK293 cell lysate was incubated with purified PRDM9 antigen peptide, Western blotted and compared to untreated HEK293 total lysate. The strongest band at ~55 kDa in the total cell lysate sample lane was blocked in the sample incubated with the antigen (Figure 6.6, red arrow).
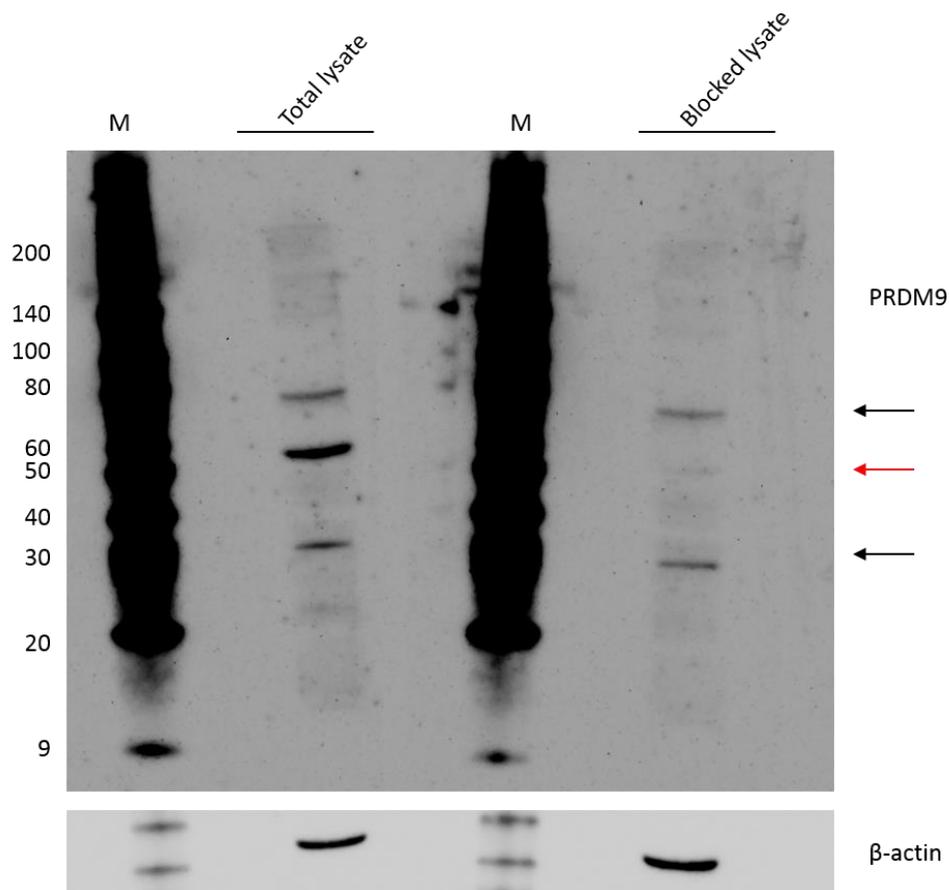
Figure 6.6 Western blot detection of PRDM9 in HEK293 cell lysate blocked with 1 μg/mL antigen peptide. Protein size marker was included on each blot (M), sizes are provided in kDa. Total lysate: 100 μg total HEK293 cell lysate was loaded and electrophoresed on a 4-20% bis-tris gel (left). Blocked lysate: after incubation with PRDM9 peptide antigen, 100 μg HEK293 cell lysate was loaded and electrophoresed on a 4-20% bis-tris gel (right).

Together, these data suggest that the protein observed at ~55 kDa is the specific target of this antibody. Although this does not confirm that full length human PRDM9 is being

detected, the protein target at ~55 kDa is detectable in HEK293 cell lysate using this antibody.

6.4.2 Detecting PRDM9 in subcellular fractions

In order to localise PRDM9 to the mitochondria, subcellular fractions were carried out using HEK293 cells. Nucleus, cytoplasm and mitochondria were isolated by differential centrifugation and each fraction lysed for protein content. Detection of PRDM9 by Western blotting showed differential protein binding in the mitochondria compared to nucleus or cytoplasm which showed similar banding profiles (Figure 6.7).
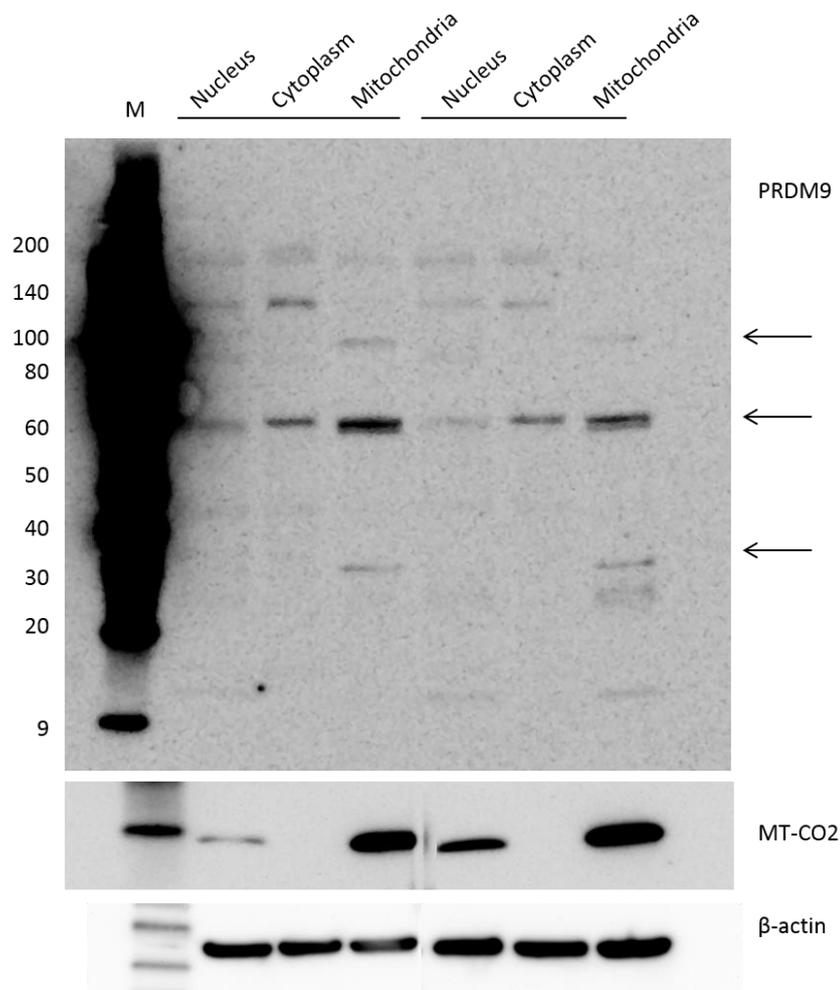
Figure 6.7 Immunoblotting of HEK293 subcellular fractions. 50 μg total protein was loaded in each sample well and electrophoresed through 4-20% bis-tris gels. Protein size marker was included (M), sizes are denoted in kDa. Membrane was incubated with anti-PRDM9 (1:500 dilution), anti-MTCO2 (1:1000) or anti-βactin (1:1000) overnight at 4°C.

Mitochondrial fractions showed enrichment of a band at ~60 kDa which was also present in the nuclear and cytoplasmic fractions (Figure 6.7). Bands at ~100 kDa and ~30 kDa were observed exclusively in the mitochondrial fractions. Although there was slight contamination of the nuclear fraction, the cytoplasmic and mitochondrial fractions were pure as shown by mitochondrial encoded cytochrome C oxidase II (MT-CO2) protein detection.

In order to purify the mitochondrial fractions further, a percoll gradient method was used to separate the mitochondria and the mitochondrial associated membrane (MAM) so that the 'pure' mitochondria could be compared to the 'crude' mitochondrial fractions (Section 3.6.3). When this subcellular fraction was immunoblotted with anti-PRDM9 there was a faint band detected in the crude mitochondrial sample lanes (Figure 6.8 A). This band was not present in the pure mitochondrial fraction or the cytosol or nuclear fractions. Purity of the fractions was tested by blotting with different antibodies specific to different cellular compartments (Figure 6.8 B). Heat shock protein family A member 5 (HSPA5/Bip1) is an endoplasmic reticulum (ER) chaperone protein of 72 kDa in size and was mostly present in the cytoplasmic fraction as shown in the left hand panel in Figure 6.8 B. The protein was also detected in the nuclear and mitochondrial fractions suggesting some contamination most likely due to the chaperone/stress response nature of Bip1 when it will shuttle from the ER to the cytosol. The structural protein α-tubulin was detected in the nuclear and cytosolic fractions at 50 kDa but was not detected in the mitochondrial fractions suggesting that these samples were not contaminated by components of the cytoskeleton. An integral component of the mitochondrial outer membrane is the voltage-dependant anion-selective channel protein 1 (VDAC1/Porin) protein which creates a channel for the diffusion of small hydrophilic molecules. Immunodetection of VDAC1 (31 kDa) showed that it was present in the mitochondrial fractions but absent in the cytoplasm (Figure 6.8 B, right hand panel). There was faint VDAC1 detection in the nuclear fraction suggesting that some mitochondrial membrane may have contaminated this. VDAC1 is also present in the plasma membrane of the cell which may suggest why a relatively small amount was present in the nuclear/membrane fraction. Finally, histone H3 protein was used to show purity of the nuclear fraction. This protein is an important structural protein involved in chromosomal packaging, transcriptional silencing and epigenetic modification within

the nucleus. Detection of this 13 kDa protein was observed in nuclear, cytosolic and crude mitochondrial fractions suggesting that the nuclear membrane may have been disrupted during initial cell fractionation or that some cell nuclei were present in the subsequent fraction steps (Figure 6.8 B, left hand panel).
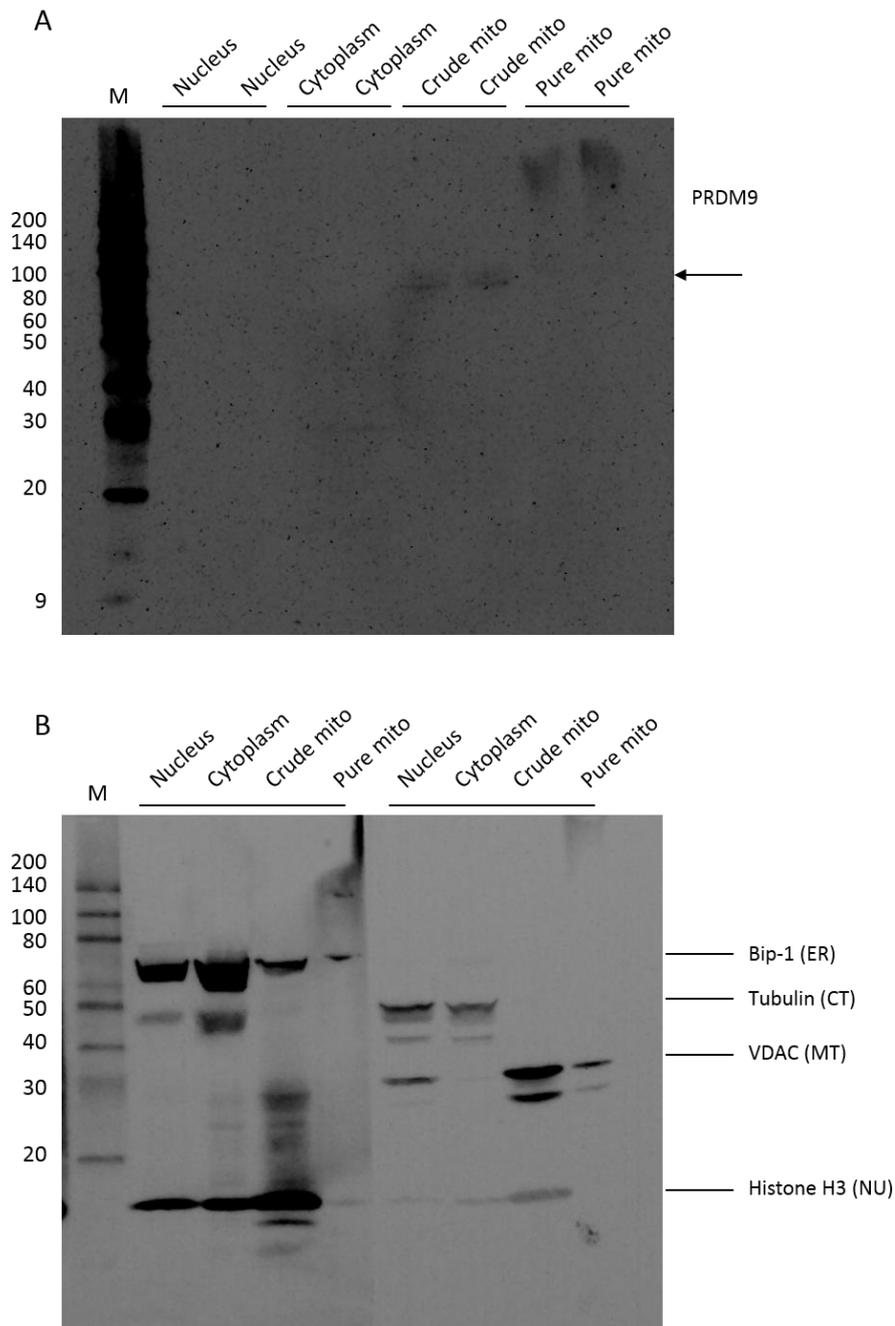
Figure 6.8 Immunoblotting of HEK293 subcellular fractions. 40 µg total protein was loaded in each sample well and electrophoresed through 4-20% bis-tris gels. Protein size marker was included (M), sizes are denoted in kDa. Membrane was incubated with anti-PRDM9 (1:500), anti-HSPA5 (Bip1, 1:1000) anti-VDAC1 (1:250) anti-α tubulin (1:1000) or anti-histone H3 (1:1000) overnight at 4°C.

136

The detection of PRDM9 by Western blotting of whole cell protein lysates and subcellular fractions proved difficult. Cell lines either showed very low protein levels or no protein at all when commercially available PRDM9 antibodies were used. Each detection method and cell line appeared to give different peptide banding profiles suggesting that the protein was not produced in any of the cell lines tested or that the available antibodies are not reliable.

6.4.3 Detecting PRDM9 expression in tissue

Several tissue types were investigated for PRDM9 expression by immunofluorescent staining. Tissues were chosen based on the likelihood that they would contain cells possibly undergoing meiosis and therefore express PRDM9: foetal gonad tissue isolated from an 11 post-conception week foetus, human placenta samples from 5, 6 and 10 post conception weeks and ovary tissue from an adult with a known mitochondrial disease (MELAS syndrome). Human skeletal muscle was used as an internal control because it is post mitotic and should not express PRDM9.

In order to increase antibody specificity a fragment antigen binding (F(ab)) antibody against human PRDM9 was kindly provided by Dr M. Vedadi (Toronto Children's Hospital, Toronto, Canada). This F(ab) fragment is an antibody which still binds to the antigen but is monovalent, lacking the Fc portion of the antibody structure. This allows for more targeted binding to the antigen without potential cell surface receptor binding by the Fc fragment. It also allows better tissue penetration since the F(ab) fragment is much smaller in size than a traditional divalent antibody structure. This antibody was raised in phage λ and requires anti-human IgG as a secondary antibody for detection. Two commercially available PRDM9 antibodies were also tested as a comparison.

Firstly, human muscle tissue was analysed for the presence of PRDM9 using the anti-PRDM9 F(ab) antibody and anti-human IgG secondary antibody (Figure 6.9). Secondary antibody specificity was tested on muscle sections at different concentrations (Figure 6.9 A, B & C). Positive staining was observed at all concentrations indicating that the secondary antibody produced a high level of non-specific binding. Tissue sections stained with both primary and secondary antibodies gave positive signal around the periphery of the individual muscle fibres (Figure 6.9 D, E & F). PRDM9 signal did not co-localise with the nuclei as shown by nuclear staining with 4'6-diamidino-2-phenylindole dihydrochloride (DAPI) which binds to AT rich regions of DNA and emits blue fluorescence. Different concentrations of secondary antibody were tested to estimate levels of background staining. All three secondary antibody concentrations showed similar staining patterns (Figure 6.9 D, E & F).
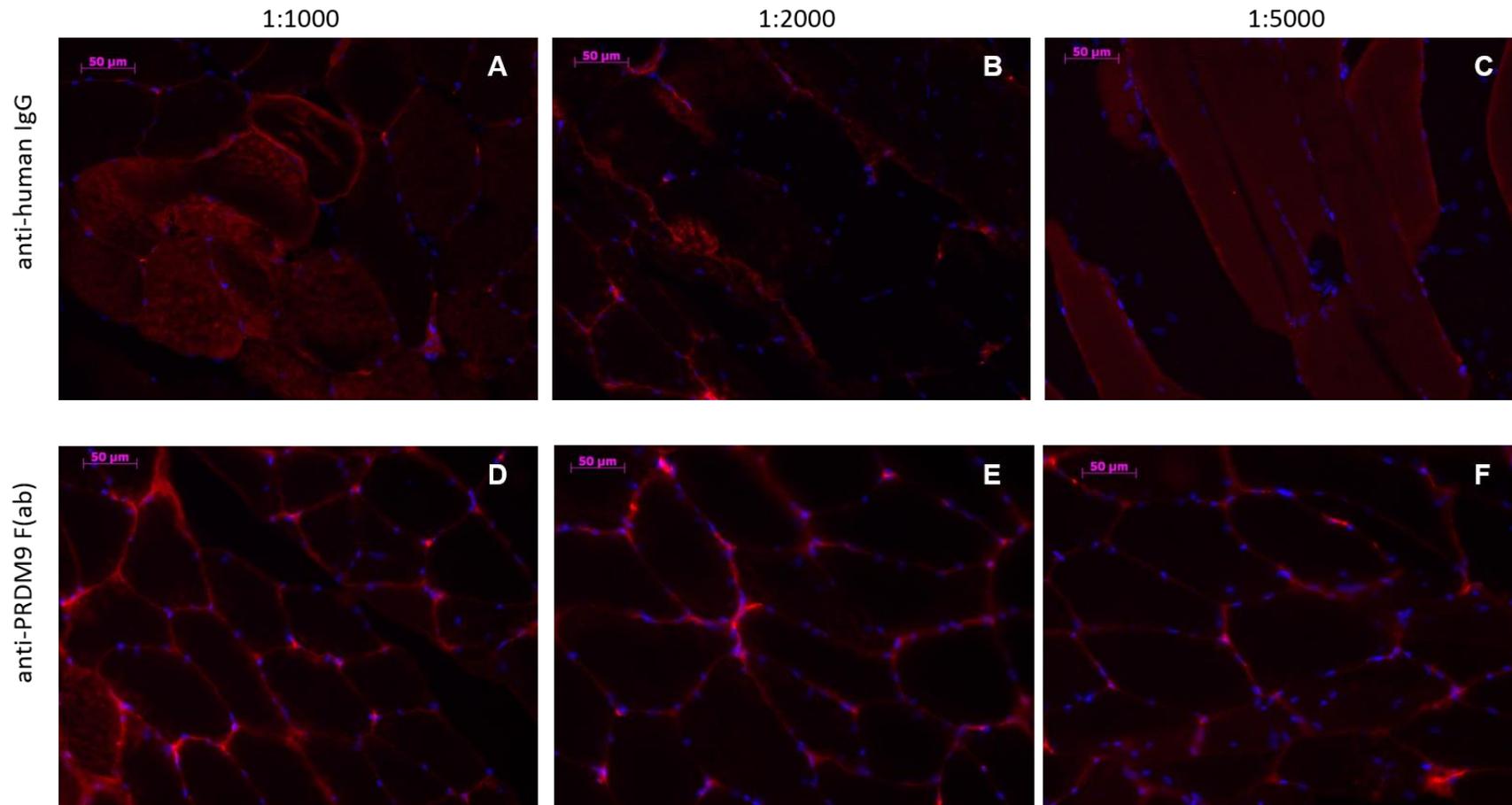
Figure 6.9 Immunofluorescent staining of human skeletal muscle tissue. A, B & C) human muscle section stained with secondary anti-human IgG antibody. (A) 1:1000 antibody dilution. (B) 1:2000 dilution. (C) 1:5000 dilution. D, E &F) human muscle section stained with anti-PRDM9 F(ab) primary and anti-human IgG secondary antibodies. (D) 1:1000 dilution. (E) 1:2000 dilution. (F) 1:5000 dilution. Nuclei are stained with DAPI in all samples.

Ovary tissue sections were assessed for PRDM9 expression by immunofluorescence (Figure 6.10). Positive staining was observed using the PRDM9 F(ab) antibody (Figure 6.10 A). However, anti-human IgG control sections showed high levels of non-specific staining using only this antibody on the tissue (Figure 6.10 B). This suggested that the anti-human IgG was not a suitable antibody for use on human tissue sections as shown by previous staining in muscle tissue (Figure 6.9).
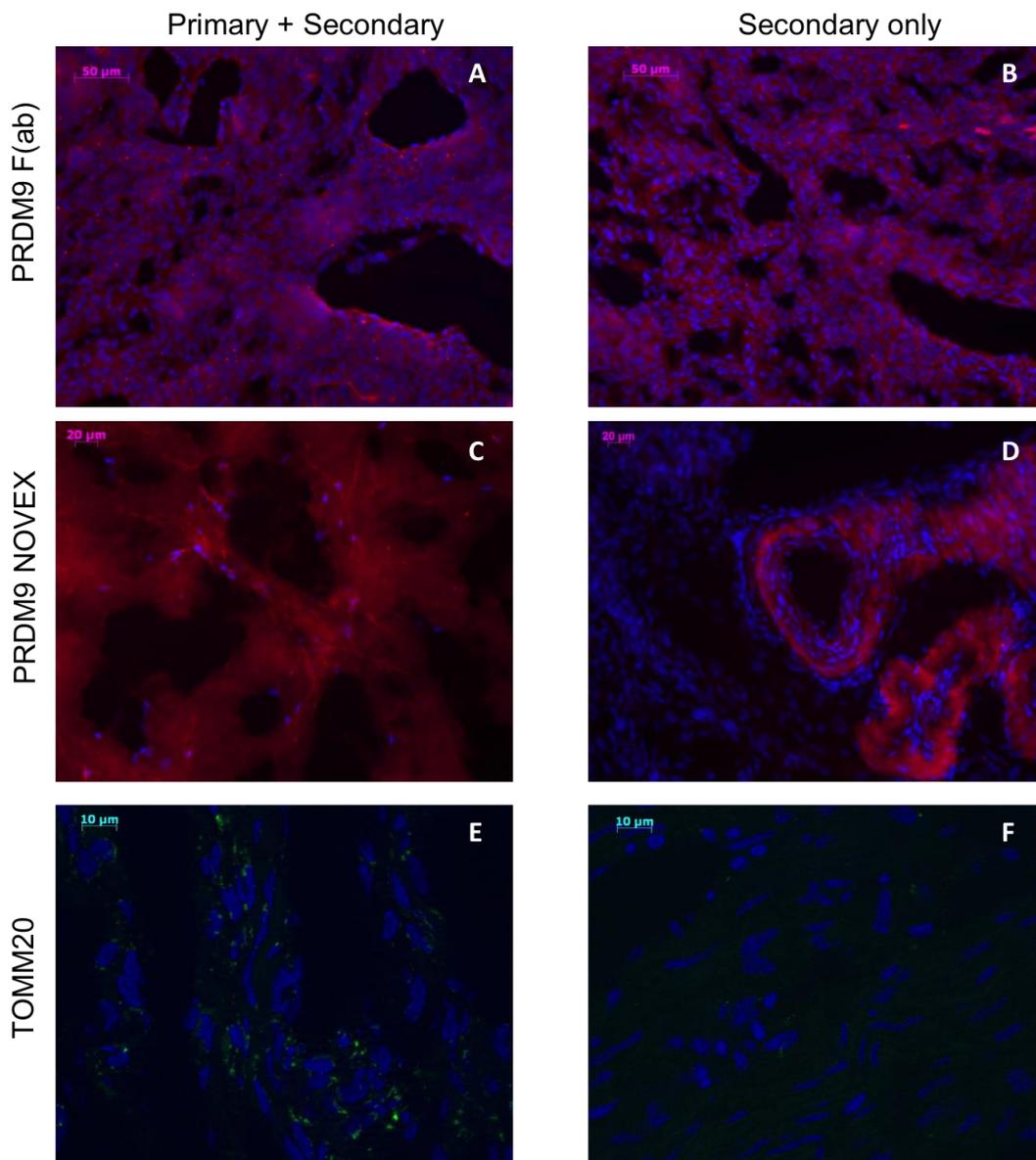
Figure 6.10 Immunofluorescent staining of human ovary tissue. All samples were imaged using the AxioImager (Leica). A & B) Serial sections were incubated with anti-PRDM9 F(ab) primary and anti-human IgG-594 secondary (A) or secondary only (B). C & D) Serial sections were incubated with anti-PRDM9 (Novex) primary and anti-mouse-594 secondary (C) or secondary only (D). E & F) Serial sections were incubated with anti-TOMM20 primary and anti-mouse-488 secondary (E) or secondary only (F). Nuclei are stained with DAPI in all samples.

Secondly, the anti-PRDM9 commercially available antibody (Novex, ThermoFisher Scientific, Loughborough, UK) was used to stain ovary tissue sections. This antibody is recommended for immunoprecipitation only but was tested as a comparison to the F(ab) antibody. Positive staining was observed (Figure 6.10 C) however, there was also a high level of background staining when using only anti-mouse IgG antibody on the tissue (Figure 6.10 D). This suggests that the ovary tissue is sequestering the secondary antibody due to some sort of physical property of the tissue rather than a true antigen binding reaction with the antibody.

Finally, an antibody targeted to translocase of outer mitochondrial membrane 20 (TOMM20) was used to stain the tissue as a positive control. Positive staining was observed (Figure 6.10 E) and appeared to be specific to the mitochondria as there was no co-immunofluorescence with the nuclear DAPI staining. Interestingly, the mitochondria appeared to be clumped within the cell and not the dynamic network that is often observed in other cell and tissue types. Moreover, secondary anti-mouse IgG2 control sections showed that the staining was not due to non-specific antigen binding (Figure 6.10 F).

Immunofluorescent staining of human ovary sections showed that the PRDM9 antibodies gave positive staining however, secondary antibody controls suggest that this positive signal is not specific. This is perhaps due to the nature of the tissue itself or the non-specificity of the antibodies. Staining of this tissue with a mitochondrial specific antibody (TOMM20) showed specific staining patterns but the mitochondrial network appeared clumpy and distorted. This could be due to the nature of the mitochondrial disease state of the female the tissue was taken from (MELAS syndrome) or due to the storage and treatment of the tissue once removed. Haematoxylin and eosin staining of the tissue showed freezing artefacts as well as degradation of the tissue structures when compared to other ovary sections in the literature (see Appendix B).

Isolated foetal gonad tissue sections were a kind gift from Dr V. Floros (MRC Mitochondrial Biology Unit, Cambridge University, UK). Sections were stained using both the PRDM9 F(ab) and anti-PRDM9 Novex antibodies (Figure 6.11). Sections stained with secondary anti-human IgG and anti-mouse IgG showed positive staining (Figure 6.11 B & E). Sections that had not come into contact with any antibody and had only been treated with DAPI also showed positive staining suggesting that this tissue has high levels of auto-fluorescence (Figure 6.11 C & F). In contrast, staining of this tissue with an antibody targeted to mitochondrial translocase of outer membrane (TOMM20) was specific to the mitochondrial network in this tissue (Figure 6.11 G, H & I). No background staining was observed in secondary antibody only or DAPI only controls (Figure 6.11 H & I).
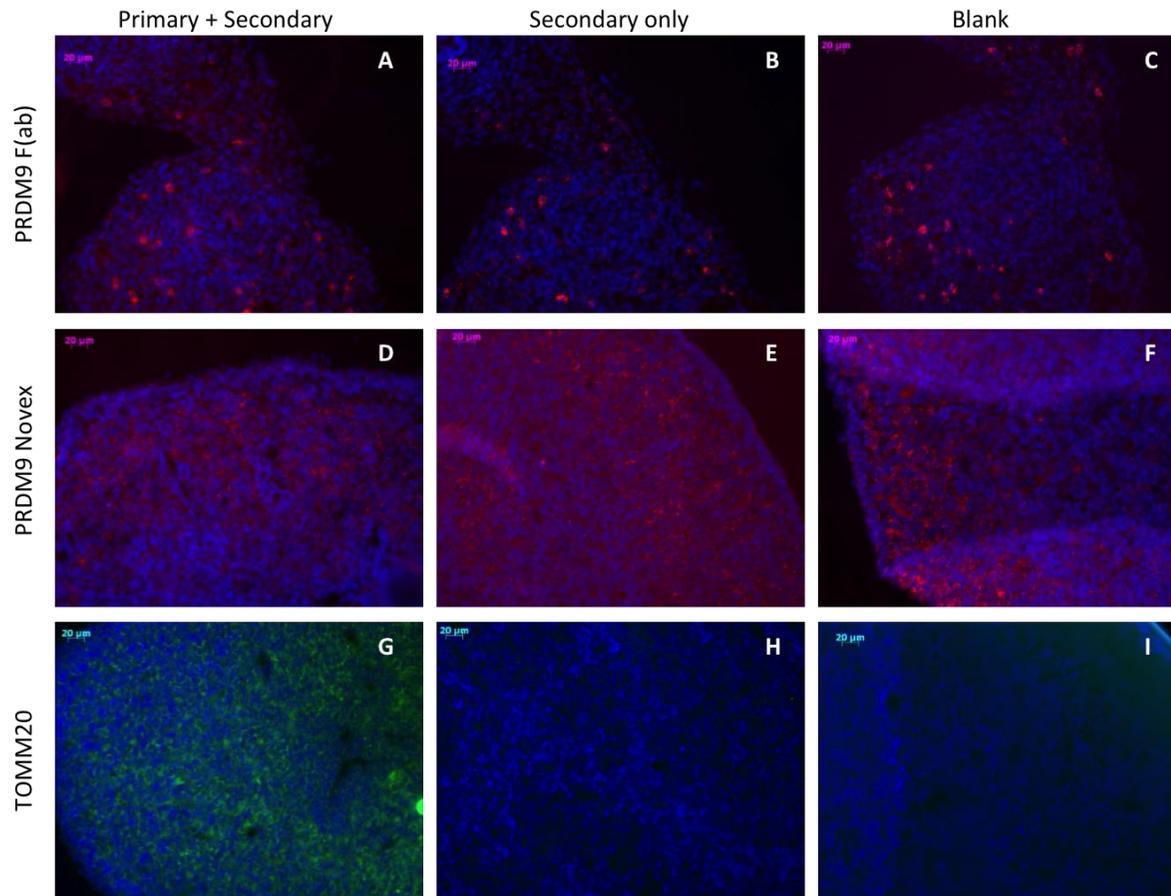
Figure 6.11 Immunofluorescent staining of foetal gonad sample. All samples were imaged using the AxioImager (Leica). A-C) Serial sections were incubated with anti-PRDM9 F(ab) primary and anti-human IgG-594 secondary (A), secondary only (B) or no antibody (C). D-E) Serial sections were incubated with anti-PRDM9 (Novex) primary and anti-mouse-594 secondary (D), secondary only (E) or no antibody (F). G-I) Serial sections were incubated with anti-TOMM20 primary and anti-mouse-488 IgG2a secondary (G), secondary only (H) or no antibody (I). Nuclei are stained with DAPI in all samples.

Immunofluorescence was also performed using placenta tissue (Figure 6.12). Staining tissue sections with PRDM9 F(ab) antibody produced a positive fluorescent stain however as shown in previous tissue sections, the anti-human IgG antibody alone also gave a similar fluorescent stain suggesting that this signal is non-specific (Figure 6.12 A & B). Staining placenta tissue with the anti-PRDM9 Novex antibody did not show any positive fluorescent stain (Figure 6.12 C). Secondary only control staining with anti-mouse antibody also showed no fluorescent signal (Figure 6.12 D). Therefore we conclude that PRDM9 is not detectable in placenta tissue using the antibodies tested in this study.
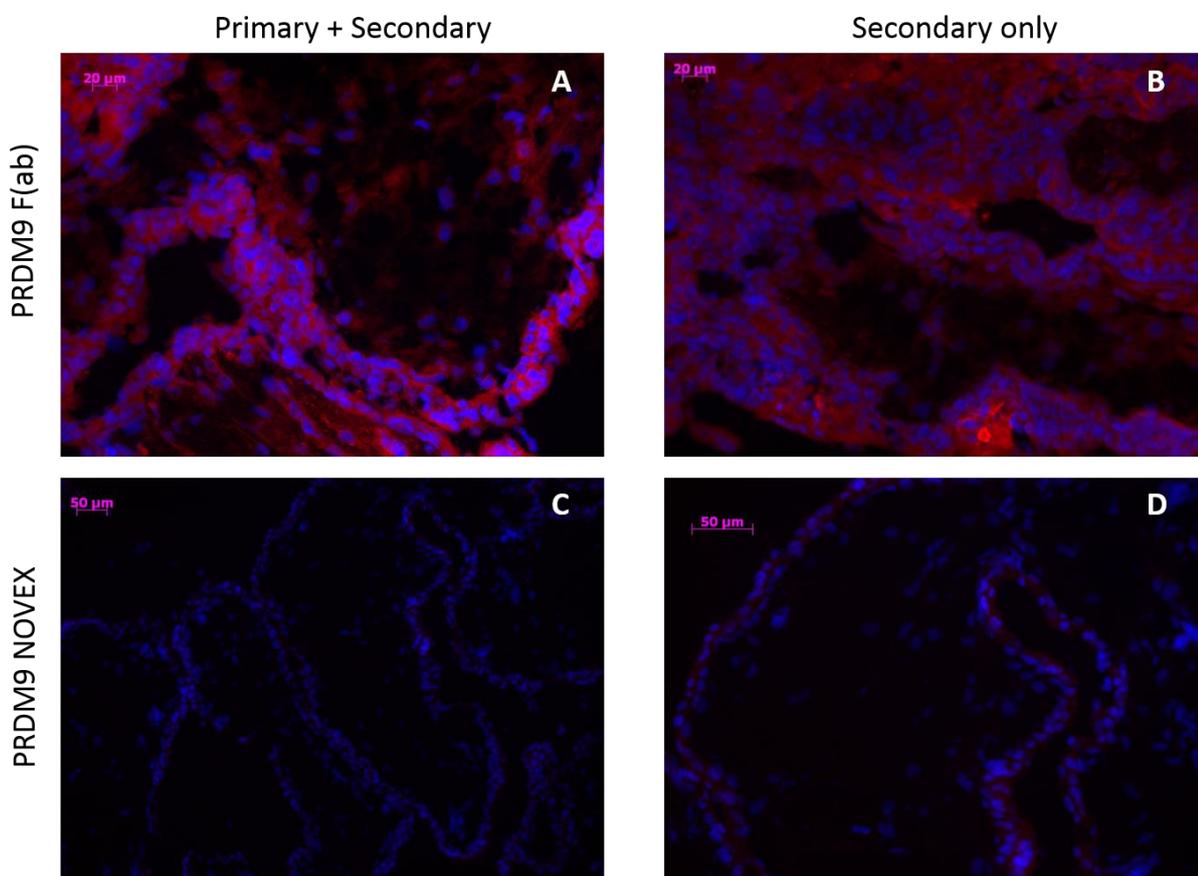
Figure 6.12 Immunofluorescence of human placenta tissue. All samples were imaged using the AxioImager (Leica). A & B) Serial sections were incubated with anti-PRDM9 F(ab) primary and anti-human IgG-594 secondary (A) or secondary only (B). C & D) Serial sections were incubated with anti-PRDM9 (Novex) primary and anti-mouse-594 secondary (C) or secondary only (D). Nuclei are stained with DAPI in all samples.

PRDM9 protein was not detected by immunofluorescent staining in the tissues used in this study. The F(ab) fragment antibody used is not appropriate for human tissue staining as it was not raised in an animal and therefore requires anti-human IgG as a secondary antibody for detection. This antibody can react with most IgG antigens present in the tissue samples therefore producing off target staining. The Novex anti-PRDM9 antibody required anti-mouse IgG secondary antibody for detection. This antibody produced a non-specific fluorescent signal when used to stain both ovary and foetal gonad tissue sections but not in the placenta sample.

Tissue samples were also analysed for protein expression by gel electrophoresis and Western blotting. Ovary tissue lysate was blotted with anti-PRDM9 giving a strong single band at a large molecular weight >200 kDa (Figure 6.13 A). Before the gel was transferred to a PVDF membrane, it was reversibly stained with InstantBlue™ Coomassie brilliant blue stain (Figure 6.13 B). Staining total protein showed a strong signal at ~180 kDa most likely due to multiple high molecular weight proteins.
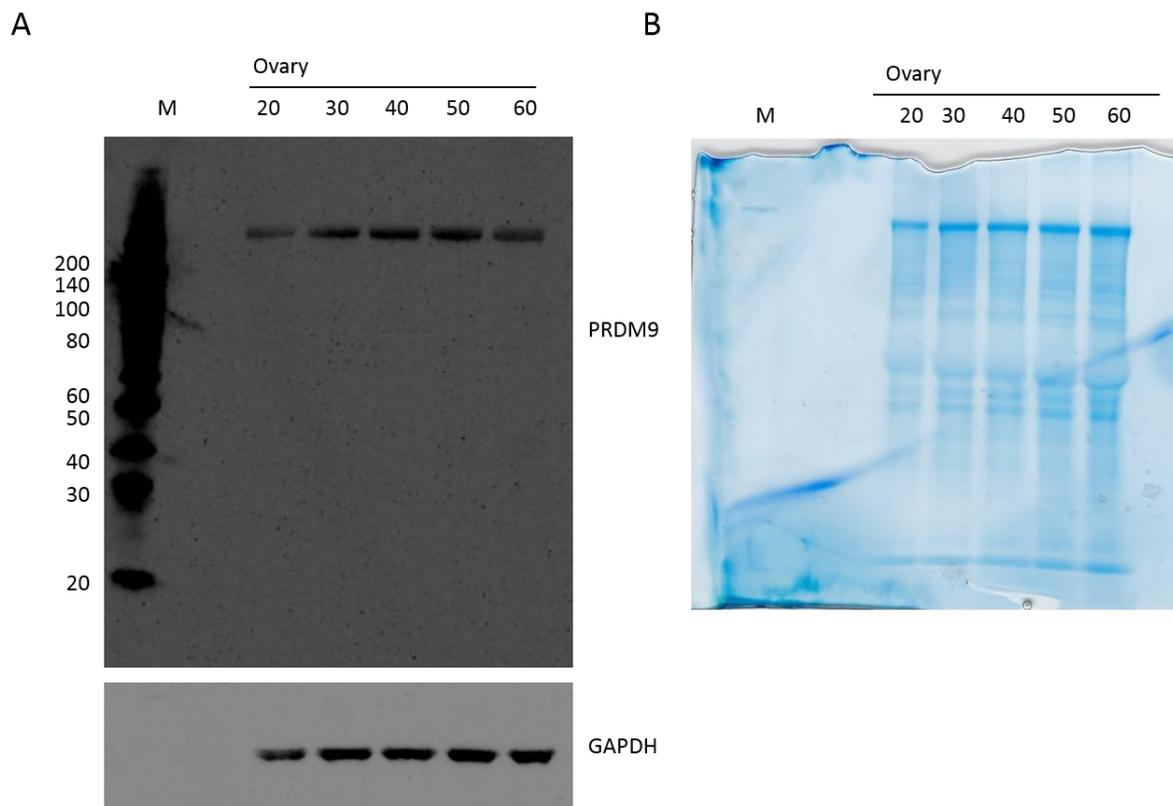
Figure 6.13 Western blot detection of PRDM9 in ovary lysate. Protein size marker was included on each blot (M), sizes are denoted in kDa. A) Ovary lysate gradient from 20-60 µg protein was loaded on a 4-20% bis-tris gel. B) Ovary lysate gradient from 20-60 µg protein was loaded on a 4-20% bis-tris gel. Gel was stained with Coomassie brilliant blue and imaged. Membrane was incubated with anti-PRDM9 (1:500 dilution) or GAPDH (1:1000 dilution) overnight at 4°C.

Foetal gonad tissue (F3) was kind gift from Dr V. Floros (MRC Mitochondrial Biology Unit, Cambridge University, UK). F3 tissue lysate was immunoblotted with anti-PRDM9 (Figure 6.14 A). No protein was detected in this sample lysate using the PRDM9 antibody, when the GAPDH antibody was used a band was detected confirming that there was protein present in the sample lanes. Staining the gel before transfer to PVDF

membrane with InstantBlue™ Coomassie brilliant blue confirmed total protein content of the F3 lysate correlates with the concentration of protein loaded (Figure 6.14 B).
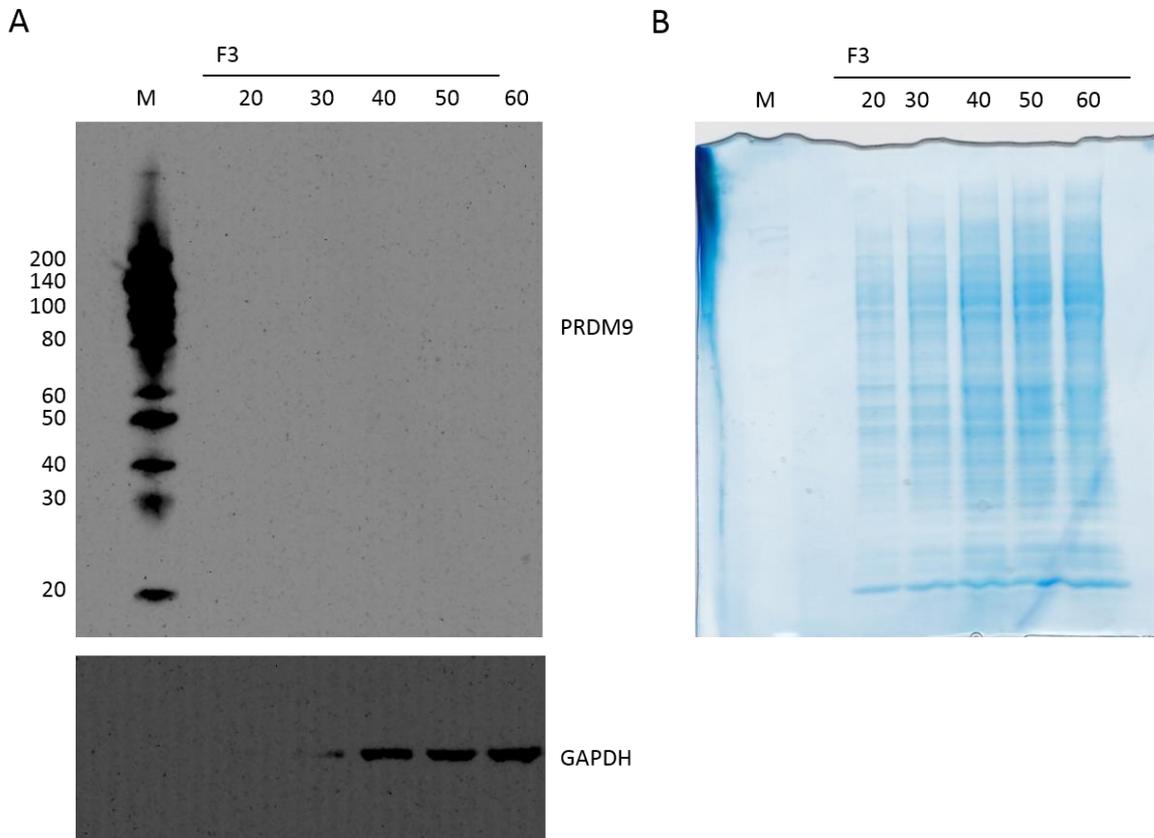
Figure 6.14 Western blot detection of PRDM9 in foetal gonad (F3) tissue lysate. Protein size marker was included on each blot (M), sizes are denoted in kDa. A) F3 lysate gradient from 20-60 μg protein was loaded on a 4-20% bis-tris gel. B) F3 lysate gradient from 20-60 μg protein was loaded on a 4-20% bis-tris gel. Gel was stained with Coomassie brilliant blue and imaged. Membrane was incubated with anti-PRDM9 (1:500 dilution) or GAPDH (1:1000 dilution) overnight at 4°C.

6.4.4 Immunoprecipitation of PRDM9

Due to the non-specific protein banding observed by Western blotting, immunoprecipitation (IP) of PRDM9 was carried out using a range of different PRDM9 antibodies. Four different anti-PRDM9 antibodies were used for IP throughout this study as detailed in Section 3.7.4 (Table 3.3).

Initial IP was carried out using the Novex anti-PRDM9 antibody. This antibody was specified as only appropriate for IP and not for other immunodetection methods, therefore it was potentially more suitable than the antibody used in previous detection methods. After incubation of total HEK293 cell lysate with anti-PRDM9 antibody the unbound, wash and final elution lysates were electrophoresed. Staining the gel with Coomassie brilliant blue showed strong protein detection in the unbound lysate lanes as expected (Figure **6.15** A & B). Weak protein staining was seen in the wash lysate lanes, which was also expected as any further unbound protein would be captured in this washing step. No protein bands were detectable in the final elution lysates suggesting that no protein had bound to the anti-PRDM9-A/G bead complexes during the assay.

As a test for sensitivity of the Coomassie stain, gels were next treated with metallic silver (Ag). Using this staining method, protein bands were visualised in the final elution sample lanes (Figure 6.15 C & D). This was also the case in the 'blank' sample lane where no antibody was present in the reaction suggesting that some protein was either contaminating this sample or that the A/G beads themselves were contributing protein during the final elution step (Figure 6.15 D). Interestingly, a band at ~100 kDa was present in the elution sample lanes where the initial protein input was 5000 and 3000 µg (Figure 6.15 C & D). As expected, a band was present at ~50 kDa in the samples where antibody was added to the reaction mix. This band corresponds to the IgG heavy chain of the antibody complex which is pulled-down in the final elution step and is present in high amounts.
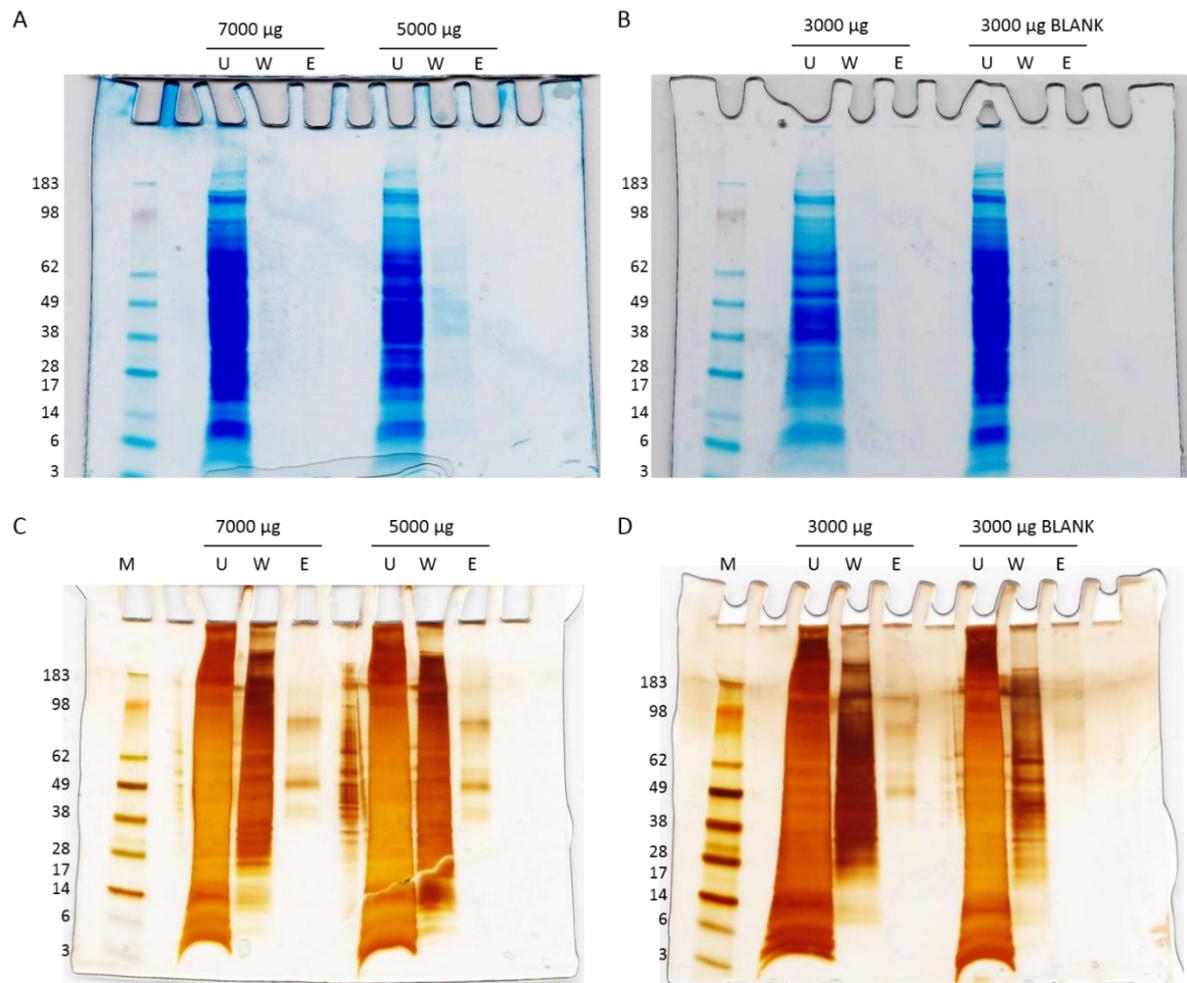
Figure 6.15 Immunoprecipitation of PRDM9 from HEK293 cell lysate. Protein lysate was incubated in the presence or absence (denoted Blank) of the Novex anti-PRDM9 antibody. Unbound protein (U), bead wash (W) and final protein elute (E) were loaded on 4-20% bis-tris gels and electrophoresed until samples had migrated through the entire gel. A & B) Gels were stained with Coomassie® blue. C & D) Gels were stained with metallic silver (Ag). Protein size markers were run on each gel in the left hand lane, sizes in kDa are denoted.

In order to determine whether the ~100 kDa band detected by silver staining in Figure 6.15 was PRDM9, Western blotting was carried out using gels run in parallel with the same IP sample lysates. After probing with anti-PRDM9 (Abcam) three distinct bands could be detected in final elution sample lanes but were absent in the blank sample lane (Figure 6.16). The band at ~50 kDa most likely corresponds to the IgG heavy chain which is detectable by secondary antibody binding. Other bands were observed at ~140 and ~45 kDa, the identity of these protein species is unknown. A faint band was detectable in the unbound protein lysate lanes at ~25 kDa, most likely corresponding to the IgG light chain.
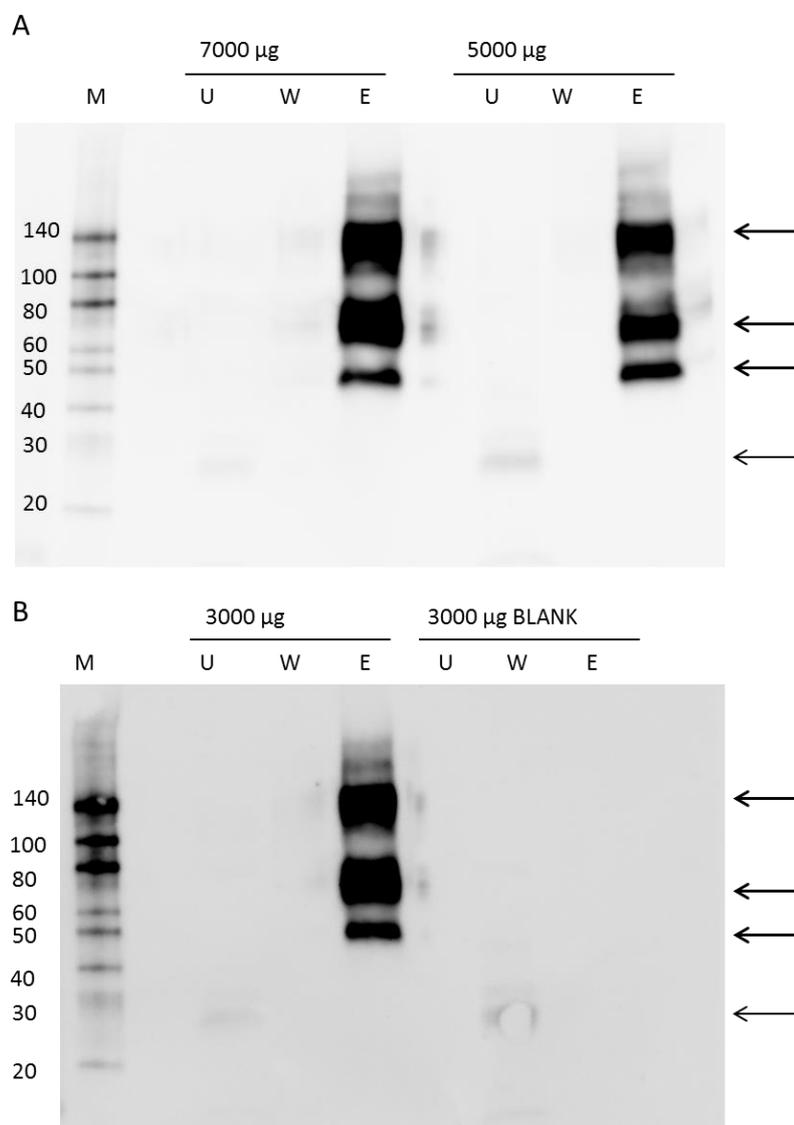
Figure 6.16 Western blot detection of immunoprecipitated PRDM9 from HEK293 cell lysate. Protein lysate was incubated in the presence or absence (denoted Blank) of the Novex anti-PRDM9 antibody. Unbound protein (U), bead wash (W) and final protein elute (E) were loaded on 4-20% bis-tris gels and transferred onto PVDF membranes.

Next, the four anti-PRDM9 antibodies (Novex, Abcam, HPA063372 and HPA059555) were compared to determine whether any were appropriate for IP of PRDM9 protein (Figure 6.17). The Novex antibody produced a single band at ~100 kDa in the final elution lysate (Figure 6.17 A). The other three antibodies gave multiple bands in the final elution lysates. The Abcam antibody gave three bands at ~50 kDa which could be the IgG heavy chain or unknown proteins of this size. As previous Western blot detections had shown this antibody to bind a species of ~55 kDa, this result was expected when using it for IP. The HPA antibodies were used at different concentrations as there was a limited amount of these samples. The results seen in Figure 6.17 B are most likely due to an excess of antibody in the HPA063372 reaction and too little antibody in the HPA059555 sample.
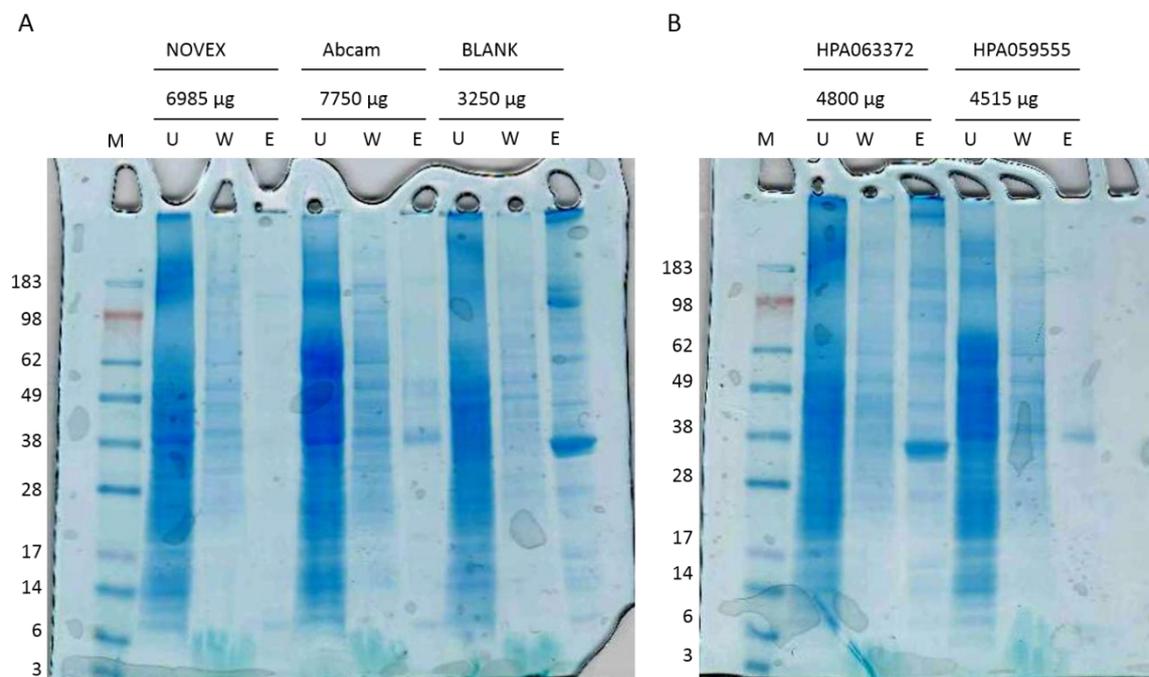
Figure 6.17 Immunoprecipitation of PRDM9 in HEK293 cell lysate. Protein lysate was incubated in the presence or absence (denoted Blank) of the Novex, Abcam, HAP063372 or HPA059555 anti-PRDM9 antibodies. Unbound protein (U), bead wash (W) and final protein elute (E) were loaded on 4-20% bis-tris gels and electrophoresed until samples had migrated through the entire gel. A & B) Gels were stained with Coomassie® blue. Protein size markers were run on each gel in the left hand lane, sizes in kDa are denoted.

Immunoprecipitation of PRDM9 was attempted using ovary lysate, which had previously shown PRDM9 detection by Western blot (Figure 6.13). Gel staining using Coomassie brilliant blue showed three protein bands at ~48, ~45 and ~12 kDa (Figure 6.18 A). Silver staining the gel also showed these three bands and additionally bands at ~100 and ~140 kDa (Figure 6.18 B). Interestingly there were some high molecular weight proteins detected in the 'blank' elution sample lanes where antibody was not added to the reaction mix (Figure 6.18 A & B).
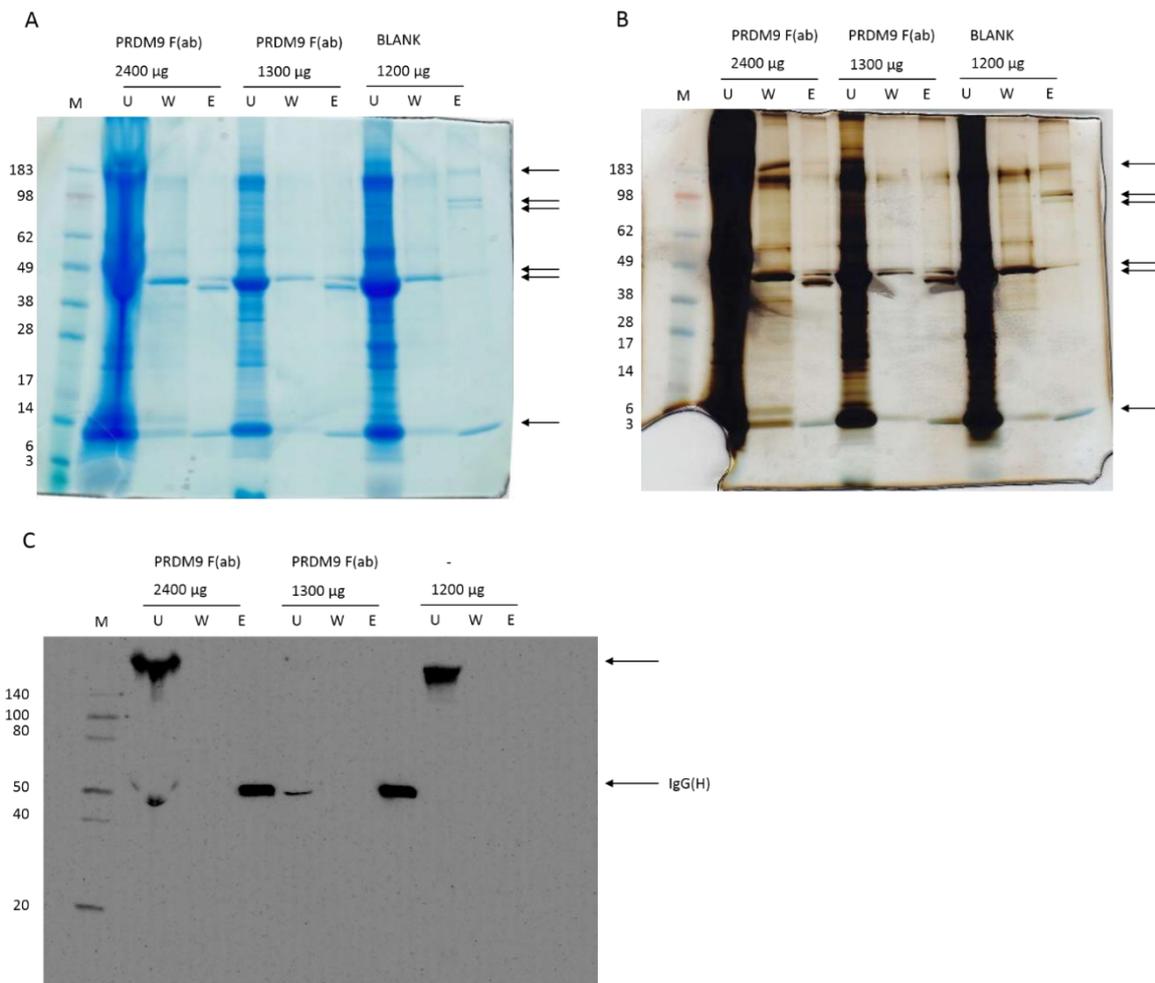
Figure 6.18 Immunoprecipitation of PRDM9 from ovary tissue lysate. Ovary lysate was incubated in the presence or absence (denoted Blank) of anti-PRDM9 F(ab). Unbound protein (U), bead wash (W) and final protein elute (E) were loaded on 4-20% bis-tris gels and electrophoresed until samples had migrated through the entire gel. A) Gel was stained with Coomassie® blue. B) Gel was stained with metallic silver (Ag). C) Gel was transferred onto PVDF membrane and Western blot detection performed using anti-PRDM9 antibody (Abcam). Protein size markers were run on each gel in the left hand lane, sizes in kDa are denoted.

Western blot detection of PRDM9 was tested for in the IP sample lysates. When incubated with anti-PRDM9 (Abcam) a single band was detected in the elution sample lanes at ~50 kDa but was absent in the blank sample lanes (Figure 6.18 C). This band corresponds to the IgG heavy chain of the antibody used in the IP reaction. A high molecular weight band was detected in the unbound lysate sample lanes of the 2400 μg and blank IP reactions. This protein species might be a non-specific binding target of the antibody or could be a peptide present in the ovary lysate containing a PRDM9 molecule in complex with other peptides making it much larger than the predicted molecular weight of 103 kDa.

In order to determine whether there were different proteins binding the anti-PRDM9 antibody in the ovary tissue and HEK293 cell lysates we directly compared ovary, HEK293 and 143B lysates (Figure 6.19). When gels were stained with Coomassie brilliant blue there were two protein bands present in the final elution lanes at ~45 and ~12 kDa (Figure 6.19 A & B). There was also a faint band at ~100 kDa in the HEK293 elution sample lane (Figure 6.19 A). The control 'blank' sample, where no antibody was added to the IP reaction mix, showed a single band at ~12 kDa which could correspond to a contaminating peptide or perhaps a peptide fragment present on the A/G beads used in the assay (Figure 6.19 B).
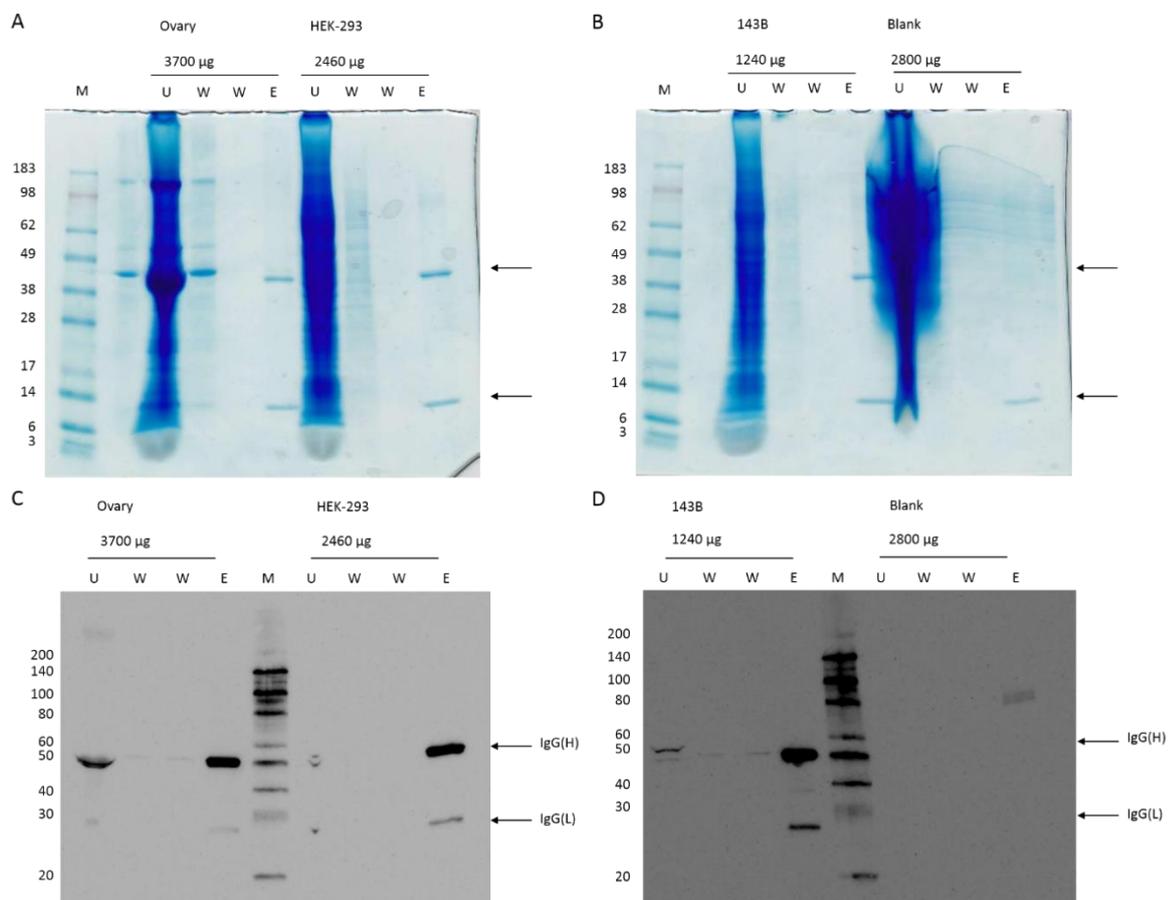
Figure 6.19 Immunoprecipitation of PRDM9 in ovary, HEK293 and 143B lysates. Lysates were incubated in the presence or absence (denoted Blank) of anti-PRDM9 F(ab). Unbound protein (U), bead wash (W) and final protein elute (E) were loaded on 4-20% bis-tris gels and electrophoresed until samples had migrated through the entire gel. A & B) Gels were stained with Coomassie® blue. C & D) Gels were transferred onto PVDF membranes and Western blot detection performed using anti-PRDM9 antibody (Abcam). Protein size markers were run on each gel in the left hand lane, sizes in kDa are denoted.

To test for PRDM9 presence in the IP samples, Western blot detection was performed on gels run in parallel using the IP sample lysates. Incubation of the membranes with anti-PRDM9 (Abcam) detected a band at ~50 kDa in the unbound and final elution sample lanes corresponding to the IgG heavy chain (Figure 6.19 C & D). A band at ~25 kDa was also observed in the final elution sample lanes which corresponds to the IgG light chain. These bands were absent in the control 'blank' sample further confirming that they are most likely IgG peptides from the antibody used in the IP reactions (Figure 6.19 D).

Ovary tissue, HEK293 and 143B cell lysates showed the same protein profile when IP was attempted using the Novex PRDM9 antibody. There was a faint band in the HEK293 elution sample at ~100 kDa which was not detected by Western blot detection suggesting that this might be contamination by another peptide of this molecular weight and most likely not PRDM9.

In almost all IP reactions there was detection of protein in the 'blank' sample where beads were incubated with protein lysate only. To determine whether this detection was due to contaminating protein from the sample lysate or from the A/G beads themselves, two different beads (Thermo Fisher Scientific, Loughborough, UK) were incubated in wash buffer in the presence or absence of PRDM9 F(ab) antibody (Figure 6.20). When incubated in the presence of PRDM9 F(ab), a protein band was detected at ~40 kDa in the elution sample from both beads (M280 and T1). There was a band detected at ~10 kDa that was also present when there was no antibody added to the reaction suggesting that the beads themselves contaminate the elution samples with a peptide species.
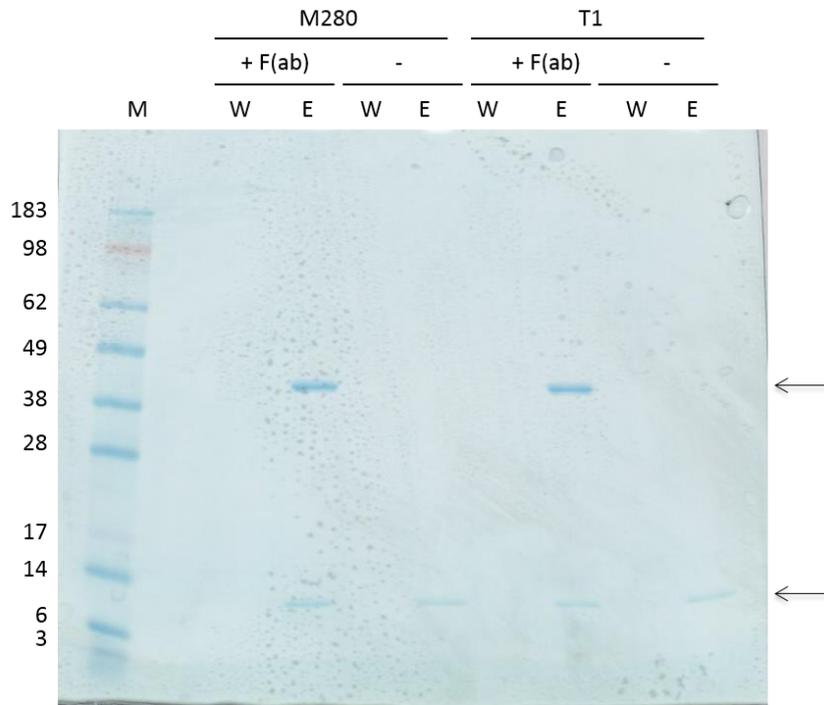
Figure 6.20 Test of A/G bead contamination in immunoprecipitation reactions. M280 or T1 beads were incubated in the presence or absence of F(ab) PRDM9. Wash (lanes W) sample and elution (lanes E) sample were electrophoresed on 4-20% bis-tris gels and stained with Coomassie brilliant blue. Protein size marker (lane M) was included on the gel, sizes are denoted in kDa.

6.4.5 Mass spectrometry analysis of PRDM9 immunoprecipitation samples

Mass spectrometry (MS) analysis was performed on bands cut out from gels containing the IP samples. Peptide fragments were analysed in a shotgun approach where amino acid sequences were re-aligned using online protein databases to identify the peptides present in the sample. Methods relating to the MS analysis are detailed in Section 3.8.5.

HEK293 cell lysate was incubated with the Novex anti-PRDM9 antibody and the gel containing the samples sent for mass spectrometry analysis (Figure 6.21). Final elution lysate showed three bands at ~100, ~45 and ~15 kDa. To identify the peptides being pulled down by this antibody, we analysed all three bands from this elute (Figure 6.21).
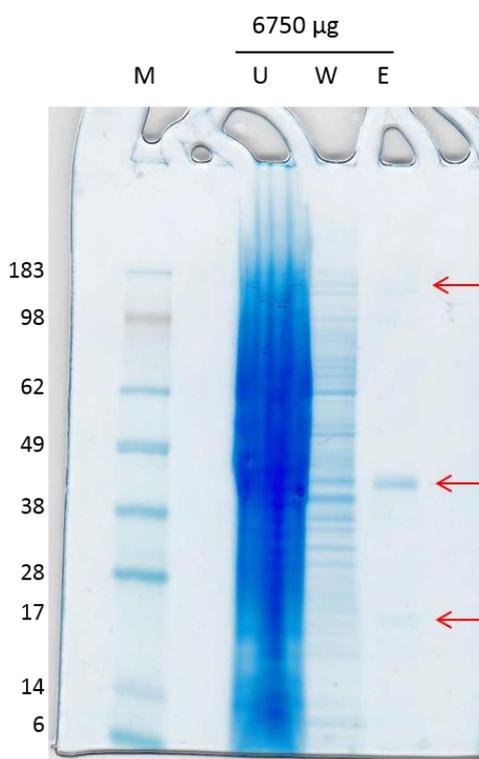
Figure 6.21 Immunoprecipitation of PRDM9 in HEK293 cell lysate. Lysate was incubated with Novex anti-PRDM9 antibody. Unbound (U), bead wash (W) and final protein elute (E) were loaded on 4-20% bis-tris gels and electrophoresed until samples had migrated through the entire gel.

The top 16 peptides, based on confidence score (log(e)), identified by MS are shown in Table 6.3. Unfortunately, PRDM9 peptide sequence was not detected in the sample. The most abundant peptides were members of the keratin protein family, from

contamination of the gel by hair or skin cells present during the sample or gel preparation. Trypsin was also found in the MS analysis, this is most likely excess enzyme remaining after digestion of the gel slices during preparation of the sample. Immunoglobulin heavy variable 3-7 protein was identified and is a fragment from the antibody used in the IP reaction. In addition, filaggrin family member 2 (FLG2) peptide was detected in the sample. This protein is primarily involved in epithelial homeostasis and is important for cornification of skin suggesting that it was likely another contaminant of the IP gel during processing (Dang *et al.*, 2016).

Other identified peptides were found in very low abundance but are also predicted to be contaminants of the pull down (Table 6.3). For example, eukaryotic translation elongation factor 1 alpha (EEF1A1), actin alpha 1 (ACTA1) and phenylalanyl-tRNA synthetase (FARSB) were all identified in the sample however this is probably due to the relative abundance of these proteins within the cell and therefore the lysate. Actin filaments and members of the transcription/translation processes are some of the most abundant cellular proteins and are in this case presumed to be carry over contamination in the IP elution.

Identifier	log(I)	rI	log(e)	pI	Mol Weight (kDa)	Description
ENSP00000310861	5.95	41	-262.0	8.1	65.4	keratin 2 IPR001664 IF IPR003054 (×6) Keratin II IPR009053 Prefoldin
ENSP00000269576	5.99	41	-198.0	5.1	58.8	keratin 10 IPR001664 IF IPR002957 (×5) Keratin I IPR009053 Prefoldin
ENSP00000252244	5.76	35	-191.0	8.1	66.0	keratin 1 IPR001664 IF IPR003054 (×6) Keratin II IPR009053 Prefoldin
ENSP00000246662	4.93	10	-84.3	5.1	62.0	keratin 9 IPR001664 IF IPR002957 (×5) Keratin I
ENSP00000252242	5.01	12	-73.2	7.6	62.3	keratin 5 IPR001664 IF IPR003054 (×6) Keratin II IPR009053 Prefoldin
ENSP00000167586	5.34	9	-57.8	5.1	51.5	keratin 14 IPR001664 IF IPR002957 (×5) Keratin I IPR009053 Prefoldin
sp|K1C15_SHEEP|	5.22	10	-45.2	4.7	48.7	Keratin, type I cytoskeletal 15; Cytokeratin-15; CK-15; Keratin-15; K15;
sp|TRYP_PIG|	5.78	11	-16.3	6.9	24.4	Trypsin; EC 3.4.21.4; Flags: Precursor;
ENSP00000375007	5.15	4	-7.7	6.2	12.9	immunoglobulin heavy variable 3-7 IPR007110
ENSP00000373370	3.74	2	-7.1	8.4	247.9	filaggrin family member 2 IPR002048
ENSP00000353542	3.75	1	-2.6	8.8	7.9	small proline-rich protein 2D
ENSP00000339063	3.48	1	-2.3	9.1	50.1	eukaryotic translation elongation factor 1 alpha 1 IPR000795 (×6)
ENSP00000355645	3.34	1	-2.3	5.2	42.0	actin, alpha 1, skeletal muscle IPR004000 (×8) Actin-related
ENSP00000357762	3.69	1	-1.8	8.7	64.1	keratinocyte proline-rich protein
ENSP00000281828	3.52	1	-1.5	6.4	66.1	phenylalanyl-tRNA synthetase, beta subunit IPR005146 (×2)
ENSP00000451870	3.61	1	-1.3	10.3	2.2	T cell receptor alpha joining 56

Table 6.3 Mass spectrometry protein read-out for HEK293 lysate IP. Identifier protein code, log (I), refractive index (rI), expectation value (log(e)), isoelectric point (pI) and molecular weight is provided along with a description of the protein.

Mass spectrometry was also performed on an IP sample using ovary tissue lysate and the PRDM9 F(ab) fragment antibody (Figure 6.22).
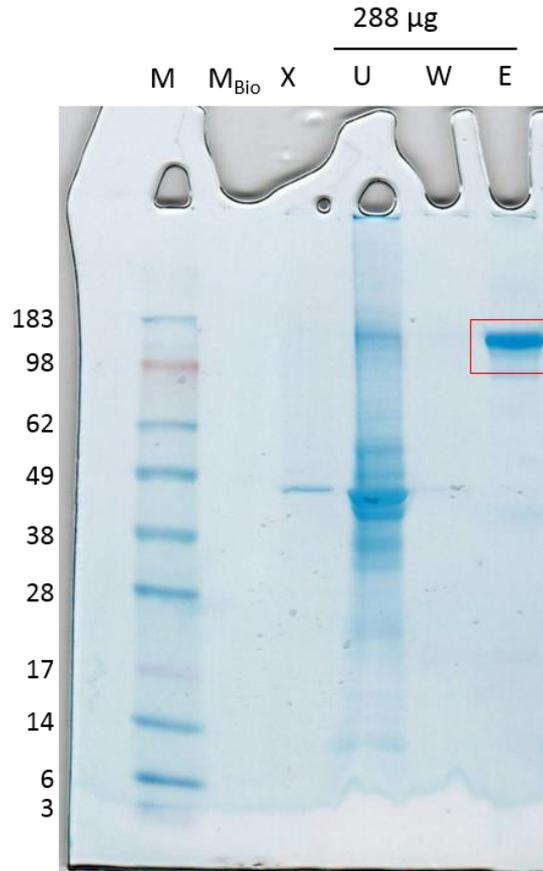
Figure 6.22 Immunoprecipitation of PRDM9 in ovary tissue lysate. Lysate was incubated with Novex anti-PRDM9 antibody. Unbound (U), bead wash (W) and final protein elute (E) were loaded on 4-20% bis-tris gels and electrophoresed until samples had migrated through the entire gel.

The most abundant peptides identified in the MS analysis were subunits of the immunoglobulin protein family (Table 6.4). The abundance of these peptides indicates that the band identified on the IP gel was most likely a large multiprotein complex mostly comprised of antibody fragments. This was caused by insufficient amounts of sample reducing agent used to linearise the peptide molecules and limit protein aggregation or insufficient denaturation of the sample by heating at 95 °C. These two steps usually effectively separate the IgG peptides from any other protein in the sample, which was clearly not the case in this reaction given the MS result (Table 6.4).

Identifier	log(I)	rI	log(e)	pI	Mr	Description
ENSP00000374990	7.17	155	-533.0	6.5	43.9	immunoglobulin heavy constant gamma 1 (G1m marker) IPR013162 (×2)
ENSP00000374993	6.93	91	-422.0	8.2	41.3	immunoglobulin heavy constant gamma 3 (G3m marker) IPR013162 (×2)
ENSP00000374987	6.79	56	-268.0	7.5	35.9	immunoglobulin heavy constant gamma 2 (G2m marker) IPR013162 (×2)
ENSP00000374985	6.87	70	-225.0	7.0	35.9	immunoglobulin heavy constant gamma 4 (G4m marker) IPR013162 (×2)
ENSP00000374777	7.28	73	-200.0	5.6	11.7	immunoglobulin kappa constant IPR013162 CD80 C2-set
ENSP00000375012	5.93	9	-92.3	8.8	12.9	immunoglobulin heavy variable 3-15 IPR007110 Ig-like dom
ENSP00000436353	6.02	24	-90.9	9.1	23.1	immunoglobulin lambda-like polypeptide 5 IPR013162 CD80 C2-set
ENSP00000374856	6.39	28	-90.3	6.7	11.3	immunoglobulin lambda constant 2 (Kern-Oz- marker) IPR013162 CD80 C2-set
ENSP00000375016	6.03	15	-77.6	8.5	12.8	immunoglobulin heavy variable 3-21 IPR007110 Ig-like dom
ENSP00000420576	5.14	9	-60.5	8.5	12.6	immunoglobulin kappa variable 1-12 IPR007110 Ig-like dom
ENSP00000420361	4.93	8	-59.1	8.5	12.7	immunoglobulin kappa variable 1-6 IPR007110 Ig-like dom
ENSP00000394447	5.72	10	-53.6	8.9	12.8	immunoglobulin heavy variable 3-74 IPR007110 Ig-like dom
ENSP00000375036	6.07	10	-49.4	8.5	12.8	immunoglobulin heavy variable 3-53 IPR007110 Ig-like dom
ENSP00000374810	5.59	5	-48.3	7.7	12.6	immunoglobulin kappa variable 1D-13 IPR007110 Ig-like dom
ENSP00000375018	5.68	8	-48.1	8.5	12.6	immunoglobulin heavy variable 3-23 IPR007110 Ig-like dom
ENSP00000375022	5.71	9	-45.0	9.1	13.0	immunoglobulin heavy variable 3-30 IPR007110 Ig-like dom
ENSP00000374778	5.31	7	-43.9	5.1	13.4	immunoglobulin kappa variable 4-1 IPR007110 Ig-like dom
ENSP00000418649	5.91	6	-41.5	4.9	12.5	immunoglobulin kappa variable 3-20 IPR007110

Table 6.4 Mass spectrometry protein read-out for ovary lysate IP. Identifier protein code, log (I), refractive index (rI), expectation value (log(e)), isoelectric point (pI) and molecular weight is provided along with a description of the protein.

6.4.6 PRDM9 mRNA expression analysis by RT q-PCR

Expression of PRDM9 mRNA was analysed in HEK293 cells by RT q-PCR. Primers were designed to align to the full length human PRDM9 mRNA transcript (NM_020227). Primer details are listed in Table 6.5. Reactions were set up as follows; 10 μL 2X iTaq™ Universal SYBR® Green supermix, 1 - 4 μM forward and reverse primer, 2 - 5 μg cDNA, made up to a final volume of 25 μL with dH$_2$O. All reaction plates were run on the Bio-Rad® iQ™5 thermocycler using the following program; denaturation at 95 °C for 30 seconds followed by 40 cycles of 95 °C for 15 seconds and 62.5 °C for 30 seconds. Immediately after product amplification melt curve analysis was performed using a 0.5 °C/2 seconds increment from 65 - 95 °C.

Primer	Forward Sequence 5'-3'	Reverse Sequence 5'-3'	Product length
PRDM9 1	AGGCTGTGAACTGCTGGTCTGG	GGCCCTTGAAATCTCCCTCTG	364 bp
PRDM9 2	GAAACCCTTGAGCCTTTGGC	AGAAGGCCCTGCTCCAATTC	77 bp
PRDM9 3	TTTGTCGTGCAGCGTGAAAC	GAAGGCCCTGCTCCAATTCT	91 bp
GAPDH	CTGACTTCAACAGCGACACC	ATGAGGTCCACCACCCTGT	94 bp

Table 6.5 Details of primers used for gene expression analysis by RT q-PCR amplification. Transcript target, forward and reverse primer sequence and amplicon product length are shown.

Amplification of a product using these primers was technically challenging. Firstly, PRDM9 primer set 1 was described in a publication assessing mRNA expression in human samples (Liu *et al.*, 2007). These primers were then tested with HEK293 whole cell RNA which was reverse transcribed to cDNA before RT q-PCR was performed. Reactions were attempted with increasing primer concentrations of 1, 2 and 4 μM using a 1 in 10 dilution series of cDNA, from a starting concentration of 2 μg. Product amplification was only achieved when 4 μM of each primer was used (Figure 6.23). Melt curve analysis confirmed that primers were specific for the PRDM9 transcript as they did not denature until temperatures of >85 °C (Figure 6.23 A & B). However, amplification was not achieved until after cycle 34 suggesting that the amount of transcript was very low in abundance (Figure 6.23 C). The PRDM9 threshold cycle (Ct) values from standard curve analysis showed that PRDM9 was being amplified and that there was no contamination in the negative control wells (Figure 6.23 D). However, the variation was large between triplicate samples and only the two highest concentration

standards showed any amplification suggesting again that the transcript level was low in the HEK293 cell sample.
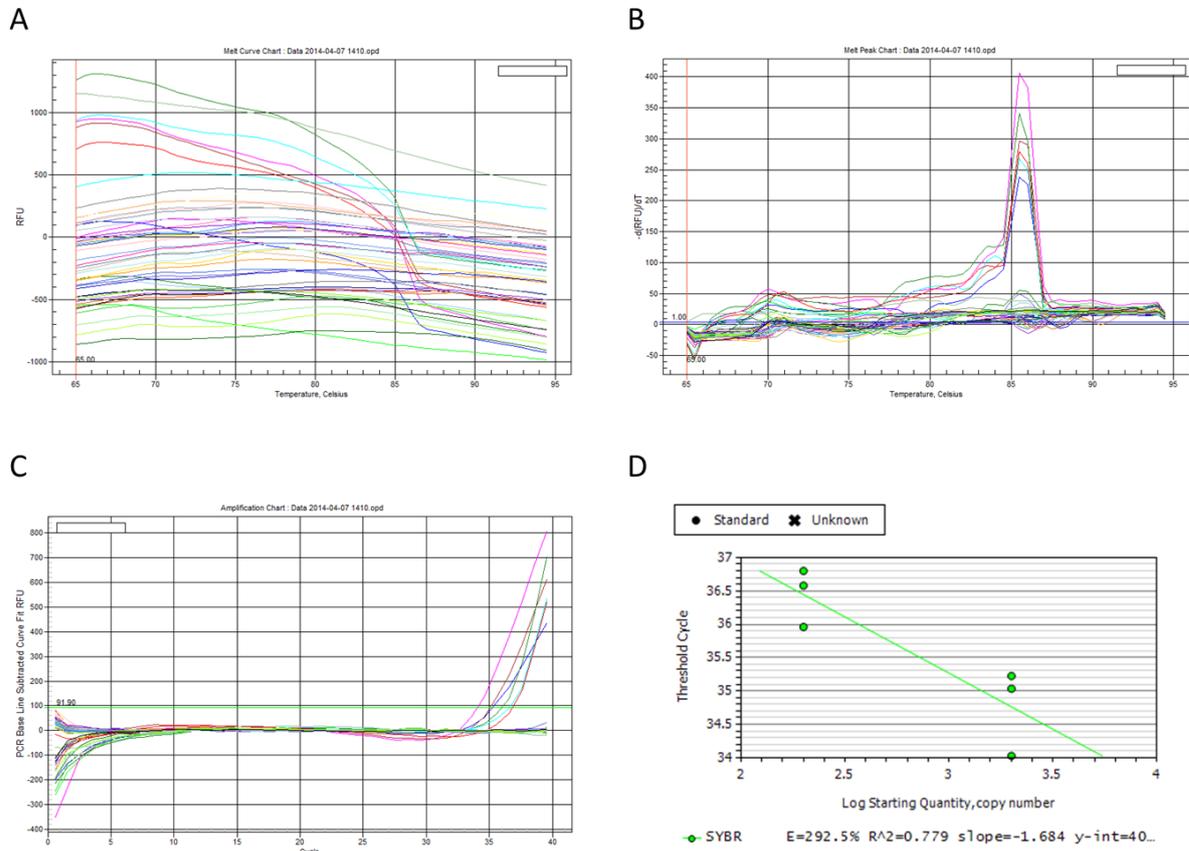
Figure 6.23 PRDM9 expression measured by RT q-PCR. Primer set 1 was used to amplify a 364bp PRDM9 product. A) Melt curve chart of primer set 1 in each sample well is shown by plotting relative fluorescence units (RFU) against temperature (°C). B) Melt peak chart of primer set 1 C) Amplification chart of primer set 1 product shown by plotting RFU against reaction cycle. D) Plot of threshold cycle values (Ct) against starting quantity of cDNA.

Next, primers for the house keeping gene glyceraldehyde 3-phosphate dehydrogenase (GAPDH) were tested using HEK293 cDNA. Efficient amplification of the product was achieved when using 1 μM of each primer and 2 μg of cDNA (Figure 6.24). Melt curve analysis showed specificity of the primers by denaturing at ~84 °C and in relation to the amount of template cDNA added to each reaction (Figure 6.24 A & B). Amplification of GAPDH product was achieved at 18 cycles and dilutions of standards showed a linear decrease in Ct values as cDNA concentration increased ten-fold (Figure 6.24 C & D).
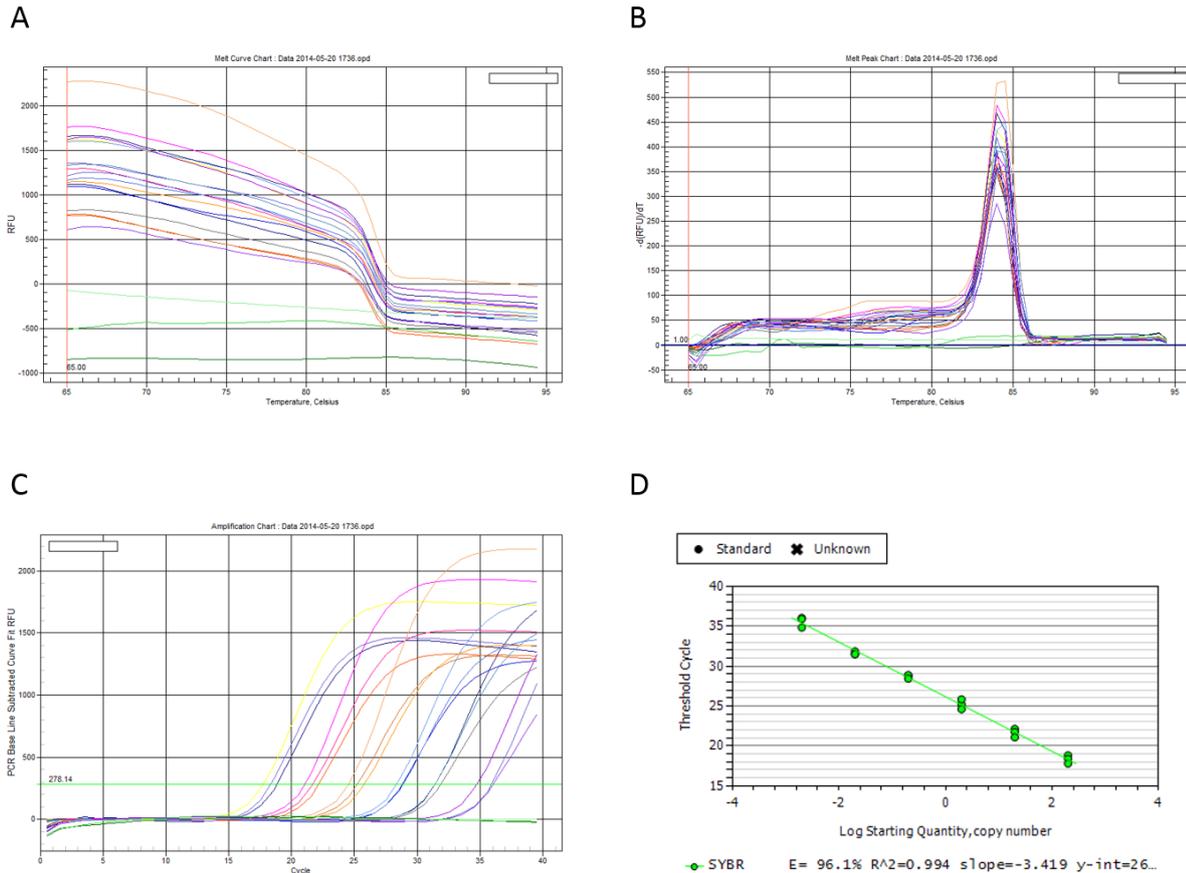
Figure 6.24 GAPDH expression measured by RT q-PCR. Primers specific for GAPDH transcript were used to amplify a product. A) Melt curve chart in each sample well is shown by plotting relative fluorescence units (RFU) against temperature (°C). B) Melt peak chart of C) Amplification chart of GAPDH product shown by plotting RFU against reaction cycle. D) Plot of threshold cycle values (Ct) against starting quantity of cDNA.

In order to increase PRDM9 product amplification, another set of primers specific for PRDM9 were designed. PRDM9 primer set 2 showed amplification when using 4 µM of each primer and 4 µg HEK293 cDNA (Figure 6.25). Amplification was achieved after 30 cycles of PCR amplification (Figure 6.25 A). This primer set was slightly less specific than primer set 1 as shown by melt curve analysis (Figure 6.25 B). This primer set did however give product amplification in four of the standard cDNA dilutions tested (Figure 6.25 C). Compared to GAPDH product amplification the PRDM9 Ct values were high suggesting that transcript abundance was low in the HEK293 cell sample (Figure 6.25 D).
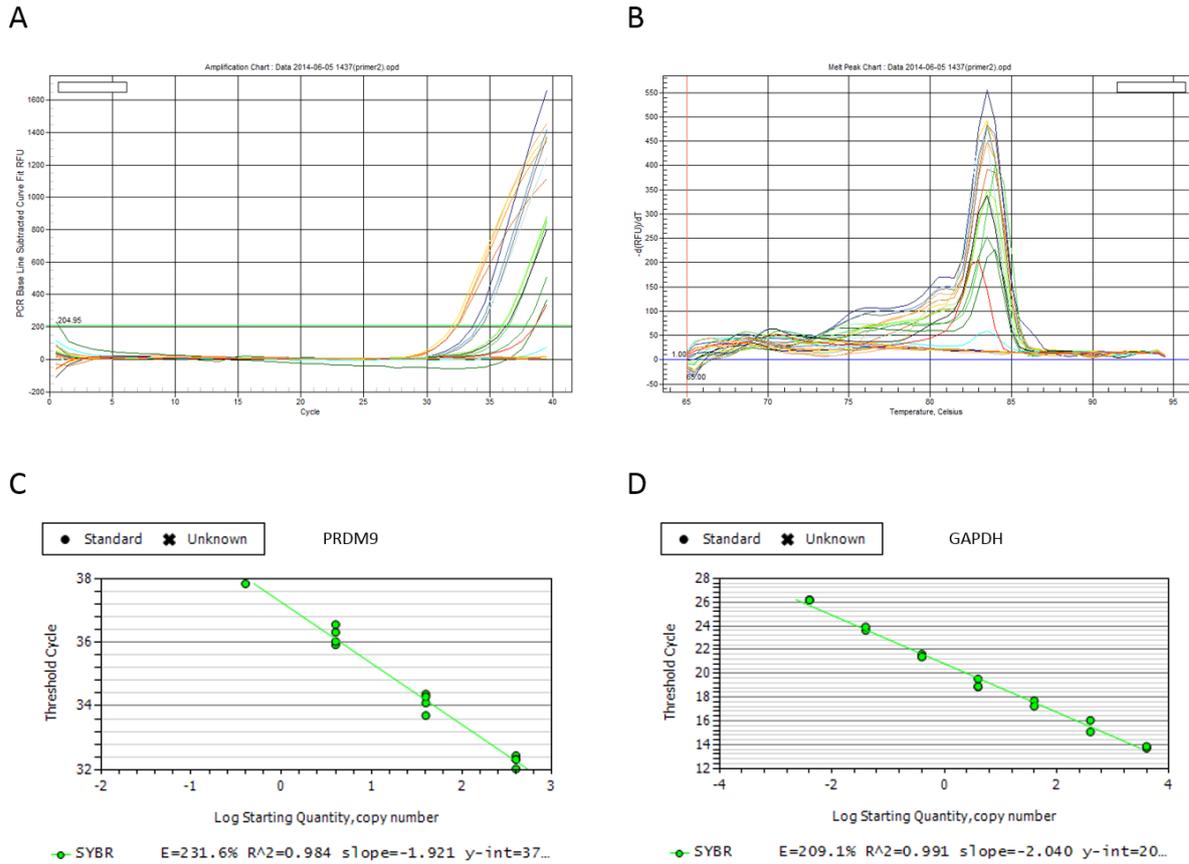
Figure 6.25 PRDM9 and GAPDH expression measured by RT q-PCR. A, B & C) PRDM9 product amplification using primer set 2. A) Amplification shown by plotting relative fluorescence units (RFU) against PCR amplification cycle number. B) Melt peak chart of PRDM9 primer set 2 shown by plotting RFU against temperature (°C). C) Plot of threshold cycle values (Ct) against starting quantity of cDNA. D) GAPDH product amplification. Plot of threshold cycle values (Ct) against starting quantity of cDNA.

A final primer set, PRDM9 primer set 3, was used and showed amplification using 4 µM of each primer with 5 µg of HEK293 cDNA (Figure 6.26). Again, amplification was not achieved until ~30 cycles of PCR amplification (Figure 6.26 A). This primer set appeared to be less specific than primer sets 1 and 2, as shown by melt curve analysis (Figure 6.26 B). Threshold values were variable between the sample triplicates (Figure 6.26 C). PRDM9 amplification was again very low when compared to the efficiency of GAPDH product amplification (Figure 6.26 D).
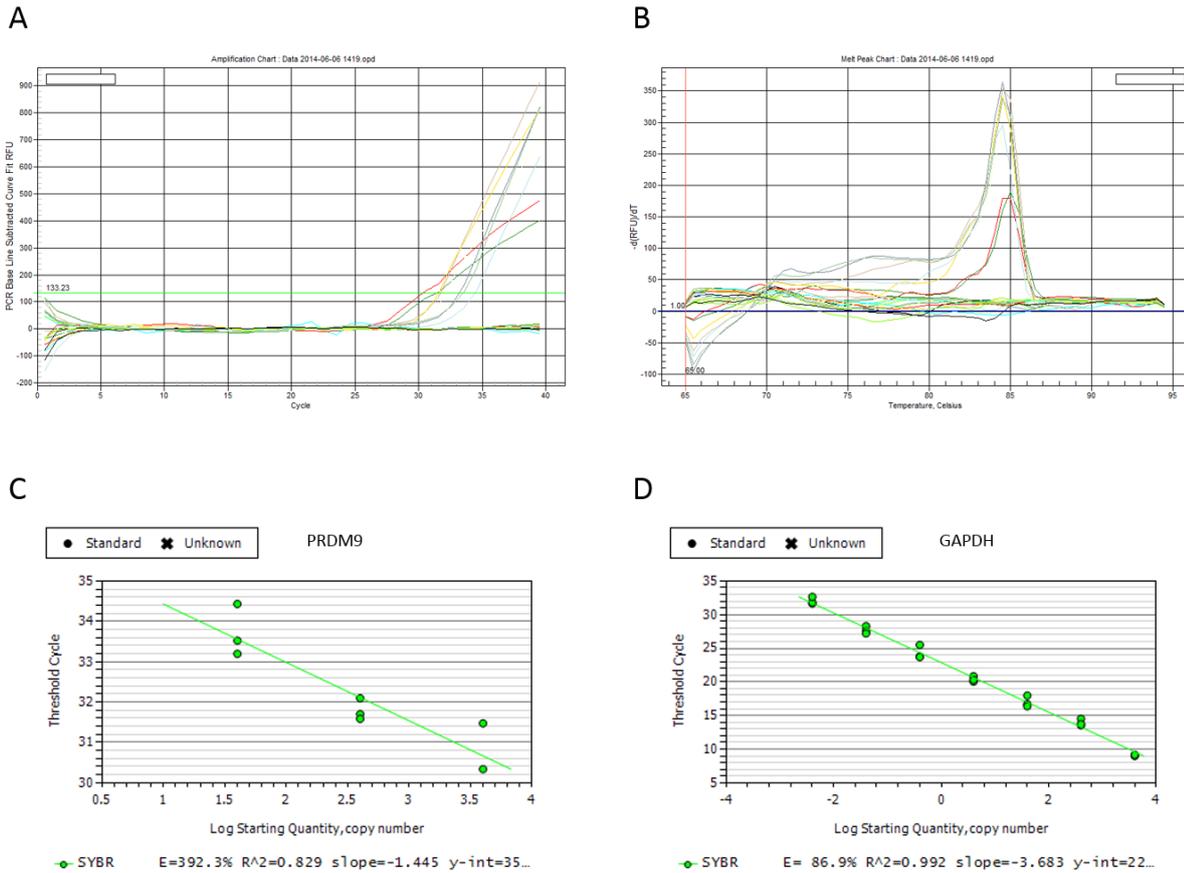
Figure 6.26 PRDM9 and GAPDH expression measured by RT q-PCR. A, B & C) PRDM9 product amplification using primer set 3. A) Amplification shown by plotting relative fluorescence units (RFU) against PCR amplification cycle number. B) Melt peak chart of PRDM9 primer set 2 shown by plotting RFU against temperature (°C). C) Plot of threshold cycle values (Ct) against starting quantity of cDNA. D) GAPDH product amplification. Plot of threshold cycle values (Ct) against starting quantity of cDNA.

Expression analysis of PRDM9 was attempted using three different primer sets designed to anneal to the mRNA transcript sequence. Primer set 2 appeared to amplify the product most efficiently however, this was only achieved when using high amounts of starting cDNA. Even when 5 μg of cDNA was used in the reaction the threshold cycle values were high suggesting that transcript levels were very low in the sample. Unfortunately due to a lack of appropriate tissue, a positive control of PRDM9 expression was not available to test the efficiency of the primer sets. Isolation of RNA from human testes tissue would be the most appropriate tissue in which to test the expression levels and compare the HEK293 cell sample.

166

6.5 Discussion

In order to detect human PRDM9 protein in cell lines and tissue samples a variety of assays were used including immunodetection, IP, IF, MS and q-PCR. Given the embryonic nature of the PRDM9 protein, tissues and cell lines were selected based on their embryonic origin. Tissue samples were selected based on the likelihood that they would contain cell types undergoing meiosis and therefore express PRDM9. Ovary, foetal gonad and placenta were tested for PRDM9 expression alongside skeletal muscle, presumed to not express PRDM9 as it is a post-mitotic tissue type.

Several cell lines available in the laboratory were tested for PRDM9 protein expression. Primary myoblast and fibroblast cell lines showed very low levels of protein expression when a PRDM9 antibody was used for detection. The nature of these cells in culture makes them slow to divide and they reach replicative senescence by 15-18 cell divisions. For this reason it was not feasible to grow sufficient amounts of cells for investigations which require high quantities of protein such as immunoprecipitation or subcellular localisation.

Protein was not detected using the PRDM9 antibody with lysate from the immortalised cell lines derived 143B and NT2. Therefore, cell investigations in this study primarily used HEK293 cells. This was due to the embryonic nature of this cell line as well as its ability to proliferate quickly and indefinitely, providing a large amount of cellular material for these investigations. When high quantities of protein were used, a protein band at ~100 kDa could be detected in HEK293 and human testis lysate leading us to believe that this cell line was appropriate for further investigation of PRDM9.

Subcellular fractions of HEK293 cells were also blotted for using PRDM9 and a range of antibodies against organelle specific proteins. Protein detected with the PRDM9 antibody showed the presence of two peptides which were exclusively localised to the mitochondrial fraction. However, multiple bands were detected using these samples and methods to separate the fractions further by dissociation of the MAM showed a completely different peptide profile to previous membranes.

Unfortunately, commercially available antibodies against PRDM9 appear to be non-specific and unreliable, giving multiple band detection on immunoblots as well as different protein banding profiles. This intra-assay variability means that any positive

PRDM9 detection is not reproducible. For this reason, PRDM9 localisation studies cannot confirm the presence of this protein in the mitochondrial fraction of HEK293 cells although enrichment of a ~60 kDa band was observed via subcellular fraction immunoblotting. Peptide blocking experiments were carried out to identify which protein band detected by the PRDM9 antibody was likely to be the desired protein. These experiments show that a ~55 kDa peptide was the target of this antibody however, this does not confirm that this commercially available antibody recognises full length human PRDM9 protein. Currently, three PRDM9 transcripts have been identified although only one is predicted to be protein encoding. The existence of two alternative transcripts could partially explain the different peptides identified by immunodetection however, further peptide characterisation would be needed to confirm this.

Immunofluorescent assays were attempted using two PRDM9 antibodies available. Staining of tissues appeared to be dependent on the condition and age of the tissue samples used. For example, ovary tissue had been stored for a long period of time whilst muscle, foetal gonad and placenta samples were frozen more recently. The F(ab) antibody was not appropriate for human tissues as the secondary (anti-human IgG) antibody reacted with multiple peptide antigens, not exclusively the anti-PRDM9 F(ab). In addition, the Novex antibody gave high levels of background staining on all samples tested apart from placenta. The IF images showed that high fluorescence was normally detected in fibrous regions of the tissues where there were connective structures but no DAPI positive cell types. This suggested that the nature of these regions prevented the antibodies (primary or secondary) from penetrating the tissue or prevented the wash out of excess antibody during the preparation of these samples.

Immunoprecipitation of PRDM9 was attempted in HEK293 and ovary tissue lysates. These IP experiments were largely unsuccessful again due to the unreliability of PRDM9 antibodies. Samples from IP were also sent for mass spectrometry analysis to confirm the presence or absence of PRDM9 protein. Unfortunately, both samples only contained contaminating peptides and antibody fragments from the reaction itself. PRDM9 was not detected in the MS analysis.

Furthermore, gene expression of PRDM9 was investigated by RT-qPCR amplification in HEK293 cells. Although the primer sets used appeared to be specific to the PRDM9

mRNA transcript there was a very low level of the product detected in the samples. Even when high quantities of cDNA were used as template the threshold cycle values appeared high suggesting that any PRDM9 transcript produced by the cell was in low abundance and could therefore not be reliably quantified. The amount of cDNA required for low level detection is achievable using immortalised cell RNA but would prove more difficult for tissue analysis.

Human PRDM9 is expressed during meiosis in gametocytes and therefore has so far only been detected in testis tissue, which continually produces new gametes in post-pubescent males as well as in foetal ovarian tissue at 23 weeks post conception (Liu *et al.*, 2007). PGCs are difficult to obtain and require specialist dissection and sorting techniques to ensure a pure cell population (Tang *et al.*, 2015). Collecting cells in this way from foetal tissue would result in very small amounts of cellular material and would therefore require high sensitivity single cell based assays to perform analysis for the presence of PRDM9. Primordial germ cell differentiation is possible using mouse ESCs and would provide a unique insight into characterising the expression profile of this protein either by transcript analysis or protein levels (Nakatsuji and Chuma, 2001). The main limiting factor of this experimental approach is the lack of a PRDM9 antibody against the murine protein.

Human testis would appear to be the most appropriate tissue in which to study PRDM9 expression given the availability of cells undergoing meiosis within this tissue. Availability of this tissue type is limited due to a lack of samples available for research. Most testicular biopsies from adult males are carried out to investigate infertility issues or suspected tumours (Ni *et al.*, 2016; Song *et al.*, 2016a; Zhang *et al.*, 2016). The amount of tissue taken at biopsy tends to be small and would therefore provide a limited amount of RNA, DNA or protein for analysis, especially after clinical investigations for which the tissue was taken have been performed.

Human PRDM9 expression analysis should either focus on utilising testicular tissue samples available for research or using primordial germ cell differentiation techniques. As shown in this study, PRDM9 protein did appear to be detectable in testis lysate and using freshly extracted protein and RNA would hopefully further confirm this finding. The low levels of PRDM9 expression coupled with rapid turnover of the protein will

make detection of PRDM9 extremely difficult. In this case, producing a cell culture based model or purifying recombinant protein would provide the largest amount of material for functional assays.

Chapter 7 Modifying PRDM9 expression levels in a cell culture model.

7.1 Overview

Previous studies attempting to elucidate the functional role and structure of the PRDM9 protein have focussed on *Mus musculus* protein as a model or partial PRDM9 constructs utilising recombinant protein technologies (Mihola *et al.*, 2009; Billings *et al.*, 2013; Eram *et al.*, 2014). These studies have interrogated the histone methyltransferase activity of the PR domain, which has homology to the well characterised SET domain; known to be essential for methyltransferase function in other protein families posessing this characteristic (Eram *et al.*, 2014; Powers *et al.*, 2016). Purified PRDM9 zinc finger array constructs have been used to assess DNA binding affinity and motif specificity (Davies *et al.*, 2016). Crystal structures of human PRDM9 have also been described, giving an insight into the potential functions of each region of the protein (Patel *et al.*, 2016). Other regions of the PRDM9 protein are predicted to have important functional roles including protein-protein interactions through the Krüpple-associated box (KRAB) domain and transcriptional regulation through the SSX repression domain (SSXRD), however these have not yet been experimentally determined. A schematic of PRDM9 protein structure is shown in Figure 7.1.
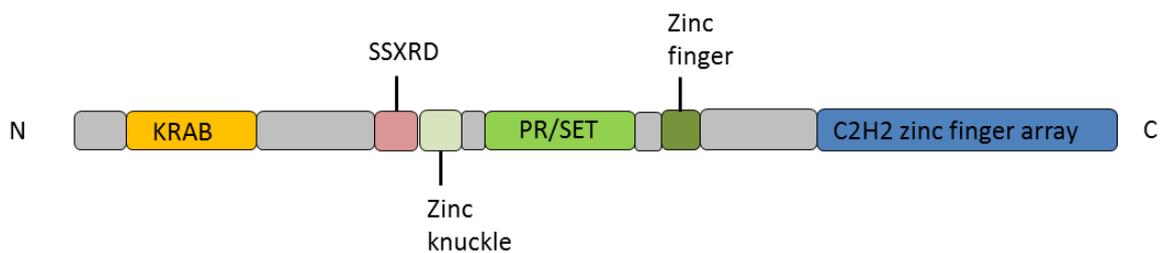
Figure 7.1 Structure of the human PRDM9 protein. KRAB, SSXRD, zinc knuckle, PR/SET, zinc finger and C2H2 zinc finger array are denoted.

7.2 Aims

The aims of this Chapter are;

1. To identify PRDM9 protein expression by immunodetection using transient knockdown.
2. To perform PRDM9 protein overexpression.

This will provide two different cell culture models for investigating some of the functions of this protein in mammalian cells.

7.3 Experimental Methods

Overlapping Methods are described in Chapter 3, Sections 3.8.6, 3.10.3 and 3.11.1.

7.3.1 PRDM9 siRNA transfection optimisation

The PRDM9 antibodies commercially available showed nonspecific binding by immunodetection methods outlined in Chapter 6. In order to further investigate whether any of the detected peptides were indeed PRDM9, knockdown of PRDM9 was attempted using short interfering RNA oligonucleotides (siRNA). This transient system allows targeted knockdown of mRNA species produced by the cell (Rao and Huang, 1979; Hillman *et al.*, 1987). Degradation of RNA by siRNA occurs through the RNA interference (RNAi) pathway involving many members of the Argonaute protein family, namely an RNAseIII-like protein (DICER) and RNA induced silencing complex (RISC) (Hammond *et al.*, 2001; Hutvagner and Zamore, 2002; Doi *et al.*, 2003). An siRNA oligo, specific to the PRDM9 transcript, was transfected into HEK293 cells in a dose dependent manner. After 72 hours of transfection, cells treated with siRNA targeted to PRDM9 displayed cytotoxicity with concentrations higher than 5 nM (Figure 7.2). Control cells treated with lipofectamine transfection agent or 20 nM non-targeting scramble siRNA did not display cytotoxic effects compared to untreated controls. Cells treated with 5 nM PRDM9 siRNA showed normal cell growth suggesting that this was the optimum concentration of this oligo in HEK293 cells (Figure 7.2).
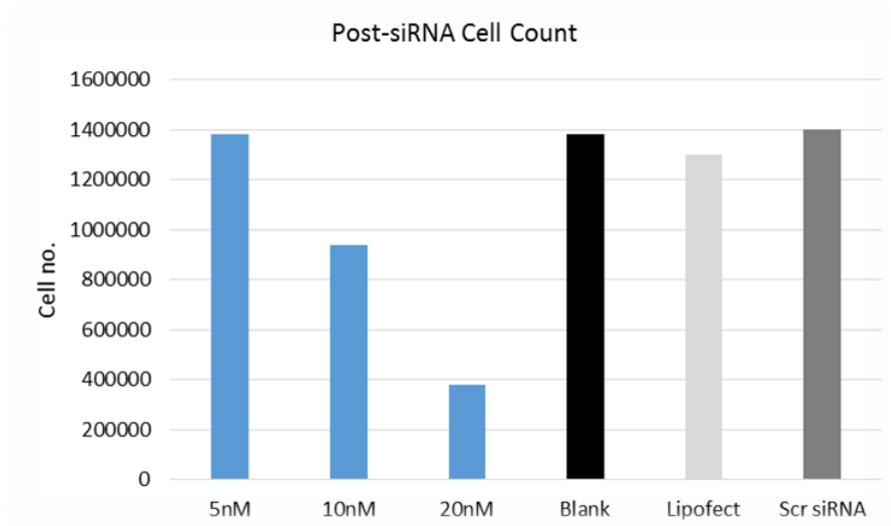
Figure 7.2 Cell counts of HEK293 cells after 72 hours of treatment with siRNA. Cells were seeded in 9.5cm^2 tissue culture plates at a density of 1.2x10^6 cells/well at the start of the experiment. Cells were either; untreated, PRDM9 siRNA treated (5, 10 or 20 nM) or non-targeting scramble siRNA (20 nM) treated for 72 hours before harvested and counted using trypan blue to assess cell death.

174

7.3.2 Overexpression plasmid validation

The transient overexpression system used in this study was the pcDNA3.1(+) plasmid containing full length *PRDM9* cDNA (Figure 7.3).

Figure 7.3 Plasmid map of pcDNA3.1(+) containing full length human PRDM9 cDNA. Cytomegalovirus (CMV) enhancer and TATA box (green), MCS denotes the multiple cloning site, BGH polyadenylation signal (pink), SV40 early polyadenylation signal (blue), ampicillin/hygromycin/kanamycin resistance elements (yellow), pUC ori sites on complementary strands (light blue).

Three different pcDNA3.1(+) plasmids were used in this study. Firstly, a pcDNA3.1(+)PRDM9 plasmid was a kind gift from Dr M.Vedadi (Toronto Childrens Hospital, Toronto, Ca). This plasmid contained full length human PRDM9 cDNA validated in the laboratory of Dr Vedadi. Secondly, an independent full length huPRDM9 pcDNA3.1(+)PRDM9 was designed (synthesised by Genscript, Ontario, CA). Thirdly, a partial huPRDM9 plasmid containing amino acids 195-385 was designed (synthesised by Genscript, Ontario, CA). These plasmids will herein be referred to as plasmid 1, plasmid 2 and plasmid 3 respectively.

All three plasmids were validated by PCR and Sanger sequencing of the PRDM9 insert sequence (Figure 7.4). Colony PCR amplification of purified plasmids 1 and 2 showed a single band at ~2700 bp corresponding to the full length *PRDM9* cDNA insert sequence (Figure 7.4 A & B). The PCR products were Sanger sequenced and compared to all available sequences using NCBI Basic Local Alignment Search Tool (BLAST). The constructs from both labs had 98% amino acid similarity to human PRDM9 protein. Plasmid 3 was also PCR amplified and as predicted, a clear band corresponding to partial *PRDM9* at ~740 bp in size was observed (Figure 7.4 B). This construct had 97% amino acid sequence similarity to human PRDM9. Colony number 1 was chosen to be purified for each plasmid (Figure 7.4; lane 1 A, lanes 1 and 5 B).
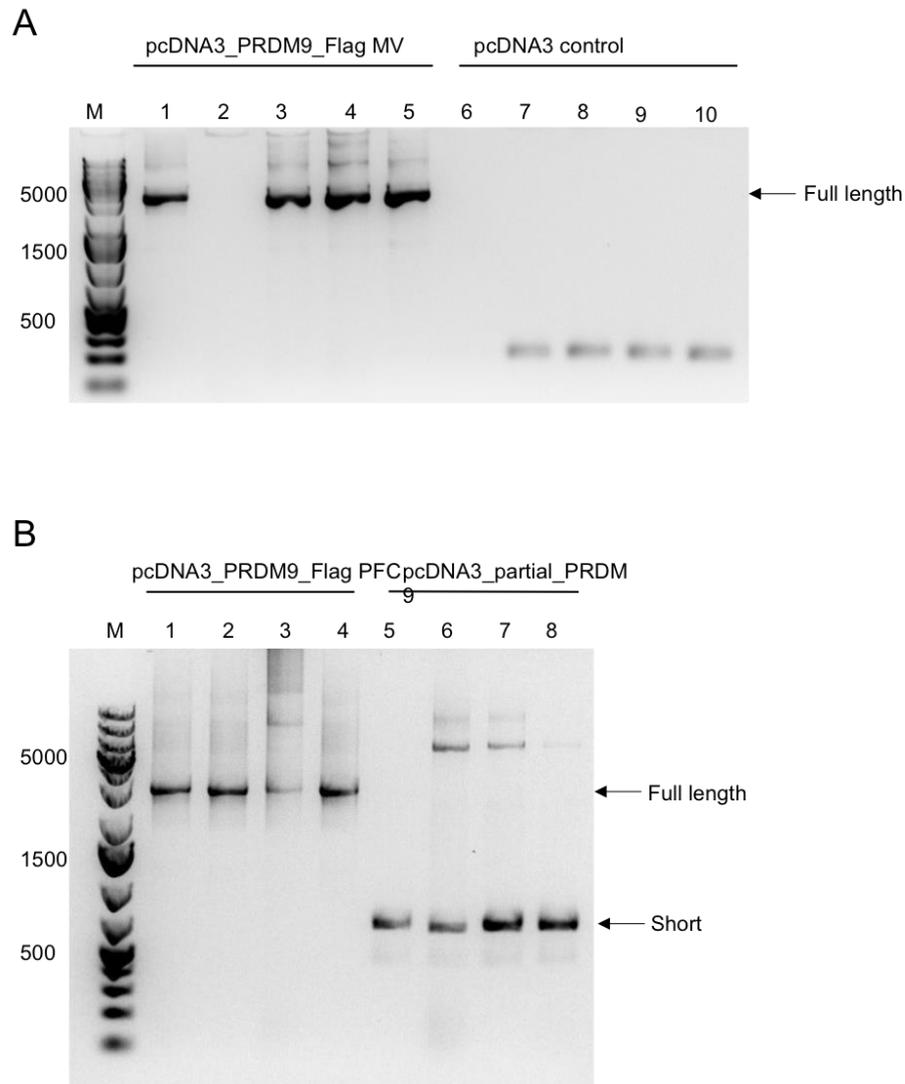
Figure 7.4 Validation of PRDM9 plasmids by PCR amplification. A) Colony PCR of pcDNA3_PRDM9_Flag plasmid 1 (provided by the Vedadi lab). Lanes 1-5: PCR product of insert DNA from colonies transformed with the pcDNA3_PRDM9_Flag construct. Lanes 6-10: PCR product of insert DNA from colonies transformed with pcDNA3 plasmid DNA as a control. B) Colony PCR of pcDNA3_PRDM9_Flag plasmid 2 and partial pcDNA3_PRDM9_Flag plasmid 3. Lanes 1-5: PCR product of insert DNA from colonies transformed with plasmid 2. Lanes 6-10: PCR product of insert DNA from colonies transformed with plasmid 3.

7.3.3 Transfection efficiency optimisation

In order to easily monitor the transfection efficiency of HEK293 cells using the pcDNA3.1 system, a plasmid expressing green fluorescent protein (GFP) was kindly provided by Professor Hanns Lochmüller (Institute of Genetic Medicine, Newcastle University, UK). The expression of this protein can be visualised whilst the cells are in culture using a GFP filter on a benchtop microscope (Zeiss Axiovert). Images were collected at 24 and 48 hours post transfection and compared (Figure 7.5). Using 6 µg of plasmid DNA for 24 hours appeared to be sufficient for a satisfactory level of transfection.
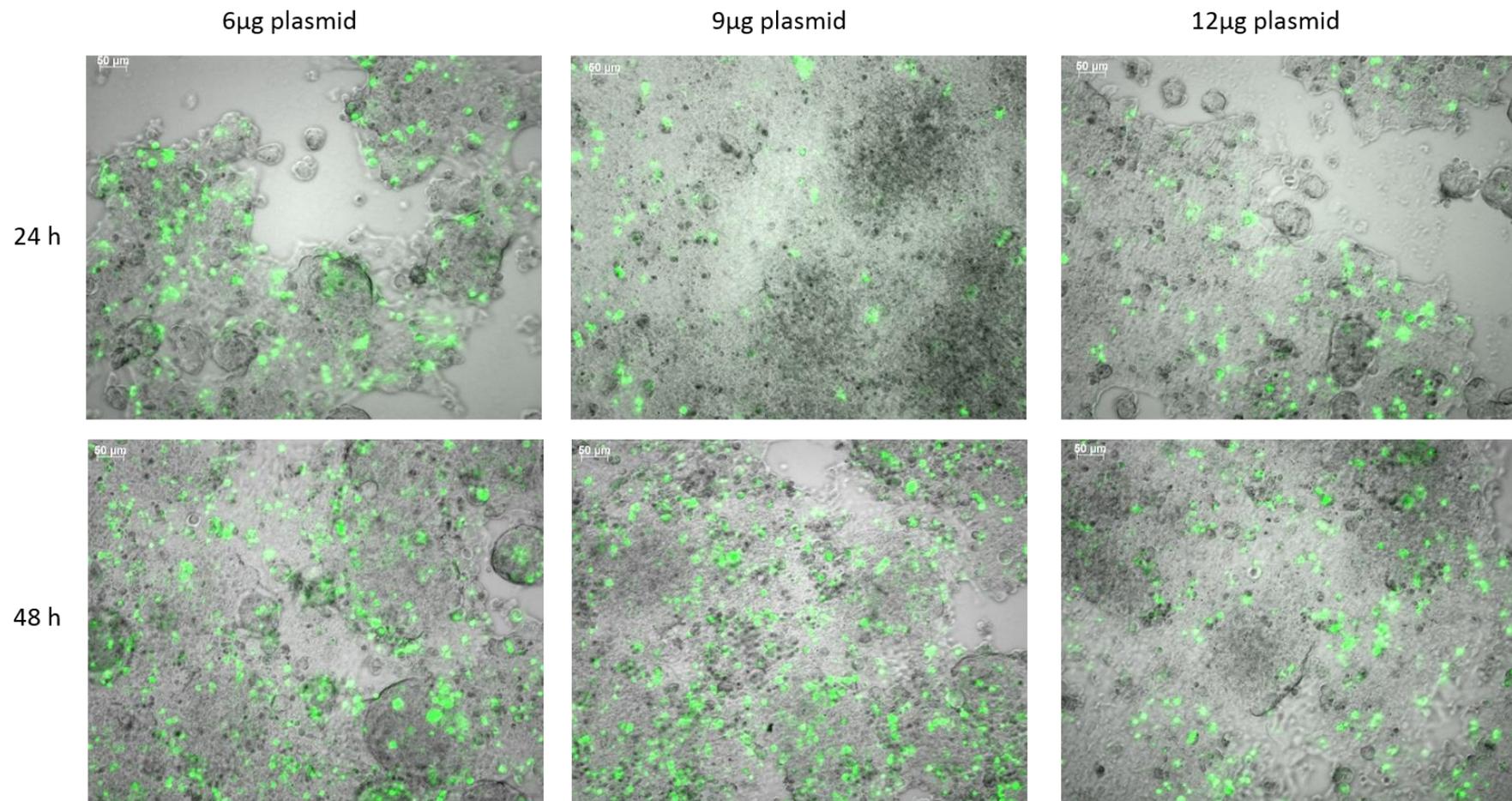
Figure 7.5 Transfection of HEK293 cells with a green fluorescent protein (GFP) expressing pcDNA3.1 plasmid. Cells were transfected with 6, 9 or 12 µg plasmid DNA for 24 hours (top panel) and 48 hours (bottom panel). Cells were imaged under normal culture conditions using the GFP filter on a bright-field microscope (Zeiss Axiovert).

As there was no difference in transfection efficiency between 6 and 12 μg plasmid DNA, the plasmid was tested at a lower amount of 3 μg (Figure 7.6). When compared to the 6 μg plasmid transfection image (Figure 7.6 B), there was less GFP signal but still a high level of transfection when using 3 μg plasmid DNA (Figure 7.6 A).
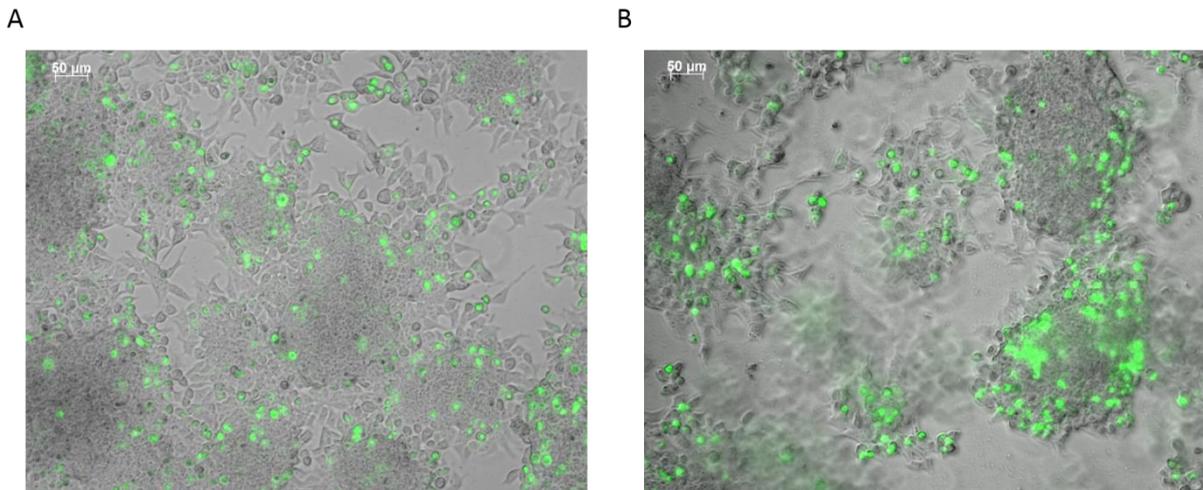
Figure 7.6 Transfection of HEK293 cells with a green fluorescent protein (GFP) expressing pcDNA3.1 plasmid. Cells were transfected with 3 (A) or 6 (B) μg plasmid DNA for 72 hours Cells were imaged under normal culture conditions using the GFP filter on a bright-field microscope (Zeiss Axiovert).

7.4 Results

7.4.1 PRDM9 knockdown by siRNA

Expression of PRDM9 protein was previously tested in HEK293 cells by various protein detection methods (Chapter 6). By using this model it was observed that multiple peptides were detected using available antibodies against PRDM9. Blocking anti-PRDM9 with immunising peptide showed that a peptide of 55 kDa was the target of this antibody (Figure 6.6). To elucidate whether any of these detected peptides were specifically PRDM9, knockdown was attempted using siRNA. Cells were transfected for 48 and 96 hours before pellets were collected and protein expression investigated by immunoblotting. Using the anti-PRDM9 antibody, three peptide bands could be detected in all PRDM9 siRNA samples as well as control treatments (Figure 7.7 A & B). These peptides were ~75, ~55 and ~30 kDa in size which correlated with previous immunoblots using this antibody (Chapter 6). After 48 hours of treatment, there was a slight reduction in the expression level of the 55 kDa band in the 20 nM siRNA treated cells compared to 5 and 10 nM treatments (Figure 7.7 A). However, none of these peptides were detected in the scramble siRNA control and only two bands were detected in cells treated with transfection agent lipofectamine only, suggesting that this was not a PRDM9 specific observation and was most likely due to variability of the antibody.

The 55 kDa band detected in the immunoblots in Figure 7.7 were quantified by measurement and normalisation of densitometric band intensity relative to β-actin (Figure 7.7 C). Detection of this band was previously shown to be blocked by addition of the PRDM9 immunising peptide indicating that it is the primary peptide bound by the commercially available anti-PRDM9 antibody. Quantification of samples after 48 hours siRNA treatment showed that this peptide was present at higher levels in the 10 nM siRNA treated cells compared to untreated or control samples (Figure 7.7 C). Quantification of samples after 96 hours siRNA treatment showed no difference in band intensity between treated or control samples (Figure 7.7 C). This observation confirms that this siRNA knockdown was either not PRDM9 specific or that PRDM9 was not present within the cell line and could therefore not be targeted for degradation.
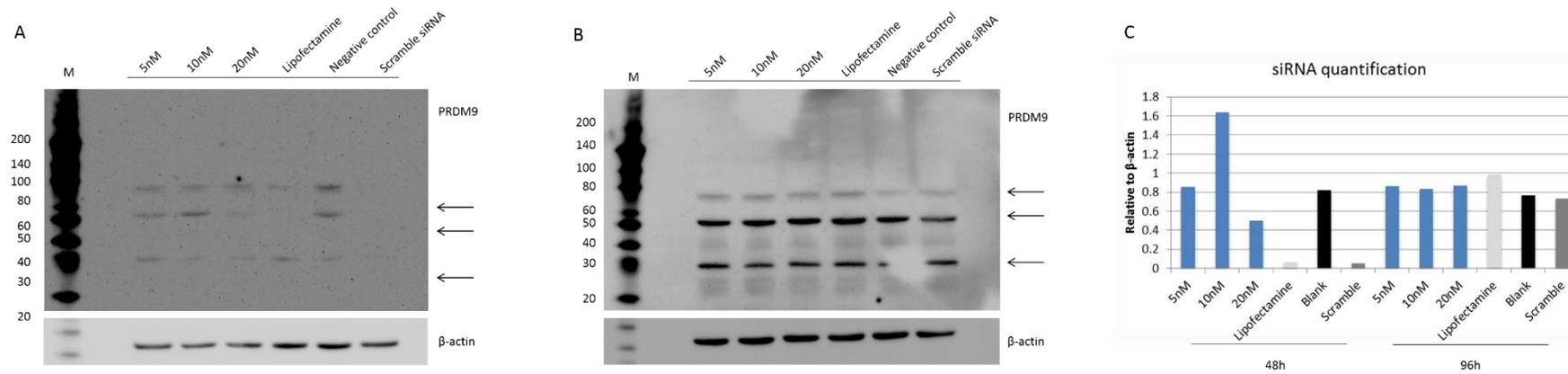
Figure 7.7 Immunoblot detection of PRDM9 in HEK293 cells treated with PRDM9 siRNA. 60 μg of protein was loaded onto 4-20% bis-tris gels. A) Cells were treated for 48 hours with siRNA targeting PRDM9 mRNA. B) Cells were treated for 96 hours with siRNA targeting PRDM9 mRNA. Membranes were incubated with anti-PRDM9 (1:500) or anti-βactin (1:1000) overnight at 4°C. C) Quantification of PRDM9 after 48 and 96 hours of treatment with siRNA. Immunoblot detection of a 55 kDa peptide band was quantified by optical density relative to β-actin.

7.4.2 Transient overexpression of PRDM9 in HEK293 cells

To determine whether PRDM9 protein could be detected in a cell culture model, transient overexpression was attempted using plasmid 1. Cells were transfected with 1 µg of Flag-PRDM9 or empty pcDNA3.1 plasmid DNA for 24, 48 or 72 hours and compared to non-transfected control HEK293 cells. Cells were then harvested and lysed for total cellular protein. Immunodetection using an antibody against Flag tag protein did not detect Flag peptide in any of the samples (Figure 7.8).
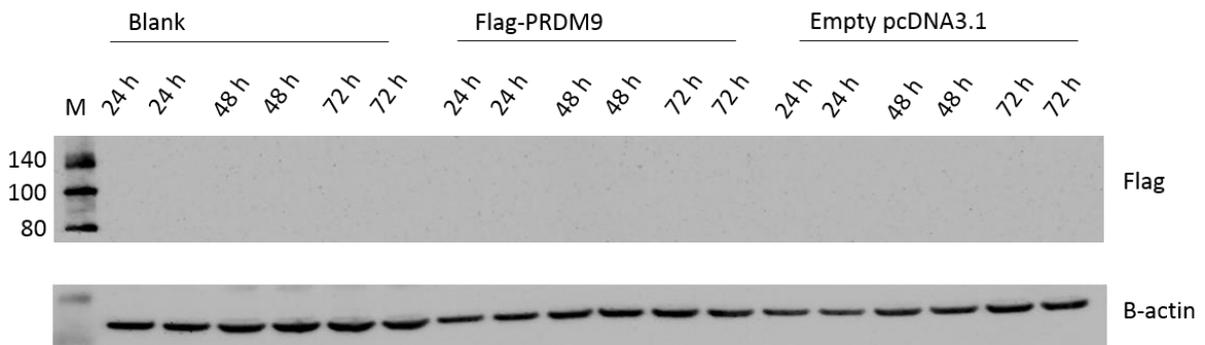
Figure 7.8 Immunodetection of Flag-PRDM9 in cells transfected with overexpression plasmid DNA. 1 µg plasmid DNA was transfected for 24, 48 or 72 hours. 75 µg protein was loaded on 4-12% bis-tris gels in duplicate. Blank lysates were untreated HEK293 cells grown in parallel with transfected cells. Vector lysates were HEK293 cells treated with pcDNA3.1 plasmid not containing any insert cDNA. Protein size marker was included on the gel (M), sizes are denoted in kDa. Membranes were incubated with anti-Flag (1:1000) or anti-βactin (1:1000) overnight at 4 °C.

Next, expression was attempted with 3 and 6 μg of plasmid 1 DNA for 72 hours in parallel, with plasmids expressing GFP and doking protein 7 (DOK7) as transfection and antibody detection controls respectively (Figure 7.9). The pcDNA3.1 plasmid containing Flag tagged DOK7 cDNA was a kind gift from Dr Juliane Muller (Institute of Genetic Medicine, Newcastle University, UK). DOK7 protein was detected at the predicted weight of ~55 kDa using 3 μg plasmid DNA with a marked increase in protein band intensity using 6 μg suggesting that the transfection efficiency using this method is satisfactory for protein overexpression. Peptide bands were detected in all samples at ~80 and ~100 kDa in size using the Flag antibody, implying that these bands were non-specific antibody binding events. This PVDF membrane was overexposed to ensure that no PRDM9 protein was present and this most likely lead to the detection of non-specific peptides (Figure 7.9).
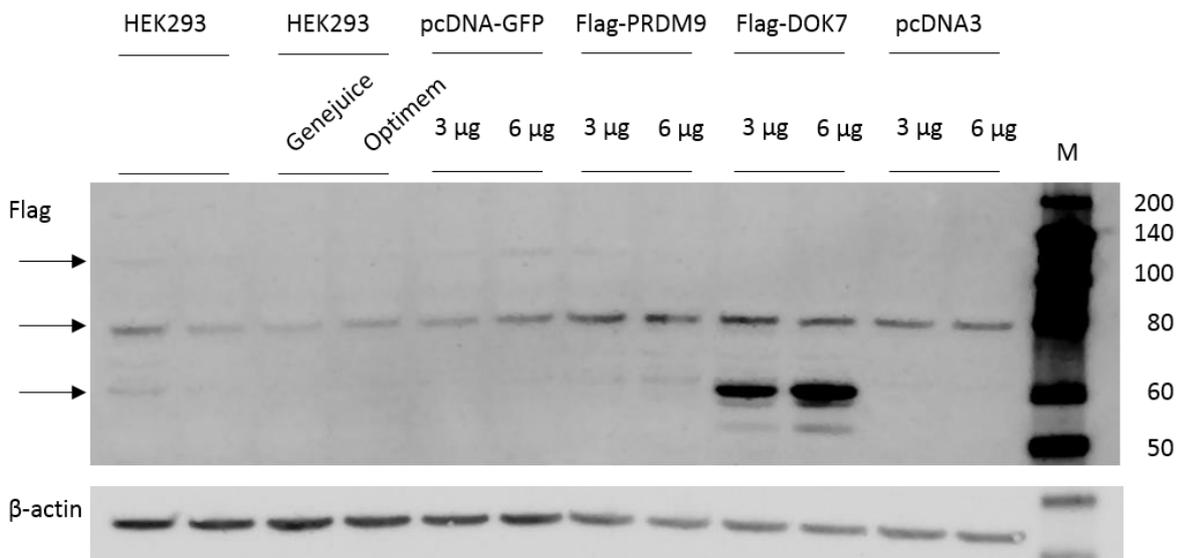
Figure 7.9 Immunoblot detection of PRDM9-Flag overexpression. 75 μg of protein was loaded onto 4-20% bis-tris gels. Cells were transfected for 72 hours with 3 or 6 μg plasmid DNA. Untreated controls are denoted as 'Blank'. Protein size marker was included on the gel (M), sizes are denoted in kDa. Membranes were incubated with anti-Flag (1:1000) or anti-βactin (1:1000) overnight at 4 °C.

184

Higher quantities of plasmid were used for cell transfection. Plasmid DNA at 6, 9, 12, 15 and 18 µg was transfected to enable detection of PRDM9 protein (Figure 7.10). When the Flag antibody was used for immunodetection, a single band was present at ~100 kDa in the PRDM9 transfected samples only (Figure 7.10). Again, DOK7 protein was also detected at a molecular weight of ~55 kDa, detectable at shorter exposure times than the 100 kDa PRDM9 band suggesting that the DOK7 expression plasmid was producing much higher amounts of detectable protein than the PRDM9 expression plasmid.
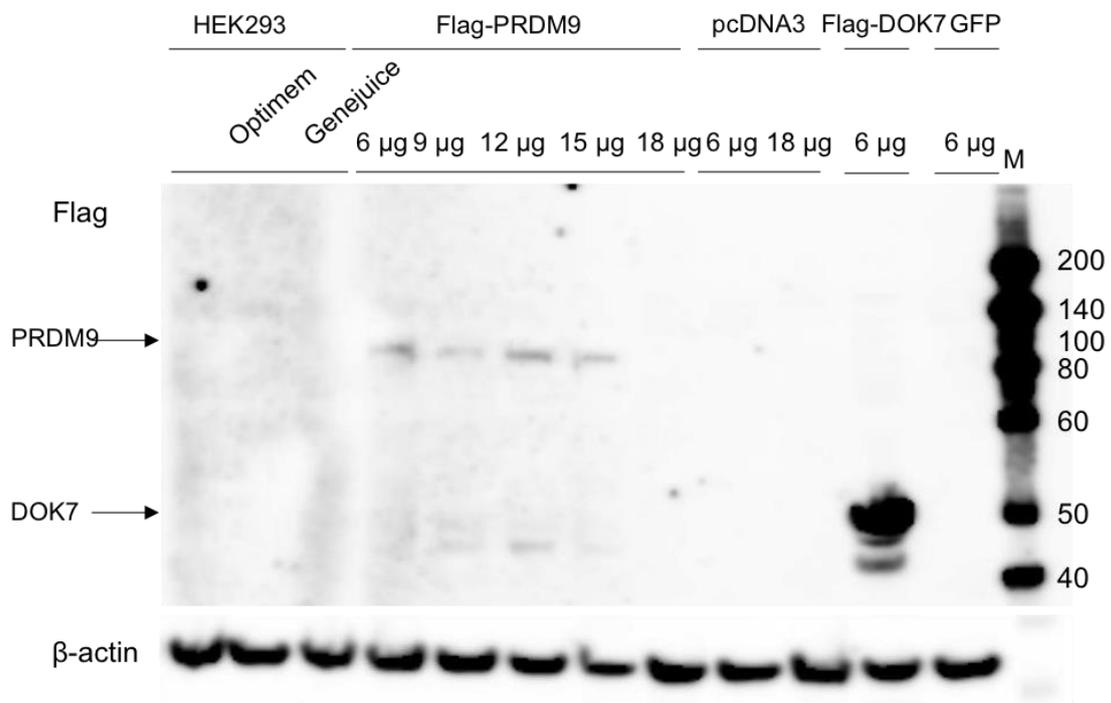
Figure 7.10 Immunodetection of PRDM9 in an overexpression cell model. Cells were transfected with 6, 9, 12 or 18 µg plasmid DNA for 72 hours. 75 µg of protein was loaded onto 4-20% bis-tris gels. Protein size marker was included on the gel (M), sizes are denoted in kDa. Membranes were incubated with anti-Flag (1:1000) or anti-βactin (1:1000) overnight at 4 °C.

In order to determine whether the inefficiency of Flag-PRDM9 overexpression level was due to the protein itself or the plasmid used, two new pcDNA3.1(+)-PRDM9 contructs were designed (synthesised by Genescript). Cells were transfected with plasmid containing full length PRDM9 cDNA (plasmid 2) or partial PRDM9 cDNA (amino acids 195-385, plasmid 3) for 72 hours before being analysed for expression by immunoblotting. Plasmids 2 and 3 were transfected in 1, 3 or 5 μg quantities to determine the optimum amount of plasmid required to induce protein expression (Figure 7.11). Immunoblot detection using the Flag antibody showed that full length PRDM9 was produced after transfection (Figure 7.11 A). A clear band was detected at ~100 kDa in all samples transfected with plasmid 2, with the strongest detection observed using 1 μg of plasmid. Weak bands were also detected at ~140, ~80 and ~70 kDa. The plasmid expressing DOK7 was used as a positive control and as expected produced a single band at ~55 kDa. Flag protein was not detected in non-transfected HEK293 cells further supporting that the peptide detected in the transfected samples is Flag-PRDM9.

As expected, a low molecular weight protein was detected after plasmid transfections with plasmid 3 (Figure 7.11 B). A single band, at ~28 kDa, was strongly detected at all plasmid 3 amounts transfected. This band was detected using a shorter exposure time and expressed at higher levels than that of the full length PRDM9 protein product suggesting that the partial protein is more easily expressed using this system. In addition, when compared to the DOK7 control plasmid, the partial PRDM9 protein overexpression appeared to be more efficient than the full length construct.
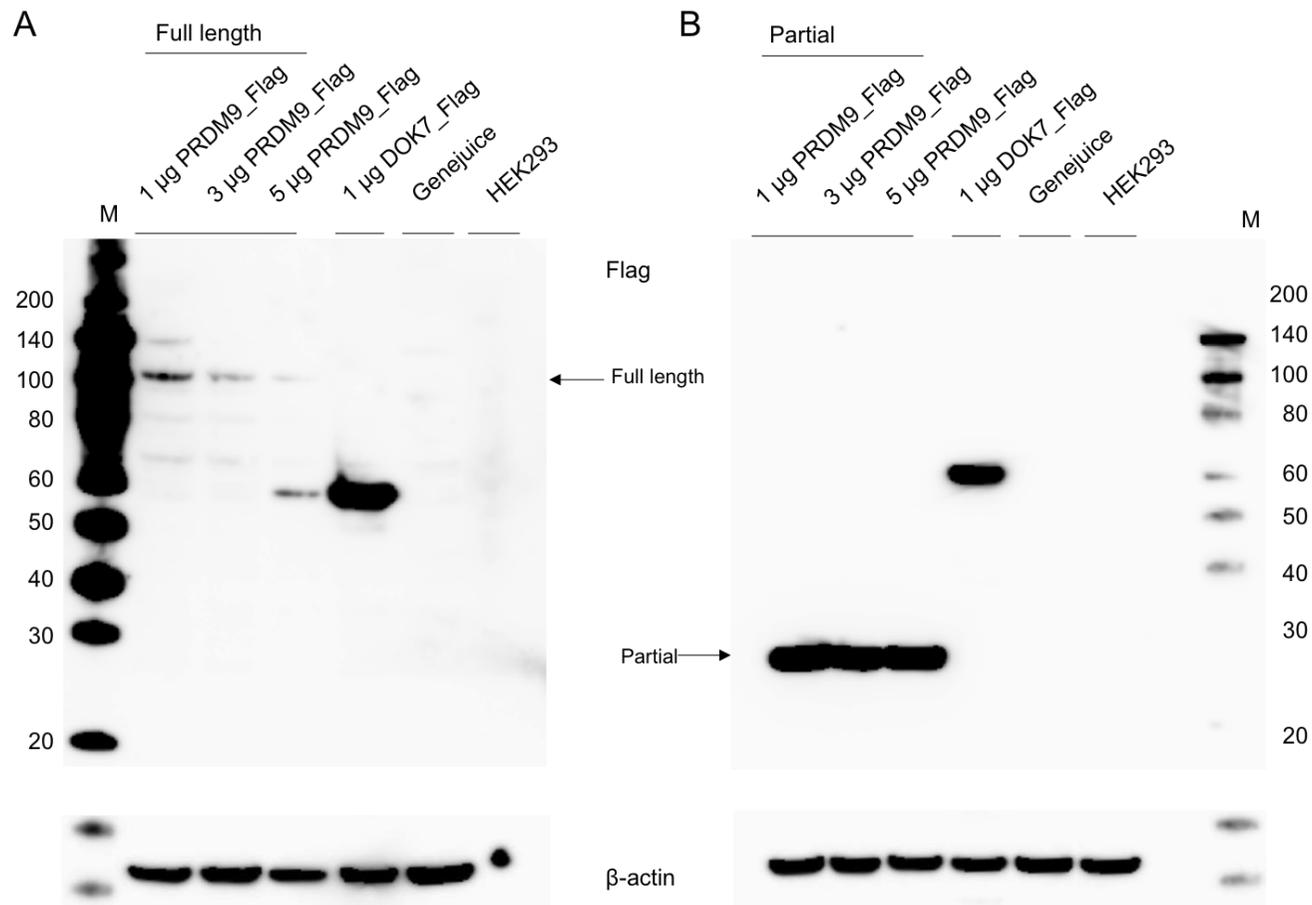
Figure 7.11 Immunodetection of PRDM9 in an overexpression cell model. A) Full length huPRDM9 protein was transiently expressed in HEK293 cells for 72 hours. DOK7 construct was also transiently expressed as a control. Cell lysate was electrophoresed in 4-20% bis-tris gels and Western blotted. B) Partial huPRDM9 protein was transiently expressed in HEK293 cells for 3 days. DOK7 construct was also transiently expressed as a control. Cell lysate was electrophoresed in 4-20% bis-tris gels and Western blotted. Protein size markers were included on each gel, sizes are denoted in kDa. Membranes were incubated with anti-Flag (1:1000) or anti-βactin (1:1000) overnight at 4 °C.

Detection of full length PRDM9 by immunoblotting was achieved twice using this transient transfection system. To assess whether low levels of transfection efficiency could be detected in single cells, immunofluorescent staining was attempted. Due to the nature of HEK293 cells, imaging protocols were developed using poly-L-lysine coated glass cover slips to enhance cell attachment and limit the loss of cells during the staining protocol. This cell line tends to clump, with poor adhesion to glass or plastics, rather than a monolayer, making it difficult to image single cells in isolation. Protocols were therefore developed to image the cells after 24 hours of transfection with plasmid DNA whilst the cells were at a low density (~60% confluent), hence forming a monolayer. Antibodies against mitochondrial proteins; heat shock protein 60 kDa (HSP60), optic atrophy 1 (OPA1), PEO1 and TFAM or Flag protein were optimised on non-transfected HEK293 control cells (Figure 7.12). Positive staining was achieved using anti-HSP60, which has mitochondrial localisation but functions as a chaperone from the cytoplasm to the inner mitochondrial matrix, therefore also appears to have cytosolic localisation (Figure 7.12 A). Staining with anti-OPA1 showed clear mitochondrial localisation and staining of the entire mitochondrial network (Figure 7.12 B). Similarly, staining with anti-PEO1 (Twinkle helicase) showed mitochondrial network staining due to its location within the mitochondrial matrix within proximity of mtDNA (Figure 7.12 C). Unfortunately, anti-TFAM stained cells did not show positive mitochondrial staining possibly due to non-specificity of this particular antibody or due to its suitability for immunofluorescent staining procedures (Figure 7.12 D). As OPA1 staining gave the clearest mitochondrial staining signal it was used in subsequent staining experiments. Staining for Flag protein did not show any positive staining as expected in this cell line (Figure 7.12 E). HEK293 cells stained with secondary antibodies only showed that there was no auto-fluorescence at antibody dilutions of 1:1000 (Figure 7.12 F).
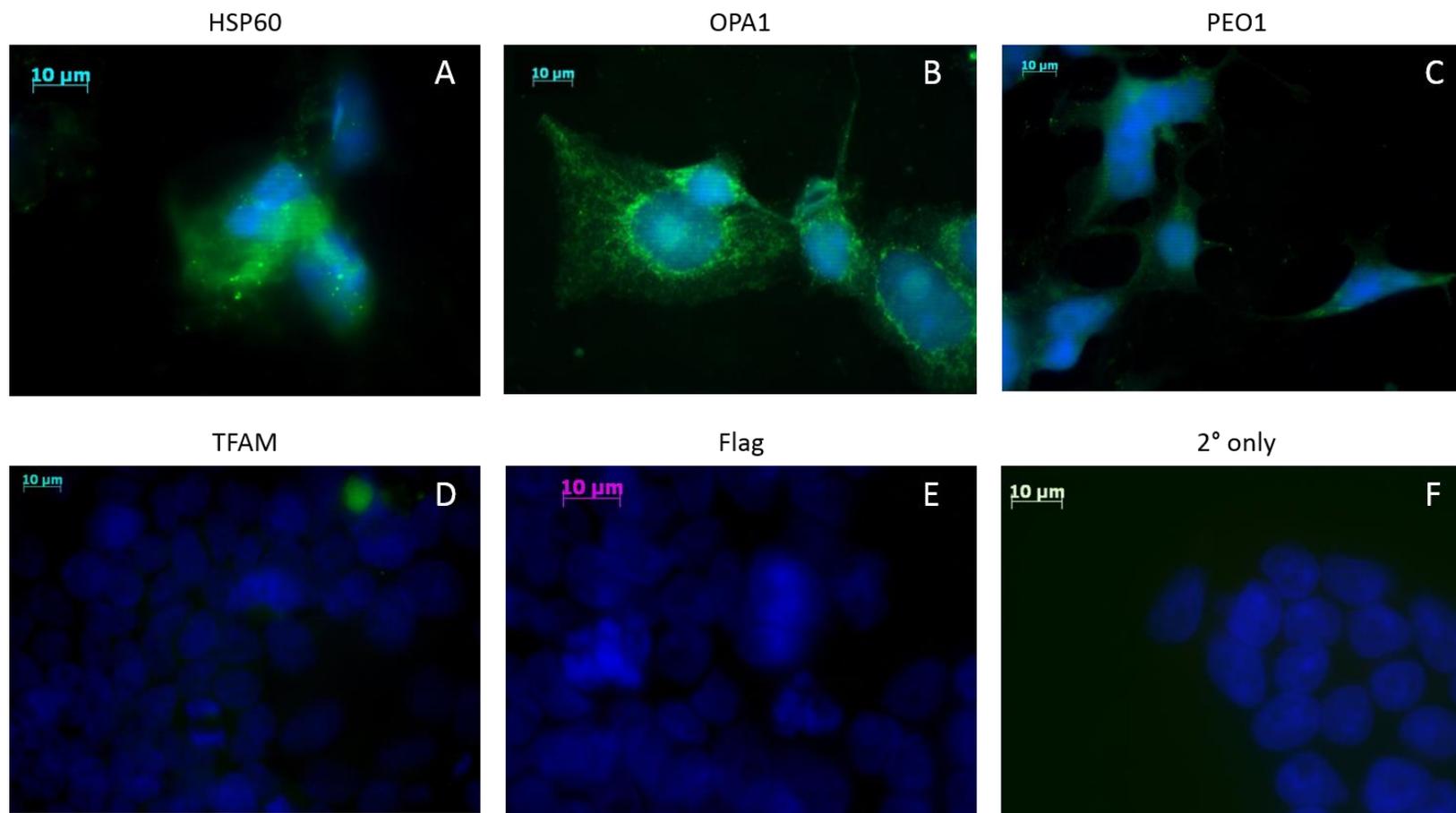
Figure 7.12 Immunofluorescent staining of HEK293 cells with antibodies against mitochondrial proteins. Cells were seeded at 2 x 10^5 cells/well and grown on poly-L-lysine coated glass coverslips for 24 hours. A) Staining with anti-PEO1 (1:100) overnight at 4 °C and anti-rabbit alexafluor®-488 secondary (green). B) Staining with anti-HSP60 (1:200) overnight at 4 °C and anti-rabbit alexafluor®-488 secondary (green). C) Staining with anti-OPA1 (1:250) overnight at 4 °C and anti-rabbit alexafluor®-488 secondary (green). D) Staining with anti-TFAM (1:100) overnight at 4 °C and anti-rabbit alexafluor®-488 secondary (green). E) Staining with anti-Flag (1:1000) overnight at 4 °C and anti-mouse alexafluor®-594 secondary (red). F) Staining with alexafluor®-488 (green) and alexafluor®-594 (red). All slides were stained with DAPI nuclear stain (blue) and imaged on the AxoImager (Zeiss).

Cells were transfected with each of the three overexpression plasmids: full length Flag-PRDM9, partial Flag-PRDM9 or control Flag-DOK7. Transfection was carried out using 1 μg plasmid DNA for 24 hours before the cells were stained with both anti-OPA1 and anti-Flag (Figure 7.13). Anti-OPA1 (green) showed mitochondrial network staining in all transfected samples although some single cells appeared to have very little visible mitochondria or clustering of mitochondria around the nucleus (shown in blue). This is most likely due to the rounded cell shape and clumped nature of HEK293 cells. Staining of Flag protein (red) showed primarily nuclear localisation of both full length and partial PRDM9 proteins (Figure 7.13 A, B, D and E). There also appeared to be Flag signal in the cytoplasm of the cells. Interestingly, Flag signal in the DOK7 expressing cells also showed nuclear localisation (Figure 7.13 C and F). This protein is known to play a key role in the function of the neuromuscular junctions of skeletal muscle cells and therefore usually localises to the plasma membrane (Okada *et al.*, 2006). As HEK293 cells are distinct in their properties from muscular and neuronal cell types it would be reasonable to assume that localisation of this protein would be different in this cell type. However, nuclear localisation of DOK7 might indicate that there is aberrant expression within the HEK293 cells. Perhaps this protein is aggregated within compartments surrounding or close to the nucleus or perhaps there is delivery of the protein into the nucleus. The images taken are from one plane and do not represent any 3D image stacks (such as Z stacks), therefore localisation cannot be reliably determined as it could be that the 3D orientation of the cell shows that any 'nuclear' localisation is in fact just cytoplasmic signal close to or surrounding the cell nuclei.
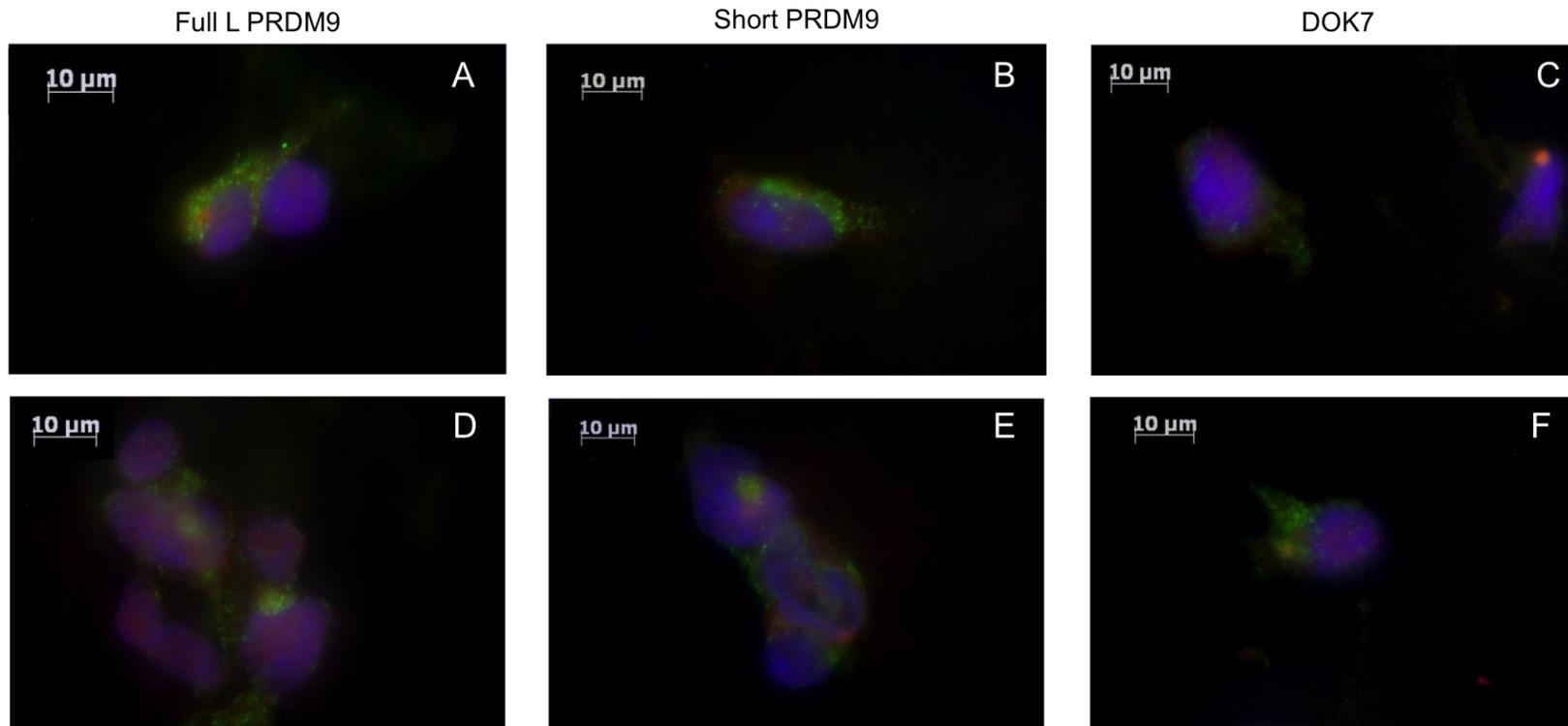
Figure 7.13 Immunofluorescent staining of transfected cells using OPA1 and Flag antibodies. Cells were seeded at 2 x 10^5 cells/well and grown on Poly-L-lysine coated glass coverslips for 6 hours before transfection with 1 μg plasmid DNA for 24 hours. Coverslips were incubated with anti-OPA1 (1:250) and anti-Flag (1:1000) overnight at 4 °C followed by secondary antibodies (Alexafluor®-488 or Alexafluor®-594) at 1:1000 dilution. A&D) Cells were transfected with full length Flag-PRDM9 pcDNA3.1 plasmid DNA. B&E) Cells were transfected with partial Flag-PRDM9 pcDNA3.1 plasmid DNA. C&F) Cells were transfected with Flag-DOK7 pcDNA3.1 plasmid DNA. DAPI was used as a nuclear stain on all coverslips (blue). All images were taken using the Axioimager (Zeiss) at x40 magnification.

7.5 Discussion

PRDM9 protein expression level was modified in a cell culture model; firstly using targeted siRNA knockdown and secondly using plasmid overexpression. Experimentally altering the PRDM9 transcript and protein levels was attempted because previous analysis showed PRDM9 was not reliably detected in several cell lines and tissue samples (Chapter 6). Creating an appropriate experimental model is essential to further understand the functional role of this protein.

To determine which peptide band was actually the target of the PRDM9 commercial antibodies described in Chapter 6, knockdown of PRDM9 was attempted in HEK293 cells using a pre-designed siRNA transfection system. Due to the inaccuracy of the PRDM9 antibodies available as well as a lack of positive control for PRDM9 expression, it was not possible to effectively assess whether the siRNA oligo had silenced the PRDM9 transcripts using this system. A band at 55 kDa was successfully blocked when incubated with the immunizing PRDM9 peptide (Chapter 6), however this band was not a specific target of the siRNA oligo used in this study. This suggests that the 55 kDa peptide detected is either not PRDM9, is a PRDM9 peptide from an alternative transcript or does not contain the region of the sequence targeted by this particular siRNA oligo.

As PRDM9 is a meiotic specific protein, it is unlikely that the full length functional transcript is expressed in HEK293 cells. In this case, there would be no target for the PRDM9 specific siRNA oligo to bind to and degrade via the DICER and RISC RNA interference pathway. To test this hypothesis, a cell line or tissue type expressing PRDM9 would need to be used in the knockdown experiment to confirm that this system is specific to PRDM9 and that it is measurable either by protein expression analysis or q-PCR of mRNA levels. The tissue expressing the highest levels of detectable PRDM9 protein is human testis and this tissue type, or cells derived from such tissue would be the most appropriate positive control for this siRNA experiment.

Overexpression of PRDM9 was attempted using three different plasmids. Firstly, transfection with plasmid 1 produced modest PRDM9 expression. Western blot detection showed a protein band at the expected size of ~100 kDa but only in cells transfected with very high quantities of plasmid. Unfortunately, this overexpression was only achieved once using this system and was therefore not reliable for use in further

experiments. It is not clear why plasmid 1 transfection was not efficient, it could be due to protein folding based on the amino acid sequence used to generate the peptide. Transfection with plasmid 2 showed expression of a band at the expected size of ~100 kDa when cells were transfected with only 1 μg of plasmid DNA. The overexpression product was produced at very low levels when compared to a control transfection using a DOK7 construct as well as a loading control. To test whether this was caused specifically by the construct cDNA sequence used, a partial PRDM9 construct was also synthesised and transfected. This smaller protein product was highly expressed in our system when compared to the DOK7 and loading control protein levels. In conclusion, partial PRDM9 constructs express detectable peptides much more efficiently than full length PRDM9 constructs. This is likely due to the smaller overall size of the plasmid being transfected as well as the faster transcription and translation rates associated with smaller DNA sequences. Following this observation, localisation experiments could be performed using partial PRDM9 peptides to elucidate whether any portion of the protein has mitochondrial localisation.

Differences in overexpression efficiency could be due to the number of cells successfully transfected with the plasmid or the stability of the full length protein product. Transfection efficiency estimated using a pcDNA3 plasmid expressing GFP, showed that only ~30% of the cells are transfected with and express the desired protein using this transfection protocol. Several key steps are involved in expression of transfected DNA constructs; accessibility of plasmid DNA (pDNA) to transcription machinery within the cytosol, which is dependent on endosomal escape and dissociation of mRNA complexes; the stability of the nucleic acids; the extent of compaction of pDNA within the lipo- or lipopoly-plex used for delivery and finally the ability to unpackage nucleic acids once inside the cytoplasm (Goncalves *et al.*, 2016). The pDNA must be delivered into the cell by either electroporation, heat shock or lipid coating, however, recent assay developments for mammalian cell culture recommend cationic lipid based delivery as the most efficient and least cytotoxic method (Jafari *et al.*, 2012; Cai *et al.*, 2016). Of course, this delivery method presents possible problems in terms of pDNA unpackaging followed by accessibility to the transcription and translation machinery required for expression. Lipid coated DNA is much less likely to be accessed by cellular components than naked DNA due to the longer time required for the pDNA to escape its protective

coating (Prasad *et al.*, 2003). Another consideration when altering expression by plasmid transfection is the role of cellular responses to the transfected pDNA. For example, clatherin-dependent or calveolae-mediated endocytosis are the mechanisms by which the pDNA is taken up by the cell and transported internally (Das *et al.*, 2016). The pDNA must be taken up efficiently by these endosomes and then released so as to be available for transcription/translation machinery. Activation of immune-response or cellular stress pathways to clear away any foreign DNA molecule might also be upregulated leading to a dampening down of the effectiveness of the transfected plasmid either by exocytosis or degradation of the molecule (Poecheim *et al.*, 2016).

Overexpression of full length PRDM9 protein using the transient pcDNA3.1 plasmid system allowed low levels of protein to be detected via immunoblotting and immunofluorescence. However, results were variable and unreliable using these two detection techniques as only three PVDF membranes and only one immunofluorescent assay successfully showed detection of PRDM9. Immunoprecipitation followed by mass spectrometry would be the best method to determine whether the positive detection observed was in fact PRDM9 protein. In addition, all transfections were performed on cells grown in 9.5 cm^2 tissue culture wells, yielding a relatively low quantity of protein, not sufficient for mitochondrial extractions, localisation or DNA binding studies.

Achieving PRDM9 expression to perform localisation studies and DNA binding assays using the system described would require large amounts of transfection agent and plasmid DNA. To overcome the technical challenges surrounding the investigation of this protein, cell lines containing stable overexpression constructs were created as described in Chapter 8.

Chapter 8 Characterisation of a PRDM9 stable overexpression cell culture model.

8.1 Overview

PRDM9 protein levels were undetectable in tissue samples or cell lines investigated previously (Chapter 6) most likely due to the fact that the expression of this protein may be highly regulated during meiosis. In addition, transient overexpression systems outlined in Chapter 7 were inconclusive and did not provide enough material for mitochondrial investigations which typically require substantial amounts of cellular material. To address this, two stable cell lines were created with full length PRDM9 cDNA constructs under the tetracycline inducible promoter (Hillen and Berens, 1994; Hinrichs *et al.*, 1994; Yao *et al.*, 1998). One cell line was created using human PRDM9 cDNA, whilst the other contained an identical PRDM9 cDNA with an N-terminal Flag-tag (Figure 8.1).
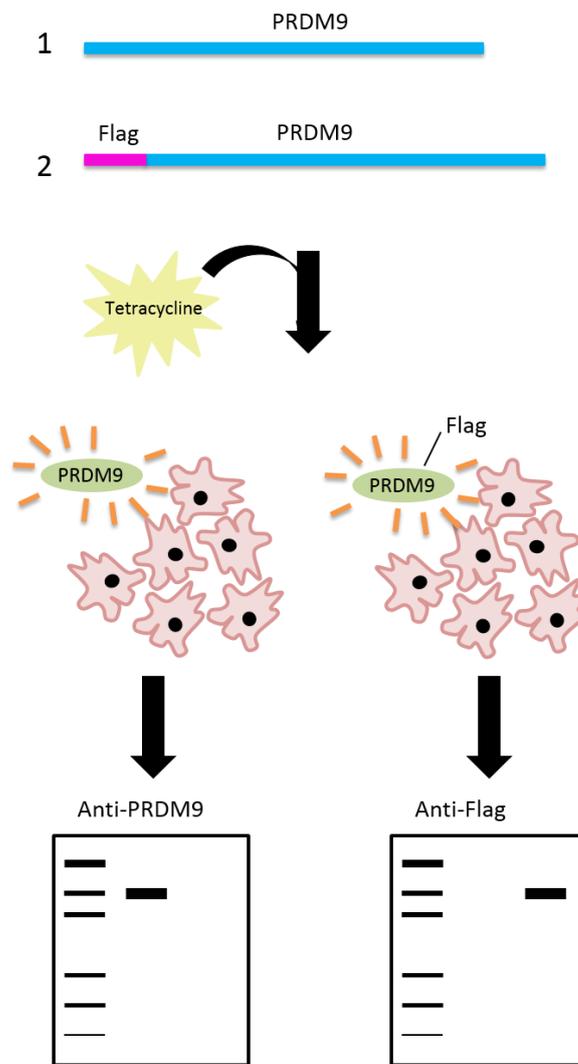
Figure 8.1 Overview of the system used in this study to stably overexpress PRDM9 cDNA in mammalian cells. Two PRDM9 cDNA constructs are depicted 1. no tag, 2. Flag tag. After integration of the constructs into the genome, tetracycline is added to the cells in culture. Expression of PRDM9 is induced. PRDM9 protein can be detected using anti-PRDM9 or anti-Flag antibodies.

8.2 Aim

The aim of this study was to create a stable cell culture model expressing PRDM9 protein using an inducible promoter to control the induction and level of expression. This is advantageous as it would allow large quantities of cells overexpressing PRDM9 to be used for experiments such as mitochondrial fractionation. Using a Flag-tagged cDNA construct allows protein detection independent of PRDM9 antibodies.

8.3 Experimental Method

Brief methods are detailed in Section 3.11.2.

8.3.1 Overexpression plasmid validation

Plasmid map of the pcDNA5/FRT/TO vector containing PRDM9 cDNA is shown below with multiple cloning sites depicted (Figure 8.1). The location of the PRDM9 cDNA insert sequence can be seen in orange; in this vector the insert is under the control of the CMV promoter but with tetracycline operator elements in between (Figure 8.1). Tetracycline must bind to these elements before expression of the insert gene can occur.

Figure 8.2 pcDNA5/FRT/TO containing the *PRDM9* cDNA insert sequence. Plasmid elements are: cytomegalovirus (CMV) enhancer and TATA box (green), multiple cloning site (MCS), BGH polyadenylation signal (pink), SV40 early polyadenylation signal (blue), ampicillin/hygromycin/kanamycin resistance elements (yellow), pUC ori sites on complementary strands (light blue), tetracycline operator elements and Flp recombination target (FRT) site are shown in black.

The PRDM9 cDNA sequence described previously (Chapter 6) was sub-cloned from the pcDNA3.1 plasmid into pcDNA5/FRT/TO plasmid by restriction enzyme digestion (Figure 8.3). To create a PRDM9 insert construct without a tag, ApaI and BamHI enzymes were used for DNA digestion (Figure 8.3 A). The PRDM9 cDNA insert sequence with an N-terminal Flag-tag was digested using XhoI and HindIII enzymes (Figure 8.3 B). Products from both digests were electrophoresed on 1% agarose gels and visualised by UV exposure to show a clear band at ~2.7 Kb corresponding to the PRDM9 insert sequence (Figure 8.3).
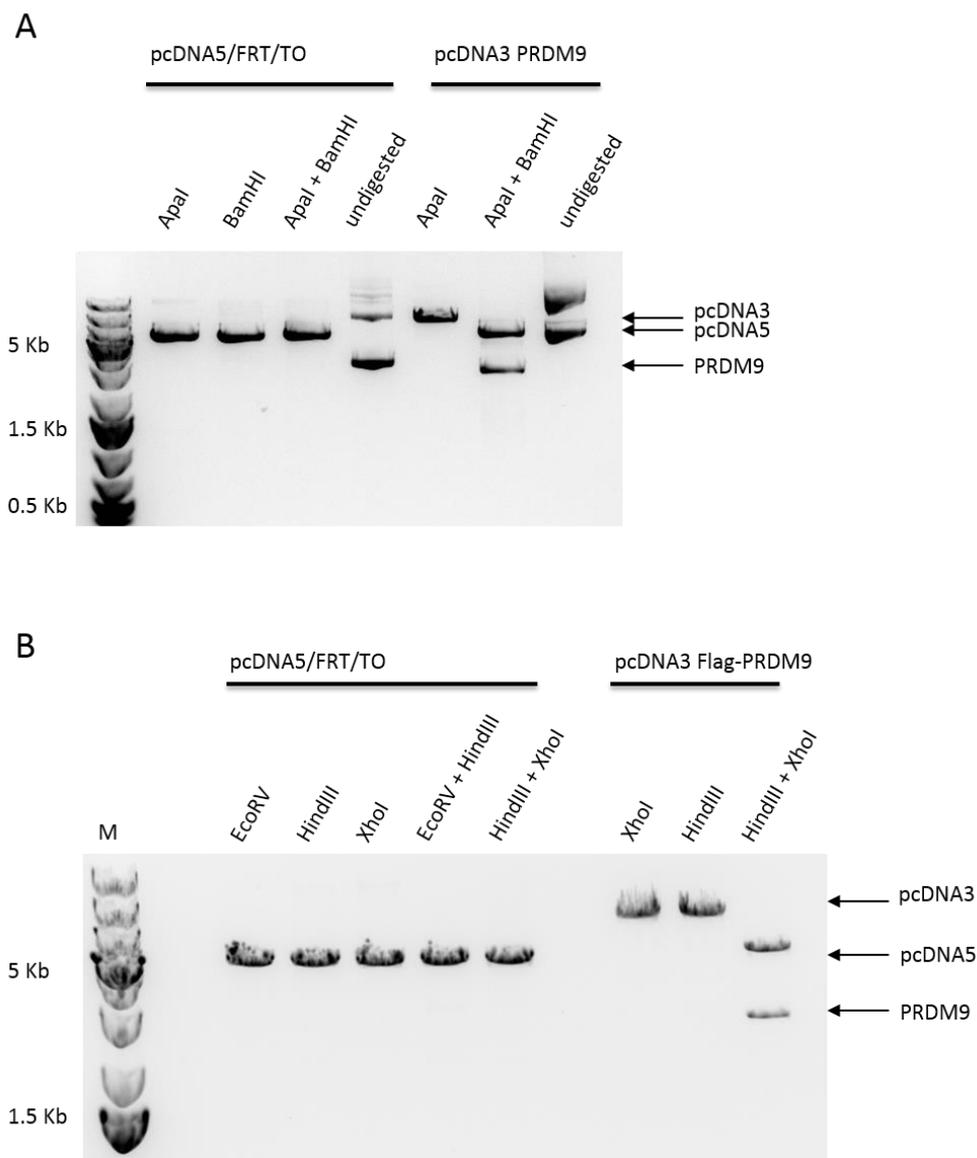
Figure 8.3 Sub-cloning of full length PRDM9 cDNA from pcDNA3 into pcDNA5/FRT/TO. A & B) Restriction enzyme digestion products were resolved on 1% agarose gels by electrophoresis. DNA size marker was included on the gel (M), sizes are given in base pairs. A) Digestion of full length PRDM9 cDNA using ApaI and BamHI enzymes. B) Digestion of Flag tagged PRDM9 cDNA using XhoI and HindIII enzymes. pcDNA3.1: ~8111 bp, pcDNA5/FRT/TO: ~7837 bp, PRDM9: ~2700 bp.

198

The PRDM9 product was gel extracted and used in subsequent ligation reactions with digested and gel extracted pcDNA5/FRT/TO plasmids (Figure 8.3). Successfully ligated plasmids were determined by colony PCR of competent *E.coli* transformed with the ligation product. Two sets of primers were used, one pair aligned to the CMV promoter sequence and pcDNA5 plasmid sequence, the other pair aligned to the CMV promoter and the BGH polyadenylation sequence (Figure 8.4).
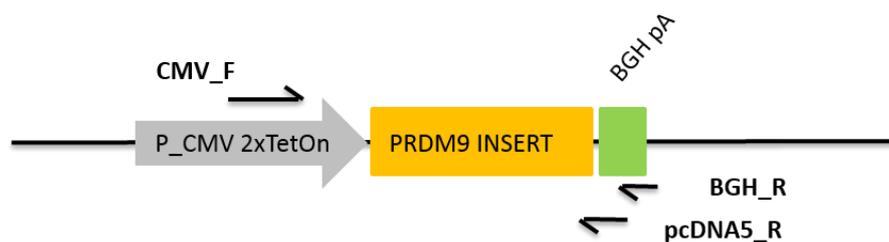
Figure 8.4 Schematic showing alignment of the primer sets used to amplify the pcDNA5 insert sequence.

Primer details are listed in Table 8.1. Temperature gradients were used to determine the optimum annealing temperature of the primer sets. PCR reaction mix was as follows; 0.5 mM betaine, 200 μM dNTPs, 1.25 U TaKaRa LA Taq polymerase, 0.2 μM primers (forward and reverse), dH$_2$O and 50 ng DNA made to final volume of 25 μL with nanopure water. Thermocycling conditions were: denaturation at 94 °C for 1 min, 98 °C for 10 secs, annealing at 62 °C for 15 min, for 30 cycles followed by a final extension for 10 min at 72 °C.

Target	Sequence 5'-3'	Annealing Temperature
CMV Forward	CGCAAATGGGCGGTAGGCGTG	62 °C
pcDNA5 Reverse	GAGGAAATTGCATCGCATTGT	62 °C
BGH Reverse	CCTCGACTGTGCCTTCTA	62 °C

Table 8.1 Primer details for amplification of insert sequences from pcDNA3.1 and pcDNA5 plasmids.

PRDM9 amplicons were visualised via electrophoresis and UV exposure, showing amplification of a band at 2.7 kb corresponding to the *PRDM9* cDNA construct (Figure 8.5).
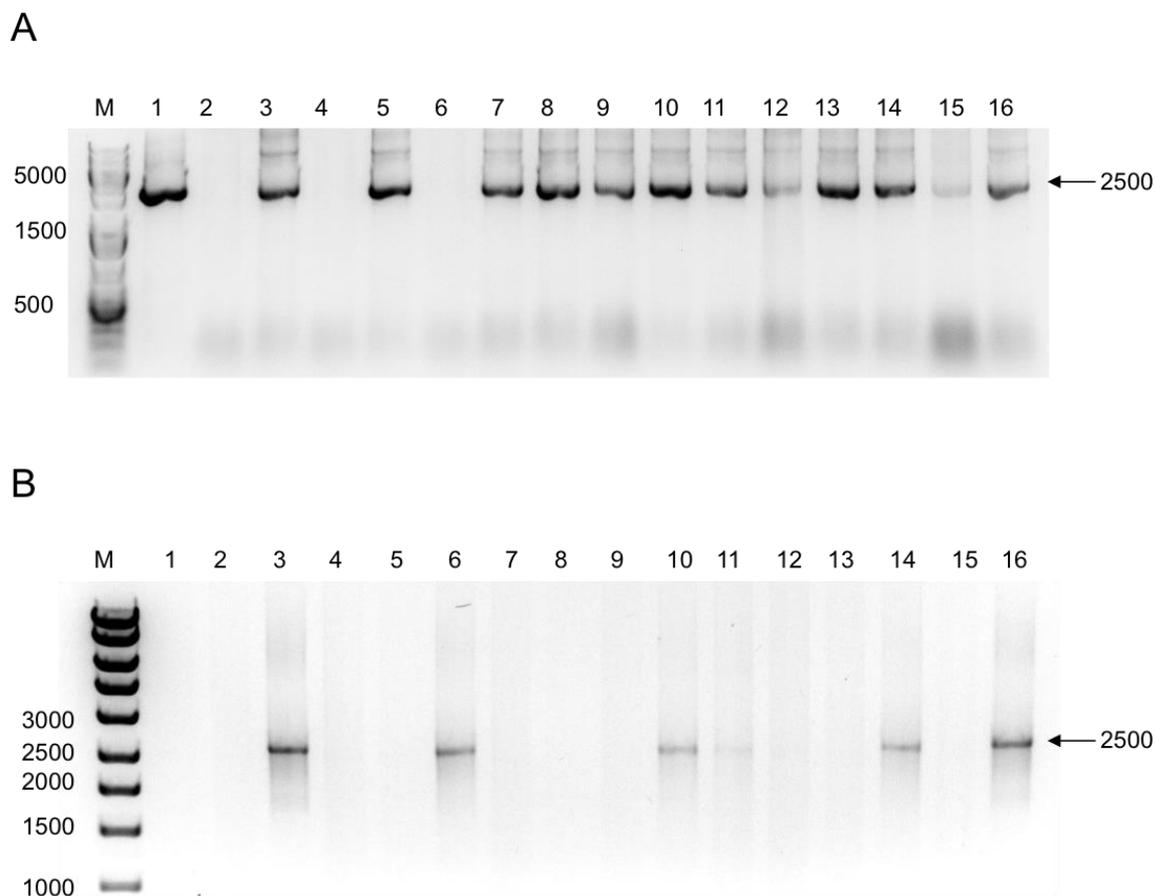
Figure 8.5 Colony PCR using plasmid specific primers. PCR products were electrophoresed on 1% agarose gels and visualised with UV exposure. DNA size marker was included on the gel (M), sizes are given in base pairs. A) Colony PCR of PRDM9 insert ligated into pcDNA5/FRT/TO, 16 colonies were screened. B) Colony PCR of Flag-PRDM9 insert ligated into pcDNA5/FRT/TO, 16 colonies were screened.

In order to successfully integrate the PRDM9 insert sequences from the pcDNA5 vector into the genome of the HEK293 cells, a second plasmid expressing Flp recombinase enzyme must be co-transfected (Figure 8.6). The enzyme catalyses recombination between FRT sites present on the pcDNA5/FRT/TO plasmid and within the TRex293™ genome. As the genomic DNA recombines and is repaired by double strand break repair complexes, the hygromycin, tetracycline operator and PRDM9 elements are incorporated between FRT sites. The cell now harbouring this gene cassette confers

hygromycin resistance and will express PRDM9 in the presence of tetracycline (Figure 8.6).
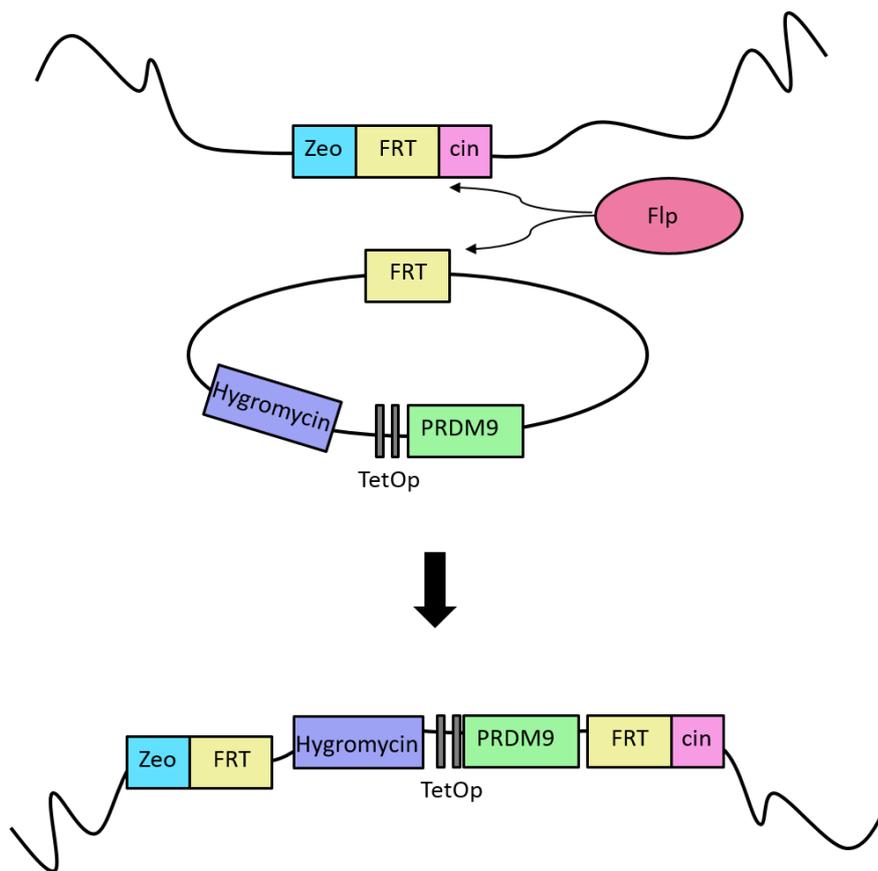
Figure 8.6 Schematic of the Flp mediated recombination event. PRDM9 cDNA on pcDNA5 plasmid is recombined into the transfected cell genome using the integrated FRT sites and Flp recombinase enzyme expressed on the pOG44 plasmid.

The plasmid containing Flp recombinase, pOG44, was a kind gift from Prof R.N. Lightowlers (Institute of Neuroscience, Newcastle University) and was expanded in our lab via transformation of competent *E.coli*. Purified plasmid DNA was digested with XbaI enzyme which cuts the plasmid at one site and KpnI which cuts the plasmid at two sites producing two different sized DNA fragments. After digestion with XbaI, a clear band could be seen at 5785 bp (Figure 8.7). After digestion with KpnI there were two clear bands at 347 bp and 5438 bp confirming that pOG44 had been successfully purified (Figure 8.7). Colony number 1 (Figure 8.7 lane 1 & 9) was taken forward for use in all related experiments.
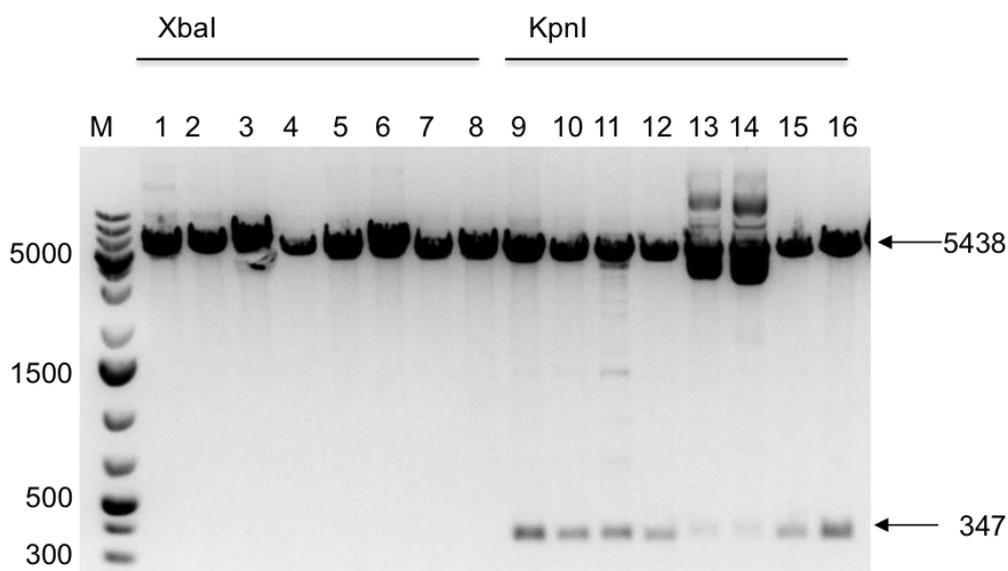
Figure 8.7 Restriction digest of pOG44 plasmid DNA purified from competent *E.coli.* 50 ng DNA was electrophoresed on a 1% agarose gel and visualised by UV exposure. DNA size marker was included on the gel (M), sizes are given in base pairs. Lanes 1-8 show DNA digested with XbaI enzyme. Lanes 9-16 show DNA digested with KpnI enzyme.

8.3.2 Creation of stable overexpression cell lines

The HEK293 cell line was previously modified to contain an FRT recombinase site within the genome (TRex™293 cells, Invitrogen, Thermo Fisher Scientific, Loughborough, UK), herein referred to as HEK293 cells. Both PRDM9 cDNA constructs were integrated into a 'transcriptionally silent' region of the genome through cell transfection as described in Section 3.11.2. Non-transfected cells were used as a control in all experiments. After 14 days of culture in antibiotic selection media, colonies of successfully transfected cells could be seen (Figure 8.8). Colonies were left to grow and colonise the well completely before being transferred into larger culture plates and flasks.
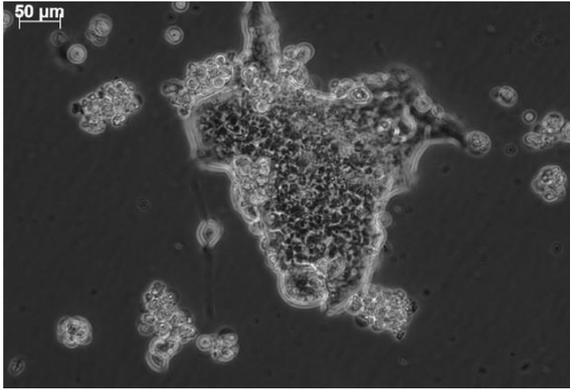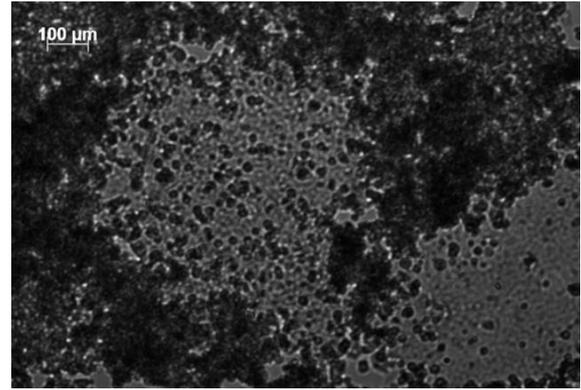
Figure 8.8 Cell colony formation following transfection of TRex™293 cells with pcDNA5_PRDM9 constructs and selection with 200 µg/mL hygromycin and blasticidin. A) Colony formed after integration of the PRDM9 pcDNA5 construct. B) Colony formed after integration of the Flag-PRDM9 pcDNA5 construct.

After 8-10 weeks of antibiotic resistance and establishment of a homogenous population, cell lines were screened for PRDM9 cDNA integration into the genome by PCR amplification as described above. PCR products were electrophoresed using 1% agarose gels and visualised via UV exposure. Amplification of a band at ~2700 bp was achieved when the CMV and pcDNA5 primers were used. This band corresponds to the PRDM9 cDNA construct and is observed in both of the overexpression cell lines but not in control HEK293 cells (Figure 8.9). The origin of the other PCR products was most likely due to non-specificity of the primers used. These primers would ordinarily be used with isolated plasmid DNA and would therefore not have off target binding. Indeed, when these primer sequences were entered into NCBI BLAST they did align to other regions of the human genome. The product at 2.7 kb was then gel extracted and Sanger sequenced and aligned to the PRDM9 mRNA sequence (NM_001310214.1) to confirm that this was indeed the PRDM9 cDNA sequence.
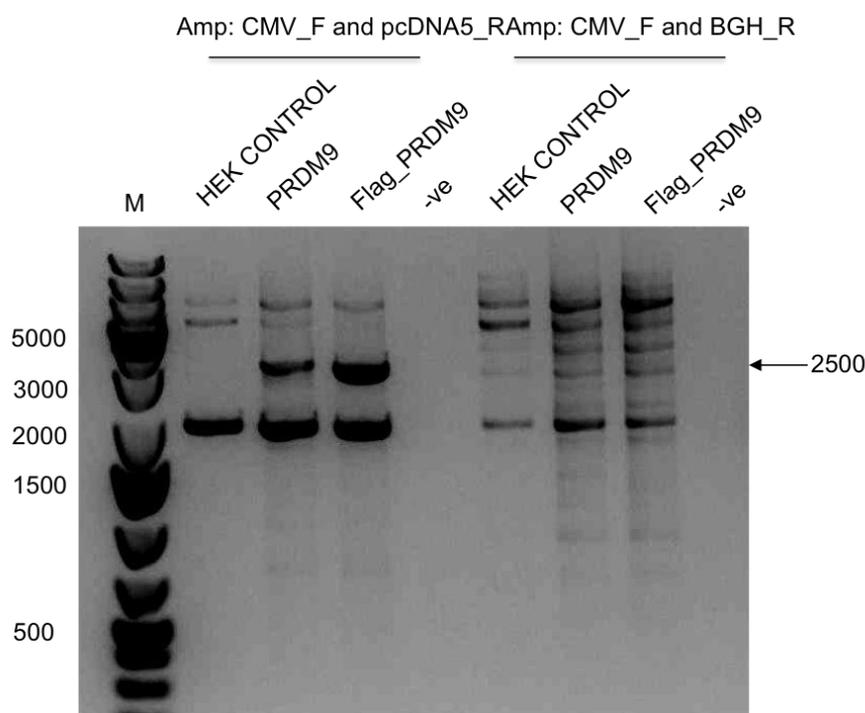
Figure 8.9 PCR amplification of PRDM9 constructs in stably transfected cell lines. PCR products were electrophoresed on a 1% agarose gel. DNA size marker was included on the gel (M), sizes are given in base pairs. Lanes 2-4 show the PCR product from each of the cell lines using primer set 1. Lanes 6-8 show the PCR product from each of the cell lines using primer set 2. Lanes 5 & 9 are negative controls where PCR mix did not contain DNA. Arrow indicates the PRDM9 insert sequence at the predicted size of ~2700 bp.

8.4 Results

8.4.1 Analysis of PRDM9 overexpression by immunoblotting

To determine whether the pcDNA5/FRT/TO stable overexpression model was successful, PRDM9 protein expression was investigated. Cells were seeded at 3 x 10^5 cells per well and when 60% confluency was reached, tetracycline was added to the growth media for 24 hours. Protein from both the PRDM9 construct and the Flag-PRDM9 construct were blotted for using either PRDM9 or Flag antibodies. After 24 hours of tetracycline treatment there was a band observed in the PRDM9 cell line at ~80 kDa in size (Figure 8.10 A). The band was present in cells treated with the highest doses of tetracycline but was not detected at lower treatment concentrations or in the untreated control (Figure 8.10 A). In the cell lines transfected with Flag-PRDM9 there was an observed band at ~100 kDa in size (Figure 8.10 B). Similarly, this band was detected in cells treated with the highest concentration of tetracycline but was not present at lower treatment concentrations or in the untreated control (Figure 8.10 B).
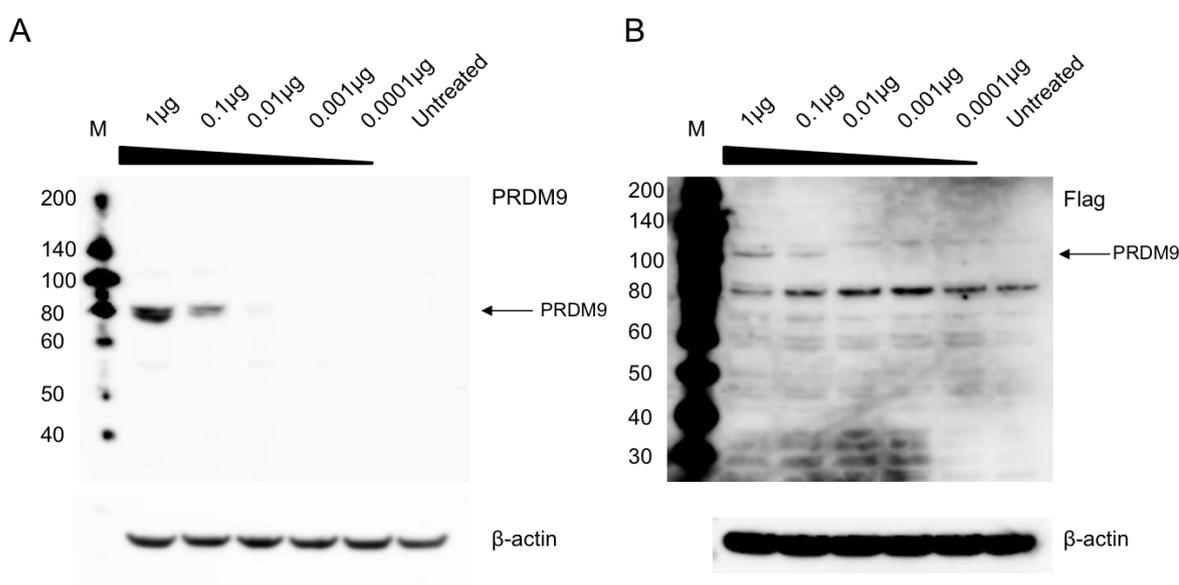
Figure 8.10 Immunodetection of PRDM9 protein in overexpression cell lines. Stably transfected cells were treated with tetracycline for 24 hours at decreasing concentrations. Blank refers to untreated cells. A) Protein extracted from PRDM9 stably transfected cells. PVDF membrane was incubated with anti-PRDM9 (1:500) overnight at 4°C. B) Protein extracted from Flag-tagged PRDM9 transfected cells. PVDF membrane was incubated with anti-Flag (1:1000) overnight at 4°C. Membranes were incubated with anti-βactin (1:1000) overnight at 4°C.

The protein bands detected in the two cell lines differed in molecular weight by ~20 kDa (Figure 8.10). Only the Flag-PRDM9 cell line showed a band at the expected size of ~100 kDa whereas the PRDM9 cell line showed a band at ~80 kDa. In Figure 8.10 B the band at ~100 kDa was faint even after the membrane was overexposed suggesting that there was a low level of detectable protein in the cells. The presence of multiple bands also points to the abundance of PRDM9 protein being low as Flag should only be detectable in cells transfected with the construct and treated with tetracycline. The presence of these bands in the untreated sample indicates that these bands are non-specific.

PRDM9 detection was achieved once although multiple passages of these cell lines were investigated by immunoblotting. This suggested that the earliest passages of the cells, after the selection process, were more likely to have PRDM9 expression induced by tetracycline treatment. To determine whether this detection issue was due to a heterogeneous cell population, transfected cells were re-selected using single cell dilution followed by hygromycin resistant clone selection. Colonies derived from a single cell containing the stably integrated constructs were assessed for PRDM9 expression by immunoblot analysis of protein lysate (Figure 8.11). Five single cell colonies were expanded from the PRDM9 containing cell line and six from the Flag-PRDM9 line. Each colony was split into two wells; one treated with 1 µg/mL tetracycline whilst the other was grown in standard culture medium as an internal control. When the samples were immunoblotted with anti-Flag there was a band detected at ~100 kDa in five of the colonies from the Flag-PRDM9 cell line but no detection in the untreated controls, the PRDM9 cell line or parental HEK293 cells (Figure 8.11). There was also a band detected at ~60 kDa which appeared in all sample lanes suggesting it is a non-specific peptide bound by anti-Flag. The detected peptide at ~100 kDa in size was very weak compared to the detected band at ~60 kDa and the loading control β-actin suggesting that if this peptide is PRDM9, the expression level is very low in this cell line. This immunoblot was also incubated with commercially available anti-PRDM9 but no peptide was detected suggesting that the antibody was not appropriate for detection in these transgenic cell lines.
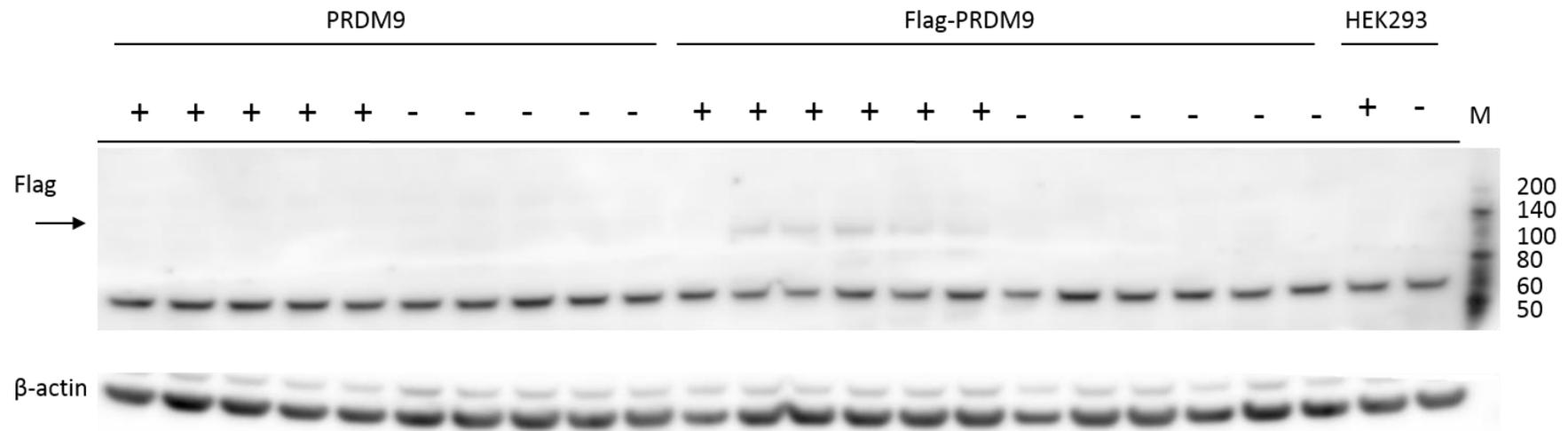
Figure 8.11 Immunodetection of PRDM9 in cells derived from hygromycin resistant single colonies. Five colonies were selected for PRDM9 construct integration, six colonies were selected for Flag-PRDM9 construct integration and compared to HEK293 parental cell line. Colonies were treated with 1 μg/mL tetracycline (+) for 24 hours or untreated (-). Protein size ladder was included on the gel (M), sizes are denoted in kDa.

Clone number 3 from each of the overexpression cell lines was then treated with 1 μg/mL tetracycline and the mitochondria extracted using a digitonin-based protocol described in Section 3.6.4. This mitochondrial extract (Mt) was then compared to whole cell protein extract (WhC) and immunodetection was carried out using anti-Flag (Figure 8.12). There was not a clear single peptide detected using anti-Flag and any faint bands detected were also present in the HEK293 control cell line.
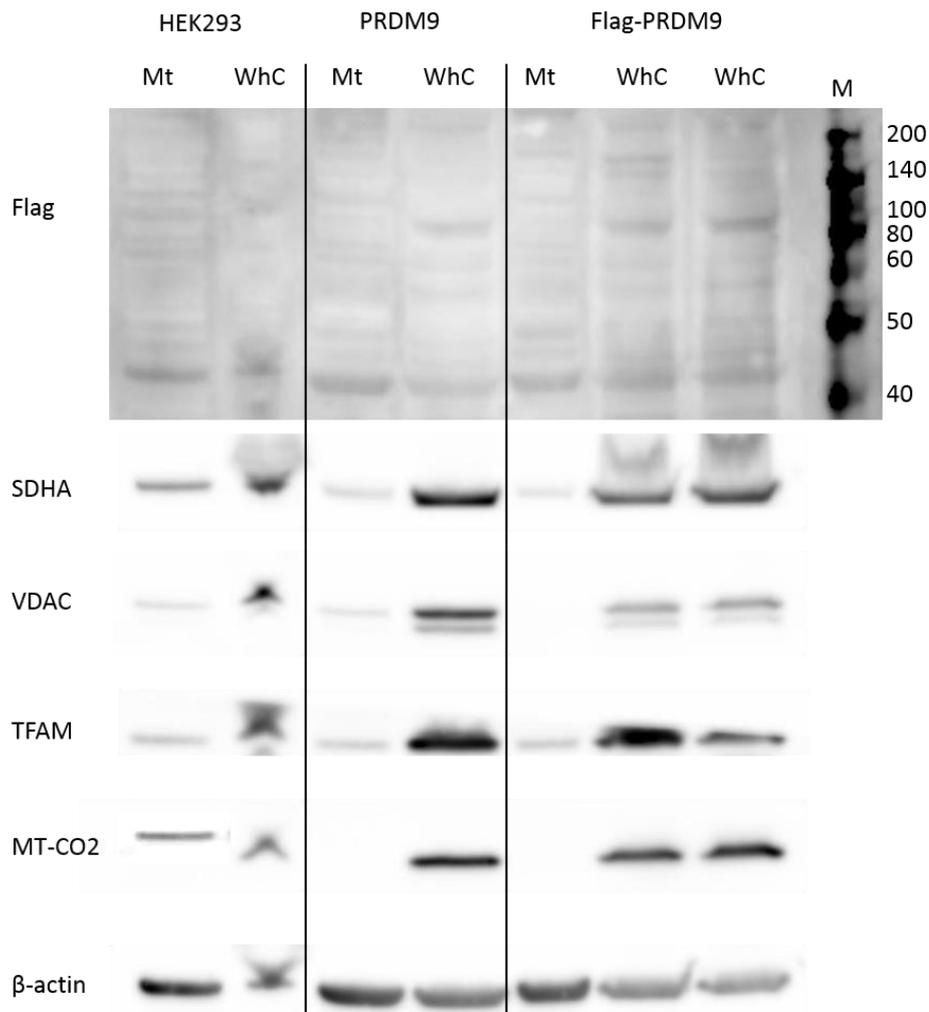
Figure 8.12 Immunodetection of Flag-PRDM9 in mitochondrial extracts from overexpression cell lines. 75 μg protein from mitochondrial (Mt) or whole cell (WhC) extracts was loaded on 4-20% bis-tris gels. Protein size marker was included on the gel (M), sizes are given in kDa. Membrane was incubated with anti-Flag (1:1000), anti-SDHA (1:2000), anti-VDAC (1:500), anti-TFAM (1:500), anti-mtCO2 (1:1000) or anti-βactin (1:1000) overnight at 4 °C.

To assess the purity of this mitochondrial fraction several antibodies against mitochondrial proteins were used (Figure 8.12). Detection of VDAC1, transcription

factor A mitochondrial (TFAM) and succinate dehydrogenase complex flavoprotein subunit A (SDHA) proteins showed that the mitochondrial extracts did contain mitochondria but at much lower levels than the whole cell protein extracts. There was also a high amount of β-actin in the mitochondrial extracts suggesting that the extraction was not pure and also contained contaminating cytosolic proteins.

8.4.2 Analysis of PRDM9 expression by RT q-PCR

Protein detection was inconclusive in the overexpression cell lines as only three blots showed possible Flag-PRDM9 peptide detection at 100 kDa. To assess whether PRDM9 transcript was being produced in the overexpression cell lines created, PRDM9 mRNA levels were measured by RT q-PCR amplification using PRDM9 specific probes. PRDM9 transcript detection was normalised to the house-keeping gene β-actin as described in Section 3.12.2. Two PRDM9 assays were tried however, one probe/primer set (Hs00360639_m1) did not detect PRDM9 amplification on several real time assay runs and was therefore not used again. The PRDM9 probe/primer set Hs01633270_s1 gave PRDM9 amplification and was used on all samples measured in triplicate on each plate, and each plate was run in triplicate to control inter-plate and intra-run assay variability.

Cycle threshold (Ct) values were very high for the PRDM9 amplicon measured in all cell lines tested, indicating low level expression. When normalised to β-actin transcript levels, the amount of PRDM9 transcript measured in the HEK293 control cell line was higher than in either of the PRDM9 overexpression cell lines (Figure 8.13). Differences in transcript expression could be seen between cell lines but also between tetracycline treated or untreated cells over the time course of 24-72 hours. Unexpectedly, in several of the sample groups tested there appeared to be higher PRDM9 transcript expression in the untreated cell lines compared to tetracycline treated, contrary to the expected result (Figure 8.13).
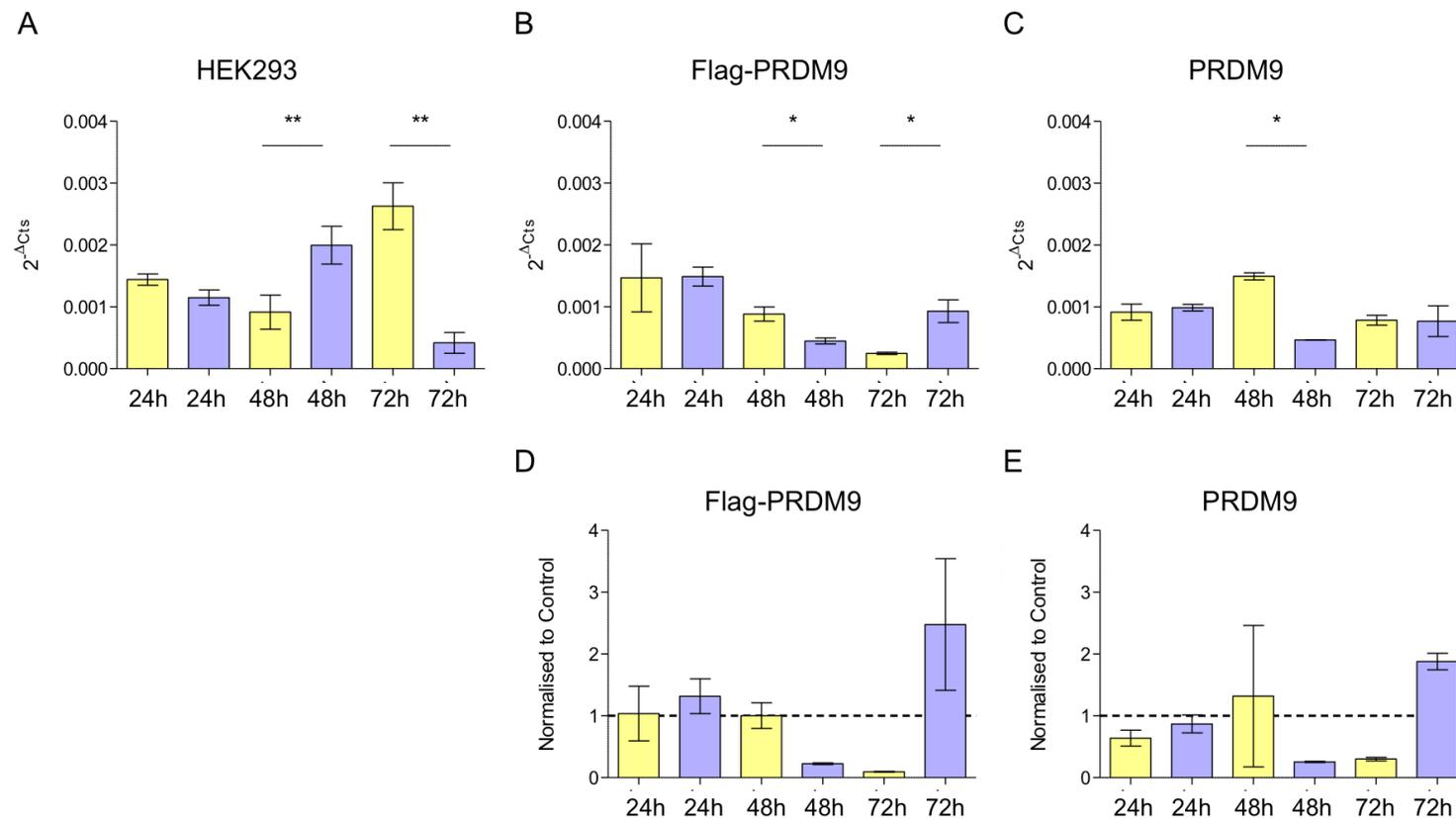
Figure 8.13 Quantification of PRDM9 transcript levels in overexpression cell lines and HEK293 controls. A, B & C) Cells were untreated (yellow) or treated (lilac) with 1 µg/mL tetracycline for 24, 48 or 72 hours. $2^{-\Delta Ct}$ values were normalised to β-actin transcript levels and plotted using Graphpad Prism analysis software. D & E) PRDM9 transcript levels ($2^{-\Delta Cts}$) in overexpression cell lines were normalised to control HEK293 transcript values, shown by the dotted line at y = 1. Students t-test was used to assess significance of each treatment group vs the un-treated partnered group. ≥0.05 = *: ≥0.005 = **: ≥0.0005 = ***. n=27

Although there were statistically significant differences between treated and untreated samples in all three cell lines (Figure 8.13 A, B and C) there was no correlation or trend in the transcript levels with respect to increased or decreased transcript levels before or after treatment. In addition, normalising $2^{-\Delta Ct}$ values of the transgenic lines to the HEK293 control cell line showed that the PRDM9 transcript level was less than or equivalent to the control at majority of time points tested (Figure 8.13 D and E).

If the transgenic cell lines created in this study were actively transcribing PRDM9 after addition of tetracycline then it would be expected that the PRDM9 mRNA transcript levels would be higher in these cell lines compared to control HEK293 cells. This was not the case, as measured in three independent experiments, suggesting that PRDM9 transcript is not overexpressed in these transgenic cell lines.

8.4.3 Immunofluorescent staining of PRDM9 in overexpression cell lines

Detection of PRDM9 by protein and mRNA detection experiments was inconclusive, therefore, immunofluorescent staining using the Flag antibody was attempted in fixed cells as described in Section 3.9.4. If PRDM9 is being expressed at low levels in a proportion of the cultured cells then single cell imaging analysis may show detectable protein expression. As a positive control, cells were stained using the mitochondrial protein OPA1 as optimised previously on HEK293 cells (Chapter 7).

Cells overexpressing Flag-PRDM9 were treated with 1 µg/mL tetracycline for 72 hours before staining with anti-Flag (red) and anti-OPA1 (green) and compared with untreated controls (Figure 8.14). All slides imaged showed positive OPA1 staining of the mitochondria in the Flag-PRDM9 cells (Figure 8.14, green signal). Positive staining was also seen with the Flag antibody (Figure 8.14, red signal) however, this signal was also seen in untreated control samples. The single cells shown (Figure 8.14) are representative of the variable pattern of anti-Flag staining observed, suggesting that this is not a specific signal. Furthermore, the presence of this staining in the untreated cells suggests that this is not Flag-PRDM9 protein, as it would not be expressed in the absence of tetracycline (Figure 8.14 B and D).
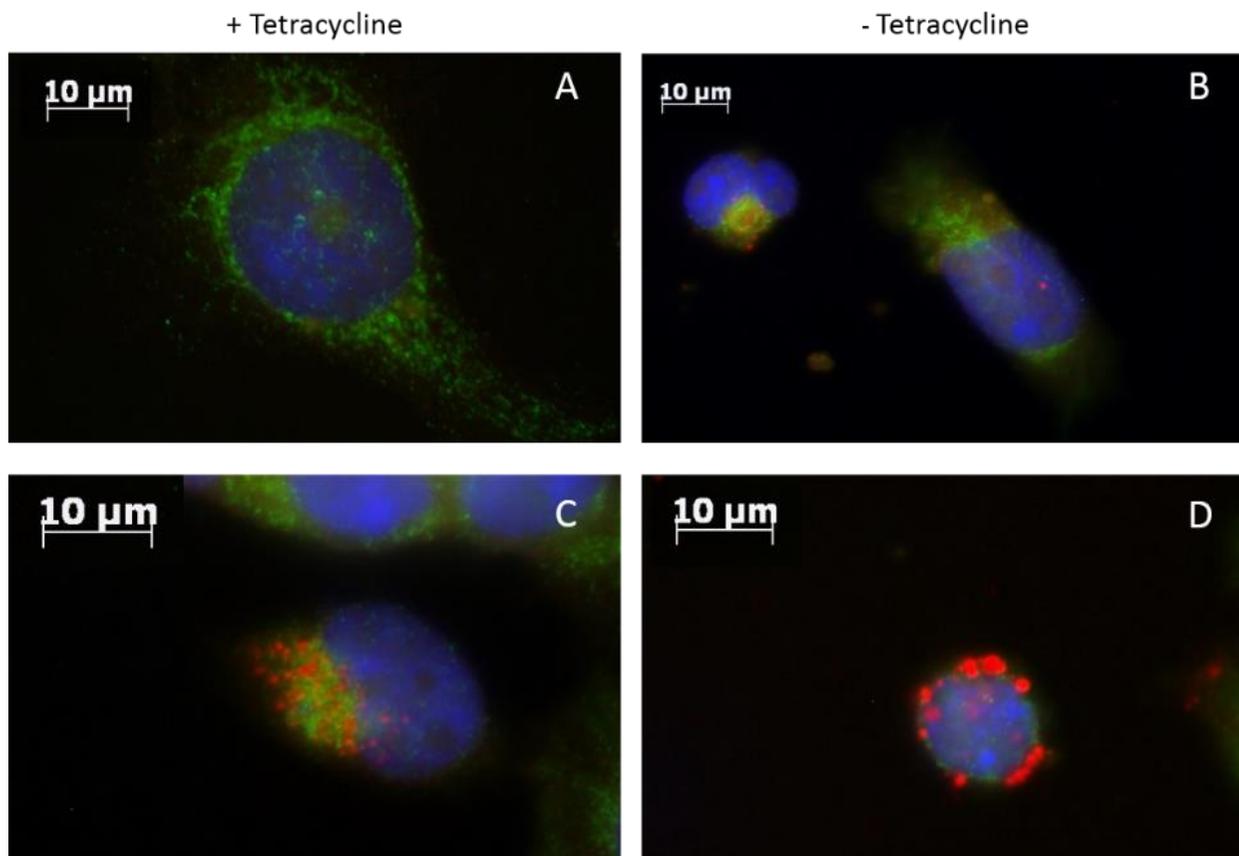
Figure 8.14 Immunofluorescent staining of Flag-PRDM9 cells. Cells were treated with 1 µg/mL tetracycline for 72 hours (A, C&E) or were untreated (B, D&F). Staining with anti-Flag (1:1000) and anti-OPA1 (1:250) overnight at 4 °C, anti-mouse alexafluor®-594 secondary (red), anti-rabbit alexafluor®-488 (green) and DAPI nuclear stain (blue). All slides were imaged on the AxioImager (Zeiss) using GFP and DAPI filter sets.

Next, the PRDM9 commercially available antibody (Abcam, Cambridge, UK) was used for staining in PRDM9 overexpressing cells. This staining procedure was carried out on untreated cells or cells treated with 1 µg/mL tetracycline for 72 hours as an internal control. In addition, the PRDM9 antibody was raised in rabbit and could not be used alongside the OPA1 antibody described previously. In this case, an antibody against TFAM, raised in mouse, was used as a positive staining control. Positive staining of TFAM was seen in both treated and untreated cell samples (Figure 8.15, red signal). Positive staining of PRDM9 (green) was observed in some of the single cells imaged (Figure 8.15 A). However, there was also green fluorescent signal observed in some of the cells not treated with tetracycline (Figure 8.15 B) suggesting that this staining was not specific.
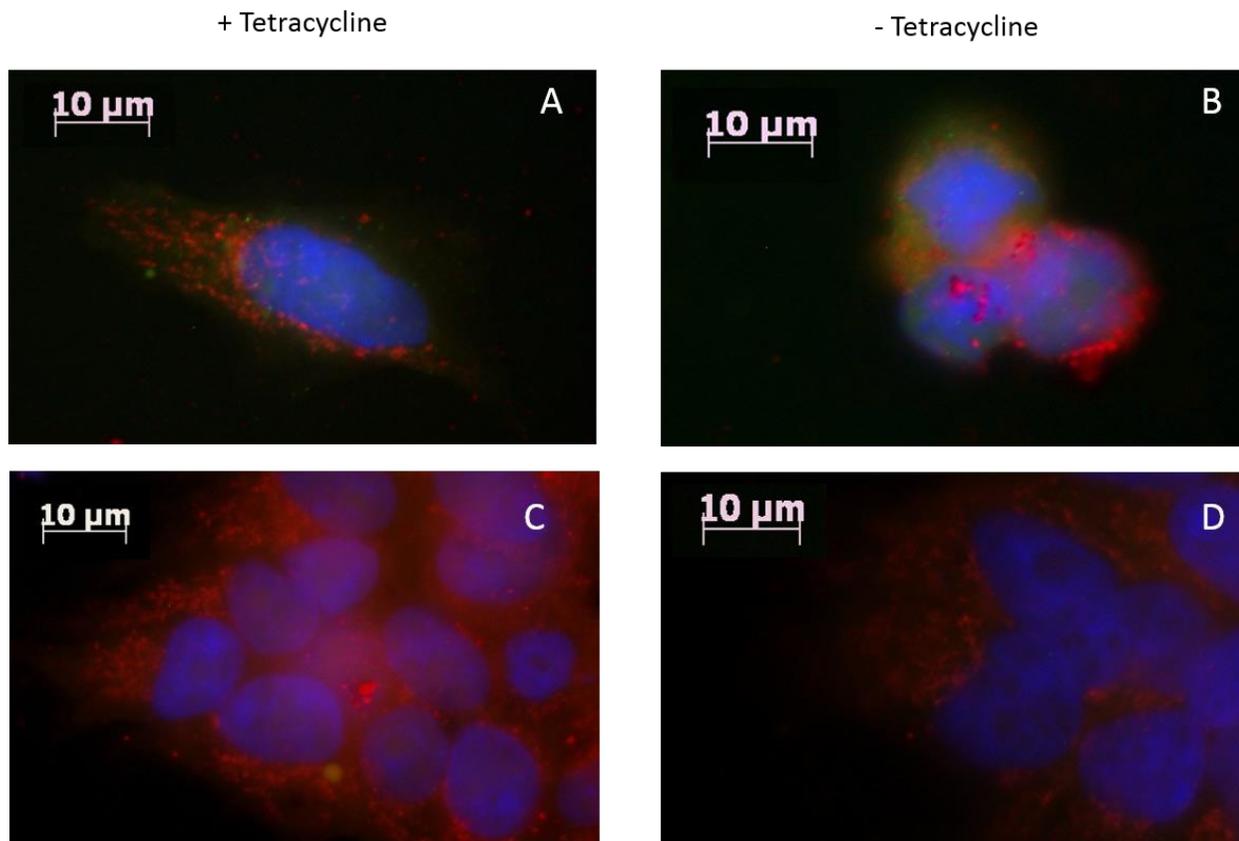
Figure 8.15 Immunofluorescent staining of PRDM9 cells. Cells were treated with 1 µg/mL tetracycline for 72 hours (A&C) or were untreated (B&D). A & B) Staining with anti-PRDM9 (1:500) and anti-TFAM (1:250) overnight at 4 °C, anti-mouse alexafluor®-594 secondary (red), anti-rabbit alexafluor®-488 (green) and DAPI nuclear stain (blue). C & D) Staining with anti-TFAM (1:250) overnight at 4 °C, anti-mouse alexafluor®-594 secondary (red) and DAPI nuclear stain (blue). All slides were imaged on the AxioImager (Zeiss) using GFP and DAPI filter sets.

8.4.4 Indirect measurement of PRDM9 activity by assessing histone methylation levels

Detection of PRDM9 mRNA and protein levels was variable in the cell lines created in this study. PRDM9 has been shown previously to mono-, di- and tri-methylate H3K4 in the nucleus (Hayashi *et al.*, 2005; Grey *et al.*, 2011; Eram *et al.*, 2014). To determine whether the transient PRDM9 protein was having any effect in the stable cell lines, H3K4 methylation status was investigated by immunoblotting (Figure 8.16). Mono-, di- and tri-methylation status was visibly increased in the PRDM9 overexpression cell lines compared to the HEK293 non-transfected control cell line.
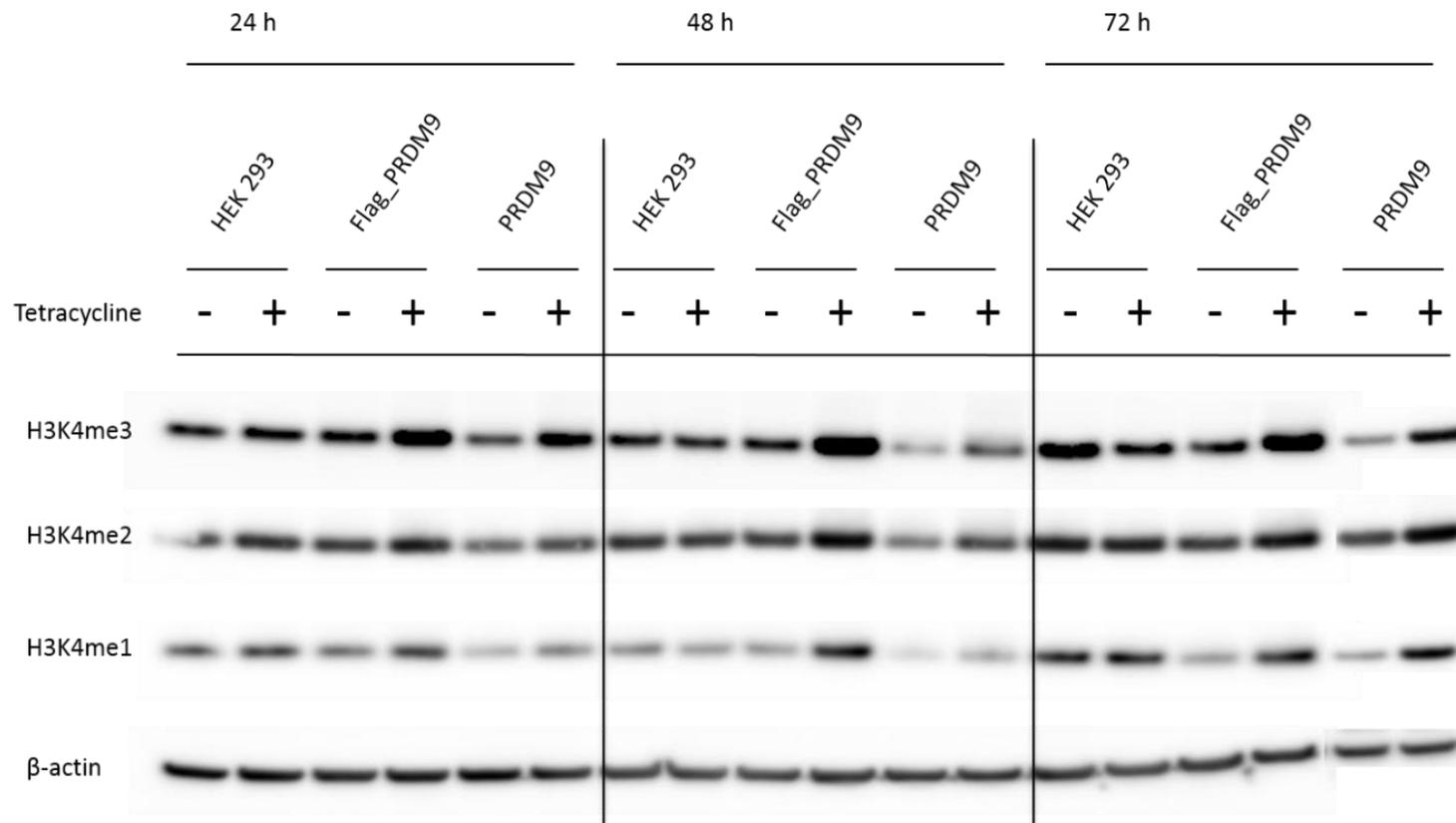
Figure 8.16 Immunodetection of histone methylation status in cell lines. PRDM9 and Flag-PRDM9 overexpression cell lines were compared to the HEK293 parental cell line. Cells were treated with 1 µg/mL tetracycline (+) or untreated (-) for 24, 48 or 72 hours. Membranes were incubated with anti-H3K4me3 (1:1000), anti-H3K4me2 (1:1000), anti-H3K4me1 (1:1000) or anti-βactin (1:1000) overnight at 4 °C.

Quantification of protein levels showed that tri-methylation of H3K4 (H3K4me3) was significantly increased in the PRDM9 transgenic cell lines at all treatment time points compared to the control cell line (Figure 8.17).
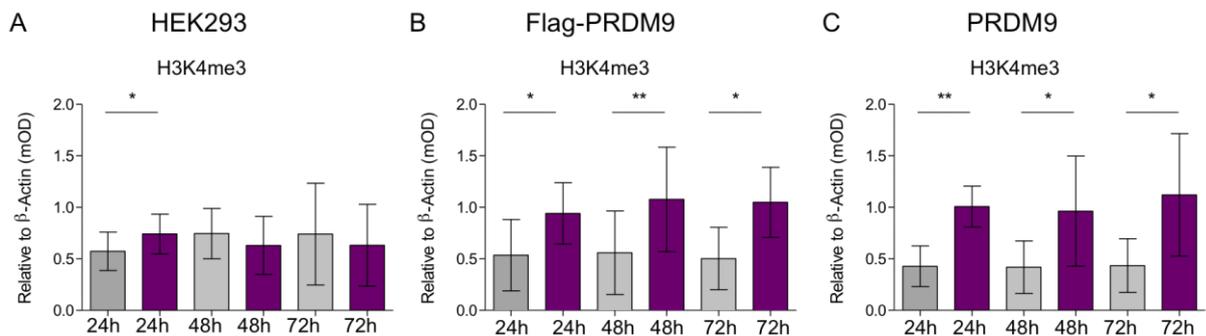
Figure 8.17 Quantification of Western blot analysis for H3K4 methylation status. Untreated samples are shown in grey, tetracycline treated samples in purple. Densitometric analysis was performed using ImageJ software. Values were normalised to β-actin levels and analysed using Graphpad analysis software. Error bars show SD. Students t-test was used to assess significance of each treatment group vs the un-treated partnered group. ≥0.05 = *: ≥0.005 = **: ≥0.0005 = ***. n=27

Di-methylation of H3K4 (H3K4me2) was also significantly increased at all treatment time points in the Flag-PRDM9 but not the PRDM9 overexpression line (Figure 8.18). Although not statistically significant, there was a trend towards increase of H3K4me2 in the PRDM9 transgenic line after tetracycline treatment (Figure 8.18 C).
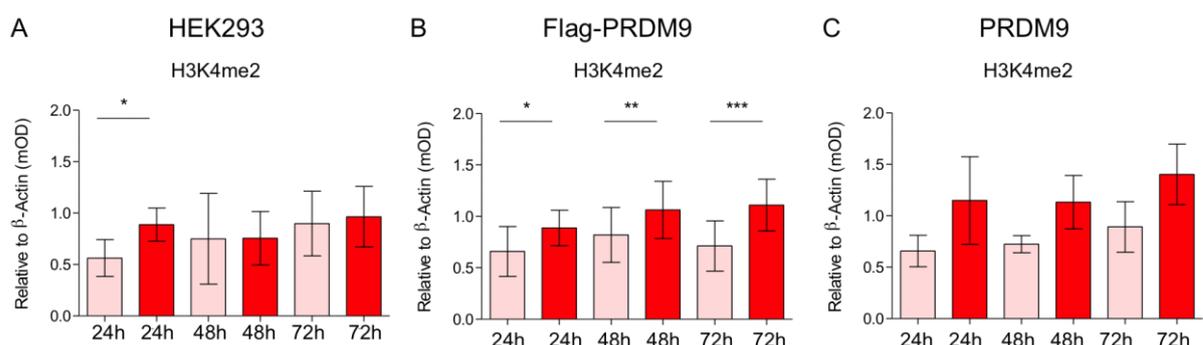
Figure 8.18 Quantification of Western blot analysis for H3K4 methylation status. Untreated samples are shown in pink, tetracycline treated samples are shown in red. Densitometric analysis was performed using ImageJ software. Values were normalised to β-actin levels and analysed using Graphpad analysis software. Error bars show SD. Students t-test was used to assess significance of each treatment group vs the un-treated partnered group. ≥0.05 = *: ≥0.005 = **: ≥0.0005 = ***. n=27

Finally, mono-methylation of H3K4 (H3K4me1) was significantly increased at 72 hours of tetracycline treatment in the Flag-PRDM9 but not the PRDM9 overexpression line. Although not significant, it does appear that levels of H3K4me1 are trending towards an increase in the PRDM9 overexpression cell line after 24 hours of tetracycline treatment (Figure 8.19 C).
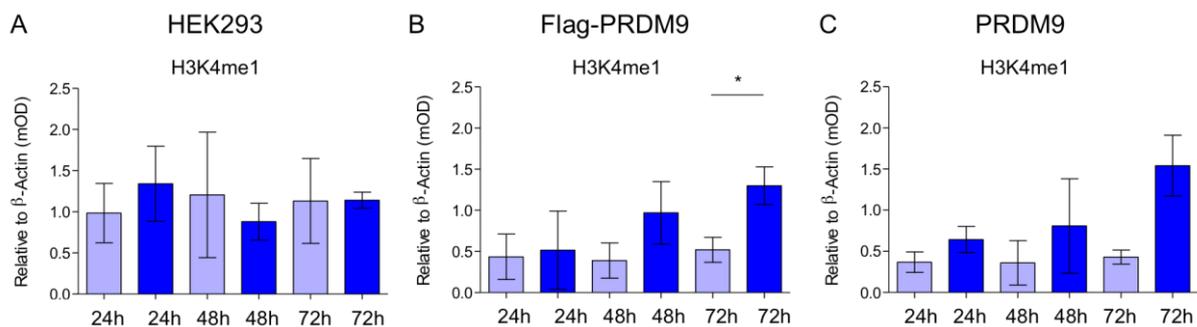
Figure 8.19 Quantification of Western blot analysis for H3K4 methylation status. Untreated samples are shown in pale blue, tetracycline treated samples are shown in dark blue. Densitometric analysis was performed using ImageJ software. Values were normalised to β-actin levels and analysed using Graphpad analysis software. Error bars show SD. Students t-test was used to assess significance of each treatment group vs the un-treated partnered group. ≥0.05 = *: ≥0.005 = **: ≥0.0005 = ***. n=27

The control cell line (HEK293) also had significantly increased levels of tri- and di-methylation of H3K4 at 24 hours of tetracycline treatment (Figure 8.17 and Figure 8.18). This could be due to the tetracycline treatment itself which might be having off target effects on the cell leading to activation of other factors controlling histone methylation. If this is the case, then PRDM9 overexpression might only be influencing the methylation status of H3K4 after 48 hours of treatment.

8.4.5 Measuring mitochondrial protein expression in overexpression cell lines

Indirect measurement of PRDM9 function was achieved by the observed increase in histone methylation status in the overexpression cell lines (Section 8.4.4). Next, mitochondrial protein levels were investigated to assess whether PRDM9 expression was affecting either nuclear encoded or mtDNA encoded mitochondrial protein levels.

TFAM plays an essential role in mtDNA transcription as well as replication and repair and is also thought to be the main protein moiety required for efficient packaging of the

plasmid mtDNA into nucleoid structures (Kukat *et al.*, 2015). In this case, TFAM protein levels cannot be used to directly quantify mtDNA copy number inside the cell but can be used to estimate whether cells are depleted of mtDNA or have increased mtDNA (Collu-Marchese *et al.*, 2015; Ikeda *et al.*, 2015; West *et al.*, 2015; Stiles *et al.*, 2016). Protein levels of TFAM were measured in the PRDM9 overexpression cell lines and the HEK293 parental cell line as a control (Figure 8.20).
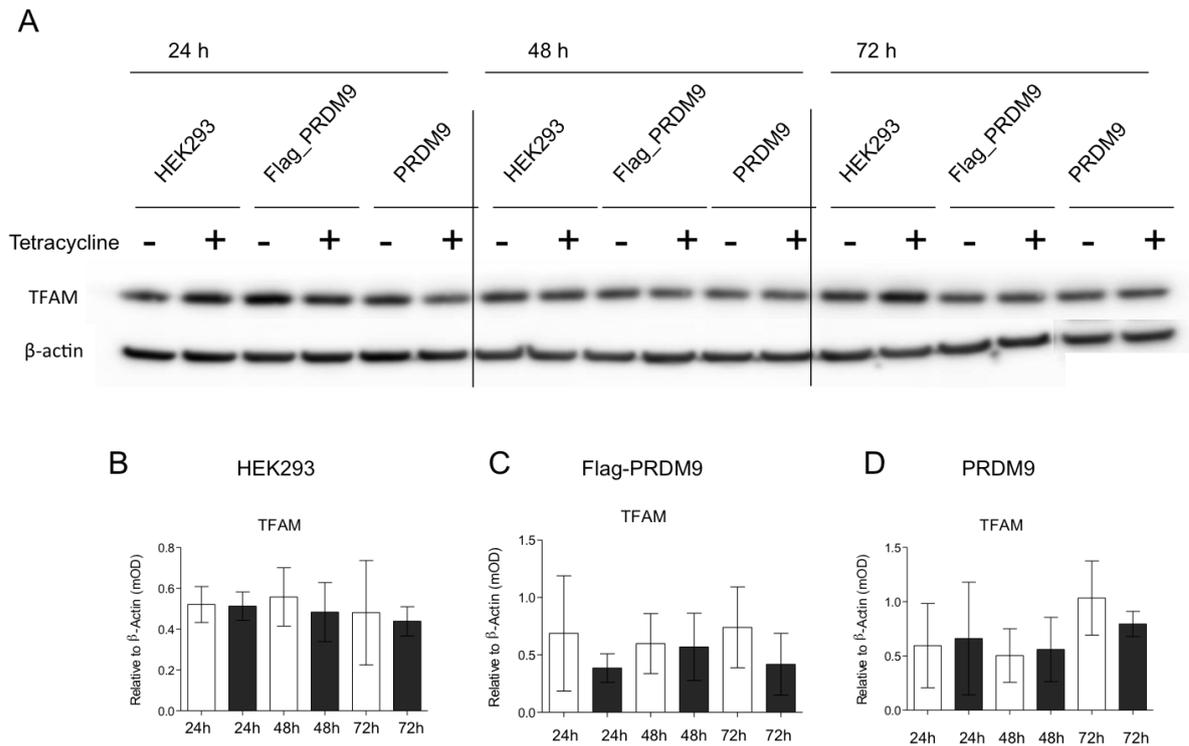
Figure 8.20 Immunodetection of TFAM protein levels in cell lines. A) PRDM9 and Flag-PRDM9 overexpression cell lines were compared to the HEK293 parental cell line. Cells were treated with 1 µg/mL tetracycline (+) or untreated (-) for 24, 48 or 72 hours. Membranes were incubated with anti-TFAM (1:500) or anti-βactin (1:1000) overnight at 4 °C. B-D) Quantification of immunodetection. Untreated samples are shown in white, tetracycline treated samples are shown in dark grey. Densitometric analysis was performed using ImageJ software. Values were normalised to β-actin levels and analysed using Graphpad analysis software. Error bars show SD. Students t-test was used to assess significance of each treatment group vs the un-treated partnered group. ≥0.05 = *: ≥0.005 = **: ≥0.0005 = ***. n=27.

Quantification of TFAM immunoblotting showed no statistically significant difference in TFAM protein expression when tetracycline treated cells were compared to untreated cells (Figure 8.20 B-D).

Next, nuclear encoded OXPHOS subunit proteins were investigated by immunodetection (Figure 8.21). There was no statistically significant differences in protein levels between tetracycline treated or untreated cells with respect to mitochondrial ATP synthase subunit alpha (CV-α), complex III subunit core 2 (CIII-core2) or complex II subunit 30 kDa (CII-30) when normalised to β-actin. There was also no difference in protein levels over the treatment time period tested (24-72 h).
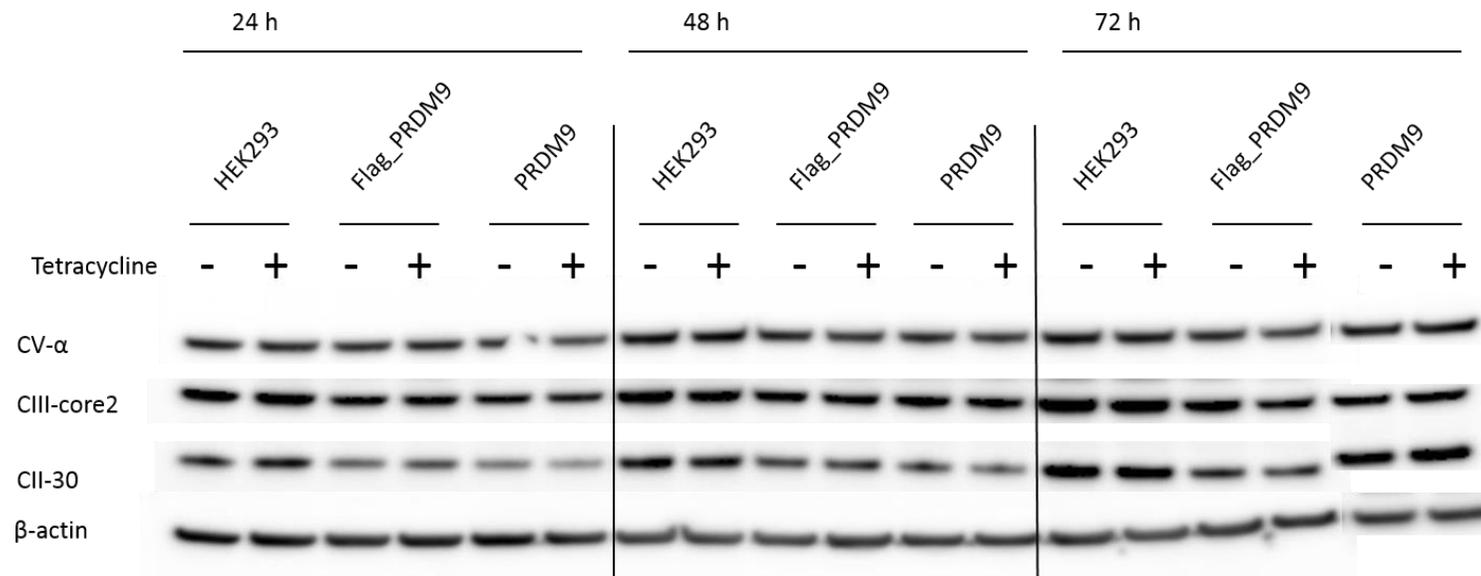
Figure 8.21 Immunodetection of mitochondrial oxidative phosphorylation subunit levels in cell lines. PRDM9 and Flag-PRDM9 overexpression cell lines were compared to the HEK293 parental cell line. Cells were treated with 1 µg/mL tetracycline (+) or untreated (-) for 24, 48 or 72 hours. Membranes were incubated with anti-OXPHOS (1:250) or anti-βactin (1:1000) overnight at 4 °C.

8.4.6 Assessing mtDNA in overexpression cell lines

As PRDM9 motif binding sites were found in mtDNA (Chapter 4), we hypothesised that overexpression of the protein might have an impact on mtDNA copy number levels. Quantification of mtDNA levels was measured by multiplex probe based assay amplification of *MT-ND1*, *MT-ND4* and the nuclear encoded house-keeping gene *RNASEP*. An mtDNA deletion control DNA sample was included on each run as a comparison. This control was previously extracted from a cell line known to harbour a large scale deletion between positions m.7982- m.15504 at a heteroplasmy level of 70% (Diaz *et al.*, 2002).

mtDNA copy number was significantly increased in the treated Flag-PRDM9 cells compared to untreated controls at all time points measured (Figure 8.22 B). This was also the case for the PRDM9 cell line but only at 48 hours of treatment. Copy number in the HEK293 control cell line was varied, with a significant increase at 24 hours tetracycline treatment followed by a decrease at 48 hours and then stabilisation by 72 hours (Figure 8.22 A).
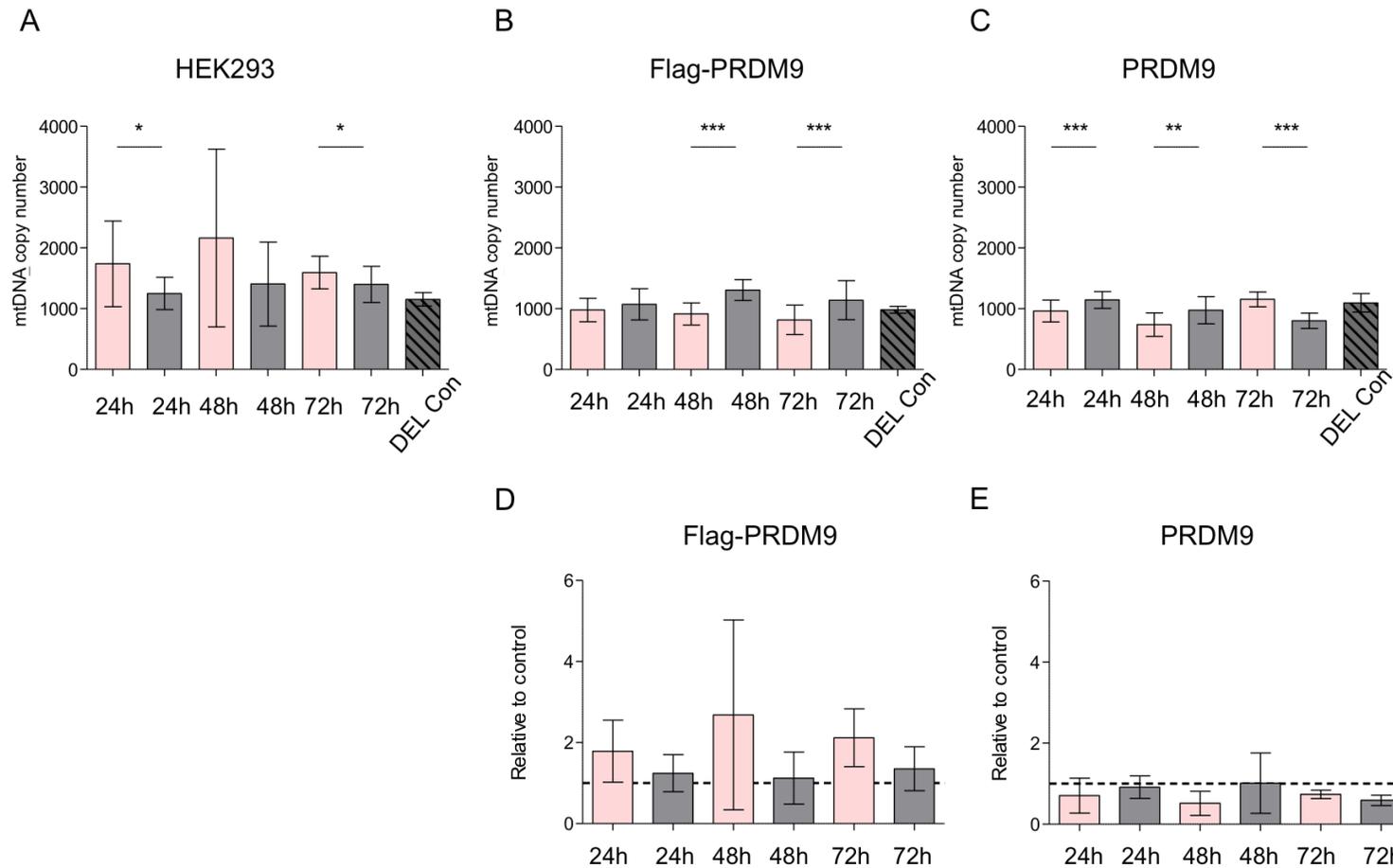
Figure 8.22 Quantification of mtDNA copy number in overexpression cell lines. A, B & C) Cells were untreated (pink) or treated (grey) with 1 μg/mL tetracycline for 24, 48 or 72 hours. $2^{-\Delta Ct}$ values were normalised to RNase P and plotted using Graphpad Prism analysis software. D & E) $2^{-\Delta Ct}$s in overexpression cell lines were normalised to control HEK293 values, shown by the dotted line at y = 1. Students t-test was used to assess significance of each treatment group vs the un-treated partnered group. ≥0.05 = *: ≥0.005 = **: ≥0.0005 = ***. n=27

As described previously (Chapters 4 & 5) we hypothesised that a functional role for PRDM9 in the mitochondria could be to mediate deletion formation, a common occurrence during the normal ageing process and in several mitochondrial disorders. Whole mitochondrial DNA was amplified using a long range PCR protocol in the cell lines overexpressing PRDM9 to test for any possible mtDNA deletion formation. Reaction mixture was as follows; 2 mM Mg^{2+} buffer, 200 µM dNTPs, 0.25 µM primer (forward and reverse) and 1.25 U GXL PrimeSTAR DNA polymerase made to a final volume of 25 µL with dH$_2$O. The reactions were carried out in a thermocycler using the following program; denaturation for 1 minute at 94 °C, followed by 30 cycles of denaturation at 98 °C for 10 minutes, annealing and extension at 68 °C for 15 minutes and final extension at 72 °C for 10 minutes. Details of mtDNA specific primers used in this PCR are detailed in Table 8.2.

Target	Primer 5'-3'	Annealing Temp
mtDNA 16F	CTCAAAGGACCTGGCGGTGCTTC	68 °C
mtDNA 16R	GTAGTGTTCTGGCGAGCAGTTTTG	68 °C

Table 8.2 Details of mtDNA long range PCR primers used in this study.

No mtDNA deletion was observed in any of the samples tested in three independent experiments (Figure 8.23). The deletion control DNA described above (~70% deletion heteroplasmy) was included in each experiment. This control shows two bands when electrophoresed, one at ~16.5 Kb and another at ~10 Kb confirming the presence of a large scale mtDNA deletion using this assay (Figure 8.23).
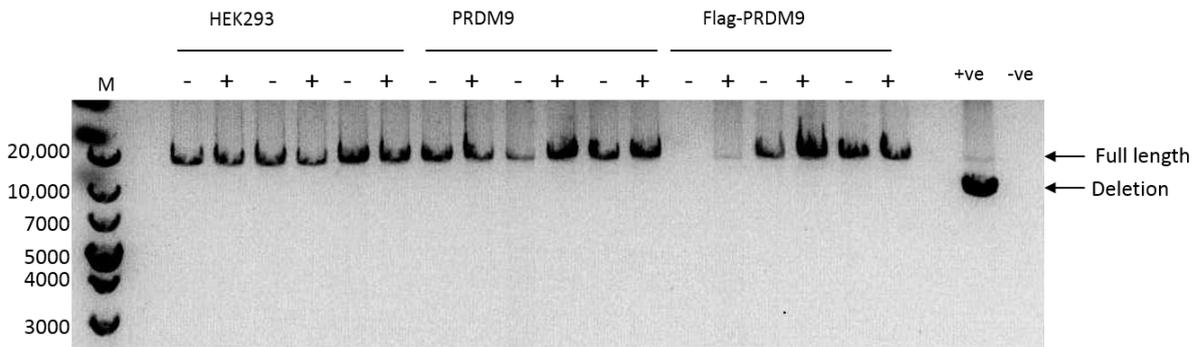
Figure 8.23 Assessment of deletions in mitochondrial DNA. Long range PCR products were electrophoresed on 1% agarose gels. HEK293 cells, PRDM9 overexpression and Flag-PRDM9 overexpression cell lines were compared against a cell line known to harbour a large mtDNA deletion at ~60% heteroplasmy. A DNA size marker was included on the gel (M), sizes are denoted in base pairs.

Measurement of mtDNA deletion by PCR can be subject to amplification bias of the larger more dominant DNA molecule, in this case the full-length mtDNA. To assess whether there were low level mtDNA deletions in the transgenic cell lines, RT-qPCR data (described in Figure 8.22) was analysed for differences in mtDNA encoded genes *MT-ND1* and *MT-ND4*. No difference was seen in the copy number ratio of the two genes suggesting no deletion of the *MT-ND4* region of the molecule.

To test whether there were deletions occurring elsewhere on the molecule, copies of *MT-ND1* and *MT-COIII* were directly compared. There was no statistically significant difference (Figure 8.24). A value of 1 corresponds to there being no difference in the number of copies of the two genes, as seen in all cell lines (Figure 8.24). The mtDNA deletion control described in Section 8.4.5 showed an increased copy number of *MT-ND1* compared to *MT-COIII*, as expected since the region of mtDNA containing *MT-COIII* is

deleted (Figure 8.24, DEL control bin). The deletion heteroplasmy level within this cell line is ~70%, as shown by the 3:1 ratio of *MT-ND1* to *MT-COIII* (Figure 8.24).

This result confirmed that there were no mtDNA deletions in the control HEK293, Flag-PRDM9 or PRDM9 transgenic cell lines.
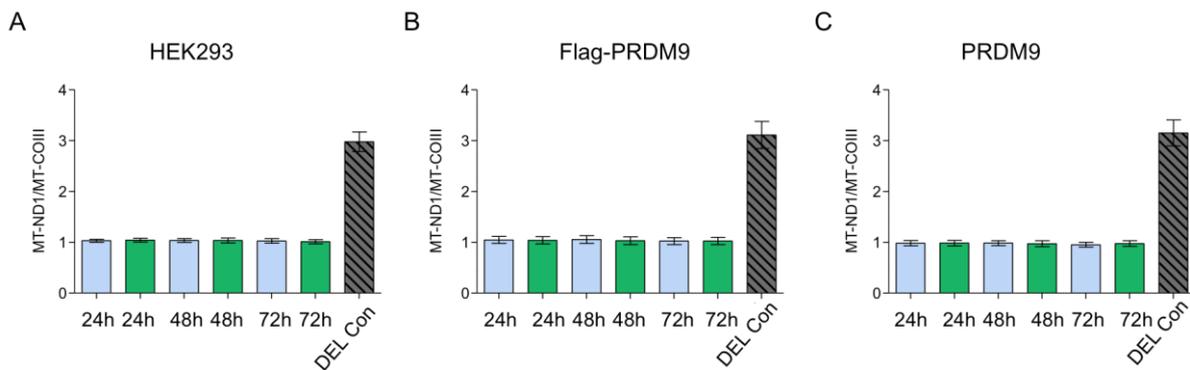
Figure 8.24 Quantification of mtDNA deletion level from qPCR data in overexpression cell lines. Cells were untreated (blue) or treated with 1 µg/mL tetracycline (green). Ratios of *MT-ND1* and *MT-COIII* amplicons were compared and plotted using Graphpad analysis software. A deletion control DNA sample was run in triplicate on each qPCR plate (DEL control). Students t-test was used to assess significance of each treatment group vs the un-treated partnered group. ≥0.05 = *: ≥0.005 = **: ≥0.0005 = ***. n=27

8.4.7 DNA damage repair pathway analysis

The function of PRDM9 protein is to mark sites of the genome for DNA double strand breakage and repair through the breast cancer type 2 susceptibility protein (BRCA2) mediated homologous recombination pathway. The BRCA2 protein interacts with Rad51 recombinase to mediate DNA repair mechanisms during DNA crossover events which are an integral part of the meiotic cell cycle (Thorslund *et al.*, 2007; Chatterjee *et al.*, 2016). The Mre11 double strand break repair nuclease protein complex is important for homologous recombination via BRCA1 mediated DNA repair mechanisms during the mitotic cell cycle (Jensen and Russell, 2016). Importantly, these two repair mechanisms act at different times due to cell type and whether the cell is undergoing meiotic or mitotic cell divisions (Roy *et al.*, 2012). In order to assess whether the PRDM9 expressed in our cell lines was influencing either pathway, protein levels of Rad51 and Mre11 were assessed through immunodetection (Figure 8.25).
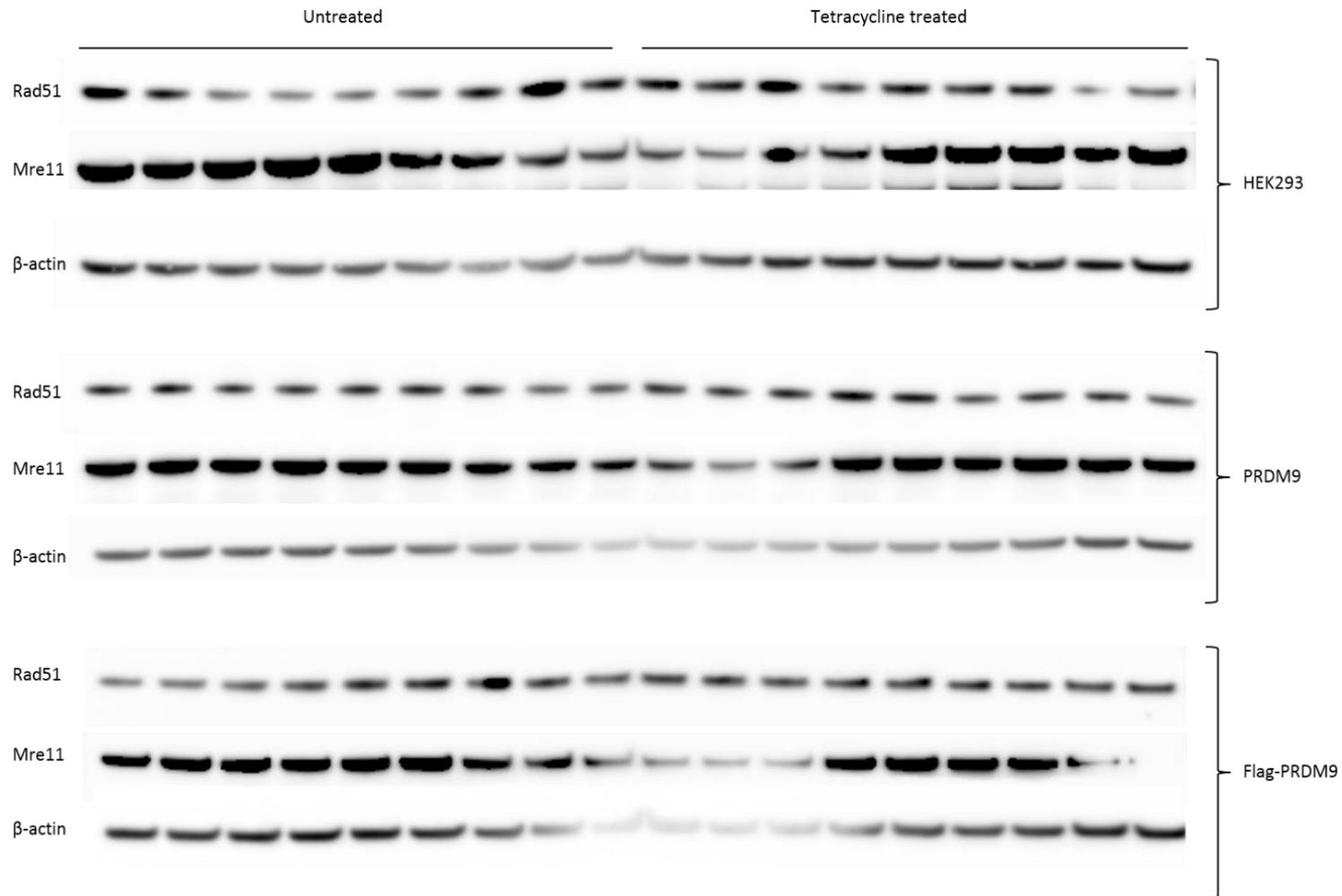
Figure 8.25 Immunodetection of Rad51 and Mre11 protein levels in cell lines. PRDM9 and Flag-PRDM9 overexpression cell lines were compared to the HEK293 parental cell line. Cells were treated with 1 μg/mL tetracycline or untreated for 72 hours. Membranes were incubated with anti-Rad51 (1:1000), anti-Mre11 (1:1000) or anti-βactin (1:1000) overnight at 4 °C.

Rad51 protein levels measured in three independent experiments were quantified (Figure 8.26). In all three cell lines protein levels were variable however there was a general trend towards a decrease in Rad51 protein levels in the treated groups after 48 and 72 hours of tetracycline treatment. However, this was only significantly different in the treated Flag-PRDM9 cell line at 48 and 72 hour time points (Figure 8.26 B).
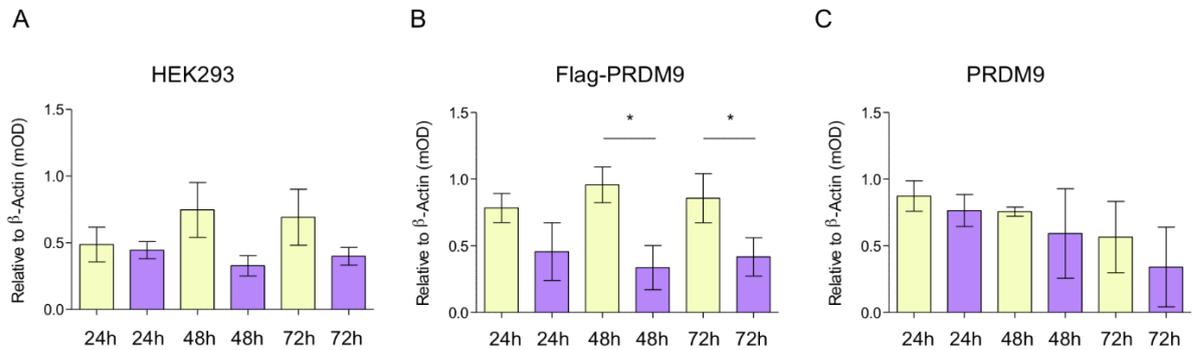
Figure 8.26 Quantification of Western blot analysis for Rad51 protein expression. Untreated samples are shown in yellow, tetracycline treated samples are shown in lilac. Densitometric analysis was performed using ImageJ software. Values were normalised to β-actin levels and analysed using Graphpad analysis software. Error bars show SD. Students t-test was used to assess significance of each treatment group vs the un-treated partnered group. ≥0.05 = *: ≥0.005 = **: ≥0.0005 = ***. n=9

Quantification of Mre11 protein levels showed that there was no significant difference in the transgenic cell lines when treated with tetracycline compared to untreated controls (Figure 8.27). There was a significant decrease in Mre11 protein level in the HEK293 control cell line at 24 hours tetracycline treatment.
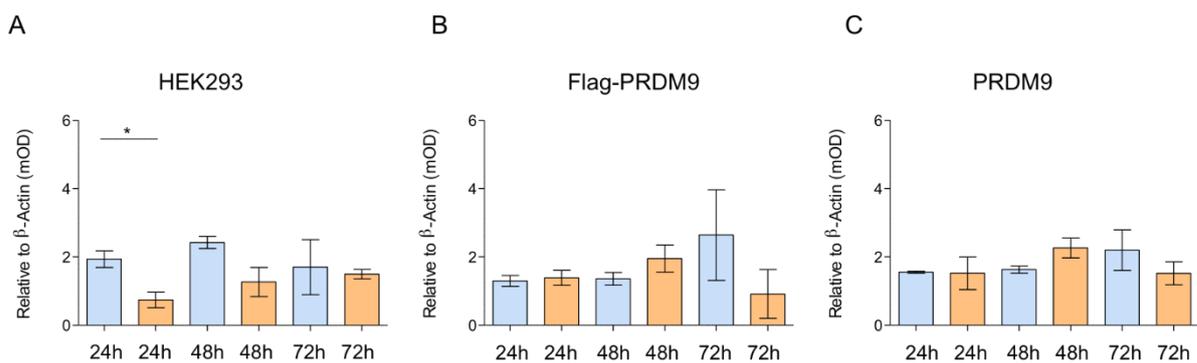
Figure 8.27 Quantification of Western blot analysis for Mre11 protein expression. Untreated samples are shown in blue, tetracycline treated samples are shown in orange. Densitometric analysis was performed using ImageJ software. Values were normalised to β-actin levels and analysed using Graphpad analysis software. Error bars show SD. Students t-test was used to assess significance of each treatment group vs the un-treated partnered group. ≥0.05 = *: ≥0.005 = **: ≥0.0005 = ***. n=9

8.5 Discussion

In order to investigate the cellular functions of human PRDM9 protein, full length cDNA constructs were stably integrated into the HEK293 cell genome under the control of a tetracycline inducible promoter. In this model, it is possible to switch on the overexpression of the cDNA construct by addition of tetracycline to the cell growth medium. This system allows for proteins which are potentially cytotoxic to be expressed in low quantities for a controlled period of time. This was a potentially advantageous system for PRDM9 protein overexpression as previous experiments indicated that this construct was difficult to express and isolate using transient expression plasmids (Chapter 7). Although it was not possible to reliably reproduce detectable full length PRDM9 protein expression using this experimental system, it was possible to indirectly measure PRDM9 function by assessing histone protein methylation status.

8.5.1 PRDM9 expression in stably transfected cell lines

When the cell lines were treated with tetracycline it appeared that PRDM9 protein expression was induced compared to untreated controls, as shown by Flag antibody immunodetection. However, this detection showed that protein expression was low and was not reliable as most lysates tested did not have detectable levels of PRDM9 protein. In addition, PRDM9 overexpression was more reliably detected in the cell line harbouring the Flag tagged construct than the untagged PRDM9 cell line. This is most likely due to the unreliable nature of the anti-PRDM9 antibody when compared to the well-established and experimentally verified anti-Flag antibody.

Quantifying PRDM9 mRNA transcript levels in the cell lines was also variable. A lack of positive PRDM9 control meant that optimisation of Taqman™ assays was difficult.

8.5.2 Indirect measurements of PRDM9 function

In order to establish whether PRDM9 was being transiently expressed at all in these two cell lines, we next asked whether it was possible to assess any functional consequences of this protein. Both murine and human PRDM9 have known methyltransferase activity and specifically target histone 3 lysine 4 (Hayashi *et al.*, 2005; Grey *et al.*, 2011; Eram *et al.*, 2014). It was possible to quantify tri-, di- and mono- methylation of H3K4 protein by

immunodetection using well established commercially available antibodies. The overexpression cell lines showed increased levels of both H3K4me3 and H3K4me2 after gene expression induction by tetracycline treatment. This effect was increased over time with cells treated for 72 hours showing the largest increase in methylation status when compared to 24 and 48 hours treatment as well as the HEK293 control samples. Previous methyltransferase activity studies have shown that recombinant PRDM9 peptide has a higher affinity for H3K4 protein when it is mono- or di- methylated suggesting that the primary function is to increase both di- and tri- methylation levels, consistent with our data. Levels of mono-methylation were increased in the Flag-PRDM9 cells treated with tetracycline suggesting that there is also the potential for the protein to act on un-methylated H3K4, consistent with previous reported data. The PRDM9 cDNA construct might be more stable when expressed with the Flag tag present than without it.

Overall, the assessment of Rad51 and Mre11 protein levels showed variation in the expression of both when cells were treated with tetracycline compared to untreated controls. Although there was an overall trend towards decreased expression of Rad51 in the Flag-PRDM9 transgenic cell line there was a similar trend in the HEK293 control cell line, although not statistically significant. This data suggests that attempting to dissect the protein family members involved in DSB repair mechanisms is complex and variable depending on treatment with tetracycline and timing of cell divisions. In this case, the effect of PRDM9 overexpression on Rad51 and Mre11 protein levels cannot be elucidated using this method.

8.5.3 Perspectives and future recommendations

A major problem with the use of this stable overexpression experimental system is that tetracycline is used as the inducer of expression by the CMV promoter element. Tetracyclines are a widely used group of antibiotics in agriculture and healthcare applications however they are also potent inhibitors of mitochondrial biogenesis, through the binding of rRNA elements of the 16S ribosome (Chatzispyrou *et al.*, 2015). However, investigations into the exact binding sites and the mechanisms by which tetracyclines protect against bacterial, viral, protozoan and non-infectious health conditions remain conflicted in the literature. The broad range of targets inhibited by

tetracycline and its derivatives suggests that RNA species present in the cell may be the elements which are bound and inhibited, in particular double stranded RNA molecules (Chukwudi, 2016). In any case, the use of tetracycline or its derivatives in laboratory research, and in particular mitochondrial based investigations, has limitations with respect to data analysis and interpretation (Chatzispyrou *et al.*, 2015; Moullan *et al.*, 2015). Any observation deemed statistically significant is likely to be confounded by the impact of the treatment itself rather than exclusively the effect of the protein/gene of interest, in this case PRDM9 overexpression. To overcome this, cell lines could be created to constitutively overexpress PRDM9 without the need for a tetracycline controlled promoter sequence. This was not attempted since experiments using transient constitutive expression did not show any detectable PRDM9 expression (Chapter 7) however this result might differ when using a stably transfected model.

In this experimental system, PRDM9 is being expressed in a mammalian cell line which is not undergoing meiosis and therefore the protein is not required by the cell. The functional role of PRDM9 as a methyltransferase is primarily to allow DSB events to occur followed by homologous chromosome crossovers. This process itself would be potentially detrimental to a post mitotic cell which has no requirement for DSB events within the nucleus. In this case, the cell would presumably try to silence any transcript or peptide which might be contributing to DSB events through several possible cellular degradation pathways.

Proteasomal degradation is a complex and highly regulated cellular pathway which targets proteins or peptide fragments for breakdown and recycling in response to several cellular signals. For example, proteins could be misfolded, aggregated, retarded or pathogenic therefore compromising cellular function in some way. In such cases, the proteins are targeted for ubiquitination followed by proteasomal degradation either through autophagosome formation or delivery by chaperone protein complexes (Hjerpe *et al.*, 2016). It is possible to experimentally block such protein degradation through several small chemical inhibitors or antibiotics. For example, chloroquine and leupeptin are well-established inhibitors of lysosomal degradation and has previously been used to show lysosome specific mRNA-protein complex degradation and endocytosis (Seguin *et al.*, 2014; Hashimoto *et al.*, 2015). The compound MG132 is a potent proteasome inhibitor which specifically targets the 26S pathway (Han *et al.*, 2009) and can be used to elucidate the proteasomal pathway responsible for degradation of proteins of interest

(Pedersen *et al.*, 2016). Brefeldin A blocks protein transport from the ER to the Golgi apparatus and can be used to elucidate the levels of protein recycling (Fernandez-Messina *et al.*, 2016). Cyclohexamide binds to the ribosome preventing the translocation step of peptide elongation in eukaryotic cells (Schneider-Poetsch *et al.*, 2010). Using cyclohexamide in pulse-chase experiments could be informative for assessing the rate of protein degradation, although it would only be specific to the protein of interest through immunoprecipitation pull down assays.

The functional role of PRDM9 in nuclear recombination is also unknown. Recent data shows that PRDM9 most likely interacts with SPO11 which is required for DNA double strand break formation (Lange *et al.*, 2016) and it is reasonable to assume that other proteins within this pathway are interacting or are affected by PRDM9 during meiosis. The cell lines created could be used to further uncover the mechanisms underlying meiotic recombination. Chromosome crossover events and global histone methylation changes could be studied in this model however investigations would still require stabilisation of overexpressed PRDM9 protein.

Although it was possible to create cell lines harbouring full length human PRDM9 cDNA under the control of a tetracycline inducible promoter, the detection of PRDM9 by immunoblotting, immunofluorescence or RT-qPCR was not reproducible. Indirectly measuring PRDM9 function through H3K4 methylation status shows that our model is consistent with previously published data with respect to its nuclear methyltransferase activity. Due to a lack of expression, it was not possible to determine PRDM9 cellular localisation with respect to the mitochondria in the cell lines created. However, through further efforts to block degradation of this protein by the proteasome, lysosome or mRNA processing pathways, it would be possible to use this model to answer further questions surrounding the localisation and function of PRDM9.

Chapter 9 General Discussion

9.1 Overview

In addition to the detailed discussion sections in each results chapter, the main findings of this thesis will be described in the context of the results as a whole. The limitations of each study will be described along with suggestions for future investigations which will be necessary to fully determine the observations outlined in this thesis.

The main aim of this thesis was to determine the presence of PRDM9 binding sites within the human mtDNA sequence and to elucidate the mechanism by which PRDM9 protein potentially binds to the molecule, or functions within the mitochondria. In order to investigate the possible link between PRDM9 and mtDNA deletion mediation, three broad approaches were undertaken; an *in silico* assessment of PRDM9 motifs within mtDNA sequences, a PRDM9 genotyping association study in a single deletion patient cohort and investigations into PRDM9 protein function using several approaches.

9.2 Main findings and experimental limitations

9.2.1 PRDM9 binding motifs are present in human mtDNA

A DNA motif was found to be enriched within recombination hotspots throughout the human genome and was later identified as the target of PRDM9 protein binding (Myers *et al.*, 2008). The function of human PRDM9 protein is of interest in a mitochondrial context due to the previous identification of a protein binding motif within the mtDNA sequence (Myers *et al.*, 2008). The presence of PRDM9 motifs within the mtDNA was further explored in this thesis.

Chapter 4 outlines the *in silico* method used to detect the presence of 11 independent PRDM9 motifs previously identified within the nuclear genome (Jeffreys *et al.*, 2013). Motifs were identified in a large cohort (n=31,554) of mtDNA sequences screened using this pipeline. Overall, the 11 motif sites showed clustering within the molecule at sites within both the major and minor arcs. Population stratification was then performed, revealing differences in motif presence between populations shown by broad haplogrouping of European, Asian and African sequences. The motifs found within the

mtDNA sequences trended towards the expected population specific *PRDM9* allele. For example, European sequences were expected to contain more motif sites recognized by allele 'A' than African sequences, which in contrast show more motif sites recognized by allele 'C'. This was expected based on previous *PRDM9* allele data from both European and African genome analysis (Berg *et al.*, 2010; Berg *et al.*, 2011; Jeffreys *et al.*, 2013). Some motif positions identified in this analysis were described previously as flanking mtDNA deletion sites (Samuels *et al.*, 2004) suggesting that there might be a functional role for PRDM9 protein within the organelle. The data suggest that PRDM9 motif sites are prevalent within the molecule and that the position and number of these sites is somewhat determined by population mtDNA haplogroup.

This analysis alone does not prove that there is any functional role for PRDM9 within the mitochondria. Available mtDNA deletion breakpoint data was therefore downloaded and aligned with the known PRDM9 motif sites to ascertain whether or not the two data sets show a pattern of overlap. It was observed that several breakpoints occur at or within PRDM9 motif sites on the mtDNA molecule suggesting that there was adequate data to prompt functional investigations into the potential role of this protein within the mitochondria *in vivo*.

9.2.2 PRDM9 alleles are not associated with increased risk of mtDNA single deletion

The *PRDM9* ZnF region is highly polymorphic due to rapidly changing hotspot usage throughout human evolution (Myers *et al.*, 2010; Lesecque *et al.*, 2014). Interestingly, SNPs within the ZnF region of this gene have been associated with several genomic rearrangement disorders, cancer and infertility (Irie *et al.*, 2009; Borel *et al.*, 2012; Hussin *et al.*, 2013), though the true pathogenic nature of these variants remains to be determined in larger cohorts. In order to determine whether *PRDM9* alleles are associated with a greater risk of mtDNA deletion formation, the ZnF region was genotyped in a cohort of 48 single deletion patients versus 50 healthy controls. Although multiple *PRDM9* alleles were identified in both case and control cohorts, there was no association between any of the alleles and deletion risk. However, there was a statistically significant difference in the number of heterozygous individuals in the case group compared to the controls. This result could be due to sample bias as only 98 individuals were genotyped and only 48 of those were patients, or it could be that increased heterozygosity at the *PRDM9* allele locus is influencing mtDNA deletion

formation in this patient group. This genotyping analysis would need to be performed using a larger cohort to determine whether this heterozygosity observation is truly statistically significant. Additionally, further information concerning characterisation of patient mtDNA could be used to better understand whether there is any difference in *PRDM9* allele distribution due to haplogroup status or deletion breakpoint sites.

9.2.3 Identifying a model system in which to study PRDM9 protein function

Expression of PRDM9 is predicted to occur during meiosis of male and female gametes (Hayashi *et al.*, 2005; Sun *et al.*, 2015). For this reason, PRDM9 protein studies in mice have focused on using testes tissue, where spermatocytes undergo maturation (Parvanov *et al.*, 2010). Little is known about the expression pattern of the human PRDM9 protein however it is assumed to be similar, if not identical, to that of the mouse. Studies into the functional role of human PRDM9 have used partial protein constructs, either using the C-terminal ZnF array or a section of the protein spanning the PR/SET domain (Wu *et al.*, 2013; Eram *et al.*, 2014; Patel *et al.*, 2016).

In this study, detection of PRDM9 protein was tested in several cell lines and human tissue samples. In order to use tissues as close as possible to embryonic origin, we established a collection of female gonadal samples including ovary, placenta and foetal gonad. The cell type explored most in this thesis was HEK293, which is embryonic in origin. Using protein lysates from tissue and cells, PRDM9 protein detection was not possible, as confirmed by mass spectrometry of immunoprecipitated protein. Several different PRDM9 antibodies were used throughout this study however, no positive PRDM9 signal could be detected. This is most likely due to two factors; the sensitivity of protein detection relative to abundance within the cell/tissue and the unreliability of appropriate antibodies used to detect this protein. Detection of PRDM9 mRNA expression was also performed using RT-qPCR. Although standard curve analysis showed that there was amplification of the target transcript, it was present at in very low quantity making detection difficult. Together these data suggest that PRDM9 transcript levels are in low abundance within the cell lines and tissue types tested and that the mRNA species are likely rapidly degraded when the cell is not undergoing active meiosis. This presents as a major limitation in studying the functional role of this protein in human cells and tissues.

9.2.4 Characterisation of a PRDM9 overexpression system

Overexpression of full length PRDM9 was attempted using a transient plasmid transfection system. This system showed partial success, as PRDM9 was detected on three occasions by immunoblotting. However, the variability of this technique in addition to the relatively low number of cells used in this experimental system lead us to create a stable overexpression model. These cell lines showed modest PRDM9 expression under the control of a tetracycline inducible promoter. One cell line, harboring PRDM9 fused to a Flag tag, showed expression of a protein at the expected size of 103 kDa, specifically detected with anti-Flag. The second cell line, expressing untagged PRDM9, did not always show expression of a protein at the expected size, possibly due to the nonspecific nature of the PRDM9 antibody.

In order to assess whether PRDM9 was not being adequately expressed or being rapidly degraded by the cell, histone H3 methylation status was assessed. Interestingly, the cell lines overexpressing PRDM9 showed an increased amount of H3K4 methylation. Eram *et al.* (2014) showed previously that overexpression of a partial human PRDM9 construct was able to carry out methyltransferase activity in HEK293 cells. This is confirmed in our model using the full length PRDM9 protein.

9.3 Future investigations

9.3.2 PRDM9 localisation

Whilst the PRDM9 overexpression cell lines showed increased H3K4 methylation, there was variable detection of the protein itself. When PRDM9 protein was detectable by immunoblot, it was expressed in low quantities compared to other proteins suggesting that PRDM9 is being actively degraded by the cell, possibly through ubiquitin pathways and the proteasome. A lack of detection of mRNA also suggests that the mRNA product could be targeted for degradation by deadenylation or endonuclease mediated pathways. These hypotheses could be experimentally tested by adding small chemical inhibitors to the cell lines in culture.

It is important to establish whether PRDM9 has any mitochondrial localization before further investigations are carried out. Overexpression of a partial PRDM9 peptide sequence was successful (as shown in Chapter 7). This experiment could be extended to

other regions of the protein to establish whether any are targeted to the mitochondria. Based on the observations from such experiments, further organelle investigations could be warranted.

9.3.4 Studies using animal models

Using animal models such as mouse or zebrafish would be beneficial to further explore the potential role of PRDM9 in the mitochondria. Zebrafish can be used as an excellent model for oogenesis and reproduction, as gametocytes can be easily harvested (Elkouby and Mullins, 2016). These qualities could be employed to further understand the timing of expression of *PRDM9* and localization of PRDM9 protein. Transgenic zebrafish lines are widely available and could be used to model whether a loss of the *PRDM9* gene has any effect on cellular function (Howe *et al.*, 2017). Localisation of PRDM9 to the nucleus was observed in mouse testes tissue (Flachs *et al.*, 2014). However, the authors concluded that PRDM9 shuttles to the cytoplasm, although its function in this compartment was not clear and has not been studied. Immunofluorescent imaging analysis of this tissue could be explored using different antibody markers to understand where PRDM9 localises to within the cytosol. Lysosomal, endoplasmic reticulum and mitochondrial specific markers would help to further understand what happens to PRDM9 when it does not translocate to the nuclear compartment.

Additionally, a putative binding site for PRDM9 protein has been identified in the mouse genome and this could be explored in the mtDNA sequence of several mouse strains using publically available sequencing data (Davies *et al.*, 2016). The *in silico* approach outlined in this study could be extended to various taxa, specifically to species closely related to *Homo sapiens* to determine whether a conserved mechanism for PRDM9 binding in mammalian mtDNA exists. This has not been explored, most likely due to a lack of identification of the PRDM9 motif in each individual species. Due to the polymorphic nature of the gene, it is probable that the binding motif varies between species.

9.3.3 Elucidating the role of PRDM9 in the nucleus

Partial recombinant PRDM9 protein has been used previously to assess methyltransferase activity and binding affinity of PRDM9 protein to oligonucleotides containing the genomic recognition motif (Wu *et al.*, 2013; Eram *et al.*, 2014). In addition, a recent study showed that PRDM9 interacts with SPO11 further supporting

the role of PRDM9 in the nDNA double strand break pathway (Lange *et al.*, 2016). The cell lines created in this project could be utilized to further investigate the mechanisms of action of the protein within the nucleus and perhaps help uncover other potential protein-protein interactions. Directed mutagenesis of PRDM9 in this cell model could be used to understand how different regions of the protein affect nuclear genomic recombination patterns.

9.4 Final conclusions

Throughout this thesis the experimental steps taken to determine whether the recombination regulator PRDM9 has involvement in the mediation of mtDNA deletions have been described.

- PRDM9 recognition motifs were identified within 99.9% of human mtDNA sequences, however the true functional role of this protein within the mitochondria remains unknown.
- Finding a model in which to study PRDM9 protein function proved difficult due to its meiotic specific expression profile.
- Two stable overexpression cell lines were created and showed promising results based on Flag-tag immunodetection and upregulated histone H3 lysine methylation. This cell culture model may provide some insights into the function of PRDM9 but must be interpreted with caution due to the biologically engineered nature of this system.
- Using animal models would allow further charactrisation of PRDM9 in the tissues and cell types in which it is normally expressed, i.e. testes and gametocytes.

Mitochondrial localization and mtDNA binding have not been determined during this study, however to the best of my knowledge this is the first time a cell culture model has been created to express full length human PRDM9 and might be of use in future experiments.

References

Al Rawi, S., Louvet-Vallee, S., Djeddi, A., Sachse, M., Culetto, E., Hajjar, C., Boyd, L., Legouis, R. and Galy, V. (2011) 'Postfertilization autophagy of sperm organelles prevents paternal mitochondrial DNA transmission', *Science*, 334(6059), pp. 1144-7.

Alhomidi, M.A., Vedicherla, B., Movva, S., Rao, P.K., Ahuja, Y.R. and Hasan, Q. (2013) 'Mitochondrial D310 instability in Asian Indian breast cancer patients', *Tumour Biol*, 34(4), pp. 2427-32.

Ancelin, K., Lange, U.C., Hajkova, P., Schneider, R., Bannister, A.J., Kouzarides, T. and Surani, M.A. (2006) 'Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells', *Nat Cell Biol*, 8(6), pp. 623-30.

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M. and Howell, N. (1999) 'Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA', *Nat Genet*, 23(2), p. 147.

Ardissone, A., Lamantea, E., Invernizzi, F., Zeviani, M., Genitrini, S., Moroni, I. and Uziel, G. (2014) 'Mitochondrial Diseases in Childhood', *Curr Mol Med*.

Auton, A., Fledel-Alon, A., Pfeifer, S., Venn, O., Segurel, L., Street, T., Leffler, E.M., Bowden, R., Aneas, I., Broxholme, J., Humburg, P., Iqbal, Z., Lunter, G., Maller, J., Hernandez, R.D., Melton, C., Venkat, A., Nobrega, M.A., Bontrop, R., Myers, S., Donnelly, P., Przeworski, M. and McVean, G. (2012) 'A fine-scale chimpanzee genetic map from population sequencing', *Science*, 336(6078), pp. 193-8.

Axelsson, E., Webster, M.T., Ratnakumar, A., Consortium, L., Ponting, C.P. and Lindblad-Toh, K. (2012) 'Death of PRDM9 coincides with stabilization of the recombination landscape in the dog genome', *Genome Res*, 22(1), pp. 51-63.

Baker, C.L., Kajita, S., Walker, M., Saxl, R.L., Raghupathy, N., Choi, K., Petkov, P.M. and Paigen, K. (2015) 'PRDM9 drives evolutionary erosion of hotspots in Mus musculus through haplotype-specific initiation of meiotic recombination', *PLoS Genet*, 11(1), p. e1004916.

Balcova, M., Faltusova, B., Gergelits, V., Bhattacharyya, T., Mihola, O., Trachtulec, Z., Knopf, C., Fotopulosova, V., Chvatalova, I., Gregorova, S. and Forejt, J. (2016) 'Hybrid Sterility Locus on Chromosome X Controls Meiotic Recombination Rate in Mouse', *PLoS Genet*, 12(4), p. e1005906.

Basha Gutierrez, J. and Nakai, K. (2016) 'A study on the application of topic models to motif finding algorithms', *BMC Bioinformatics*, 17(Suppl 19), p. 502.

Basu, A., Chaudhuri, P. and Majumder, P.P. (2005) 'Identification of polymorphic motifs using probabilistic search algorithms', *Genome Res*, 15(1), pp. 67-77.

Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., Coop, G. and de Massy, B. (2010) 'PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice', *Science*, 327(5967), pp. 836-40.

Baum, H. and Rieske, J.S. (1966) 'Involvement of water in electron transport in complexes III and IV of the mitochondrial electron transport chain', *Biochem Biophys Res Commun*, 24(1), pp. 1-9.

Bellizzi, D., D'Aquila, P., Scafone, T., Giordano, M., Riso, V., Riccio, A. and Passarino, G. (2013) 'The control region of mitochondrial DNA shows an unusual CpG and non-CpG methylation pattern', *DNA Res*, 20(6), pp. 537-47.

Berg, I.L., Neumann, R., Lam, K.W., Sarbajna, S., Odenthal-Hesse, L., May, C.A. and Jeffreys, A.J. (2010) 'PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans', *Nat Genet*, 42(10), pp. 859-63.

Berg, I.L., Neumann, R., Sarbajna, S., Odenthal-Hesse, L., Butler, N.J. and Jeffreys, A.J. (2011) 'Variants of the protein PRDM9 differentially regulate a set of human meiotic recombination hotspots highly active in African populations', *Proc Natl Acad Sci U S A*, 108(30), pp. 12378-83.

Berger, S.L., Kouzarides, T., Shiekhattar, R. and Shilatifard, A. (2009) 'An operational definition of epigenetics', *Genes Dev*, 23(7), pp. 781-3.

Beri, S., Bonaglia, M.C. and Giorda, R. (2013) 'Low-copy repeats at the human VIPR2 gene predispose to recurrent and nonrecurrent rearrangements', *Eur J Hum Genet*, 21(7), pp. 757-61.

Billings, T., Parvanov, E.D., Baker, C.L., Walker, M., Paigen, K. and Petkov, P.M. (2013) 'DNA binding specificities of the long zinc-finger recombination protein PRDM9', *Genome Biol*, 14(4), p. R35.

Blazer, L.L., Lima-Fernandes, E., Gibson, E., Eram, M.S., Loppnau, P., Arrowsmith, C.H., Schapira, M. and Vedadi, M. (2016) 'PR Domain-containing Protein 7 (PRDM7) Is a Histone 3 Lysine 4 Trimethyltransferase', *J Biol Chem*, 291(26), pp. 13509-19.

Bogenhagen, D., Gillum, A.M., Martens, P.A. and Clayton, D.A. (1979) 'Replication of mouse L-cell mitochondrial DNA', *Cold Spring Harb Symp Quant Biol*, 43 Pt 1, pp. 253-62.

Bogenhagen, D.F. (2012) 'Mitochondrial DNA nucleoid structure', *Biochim Biophys Acta*, 1819(9-10), pp. 914-20.

Borel, C., Cheung, F., Stewart, H., Koolen, D.A., Phillips, C., Thomas, N.S., Jacobs, P.A., Eliez, S. and Sharp, A.J. (2012) 'Evaluation of PRDM9 variation as a risk factor for recurrent genomic disorders and chromosomal non-disjunction', *Hum Genet*, 131(9), pp. 1519-24.

Boulton, A., Myers, R.S. and Redfield, R.J. (1997) 'The hotspot conversion paradox and the evolution of meiotic recombination', *Proc Natl Acad Sci U S A*, 94(15), pp. 8058-63.

Bradford, M.M. (1976) 'A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding', *Anal Biochem*, 72, pp. 248-54.

Brick, K., Smagulova, F., Khil, P., Camerini-Otero, R.D. and Petukhova, G.V. (2012) 'Genetic recombination is directed away from functional genomic elements in mice', *Nature*, 485(7400), pp. 642-5.

Brierley, E.J., Johnson, M.A., Lightowlers, R.N., James, O.F. and Turnbull, D.M. (1998) 'Role of mitochondrial DNA mutations in human aging: implications for the central nervous system and muscle', *Ann Neurol*, 43(2), pp. 217-23.

Brown, M.D., Hosseini, S.H., Torroni, A., Bandelt, H.J., Allen, J.C., Schurr, T.G., Scozzari, R., Cruciani, F. and Wallace, D.C. (1998) 'mtDNA haplogroup X: An ancient link between Europe/Western Asia and North America?', *Am J Hum Genet*, 63(6), pp. 1852-61.

Brown, M.D., Sun, F. and Wallace, D.C. (1997) 'Clustering of Caucasian Leber hereditary optic neuropathy patients containing the 11778 or 14484 mutations on an mtDNA lineage', *Am J Hum Genet*, 60(2), pp. 381-7.

Brown, T.A., Cecconi, C., Tkachuk, A.N., Bustamante, C. and Clayton, D.A. (2005) 'Replication of mitochondrial DNA occurs by strand displacement with alternative light-strand origins, not via a strand-coupled mechanism', *Genes Dev*, 19(20), pp. 2466-76.

Bua, E., Johnson, J., Herbst, A., Delong, B., McKenzie, D., Salamat, S. and Aiken, J.M. (2006) 'Mitochondrial DNA-deletion mutations accumulate intracellularly to detrimental levels in aged human skeletal muscle fibers', *Am J Hum Genet*, 79(3), pp. 469-80.

Buard, J., Rivals, E., Dunoyer de Segonzac, D., Garres, C., Caminade, P., de Massy, B. and Boursot, P. (2014) 'Diversity of Prdm9 zinc finger array in wild mice unravels new facets of the evolutionary turnover of this coding minisatellite', *PLoS One*, 9(1), p. e85021.

Cai, J., Yue, Y., Wang, Y., Jin, Z., Jin, F. and Wu, C. (2016) 'Quantitative study of effects of free cationic chains on gene transfection in different intracellular stages', *J Control Release*, 238, pp. 71-79.

Calvo, S.E., Clauser, K.R. and Mootha, V.K. (2016) 'MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins', *Nucleic Acids Res*, 44(D1), pp. D1251-7.

Campbell, C.L., Bherer, C., Morrow, B.E., Boyko, A.R. and Auton, A. (2016) 'A Pedigree-Based Map of Recombination in the Domestic Dog Genome', *G3 (Bethesda)*.

Campbell, G., Krishnan, K.J., Deschauer, M., Taylor, R.W. and Turnbull, D.M. (2014) 'Dissecting the mechanisms underlying the accumulation of mitochondrial DNA deletions in human skeletal muscle', *Hum Mol Genet*.

Cann, R.L., Stoneking, M. and Wilson, A.C. (1987) 'Mitochondrial DNA and human evolution', *Nature*, 325(6099), pp. 31-6.

Cao, L., Shitara, H., Horii, T., Nagao, Y., Imai, H., Abe, K., Hara, T., Hayashi, J. and Yonekawa, H. (2007) 'The mitochondrial bottleneck occurs without reduction of mtDNA content in female mouse germ cells', *Nat Genet*, 39(3), pp. 386-90.

Cardon, L.R., Burge, C., Clayton, D.A. and Karlin, S. (1994) 'Pervasive CpG suppression in animal mitochondrial genomes', *Proc Natl Acad Sci U S A*, 91(9), pp. 3799-803.

Carelli, V., Achilli, A., Valentino, M.L., Rengo, C., Semino, O., Pala, M., Olivieri, A., Mattiazzi, M., Pallotti, F., Carrara, F., Zeviani, M., Leuzzi, V., Carducci, C., Valle, G., Simionati, B., Mendieta, L., Salomao, S., Belfort, R., Jr., Sadun, A.A. and Torroni, A. (2006) 'Haplogroup effects and recombination of mitochondrial DNA: novel clues from the analysis of Leber hereditary optic neuropathy pedigrees', *Am J Hum Genet*, 78(4), pp. 564-74.

Chaban, Y., Boekema, E.J. and Dudkina, N.V. (2014) 'Structures of mitochondrial oxidative phosphorylation supercomplexes and mechanisms for their stabilisation', *Biochim Biophys Acta*, 1837(4), pp. 418-26.

Chang, D.D., Hauswirth, W.W. and Clayton, D.A. (1985) 'Replication priming and transcription initiate from precisely the same site in mouse mitochondrial DNA', *EMBO J*, 4(6), pp. 1559-67.

Chatterjee, G., Jimenez-Sainz, J., Presti, T., Nguyen, T. and Jensen, R.B. (2016) 'Distinct binding of BRCA2 BRC repeats to RAD51 generates differential DNA damage sensitivity', *Nucleic Acids Res*, 44(11), pp. 5256-70.

Chatzispyrou, I.A., Held, N.M., Mouchiroud, L., Auwerx, J. and Houtkooper, R.H. (2015) 'Tetracycline antibiotics impair mitochondrial function and its experimental use confounds research', *Cancer Res*, 75(21), pp. 4446-9.

Chen, H., Vermulst, M., Wang, Y.E., Chomyn, A., Prolla, T.A., McCaffery, J.M. and Chan, D.C. (2010) 'Mitochondrial fusion is required for mtDNA stability in skeletal muscle and tolerance of mtDNA mutations', *Cell*, 141(2), pp. 280-9.

Chen, Y.S., Torroni, A., Excoffier, L., Santachiara-Benerecetti, A.S. and Wallace, D.C. (1995) 'Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups', *Am J Hum Genet*, 57(1), pp. 133-49.

Chinnery, P.F., DiMauro, S., Shanske, S., Schon, E.A., Zeviani, M., Mariotti, C., Carrara, F., Lombes, A., Laforet, P., Ogier, H., Jaksch, M., Lochmuller, H., Horvath, R., Deschauer, M., Thorburn, D.R., Bindoff, L.A., Poulton, J., Taylor, R.W., Matthews, J.N. and Turnbull, D.M. (2004) 'Risk of developing a mitochondrial DNA deletion disorder', *Lancet*, 364(9434), pp. 592-6.

Chukwudi, C.U. (2016) 'rRNA Binding Sites and the Molecular Mechanism of Action of the Tetracyclines', *Antimicrob Agents Chemother*, 60(8), pp. 4433-41.

Collu-Marchese, M., Shuen, M., Pauly, M., Saleem, A. and Hood, D.A. (2015) 'The regulation of mitochondrial transcription factor A (Tfam) expression during skeletal muscle cell differentiation', *Biosci Rep*, 35(3).

Cooper, C. and Lehninger, A.L. (1957) 'Oxidative phosphorylation by an enzyme complex from extracts of mitochondria. V. The adenosine triphosphate-phosphate exchange reaction', *J Biol Chem*, 224(1), pp. 561-78.

Copeland, W.C. and Longley, M.J. (2014) 'Mitochondrial genome maintenance in health and disease', *DNA Repair (Amst)*.

Corneo, G., Moore, C., Sanadi, D.R., Grossman, L.I. and Marmur, J. (1966) 'Mitochondrial DNA in yeast and some mammalian species', *Science*, 151(3711), pp. 687-9.

Coto, E., Gomez, J., Alonso, B., Corao, A.I., Diaz, M., Menendez, M., Martinez, C., Calatayud, M.T., Moris, G. and Alvarez, V. (2011) 'Late-onset Alzheimer's disease is associated with mitochondrial DNA 7028C/haplogroup H and D310 poly-C tract heteroplasmy', *Neurogenetics*, 12(4), pp. 345-6.

Crane, F.L., Lester, R.L., Widmer, C. and Hatefi, Y. (1959) 'Studies on the electron transport system. XVIII. Isolation of coenzyme Q (Q275) from beef heart and beef heart mitochondria', *Biochim Biophys Acta*, 32(1), pp. 73-9.

Cree, L.M., Samuels, D.C., de Sousa Lopes, S.C., Rajasimha, H.K., Wonnapinij, P., Mann, J.R., Dahl, H.H. and Chinnery, P.F. (2008) 'A reduction of mitochondrial DNA molecules during embryogenesis explains the rapid segregation of genotypes', *Nat Genet*, 40(2), pp. 249-54.

Damas, J., Carneiro, J., Amorim, A. and Pereira, F. (2014) 'MitoBreak: the mitochondrial DNA breakpoints database', *Nucleic Acids Res*, 42(Database issue), pp. D1261-8.

Dang, N., Ma, X., Meng, X., An, L. and Pang, S. (2016) 'Dysregulated function of normal human epidermal keratinocytes in the absence of filaggrin', *Mol Med Rep*, 14(3), pp. 2566-72.

Das, J., Han, J.W., Choi, Y.J., Song, H., Cho, S.G., Park, C., Seo, H.G. and Kim, J.H. (2016) 'Cationic lipid-nanoceria hybrids, a novel nonviral vector-mediated gene delivery into mammalian cells: investigation of the cellular uptake mechanism', *Sci Rep*, 6, p. 29197.

das Neves, R.P., Jones, N.S., Andreu, L., Gupta, R., Enver, T. and Iborra, F.J. (2010) 'Connecting variability in global transcription rate to mitochondrial variability', *PLoS Biol*, 8(12), p. e1000560.

Davies, B., Hatton, E., Altemose, N., Hussin, J.G., Pratto, F., Zhang, G., Hinch, A.G., Moralli, D., Biggs, D., Diaz, R., Preece, C., Li, R., Bitoun, E., Brick, K., Green, C.M., Camerini-Otero, R.D., Myers, S.R. and Donnelly, P. (2016) 'Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice', *Nature*, 530(7589), pp. 171-6.

Dawid, I.B. (1974) '5-methylcytidylic acid: absence from mitochondrial DNA of frogs and HeLa cells', *Science*, 184(4132), pp. 80-1.

Dayama, G., Emery, S.B., Kidd, J.M. and Mills, R.E. (2014) 'The genomic landscape of polymorphic human nuclear mitochondrial insertions', *Nucleic Acids Res*, 42(20), pp. 12640-9.

de Vries, S., Albracht, S.P., Berden, J.A. and Slater, E.C. (1981) 'A new species of bound ubisemiquinone anion in QH2: cytochrome c oxidoreductase', *J Biol Chem*, 256(23), pp. 11996-8.

Derenko, M.V., Malyarchuk, B.A., Dambueva, I.K., Shaikhaev, G.O., Dorzhu, C.M., Nimaev, D.D. and Zakharov, I.A. (2000) 'Mitochondrial DNA variation in two South Siberian

Aboriginal populations: implications for the genetic history of North Asia', *Hum Biol*, 72(6), pp. 945-73.

Diaz, F., Bayona-Bafaluy, M.P., Rana, M., Mora, M., Hao, H. and Moraes, C.T. (2002) 'Human mitochondrial DNA with large deletions repopulates organelles faster than full-length genomes under relaxed copy number control', *Nucleic Acids Res*, 30(21), pp. 4626-33.

Dittwald, P., Gambin, T., Gonzaga-Jauregui, C., Carvalho, C.M., Lupski, J.R., Stankiewicz, P. and Gambin, A. (2013) 'Inverted low-copy repeats and genome instability--a genome-wide analysis', *Hum Mutat*, 34(1), pp. 210-20.

Dobzhansky, T. (1936) 'Studies on Hybrid Sterility. II. Localization of Sterility Factors in Drosophila Pseudoobscura Hybrids', *Genetics*, 21(2), pp. 113-35.

Doi, N., Zenno, S., Ueda, R., Ohki-Hamazaki, H., Ui-Tei, K. and Saigo, K. (2003) 'Short-interfering-RNA-mediated gene silencing in mammalian cells requires Dicer and eIF2C translation initiation factors', *Curr Biol*, 13(1), pp. 41-6.

Doynova, M.D., Berretta, A., Jones, M.B., Jasoni, C.L., Vickers, M.H. and O'Sullivan, J.M. (2016) 'Interactions between mitochondrial and nuclear DNA in mammalian cells are non-random', *Mitochondrion*.

Ebner, S., Mangge, H., Langhof, H., Halle, M., Siegrist, M., Aigner, E., Paulmichl, K., Paulweber, B., Datz, C., Sperl, W., Kofler, B. and Weghuber, D. (2015) 'Mitochondrial Haplogroup T Is Associated with Obesity in Austrian Juveniles and Adults', *PLoS One*, 10(8), p. e0135622.

Elkouby, Y.M. and Mullins, M.C. (2016) 'Methods for the analysis of early oogenesis in Zebrafish', *Dev Biol*.

Elstner, M., Andreoli, C., Klopstock, T., Meitinger, T. and Prokisch, H. (2009) 'The mitochondrial proteome database: MitoP2', *Methods Enzymol*, 457, pp. 3-20.

Emery, L.S., Magnaye, K.M., Bigham, A.W., Akey, J.M. and Bamshad, M.J. (2015) 'Estimates of continental ancestry vary widely among individuals with the same mtDNA haplogroup', *Am J Hum Genet*, 96(2), pp. 183-93.

Eom, G.H., Kim, K., Kim, S.M., Kee, H.J., Kim, J.Y., Jin, H.M., Kim, J.R., Kim, J.H., Choe, N., Kim, K.B., Lee, J., Kook, H., Kim, N. and Seo, S.B. (2009) 'Histone methyltransferase PRDM8 regulates mouse testis steroidogenesis', *Biochem Biophys Res Commun*, 388(1), pp. 131-6.

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. and Vedadi, M. (2014) 'Trimethylation of Histone H3 Lysine 36 by Human Methyltransferase PRDM9', *J Biol Chem*.

Farha, S., Hu, B., Comhair, S., Zein, J., Dweik, R., Erzurum, S.C. and Aldred, M.A. (2016) 'Mitochondrial Haplogroups and Risk of Pulmonary Arterial Hypertension', *PLoS One*, 11(5), p. e0156042.

Feichtinger, J., Aldeailej, I., Anderson, R., Almutairi, M., Almatrafi, A., Alsiwiehri, N., Griffiths, K., Stuart, N., Wakeman, J.A., Larcombe, L. and McFarlane, R.J. (2012) 'Meta-analysis of clinical data using human meiotic genes identifies a novel cohort of highly restricted cancer-specific marker genes', *Oncotarget*, 3(8), pp. 843-53.

Felsenstein, J. (1974) 'The evolutionary advantage of recombination', *Genetics*, 78(2), pp. 737-56.

Fernandez-Messina, L., Reyburn, H.T. and Vales-Gomez, M. (2016) 'A short half-life of ULBP1 at the cell surface due to internalization and proteosomal degradation', *Immunol Cell Biol*, 94(5), pp. 479-85.

Fernandez-Vizarra, E., Ferrin, G., Perez-Martos, A., Fernandez-Silva, P., Zeviani, M. and Enriquez, J.A. (2010) 'Isolation of mitochondria for biogenetical studies: An update', *Mitochondrion*, 10(3), pp. 253-62.

Finnila, S., Lehtonen, M.S. and Majamaa, K. (2001) 'Phylogenetic network for European mtDNA', *Am J Hum Genet*, 68(6), pp. 1475-84.

Flachs, P., Bhattacharyya, T., Mihola, O., Pialek, J., Forejt, J. and Trachtulec, Z. (2014) 'Prdm9 incompatibility controls oligospermia and delayed fertility but no selfish transmission in mouse intersubspecific hybrids', *PLoS One*, 9(4), p. e95806.

Flachs, P., Mihola, O., Simecek, P., Gregorova, S., Schimenti, J.C., Matsui, Y., Baudat, F., de Massy, B., Pialek, J., Forejt, J. and Trachtulec, Z. (2012) 'Interallelic and intergenic incompatibilities of the Prdm9 (Hst1) gene in mouse hybrid sterility', *PLoS Genet*, 8(11), p. e1003044.

Fog, C.K., Galli, G.G. and Lund, A.H. (2012) 'PRDM proteins: important players in differentiation and disease', *Bioessays*, 34(1), pp. 50-60.

Forejt, J., Vincek, V., Klein, J., Lehrach, H. and Loudova-Mickova, M. (1991) 'Genetic mapping of the t-complex region on mouse chromosome 17 including the Hybrid sterility-1 gene', *Mamm Genome*, 1(2), pp. 84-91.

French, S.L., Santangelo, T.J., Beyer, A.L. and Reeve, J.N. (2007) 'Transcription and translation are coupled in Archaea', *Mol Biol Evol*, 24(4), pp. 893-5.

Freyer, C., Cree, L.M., Mourier, A., Stewart, J.B., Koolmeister, C., Milenkovic, D., Wai, T., Floros, V.I., Hagstrom, E., Chatzidaki, E.E., Wiesner, R.J., Samuels, D.C., Larsson, N.G. and Chinnery, P.F. (2012) 'Variation in germline mtDNA heteroplasmy is determined prenatally but modified during subsequent transmission', *Nat Genet*, 44(11), pp. 1282-5.

Fuhrmann, G., Chung, A.C., Jackson, K.J., Hummelke, G., Baniahmad, A., Sutter, J., Sylvester, I., Scholer, H.R. and Cooney, A.J. (2001) 'Mouse germline restriction of Oct4 expression by germ cell nuclear factor', *Dev Cell*, 1(3), pp. 377-87.

Fuste, J.M., Wanrooij, S., Jemt, E., Granycome, C.E., Cluett, T.J., Shi, Y., Atanassova, N., Holt, I.J., Gustafsson, C.M. and Falkenberg, M. (2010) 'Mitochondrial RNA polymerase is needed for activation of the origin of light-strand DNA replication', *Mol Cell*, 37(1), pp. 67-78.

Ghosh, S., Sengupta, S. and Scaria, V. (2014) 'Comparative analysis of human mitochondrial methylomes shows distinct patterns of epigenetic regulation in mitochondria', *Mitochondrion*, 18, pp. 58-62.

Giordano, C., Iommarini, L., Giordano, L., Maresca, A., Pisano, A., Valentino, M.L., Caporali, L., Liguori, R., Deceglie, S., Roberti, M., Fanelli, F., Fracasso, F., Ross-Cisneros, F.N., D'Adamo, P., Hudson, G., Pyle, A., Yu-Wai-Man, P., Chinnery, P.F., Zeviani, M., Salomao, S.R., Berezovsky, A., Belfort, R., Jr., Ventura, D.F., Moraes, M., Moraes Filho, M., Barboni, P., Sadun, F., De Negri, A., Sadun, A.A., Tancredi, A., Mancini, M., d'Amati, G., Loguercio Polosa, P., Cantatore, P. and Carelli, V. (2014) 'Efficient mitochondrial biogenesis drives incomplete penetrance in Leber's hereditary optic neuropathy', *Brain*, 137(Pt 2), pp. 335-53.

Gitschlag, B.L., Kirby, C.S., Samuels, D.C., Gangula, R.D., Mallal, S.A. and Patel, M.R. (2016) 'Homeostatic Responses Regulate Selfish Mitochondrial Genome Dynamics in C. elegans', *Cell Metab*, 24(1), pp. 91-103.

Gomez-Duran, A., Pacheu-Grau, D., Lopez-Gallardo, E., Diez-Sanchez, C., Montoya, J., Lopez-Perez, M.J. and Ruiz-Pesini, E. (2010) 'Unmasking the causes of multifactorial disorders: OXPHOS differences between mitochondrial haplogroups', *Hum Mol Genet*, 19(17), pp. 3343-53.

Gomez-Duran, A., Pacheu-Grau, D., Martinez-Romero, I., Lopez-Gallardo, E., Lopez-Perez, M.J., Montoya, J. and Ruiz-Pesini, E. (2012) 'Oxidative phosphorylation differences between mitochondrial DNA haplogroups modify the risk of Leber's hereditary optic neuropathy', *Biochim Biophys Acta*, 1822(8), pp. 1216-22.

Goncalves, C., Akhter, S., Pichon, C. and Midoux, P. (2016) 'Intra-cellular availability of pDNA and mRNA after transfection: A comparative study in between polyplexes, lipoplexes and lipopolyplexes', *Mol Pharm*.

Goto, Y., Nonaka, I. and Horai, S. (1990) 'A mutation in the tRNA(Leu)(UUR) gene associated with the MELAS subgroup of mitochondrial encephalomyopathies', *Nature*, 348(6302), pp. 651-3.

Gould, M.P., Bosworth, C.M., McMahon, S., Grandhi, S., Grimerg, B.T. and LaFramboise, T. (2015) 'PCR-Free Enrichment of Mitochondrial DNA from Human Blood and Cell Lines for High Quality Next-Generation DNA Sequencing', *PLoS One*, 10(10), p. e0139253.

Grady, J.P., Campbell, G., Ratnaike, T., Blakely, E.L., Falkous, G., Nesbitt, V., Schaefer, A.M., McNally, R.J., Gorman, G.S., Taylor, R.W., Turnbull, D.M. and McFarland, R. (2014) 'Disease progression in patients with single, large-scale mitochondrial DNA deletions', *Brain*, 137(Pt 2), pp. 323-34.

Grey, C., Barthes, P., Chauveau-Le Friec, G., Langa, F., Baudat, F. and de Massy, B. (2011) 'Mouse PRDM9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination', *PLoS Biol*, 9(10), p. e1001176.

Groeneveld, L.F., Atencia, R., Garriga, R.M. and Vigilant, L. (2012) 'High diversity at PRDM9 in chimpanzees and bonobos', *PLoS One*, 7(7), p. e39064.

Guantes, R., Rastrojo, A., Neves, R., Lima, A., Aguado, B. and Iborra, F.J. (2015) 'Global variability in gene expression and alternative splicing is modulated by mitochondrial content', *Genome Res*, 25(5), pp. 633-44.

Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R. and Hannon, G.J. (2001) 'Argonaute2, a link between genetic and biochemical analyses of RNAi', *Science*, 293(5532), pp. 1146-50.

Han, Y.H., Moon, H.J., You, B.R. and Park, W.H. (2009) 'The effect of MG132, a proteasome inhibitor on HeLa cells in relation to cell growth, reactive oxygen species and GSH', *Oncol Rep*, 22(1), pp. 215-21.

Hashimoto, Y., Kondo, C. and Katunuma, N. (2015) 'An Active 32-kDa Cathepsin L Is Secreted Directly from HT 1080 Fibrosarcoma Cells and Not via Lysosomal Exocytosis', *PLoS One*, 10(12), p. e0145067.

Hayashi, K., Yoshida, K. and Matsui, Y. (2005) 'A histone H3 methyltransferase controls epigenetic events required for meiotic prophase', *Nature*, 438(7066), pp. 374-8.

Hayflick, L. (1965) 'The Limited in Vitro Lifetime of Human Diploid Cell Strains', *Exp Cell Res*, 37, pp. 614-36.

Hazkani-Covo, E. and Covo, S. (2008) 'Numt-mediated double-strand break repair mitigates deletions during primate genome evolution', *PLoS Genet*, 4(10), p. e1000237.

He, X.J., Ruan, J., Du, W.D., Cao, Y.X., Chen, G., Zuo, X.B., Peng, Y.W., Wu, H., Song, B. and Zhang, X.J. (2013) 'PRDM9 gene polymorphism may not be associated with defective spermatogenesis in the Chinese Han population', *Syst Biol Reprod Med*, 59(1), pp. 38-41.

Hillen, W. and Berens, C. (1994) 'Mechanisms underlying expression of Tn10 encoded tetracycline resistance', *Annu Rev Microbiol*, 48, pp. 345-69.

Hillman, B.I., Carrington, J.C. and Morris, T.J. (1987) 'A defective interfering RNA that contains a mosaic of a plant virus genome', *Cell*, 51(3), pp. 427-33.

Hinch, A.G., Tandon, A., Patterson, N., Song, Y., Rohland, N., Palmer, C.D., Chen, G.K., Wang, K., Buxbaum, S.G., Akylbekova, E.L., Aldrich, M.C., Ambrosone, C.B., Amos, C., Bandera, E.V., Berndt, S.I., Bernstein, L., Blot, W.J., Bock, C.H., Boerwinkle, E., Cai, Q., Caporaso, N., Casey, G., Cupples, L.A., Deming, S.L., Diver, W.R., Divers, J., Fornage, M., Gillanders, E.M., Glessner, J., Harris, C.C., Hu, J.J., Ingles, S.A., Isaacs, W., John, E.M., Kao, W.H., Keating, B., Kittles, R.A., Kolonel, L.N., Larkin, E., Le Marchand, L., McNeill, L.H., Millikan, R.C., Murphy, A., Musani, S., Neslund-Dudas, C., Nyante, S., Papanicolaou, G.J., Press, M.F., Psaty, B.M., Reiner, A.P., Rich, S.S., Rodriguez-Gil, J.L., Rotter, J.I., Rybicki, B.A., Schwartz,

A.G., Signorello, L.B., Spitz, M., Strom, S.S., Thun, M.J., Tucker, M.A., Wang, Z., Wiencke, J.K., Witte, J.S., Wrensch, M., Wu, X., Yamamura, Y., Zanetti, K.A., Zheng, W., Ziegler, R.G., Zhu, X., Redline, S., Hirschhorn, J.N., Henderson, B.E., Taylor, H.A., Jr., Price, A.L., Hakonarson, H., Chanock, S.J., Haiman, C.A., Wilson, J.G., Reich, D. and Myers, S.R. (2011) 'The landscape of recombination in African Americans', *Nature*, 476(7359), pp. 170-5.

Hinrichs, W., Kisker, C., Duvel, M., Muller, A., Tovar, K., Hillen, W. and Saenger, W. (1994) 'Structure of the Tet repressor-tetracycline complex and regulation of antibiotic resistance', *Science*, 264(5157), pp. 418-20.

Hjerpe, R., Bett, J.S., Keuss, M.J., Solovyova, A., McWilliams, T.G., Johnson, C., Sahu, I., Varghese, J., Wood, N., Wightman, M., Osborne, G., Bates, G.P., Glickman, M.H., Trost, M., Knebel, A., Marchesi, F. and Kurz, T. (2016) 'UBQLN2 Mediates Autophagy-Independent Protein Aggregate Clearance by the Proteasome', *Cell*.

Howe, D.G., Bradford, Y.M., Eagle, A., Fashena, D., Frazer, K., Kalita, P., Mani, P., Martin, R., Moxon, S.T., Paddock, H., Pich, C., Ramachandran, S., Ruzicka, L., Schaper, K., Shao, X., Singer, A., Toro, S., Van Slyke, C. and Westerfield, M. (2017) 'The Zebrafish Model Organism Database: new support for human disease models, mutation details, gene expression phenotypes and searching', *Nucleic Acids Res*, 45(D1), pp. D758-D768.

Howell, N., Herrnstadt, C., Shults, C. and Mackey, D.A. (2003) 'Low penetrance of the 14484 LHON mutation when it arises in a non-haplogroup J mtDNA background', *Am J Med Genet A*, 119A(2), pp. 147-51.

Huang, C.Y., Grunheit, N., Ahmadinejad, N., Timmis, J.N. and Martin, W. (2005) 'Mutational decay and age of chloroplast and mitochondrial genomes transferred recently to angiosperm nuclear chromosomes', *Plant Physiol*, 138(3), pp. 1723-33.

Huang, S., Shao, G. and Liu, L. (1998) 'The PR domain of the Rb-binding zinc finger protein RIZ1 is a protein binding interface and is related to the SET domain functioning in chromatin-mediated gene expression', *J Biol Chem*, 273(26), pp. 15933-9.

Hudson, G., Carelli, V., Spruijt, L., Gerards, M., Mowbray, C., Achilli, A., Pyle, A., Elson, J., Howell, N., La Morgia, C., Valentino, M.L., Huoponen, K., Savontaus, M.L., Nikoskelainen, E., Sadun, A.A., Salomao, S.R., Belfort, R., Jr., Griffiths, P., Yu-Wai-Man, P., de Coo, R.F., Horvath, R., Zeviani, M., Smeets, H.J., Torroni, A. and Chinnery, P.F. (2007) 'Clinical expression of Leber hereditary optic neuropathy is affected by the mitochondrial DNA-haplogroup background', *Am J Hum Genet*, 81(2), pp. 228-33.

Hudson, G., Gomez-Duran, A., Wilson, I.J. and Chinnery, P.F. (2014) 'Recent mitochondrial DNA mutations increase the risk of developing common late-onset human diseases', *PLoS Genet*, 10(5), p. e1004369.

Hunter, S.E., Gustafson, M.A., Margillo, K.M., Lee, S.A., Ryde, I.T. and Meyer, J.N. (2012) 'In vivo repair of alkylating and oxidative DNA damage in the mitochondrial and nuclear genomes of wild-type and glycosylase-deficient Caenorhabditis elegans', *DNA Repair (Amst)*, 11(11), pp. 857-63.

Hussin, J., Sinnett, D., Casals, F., Idaghdour, Y., Bruat, V., Saillour, V., Healy, J., Grenier, J.C., de Malliard, T., Busche, S., Spinella, J.F., Lariviere, M., Gibson, G., Andersson, A., Holmfeldt, L., Ma, J., Wei, L., Zhang, J., Andelfinger, G., Downing, J.R., Mullighan, C.G. and Awadalla, P. (2013) 'Rare allelic forms of PRDM9 associated with childhood leukemogenesis', *Genome Res*, 23(3), pp. 419-30.

Hutvagner, G. and Zamore, P.D. (2002) 'A microRNA in a multiple-turnover RNAi enzyme complex', *Science*, 297(5589), pp. 2056-60.

Ikeda, M., Ide, T., Fujino, T., Arai, S., Saku, K., Kakino, T., Tyynismaa, H., Yamasaki, T., Yamada, K., Kang, D., Suomalainen, A. and Sunagawa, K. (2015) 'Overexpression of TFAM or twinkle increases mtDNA copy number and facilitates cardioprotection associated with limited mitochondrial oxidative stress', *PLoS One*, 10(3), p. e0119687.

Irie, S., Tsujimura, A., Miyagawa, Y., Ueda, T., Matsuoka, Y., Matsui, Y., Okuyama, A., Nishimune, Y. and Tanaka, H. (2009) 'Single-nucleotide polymorphisms of the PRDM9 (MEISETZ) gene in patients with nonobstructive azoospermia', *J Androl*, 30(4), pp. 426-31.

Ishmukhametova, A., Chen, J.M., Bernard, R., de Massy, B., Baudat, F., Boyer, A., Mechin, D., Thorel, D., Chabrol, B., Vincent, M.C., Khau Van Kien, P., Claustres, M. and Tuffery-Giraud, S. (2013) 'Dissecting the structure and mechanism of a complex duplication-triplication rearrangement in the DMD gene', *Hum Mutat*, 34(8), pp. 1080-4.

Jafari, M., Soltani, M., Naahidi, S., Karunaratne, D.N. and Chen, P. (2012) 'Nonviral approach for targeted nucleic acid delivery', *Curr Med Chem*, 19(2), pp. 197-208.

Jeffreys, A.J., Cotton, V.E., Neumann, R. and Lam, K.W. (2013) 'Recombination regulator PRDM9 influences the instability of its own coding sequence in humans', *Proc Natl Acad Sci U S A*, 110(2), pp. 600-5.

Jensen, K.L. and Russell, P. (2016) 'Ctp1-dependent clipping and resection of DNA double-strand breaks by Mre11 endonuclease complex are not genetically separable', *Nucleic Acids Res*.

Jia, L., Li, J., He, B., Jia, Y., Niu, Y., Wang, C. and Zhao, R. (2016) 'Abnormally activated one-carbon metabolic pathway is associated with mtDNA hypermethylation and mitochondrial malfunction in the oocytes of polycystic gilt ovaries', *Sci Rep*, 6, p. 19436.

Johns, D.R., Neufeld, M.J. and Park, R.D. (1992) 'An ND-6 mitochondrial DNA mutation associated with Leber hereditary optic neuropathy', *Biochem Biophys Res Commun*, 187(3), pp. 1551-7.

Ju, Y.S., Alexandrov, L.B., Gerstung, M., Martincorena, I., Nik-Zainal, S., Ramakrishna, M., Davies, H.R., Papaemmanuil, E., Gundem, G., Shlien, A., Bolli, N., Behjati, S., Tarpey, P.S., Nangalia, J., Massie, C.E., Butler, A.P., Teague, J.W., Vassiliou, G.S., Green, A.R., Du, M.Q., Unnikrishnan, A., Pimanda, J.E., Teh, B.T., Munshi, N., Greaves, M., Vyas, P., El-Naggar, A.K., Santarius, T., Collins, V.P., Grundy, R., Taylor, J.A., Hayes, D.N., Malkin, D., Group, I.B.C., Group, I.C.M.D., Group, I.P.C., Foster, C.S., Warren, A.Y., Whitaker, H.C., Brewer, D., Eeles, R., Cooper, C., Neal, D., Visakorpi, T., Isaacs, W.B., Bova, G.S., Flanagan, A.M.,

Futreal, P.A., Lynch, A.G., Chinnery, P.F., McDermott, U., Stratton, M.R. and Campbell, P.J. (2014) 'Origins and functional consequences of somatic mitochondrial DNA mutations in human cancer', *Elife*, 3.

Ju, Y.S., Tubio, J.M., Mifsud, W., Fu, B., Davies, H.R., Ramakrishna, M., Li, Y., Yates, L., Gundem, G., Tarpey, P.S., Behjati, S., Papaemmanuil, E., Martin, S., Fullam, A., Gerstung, M., Group, I.P.C.W., Group, I.B.C.W., Group, I.B.C.W., Nangalia, J., Green, A.R., Caldas, C., Borg, A., Tutt, A., Lee, M.T., van't Veer, L.J., Tan, B.K., Aparicio, S., Span, P.N., Martens, J.W., Knappskog, S., Vincent-Salomon, A., Borresen-Dale, A.L., Eyfjord, J.E., Flanagan, A.M., Foster, C., Neal, D.E., Cooper, C., Eeles, R., Lakhani, S.R., Desmedt, C., Thomas, G., Richardson, A.L., Purdie, C.A., Thompson, A.M., McDermott, U., Yang, F., Nik-Zainal, S., Campbell, P.J. and Stratton, M.R. (2015) 'Frequent somatic transfer of mitochondrial DNA into the nuclear genome of human cancer cells', *Genome Res*, 25(6), pp. 814-24.

Kajander, O.A., Rovio, A.T., Majamaa, K., Poulton, J., Spelbrink, J.N., Holt, I.J., Karhunen, P.J. and Jacobs, H.T. (2000) 'Human mtDNA sublimons resemble rearranged mitochondrial genoms found in pathological states', *Hum Mol Genet*, 9(19), pp. 2821-35.

Kang, D., Miyako, K., Kai, Y., Irie, T. and Takeshige, K. (1997) 'In vivo determination of replication origins of human mitochondrial DNA by ligation-mediated polymerase chain reaction', *J Biol Chem*, 272(24), pp. 15275-9.

Kenney, M.C., Chwa, M., Atilano, S.R., Falatoonzadeh, P., Ramirez, C., Malik, D., Tarek, M., Del Carpio, J.C., Nesburn, A.B., Boyer, D.S., Kuppermann, B.D., Vawter, M.P., Jazwinski, S.M., Miceli, M.V., Wallace, D.C. and Udar, N. (2014) 'Molecular and bioenergetic differences between cells with African versus European inherited mitochondrial DNA haplogroups: implications for population susceptibility to diseases', *Biochim Biophys Acta*, 1842(2), pp. 208-19.

Kim, K.C., Geng, L. and Huang, S. (2003) 'Inactivation of a histone methyltransferase by mutations in human cancers', *Cancer Res*, 63(22), pp. 7619-23.

Kirby, D.M., Thorburn, D.R., Turnbull, D.M. and Taylor, R.W. (2007) 'Biochemical assays of respiratory chain complex activity', *Methods Cell Biol*, 80, pp. 93-119.

Kishita, Y., Pajak, A., Bolar, N.A., Marobbio, C.M., Maffezzini, C., Miniero, D.V., Monne, M., Kohda, M., Stranneheim, H., Murayama, K., Naess, K., Lesko, N., Bruhn, H., Mourier, A., Wibom, R., Nennesmo, I., Jespers, A., Govaert, P., Ohtake, A., Van Laer, L., Loeys, B.L., Freyer, C., Palmieri, F., Wredenberg, A., Okazaki, Y. and Wedell, A. (2015) 'Intra-mitochondrial Methylation Deficiency Due to Mutations in SLC25A26', *Am J Hum Genet*, 97(5), pp. 761-8.

Kloss-Brandstatter, A., Pacher, D., Schonherr, S., Weissensteiner, H., Binna, R., Specht, G. and Kronenberg, F. (2011) 'HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups', *Hum Mutat*, 32(1), pp. 25-32.

Komulainen, T., Hautakangas, M.R., Hinttala, R., Pakanen, S., Vahasarja, V., Lehenkari, P., Olsen, P., Vieira, P., Saarenpaa-Heikkila, O., Palmio, J., Tuominen, H., Kinnunen, P., Majamaa, K., Rantala, H. and Uusimaa, J. (2015) 'Mitochondrial DNA Depletion and

Deletions in Paediatric Patients with Neuromuscular Diseases: Novel Phenotypes', *JIMD Rep*, 23, pp. 91-100.

Kong, Q.P., Yao, Y.G., Sun, C., Bandelt, H.J., Zhu, C.L. and Zhang, Y.P. (2003) 'Phylogeny of east Asian mitochondrial DNA lineages inferred from complete sequences', *Am J Hum Genet*, 73(3), pp. 671-6.

Koopman, W.J., Distelmaier, F., Smeitink, J.A. and Willems, P.H. (2013) 'OXPHOS mutations and neurodegeneration', *EMBO J*, 32(1), pp. 9-29.

Krishnan, K.J. and Birch-Machin, M.A. (2006) 'The incidence of both tandem duplications and the common deletion in mtDNA from three distinct categories of sun-exposed human skin and in prolonged culture of fibroblasts', *J Invest Dermatol*, 126(2), pp. 408-15.

Krishnan, K.J., Reeve, A.K., Samuels, D.C., Chinnery, P.F., Blackwood, J.K., Taylor, R.W., Wanrooij, S., Spelbrink, J.N., Lightowlers, R.N. and Turnbull, D.M. (2008) 'What causes mitochondrial DNA deletions in human cells?', *Nat Genet*, 40(3), pp. 275-9.

Ku, C., Nelson-Sathi, S., Roettger, M., Sousa, F.L., Lockhart, P.J., Bryant, D., Hazkani-Covo, E., McInerney, J.O., Landan, G. and Martin, W.F. (2015) 'Endosymbiotic origin and differential loss of eukaryotic genes', *Nature*, 524(7566), pp. 427-32.

Kuhl, I., Miranda, M., Posse, V., Milenkovic, D., Mourier, A., Siira, S.J., Bonekamp, N.A., Neumann, U., Filipovska, A., Polosa, P.L., Gustafsson, C.M. and Larsson, N.G. (2016) 'POLRMT regulates the switch between replication primer formation and gene expression of mammalian mtDNA', *Sci Adv*, 2(8), p. e1600963.

Kukat, C., Davies, K.M., Wurm, C.A., Spahr, H., Bonekamp, N.A., Kuhl, I., Joos, F., Polosa, P.L., Park, C.B., Posse, V., Falkenberg, M., Jakobs, S., Kuhlbrandt, W. and Larsson, N.G. (2015) 'Cross-strand binding of TFAM to a single mtDNA molecule forms the mitochondrial nucleoid', *Proc Natl Acad Sci U S A*, 112(36), pp. 11288-93.

Laemmli, U.K. (1970) 'Cleavage of structural proteins during the assembly of the head of bacteriophage T4', *Nature*, 227(5259), pp. 680-5.

Lange, J., Yamada, S., Tischfield, S.E., Pan, J., Kim, S., Zhu, X., Socci, N.D., Jasin, M. and Keeney, S. (2016) 'The Landscape of Mouse Meiotic Double-Strand Break Formation, Processing, and Repair', *Cell*, 167(3), pp. 695-708 e16.

Lapuente-Brun, E., Moreno-Loshuertos, R., Acin-Perez, R., Latorre-Pellicer, A., Colas, C., Balsa, E., Perales-Clemente, E., Quiros, P.M., Calvo, E., Rodriguez-Hernandez, M.A., Navas, P., Cruz, R., Carracedo, A., Lopez-Otin, C., Perez-Martos, A., Fernandez-Silva, P., Fernandez-Vizarra, E. and Enriquez, J.A. (2013) 'Supercomplex assembly determines electron flux in the mitochondrial electron transport chain', *Science*, 340(6140), pp. 1567-70.

Larsson, N.G. and Holme, E. (1992) 'Multiple short direct repeats associated with single mtDNA deletions', *Biochim Biophys Acta*, 1139(4), pp. 311-4.

Lesecque, Y., Glemin, S., Lartillot, N., Mouchiroud, D. and Duret, L. (2014) 'The red queen model of recombination hotspots evolution in the light of archaic and modern human genomes', *PLoS Genet*, 10(11), p. e1004790.

Lester, R.L., Hatefi, Y., Widmer, C. and Crane, F.L. (1959) 'Studies on the electron transport system. XX. Chemical and physical properties of the coenzyme Q family of compounds', *Biochim Biophys Acta*, 33(1), pp. 169-85.

Li, M., Schroeder, R., Ko, A. and Stoneking, M. (2012) 'Fidelity of capture-enrichment for mtDNA genome sequencing: influence of NUMTs', *Nucleic Acids Res*, 40(18), p. e137.

Lightowlers, R.N., Taylor, R.W. and Turnbull, D.M. (2015) 'Mutations causing mitochondrial disease: What is new and what challenges remain?', *Science*, 349(6255), pp. 1494-9.

Lin, I.Y., Chiu, F.L., Yeang, C.H., Chen, H.F., Chuang, C.Y., Yang, S.Y., Hou, P.S., Sintupisut, N., Ho, H.N., Kuo, H.C. and Lin, K.I. (2014) 'Suppression of the SOX2 Neural Effector Gene by PRDM1 Promotes Human Germ Cell Fate in Embryonic Stem Cells', *Stem Cell Reports*, 2(2), pp. 189-204.

Lin, Y.F., Schulz, A.M., Pellegrino, M.W., Lu, Y., Shaham, S. and Haynes, C.M. (2016) 'Maintenance and propagation of a deleterious mitochondrial genome by the mitochondrial unfolded protein response', *Nature*, 533(7603), pp. 416-9.

Liu, B., Du, Q., Chen, L., Fu, G., Li, S., Fu, L., Zhang, X., Ma, C. and Bin, C. (2016) 'CpG methylation patterns of human mitochondrial DNA', *Sci Rep*, 6, p. 23421.

Liu, Y., Wu, C., Lyu, Q., Yang, D., Albertini, D.F., Keefe, D.L. and Liu, L. (2007) 'Germline stem cells and neo-oogenesis in the adult human ovary', *Dev Biol*, 306(1), pp. 112-20.

Logan, A., Shabalina, I.G., Prime, T.A., Rogatti, S., Kalinovich, A.V., Hartley, R.C., Budd, R.C., Cannon, B. and Murphy, M.P. (2014) 'In vivo levels of mitochondrial hydrogen peroxide increase with age in mtDNA mutator mice', *Aging Cell*.

Longley, M.J., Ropp, P.A., Lim, S.E. and Copeland, W.C. (1998) 'Characterization of the native and recombinant catalytic subunit of human DNA polymerase gamma: identification of residues critical for exonuclease activity and dideoxynucleotide sensitivity', *Biochemistry*, 37(29), pp. 10529-39.

Lopez-Gallardo, E., Lopez-Perez, M.J., Montoya, J. and Ruiz-Pesini, E. (2009) 'CPEO and KSS differ in the percentage and location of the mtDNA deletion', *Mitochondrion*, 9(5), pp. 314-7.

Luo, S.M., Ge, Z.J., Wang, Z.W., Jiang, Z.Z., Wang, Z.B., Ouyang, Y.C., Hou, Y., Schatten, H. and Sun, Q.Y. (2013) 'Unique insights into maternal mitochondrial inheritance in mice', *Proc Natl Acad Sci U S A*, 110(32), pp. 13038-43.

Ma, H., Marti Gutierrez, N., Morey, R., Van Dyken, C., Kang, E., Hayama, T., Lee, Y., Li, Y., Tippner-Hedges, R., Wolf, D.P., Laurent, L.C. and Mitalipov, S. (2016) 'Incompatibility

between Nuclear and Mitochondrial Genomes Contributes to an Interspecies Reproductive Barrier', *Cell Metab*, 24(2), pp. 283-94.

Mackey, D. and Howell, N. (1992) 'A variant of Leber hereditary optic neuropathy characterized by recovery of vision and by an unusual mitochondrial genetic etiology', *Am J Hum Genet*, 51(6), pp. 1218-28.

Magner, M., Kolarova, H., Honzik, T., Svandova, I. and Zeman, J. (2015) 'Clinical manifestation of mitochondrial diseases', *Dev Period Med*, 19(4), pp. 441-9.

Magnusdottir, E., Dietmann, S., Murakami, K., Gunesdogan, U., Tang, F., Bao, S., Diamanti, E., Lao, K., Gottgens, B. and Azim Surani, M. (2013) 'A tripartite transcription factor network regulates primordial germ cell specification in mice', *Nat Cell Biol*, 15(8), pp. 905-15.

Malik, A.N., Czajka, A. and Cunningham, P. (2016) 'Accurate quantification of mouse mitochondrial DNA without co-amplification of nuclear mitochondrial insertion sequences', *Mitochondrion*, 29, pp. 59-64.

Mancuso, M., Orsucci, D., Angelini, C., Bertini, E., Carelli, V., Comi, G.P., Donati, M.A., Federico, A., Minetti, C., Moggio, M., Mongini, T., Santorelli, F.M., Servidei, S., Tonin, P., Toscano, A., Bruno, C., Bello, L., Caldarazzo Ienco, E., Cardaioli, E., Catteruccia, M., Da Pozzo, P., Filosto, M., Lamperti, C., Moroni, I., Musumeci, O., Pegoraro, E., Ronchi, D., Sauchelli, D., Scarpelli, M., Sciacco, M., Valentino, M.L., Vercelli, L., Zeviani, M. and Siciliano, G. (2015) 'Redefining phenotypes associated with mitochondrial DNA single deletion', *J Neurol*, 262(5), pp. 1301-9.

Marchington, D.R., Hartshorne, G.M., Barlow, D. and Poulton, J. (1997) 'Homopolymeric tract heteroplasmy in mtDNA from tissues and single oocytes: support for a genetic bottleneck', *Am J Hum Genet*, 60(2), pp. 408-16.

McClintock, B. (1939) 'The Behavior in Successive Nuclear Divisions of a Chromosome Broken at Meiosis', *Proc Natl Acad Sci U S A*, 25(8), pp. 405-16.

Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J.C. and Forejt, J. (2009) 'A mouse speciation gene encodes a meiotic histone H3 methyltransferase', *Science*, 323(5912), pp. 373-5.

Mita, S., Rizzuto, R., Moraes, C.T., Shanske, S., Arnaudo, E., Fabrizi, G.M., Koga, Y., DiMauro, S. and Schon, E.A. (1990) 'Recombination via flanking direct repeats is a major cause of large-scale deletions of human mitochondrial DNA', *Nucleic Acids Res*, 18(3), pp. 561-7.

Miyamoto, T., Koh, E., Sakugawa, N., Sato, H., Hayashi, H., Namiki, M. and 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', *J Assist Reprod Genet*, 25(11-12), pp. 553-7.

Moullan, N., Mouchiroud, L., Wang, X., Ryu, D., Williams, E.G., Mottis, A., Jovaisaite, V., Frochaux, M.V., Quiros, P.M., Deplancke, B., Houtkooper, R.H. and Auwerx, J. (2015)

'Tetracyclines Disturb Mitochondrial Function across Eukaryotic Models: A Call for Caution in Biomedical Research', *Cell Rep.*

Muller, H.J. (1964) 'The Relation of Recombination to Mutational Advance', *Mutat Res*, 106, pp. 2-9.

Munoz-Fuentes, V., Di Rienzo, A. and Vila, C. (2011) 'Prdm9, a major determinant of meiotic recombination hotspots, is not functional in dogs and their wild relatives, wolves and coyotes', *PLoS One*, 6(11), p. e25498.

Murakami, K., Gunesdogan, U., Zylicz, J.J., Tang, W.W., Sengupta, R., Kobayashi, T., Kim, S., Butler, R., Dietmann, S. and Surani, M.A. (2016) 'NANOG alone induces germ cells in primed epiblast in vitro by activation of enhancers', *Nature*, 529(7586), pp. 403-7.

Myers, S., Bottolo, L., Freeman, C., McVean, G. and Donnelly, P. (2005) 'A fine-scale map of recombination rates and hotspots across the human genome', *Science*, 310(5746), pp. 321-4.

Myers, S., Bowden, R., Tumian, A., Bontrop, R.E., Freeman, C., MacFie, T.S., McVean, G. and Donnelly, P. (2010) 'Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination', *Science*, 327(5967), pp. 876-9.

Myers, S., Freeman, C., Auton, A., Donnelly, P. and McVean, G. (2008) 'A common sequence motif associated with recombination hot spots and genome instability in humans', *Nat Genet*, 40(9), pp. 1124-9.

Nakatsuji, N. and Chuma, S. (2001) 'Differentiation of mouse primordial germ cells into female or male germ cells', *Int J Dev Biol*, 45(3), pp. 541-8.

Narasimhan, V.M., Hunt, K.A., Mason, D., Baker, C.L., Karczewski, K.J., Barnes, M.R., Barnett, A.H., Bates, C., Bellary, S., Bockett, N.A., Giorda, K., Griffiths, C.J., Hemingway, H., Jia, Z., Kelly, M.A., Khawaja, H.A., Lek, M., McCarthy, S., McEachan, R., O'Donnell-Luria, A., Paigen, K., Parisinos, C.A., Sheridan, E., Southgate, L., Tee, L., Thomas, M., Xue, Y., Schnall-Levin, M., Petkov, P.M., Tyler-Smith, C., Maher, E.R., Trembath, R.C., MacArthur, D.G., Wright, J., Durbin, R. and van Heel, D.A. (2016) 'Health and population effects of rare gene knockouts in adult humans with related parents', *Science*, 352(6284), pp. 474-7.

Nass, M.M. (1973) 'Differential methylation of mitochondrial and nuclear DNA in cultured mouse, hamster and virus-transformed hamster cells. In vivo and in vitro methylation', *J Mol Biol*, 80(1), pp. 155-75.

Ni, K., Dansranjavin, T., Rogenhofer, N., Oeztuerk, N., Deuker, J., Bergmann, M., Schuppe, H.C., Wagenlehner, F., Weidner, W., Steger, K. and Schagdarsurengin, U. (2016) 'TET enzymes are successively expressed during human spermatogenesis and their expression level is pivotal for male fertility', *Hum Reprod*, 31(7), pp. 1411-24.

Nikkanen, J., Forsstrom, S., Euro, L., Paetau, I., Kohnz, R.A., Wang, L., Chilov, D., Viinamaki, J., Roivainen, A., Marjamaki, P., Liljenback, H., Ahola, S., Buzkova, J., Terzioglu, M., Khan, N.A., Pirnes-Karhu, S., Paetau, A., Lonnqvist, T., Sajantila, A., Isohanni, P., Tyynismaa, H., Nomura, D.K., Battersby, B.J., Velagapudi, V., Carroll, C.J. and Suomalainen, A. (2016)

'Mitochondrial DNA Replication Defects Disturb Cellular dNTP Pools and Remodel One-Carbon Metabolism', *Cell Metab*, 23(4), pp. 635-48.

Nunes, M.D., Dolezal, M. and Schlotterer, C. (2013) 'Extensive paternal mtDNA leakage in natural populations of Drosophila melanogaster', *Mol Ecol*, 22(8), pp. 2106-17.

Nunnari, J. and Suomalainen, A. (2012) 'Mitochondria: in sickness and in health', *Cell*, 148(6), pp. 1145-59.

Ohinata, Y., Payer, B., O'Carroll, D., Ancelin, K., Ono, Y., Sano, M., Barton, S.C., Obukhanych, T., Nussenzweig, M., Tarakhovsky, A., Saitou, M. and Surani, M.A. (2005) 'Blimp1 is a critical determinant of the germ cell lineage in mice', *Nature*, 436(7048), pp. 207-13.

Ojala, D., Montoya, J. and Attardi, G. (1981) 'tRNA punctuation model of RNA processing in human mitochondria', *Nature*, 290(5806), pp. 470-4.

Okada, K., Inoue, A., Okada, M., Murata, Y., Kakuta, S., Jigami, T., Kubo, S., Shiraishi, H., Eguchi, K., Motomura, M., Akiyama, T., Iwakura, Y., Higuchi, O. and Yamanashi, Y. (2006) 'The muscle protein Dok-7 is essential for neuromuscular synaptogenesis', *Science*, 312(5781), pp. 1802-5.

Otten, A.B., Theunissen, T.E., Derhaag, J.G., Lambrichs, E.H., Boesten, I.B., Winandy, M., van Montfoort, A.P., Tarbashevich, K., Raz, E., Gerards, M., Vanoevelen, J.M., van den Bosch, B.J., Muller, M. and Smeets, H.J. (2016) 'Differences in Strength and Timing of the mtDNA Bottleneck between Zebrafish Germline and Non-germline Cells', *Cell Rep*, 16(3), pp. 622-30.

P, Y.W.-M., Griffiths, P.G., Brown, D.T., Howell, N., Turnbull, D.M. and Chinnery, P.F. (2003) 'The epidemiology of Leber hereditary optic neuropathy in the North East of England', *Am J Hum Genet*, 72(2), pp. 333-9.

Pagliarini, D.J., Calvo, S.E., Chang, B., Sheth, S.A., Vafai, S.B., Ong, S.E., Walford, G.A., Sugiana, C., Boneh, A., Chen, W.K., Hill, D.E., Vidal, M., Evans, J.G., Thorburn, D.R., Carr, S.A. and Mootha, V.K. (2008) 'A mitochondrial protein compendium elucidates complex I disease biology', *Cell*, 134(1), pp. 112-23.

Papa, S., Martino, P.L., Capitanio, G., Gaballo, A., De Rasmo, D., Signorile, A. and Petruzzella, V. (2012) 'The oxidative phosphorylation system in mammalian mitochondria', *Adv Exp Med Biol*, 942, pp. 3-37.

Parfrey, L.W., Lahr, D.J., Knoll, A.H. and Katz, L.A. (2011) 'Estimating the timing of early eukaryotic diversification with multigene molecular clocks', *Proc Natl Acad Sci U S A*, 108(33), pp. 13624-9.

Park, J., Lai, L., Samuel, M.S., Wax, D., Prather, R.S. and Tian, X. (2015) 'Disruption of Mitochondrion-To-Nucleus Interaction in Deceased Cloned Piglets', *PLoS One*, 10(6), p. e0129378.

Parvanov, E.D., Petkov, P.M. and Paigen, K. (2010) 'Prdm9 controls activation of mammalian recombination hotspots', *Science*, 327(5967), p. 835.

Patel, A., Horton, J.R., Wilson, G.G., Zhang, X. and Cheng, X. (2016) 'Structural basis for human PRDM9 action at recombination hot spots', *Genes Dev*, 30(3), pp. 257-65.

Paulsson, K., An, Q., Moorman, A.V., Parker, H., Molloy, G., Davies, T., Griffiths, M., Ross, F.M., Irving, J., Harrison, C.J., Young, B.D. and Strefford, J.C. (2009) 'Methylation of tumour suppressor gene promoters in the presence and absence of transcriptional silencing in high hyperdiploid acute lymphoblastic leukaemia', *Br J Haematol*, 144(6), pp. 838-47.

Pavlakis, S.G., Phillips, P.C., DiMauro, S., De Vivo, D.C. and Rowland, L.P. (1984) 'Mitochondrial myopathy, encephalopathy, lactic acidosis, and strokelike episodes: a distinctive clinical syndrome', *Ann Neurol*, 16(4), pp. 481-8.

Payne, B.A., Wilson, I.J., Yu-Wai-Man, P., Coxhead, J., Deehan, D., Horvath, R., Taylor, R.W., Samuels, D.C., Santibanez-Koref, M. and Chinnery, P.F. (2013) 'Universal heteroplasmy of human mitochondrial DNA', *Hum Mol Genet*, 22(2), pp. 384-90.

Pedersen, N.M., Thorvaldsen, T.E., Schultz, S.W., Wenzel, E.M. and Stenmark, H. (2016) 'Formation of Tankyrase Inhibitor-Induced Degradasomes Requires Proteasome Activity', *PLoS One*, 11(8), p. e0160507.

PhyloTree *www.phylotree.org*.

Poecheim, J., Barnier-Quer, C., Collin, N. and Borchard, G. (2016) 'Ag85A DNA Vaccine Delivery by Nanoparticles: Influence of the Formulation Characteristics on Immune Responses', *Vaccines (Basel)*, 4(3).

Pohjoismaki, J.L., Holmes, J.B., Wood, S.R., Yang, M.Y., Yasukawa, T., Reyes, A., Bailey, L.J., Cluett, T.J., Goffart, S., Willcox, S., Rigby, R.E., Jackson, A.P., Spelbrink, J.N., Griffith, J.D., Crouch, R.J., Jacobs, H.T. and Holt, I.J. (2010) 'Mammalian mitochondrial DNA replication intermediates are essentially duplex but contain extensive tracts of RNA/DNA hybrid', *J Mol Biol*, 397(5), pp. 1144-55.

Pollack, Y., Kasir, J., Shemer, R., Metzger, S. and Szyf, M. (1984) 'Methylation pattern of mouse mitochondrial DNA', *Nucleic Acids Res*, 12(12), pp. 4811-24.

Poulton, J., Deadman, M.E., Bindoff, L., Morten, K., Land, J. and Brown, G. (1993) 'Families of mtDNA re-arrangements can be detected in patients with mtDNA deletions: duplications may be a transient intermediate form', *Hum Mol Genet*, 2(1), pp. 23-30.

Poulton, J., Deadman, M.E., Ramacharan, S. and Gardiner, R.M. (1991) 'Germ-line deletions of mtDNA in mitochondrial myopathy', *Am J Hum Genet*, 48(4), pp. 649-53.

Powers, N.R., Parvanov, E.D., Baker, C.L., Walker, M., Petkov, P.M. and Paigen, K. (2016) 'The Meiotic Recombination Activator PRDM9 Trimethylates Both H3K36 and H3K4 at Recombination Hotspots In Vivo', *PLoS Genet*, 12(6), p. e1006146.

Prasad, T.K., Gopal, V. and Madhusudhana Rao, N. (2003) 'Structural changes in DNA mediated by cationic lipids alter in vitro transcriptional activity at low charge ratios', *Biochim Biophys Acta*, 1619(1), pp. 59-69.

Pyle, A., Hudson, G., Wilson, I.J., Coxhead, J., Smertenko, T., Herbert, M., Santibanez-Koref, M. and Chinnery, P.F. (2015) 'Extreme-Depth Re-sequencing of Mitochondrial DNA Finds No Evidence of Paternal Transmission in Humans', *PLoS Genet*, 11(5), p. e1005040.

Rajala, N., Hensen, F., Wessels, H.J., Ives, D., Gloerich, J. and Spelbrink, J.N. (2015) 'Whole cell formaldehyde cross-linking simplifies purification of mitochondrial nucleoids and associated proteins involved in mitochondrial gene expression', *PLoS One*, 10(2), p. e0116726.

Rao, D.D. and Huang, A.S. (1979) 'Synthesis of a small RNA in cells coinfected by standard and defective interfering particles of vesicular stomatitis virus', *Proc Natl Acad Sci U S A*, 76(8), pp. 3742-5.

Raymann, K., Brochier-Armanet, C. and Gribaldo, S. (2015) 'The two-domain tree of life is linked to a new root for the Archaea', *Proc Natl Acad Sci U S A*, 112(21), pp. 6670-5.

Reyes, A., Kazak, L., Wood, S.R., Yasukawa, T., Jacobs, H.T. and Holt, I.J. (2013) 'Mitochondrial DNA replication proceeds via a 'bootlace' mechanism involving the incorporation of processed transcripts', *Nucleic Acids Res*, 41(11), pp. 5837-50.

Ricchetti, M., Fairhead, C. and Dujon, B. (1999) 'Mitochondrial DNA repairs double-strand breaks in yeast chromosomes', *Nature*, 402(6757), pp. 96-100.

Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., Villems, R., Thomas, M., Rychkov, S., Rychkov, O., Rychkov, Y., Golge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Cali, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C., Stefanescu, G., Hatina, J., Belledi, M., Di Rienzo, A., Novelletto, A., Oppenheim, A., Norby, S., Al-Zaheri, N., Santachiara-Benerecetti, S., Scozari, R., Torroni, A. and Bandelt, H.J. (2000) 'Tracing European founder lineages in the Near Eastern mtDNA pool', *Am J Hum Genet*, 67(5), pp. 1251-76.

Roy, R., Chun, J. and Powell, S.N. (2012) 'BRCA1 and BRCA2: different roles in a common pathway of genome protection', *Nat Rev Cancer*, 12(1), pp. 68-78.

Sadikovic, B., Wang, J., El-Hattab, A., Landsverk, M., Douglas, G., Brundage, E.K., Craigen, W.J., Schmitt, E.S. and Wong, L.J. (2010) 'Sequence homology at the breakpoint and clinical phenotype of mitochondrial DNA deletion syndromes', *PLoS One*, 5(12), p. e15687.

Sallevelt, S.C., de Die-Smulders, C.E., Hendrickx, A.T., Hellebrekers, D.M., de Coo, I.F., Alston, C.L., Knowles, C., Taylor, R.W., McFarland, R. and Smeets, H.J. (2016) 'De novo mtDNA point mutations are common and have a low recurrence risk', *J Med Genet*.

Samuels, D.C., Schon, E.A. and Chinnery, P.F. (2004) 'Two direct repeats cause most human mtDNA deletions', *Trends Genet*, 20(9), pp. 393-8.

Sarbajna, S., Denniff, M., Jeffreys, A.J., Neumann, R., Soler Artigas, M., Veselis, A. and May, C.A. (2012) 'A major recombination hotspot in the XqYq pseudoautosomal region gives new insight into processing of human gene conversion events', *Hum Mol Genet*, 21(9), pp. 2029-38.

Sasaki, K., Yokobayashi, S., Nakamura, T., Okamoto, I., Yabuta, Y., Kurimoto, K., Ohta, H., Moritoki, Y., Iwatani, C., Tsuchiya, H., Nakamura, S., Sekiguchi, K., Sakuma, T., Yamamoto, T., Mori, T., Woltjen, K., Nakagawa, M., Yamamoto, T., Takahashi, K., Yamanaka, S. and Saitou, M. (2015) 'Robust In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells', *Cell Stem Cell*, 17(2), pp. 178-94.

Sato, M. and Sato, K. (2011) 'Degradation of paternal mitochondria by fertilization-triggered autophagy in C. elegans embryos', *Science*, 334(6059), pp. 1141-4.

Schaefer, A.M., Taylor, R.W., Turnbull, D.M. and Chinnery, P.F. (2004) 'The epidemiology of mitochondrial disorders--past, present and future', *Biochim Biophys Acta*, 1659(2-3), pp. 115-20.

Scheible, M., Just, R., Sturk-Andreaggi, K., Saunier, J., Parson, W., Parsons, T., Coble, M. and Irwin, J. (2016) 'The mitochondrial landscape of African Americans: An examination of more than 2500 control region haplotypes from 22 U.S. locations', *Forensic Sci Int Genet*, 22, pp. 139-48.

Schneider-Poetsch, T., Ju, J., Eyler, D.E., Dang, Y., Bhat, S., Merrick, W.C., Green, R., Shen, B. and Liu, J.O. (2010) 'Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin', *Nat Chem Biol*, 6(3), pp. 209-217.

Schon, E.A., DiMauro, S. and Hirano, M. (2012) 'Human mitochondrial DNA: roles of inherited and somatic mutations', *Nat Rev Genet*, 13(12), pp. 878-90.

Schwartz, M. and Vissing, J. (2002) 'Paternal inheritance of mitochondrial DNA', *N Engl J Med*, 347(8), pp. 576-80.

Seguin, S.J., Morelli, F.F., Vinet, J., Amore, D., De Biasi, S., Poletti, A., Rubinsztein, D.C. and Carra, S. (2014) 'Inhibition of autophagy, lysosome and VCP function impairs stress granule assembly', *Cell Death Differ*, 21(12), pp. 1838-51.

Shimojima, K., Mano, T., Kashiwagi, M., Tanabe, T., Sugawara, M., Okamoto, N., Arai, H. and Yamamoto, T. (2012) 'Pelizaeus-Merzbacher disease caused by a duplication-inverted triplication-duplication in chromosomal segments including the PLP1 region', *Eur J Med Genet*, 55(6-7), pp. 400-3.

Shock, L.S., Thakkar, P.V., Peterson, E.J., Moran, R.G. and Taylor, S.M. (2011) 'DNA methyltransferase 1, cytosine methylation, and cytosine hydroxymethylation in mammalian mitochondria', *Proc Natl Acad Sci U S A*, 108(9), pp. 3630-5.

Shoffner, J.M., Lott, M.T., Lezza, A.M., Seibel, P., Ballinger, S.W. and Wallace, D.C. (1990) 'Myoclonic epilepsy and ragged-red fiber disease (MERRF) is associated with a mitochondrial DNA tRNA(Lys) mutation', *Cell*, 61(6), pp. 931-7.

Shriver, M.D. and Kittles, R.A. (2004) 'Genetic ancestry and the search for personalized genetic histories', *Nat Rev Genet*, 5(8), pp. 611-8.

Silva, M., Alshamali, F., Silva, P., Carrilho, C., Mandlate, F., Jesus Trovoada, M., Cerny, V., Pereira, L. and Soares, P. (2015) '60,000 years of interactions between Central and

Eastern Africa documented by major African mitochondrial haplogroup L2', *Sci Rep*, 5, p. 12526.

Singhal, N.K., Li, S., Arning, E., Alkhayer, K., Clements, R., Sarcyk, Z., Dassanayake, R.S., Brasch, N.E., Freeman, E.J., Bottiglieri, T. and McDonough, J. (2015) 'Changes in Methionine Metabolism and Histone H3 Trimethylation Are Linked to Mitochondrial Defects in Multiple Sclerosis', *J Neurosci*, 35(45), pp. 15170-86.

Smagulova, F., Brick, K., Pu, Y., Camerini-Otero, R.D. and Petukhova, G.V. (2016) 'Erratum: The evolutionary turnover of recombination hot spots contributes to speciation in mice', *Genes Dev*, 30(7), p. 871.

Smeets, H.J. (2013) 'Preventing the transmission of mitochondrial DNA disorders: selecting the good guys or kicking out the bad guys', *Reprod Biomed Online*, 27(6), pp. 599-610.

Song, W., Zhao, W., Yang, Q., Wang, X., Jin, H., Yao, G., Peng, Z., Shi, S., Yang, H. and Sun, Y. (2016a) 'Effect of rapid cryopreservation on meiotic recombination in human spermatocytes', *Microsc Res Tech*.

Song, W.H., Yi, Y.J., Sutovsky, M., Meyers, S. and Sutovsky, P. (2016b) 'Autophagy and ubiquitin-proteasome system contribute to sperm mitophagy after mammalian fertilization', *Proc Natl Acad Sci U S A*.

Soto-Calderon, I.D., Lee, E.J., Jensen-Seaman, M.I. and Anthony, N.M. (2012) 'Factors affecting the relative abundance of nuclear copies of mitochondrial DNA (numts) in hominoids', *J Mol Evol*, 75(3-4), pp. 102-11.

Spang, A., Saw, J.H., Jorgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E., van Eijk, R., Schleper, C., Guy, L. and Ettema, T.J. (2015) 'Complex archaea that bridge the gap between prokaryotes and eukaryotes', *Nature*, 521(7551), pp. 173-9.

Speijer, D., Lukes, J. and Elias, M. (2015) 'Sex is a ubiquitous, ancient, and inherent attribute of eukaryotic life', *Proc Natl Acad Sci U S A*, 112(29), pp. 8827-34.

Spelbrink, J.N., Li, F.Y., Tiranti, V., Nikali, K., Yuan, Q.P., Tariq, M., Wanrooij, S., Garrido, N., Comi, G., Morandi, L., Santoro, L., Toscano, A., Fabrizi, G.M., Somer, H., Croxen, R., Beeson, D., Poulton, J., Suomalainen, A., Jacobs, H.T., Zeviani, M. and Larsson, C. (2001) 'Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localized in mitochondria', *Nat Genet*, 28(3), pp. 223-31.

Starikovskaya, Y.B., Sukernik, R.I., Schurr, T.G., Kogelnik, A.M. and Wallace, D.C. (1998) 'mtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of Ancient Beringia and the peopling of the New World', *Am J Hum Genet*, 63(5), pp. 1473-91.

Steiner, C.C. and Ryder, O.A. (2013) 'Characterization of Prdm9 in equids and sterility in mules', *PLoS One*, 8(4), p. e61746.

Stiles, A.R., Simon, M.T., Stover, A., Eftekharian, S., Khanlou, N., Wang, H.L., Magaki, S., Lee, H., Partynski, K., Dorrani, N., Chang, R., Martinez-Agosto, J.A. and Abdenur, J.E. (2016) 'Mutations in TFAM, encoding mitochondrial transcription factor A, cause neonatal liver failure associated with mtDNA depletion', *Mol Genet Metab*.

Stirling, D. (2003) 'Technical notes for sequencing difficult templates', *Methods Mol Biol*, 226, pp. 401-2.

Stock, D., Leslie, A.G. and Walker, J.E. (1999) 'Molecular architecture of the rotary motor in ATP synthase', *Science*, 286(5445), pp. 1700-5.

Sumegi, B. and Srere, P.A. (1984) 'Complex I binds several mitochondrial NAD-coupled dehydrogenases', *J Biol Chem*, 259(24), pp. 15040-5.

Sun, F., Fujiwara, Y., Reinholdt, L.G., Hu, J., Saxl, R.L., Baker, C.L., Petkov, P.M., Paigen, K. and Handel, M.A. (2015) 'Nuclear localization of PRDM9 and its role in meiotic chromatin modifications and homologous synapsis', *Chromosoma*, 124(3), pp. 397-415.

Suzuki, H., Owada, Y., Watanabe, Y., Inoue, T., Fukuharav, M., Yamaura, T., Mutoh, S., Okabe, N., Yaginuma, H., Hasegawa, T., Yonechi, A., Ohsugi, J., Hoshino, M., Higuchi, M., Shio, Y. and Gotoh, M. (2013) 'Recent advances in immunotherapy for non-small-cell lung cancer', *Hum Vaccin Immunother*, 10(2).

Tang, S., Wang, J., Zhang, V.W., Li, F.Y., Landsverk, M., Cui, H., Truong, C.K., Wang, G., Chen, L.C., Graham, B., Scaglia, F., Schmitt, E.S., Craigen, W.J. and Wong, L.J. (2013) 'Transition to next generation analysis of the whole mitochondrial genome: a summary of molecular defects', *Hum Mutat*, 34(6), pp. 882-93.

Tang, W.W., Dietmann, S., Irie, N., Leitch, H.G., Floros, V.I., Bradshaw, C.R., Hackett, J.A., Chinnery, P.F. and Surani, M.A. (2015) 'A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development', *Cell*, 161(6), pp. 1453-67.

Taylor, R.W. and Turnbull, D.M. (2005) 'Mitochondrial DNA mutations in human disease', *Nat Rev Genet*, 6(5), pp. 389-402.

Thomas, J.H., Emerson, R.O. and Shendure, J. (2009) 'Extraordinary molecular evolution in the PRDM9 fertility gene', *PLoS One*, 4(12), p. e8505.

Thorslund, T., Esashi, F. and West, S.C. (2007) 'Interactions between human BRCA2 protein and the meiosis-specific recombinase DMC1', *EMBO J*, 26(12), pp. 2915-22.

Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus, M.L. and Wallace, D.C. (1996) 'Classification of European mtDNAs from an analysis of three European populations', *Genetics*, 144(4), pp. 1835-50.

Torroni, A., Lott, M.T., Cabell, M.F., Chen, Y.S., Lavergne, L. and Wallace, D.C. (1994) 'mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region', *Am J Hum Genet*, 55(4), pp. 760-76.

Torroni, A., Richards, M., Macaulay, V., Forster, P., Villems, R., Norby, S., Savontaus, M.L., Huoponen, K., Scozzari, R. and Bandelt, H.J. (2000) 'mtDNA haplogroups and frequency patterns in Europe', *Am J Hum Genet*, 66(3), pp. 1173-7.

Torroni, A., Schurr, T.G., Yang, C.C., Szathmary, E.J., Williams, R.C., Schanfield, M.S., Troup, G.A., Knowler, W.C., Lawrence, D.N., Weiss, K.M. and et al. (1992) 'Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations', *Genetics*, 130(1), pp. 153-62.

Torroni, A., Sukernik, R.I., Schurr, T.G., Starikorskaya, Y.B., Cabell, M.F., Crawford, M.H., Comuzzie, A.G. and Wallace, D.C. (1993) 'mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans', *Am J Hum Genet*, 53(3), pp. 591-608.

Tranah, G.J., Santaniello, A., Caillier, S.J., D'Alfonso, S., Martinelli Boneschi, F., Hauser, S.L. and Oksenberg, J.R. (2015) 'Mitochondrial DNA sequence variation in multiple sclerosis', *Neurology*, 85(4), pp. 325-30.

Tsuji, J., Frith, M.C., Tomii, K. and Horton, P. (2012) 'Mammalian NUMT insertion is non-random', *Nucleic Acids Res*, 40(18), pp. 9073-88.

Tyynismaa, H., Sembongi, H., Bokori-Brown, M., Granycome, C., Ashley, N., Poulton, J., Jalanko, A., Spelbrink, J.N., Holt, I.J. and Suomalainen, A. (2004) 'Twinkle helicase is essential for mtDNA maintenance and regulates mtDNA copy number', *Hum Mol Genet*, 13(24), pp. 3219-27.

van der Walt, E.M., Smuts, I., Taylor, R.W., Elson, J.L., Turnbull, D.M., Louw, R. and van der Westhuizen, F.H. (2012) 'Characterization of mtDNA variation in a cohort of South African paediatric patients with mitochondrial disease', *Eur J Hum Genet*, 20(6), pp. 650-6.

Van Goethem, G., Dermaut, B., Lofgren, A., Martin, J.J. and Van Broeckhoven, C. (2001) 'Mutation of POLG is associated with progressive external ophthalmoplegia characterized by mtDNA deletions', *Nat Genet*, 28(3), pp. 211-2.

van Oven, M. and Kayser, M. (2009) 'Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation', *Hum Mutat*, 30(2), pp. E386-94.

Vincent, S.D., Dunn, N.R., Sciammas, R., Shapiro-Shalef, M., Davis, M.M., Calame, K., Bikoff, E.K. and Robertson, E.J. (2005) 'The zinc finger transcriptional repressor Blimp1/Prdm1 is dispensable for early axis formation but is required for specification of primordial germ cells in the mouse', *Development*, 132(6), pp. 1315-25.

Wallace, D.C. (1992) 'Mitochondrial genetics: a paradigm for aging and degenerative diseases?', *Science*, 256(5057), pp. 628-32.

Wallace, D.C., Garrison, K. and Knowler, W.C. (1985) 'Dramatic founder effects in Amerindian mitochondrial DNAs', *Am J Phys Anthropol*, 68(2), pp. 149-55.

Weiss, J., Hurley, L.A., Harris, R.M., Finlayson, C., Tong, M., Fisher, L.A., Moran, J.L., Beier, D.R., Mason, C. and Jameson, J.L. (2012) 'ENU mutagenesis in mice identifies candidate genes for hypogonadism', *Mamm Genome*, 23(5-6), pp. 346-55.

Wessels, H.J., Vogel, R.O., Lightowlers, R.N., Spelbrink, J.N., Rodenburg, R.J., van den Heuvel, L.P., van Gool, A.J., Gloerich, J., Smeitink, J.A. and Nijtmans, L.G. (2013) 'Analysis of 953 human proteins from a mitochondrial HEK293 fraction by complexome profiling', *PLoS One*, 8(7), p. e68340.

West, A.P., Khoury-Hanold, W., Staron, M., Tal, M.C., Pineda, C.M., Lang, S.M., Bestwick, M., Duguay, B.A., Raimundo, N., MacDuff, D.A., Kaech, S.M., Smiley, J.R., Means, R.E., Iwasaki, A. and Shadel, G.S. (2015) 'Mitochondrial DNA stress primes the antiviral innate immune response', *Nature*, 520(7548), pp. 553-7.

Wieckowski, M.R., Giorgi, C., Lebiedzinska, M., Duszynski, J. and Pinton, P. (2009) 'Isolation of mitochondria-associated membranes and mitochondria from animal tissues and cells', *Nat Protoc*, 4(11), pp. 1582-90.

Williams, T.A., Foster, P.G., Cox, C.J. and Embley, T.M. (2013) 'An archaeal origin of eukaryotes supports only two primary domains of life', *Nature*, 504(7479), pp. 231-6.

Wilson, I.J., Carling, P.J., Alston, C.L., Floros, V.I., Pyle, A., Hudson, G., Sallevelt, S.C., Lamperti, C., Carelli, V., Bindoff, L.A., Samuels, D.C., Wonnapinij, P., Zeviani, M., Taylor, R.W., Smeets, H.J., Horvath, R. and Chinnery, P.F. (2016) 'Mitochondrial DNA sequence characteristics modulate the size of the genetic bottleneck', *Hum Mol Genet*, 25(5), pp. 1031-41.

Wockner, L.F., Morris, C.P., Noble, E.P., Lawford, B.R., Whitehall, V.L., Young, R.M. and Voisey, J. (2015) 'Brain-specific epigenetic markers of schizophrenia', *Transl Psychiatry*, 5, p. e680.

Woodward, E.L., Olsson, M.L., Johansson, B. and Paulsson, K. (2014) 'Allelic variants of PRDM9 associated with high hyperdiploid childhood acute lymphoblastic leukaemia', *Br J Haematol*.

Wu, H., Mathioudakis, N., Diagouraga, B., Dong, A., Dombrovski, L., Baudat, F., Cusack, S., de Massy, B. and Kadlec, J. (2013) 'Molecular basis for the regulation of the H3K4 methyltransferase activity of PRDM9', *Cell Rep*, 5(1), pp. 13-20.

Yamaji, M., Seki, Y., Kurimoto, K., Yabuta, Y., Yuasa, M., Shigeta, M., Yamanaka, K., Ohinata, Y. and Saitou, M. (2008) 'Critical function of Prdm14 for the establishment of the germ cell lineage in mice', *Nat Genet*, 40(8), pp. 1016-22.

Yamashita, S., Nishino, I., Nonaka, I. and Goto, Y. (2008) 'Genotype and phenotype analyses in 136 patients with single large-scale mitochondrial DNA deletions', *J Hum Genet*, 53(7), pp. 598-606.

Yang, M.Y., Bowmaker, M., Reyes, A., Vergani, L., Angeli, P., Gringeri, E., Jacobs, H.T. and Holt, I.J. (2002) 'Biased incorporation of ribonucleotides on the mitochondrial L-strand accounts for apparent strand-asymmetric DNA replication', *Cell*, 111(4), pp. 495-505.

Yao, F., Svensjo, T., Winkler, T., Lu, M., Eriksson, C. and Eriksson, E. (1998) 'Tetracycline repressor, tetR, rather than the tetR-mammalian cell transcription factor fusion derivatives, regulates inducible gene expression in mammalian cells', *Hum Gene Ther*, 9(13), pp. 1939-50.

Yu-Wai-Man, P. and Chinnery, P.F. (2012) 'Dysfunctional mitochondrial maintenance: what breaks the circle of life?', *Brain*, 135(Pt 1), pp. 9-11.

Zelazowski, M.J. and Cole, F. (2016) 'X marks the spot: PRDM9 rescues hybrid sterility by finding hidden treasure in the genome', *Nat Struct Mol Biol*, 23(4), pp. 267-9.

Zhang, C., Xue, P., Gao, L., Chen, X., Lin, K., Yang, X., Dai, Y. and Xu, E.Y. (2016) 'Highly conserved epigenetic regulation of BOULE and DAZL is associated with human fertility', *FASEB J*.

Zhang, D., Mott, J.L., Chang, S.W., Denniger, G., Feng, Z. and Zassenhaus, H.P. (2000) 'Construction of transgenic mice with tissue-specific acceleration of mitochondrial DNA mutagenesis', *Genomics*, 69(2), pp. 151-61.

Zheng, W., Khrapko, K., Coller, H.A., Thilly, W.G. and Copeland, W.C. (2006) 'Origins of human mitochondrial point mutations as DNA polymerase gamma-mediated errors', *Mutat Res*, 599(1-2), pp. 11-20.

Zhou, Q., Li, H., Li, H., Nakagawa, A., Lin, J.L., Lee, E.S., Harry, B.L., Skeen-Gaar, R.R., Suehiro, Y., William, D., Mitani, S., Yuan, H.S., Kang, B.H. and Xue, D. (2016) 'Mitochondrial endonuclease G mediates breakdown of paternal mitochondria upon fertilization', *Science*, 353(6297), pp. 394-9.

Ziegler, D.M., Green, D.E. and Doeg, K.A. (1959) 'Studies on the electron transfer system. XXV. The isolation and properties of a lipoflavoprotein with diaphorase activity from beef heart mitochondria', *J Biol Chem*, 234(7), pp. 1916-21.

Zoratti, M., Pietrobon, D. and Azzone, G.F. (1982) 'On the relationship between rate of ATP synthesis and H+ electrochemical gradient in rat-liver mitochondria', *Eur J Biochem*, 126(3), pp. 443-51.

Appendix A: Alignment of mtDNA deletion breakpoint data with PRDM9 motif sites

Seven figures are presented below which show visual representations of alignment between the mtDNA deletion breakpoint data and PRDM9 motif sites as described in Section 4.3.3.
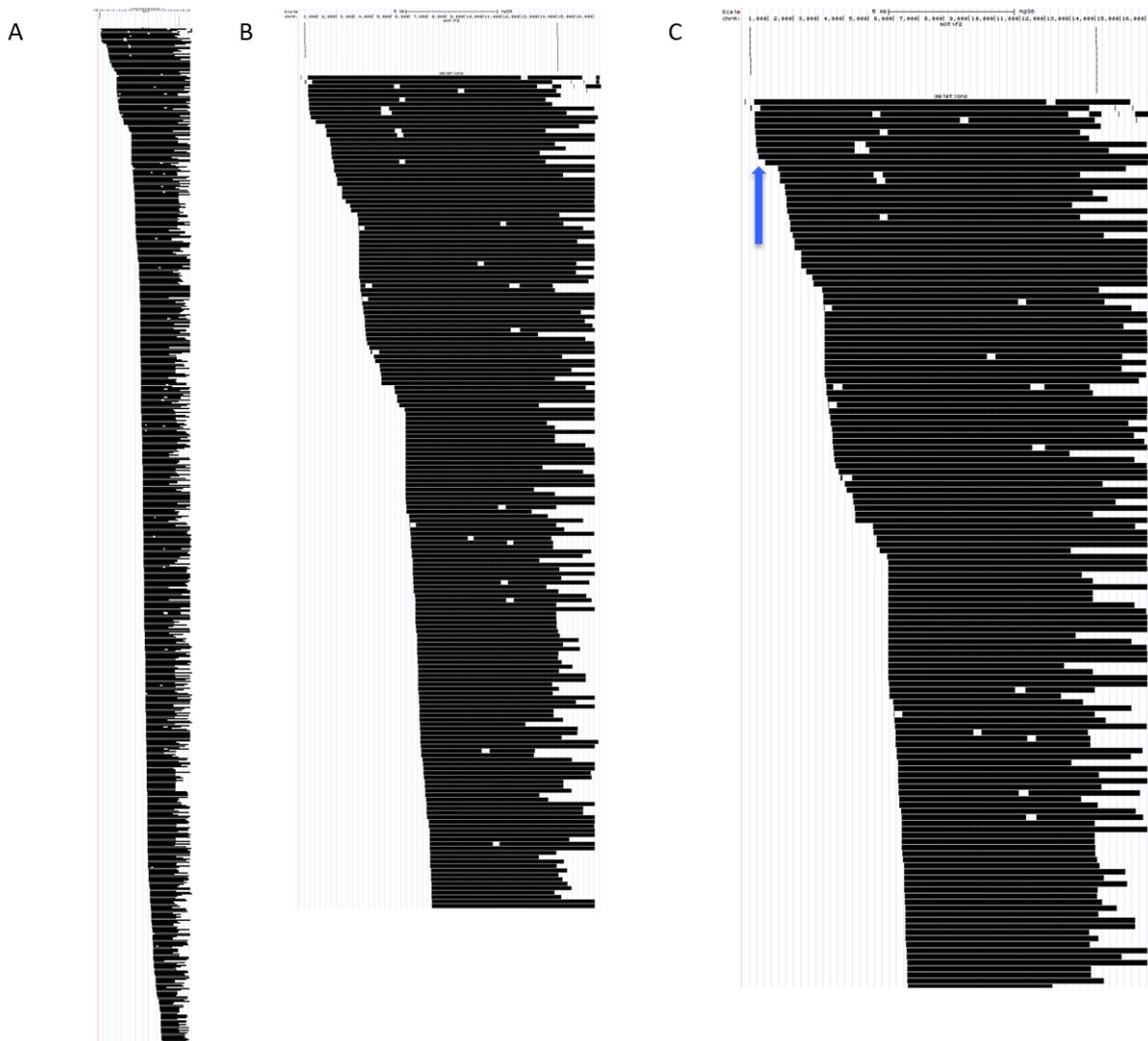
Figure A.1 Visual representation of the alignment between deletion breakpoints and PRDM9 motif 2 sites within the mtDNA. Images were generated using the UCSC genome browser. A) The whole image generated using this online tool. B) Magnified version of image A. C) Magnified version of image C. Blue arrows represent regions where there is visual alignment of the breakpoints and PRDM9 motif 2 sites.
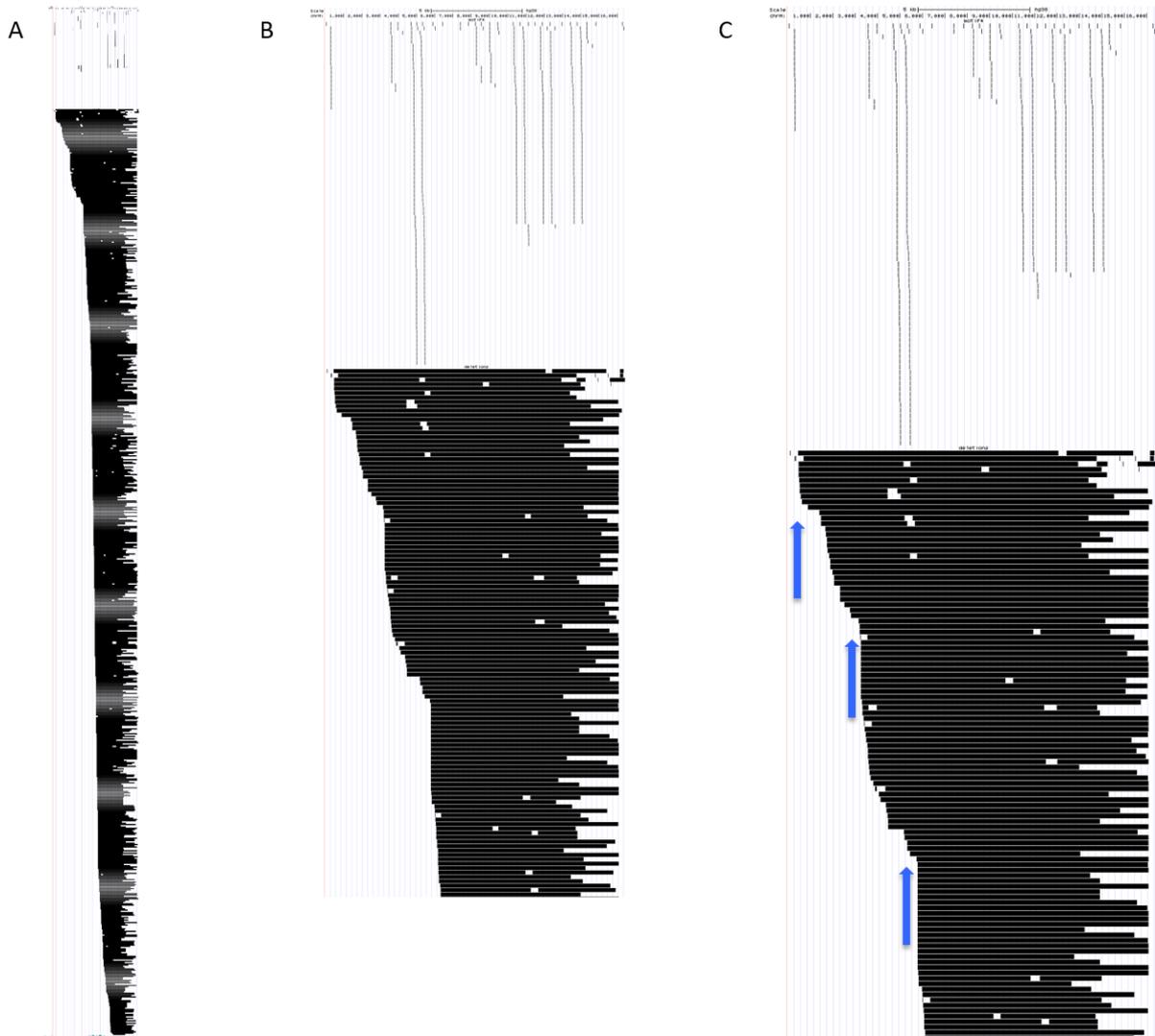
Figure A.2 Visual representation of the alignment between deletion breakpoints and PRDM9 motif 4 sites within the mtDNA. Images were generated using the UCSC genome browser. A) The whole image generated using this online tool. B) Magnified version of image A. C) Magnified version of image C. Blue arrows represent regions where there is visual alignment of the breakpoints and PRDM9 motif 4 sites.
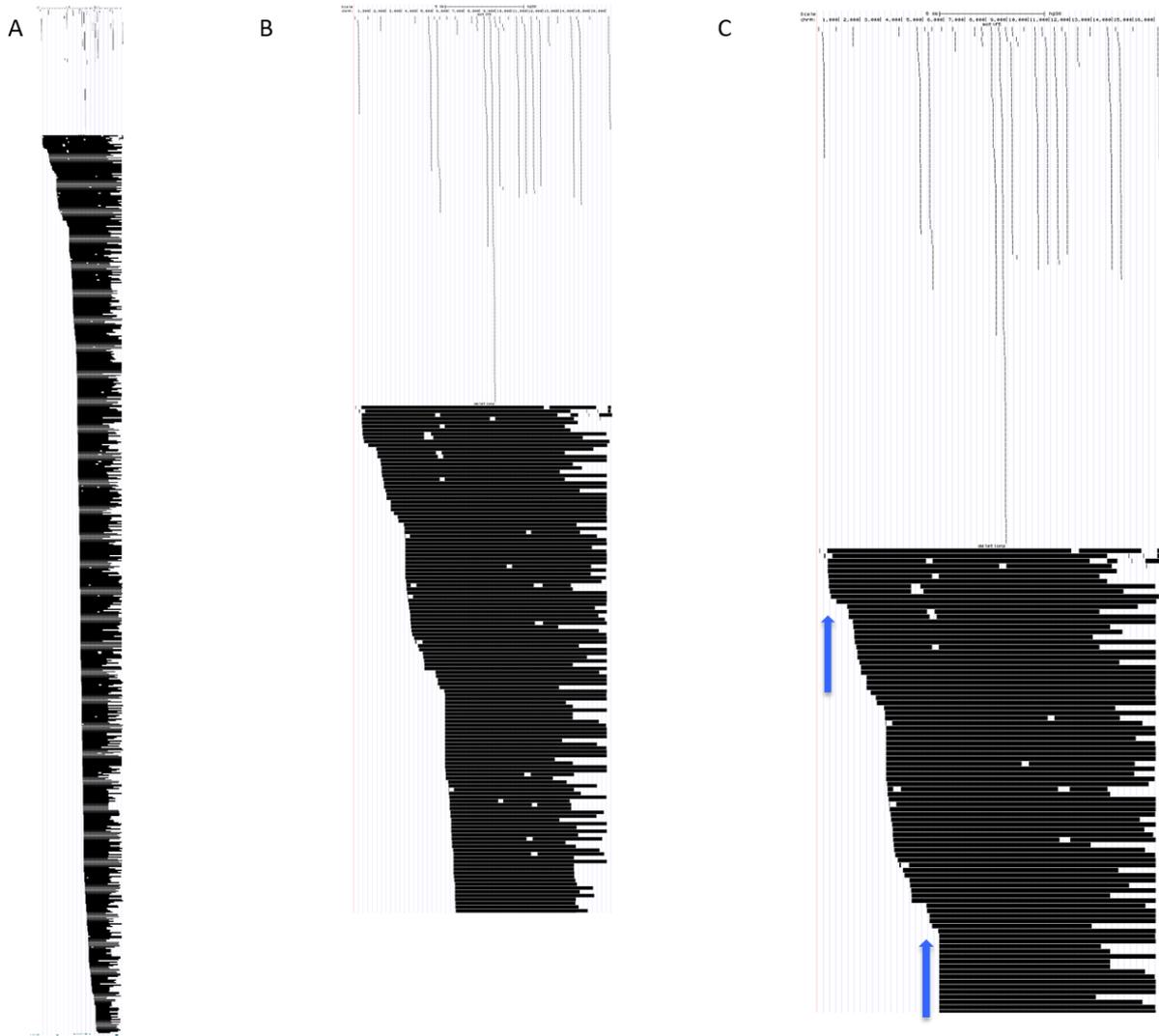
Figure A.3 Visual representation of the alignment between deletion breakpoints and PRDM9 motif 5 sites within the mtDNA. Images were generated using the UCSC genome browser. A) The whole image generated using this online tool. B) Magnified version of image A. C) Magnified version of image C. Blue arrows represent regions where there is visual alignment of the breakpoints and PRDM9 motif 5 sites.
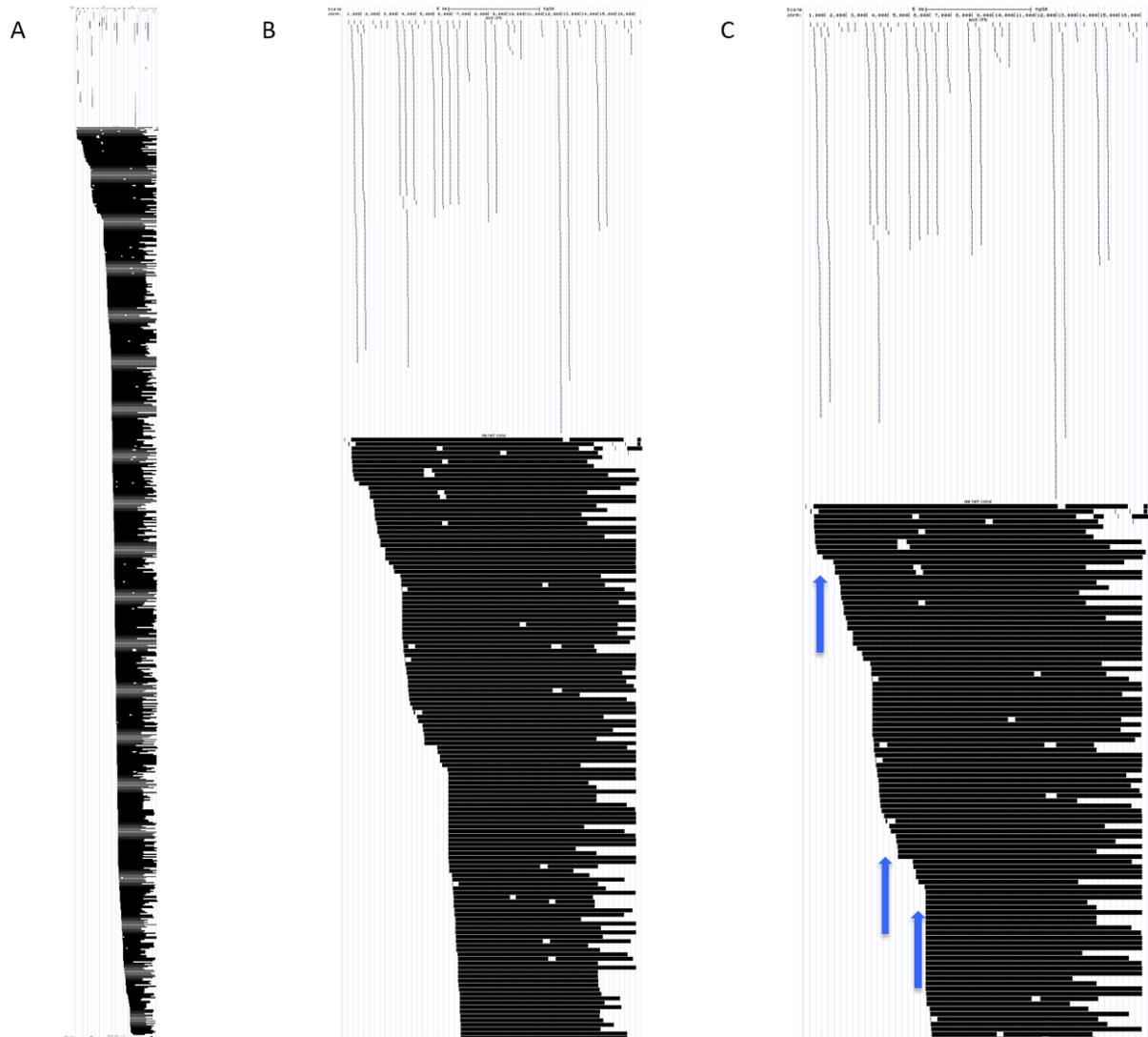
Figure A.4 Visual representation of the alignment between deletion breakpoints and PRDM9 motif 6 sites within the mtDNA. Images were generated using the UCSC genome browser. A) The whole image generated using this online tool. B) Magnified version of image A. C) Magnified version of image C. Blue arrows represent regions where there is visual alignment of the breakpoints and PRDM9 motif 6 sites.
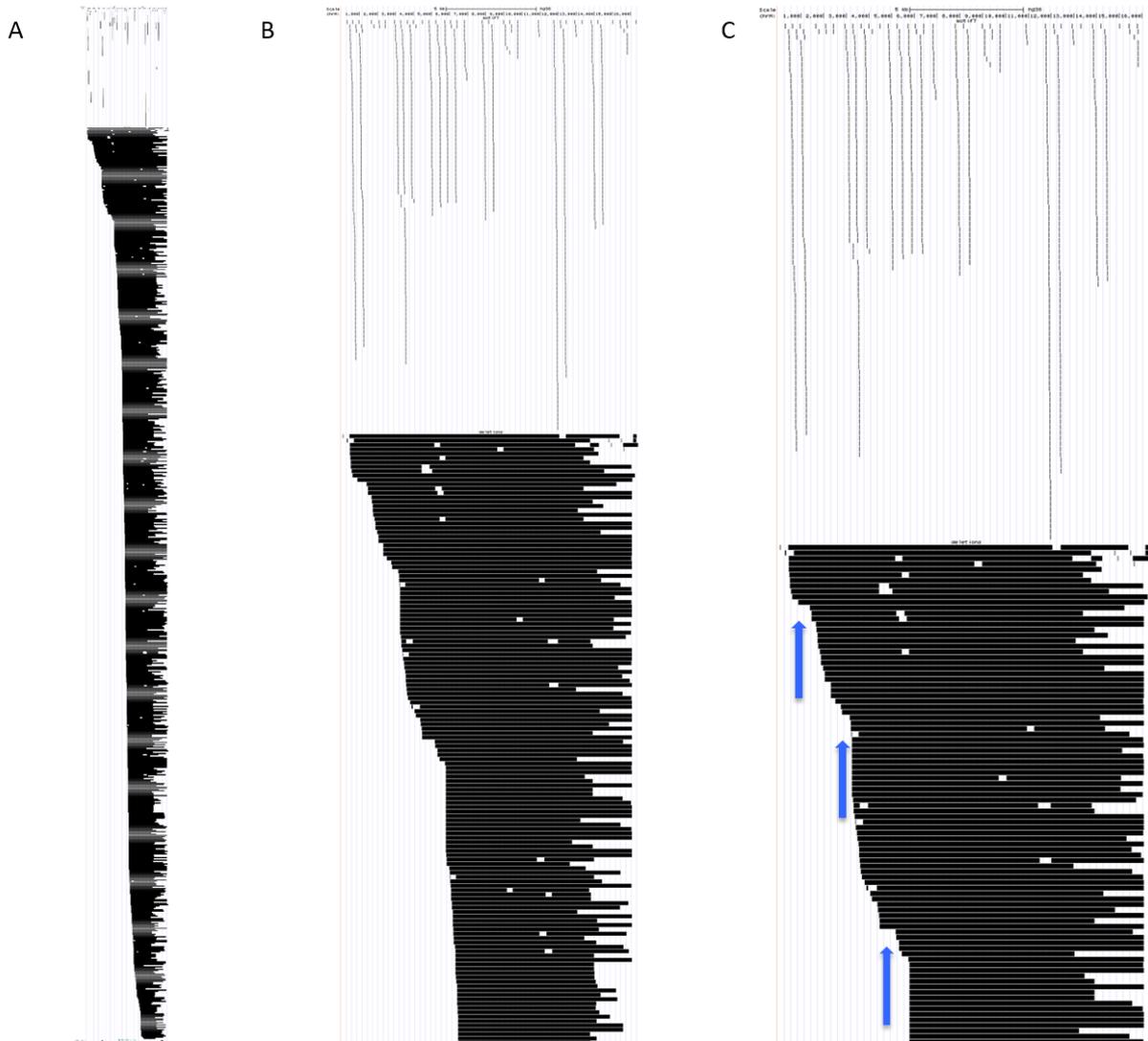
Figure A.5 Visual representation of the alignment between deletion breakpoints and PRDM9 motif 7 sites within the mtDNA. Images were generated using the UCSC genome browser. A) The whole image generated using this online tool. B) Magnified version of image A. C) Magnified version of image C. Blue arrows represent regions where there is visual alignment of the breakpoints and PRDM9 motif 7 sites.
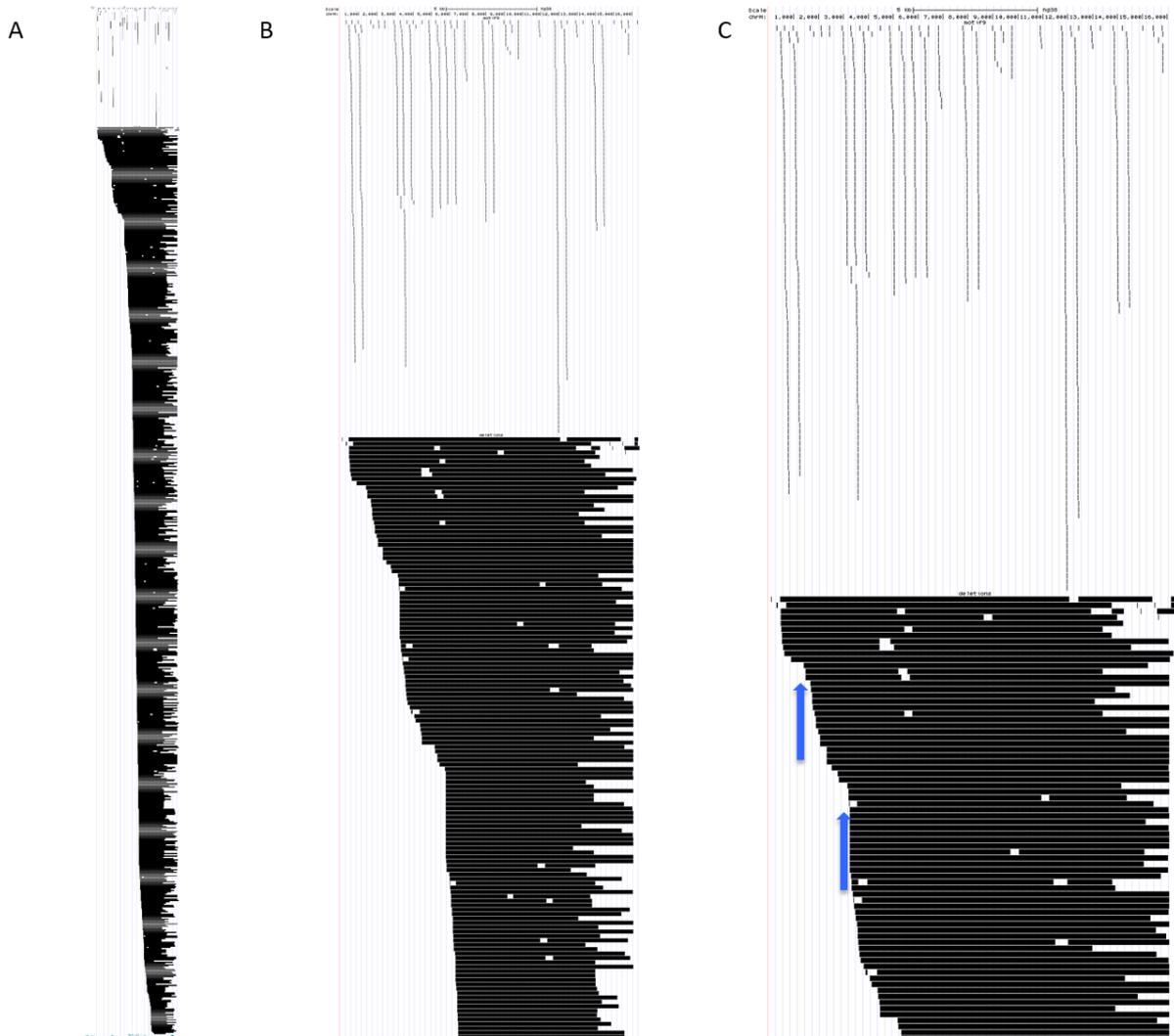
Figure A.6 Visual representation of the alignment between deletion breakpoints and PRDM9 motif 9 sites within the mtDNA. Images were generated using the UCSC genome browser. A) The whole image generated using this online tool. B) Magnified version of image A. C) Magnified version of image C. Blue arrows represent regions where there is visual alignment of the breakpoints and PRDM9 motif 9 sites.
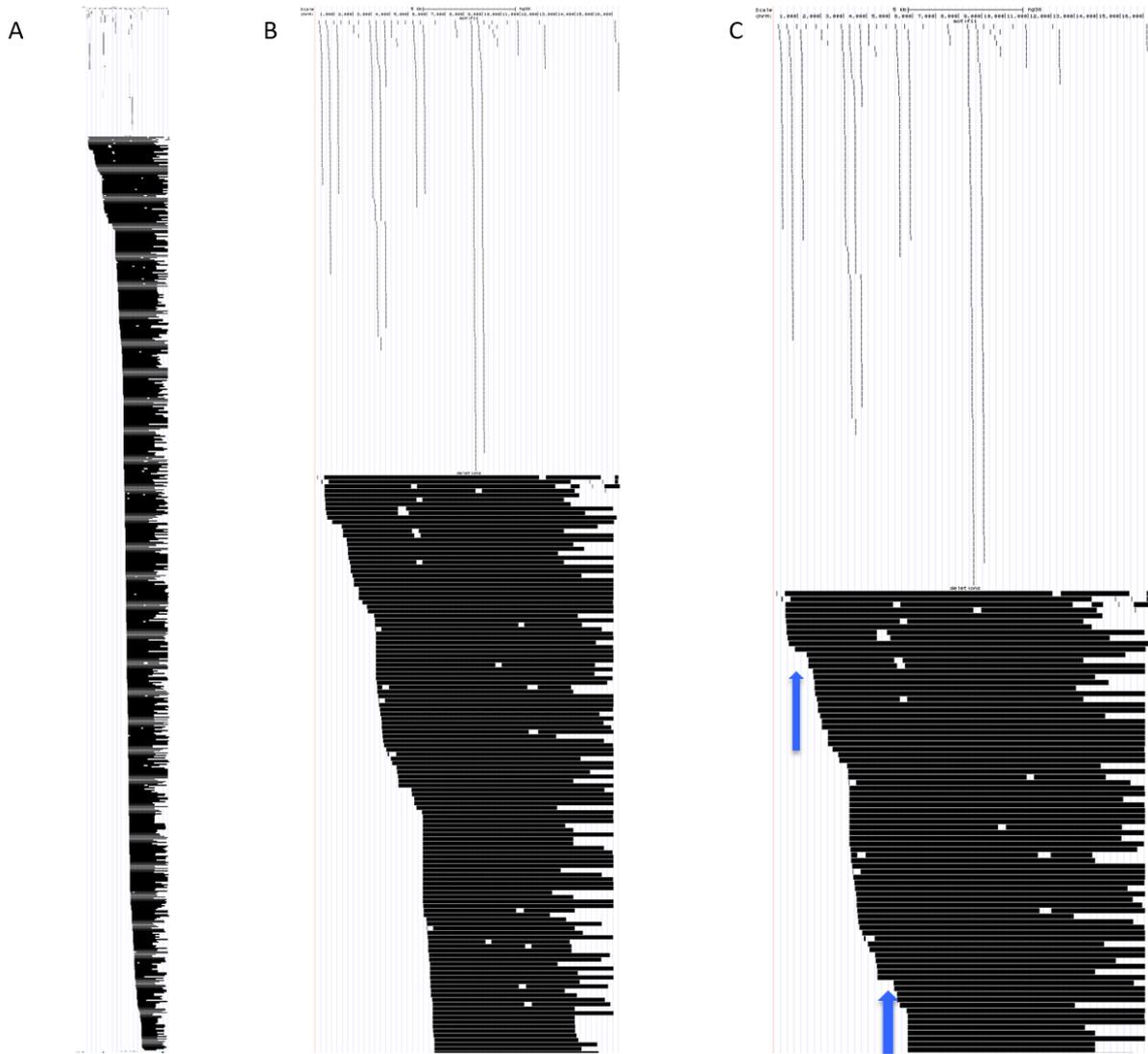
Figure A.7 Visual representation of the alignment between deletion breakpoints and PRDM9 motif 11 sites within the mtDNA. Images were generated using the UCSC genome browser. A) The whole image generated using this online tool. B) Magnified version of image A. C) Magnified version of image C. Blue arrows represent regions where there is visual alignment of the breakpoints and PRDM9 motif 11 sites.

Two tables are presented below showing the genotyping data for the mtDNA single deletion patient cohort (Table B.1) and the control cohort (Table B.2) described in Chapter 5.

					Repeat No.											
Patient ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Allele status
10	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	T	P	F	Q	J			L24
23	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	I	J			A
132	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	I	J			A
143	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	I	J			A
172	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	C	C	F	G	H	F	I	J			B
192	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	G	F	I	J			A
206	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	I	J			A
210	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	I	J			A
241	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	I	J			A
243	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	I	J			A
244	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	C	C	F	K	H	L	H	I	J		C
255	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	I	J			B
289	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	I	J			A
319	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	I	J			A
347	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	K	F	Q	J			L20
352	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	K	G	H	F	I	J		D
377	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	I	J			A

#															Result
379	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
380	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
381	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	T	P	F	Q	J		L24
382	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
387	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
412	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	T	P	F	Q	J		L24
449	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	C	C	F	G	H	F	I	J		B
461	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
496	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	P	F	Q	J		L9
498	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	C	F	K	H	L	H	F	I	J	C
499	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
527	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	H	F	I	J							E
564	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
590	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
612	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	P	F	Q	J		L9
616	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
632	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
702	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
705	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
803	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
899	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
918	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
923	A	B	C	D	D	E	C	F	G	H	F	I	J		A

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Allele status
	A	B	C	D	D	E	C	F	T	P	F	Q	J			L24
940	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	I	J			A
1020	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	I	J			A
1022	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	I	J			A
1051	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	I	J			A
1253	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	I	J			A
1360	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	I	J			A
1627	A	B	C	D	D	D	F	C	F	G	H	F	I	J		C
	A	B	C	D	D	E	C	F	G	H	F	I	J			A
869	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	Q	J			A
103	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	F	J			A

Table B.1 Genotyping analysis of single deletion patient cohort. PRDM9 ZnF repeat types are given in the order they appear on the amplicon and were designated a repeat type according to Berg *et al* 2012.

					Repeat No.											
Control ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Allele status
501																
1	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	I	J			A
2	A	B	C	D	D	E	C	F	T	P	F	Q	J			L24
	A	B	C	D	D	C	C	F	K	H	L	H	I	J		C
3	A	B	C	D	D	E	C	F	G	P	F	Q	J			L9
	A	B	C	D	D	E	C	F	G	H	F	I	J			A
4	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	I	J			A
5	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	I	J			A
6	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	I	J			A
7	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	I	J			A
8	A	B	C	D	D	E	C	F	G	H	F	I	J			A
	A	B	C	D	D	E	C	F	G	H	F	I	J			A

9	A	B	C	D	D	E	C	F	G	H	F	I	J	A
	A	B	C	D	D	E	C	F	G	K	F	Q	J	L20
10	A	B	C	D	D	E	C	F	G	H	F	I	J	A
	A	B	C	D	D	E	C	F	G	H	F	I	J	A
11	A	B	C	D	D	E	C	F	G	H	F	Q	J	A
	A	B	C	D	D	C	C	F	G	H	F	I	J	B
12	A	B	C	D	D	E	C	F	G	H	F	I	J	A
	A	B	C	D	D	E	C	F	G	H	F	I	J	A
13	A	B	C	D	D	E	C	F	G	H	F	I	J	A
	A	B	C	D	D	E	C	F	G	H	F	I	J	A
14	A	B	C	D	D	E	C	F	G	H	F	I	J	A
	A	B	C	D	D	E	C	F	G	H	F	I	J	A
15	A	B	C	D	D	E	C	F	G	H	F	I	J	A
	A	B	C	D	D	E	C	F	G	H	F	I	J	A
16	A	B	C	D	D	E	C	F	G	H	F	I	J	A
	A	B	C	D	D	E	C	F	G	H	F	I	J	A
17	A	B	C	D	D	E	C	F	G	H	F	I	J	A
	A	B	C	D	D	E	C	F	G	H	F	I	J	A
18	A	B	C	D	D	E	C	F	G	H	F	I	J	A
	A	B	C	D	D	E	C	F	G	H	F	I	J	A
19	A	B	C	D	D	E	C	F	G	H	F	I	J	A
	A	B	C	D	D	E	C	F	T	P	F	Q	J	L24
20	A	B	C	D	D	E	C	F	G	H	F	I	J	A
	A	B	C	D	D	E	C	F	G	H	F	I	J	A
21	A	B	C	D	D	E	C	F	G	H	F	I	J	A
	A	B	C	D	D	E	C	F	G	H	F	I	J	A
22	A	B	C	D	D	E	C	F	G	H	F	I	J	A
	A	B	C	D	D	E	C	F	G	H	F	I	J	A
23	A	B	C	D	D	E	C	F	G	H	F	I	J	A
	A	B	C	D	D	E	C	F	T	P	F	Q	J	L24
24	A	B	C	D	D	E	C	F	G	H	F	I	J	A
	A	B	C	D	D	E	C	F	G	H	F	I	J	A
601														
1	A	B	C	D	D	E	C	F	G	H	F	I	J	A
	A	B	C	D	D	C	C	F	G	H	F	I	J	B
2	A	B	C	D	D	E	C	F	G	H	F	I	J	A
	A	B	C	D	D	E	C	F	G	H	F	I	J	A
3	A	B	C	D	D	E	C	F	G	H	F	I	J	A
	A	B	C	D	D	E	C	F	G	H	F	I	J	A
4	A	B	C	D	D	E	C	F	G	H	F	I	J	A
	A	B	C	D	D	E	C	F	G	H	F	I	J	A
5	A	B	C	D	D	E	C	F	G	H	F	I	J	A
	A	B	C	D	D	E	C	F	G	H	F	I	J	A
6	A	B	C	D	D	E	C	F	G	H	F	I	J	A
	A	B	C	D	D	E	C	F	G	H	F	I	J	A

7	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
8	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
9	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
10	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
11	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	Q	J		A
12	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	Q	J		A
13	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	Q	J		A
14	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
15	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	P	F	Q	J		L9
16	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
17	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
18	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
19	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
20	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
21	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	Q	J		A
22	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
23	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A
24	A	B	C	D	D	E	C	F	G	H	F	I	J		A
	A	B	C	D	D	E	C	F	G	H	F	I	J		A

Table B.2 Genotyping analysis of control cohort. PRDM9 ZnF repeat types are given in the order they appear on the amplicon and were designated a repeat type according to Berg *et al* 2012.

Appendix C: Histological staining

To assess the quality of the ovary tissue samples compared to healthy control muscle tissue, hematoxylin and eosin (H & E) staining was performed (Figure C.1). Hematoxylin stains the cell nuclei whilst eosin counterstains, showing tissue morphology. Staining cryosections of the ovary samples described in Chapter 6 show that the gross morphology of the tissue was not as would be expected (Figure C.1 A, B, C & D). The state of the tissue suggests that perhaps multiple freeze/thaw cycles or poor tissue collection resulted in freezing artefacts. In contrast, healthy muscle tissue from a young individual showed expected muscle tissue morphology (Figure C.1 E & F). This tissue biopsy was taken recently and has only been through 2 freeze/thaw cycles when used for cryosectioning. Overall, the morphology observed from the H & E staining suggests that degraded ovary tissue might explain why immunofluorescent stainings described in Chapter 6 had a high level of background immunofluorescent signal.
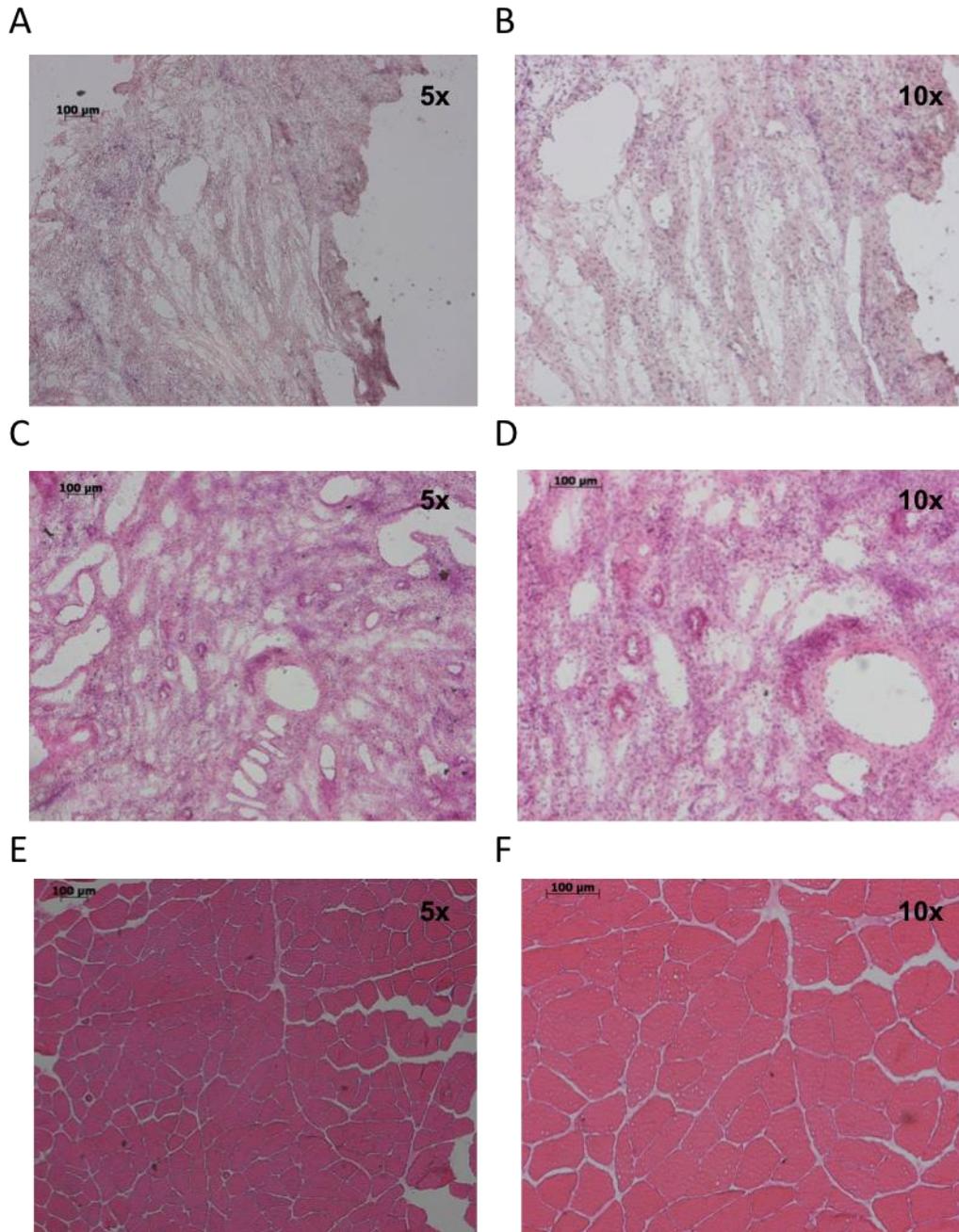
Figure C.9.1 Hematoxylin and eosin staining of tissue cryosections. A & B: section from ovary sample PFC-H-01-07 at 5X magnification (A) and 10X magnification (B). C & D: section from ovary sample PFC-H-02-07 at 5X magnification (C) and 10X magnification (D). E & F: section from muscle sample PFC-H-351-126 at 5X magnification (E) and 10X magnification (F).

Appendix D: Lentivirus overexpression system

PRDM9 overexpression was also attempted by lentiviral plasmid overexpression, described in detail below.

D.1 Lentivirus overexpression system

In order to overexpress PRDM9 in cell lines other than the modified HEK293 line, for example 143B osteosarcoma or HeLa, lentiviral vector overexpression was attempted. Cloning of the full length PRDM9 cDNA construct into the pWPXLd plasmid was attempted using the pWPXLd vector (Figure D.1). Expression using this vector system relies on the retroviral elements derived from HIV-1 which allow stable integration of the gene of interest into the genome of both dividing and non-dividing cells. This system splits retroviral packaging elements over two plasmids, the transfer plasmid containing the gene of interest is shown in Figure D.1.
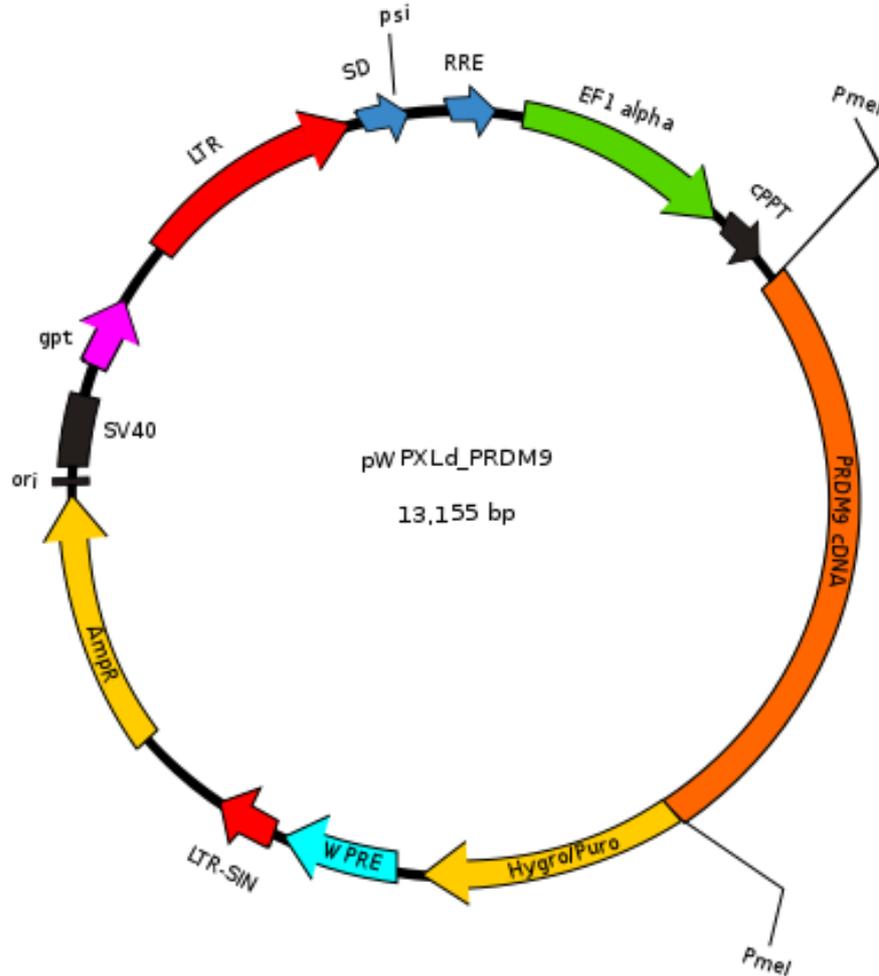
Figure D.9.1 Schematic of pW PXLd plasmid containing full length human *PRDM9* cDNA insert sequence. Plasmid elements shown are: SV40 early polyadenylation signal (black), stem loop 2 HIV1 SD for HIV1 splicing (dark blue), PsiΨ RNA binding site for packaging, Rev Response Elelment (dark blue), EF1α promoter element (green), Central Polypurine Tract for DNA transcription (black arrow), Woodchuck hepatitis virus post-transcriptional regulatory element WPRE (light blue), 5' and 3' Long Terminal Repeats (red), ampicillin/hygromycin/puromycin resistance elements (yellow) and gpt plasmid selective survival element (pink).

The cloning strategy used for this expression system was to digest the Flag-PRDM9 cDNA construct from the pcDNA3.1 plasmid used in previous experiments and sub-clone the product into pWPXLd using restriction enzyme digestion. The pcDNA3.1 Flag-PRDM9 plasmid sequence was checked for all possible restriction enzyme sites using the NEBcutter V2.0 (http://nc2.neb.com/NEBcutter2/). PmeI sites were found to flank either end of the PRDM9 insert sequence (Figure D.2 A) and could also be used to digest the pWPXLd vector. Digestion of pcDNA3 yielded a product at the predicted size of ~2700 bp as shown by agarose gel electrophoresis (Figure D.2 B).

281
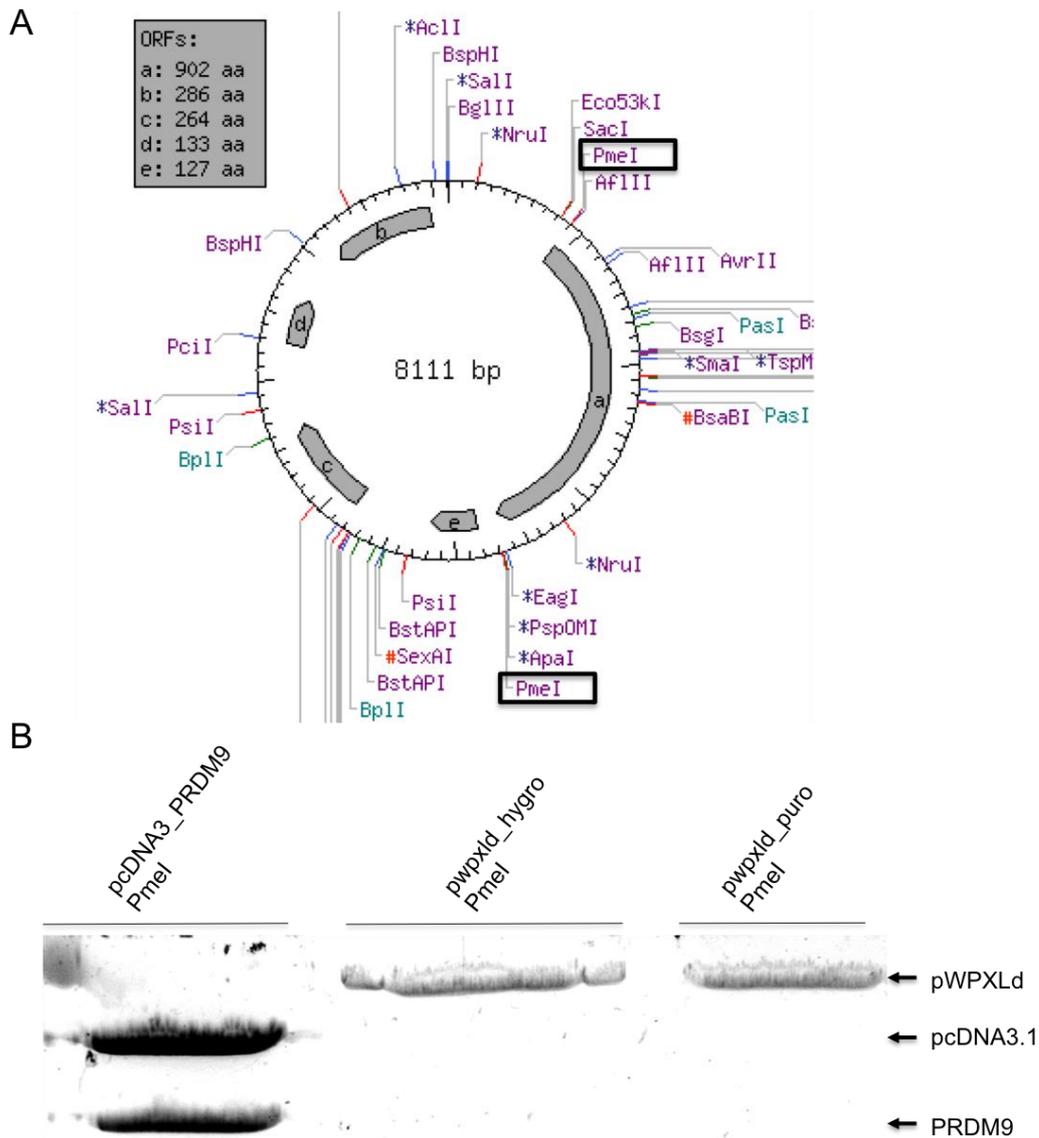
Figure D.1.2 Sub-cloning of Flag-PRDM9 cDNA sequence into pWPXLd vectors. A) Restriction enzyme sites on the pcDNA3.1 Flag-PRDM9, PmeI sites are highlighted in boxes and flank the PRDM9 insert sequence (a). B) Agarose gel size separation of restriction enzyme digest products. pcDNA3.1, Flag-PRDM9 and pWPXLd digest products are indicated by arrows on the right hand side.

Two pWPXLd plasmids were digested, one with hygromycin and the other with puromycin antibiotic resistance genes, both were successfully digested with PmeI in the presence of shrimp alkaline phosphatase (Figure D.2 B). The DNA products were purified from the gel shown in Figure D.2 B and used for ligation reactions.

A vector molar ratio of 1:3 was used for ligation of the Flag-PRDM9 cDNA insert into pWPXLd vectors. Briefly, ligation reactions were set up as follows; 5 µL T4 ligase, 2 µL 2X ligase buffer, 50 ng vector DNA and 66.7 ng insert DNA were mixed by gentle pipetting and incubated at 16 °C for 16 hours. Due to the large plasmid size, several competent cells were used to produce expansion of the ligation product (detailed in Section 3.10) without successful colony growth. Next, a vector to insert ratio of 1:1 was used in ligation reactions, also without successful bacterial colony growth.

Finally, to eliminate any possible negative effects of using only one restriction enzyme in the cloning procedure, such as self-religation, the PRDM9 cDNA sequence was PCR amplified using pcDNA3.1 Flag-PRDM9 as template. Primers for PRDM9 cDNA were designed to introduce a BamHI restriction site in the reverse prime site (Table D.1). This would allow the use of BamHI and PmeI enzymes in the digest reactions and limit self-ligation of the plasmids which is a risk when using only one enzyme.

Primer	Sequence 5'-3'	Annealing Temp. (°C)
Forward PmeI	GCTAGCGTTTAAACTTAAGCTT	72
Reverse BamHI	AGAATTCTGCAGGATCCCAGCACAGTGGCGG	72

Table D.1 Details of primers used to amplify the Flag-PRDM9 insert sequence.

PCR reaction mix was as follows; 1X reaction buffer, 250 µM dNTPs, 0.5 µM primer (forward and reverse), 1 U Phusion DNA polymerase, 10 ng DNA made up to a final volume of 20 µL with dH_2O. Reactions were placed in a thermocycler using the following programme; denaturation at 98 °C for 30 seconds, followed by 30 cycles of denaturation at 98 °C for 10 seconds, annealing at 72 °C for 30 seconds, extension at 72 °C for 90 seconds and final extension at 72 °C for 10 minutes. PCR products were visualised by 1% agarose gel electrophoresis followed by UV exposure. Temperature gradients were performed using two different reaction buffers provided. No amplification was achieved using a gradient PCR from 50-60 °C using either buffer. Amplification did not occur using the High Fidelity PCR buffer at annealing temperatures ranging from 62-72 °C (Figure D.3). Amplification was achieved using the GC Rich Template PCR buffer at all

temperatures but was most specific at 70-72 °C (Figure D.3). A band at 2700 bp corresponds to the length of the PRDM9 insert sequence however bands were also detected at ~2800 and ~125 bp. These are most likely artefacts from the complex repetitive nature of the PRDM9 zinc finger array which could lead to collapsed fragments or mispriming events producing smaller or longer products than the expected 2700 bp amplicon.
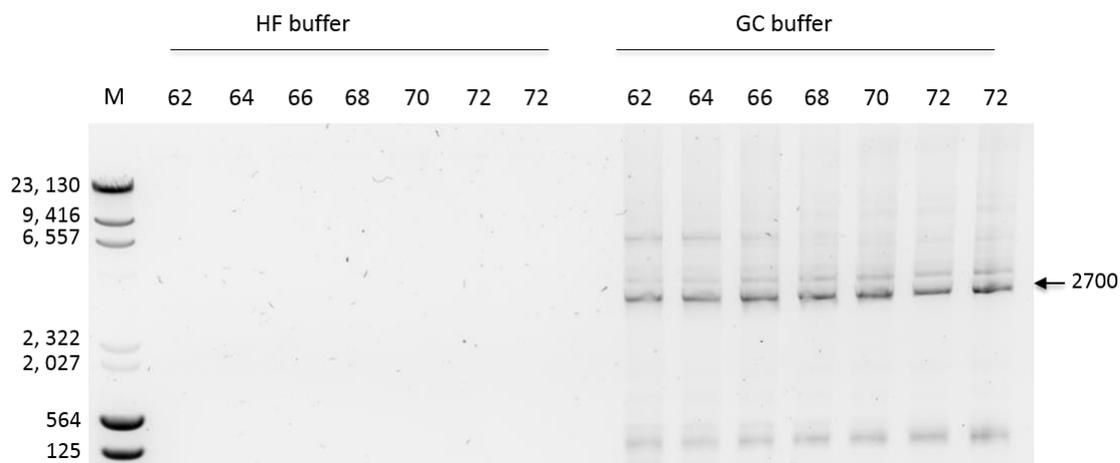

Figure D.1.3 PCR amplification of Flag_PRDM9 cDNA using Phusion High Fidelity Taq polymerase. Temperature gradients were performed using High Fidelity (HF) buffer or GC rich template buffer. Size marker in base pairs is shown in lane M.

After gel extraction of the PCR product at 2700 bp, adenosine nucleotide overhangs were added using the Bioline polymerase kit as follows; 0.2 mM dATP, 1X PCR buffer, 1 mM $MgCl_2$, 1 U DNA Taq polymerase, 25 μL PCR product made up to final volume of 50 μL with dH_20. This reaction mix was incubated at 72 °C for 20 minutes. Next, to ensure that the size of the pWPXLd was the limiting step for successful bacterial transformation, the PCR product was ligated into the much smaller 4 Kb TOPO cloning vector (Thermo Fisher Scientific, Loughborough, UK). Briefly, 4 μL PCR product (with A overhangs), 1 mM NaCl and 50 ng TOPO vector DNA was incubated at room temperature for 5 minutes before transformation of competent cells, as described in Section 3.10.3. Unfortunately only three colonies had grown from the plated out transformed cells. One colony survived through antibiotic selection when grown in antibiotic selection medium and was used for maxi-prep DNA extraction.

Purified TOPO-PRDM9 plasmid was digested using BamHI and PmeI enzymes which had been incorporated into the 3' and 5' ends of the PRDM9 sequence by PCR. Digested DNA product was analysed by gel electrophoresis (Figure D.4). The pWPXLd plasmids were digested and gave a single band at ~11,000 bp as expected when digested with BamHI or PmeI alone and as a double digest. Digestion with PmeI or double enzyme digest showed a single band at ~6,000 bp as expected, however, the TOPO-PRDM9 plasmid gave two small bands when digested with BamHI alone suggesting that there was incorporation of more than one BamHI restriction site.
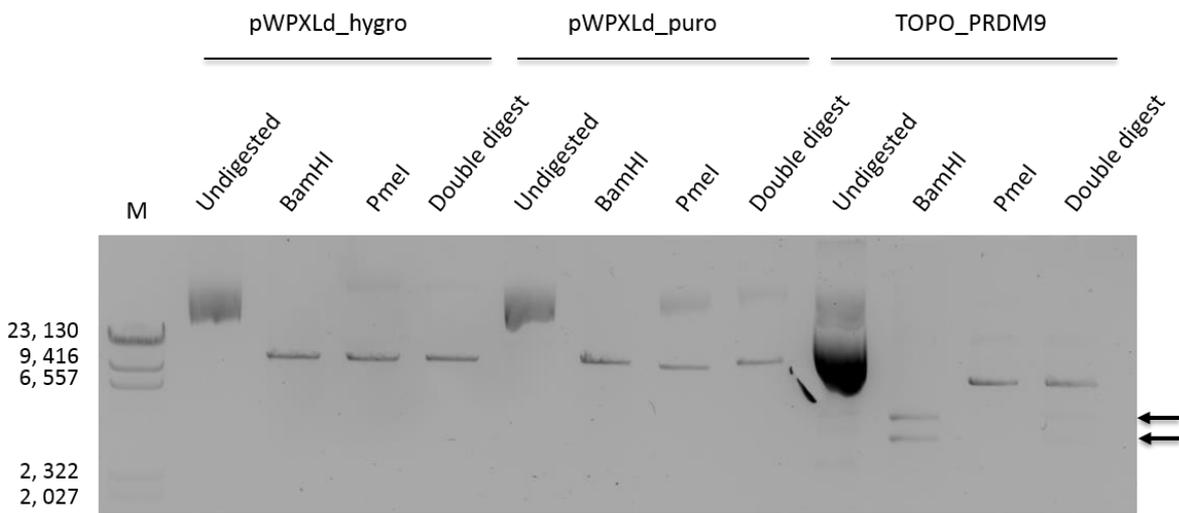

Figure D.1.4 Restriction digest of plasmid DNA purified from competent *E.coli.* 25 ng DNA was electrophoresed on a 1% agarose gel and visualised by UV exposure. DNA size marker was included on the gel (M), sizes are given in base pairs. pWPXLd vectors containing either hygromycin or puromycin resistance genes were digested with BamHI, PmeI or both (Double digest). The TOPO-PRDM9 plasmid was also digested with BamHI, PmeI or both (Double digest).

Sub-cloning the full length PRDM9 cDNA insert into the pWPXLd plasmid was not successful. Sub-cloning full length PRDM9 cDNA into pWPXLd was unsuccessful and some optimisation of the methodology was required. A number of different approaches were tested, such as; lowering incubation temperature (16, 25 and 37 °C), using low-salt growth media and agar plates (0.5%), transformation of several competent cell lines recommended for large plasmid transformations (TOP10, DH5α and Stbl3) as well as trying restriction digest and PCR amplification strategies. Due to the unreliability of this

insert incorporation and the difficulties in obtaining positively transformed *E.coli* colonies this lentivirus strategy was not continued.

Appendix E: Analysis of NuMt sites at recombination hotspots

E.1 NuMt sites in relation to genomic recombination hotspot regions

Genomic recombination 'hotspots' are defined as regions of the genome where chromosome crossovers have occurred, in accordance with genome reshuffling during meiosis. Hotspot location data was downloaded from the HapMap database for each chromosome. Nuclear mitochondrial (NuMt) sequence locations were downloaded from (Li *et al.*, 2012). NuMt sequence locations were incorporated into the hotspot data sets for each chromosome and positions compared. Proximity of NuMts to hotspots was calculated by subtracting the start positions from each other and alignments were verified by eye. In total, 36 NuMts were identified within 20 hotspot regions (Table E.1). These regions were present on 13 chromosomes with chromosome 2 containing the highest number (4 different hotspot regions). Interestingly, single hotspots containing multiple NuMts were found on chromosomes 11 and 14 (Table E.1).

Chromosome	Start	End	Type	Length	Distance from HS start
1	147327001	147350001	HS	23000	
1	147332804	147332915	NUMT	111	5803
1	205442001	205445001	HS	3000	
1	205444544	205444632	NUMT	88	2543
2	68485001	68491001	HS	6000	
2	68487872	68488036	NUMT	164	2871
2	85294001	85301001	HS	7000	
2	85295952	85296153	NUMT	201	1951
2	133179001	133183001	HS	4000	
2	133182673	133182730	NUMT	57	3672
2	180604001	180608001	HS	4000	
2	180604074	180604289	NUMT	215	73
2	180604379	180604467	NUMT	88	378
3	153370001	153378001	HS	8000	
3	153376674	153376776	NUMT	102	6673
3	169653001	169656001	HS	3000	
3	169654585	169654656	NUMT	71	1584
4*	14508001	14513001	HS	5000	
4*	14507530	14507742	NUMT	212	-471
4*	14507834	14508073	NUMT	239	-167
4	182155001	182159001	HS	4000	
4	182158556	182158693	NUMT	137	3555
5	93895001	93920001	HS	25000	
5	93903161	93906623	NUMT	3462	8160
6	133470001	133474001	HS	4000	
6	133471710	133471933	NUMT	223	1709
9	85041001	85047001	HS	6000	
9	85042306	85042806	NUMT	500	4695
9	85042830	85042944	NUMT	114	4171
11	103271001	103279001	HS	8000	
11	103272857	103273350	NUMT	493	6144
11	103274883	103275102	NUMT	219	4118
11	103275372	103276049	NUMT	677	3629
11	103276576	103276694	NUMT	118	2425
11	103276718	103276943	NUMT	225	2283
11	103277402	103277483	NUMT	81	1599
11	103277548	103277680	NUMT	132	1453
11	103277962	103278170	NUMT	208	1039
11	103278536	103278708	NUMT	172	465
11	103278904	103279035	NUMT	131	97
14	84634000	84640000	HS	6000	
14	84637696	84638079	NUMT	383	2304
14	84638350	84638625	NUMT	275	1650
14	84638769	84639028	NUMT	259	1231
14	84639090	84639184	NUMT	94	910
14	84639187	84639382	NUMT	195	813
15	34685000	34689000	HS	4000	
15	34686922	34687072	NUMT	150	1922
15	34833000	34837000	HS	4000	
15	34833143	34833293	NUMT	150	143
18	59539001	59542001	HS	3000	
18	59541804	59542118	NUMT	314	2803
21	46786000	46797000	HS	11000	
21	46796121	46796299	NUMT	178	10121
22	46863000	46867000	HS	4000	
22	46866181	46866230	NUMT	49	3181

Table E.1 Alignment data of genomic hotspots and NuMt positions. HS denotes 'hotspot'; NUMT denotes 'nuclear mitochondrial'. Chromosome location and start and end positions are provided. Length of HS or NuMt is shown along with the distance of the NuMt sequence from the start of the HS region.

In contrast, a hotspot on chromosome 4 was identified as being downstream of two NuMt regions, one overlapping the beginning of the hotspot (Table E.1).

The HapMap Phase III data consists of genome data from 1301 individuals from 11 population groups. The Phase III build shows the most common 'events' based on the combined data to give an overall view of general human genome patterns. This makes the hotspot data unbiased as it is not population specific. The NuMt data used in this study was generated by Li *et al* using comparisons between the mitochondrial revised Cambridge Reference Sequence (rCRS) and the HG19 genome build. This genome is a composite of sequences from multiple individuals. The group also analysed data from artificial mixes of mtDNA sequences which differed at 34 positions along the molecule. These methods ensure that the NuMt data provided is reliable and not population biased. Although these data sets have been compiled with efforts to minimise confounding factors such as population genetics, there will undoubtedly be differences between individuals and this analysis is perhaps not applicable to all genomes. However, it provides an overview of the alignment between NuMt sequences and recombination hotspots in a model of a generic human genome which represents a large proportion of the population.

Appendix F: Attendance at scientific meetings

The content of this thesis has been presented in the form of a poster presentation at three scientific meetings in 2016:

1. Mitochondrial Medicine: Developing New Treatments for Mitochondrial Disease, May 2016, Cambridge, UK.

2. United Mitochondrial Disease Foundation: Mitochondrial Medicine 2016, June 2016, Seattle, USA.

3. Newcastle University Institute of Ageing: Postgraduate Research Day 2016, June 2016, Newcastle, UK.

F1 Abstract

The potential role of recombination regulator PRDM9 in mitochondria.

Emily McILwaine[1], Aurora Gomez-Duran[1,2] Helen Griffin[1], Gavin Hudson[1], Rita Horvath[1] & Patrick F. Chinnery[1,2]

[1]Wellcome Trust Centre for Mitochondrial Research, Institute of Genetic Medicine, Newcastle University, UK [2]Medical Research Council Mitochondrial Biology Unit, Cambridge, UK

email: e.mcilwaine@ncl.ac.uk

At present 805 mitochondrial DNA (mtDNA) deletions have been described. Short direct repeat regions of DNA flank many of these deletions, suggesting that specific regions of the mtDNA molecule have a susceptibility to deletion formation. PRDM9 is a meiotic-specific protein responsible for determining where recombination hotspots will occur in the nuclear genome. It binds a specific DNA consensus sequence through its zinc finger repeat region.

The aim of this project is to determine whether the PRDM9 binding motif is present in the mtDNA molecule and investigate whether PRDM9 plays a role in mtDNA deletion formation.

A bioinformatic approach was used to screen mtDNA sequences from 31,000 individuals for the presence of the PRDM9 binding motif. The motif was found in 99.9% of the

sequences searched at a position known to flank a previously described deletion (m.5327). Several sequences contained the motif more than once with some sequences containing up to 3 motif sites in total. Classifying the sequences by broad haplogrouping (African, Asian and European) showed population differences with respect to the number of motifs present. Further analysis of the European sequences showed differences between sub-haplogroup status and the number of motifs present

Genotyping via Sanger sequencing of the *PRDM9* zinc finger repeat region in a cohort of 48 single deletion patients and 50 healthy controls showed no correlation between *PRDM9* allele status and the presence of the 'common' 4977bp mtDNA deletion.

Subcellular fractions of HEK-293T cells followed by SDS-PAGE and immunoblotting with anti-PRDM9 antibody showed PRDM9 protein is enriched in mitochondria. However, the antibody gave unreliable multiple banding on Western blot membranes. To address this, stable cell lines overexpressing FLAG-tagged PRDM9 under control of the tetracycline inducible promoter were created. PRDM9 expression was detected by immunoblotting. PRDM9 was previously reported to directly methylate the lysine 4 residue on histone 3 and measurement of H3K4me3 protein levels in these cell lines showed a significant increase.

The PRDM9 recognition motif previously described is present in human mtDNA. PRDM9 protein appears to be enriched in the mitochondrial fraction of HEK-293 cells suggesting that PRDM9 may play a role in mediating mtDNA maintenance during development.

F2 Poster

The potential role of recombination regulator PRDM9 in mitochondria

Emily McIlwaine[1], Aurora Gomez-Duran[1,2], Gavin Hudson[1], Rita Horvath[1] & Patrick F. Chinnery[1,2]

[1]Wellcome Trust Centre for Mitochondrial Research, Newcastle University, UK   [2]Medical Research Council, Mitochondrial Biology Unit, Cambridge University, UK

email: e.mcilwaine@ncl.ac.uk

INTRODUCTION

Mitochondrial DNA deletions form the basis of several mitochondrial disorders in humans. How these deletions are formed is unclear and is speculated to be during mtDNA replication and maintenance. PRDM9 is a nuclear encoded protein involved in meiotic recombination . It has a zinc finger repeat region which recognises a 13bp repeat motif near recombination hotspots. This motif is also present in mtDNA at a known site of deletion formation.

Figure 1 mtDNA deletions result in the formation of smaller mtDNA molecules within the cell leading to a phenomenon known as heteroplasmy. Females with oocytes heteroplasmic for mtDNA deletions can have offspring carrying different deletion burden and can be unaffected or affected to varying degrees.

RESULTS

PRDM9 binding motif is differentially represented between populations
n = 30,531
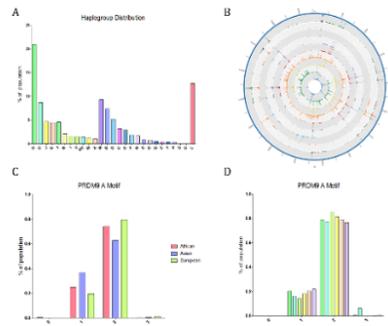
Figure 2 Bioinformatic search for PRDM9 allele A recognition motif in 31,531 mtDNA sequences. A) Distribution of haplogroups present in the cohort analysed. B) Circos plot of the frequency of PRDM9 motifs identified and their position on the mtDNA molecule. C) Distribution of the number of motifs present in mtDNA sequences from each population (African, Asian & European). D) Distribution of the number of motifs present in each European haplogroup.
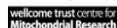
PRDM9 alleles are not associated with mtDNA single deletions
n = 98
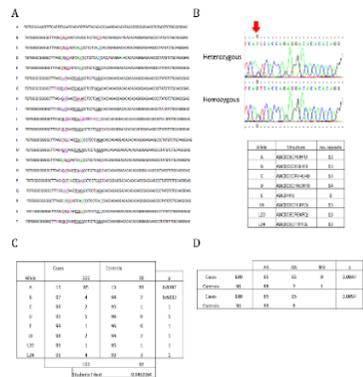
Figure 3 PRDM9 allele status in single deletion patients. A) PRDM9 zinc finger repeat sequences found in the human population. SNVs are highlighted in colour.B) Representative Sanger sequencing traces in a heterozygous individual (top) and a homozygous individual (bottom) from the cohort analysed. C) Chi-squared results of the patient cohort (n=48) compared to the control cohort (n=50).

### PRDM9 localises to the mitochondria in HEK293 cells
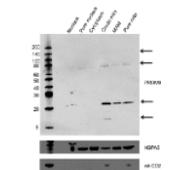
Figure 4 Immunodetection of PRDM9 in HEK-293 cell lysate. Nuclear, cytoplasmic, mitochondrial and mitochondrial associated membrane fractions were blotted. HSPA5 and mt-CO2 antibodies were used to assess the purity of the fractions.

PRDM9 overexpression

Figure 5 Validation of PRDM9 TRex293 expression cell lines. A)PCR amplification of the PRDM9 insert sequence in stably transfected TRex293 cells. Western blot of PRDM9 (B) and FLAG_PRDM9 (C) in cell lysate from stably transfected TRex293 cell lines.

### Overexpressed PRDM9 functions as a methyl transferase

Figure 6 Detection of histone H3 lysine 4 methylation status. A) Western blot of H3K4me1, me2 and me3 in HEK293 control, PRDM9_FLAG and PRDM9 overexpression lines. Treated and non-treated cells were harvested at 24, 48 and 72h. B) Densitometric quantification of H3K4me3 levels relative to β-actin (n=4) for each cell line; HEK293 (left), FLAG PRDM9 (middle) and PRDM9 (right). Error bars show SD values.

Mitochondrial DNA and protein levels in PRDM9 overexpression lines
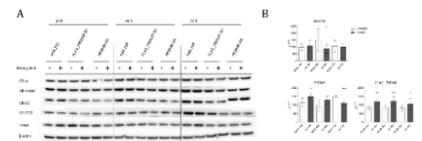

Figure 7 Mitochondrial protein levels and mtDNA copy number quantification. Treated and non-treated cells were harvested at 24, 48 and 72h. A) Western blotting of several known mitochondrial proteins showed no differences in protein level in any of the cell lines, HEK293 control, FLAG PRDM9 or PRDM9 overexpression. B) RT-qPCR was performed using Taqman chemistry. RNase P copy number was used as an estimation of genomic DNA, mtCOIII was used to quantify the number of mtDNA molecules in each sample. Error bars show SD values.

CONCLUSIONS

❖ PRDM9 motifs are present in the mtDNA sequences in 99% of individuals

❖ PRDM9 allele status is not associated with the presence of mtDNA single deletions PRDM9 localises to mitochondrial fractions in HEK293 cells

❖ Overexpression of PRDM9 shows that the protein functions as a methyl transferase

❖ Mitochondrial protein levels are not affected by PRDM9 overexpression but mtDNA copy number is increased